From Poisson to Silicon - Advancing Compact SPICE Models for IC Design

Sriramkumar Venugopalan

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2013-166
http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-166.html

October 8, 2013

From Poisson to Silicon - Advancing Compact SPICE Models for IC Design

by

Sriramkumar Venugopalan

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

in

Engineering – Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Chenming Hu, Chair
Professor Ali M. Niknejad
Professor Paul K. Wright

Fall 2013

From Poisson to Silicon - Advancing Compact SPICE Models for IC Design

Abstract

From Poisson to Silicon - Advancing Compact SPICE Models for IC Design

by

Sriramkumar Venugopalan

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Chenming Hu, Chair

The semiconductor industry has relied on accurate device models for analyzing, predicting and innovating integrated circuit design. Multi-gate MOSFET device architectures like FinFETs are beginning to replace their planar MOSFET counterparts at the 22 nm technology node to enable continued technology scaling. Vertical cylindrical gate (CG) MOSFET are touted to replace planar MOSFETs as the memory device for DRAM and NAND Flash offering increased area density. New device architectures together with relentless scaling of MOSFETs for performance mean increased complexity and new device physics that need to be comprehended. This new understanding needs to be translated into device models for technology progress. Newer device models also require newer methodologies for model creation process and usage for circuit design.

In this thesis we develop a comprehensive compact SPICE model for a CG MOSFET. Relying on fundamental physics based electrostatics description (Poisson Equation) of the device analytic equations for terminal current and capacitance are derived forming the core model. Including all requisite real device effects we validate this model to both numerical simulations (TCAD) and hardware silicon data showing $< 1\%$ RMS error when the model is tuned to the data. For channel diameters < 20 nm quantum mechanical confinement effects tend to dominate. The complex bias and geometry dependence of the inversion charge centroid is captured through a phenomenological model. This model helps accurate prediction of the reduction in gate capacitance of a CG MOSFET. This model was also extended to carrier confinement in thin channels such as the double gate FET or FinFET. The vertical CG MOSFET exhibits asymmetry w.r.t. source and drain. With the aid of TCAD we propose that non-uniform vertical channel doping and structural differences in the top and bottom (source/drain) junction regions as the major contributors to the asymmetric behavior. We then create a mathematical framework to capture these asymmetries in the compact model developed above. We validate this approach by showing excellent agreement to hardware silicon data from a high voltage vertical CG MOSFET technology. All these models have been incorporated in BSIM-CMG the first industry standard multi-gate MOSFET model. Despite including many complex physical effects the resultant model can be executed in the

order of few 10's of $\mu secs$ (per operating point) enabling rapid very large scale integrated circuit design.

A compact SPICE model maintains a balance of predictive nature and flexibility with many sub-components describing various physics and tunable parameters in order to capture data from various sources accurately. This could quickly become unmanageable during a model creation process. For this we propose a RF model extraction procedure that does not require any additional sub-circuit elements and takes advantage of advances in parameter optimization tools available today in an efficient manner. We demonstrate this procedure on high frequency data from multiple planar MOSFET technologies discussing various use cases. Using BSIM6, a bulk planar MOSFET compact model the resultant procedure was able to capture silicon data even beyond the cut-off frequency of the MOSFET and predict various RF circuit design figure of merits with great accuracy.

To my Parents, Sis, extended family,

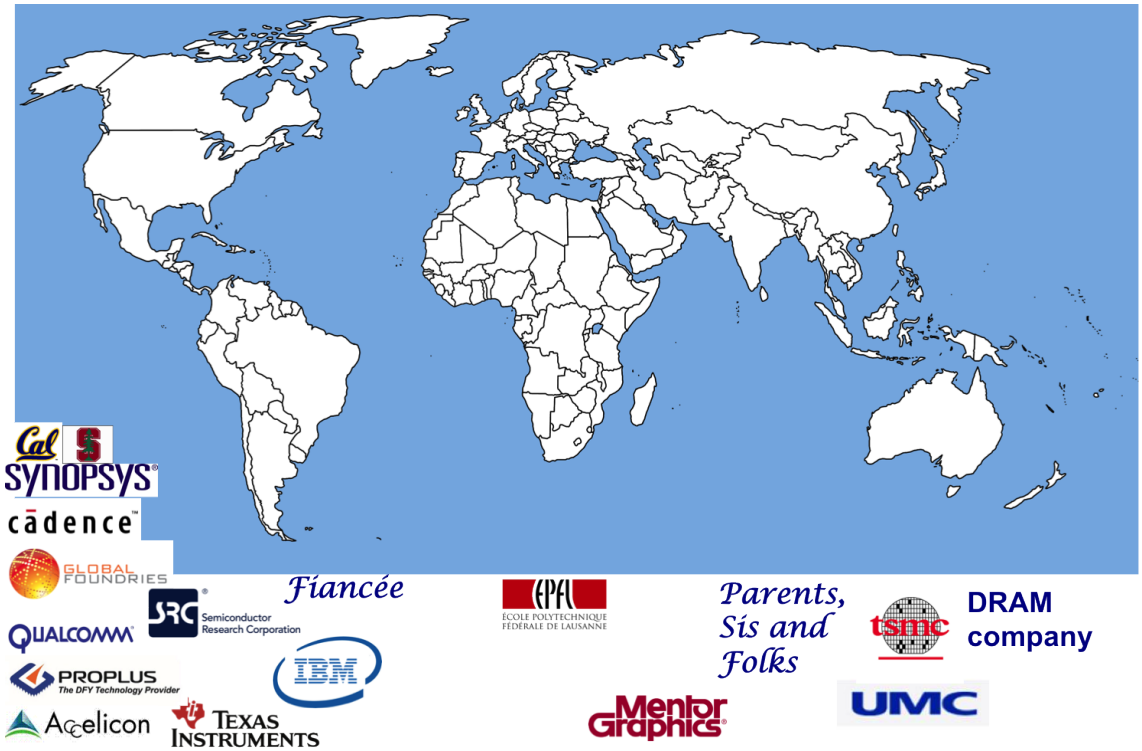
and some wonderful Mentors.

Contents

List of Figures

Acknowledgments

This dissertation thesis stands on the shoulders of two people who made it possible for me - Prof. Chenming Hu and Prof. Ali Niknejad who both accepted me as a graduate student five years ago. I would like to thank them for allowing me to work on a high visibility project. I am indebted to them for the inculcating in me the art of innovation and in-depth analysis. Prof. Hu has been a very patient teacher in my formative years and provided me with endless support throughout my graduate school. Prof. Niknejad's thoughtful criticism and constant encouragement have helped me immensely in my research. He would often times amaze me by bringing up a view point I wouldn't have though about despite all the rigor of research I would have put in. I also thank UC Berkeley (for the Frank and Lucas Margaret Fellowship for my first year of graduate studies), Semiconductor Research Corporation and the Compact Model Council for funding my research and supporting me generously.



Thanks to the people that kept me going and made this dissertation possible.

I would like to thank Prof. Sayeef Salahuddin and Prof. Junqiao Wu for taking off time from their busy schedules to serve on my qualification exam committee. Their valuable suggestions on my on-going research was timely for mid-course corrections to my work. I would also like to thank Prof. Paul Wright for his feedback and help shaping this dissertation. Prof. Jaijeet's deep insights in design automation helped me understand how the models

interacted with the SPICE simulators and I am thankful to him to allow me to contribute to his research as well in this process.

Over the course of five years I was fortunate to have interacted with numerous members from semiconductor industry across the world who served as wonderful mentors with their critical feedback 24/7. Early morning telecons to late night email exchanges kept my determination consistent. My day begun with the rise of sun in Japan (not the US) with Peter Lee and moved on to interactions with the modeling group at TSMC Taiwan (Min-Chie Jeng, Ke Wei Su and Wai Kit Lee), the Mentor Graphics SPICE modeling team at Cairo, Egypt (Ahmed Ramadan, Selim Mohamed and Abo-Elhadid Ahmed), the EKV modeling team at EPFL Switzerland, our earliest adopters and well-wishers Chung-Hsun Lin and Richard Williams at IBM New York, Sagnik Dey at Texas Instruments Dallas, our supporters at Qualcomm San Diego, GLOBALFOUNDRIES Sunnyvale and Proplus at San Jose. At a certain period of time the BSIM-CMG implementation team, in my opinion consisted of Weidong Liu, Joddy Wang and Qingxue Wang at Synopsys, Fulong Zhao and Jushan Xie at Cadence and myself at Berkeley. We owe the code quality control to Brian Chen at Accelicon (now Agilent). Special thanks to all of them for showing the greater picture of the SPICE model ecosystem in the semiconductor industry. It was great interacting with you all and owe all my learning outside academia to you.

I would like to thank the Texas Instruments SPICE Modeling team (Xin Zhang, Krish Dandu, Sagnik Dey) for my internship experience that provided me a chance to apply the developed models to realistic scenarios. I am grateful to Sriram Balasubramanian for giving me an opportunity to work at GLOBALFOUNDRIES as an intern on leading edge CMOS technologies.

Thanks to my peers in Device Group and IC Group (BWRC) who made my experience at Berkeley most special - Darsen and Reinaldo for patiently spending a lot of time bringing me up to speed in my research in device physics and modeling, Jiashu, Siva and Steven for guiding me through my first IC tape-out process, Angada, Asif, Byron, Juan, Kartik, Khalid, Nattapol, Navid, Nuo, Rumi, Shijing, Sung and Tanvir for the invaluable and endless discussions on different research topics. Thanks to all my friends beyond Berkeley who kept my sanity levels in check and kept my life fun and interesting all through my graduate school - Ashwin, Krishna Teja, Shrey, Saaransh, Sudeep, Ruchi and Vikas. My stint at SPICMACAY wouldn't have been memorable if not for the good camaraderie with Aditya, Anuj, Debanjan, Sharanya, Venky and Yasaswini.

Thanks to my apartment-mates at Berkeley Chintan and Siva for their unconditional support through the ebbs and tides of graduate school and for being partners in kitchen crime (a.k.a. cooking).

Finally I would like to express my gratitude to my mom Usha and dad Venu for their immense sacrifices to ensure I received the best possible education. My aunts, uncles and grandma for their constant encouragement, guidance and support throughout my 25 years in school. They all played an important role in goal setting in my life. To my sister Sriranjani for always being there for me. Lastly I would like to thank my fiancèe Soundarya for being patient, supportive and for making sure I stayed on course while I write this thesis.

Chapter 1

Introduction

The semiconductor industry is one of the few engineering areas where *first time right* product design is expected. This creates the need for complete understanding of underlying physical phenomenon of the devices and process technology. Also, today the technology cadence of many consumer electronic products have rapid refresh rates that their shelf life seem to be approaching that of a banana. These rapid refresh rates require faster design cycles which in turn need speedy and accurate simulations to predict design performance improvements. Compact SPICE Models take a seat right in between these two paradigms. This dissertation will present some key advances created in the area of Compact SPICE Models and discuss their benefits for integrated circuit (IC) design.

1.1 Industry Relevance

With increasing technology complexity and ever-decreasing time to market for electronic products from newer technology nodes, IC design and technology development cannot remain de-coupled. Unlike in the past where product development happened after process technology maturity to benefit from high yields, today IC design houses closely collaborate and engage with their foundry partners as much as 3 years prior. Product definition and intellectual property (IP) development begin even before the availability of first silicon hardware results. Compact SPICE models especially BSIM family of models are very fundamental to this transaction between the IC design houses and foundries. The exchange of information between the two parties happen at a number of stages from product definition to final product validation through a set of technology definitions called a process design kit (PDK), Fig. 1.1. Compact SPICE models are an integral part of this PDK enabling accurate IC design simulations.

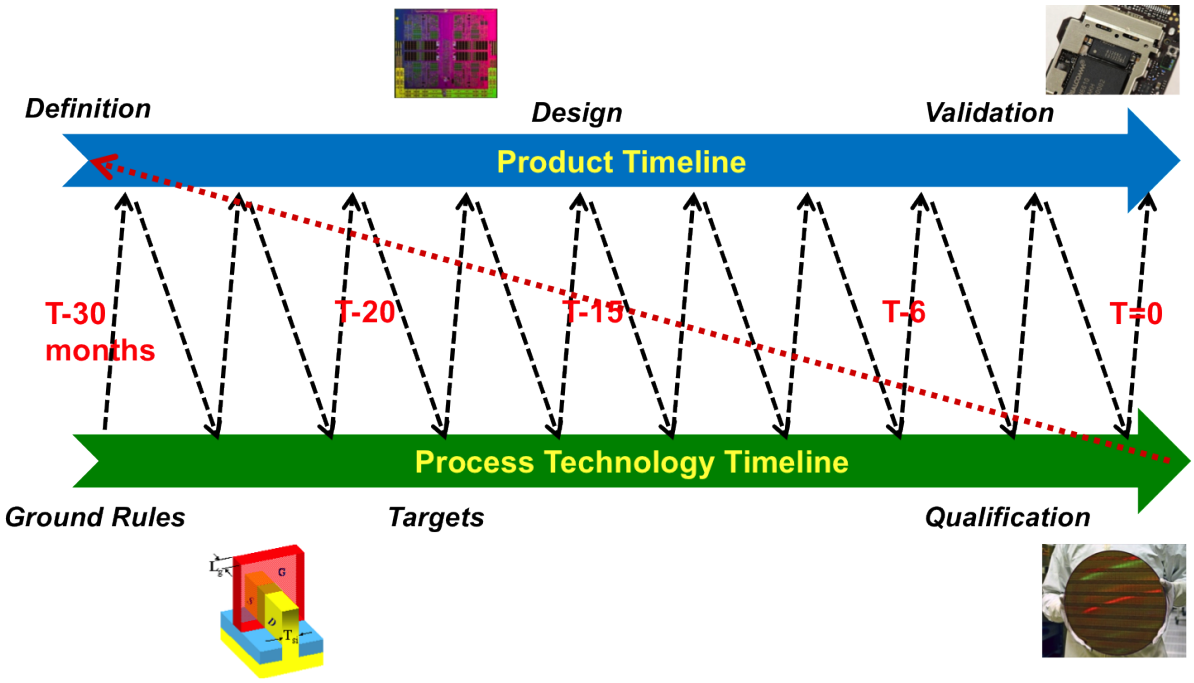


Figure 1.1: Semiconductor technology-IC design interaction along the different phases of product development. The dotted lines represent exchange of information in the form of Compact Models for SPICE.

1.2 Advanced MOSFET Device Architectures

The legacy bulk planar MOSFETs have hit the limits in terms of scaling, [1]. With ever shrinking scaling length for increasing MOSFET performance (speed) the gate terminal of the MOSFET was no longer in control of the channel yielding the control more to the drain terminal. This led to increasing sub-surface leakage current that determined the off-state / stand-by power consumption of the system which has become more important with the advent of always-on mobile devices. In order to stem this leakage current one could enhance the gate control over the channel by either increasing the gate capacitance or enhancing the doping in the channel. An increase in gate capacitance can be achieved either by thinning the gate oxide or by introducing materials with higher dielectric constant (κ). As thinning the silicon oxide hit the limits (due to increased direct tunneling based gate leakage) around the 65 nm technology node, high-κ materials were introduced in the 45 nm technology node, [2]. However even this solution begun to hit the limit when approaching sub-20 nm node. The limit for how much high a doping can be introduced in the channel is set by the mobility degradation due to scattering, [3]. Strained channel that was introduced 90 nm technology node as a mobility enhancement technique also ran out of steam as there is a limit to the amount of stress the channel can experience before material defects arise degrading the performance thereon. [4].

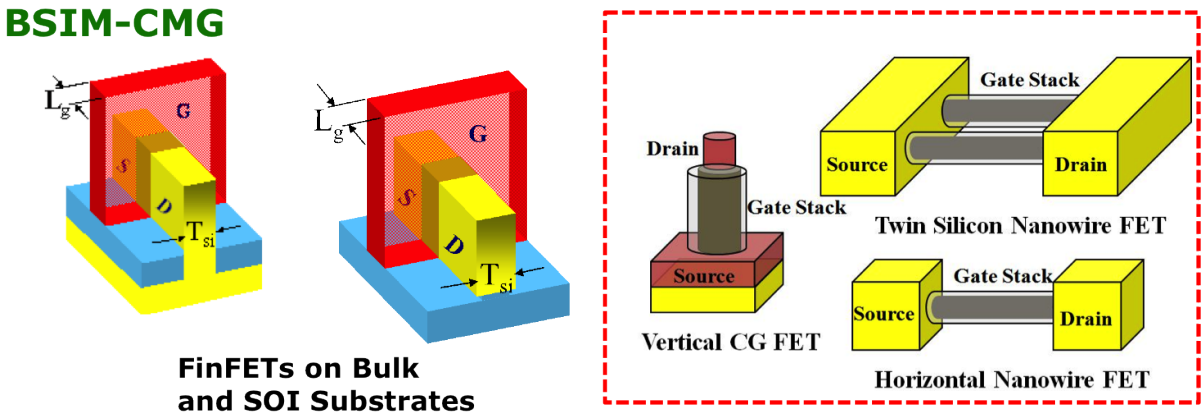


Figure 1.2: Various Symmetric Multi-Gate Transistor architectures

1.2.1 Symmetric Multi-Gate FET - FinFETs

At this juncture there was a need for novel device architectures that help continue the device scaling. Three dimensional multi-gate device architectures were foreseen as the natural replacement to the two dimensional planar device counterparts, [5]. Symmetric multi-gate transistors comprise of multiple gate terminals on more than one side of the channel controlled by a single voltage, Fig.1.2. FinFETs (an extension of the double-gate FET) were first demonstrated in silicon a decade ago, [6]. They have been adopted by the industry as a solution for the sub-22 nm node recently for logic / digital circuit design technology where higher performance and lower power are of prime importance, [7]. These multi-gate devices present the following advantages for circuit design,

- No doping required in the channel due to better gate control of the channel. The threshold voltage, V_{th} is in-turn set by the metal gate work-function.
- Higher channel mobility due to lower scattering in the channel (no dopants in the channel). The mobility degradation due to vertical field is also lower for achieving the same amount of inversion level as a bulk-planar MOSFET.
- Better short channel control by the gate leads to ≈ 60 mV per decade sub-threshold swing thereby making the transistor a better switch.
- Better short channel control also leads to lower drain-induced barrier lowering especially V_{th} shift due to drain bias.
- The optimal device design is achieved where there is a slight underlap of the gate terminal over the source-drain region leading to lower gate-induced drain leakage or lower off-current. The on-current due to increased access resistance is traded-off with off-current due to leakage.

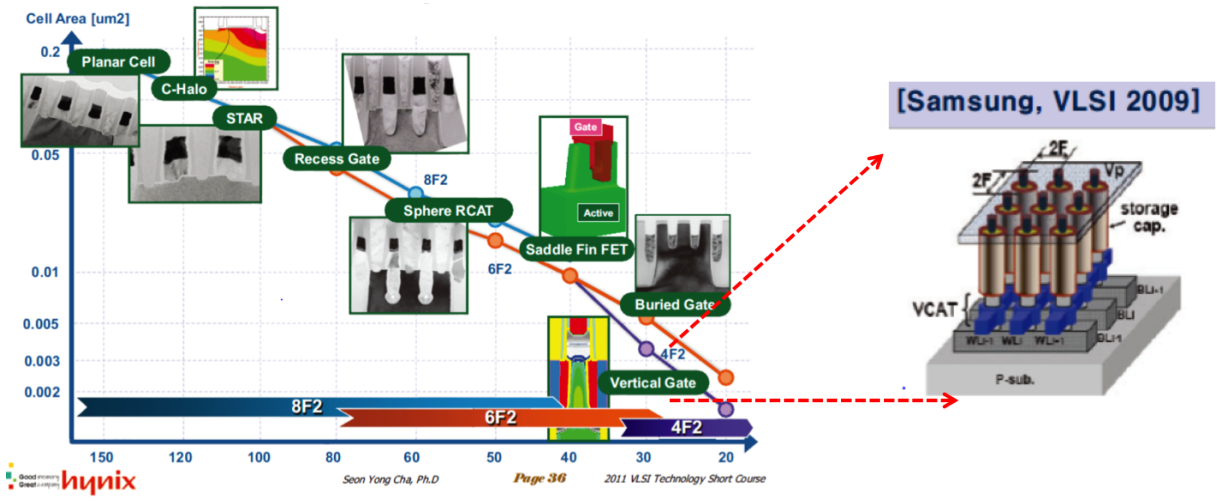


Figure 1.3: DRAM transistor device technology progress over the years.

Due to very little doping in the channel these devices are fully-depleted unlike the bulk-planar MOSFETs that operate in the partially depleted regime. Strained channel and high-κ materials as used in bulk planar devices can also be used here. In realistic FinFETs some amount of body-doping is required to be employed to cater to multiple-V_{th} that are required for contemporary system-on-chip (SoC) technologies for better power-performance-area trade-off without affecting the device variability due to random dopant fluctuations [8]. The next-generation FinFETs with shorter channel lengths for better performance and short-channel control can be obtained by thinning the fin/channel which can be controlled by advances in lithography.

1.2.2 Symmetric Multi-Gate FET - Surround Gate / Gate-All-Around FETs

The ultimate architecture in terms of multi-gate devices would be the surround gate / gate-all-around / quadruple gate / nano-wire transistors where the gate surrounds the channel on all four sides, Fig.1.2. These devices have the same benefits as the FinFETs with even better short-channel control. Horizontal nanowire transistors though demonstrated in research could pose a challenge for large scale manufacturing mainly due to increasing complexity of having to undercut the channel in order to wrap the gate below and controlling the channel thickness uniformly across the wafer. The DRAM bit-cell transistor has been following its own course in terms of device architecture seeking a different set of trade-offs than what is required for logic technology devices, [9]. Bulk planar MOSFETs to recessed channel MOSFETs to saddle MOSFETs (similar to FinFETs) have all been used in an attempt to contain the transistor leakage that directly affects the bit retention rate of the DRAM bit cell. Vertical cylindrical gate (CG) FETs are seen as the replacement for the current generation transistors, Fig.1.2. A DRAM bit-cell consists of a capacitor where the memory

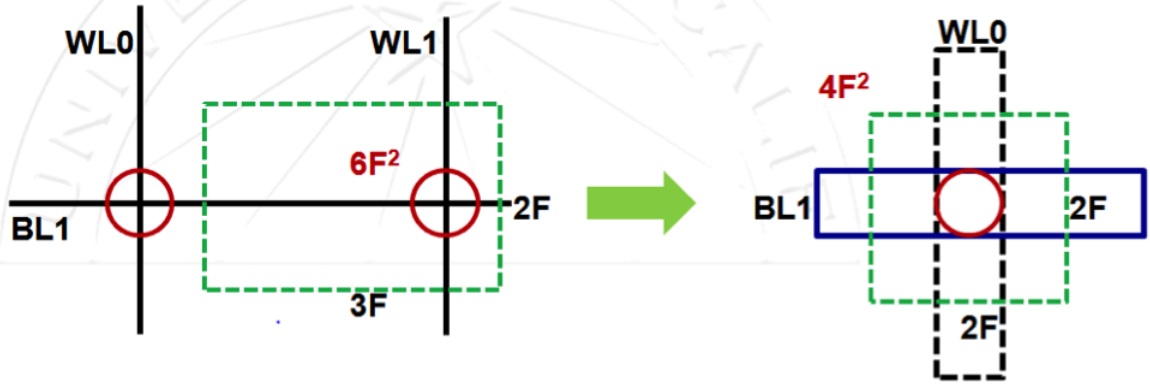


Figure 1.4: Layout of an one-transistor one-capacitor DRAM technology demonstrating the move from a $6F^2$ to a more compact $4F^2$ per bit cell enhancing DRAM area density.

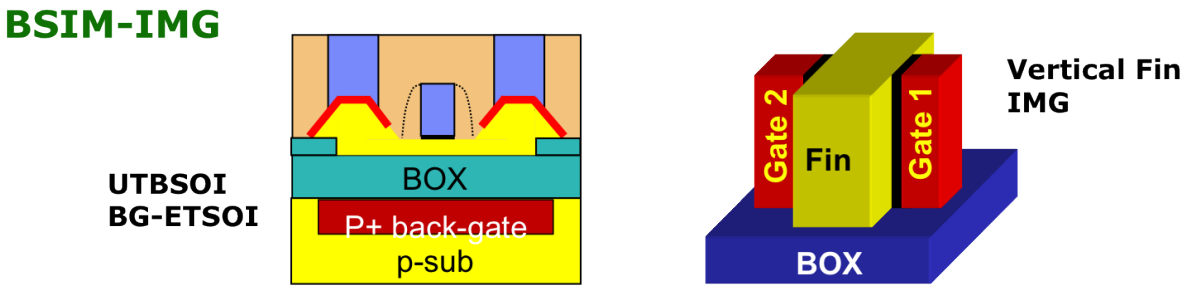


Figure 1.5: Various Asymmetric Multi-Gate Transistor architectures

bit is stored and a transistor that behaves as a switch to control the read and write to the capacitor. The transistor together with the capacitor if placed beside each other would require about $8F^2$ are where F is the minimum feature size of a technology determined by lithography patterning limits. In an attempt to increase the DRAM bit-cell area density a staggered bit-line word-line approach (as opposed to orthogonal bit-line and word-line in $8F^2$ technology) reduced the per cell area to $6F^2$. The vertical CG-FET architecture provides the benefit of being able to reduce this area even further to $4F^2$, Fig.1.4. A cross-bar type of memory array architecture could now be employed where in the bit-cell capacitor can now be stacked over the transistor itself.

1.2.3 Asymmetric Multi-Gate FETs

Another multi-gate architecture that is also currently being used in IC design is the asymmetric multi-gate transistor, Fig.1.5. The asymmetry could arise from either the device structure (different oxide thickness, different gate work-functions) or through application of

different gate voltages at the different gate terminals. Ultra-thin body transistors that have also been touted as a replacement for bulk-planar transistors is another form of an asymmetric multi-gate device, [10]. Here the second gate is present directly below the channel and could be controlled by a separate contact. These devices also address sub-surface leakage in a similar way as the FinFETs where in the channel is thinned down cutting off the leakage paths. The channel here is horizontal like the planar MOSFET, unlike the FinFET which is vertical. These devices have also seen the light of the day recently, [11]. FinFETs with the top gate etched away results in a channel that is controlled by two different gates on either sides. However for asymmetric multi-gate transistors there is a resultant area impact in having to contact both the gates separately. These devices have not yet been used to their fullest potential. Dynamic biasing of both the gates could have some interesting applications. For digital logic technology, the second gate could be used to dynamically control the V_{th} of the device in-order to achieve better control of power, performance and process variations in a SoC technology. One could also visualize a single device RF mixer where both the gates could be excited by two different frequency tones and the drain terminal current would then contain the various frequency modulated components.

1.3 Compact SPICE Models

The maturity and wide acceptance of SPICE (an integrated circuit simulator tool) led to the need for compact device models. The initial research for advanced FET compact device models for SPICE began at Berkeley, [12, 13]. With better understanding of the MOSFET and increasing complexity the compact device models became an important area of research in itself leading to the Berkeley SPICE Igfet Model, BSIM. For more than two decades the BSIM group has been advancing the field with research and developments leading to introduction of various industry standard models, Fig.1.6, [14].

A compact model for a semiconductor device is a concise mathematical description of its complex behavior, and it is usually implemented in a computer programming language like C or Verilog-A. Despite the fact that the implementation might consist of thousands of lines of codes, it takes only a fraction of a millisecond for computer simulation tools like SPICE to run the code (for one transistor at a single bias point). In addition, the model's accuracy and quality is of very high importance. The speed and accuracy (1% RMS error after calibration to experimental data) enable simulation tools to verify the functionality and performance of ICs (containing millions of transistors) before an expensive fabrication process takes place. The models are also flexible enough to accommodate technology modifications from multiple foundries. The models undergo continuous innovation with fundamental contributions from universities and important incremental contributions from industry leading to their deployment. Compact Model Council is the industry consortium that proposes, approves and ensures the quality of today's industry standard compact models, [15].

Contemporary MOSFET compact models constitute a core model that contains analytic equations for the terminal currents and charges for an ideal long channel MOSFET, Fig.1.7.

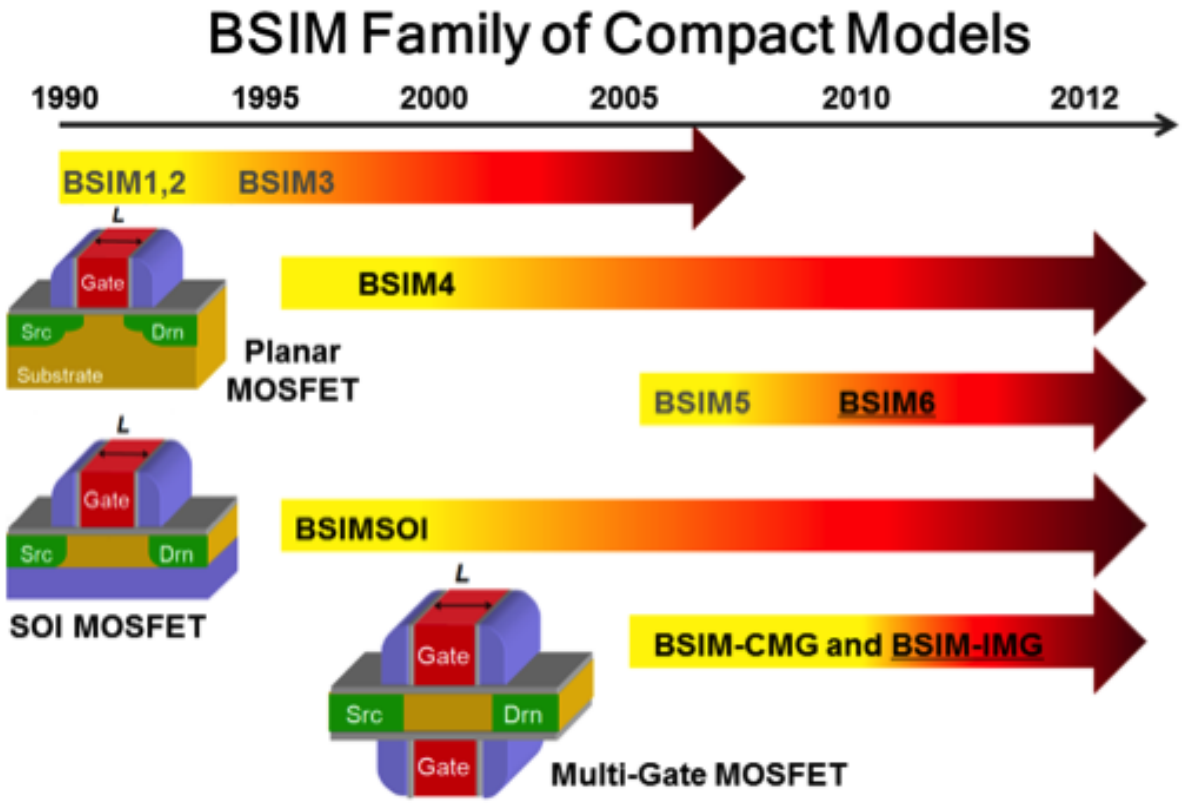


Figure 1.6: Timeline of the BSIM family of compact models. The models in gray (BSIM1-3 and BSIM5) are not supported by the group any longer. The underlined model (BSIM-IMG) is under evaluation/standardization by CMC.

Terminal charges are implemented rather than the device capacitance directly as it would help visualize and maintain charge neutrality in the device. The simulators convert the terminal charges to terminal capacitance. These currents and capacitance are obtained as a function of either the terminal voltages or the source and drain end charge / surface potentials. The currents and capacitance equations are further augmented by various sub-models that capture the complex physics of the real device. Some of these sub-models are shown in Fig.1.7 (Also see Appendix A). Some examples of compact models for MOSFET include - BSIM, PSP, EKV and HiSIM for bulk planar MOSFETs, HiSIM-HV for high-voltage MOSFETs, BSIM-CMG for symmetric multi-gate MOSFETs, BSIM-SOI and HiSIM-SOI for MOSFETs built with Silicon on Insulator Technology, [15]. Most compact models have a certain degree of inherent scalability along important physical device parameters like length, width, gate oxide thickness and channel doping etc. All compact models contain many parameters to be tuned/extracted that render just enough flexibility to the model to be able to capture data from multiple foundries and industries. These parameters are estimated from a small set of hardware data and the so tuned compact model's inherent scalablility across geometry helps predict device performance beyond that of available data

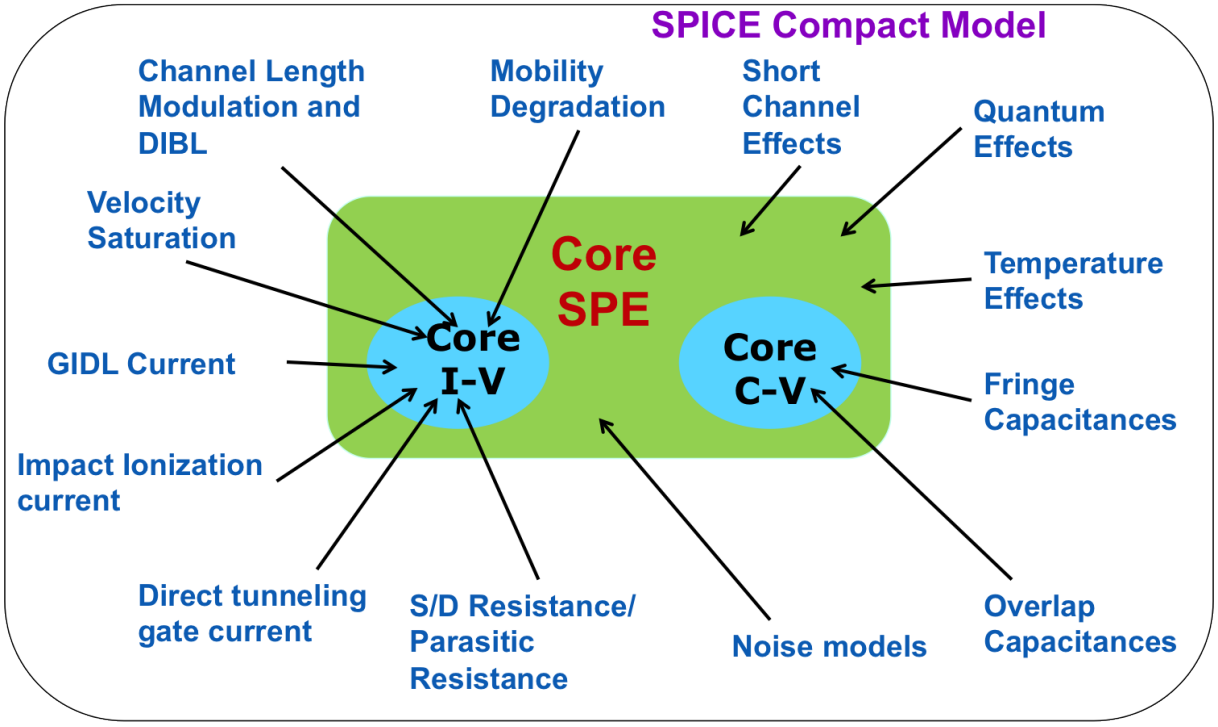


Figure 1.7: The various components and structure of contemporary Compact SPICE transistor models.

by extrapolation.

1.3.1 BSIM-CMG

BSIM-CMG (complete multi-gate) is a compact SPICE model developed for symmetric multi-gate FET architectures, Fig.1.2. The model is very versatile and supports a variety of multi-gate FET architectures including double-gate (or a FinFET with a hard-mask on the top), triple-gate (gate wrapped around on three sides), quadruple-gate (gate on all sides with a square or rectangular channel cross-section) and cylindrical gate (gate wrapped around a circular channel cross-section) devices. This dissertation concentrates on the development of a core model for cylindrical gate devices. The details of the core model for the other architectures can be found in [16]. All the core models are physical charge/surface potential based and capture the volume inversion phenomenon of fully-depleted channels. The short channel effects, quantum mechanical confinement effects (also developed in this dissertation), source/drain parasitic resistance and capacitance were all specially developed keeping in view the 3-D device geometry, [17]. Through work in this dissertation, BSIM-CMG also models asymmetric devices such as the vertical CG FET device. The model is continuous (drain current is a single piece analytic equation) and has no discontinuity in its higher order derivatives and is thus usable for analog/RF circuit design. As a result,

BSIM-CMG was voted the first industry standard multi-gate model by the Compact Model Council in 2012.

1.3.2 BSIM6

BSIM6 is a RF relevant update to the legacy industry standard model for bulk-planar MOSFETs, BSIM4, [18]. BSIM4, a threshold voltage based model and the industry work-horse for digital and analog circuit designs was known to have issues in predicting RF circuit blocks like passive mixers due to discontinuity in the model, [19]. BSIM6 rectifies all the issues and is able to satisfy all RF relevant quality tests. We will discuss more about this in this dissertation. At its core it consists of a physical charge based model adopted from the EKV MOSFET model, [20]. Other than the usual sub-models for capturing a real device (such as mobility degradation, velocity saturation, and output conductance etc.) BSIM6 also has all RF relevant components embedded within the model package itself. We will highlight all of them in this dissertation. BSIM6 shares many of its parameter names with BSIM4 to retain the familiarity. We will discuss a RF CMOS model development procedure in this dissertation using BSIM6 as an example. BSIM6 was accepted as an update by the Compact Model Council in 2013.

1.3.3 BSIM-IMG

BSIM-IMG (independent multi-gate) model is being developed for asymmetric multi-gate devices, Fig.1.5. Similar to BSIM-CMG, BSIM-IMG is also a surface potential based model. The core model assumes a asymmetric double gate FET and supports asymmetry in both gate-oxide thickness and gate work-function, [17, 21]. Owing to its similarity to BSIM-CMG, BSIM-IMG can share all the real device effect models taking into account the back-gate bias dependence. Currently BSIM-IMG supports channel inversion on one side and treats the other gate as a threshold voltage tuning knob, i.e. it does not support inversion on both sides of the channel. Hence this model is more suited for Ultra-thin body transistor technology. Inclusion of dual side inversion has been dealt in [22, 23]. Advances to BSIM-IMG is beyond the goals set for this dissertation.

1.4 Compact Models Usage

Compact MOSFET models are used in various ways. Models tuned to the current technology node are applied for predicting the performance of the next technology node through a process commonly known as technology targeting. Targets for the next generation node are set by performance extrapolation or TCAD device simulations. Compact models thus created are distributed to circuit designer for an early look-ahead of what lies in the next generation. For a digital designer, the MOSFET behaves as a non-ideal switch. The intrinsic delay in switching given by C_gV/I_d, where C_g is the gate capacitance, V is the operation

voltage and I_d is the drain current of the device is an important metric. Compact models are thus required to describe the terminal currents and terminal capacitance accurately in order for precise estimation of power and performance of digital circuits. For analog/ mixed signal designers the trans-conductance efficiency - cut-off frequency product to minimize power and ensure adequate bandwidth given by $G_m/I_d \times F_t$ is an important metric where G_m is the device trans-conductance. For this the compact model is required to predict not only the currents and capacitance but also the first derivatives accurately (the gain of an amplifier is given by G_m/G_ds where G_{ds} is the output conductance). In addition for some circuits device noise is also important and hence the compact models are required to capture all forms of device noise including thermal, channel, flicker and spot noise. RF wireless designs typically contain fewer transistors per chip and heavily rely on the compact models. Important complex metrics that affect the system behavior such as linearity, higher order distortion, frequency mixing and noise folding etc. are all expected to be predicted by the compact models very accurately. This entails the model be able to capture not only the drain currents and capacitance but also predict its derivatives up to third - fifth order. The model should also exhibit no discontinuity in these derivatives. The compact models are implemented and distributed as a part of commercially available SPICE simulators by the Electronic Design Automation (EDA) industry (such as HSPICE, Spectre, and SmartSPICE etc.).

1.5 Scope and Organization of Dissertation

The goal of this dissertation is to advance compact SPICE MOSFET models in light of recent developments in multi-gate device architectures which have become a replacement for bulk planar MOSFETs. In addition there will be a special emphasis on RF CMOS compact models for Analog/RF circuit design.

After a brief review of progress in multi-gate MOSFET core models, in Chapter 2 we will develop a comprehensive core model for a cylindrical gate (CG) FET. The physical charge based core model will include effects of gate poly-depletion and channel doping as well. We will present analytic derivations for ideal long channel drain current and terminal charge / capacitance (the short channel equivalents will be discussed in light of a generic charge based model in Chapter 5). After incorporating the developed core model in BSIM-CMG along with other real device effects we will validate the CG-FET model to both TCAD and hardware data.

In Chapter 3 we will discuss two effects that arise due to the 3-D nature of the symmetric multi-gate devices. In particular we will first develop a simple phenomenological model for the complex geometry and bias dependence of charge centroid due to quantum confinement effects of the carriers in a fin/nano-wire channel. We will validate this model by demonstrating it's importance to predicting moderate inversion region gate capacitance. Next we will look into a subtle yet important double source/drain junction that arises in a FinFET on bulk substrate. We will extend the single junction model found in BSIM4 and validate it

with TCAD simulation results.

In Chapter 4 we will delve deeper into the vertical CG-FET architecture and understand the asymmetric device behavior. We will translate the understanding into compact models to be incorporated in BSIM-CMG. In addition, a mathematical framework to incorporate any source-drain asymmetry in a compact model while retaining the physical nature and continuity of higher order derivatives of the drain current is presented. We then validate this model to hardware data from a 2.2 V vertical CG-FET technology.

In view of the recently developed BSIM6, in Chapter 5 we will establish a generic RF CMOS Model development procedure. First we will critically review all the RF components of the BSIM6 model and understand their RF related effects and limitations. As a part of the critical review, we will also develop a short channel terminal charge / capacitance model applicable for any charge based MOSFET model and highlight its importance in predicting high-frequency behavior. We will also present a detailed discussion on the models that capture the non-quasi static effects of a MOSFET and discuss its correct usage. Finally, a physical model extraction procedure is developed that requires no additional sub-circuit elements. The procedure is validated using hardware S-parameter data from multiple technologies for a bulk planar MOSFET highlighting the benefits of the BSIM6 model components.

We finally summarize the key findings in Chapter 6 and present some suggestions for future work in this area.

Chapter 2

Core Models for Multi-Gate MOSFETs in BSIM-CMG

Contemporary SPICE compact models for FET contain a physical core model that describes the electrostatics and transport of the channel carriers for an ideal long channel transistor. They fall under two varieties -

- Threshold voltage based model like that in industry standard bulk planar transistor model BSIM3/4. Threshold voltage, V_{th} (the gate voltage at the onset of inversion) is the primary unknown for a FET device in this model. The charge in the channel is explicitly expressed as a function of the terminal voltages and V_{th}.

- Charge/Surface Potential based model like that in MOS11, EKV, PSP, HiSIM etc. An analytic solution of Poisson equation under boundary conditions set by the device architecture and suitable approximations leads to an implicit equation of the charge/surface potential in the channel as a function of terminal voltage and other physical device parameters. This implicit equation is solved for to obtain the charge/surface potential in the channel.

FET devices that are currently being used in design are not operating in the quantum regime. The transport is therefore accurately described by Boltzmann transport equations that leads to a drift-diffusion based description for drain current model. The drain current is a function of the source-end and drain-end channel charge obtained previously. In order to predict the transient and small-signal operation conditions device capacitance/trans-capacitance is described using a terminal charge model. The terminal charges are also obtained as an analytic function of the source-end and drain-end channel charge using suitable partition schemes to describe the source and drain charge. Recently charge/surface potential based core models have gained popularity owing to their physical and single-piece smooth drain current and capacitance behavior covering all operation regions. Threshold voltage based model required stitching together of drift and diffusion currents with suitable smoothing

functions that on one hand rendered flexibility but on the other hand led to non-physical capacitance behavior.

In this chapter we will review the different symmetric multi-gate core models available in literature. We will then develop a comprehensive core model for cylindrical/surround gate transistor (CG-FET) covering both accumulation region capacitance and poly-silicon gate-depletion. We will validate the developed model using both TCAD and hardware silicon data to demonstrate the physical predictive nature of the core model across FET geometry, terminal bias and process parameters.

2.1 Symmetric Multi-Gate Core Models - A Review

The electrostatics governing the carriers in a FET is described by the 3-D Poisson's equation in Cartesian co-ordinates as follows,

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}=\frac{qn_i}{\epsilon_{si}}\exp\left(\frac{-\psi}{V_t}\right)+\frac{qn_i}{\epsilon_{si}}\exp\left(\frac{\psi-V_{ch}}{V_t}\right)+\frac{qN_a}{\epsilon_{si}} \tag{2.1}$$

where ψ is the potential at *(x,y,z),*$\epsilon_s i$ is the permittivity of silicon channel, n_i is the intrinsic carrier concentration of the channel material (here assumed silicon), N_a is the channel doping concentration and V_{ch} is the channel potential at a distance y from the source. It is also assumed that z is in a dimension in and out of the page, and x is along the gate-insulator-channel direction. As much of the action for an ideal device are in the x and y dimension, the $\partial^{\psi}/\partial z^2$ term is ignored here. Solving the 2-D Poisson equation with appropriate boundary conditions also turns out to be cumbersome and not leading to analytic expressions. It is usual practice to assume that the electric field along x dominates over the electric field along the channel direction, y. This assumption then allows us to convert Eqn.(2.1) to a 1-D Poisson equation. This assumption known as gradual channel approximation is true for a long channel device where the source and drain do not influence the channel. Also in a compact model framework we recover the 2-D effect (commonly known as short channel effects) by other sub-models.

The three terms in the RHS of the Eqn.(2.1) represent the holes, the electrons and the fixed charge in the channel. The 1-D Poisson equation in x with all the three terms does not still give us analytic closed form solutions conducive for compact models. Unlike bulk MOSFETs where both electric field and potential are known to be zero at the edge of depletion, symmetric multi-gate FETs have only one boundary condition in terms of the electric field at the center of the fin/body (electric field is zero due to symmetry). This in turn means one would have to integrate the Poisson equation twice (just once suffices for bulk MOSFET) to get a complete solution. However for doing so one requires the use of incomplete elliptic integrals to describe the solution which are not analytic. So another approximation is made by ignoring the holes and considering only the above flat-band voltage operation of the device. With advent of metal gate and the foresight of a fully-depleted undoped channel having better mobility, FinFETs and CG FETs were touted to be made of near intrinsic

silicon channel. For this case the Poisson equation simplifies to (ignoring even the fixed charge term)

$$\frac{\partial^2\psi}{\partial x^2} = \frac{qn_i}{\epsilon_{si}} \exp\left(\frac{\psi - V_{ch}}{V_t}\right) \tag{2.2}$$

For this case, Taur showed that for a double-gate (DG) FET with a symmetric boundary condition imposed ($d\psi/dx = 0$ @ center of the fin) results in a closed form solution for $\psi(x)$ as follows,

$$\psi(x) = V_{ch} - \frac{2kT}{q} ln\left[\frac{T_{FIN}}{2\beta}\sqrt{\left(\frac{q^2 n_i}{2\epsilon_{si}kT}\right)}\cos\left(\frac{2\beta x}{T_{FIN}}\right)\right] \tag{2.3}$$

where the symbols carry their usual meaning and β is to be determined from the boundary condition at the surface using Gauss law, [24]. This leads to an implicit equation in β as follows,

$$\frac{V_g - V_{fb} - V_{ch}}{2kT/q} = ln(\beta) - ln(\cos(\beta)) + \frac{2\epsilon_{si}t_{ox}}{\epsilon_{ox}T_{FIN}}\beta\tan(\beta) \tag{2.4}$$

Solving for β we can obtain the channel charge as follows,

$$Q_i = \frac{8\epsilon_{si}kT}{qT_{FIN}}\beta\tan(\beta) \tag{2.5}$$

We can observe that Eqn.(2.4) cannot be expressed implicitly in terms of channel charge. However making a linear approximation of β w.r.t Q_i in Eqn.(2.5), the EKV team has developed an implicit equation in channel charge, [25]. Results after the approximation have been validated with TCAD and shown favorable in integrating the resultant core model into the existing EKV Bulk model framework.

In a modern digital IC technology one can routinely find the use of multiple threshold voltage devices to trade off power-performance over the chip. Among the knobs available, introducing body-doping to tune the threshold voltage of a device works the best. The previous solutions for FinFET core models are valid for intrinsic to low channel doping. However unlike the case of intrinsic channel, the Poisson's equation with channel doping does not have a closed form solution and hence one needs to resort to approximations for achieving analytic equations a compact model. Dunga *et al.* used a perturbation approach to tackle body-doping [16] assuming that the body is fully-depleted. Given that the fin thickness in today's technology is < 30 nm it is reasonable to assume the fin is always fully-depleted for moderate amounts of doping (optimum fin thickness is set by minimum channel length to control short channel effects). Partial depletion operation is not a preferred mode for these devices due to degraded subthreshold swing and mobility. The solution to the Poisson equation

$$\frac{\partial^2\psi}{\partial x^2} = \frac{qn_i}{\epsilon_{si}} \exp\left(\frac{\psi - V_{ch}}{V_t}\right) + \frac{qN_a}{\epsilon_{si}} \tag{2.6}$$

is given by

$$\psi = \psi_1 + \psi_2 \tag{2.7}$$

where, ψ_1 is the contribution due to inversion carriers and ψ_2 is the perturbation due to body-doping. The solution to ψ_1 is the same as that obtained previously in Eqn. (2.3). This is used as a first estimate to then solve for ψ from Eqn. (2.6). Using the intrinsic channel solution as a first guess aids in arriving at an analytic solution for ψ, which then with boundary conditions is re-cast into a surface potential equation implicit in β (which is in turn a function of ψ) as follows

$$\begin{aligned}\frac{V_{gs}-V_{fb}-V_{ch}-\psi_{pert}}{2kT/q}+ln\left(\sqrt{\frac{q^2n_i^2}{2\epsilon_{si}kTN_a}}\frac{T_{FIN}}{2}\right)=ln(\beta)-ln(cos(\beta))...\\ +\frac{2\epsilon_{si}t_{ox}}{\epsilon_{ox}T_{FIN}}\sqrt{\beta^2\left(\frac{\exp^{\frac{\psi_{pert}}{kT/q}}}{\cos^2(\beta)}-1\right)+\frac{\psi_{pert}}{(kT/q)^2}\cdot\left(\psi_{pert}-2\frac{kT}{q}ln(\cos(\beta))\right)}\end{aligned} \tag{2.8}$$

$$\psi_{pert}=\frac{qN_aT_{FIN}^2}{8\epsilon_{si}}$$

Feng Liu *et al.* went further to use the above obtained ψ as an updated guess for ψ and solved the Poisson equation in Eqn.(2.6) again to obtain a more accurate ψ, [26]. The so obtained ψ contain error functions and imaginary error functions that need some kind of polynomial approximations for compact model implementation. Here we note that compared to the solution for intrinsic channel, the solutions in [16] and [26] are increasingly computationally expensive.

For a CG-FET the Poisson equation is written in cylindrical co-ordinates for convenience as follows,

$$\frac{\partial^2\psi}{\partial z^2}+\frac{1}{r^2}\frac{\partial^2\psi}{\partial\theta^2}+\frac{1}{r}\cdot\frac{\partial}{\partial r}\left(r\frac{\partial\psi}{\partial r}\right)=\frac{qn_i}{\epsilon_{si}}\exp\left(\frac{-\psi}{kT/q}\right)+\frac{qN_a}{\epsilon_{si}}+\frac{qn_i}{\epsilon_{si}}\exp\left(\frac{\psi-V_{ch}}{kT/q}\right) \tag{2.9}$$

With assumptions and arguments similar to the DG-FET case, *i.e.* we will take advantage of the symmetry along θ, consider only an ideal long-channel device, assume gradual channel approximation (neglect z), and neglect holes and channel doping, Eqn.(2.9) boils down to

$$\frac{1}{r}\cdot\frac{\partial}{\partial r}\left(r\frac{\partial\psi}{\partial r}\right)=\frac{qn_i}{\epsilon_{si}}\exp\left(\frac{\psi-V_{ch}}{kT/q}\right) \tag{2.10}$$

B. Iniguez *et al.* have shown that an analytic solution exists for ψ as follows,

$$\psi(r)=V_{ch}+\frac{kT}{q}lnleft(frac{-8BkT\epsilon_{si}q^2n_i}{(1+Br^2)^2}right) \tag{2.11}$$

where the symbols take their usual meaning and B is the constant of integration obtained by using a boundary condition set by Gauss law at the interface [27]. This then turns into a

surface potential equation implicit in β as follows,

$$\frac{V_{gs} - V_{fb} - V_{ch}}{kT/q} - ln\left(\frac{8kT\epsilon_{si}}{q^2 n_i R^2}\right) = ln(1-\beta) - ln(\beta^2) + \frac{4\epsilon_{si}}{C_{ox}R}\left(\frac{1-\beta}{\beta}\right) \tag{2.12}$$

The charge in the channel can be obtained as $Q_i = 4\epsilon_{si}kT/q \cdot (1-\beta)/\beta$. With this insight, unlike the case of a DG-FET solution, the whole equation, Eqn.(2.12) can be expressed in terms of channel charge allowing for easy incorporation in any charge based model.

$$V_{gs} - V_{fb} - V_{ch} - \frac{kT}{q}ln\left(\frac{8kT\epsilon_{si}}{q^2 n_i R^2}\right) = \frac{Q_i}{C_{ox}} + \frac{kT}{q}ln\left(\frac{Q_i}{Q_0}\right) + \frac{kT}{q}ln\left(\frac{Q_i}{Q_0}+1\right) \tag{2.13}$$

$$Q_0 = \frac{4\epsilon_{si}}{R}\frac{kT}{q}$$

Feng Liu *et al.* extended the above solution for a CG-FET to include channel doping as well [28]. Using the analytic solution for the Poisson equation in weak-inversion region (neglecting inversion carriers and considering only the depletion charge) as an initial guess, they obtain a equation implicit in channel charge. This method was chosen for our work towards building a comprehensive core model. We will discuss the intricacies of the derivation in the next section.

Some recent developments that occurred later in time to when the current work was pursued merit mention. Combining the solutions for DG-FET in [25] and for CG-FET [27], one could visualize a unified channel charge based implicit equation that describes both the architectures. H. Lu *et al.* have pursued this unification and provided the convenience of being able to utilize/create various sub-models (for mobility degradation, current saturation etc.) based on channel charge. The PSP Compact model team, Dessai *et al.* have improved on the accuracy of the approximations involved in converting the DG-FET surface potential equation to one implicit in channel charge [29]. More recently J. P. Duarte *et al.* have even furthered the unification by coming up with a channel charge based implicit equation that encompasses all architectures of symmetric multi-gate FETs as follows,

$$\begin{aligned} V_g - V_{fb} - V_{ch} - \frac{Q_d}{C_g} = \frac{Q_i}{C_g} + \frac{kT}{q}ln\left(\frac{Q_i}{q\frac{n_i^2}{N_a}A_{ch}}\right)... \\ + \frac{kT}{q}ln\left(\frac{(Q_i+Q_{dep})/C_{ch}/(kT/q)}{1-\exp^{\frac{(Q_i+Q_{dep})}{C_{ch}(kT/q)}}}\right) \end{aligned} \tag{2.14}$$

where Q_{dep} is the depletion charge in the channel per unit length, C_g is the gate capacitance per unit length, C_{ch} is the channel capacitance per unit length and A_{ch} is the cross-section area of the channel perpendicular to the flow of current [30]. All the parameters (Q_{dep}, C_g, C_{ch} and A_{ch}) are functions of the geometry of the respective device architecture.

In all above, the authors have concentrated on the unipolar case of operation where only one carrier dominated (electrons in NMOS). This is acceptable for devices that are built on Silicon-on-Insulator (SOI) substrate where there is the bulk terminal supplying the holes is absent. However devices that are built on bulk-substrates like FinFETs, do have accumulation mode of operation and hence holes must be considered. Solving for the Poisson equation with holes (together with body-doping) is cumbersome as discussed before. Dessai *et al.* have rigorously tackled a case where channel doping was neglected and considering both electrons and holes in the channel by making some approximations to avoid incomplete elliptic integrals [31]. However we found that it incurs higher computational expense to solve the implicit equation obtained in this manner. We will provide a simpler alternative to incorporating holes for a NMOS device on bulk substrate (or electrons for PMOS device) in this work.

In a compact model framework the implicit equation discussed for different cases above is usually solved with analytic approximations. Using Newton-Raphson iterations until an error convergence criterion is met is avoided owing to long and unequal computation time for various terminal bias conditions. A two-iteration Householder's method ($3^{rd} order Newton-Raphson$ is typically used with a good initial guess to obtain a solution [32], [33]. This approach has been shown adequate for many cases above by different authors. In all above discussed the respective authors have also shown analytic expressions for drift-diffusion based drain current and Ward-Dutton partition based terminal charge model (either in the cited work or in their subsequent publications). To the knowledge of the author, no work has concentrated on considering the body-effect (threshold voltage modulation by body voltage) that could be observed in devices on bulk substrate so far.

2.2 Comprehensive Core model for Cylindrical/Surround Gate FET

The coordinate system that we will use for derivations in this work is shown in Fig.2.1. Symbols R, L, t_{ox}, t_{poly}, N_a, and N_{poly} represent the radius of the channel, channel length, gate oxide thickness, poly-gate thickness, channel doping and polysilicon gate doping respectively.

2.2.1 Poly-Depletion Effect

The usage of metal as a gate material for CG transistors is likely as the effective oxide thickness (EOT) for sub-22 nm technology node is speculated to be ≤ 1 nm [1]. Using highly doped (at dopant solubility limit) poly-silicon as a gate material will still be a deterrent to achieving this EOT as poly-depletion contributes to the increase in EOT. However for some cost constrained applications like memories it is more likely that polysilicon will still be employed as a gate material. Also for vertical device orientation scaling the length of the device is not as critical and hence EOT scaling can be relaxed. We shall model poly-depletion for

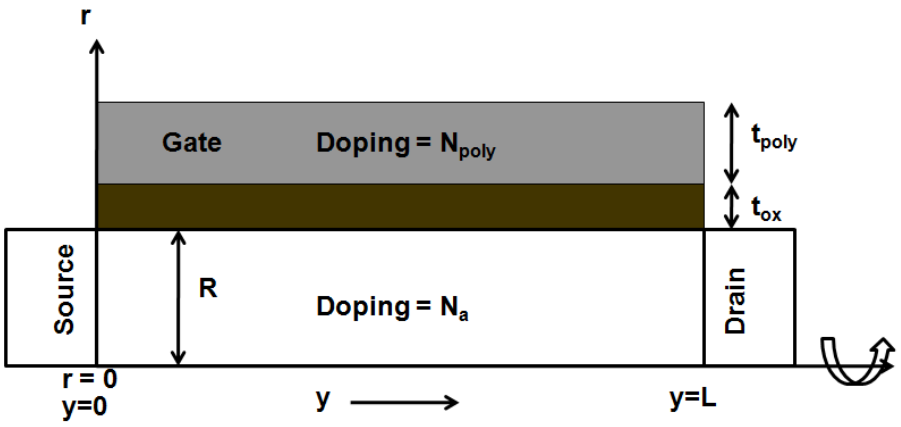


Figure 2.1: Coordinate system convention and variables for the cylindrical gate FET

cylindrical gate and incorporate it in the surface potential equation later.

The Poisson's equation in the oxide assuming no oxide charges is as follows

$$\frac{1}{r} \cdot \frac{d}{dr}\left(r\frac{d\psi}{dr}\right) = 0 \tag{2.15}$$

This leads to the electric field at an arbitrary point $R < r < R + t_{ox}$ in the oxide to be

$$E_{ox}(r) = E_{ox}(R) \cdot \frac{R}{r} \tag{2.16}$$

The voltage drop in the oxide can be derived by integrating $E_{ox}(r)$ w.r.t. r within the oxide.

$$V_{ox} = E_{ox}(R) \cdot R \cdot ln\left(1 + \frac{t_{ox}}{R}\right) = E_{ox}(R) \cdot EOT \tag{2.17}$$

The electric field in the gate at the gate-oxide interface, E_{polyox} is given by

$$E_{polyox} = \frac{\epsilon_{ox}}{\epsilon_{gate}} \cdot E_{ox}(R + t_{ox}) = \frac{\epsilon_{ox}}{\epsilon_{gate}} \cdot \frac{R}{R + t_{ox}} \cdot \frac{V_{ox}}{EOT} \tag{2.18}$$

where ϵ_{gate} and ϵ_{ox} are the dielectric constants of the gate and the oxide respectively.

The Poisson's equation in the gate assuming a poly-doping of N_{poly} is

$$\frac{1}{r} \cdot \frac{d}{dr}\left(\frac{d\psi}{dr}\right) = -\frac{qN_{poly}}{\epsilon_{gate}} \tag{2.19}$$

Integrating it once, the electric field in the poly at r, $E_{poly}(r)$ can be obtained from the below equation.

$$-E_{poly}(r) \cdot r + E_{polyox} \cdot (R + t_{ox}) = \frac{qN_{poly}}{2\epsilon_{gate}}\left[r^2 - (R + t_{ox})^2\right] \tag{2.20}$$

The electric field goes to zero at the edge of the depletion region in the polysilicon gate, i.e. at $r = R + t_{ox} + X_{dpoly}$. This gives us the value of E_{polyox} in terms of X_{dpoly} as

$$E_{polyox} \cdot (R + t_{ox}) = \frac{qN_{poly}}{2\epsilon_{gate}} \left[2(R + t_{ox}) + X_{dpoly}\right] X_{dpoly} \tag{2.21}$$

The voltage drop in the polysilicon gate, V_{poly} can now be obtained by integrating Eqn.(2.20) w.r.t r from the gate-oxide boundary to the edge of polysilicon depletion.

$$\begin{aligned} V_{poly} = {} & E_{polyox} \cdot (R + t_{ox}) \cdot ln\left(1 + \frac{X_{dpoly}}{R + t_{ox}}\right) - \frac{qN_{poly}}{4\epsilon_{gate}} \left[2(R + t_{ox}) + X_{dpoly}\right] X_{dpoly} \\ & + \frac{qN_{poly}}{2\epsilon_{gate}} (R + t_{ox})^2 \cdot ln\left(1 + \frac{X_{dpoly}}{R + t_{ox}}\right) \end{aligned} \tag{2.22}$$

Eliminating X_{dpoly} from Eqn.(2.21) and Eqn.(2.22), we obtain an implicit equation in V_{poly} and E_{polyox}. But from Eqn.(2.18), E_{polyox} can be expressed as function of V_{ox}. This implies that one would end up with an implicit equation that relates V_{poly} and V_{ox}. For the purpose of compact modeling it would become computationally expensive if one were to solve this equation using iterations as a part of the transistor model. So we will resort to an approximation in order to get an explicit relation between V_{poly} and V_{ox}. Using $ln(1+x) = x - x^2/2$, it can be shown that

$$V_{poly} = \frac{\epsilon_{ox}^2}{2qN_{poly}\epsilon_{gate}EOT^2} \cdot \left(\frac{R}{R + t_{ox}}\right)^2 \cdot V_{ox}^2 = c_{poly}V_{ox}^2 \tag{2.23}$$

Fig.2.2 shows V_{poly} as a function of E_{polyox}, where the symbols represent the solution to Eqn.(2.21) and Eqn.(2.22) solved using iterations, and the solid line represents the same after the usage of above approximation. We observe that the curves overlay on each other and can conclude that the above approximation is valid in this case.

2.2.2 Surface Potential / Charge Equation

Using the gradual channel approximation ignoring the electric field component along the channel in the device, the 1-D Poisson's equation in a cylindrical coordinate system can be written as follows,

$$\frac{1}{r} \cdot \frac{d}{dr}\left(r\frac{d\psi}{dr}\right) = \frac{qN_a}{\epsilon_{si}} + \frac{qN_a}{\epsilon_{si}} \exp\left(\frac{\psi - V_{ch} - 2\phi_b}{V_t}\right) \tag{2.24}$$

where $\psi(r)$ is potential at a distance r from the center of the cylindrical channel, q is the electronic charge, ϵ_{si} is the permittivity of silicon, N_a is the channel doping, V_t is the thermal voltage and $V_{ch}(y)$ is the channel potential at a distance y from the source. The concentration of holes is ignored in this case and will be accounted for later in this paper. The first term denotes the depletion charge while the second term denotes the mobile charges due

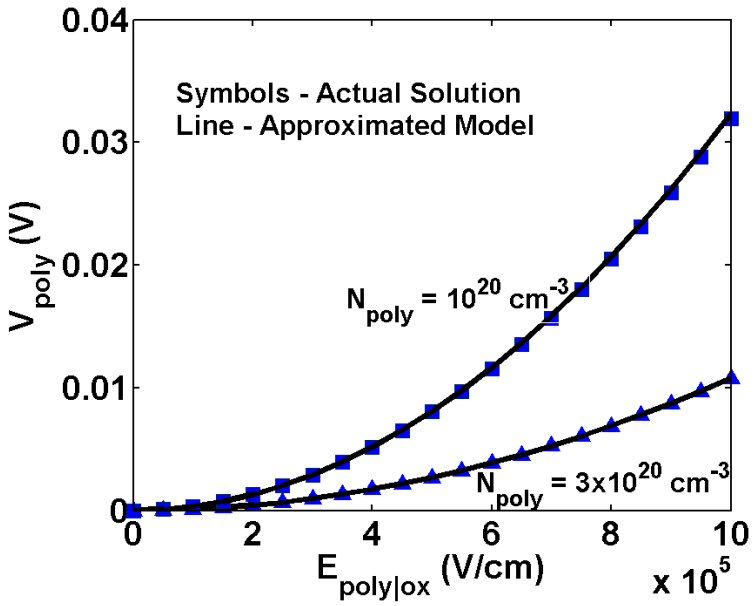


Figure 2.2: Voltage drop in the polysilicon gate obtained from the iterative numerical solution compared against approximate solution for a given electric field at the gate and oxide interface

to band-bending in the silicon. There is no known analytic solution to Eqn.(2.24). However a closed form solution for the electric field at the surface can be obtained under weak-inversion condition which shall form out first educated guess, [28].In weak inversion one can neglect the contribution of mobile carriers. In this case the inversion charge density can be written as $n_0 = N_a \cdot e^{\frac{\psi_0 - V_{ch}(y) - 2\phi_B}{V_t}}$, where n_0 and ψ_0 are concentration and potential at the center, $r = 0$. The field in the channel under weak inversion can then be obtained by integrating Eqn.(2.24) once w.r.t. r as follows,

$$\frac{d\psi}{dr} = \left(\frac{Q_{dep} + Q_i}{R\epsilon_{si}}\right) r. \tag{2.25}$$

$Q_{dep} = q.N_a.R/2$ and $Q_i = q.n_0.R/2$ are the depletion and inversion charge per unit gate area. Using the value of rdr from Eqn.(2.25) to change the variable of integration for the mobile charge term in Eqn.(2.24) from r to ψ, we perform a second iteration of the 1-D Poisson equation. Integrating Eqn.(2.24) once again, the electric field can be re-derived as follows

$$\frac{d\psi}{dr}(r = R) = \frac{Q_{dep}}{\epsilon_{si}} + \frac{2Q_{dep}}{Q_{dep} + Q_i}\frac{V_t}{R}\exp\left(\frac{\psi_s - V_{ch} - 2\phi_b}{V_t}\right)\left[1 - \exp\left(\frac{\psi_0 - \psi_s}{V_t}\right)\right] \tag{2.26}$$

where ψ_s is the potential at the surface.

Using Gauss law (the electric field at the surface is due to the sum contributions of both the depletion and the inversion charges) the following equation can be directly obtained from

Eqn.(2.26).

$$\frac{Q_i}{\epsilon_{si}} = \frac{2Q_{dep}}{Q_{dep}+Q_i}\frac{V_t}{R}\exp\left(\frac{\psi_s - V_{ch} - 2\phi_b}{V_t}\right)\left[1-\exp\left(\frac{\psi_0-\psi_s}{V_t}\right)\right] \tag{2.27}$$

Taking the logarithm of Eqn.(2.27), we have

$$\frac{\psi_s}{V_t} = \frac{V_{ch}+2\phi_b}{V_t} + ln\frac{R}{2\epsilon_{si}V_t} - ln\left[1-\exp\left(\frac{\psi_0-\psi_s}{V_t}\right)\right] + ln(Q_i) + ln\left(1+H\cdot\frac{Q_i}{C_{ox}V_t}\right) \tag{2.28}$$

where C_{ox} is the oxide capacitance per unit area for an oxide thickness, t_{ox} and is given by $C_{ox} = \epsilon_{ox}/(R.ln(1+t_{ox}/R)) = \epsilon_{ox}/EOT$ and H $= V_t.C_{ox}/Q_{dep}$. From the continuity of electric field flux at the silicon - gate oxide interface we can write

$$(V_{gs} - V_{fb} - V_{poly} - \psi_s) = V_{ox} \tag{2.29}$$

where V_{gs} is the gate to source voltage and V_{fb} is the flat-band voltage of the MOSFET under consideration. From Gauss law, it is easy to see that $V_{ox} = (Q_{dep}+Q_i)/C_{ox}$. Using Eqns.(2.23),(2.28) and (2.29) we can write the surface potential equation as follows.

$$V_{gs} - V_{th0} - \Delta V_{th} - V_{ch} = c_{poly}\cdot\left(\frac{Q_i+Q_{dep}}{C_{ox}}\right)^2 + \frac{Q_i}{C_{ox}} + V_t\cdot ln\frac{Q_i}{V_tC_{ox}} + V_t\cdot ln\left(1+H\cdot\frac{Q_i}{C_{ox}V_t}\right) \tag{2.30}$$

where

$$V_{th0} = V_{fb} + 2\phi_b + \frac{Q_{dep}}{C_{ox}} - V_t\cdot ln\left(\frac{2qN_a\epsilon_{si}}{V_tC_{ox}^2}\right)$$

$$\Delta V_{th} = -V_t\cdot ln\left[\frac{V_tC_{ox}}{qN_aR}\cdot\left(1-exp\left(-\frac{qN_aR^2}{4\epsilon_{si}V_t}\right)\right)\right]$$

It is intuitive to note that the second, third and fourth terms of the R.H.S. correspond to the strong, weak and moderate inversion regions respectively. The first term is representative of the correction for poly-depletion. Note that in the above equation simply setting $c_{poly} = 0$ is sufficient to convert the equation for the metal gate case. It is assumed the device operates in fully-depleted mode irrespective of the doping. At high gate voltages, the inversion charge tends to move closer to the silicon-oxide interface and the behavior deviates to a normal bulk MOSFET like operation away from the volume inversion that is observed in sub-threshold operation of multi-gate MOSFETs. Setting the value of parameter $H = exp(-\Delta V_{th}/V_t)$, instead of $H = V_t.C_{ox}/Q_{dep}$ makes the surface potential equation similar to what is obtained for bulk MOSFETs at strong inversion [28]. For high Q_i, inversion charge the third term can be approximated as $ln(H)+ln(Q_i)$. The $ln(H)$ term tends to cancel away the extra threshold voltage shift ΔV_{th} (due to volume inversion) under strong inversion. This type of correction within the surface potential equation itself, as opposed to other extraneous means such as shifting threshold voltage and modifying inversion region of current characteristics leads to

accurate prediction of trans-conductance (g_m) and its derivatives (g'_m, g''_m etc.) which are important for analog/RF models. The term ΔV_{th} also contains another simple but accurate approximation for $\psi_s - \psi_0$ present in Eqn.(2.28).

Following the method outlined for solving the above equation for the intrinsic case in [33], Eqn.(2.30) can be rewritten as follows,

$$\frac{V_{gs} - V_{th0} - \Delta V_{th} - V_{ch}}{2V_t} = r2 \cdot (z + z_{dep})^2 + r1 \cdot z + 0.5 \cdot ln(z) + 0.5 \cdot ln(1 + c_{dop} z) \quad (2.31)$$

where

$$z = \frac{Q_i}{2V_t C_{ox}}; r1 = \frac{2\epsilon_{si}}{RC_{ox}}; r2 = 2V_t c_{poly} r1^2; z_{dep} = \frac{Q_{dep} R}{Q_0}; Q_0 = \frac{4V_t \epsilon_{si}}{R}; c_{dop} = 2r1 \cdot H$$

It was observed that the relative numerical behavior of the individual terms in Eqn.(2.31) was very similar to the one used in [33] for surrounding gate MOSFETs. Hence the same initial guess and 3^{rd}-order Householders method can be extended to this case too [32]. In the above equation both V_{ch} and z are function of y i.e. the distance from the source along the channel. At an arbitrary position in the channel y, from the solution of z for given V_{ch}, the surface potential can be obtained from,

$$\psi_s = V_{ch} + V_t \cdot ln(z) + V_t \cdot ln(1 + c_{dop} z) \quad (2.32)$$

The inversion charge at this y is given by,

$$Q_i = \frac{4\epsilon_{si} V_t}{R} z = Q_0 \cdot z \quad (2.33)$$

and the voltage drop in the poly-silicon gate would be

$$V_{poly} = 2V_t r_2 (z + z_{dep})^2 \quad (2.34)$$

For the purpose of formulating drain current and charge-voltage relationship Eqn.(2.31) is solved at the source end ($V_{ch} = V_s$) and the drain end ($V_{ch} = V_d$) to obtain z_s ,z_d, and the corresponding inversion charges (Q_{is}, Q_{id}).

2.2.3 Drain Current Equation

Given that we now have the inversion charges at the source side and the drain side from solving the SPE, an analytic derivation of the drain current can be made. We will assume constant mobility, μ for simplicity. The bias dependence of the mobility will however have to be included in a full-fledged compact model. The drain current in the channel is given by,

$$I_{ds}(y) = \mu \cdot 2\pi R \cdot Q_i(y) \frac{dV_{ch}(y)}{dy} \quad (2.35)$$

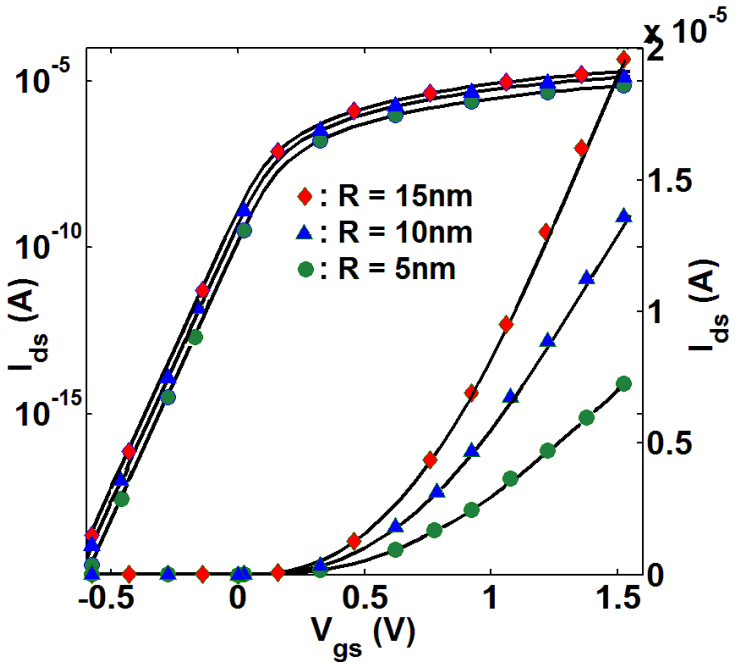


Figure 2.3: Drain current for a long channel FET ($L = 10\,\mu$m) with different channel radii showing scalability of the model. (Symbols : TCAD, Lines : Model), $t_{ox} = 2$ nm

Using Eqn.(2.30), (Note that $c_{dop}/Q_0 = H/(V_t C_{ox})$) dV_{ch}/dy can be replaced in terms of $Q_i(y)$ and dQ_i/dy by

$$\frac{dV_{ch}}{dy} = -\left(2c_{poly}\frac{Q_i + Q_{dep}}{C_{ox}^2} + \frac{1}{C_{ox}} + \frac{V_t}{Q_i} + V_t\frac{c_{dop}/Q_0}{1 + c_{dop}Q_i/Q_0}\right)\cdot\frac{dQ_i}{dy} \tag{2.36}$$

Assuming current continuity along the channel and integrating the above equation from source to drain i.e. from Q_{is} to Q_{id}, the closed form expression for the drain current is,

$$I_{ds} = \mu\cdot\frac{2\pi R}{L}\cdot\left[T_{poly} + \left(\frac{Q_{is}^2 - Q_{id}^2}{2C_{ox}}\right) + 2V_t(Q_{is} - Q_{id}) + \frac{Q_0 V_t}{c_{dop}} ln\left(\frac{Q_0 + c_{dop}Q_{id}}{Q_0 + c_{dop}Q_{is}}\right)\right] \tag{2.37}$$

where

$$T_{poly} = 2c_{poly}\frac{Q_{is}^3 - Q_{id}^3}{3C_{ox}^2} + \frac{2c_{poly}Q_{dep}}{C_{ox}}\left(\frac{Q_{is}^2 - Q_{id}^2}{2C_{ox}}\right) \tag{2.38}$$

The term T_{poly} can be visualized as a correction term for poly-depletion which would reduce to zero in the case of a metal gate ($c_{poly} = 0$). Device simulations were performed for a long channel transistor with $L = 10\,\mu$m, in order to decouple any short channel effects [34]. SiO_2 of 2 nm thickness was used as the gate oxide. A mid-gap metal gate unless otherwise mentioned to be poly-silicon gate with appropriate doping was used as the gate material. Fig.2.3 shows the scalability of the model as the drain current matches well to Sentaurus TCAD simulations for various channel radii. This model also retains the same predictability

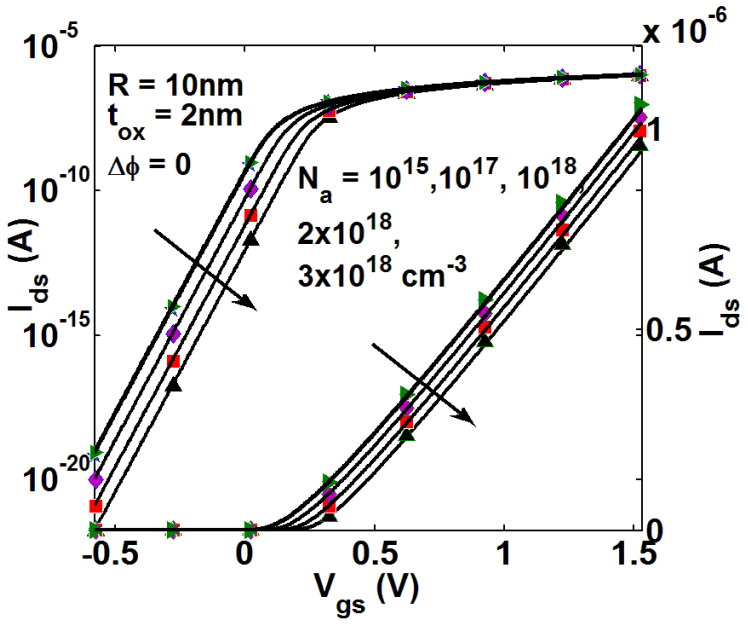


Figure 2.4: Drain current characteristics for a lightly doped and a heavilly doped long channel transistor. ($L = 10\,\mu$m) with a 2 nm gate oxide. This model captures doping in the channel as well. (Symbols : TCAD, Lines : Model)

as the model given by [28] for channel doping, Fig.2.4. This model was additionally extended to include poly-depletion which is well captured for different polysilicon doping values, see inset of Fig.2.5. The drain current characteristics as a function of drain bias voltage for different gate bias values also agrees well with device simulations, Fig.2.6. Adapting the new expression for H in the surface potential equation to align it with bulk-MOSFET like behavior in strong inversion reduced the error in the drain current from about $6-8\%$ down to $< 2\%$ error in the moderate and strong inversion region when compared to TCAD, for example, see Fig.2.7. This was verified for a wide range of channel radii (up to 100 nm) and for a large channel doping (for that radius) such that the fully-depleted channel approximation is still valid in the bias range of interest.

2.2.4 Charge and Capacitance Models

The terminal charges are required to compute the capacitances of the transistor. The capacitances are important while performing transient (time-varying) or AC (small signal) analysis of a circuit.

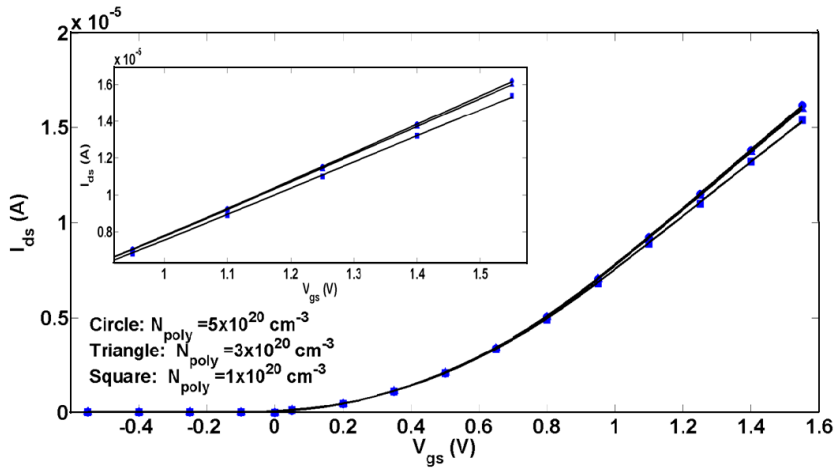


Figure 2.5: $I_{ds} - V_{gs}$ characteristics for three different values of gate polysilicon doping. (Symbols : TCAD, Lines : Model) with $L = 10\,\mu$m,$R = 10\,$nm and $t_{ox} = 2\,$nm

2.2.4.1 Inversion Region

The total charge on the gate is obtained by integrating the mobile inversion charge in the channel with the appropriate sign.

$$Q_{g,inv} = 2\pi R \int_0^L Q_i dy \tag{2.39}$$

Using Eqn.(2.35) the integration w.r.t variable y can be replaced by V_{ch} which is the channel voltage at the position y. Assuming current continuity, i.e. $I_{ds}(y) = I_{ds}$, the above integral can be rewritten as,

$$Q_{g,inv} = (2\pi R)^2 \frac{\mu}{I_{ds}} \int_0^{V_{ds}} Q_i^2 dV_{ch} \tag{2.40}$$

For the purpose of evaluating this integral we shall make an assumption as follows

$$I_{ds,CV} = \mu \cdot \frac{2\pi R}{L} \cdot \left[T_{poly} + \left(\frac{Q_{is}^2 - Q_{id}^2}{2C_{ox}} \right) + 2V_t(Q_{is} - Q_{id}) \right] \tag{2.41}$$

$$\frac{dV_{ch}}{dy} = -\left(2c_{poly} \frac{Q_i + Q_{dep}}{C_{ox}^2} + \frac{1}{C_{ox}} + \frac{V_t}{Q_i} \right) \cdot \frac{dQ_i}{dy} \tag{2.42}$$

The above assumption is valid in the inversion region as can be seen in the Fig.2.8 that compares $I_{ds,CV}$ and I_{ds}. The error in the inversion region is negligible. This would introduce errors in the capacitances in sub-threshold region. However values of intrinsic capacitances are relatively small and the parasitic (overlap and fringe components) capacitances dominate in this region. As a part of a compact model one could always introduce tuning parameters to capture back the accuracy in these regions. This assumption also tends to throw out

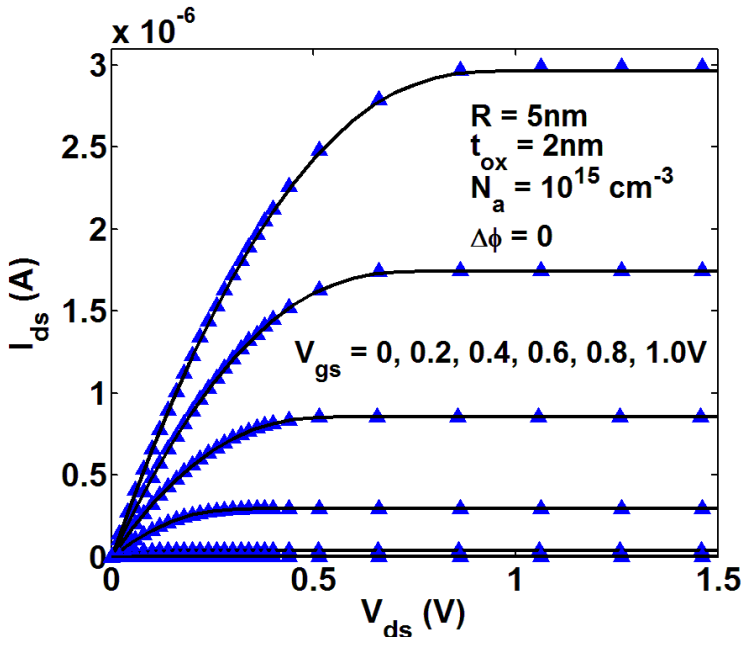


Figure 2.6: $I_{ds} - V_{ds}$ characteristics for a long channel device with $L = 10\mu\,\mathrm{m}$,$R = 5\,\mathrm{nm}$ and 2 nm oxide thickness. (Symbols : TCAD, Lines : Model)

computationally expensive terms (like multiple number of logarithm terms) from the final analytic charge expressions. Using Eqn.(2.42) in Eqn.(2.40), and canceling the common term $(Q_{is} - Q_{id})$ in the numerator and denominator, the analytic expression for the gate charge can be written as follows,

$$Q_{g,inv} = \frac{2\pi R \cdot L}{i_{cv}} \left[\frac{c_{poly}}{2C_{ox}^2}(Q_{is}^2 + Q_{id}^2)(Q_{is} + Q_{id}) + V_t(Q_{is} + Q_{id})... \right. \\ \left. + \left(1 + \frac{2c_{poly}Q_{dep}}{C_{ox}}\right) \cdot \left(\frac{Q_{is}^2 + Q_{is}Q_{id} + Q_{id}^2}{3C_{ox}}\right) \right] \tag{2.43}$$

where

$$i_{cv} = \frac{2c_{poly}}{3C_{ox}^2}(Q_{is}^2 + Q_{is}Q_{id} + Q_{id}^2) + \left(1 + \frac{2c_{poly}Q_{dep}}{C_{ox}}\right) \cdot \left(\frac{Q_{is} + Q_{id}}{2C_{ox}}\right) + 2V_t$$

Factoring out the common term $(Q_{is} - Q_{id})$ prevents the equation from non-convergence or discontinuity close to $V_{ds} = 0\,\mathrm{V}$.

The drain side charge can be estimated by using the Ward-Dutton charge partitioning scheme [35] using the following integral,

$$Q_d = -2\pi R \int_0^L \frac{y}{L} Q_i dy \tag{2.44}$$

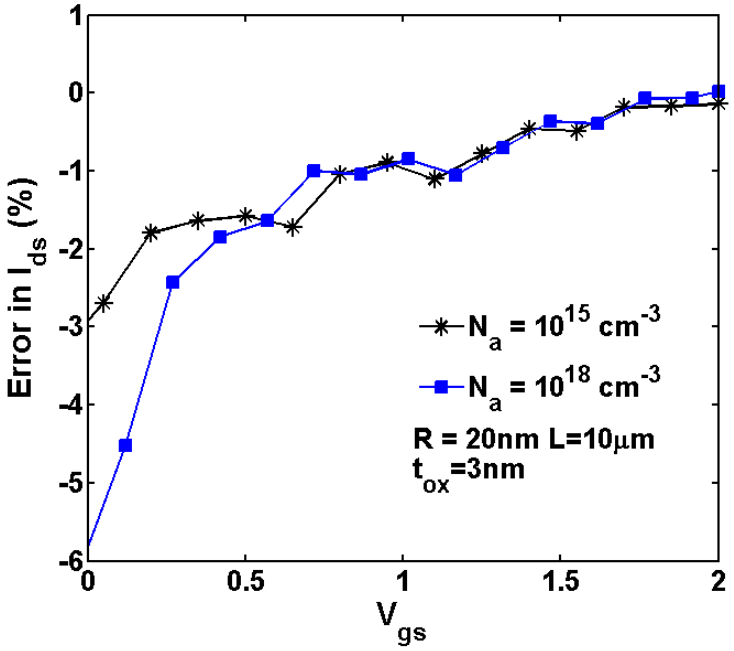


Figure 2.7: Percentage error in drain current versus gate voltage is below 2% in the moderate and strong inversion region for a wide range of channel radii and channel doping when compared to TCAD

The current continuity can be used to replace y/L by $I_{ds,CV}(y)/I_{ds,CV}$, where $I_{ds,CV}$ is the Eqn.(2.41) with Q_{id} replaced by Q_i the inversion charge density at position y along the channel. Following a similar approach as for the gate charge, the integral can be re-written as

$$Q_d = -\frac{(2\pi R)^3 \mu^2}{L \cdot I_{ds,CV}^2} \int_0^{V_{ds}} I_{ds,CV}(y) \cdot Q_i^2 dV_{ch} \tag{2.45}$$

The above integral together with Eqn.(2.42) would give rise to an analytic expression for the

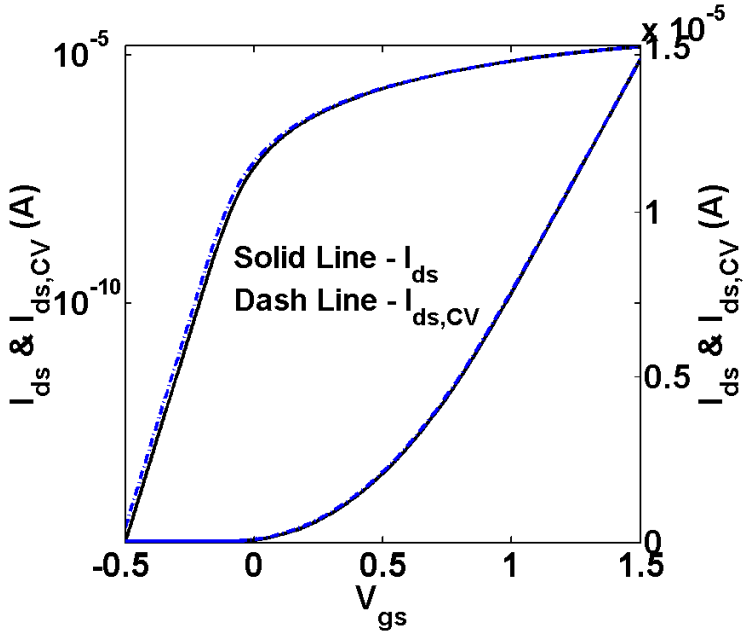


Figure 2.8: Comparison of the actual drain current and the approximation used for charge model showing negligible error in inversion

drain side charge as follows,

$$Q_d = -\frac{2\pi R \cdot L}{i_{cv}^2}\left[\frac{1}{30C_{ox}^2}\cdot(3Q_{id}^3+6Q_{id}^2Q_{is}+4Q_{id}Q_{is}^2+2Q_{is}^3)+\frac{2}{3}\cdot V_t^2\cdot(2Q_{id}+Q_{is})\ldots\right.$$
$$+\frac{V_t}{12C_{ox}}\cdot\left(1+\frac{2c_{poly}Q_{dep}}{C_{ox}}\right)\cdot(9Q_{id}^2+10Q_{id}Q_{is}+5Q_{is}^2)\ldots$$
$$+\frac{c_{poly}^2}{21C_{ox}^4}\cdot(4Q_{id}^5+8Q_{id}^4Q_{is}+12Q_{id}^3Q_{is}^2+9Q_{id}^2Q_{is}^3+6Q_{id}Q_{is}^4+3Q_{is}^5)\ldots$$
$$+\frac{c_{poly}}{36C_{ox}^3}\cdot(10Q_{id}^4+20Q_{id}^3Q_{is}+21Q_{id}^2Q_{is}^2+14Q_{id}Q_{is}^3+7Q_{is}^4)\ldots$$
$$\left.+\frac{c_{poly}}{5C_{ox}^2}\cdot V_t\cdot(8Q_{id}^3+11Q_{id}^2Q_{is}+14Q_{id}Q_{is}^2+7Q_{is}^3)\right] \tag{2.46}$$

The source side charge using appropriate sign is given by,

$$Q_s = -Q_{g,inv} - Q_d \tag{2.47}$$

In case of devices where the channel is still in electrical contact with the bulk, the bulk charge should also be taken into account for. The bulk charge is then given by,

$$Q_b = -2\pi R \cdot L \cdot Q_{dep} \tag{2.48}$$

For charge neutrality and equal and opposite charge is assigned to the gate, and the total gate charge becomes,

$$Q_g = Q_{g,inv} + 2\pi R \cdot L \cdot Q_{dep} \tag{2.49}$$

However in accounting for the bulk-charge we still adhere to the fully-depleted approximation (i.e. all of the channel region gets depleted for gate voltages slightly above flat-band condition), and hence assume a constant bulk-charge.

2.2.4.2 Accumulation Region

In the case of a thick (large radius), vertical cylindrical gate structure, it is possible for the bottom region doping to not fully isolate the channel from the bulk substrate. One can draw parallels of this multi-gate structure to that of a FinFET on bulk-substrate. In these cases there is pathway for holes to approach the channel interface from the bulk-contact and the device would exhibit accumulation capacitance like that of a bulk-planar transistor. For other small radius or horizontal cylindrical gate structures, there will be no accumulation as the n-doped source and drain region isolate the channel and also provide negligible to no supply of holes into the channel. Until now we have disregarded the holes for simplicity and increased viability of creating computationally inexpensive analytic expressions for the inversion region. We will begin with a Poisson equation that considers just holes (neglects electrons or bulk charge) as follows,

$$\frac{1}{r} \cdot \frac{d}{dr}\left(r\frac{d\psi}{dr}\right) = -\frac{qn_i}{\epsilon_{si}} \exp\left(\frac{-\psi}{V_t}\right) \tag{2.50}$$

where n_i is the intrinsic carrier concentration. The hole quasi-fermi level is almost constant throughout the device and hence the source or drain voltage has no influence to the above equation. Imposing the following boundary conditions that,

$$\frac{d\psi}{dr}(r=0) = 0, \psi(r=R) = \psi_s \tag{2.51}$$

the solution to Eqn.(2.50) is easily obtained by observing the fact that the equation is the same as that for inversion side modeling for the intrinsic doping case for CG transistors, except that the sign of the surface potential is reversed. The solution for the surface is given by [27],

$$\psi(r) = -\frac{kT}{q} ln\left(\frac{-8B \cdot kT \cdot \epsilon_{si}}{q^2 n_i (1+Br^2)^2}\right) \tag{2.52}$$

where B the constant of integration, can be obtained by using the boundary condition for the potential at the surface. We will leave it to the reader to construct a mirror surface potential equation along the lines of [27]. If the solution to the surface potential equation was ψ_{acc}, the gate and the body charge are given by,

$$Q_{g,acc} = 2\pi R \cdot L \cdot (V_{gb} - V_{fb} - \psi_{acc}) \tag{2.53}$$

$$Q_{b,acc} = -Q_{g,acc} \tag{2.54}$$

where V_{gb} is the gate to body voltage. The total gate and body charge for the device would just be the sum of the inversion and the accumulation region charges. The above separate unipolar treatment has two advantages. One, it avoids the requirement of first constructing single piece complex solutions (when taking both electrons and holes into consideration) and then later making assumptions to include them in a compact model framework. Two, it gives the flexibility to fit to real device data (which is one of the goals) with a separate set of device parameters like oxide thickness and flat-band voltage. The minor caveat is the inaccuracy observed in the model for a region close to flat-band voltage due to the separate treatment. But we predict that parasitic capacitances will dominate this area and thus minor inaccuracies in the intrinsic model is acceptable.

Once the charges have been computed it is easy to evaluate the quasi-static capacitances by differentiating the charge with the node voltages. This can be written as

$$C_{ij} = \begin{cases} \frac{dQ_i}{dV_j} & i \neq j \\ -\frac{dQ_i}{dV_j} & i = j \end{cases} \tag{2.55}$$

From conservation of charge neutrality it should be noted that,

$$\sum_i C_{ij} = \sum_j C_{ij} = 0 \tag{2.56}$$

Though it is not recommended to use a lowly doped polysilicon for gate as it would increase EOT significantly, a transistor with polysilicon doping of $10^{20}\,\mathrm{cm}^{-3}$ was used for illustrative purposes to show that the model captures the poly-depletion phenomenon well. A long channel with short source drain regions were used for simulations reduce effect of parasitic capacitances. Figs2.9 and 2.10 shows that the analytic model is in excellent agreement with TCAD simulation results without the use of any fitting parameters. The core model is symmetric as indicated by the equality of capacitances at $V_{ds} = 0\,\mathrm{V}$ (i.e. $C_{ij} = Cji$) in the capacitance versus drain voltage plot, Fig.2.10. Fig.2.11 shows that the accumulation capacitance has also been captured well by the simple model. The first and second derivatives of C_{gg} (obtained by the addition of the contribution of $Q_{g,inv}$ and $Q_{g,acc}$) are shown in Fig.2.12. They have sufficient smoothness though the inversion and accumulation regions were treated separately and added up.

2.3 Model Verification and Quality Assurance

In order to be able to fit to real silicon data, it requires more than just what has been presented above. For this purpose plethora of other details were added to the model. These were adapted from the BSIM-CMG model for FinFETs [16, 36] with appropriate modifications for the cylindrical geometry. These include bias independent features like temperature

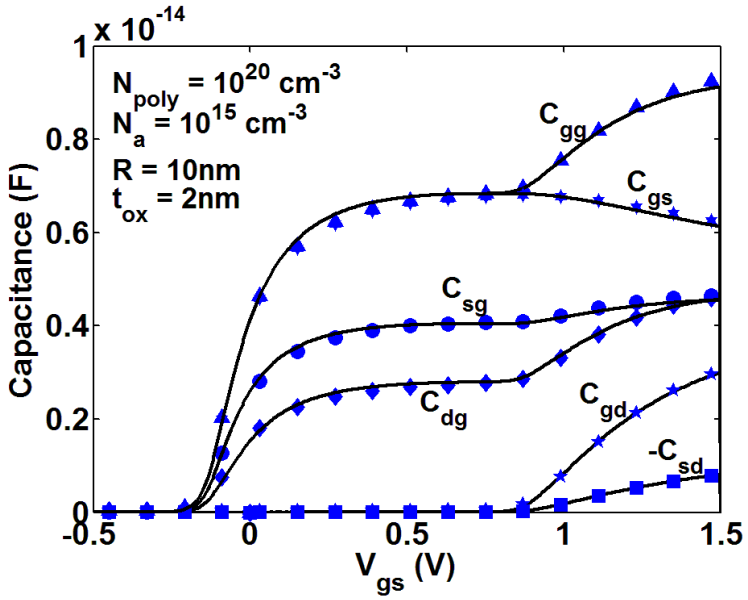


Figure 2.9: Capacitances calculated using the Q-V model (Lines) developed including poly-depletion as a function of gate voltage, $V_{ds} = 1\,\mathrm{V}$ shown for a low polysilicon doping overlayed against those obtained from TCAD (Symbols) simulations. The model agrees well with simulated results.

scaling, various bias dependent features such as mobility, source-drain resistance, channel length modulation and velocity saturation, etc. A scaling length based short channel threshold voltage roll-off and sub-threshold swing degradation was also included [37, 38]. Leakage currents such as gate tunneling leakage, gate induced source/drain leakage (GIDL/GISL) and junction leakage currents, etc. have been modeled as well. Models to capture parasitic capacitances have also been included. Additional parameters / variability handles are also provided to account for any process induced variations. This model was written in Verilog-A language, [39]. The model was shown to successfully describe measured drain current, transconductance and output conductance of silicon device data. For more information regarding these additional effects and models the reader should refer to Appendix A.

The cylindrical gate FETs were fabricated on bulk silicon wafers. The channel was vertically oriented with the source and drain contacts at the top and bottom while a polysilicon gate wraps around the channel. The vertical channel was 150 nm tall with a diameter of 80 nm and was doped adequately to ensure that that device was fully depleted for all operating bias conditions. The gate oxide was 3 nm thick. These devices were one of the first known multi-gate devices to exhibit asymmetric drain current data when the source and drain were swapped between top and bottom. For this purpose we used different source and drain resistance values. Additionally an asymmetric drain induced barrier lowering (DIBL) model was also introduced to capture the asymmetric threshold voltage shift for high drain bias. From the saturation region drain current versus gate voltage plot in Fig.2.13, we

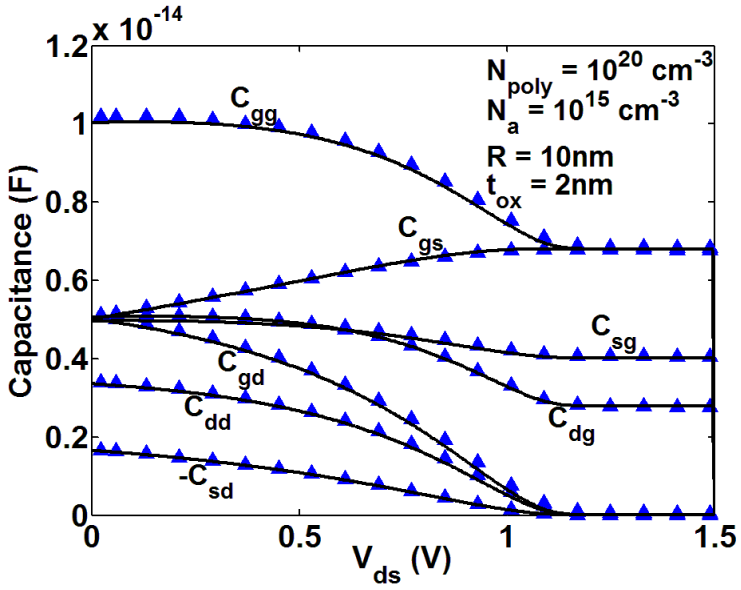


Figure 2.10: Capacitances calculated using the Q-V model (Lines) developed including poly-depletion as a function of drain bias voltage, shown for a low polysilicon doping overlayed against those obtained from TCAD (Symbols) simulations. The reciprocity of the capacitances at $V_{ds} = 0\,\mathrm{V}$ shows that the developed core model is inherently symmetric w.r.t. source and drain terminal charges.

note that this phenomenon has been well captured by this model. For clarity the rest of the plots show only data that corresponds to using the top of the channel as drain. The reverse condition has been well captured too. Fig.2.14 shows the fitting for drain current in both linear and saturation conditions from which one can deduce that short channel features like DIBL have been well modeled. The trans-conductance plot further validates the model, Fig.2.15. In Fig.2.16, drain currents in the linear region versus gate voltage for temperatures from $-25°\mathrm{C}$ to $125°\mathrm{C}$ are presented. This model also includes equations for temperature dependencies for various parameters. Fig.2.17 illustrate the drain current against drain voltage. This model has captured the channel length modulation and velocity saturation effects well as observed by excellent fit of model output conductance to data, Fig.2.18. The model passes all stringent quality assurance tests exhibiting convergence of circuit simulations for coupled ring-oscillators, sense-amplifiers and phase-locked loop etc. The model passes the Gummel test, Fig.2.19 and the AC Symmetry test, Fig.2.20(b) when a symmetric set of parameters are used for the source and drain side,[40].

2.4 Summary

A complete yet lean compact model for cylindrical/surround gate transistors have been developed. A surface potential equation that captures enhanced features such as poly-depletion

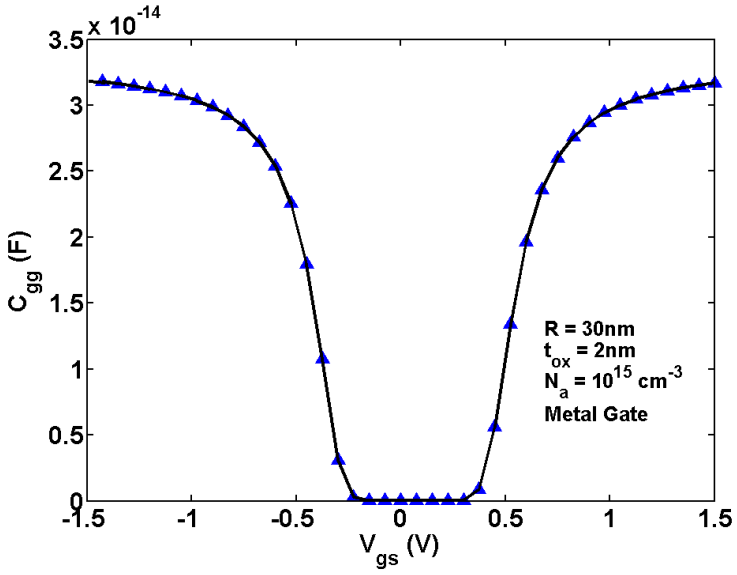


Figure 2.11: (a) Gate capacitance as a function of gate voltage obtained for a device exhibiting accumulation capacitance. (Lines:Model, Symbols:TCAD), $V_{ds} = 0\,\mathrm{V}$.

was outlined. Analytic expression for drain current was derived without any charge sheet approximation. This expression was shown to agree well with numerical device simulation results for various channel radii, and body doping up to $\approx 10^{18}\,\mathrm{cm}^{-3}$ without any fitting parameters. Expressions for terminal charges as a function of bias were derived and used to model the capacitance of the device. The capacitances and trans-capacitances were shown to be accurate for both metal and poly-silicon gates. Accumulation region capacitance was also modeled for devices whose channel might not be electrically isolated from the bulk contact or built over bulk silicon substrates. The core model was then expanded to a full-fledged SPICE model by including various sub-models for real device effects. The model was also verified against experimental data set for an asymmetric vertical channel cylindrical gate device technology.

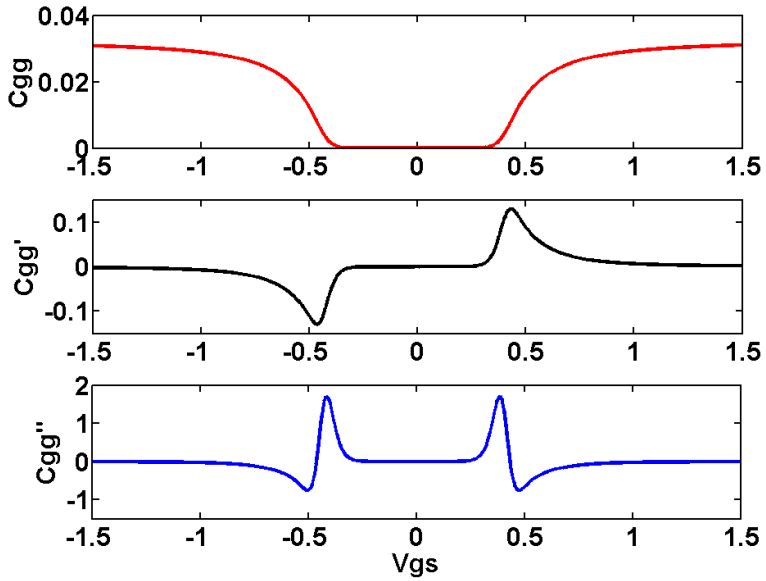


Figure 2.12: Derivatives of gate capacitance exhibiting model smoothness.

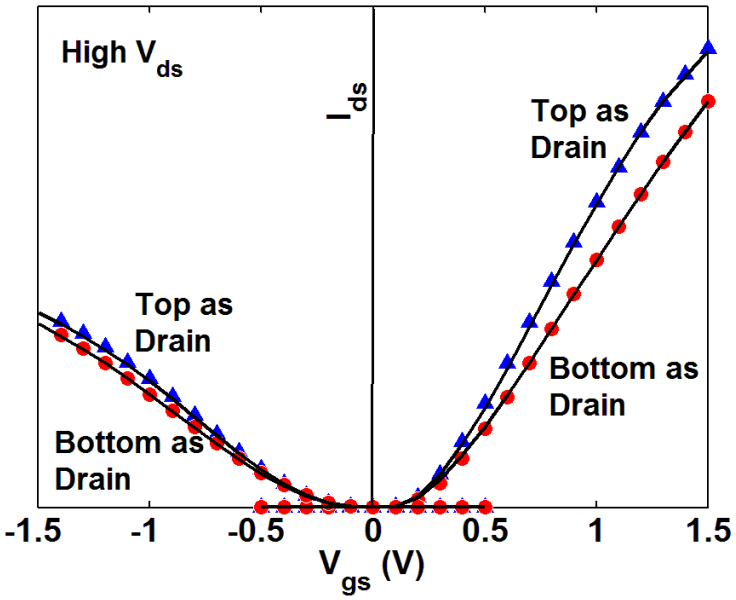


Figure 2.13: Linear I_{ds} for a low drain bias voltage shown for both top as drain (triangles) and bottom (circles) as drain conditions. The asymmetry in the device is captured well using an asymmetric source-drain resistance model. (Symbols :Measurement data, Lines : Model)

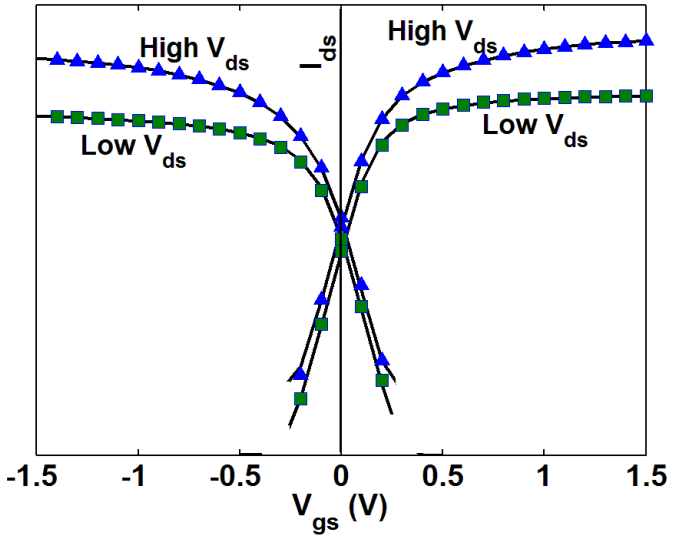


Figure 2.14: (a) Log I_{ds} as a function of gate voltage for high and low drain bias conditions. The subthreshold slope and DIBL induced threshold voltage shift have been modeled accurately for both NMOS and PMOS. (Symbols : Measurement data, Lines : Model)

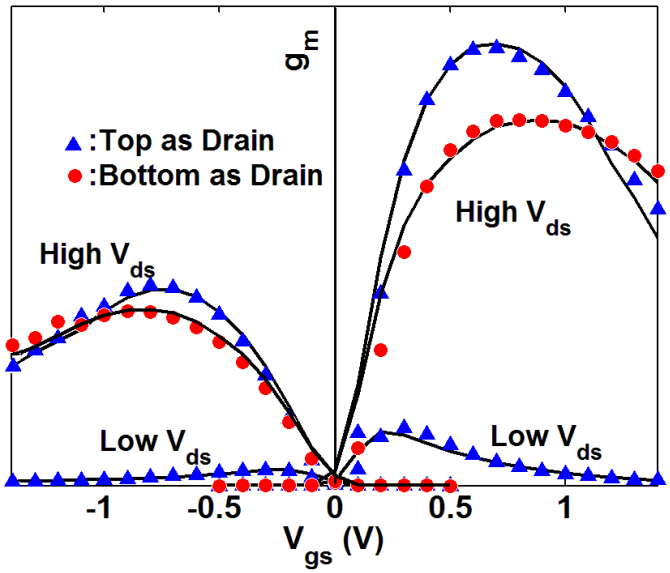


Figure 2.15: Transconductance (g_m) plot further proves that phenomenon like mobility degradation and velocity saturation have been modeled accurately. g_m for both top as drain and bottom as drain conditions are being shown at high V_{ds}. (Symbols : Measurement data, Lines : Model)

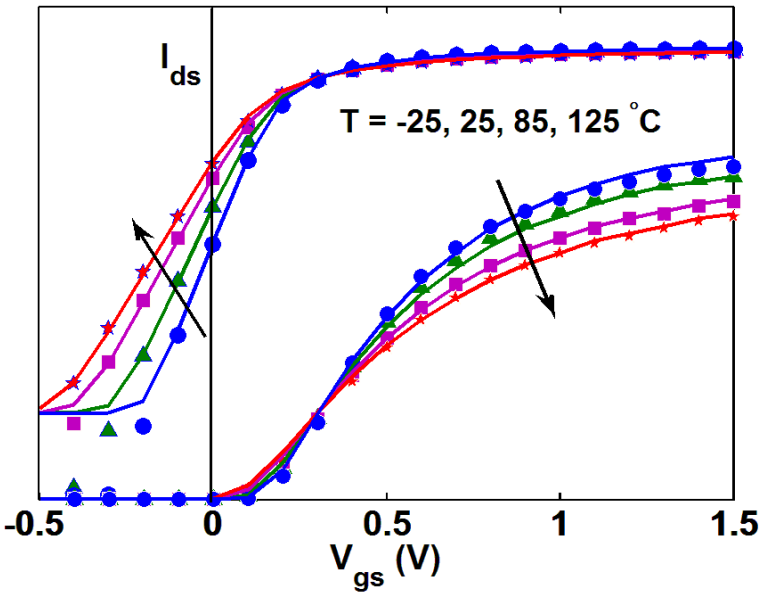


Figure 2.16: Temperature dependence of $I_{ds} - V_{gs}$ characteristics for $T = -25°C$ to $T = 125°C$ predicted by the model (Lines) agrees well with the measurement data (Symbols)

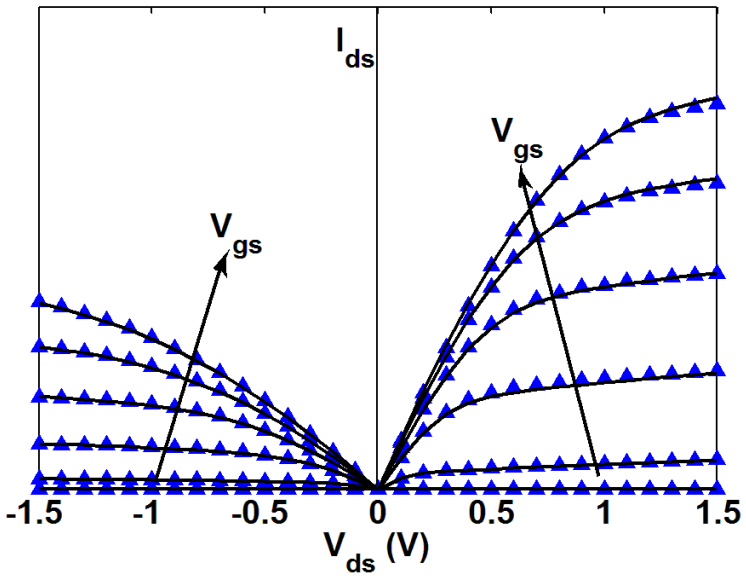


Figure 2.17: Drain current characteristics as a function of drain voltage for both NMOS and PMOS.(Symbols : Measurement data, Lines : Model)

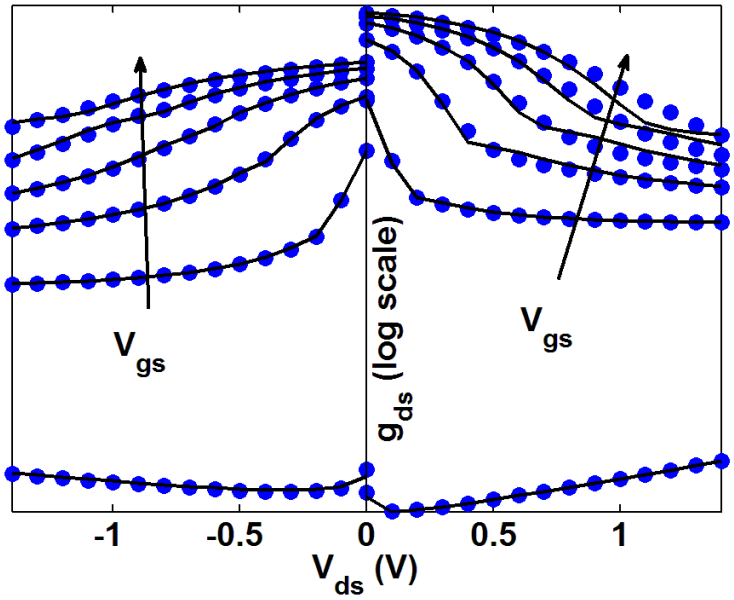


Figure 2.18: Output conductance (g_{ds}) overlay of model against measurement data. (Symbols : Measurement data, Lines : Model)

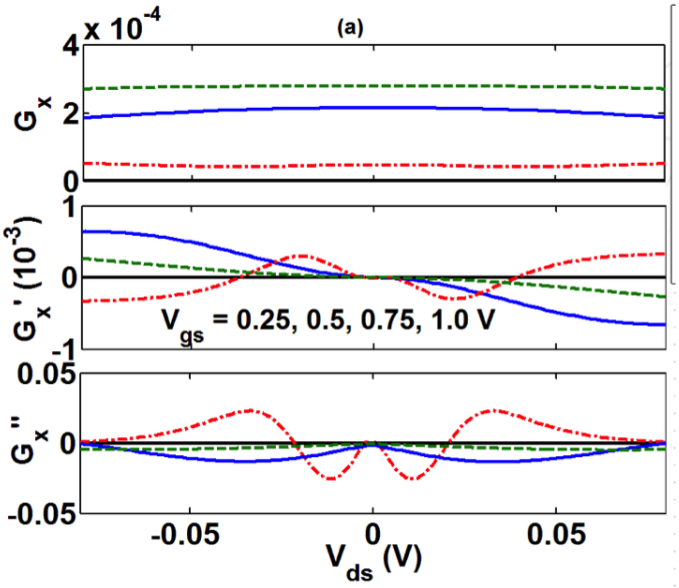


Figure 2.19: Gummel Test: 1^{st}, 2^{nd} and 3^{rd} derivatives of drain current are continuous and symmetric when a symmetric voltage is applied across the source and drain and a symmetric set of parameters is used in the model.

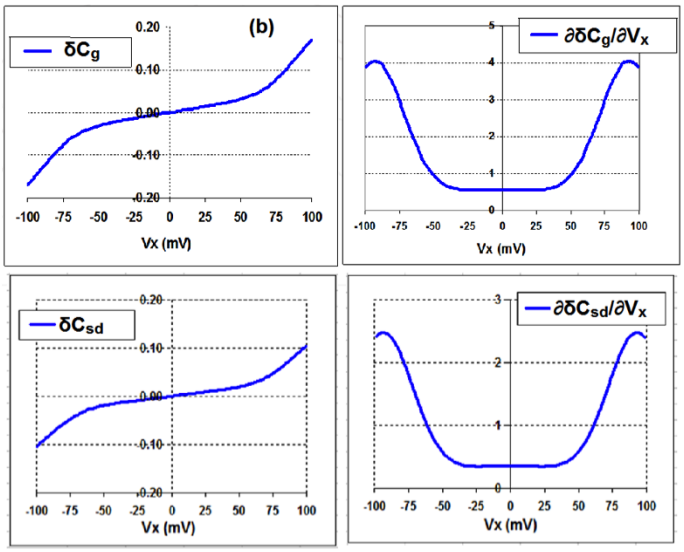


Figure 2.20: AC Symmetry: δ_{cg} and δ_{csd} and their derivatives show the continuity and the symmetry in the charge/capacitance model. For definitions see [40].

Chapter 3

3-D Device Effects Models for Multi-Gate FETs

Multi-gate transistor architecture differ from their planar transistor counterparts in an obvious way − multi-gate transistors are three-dimensional (3-D) in nature. The current transits from source to drain terminals through a 3-D channel. Some new real device effects not observed or dominant in planar transistor architecture arise. These device effects could be demarcated into two categories −

- Phenomenon whose effect can be mapped to their equivalent 2-D device effect. Effects such as mobility degradation, velocity saturation (trans-conductance), bias dependence of drain induced barrier lowering (output conductance) fall into this category. Regardless of which crystal plane is utilized as the channel of a transistor, mobility degradation happens due to the same underlying physics *viz.* Coulomb scattering, phonon scattering and surface roughness scattering. These have been well studied in planar transistor counterparts. For these effects we could take advantage of established industry standard models like BSIM4, PSP etc.
- Phenomenon whose effects cannot be mapped to their 2-D analogue. Short channel effects, quantum mechanical confinement in a 3-D channel, structural asymmetries, 3-D fringe capacitances and spreading resistances etc. fall into this category. The underlying physics that create these effects render them with dependence on device parameters that arise from the inherent 3-D architecture such as fin thickness, fin height, fin pitch and nanowire radius etc. The dependence on such parameters may not be obvious to be captured in a simple model and therefore a theoretical study of the same is required.

In this chapter we will look into a set of device phenomenon that fall into the latter category above. We will study the quantum mechanical confinement of the carriers in the 3-D channel and their effect on the overall gate capacitance. We will also look into

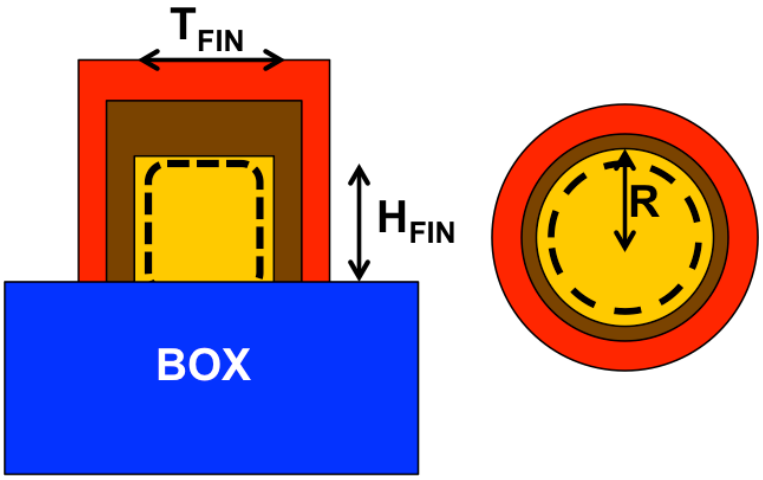


Figure 3.1: Channel cross-section of a FinFET on SOI substrate and a Gate-All-Around FET. The dotted lines represent the presence of the charge centroid being away from the oxide-channel interface due to QM the carriers in the channel.

a unique double-junction that arises in source / drain of FinFETs that under-go punch-through prevent implant (ground plane doping). In the Chapter 4 we will focus on vertical cylindrical/surround gate (CG) transistor architecture to study the impact of device asymmetry on its electrical characteristics. These effects require new models that are conducive to incorporating them back into existing multi-gate transistor models.

3.1 Quantum mechanical effects in Fins and Pillars

At dimensions below 20 nm of fin thickness in the case of double gate (DG) like structures or channel diameter in case of a cylindrical gate (CG) structure, the multi-gate devices exhibit both structural and electrical quantum mechanical (QM) confinement of carriers in the channel, Fig. 3.1. Structural confinement occurs due to the carriers being constrained between the two gates of a DG FET or within a cylindrical well region in a CG FET. From quantum mechanics we know this leads to two observable effects –
(a) threshold voltage shift due to band-gap widening and
(b) a reduced effective gate capacitance due to the centroid of the charge distribution in the channel being away from the gate-oxide channel interface [41],[42].
Moreover with increasing gate terminal bias electrical confinement begins to dominate. Electrical confinement arises even in a planar transistor due to carriers being trapped within a triangular potential well near the gate oxide - channel interface. As a result the charge centroid shifts towards the interface rendering a bias dependence to the gate capacitance in addition to the intrinsic device capacitance behavior.

In a circuit simulation environment (involving thousands of transistors) one cannot afford

to run computationally expensive Schrodinger-Poisson solvers to capture the QM confinement effect. A simple yet predictive compact model is required for this purpose. Our goal is to create a charge centroid model that can be integrated into charge based compact model for multi-gate FETs like BSIM-CMG [39]. Lixin Ge *et al.* have used a variational approach to solve Poisson and Schrodinger equation in an analytic self-consistent manner for a DG structure [42]. While this method delivers accurate and excellent insights of the device behavior, the final model is computationally expensive in light of requirement of industry standard compact models. Y.S. Wu and Pin Su have derived an analytic model from Poisson and Schrodinger equations for the charge density distribution in the channel for a CG structure [43]. The so derived expression for charge in the channel is a series summation of 20 terms and is not conducive for obtaining a simple expression for the charge centroid. J. B. Roldan et al. have developed an empirical model for charge centroid of a cylindrical gate FET based on TCAD simulations [44]. However certain key physical insights that render predictive nature across device's physical dimensions (fin thickness) are missing. In this section we develop a first principles physics based model for intrinsic channel CG and DG FET structure to capture both the geometry and the bias dependence of the charge centroid. We validate the model with TCAD simulations of extracted charge centroid and with linear regime FET gate capacitance simulations.

3.1.1 Charge Centroid Model

For this work the developed model was verified using TCAD device simulations wherein Schrodinger-Poisson-Continuity equations were solved self consistently [34]. 3-D simulations were performed for long ($L = 10\,\mu\text{m}$) intrinsic silicon channel DG and CG structures (silicon oxide gate dielectric and mid-gap work-function metal gate) assuming constant mobility and with abrupt source-drain junctions in order to decouple any short channel effects. The charge distribution within the channel was observed for different gate bias conditions at low drain terminal bias ($V_{ds} = 50\,\text{mV}$) and the charge centroid was numerically extracted from the profile.

The structural confinement component was extracted from the charge profile in the channel by varying the channel thickness. The device was biased in sub-threshold region ($V_{gs} - V_{fb}$ 100 mV) of operation. The charge profile for varying channel radii of a CG structure is shown in Fig.3.2. The plot has been normalized along the radius, R to illustrate that the charge centroid at low gate bias is a function of the radius. A similar statement holds true for DG structure too.

The electrical confinement component is extracted by sweeping the gate bias until strong inversion in the channel. In Fig.3.3, the charge density distribution for varying gate bias is shown. Although the total inversion charge in the channel increases in magnitude by various orders, we have normalized the charge for each gate bias to show the bias dependence of the charge centroid.

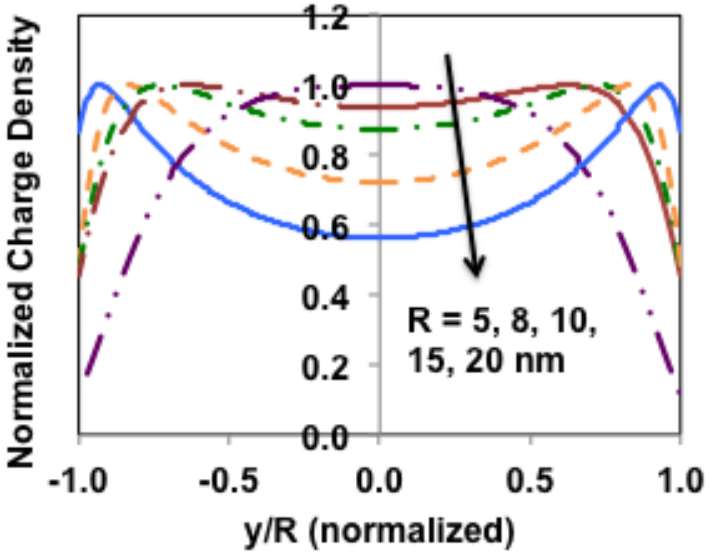


Figure 3.2: Simulated charge density distribution in the channel along the diameter of cylindrial gate (CG) FET strucutre in sub-threshold region for different channel radii. Lines: TCAD simulation

3.1.1.1 Geometry Dependence

For the DG device, we have observed that the centroid position at low gate bias is a function of the channel thickness, T_{FIN}. This can be explained considering two asymptotic cases a very thin fin and a thick fin. For a very thin fin, structural confinement of carriers would lead to an increased curvature of the wave-functions and hence an increased bound state's energies (energy goes as $1/T_{FIN}^2$). The carriers tend to occupy only the first sub-band. From the Schrodinger equation solution to an 1-D quantum well we know that the distribution of the charge in the channel goes as $\cos^2(\pi x/T_{FIN})$. The centroid for this distribution is $0.351 \times T_{FIN}$ away from the interface.

$$\text{Thin Fin: } \frac{T_{cen}}{T_{FIN}} = 0.5 - \frac{\int_0^{0.5} |\cos(\pi x)|^2 x.dx}{\int_0^{0.5} |\cos(\pi x)|^2 .dx} = 0.351 \tag{3.1}$$

For the case of a thick fin, quantization effect for structural confinement can be neglected and all the bound states merge to make a continuum of energies leading to a uniform charge distribution. The centroid would be $0.25 \times T_{FIN}$ away from interface.

$$\text{Thick Fin: } \frac{T_{cen}}{T_{FIN}} = 0.5 - \frac{\int_0^{0.5} 1.x.dx}{\int_0^{0.5} 1.dx} = 0.25 \tag{3.2}$$

For intermediate fin thicknesses the carriers occupy multiple sub-bands and it is not feasible to obtain a simple analytic expression for distribution or the centroid itself. We resort to an

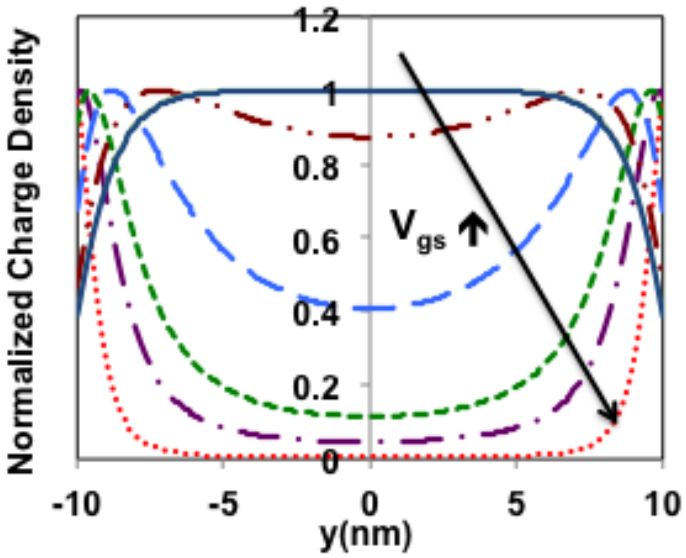


Figure 3.3: Tracking the change in charge density distribution in the channel for a R=10 nm CG FET for increasing gate bias from 0 V to 2 V. Lines: TCAD simulations

empirical function of the following form for this purpose,

$$\frac{T_{cen0}}{T_{FIN}} = 0.25 + (0.351 - 0.25).\exp\left(\frac{T_{FIN}}{T_0}\right) \tag{3.3}$$

where T_0 is an empirical parameter that was obtained from fitting to TCAD data.

Similar arguments can be made for a CG device with a circular cross-section. The Schrodinger equation solution for the charge distribution in a cylindrical quantum well under isotropic mass approximation is a Bessel function of first kind and order zero, $J_0(k.r)$, [45]. The wave-number k satisfies the condition $J_0(k.R) = 0$, i.e. the wave-function goes to zero at the boundary of the well $r = R$, where R is the radius of the channel. The asymptotic values for the centroid position are calculated to be $0.576 \times R$ and $0.334 \times R$ away from the interface for a thin and thick diameter channel respectively.

$$\text{Thick Diameter: } \frac{T_{cen}}{R} = 1 - \frac{\int_0^R 1.r^2.dr}{\int_0^R 1.r.dr} = 0.334 \tag{3.4}$$

$$\text{Thin Diameter: } \frac{T_{cen}}{R} = 1 - \frac{\int_0^R [J_0(k.R)]^2.r^2.dr}{\int_0^R [J_0(k.R)]^2.r.dr} = 0.576 \tag{3.5}$$

The empirical function assumes the form,

$$\frac{T_{cen0}}{R} = 0.334 + (0.576 - 0.25).\exp\left(\frac{R}{R_0}\right) \tag{3.6}$$

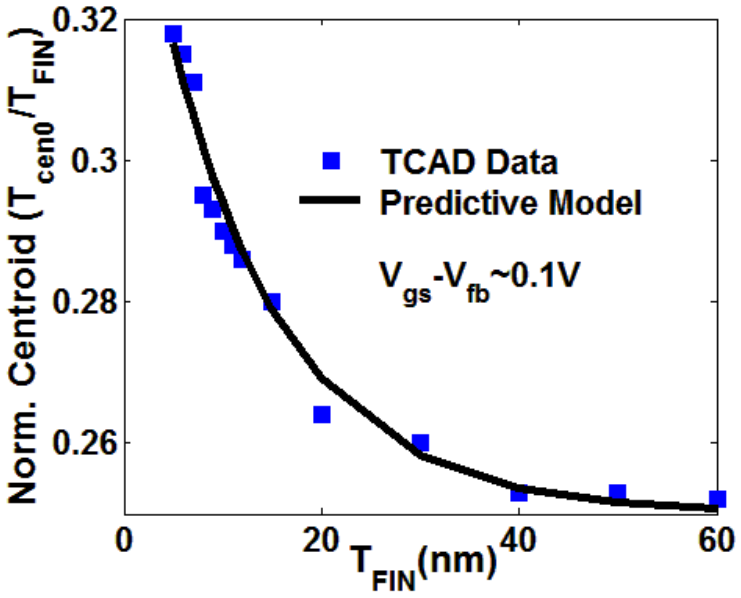


Figure 3.4: Charge centroid for low gate bias has a geometric dependence and takes values in between the asymptotic values calculated in Eqn.(3.3) for a DG FET with rectangular cross-section. Value extracted for $T_0 = 12\,\text{nm}$.

where R_0 is again an empirical parameter that was obtained from fitting to TCAD data.

In Figs.3.4 and 3.5, the empirical model obtained in Eqns.(3.3) and (3.6) are verified using TCAD data respectively. The centroid was extracted from the charge density distribution curves in the channel for low gate bias such as Fig.3.2 for CG structure. We can observe that not only the simple exponential function seems to agree well, but also the asymptotic values calculated from first principles matches well at the two extreme values of channel thicknesses. Also we note that these calculated asymptotic limit values for both CG and DG structures are *independent* of the channel material and orientation.

3.1.1.2 Bias Dependence

As the gate terminal bias is increased more carriers are introduced in the channel. The channel gets inverted and the band bending increases. With increase of band bending, the potential well gets narrower and hence carrier's centroid tends to move towards the gate-oxide channel interface, Fig.3.3. We continue to track this movement of charge centroid towards the interface with increasing gate bias using TCAD simulations. Instead of the usual plot between charge centroid and gate bias, we plot the charge centroid vs. the total charge in the channel to raise a subtle point in Fig.3.6. We observe (for a DG FET) that the charge centroid remains almost constant at the value calculated in Eqn.(3.3) for low gate bias until a critical charge density, Q_0 in the channel. Instead of resorting to a tedious analytic approach we recall the Debye screening length concept to derive an expression for

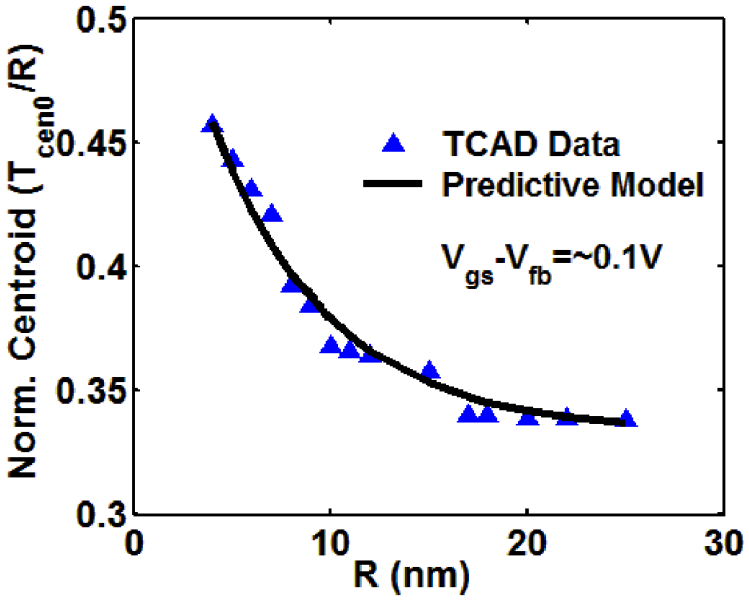


Figure 3.5: Charge centroid for low gate bias has a geometric dependence and takes values in between the asymptotic values calculated in Eqn.(3.6) for a CG FET with circular cross-section. Value extracted for $R_0 = 6\,\text{nm}$.

Q_0 (in C/cm^2). The critical charge is defined as the charge density for which the screening length, T_{debye} approaches the device confinement dimension. We derive,

$$Q_0 = q.N_{ch}.T_{debye} \tag{3.7}$$

$$T_{debye} = \sqrt{\frac{\epsilon_0 \epsilon_{ch} kT}{q^2 N_{ch}}} \tag{3.8}$$

where ϵ_0 is the permittivity of free space, ϵ_{ch} is the dielectric constant of channel material, kT/q is the thermal voltage and N_{ch} is the critical carrier density in the channel (in cm^{-3}). For a DG FET device the confinement is along the fin thickness and hence $T_{debye} = T_{FIN}$. Thus, using Eqn.(3.8) we observe that Q_0 is inversely proportional to T_{FIN}. A semi-empirical form for the charge centroid can now be written as follows,

$$T_{cen} = \frac{T_{cen0}}{1 + \left(\frac{Q_{inv}}{Q_0}\right)^{\alpha}} = \frac{T_{cen0}}{1 + \left(\frac{T_{FIN} Q_{inv}}{Q_1}\right)^{\alpha}} \tag{3.9}$$

where T_{cen0} is the upper bound of centroid obtained in Eqn.(3.3) and Q_{inv} (in C/cm^2) is the inversion charge in the channel. For compact model purposes we find Eqn.(3.9) is sufficient for capturing the geometry dependence along confinement direction (through T_{FIN} in numerator and denominator) and the bias dependence (through Q_{inv}) of the centroid

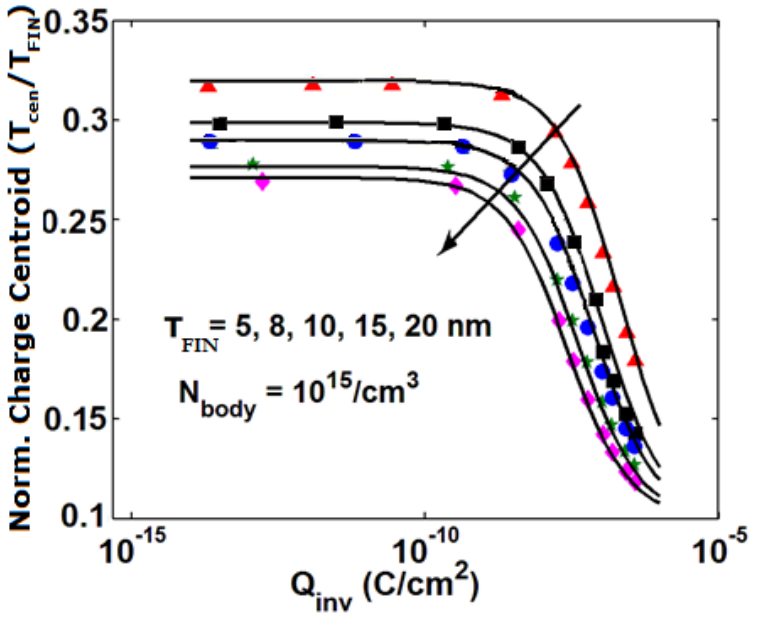


Figure 3.6: After a certain critical charge in the channel, the centroid shifts towards the gate-oxide channel interface with increasing gate bias. Symbols: TCAD simulations, Lines: Proposed compact model

adequately. Unlike the empirical expression proposed in [44], invoking the Debye length concept has rendered an accurate form for T_{FIN} and temperature dependence (T in kT/q) for the charge centroid. Q_{inv} is already available for use in a Verilog-A code for a charge-based model like BSIM-CMG. A more rigorous derivation based on Schrodinger and Poisson equations would show that parameter Q_1 is inversely proportional to the effective mass and thus the channel orientation [42]. To accommodate this deficiency of the semi-classical approach and in order to support holes as well as electrons as carriers, the usage of different channel materials in the transistor and process induced variations in the shape of the fin T_0, Q_1 and α will be used as tuning parameters to capture any inadequacies. From [42] we also gather that T_0 would be weak functions of channel orientation.

Along similar lines, an equation can be obtained for the CG FET device too. The equation would look just like Eqn.(3.9) except with T_{FIN} replaced by radius R, and T_{cen0} obtained from Eqn.(3.6) will be used.

In Fig.3.6, we present an overlay of model as in Eqn.(3.9) tuned to TCAD data for a DG FET device. For this case the following values were extracted for the tuning parameters in the proposed model, $T_0 = 12\,\text{nm}$, $\alpha = 1.3$ and $Q_1 = 10^{-16}\,\text{C/cm}$. This simple model shows excellent match with TCAD and retains the necessary predictive nature across geometry and bias. The CG FET device also shows similar fit, see $N_{body} = 10^{15}\,\text{cm}^{-3}$ case in Fig.3.7.

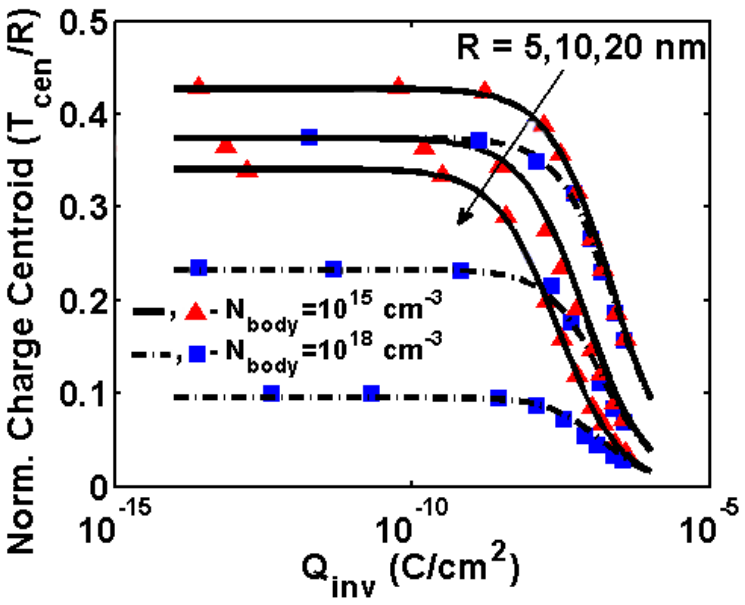


Figure 3.7: Channel doping sets up a varying potential across the channel re-distributing the inversion charge closer to the interface even for low gate bias. Inversion charge centroid for a CG FET structure for intrinsic and moderate channel doping. Symbols: TCAD simulations, Lines: Proposed compact model

3.1.1.3 Impact of Channel Doping

Low power circuit design optimization requires transistors with multiple threshold voltages. Usage of channel doping is still sought after as the most practical approach despite a preference to use intrinsic channel for higher mobility and lower variability [46]. However in order to retain the advantages of multi-gate structures just enough doping is used to still stay within a fully depleted channel regime of device operation. Under this condition, even at low gate voltages (negligible inversion carriers) the depletion charge sets up a non-zero electric field across the channel. Unlike the intrinsic channel case where the initial field in the channel was negligible and the potential was roughly constant in the quantum well, with considerable doping, for example in a DG FET structure there exists a linear electric field and a parabolic potential. The inversion charge re-distributes itself closer to the interface even at low gate voltages. Similar arguments can be made for a CG FET structure as well where in a parabolic field and cubic potential is set up at low gate voltages as obtained from solution to a Poisson equation in cylindrical coordinate. Eqn.(3.9) can be extended to accommodate this phenomenon as follows,

$$T_{cen} == \frac{T_{cen0}}{1 + \left(T_{FIN}\frac{Q_{inv}+\eta.Q_{dep}}{Q_1}\right)^{\alpha}} \tag{3.10}$$

where Q_{dep} is the depletion charge given by $qN_{body}T_{FIN}$ for a DG FET structure and $qN_{body}R/2$ for a CG FET structure. N_{body} is the channel doping in cm^{-3}. η is a tuning parameter introduced for fitting purposes. In Fig.3.7, Eqn.(3.10) (with T_{FIN} replaced by R and T_{cen0} from Eqn.(3.6)) is validated for a CG FET with two different channel doping values. A consistent set of parameters were extracted across all radii and channel doping values $R_0 = 6\,\mathrm{nm}$, $\alpha = 0.834$, $Q_1 = 10^{-16}\,\mathrm{C/cm}$ and $\eta = 0.15$. Similar results were obtained for DG FET structure at $N_{body} = 10^{18}\,\mathrm{cm}^{-3}$ with $\eta = 0.25$ and other parameters as extracted for Fig.3.6 (result not shown).

3.1.2 Effective Gate Capacitance Model

The position of the charge centroid obtained in Eqns.(3.9) and (3.10) can now be used to create a modified expression for effective gate capacitance. The capacitance is a series combination of the oxide capacitance and a bias dependent channel capacitance component due to the centroid. For a DG FET device the expression is as follows

$$\frac{1}{C_{oxeff}} = \frac{T_{ox}}{\epsilon_0\epsilon_{ox}} + \frac{T_{cen}}{\epsilon_0\epsilon_{ch}} \tag{3.11}$$

and for a CG FET device the expression is

$$\frac{1}{C_{oxeff}} = \frac{1}{\epsilon_0\epsilon_{ox}} ln\left(\frac{R+T_{ox}}{T_{ox}}\right) + \frac{1}{\epsilon_0\epsilon_{ch}} ln\left(\frac{R}{R-T_{cen}}\right) \tag{3.12}$$

where ϵ_{ox} is the dielectric constant and T_{ox} is the physical thickness of the gate-oxide. The small signal (AC) gate capacitance of a DG structure with intrinsic silicon long channel and silicon-oxide gate dielectric ($T_{FIN} = 15\,\mathrm{nm}$, $T_{ox} = 1.5\,\mathrm{nm}$ and $L = 10\,\mu\mathrm{m}$) was simulated using TCAD, Fig.3.8. The device was biased in linear region (drain voltage, $V_{ds} = 50\,\mathrm{mV}$). Two cases were simulated to assess the effect of QM charge centroid one without QM effects (in which case the Poisson and Continuity equations are solved for self-consistently) and another with the QM effects present (in which case Schrodinger equations are included self-consistently as well). As expected, the gate capacitance is reduced when QM confinement effects are taken in to consideration. The gate capacitance approaches its maximum value (of $C_{ox} = \epsilon_0\epsilon_{ox}.2.H_{FIN}.L/T_{ox}$, where H_{FIN} is the height of the fin) in a more gradual manner due to the charge centroid being away from the channel interface. The modified expressions in Eqns.(3.11) and (3.12) are incorporated in the BSIM-CMG charge model [39]. The resultant model now predicts the gate capacitance accurately. As the charge centroid moves towards the interface for higher gate bias, the gate capacitance approaches its maximum value.

In a usual compact model parameter extraction flow, the parameters (T_0 or R_0, Q_1, η and α) introduced can be directly extracted from either C_{gg} vs. V_{gs} curve for the linear region or from a split-CV measurement of a large device (long channel, multi-fin, multi-finger).

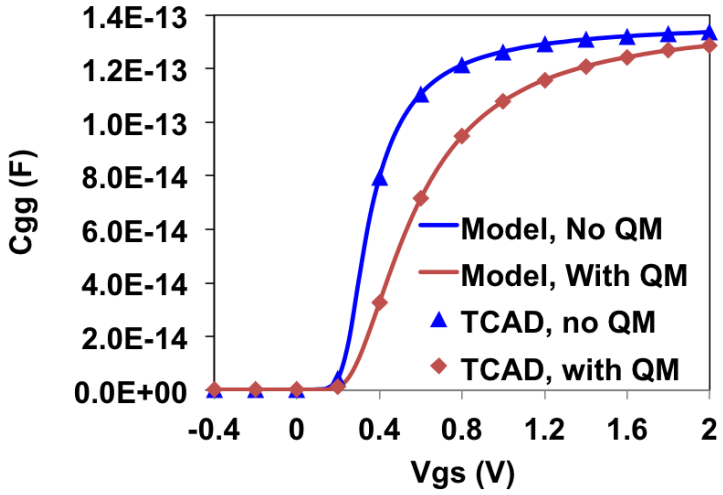


Figure 3.8: Small signal gate capacitance vs. gate terminal bias for a long channel double gate FET (T_{FIN} = 15 nm, Tox = 1.5 nm). Symbols: TCAD simulations, Lines: Model in Eqn.(3.10) incorporated in BSIM-CMG

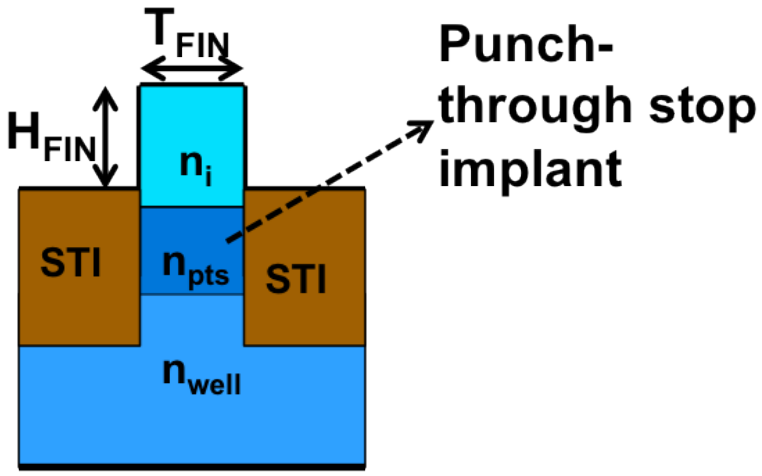


Figure 3.9: Cross-section of a bulk p-FinFET showing (a) punch-through stop implant below the fin. Gate and oxide regions not shown.

3.2 Double Junction capacitance modeling for Bulk FinFETs

As the channel length scaled down the MOSFET incurs an increased amount of undesired sub-surface leakage current from the source to drain. This region below the interface is less

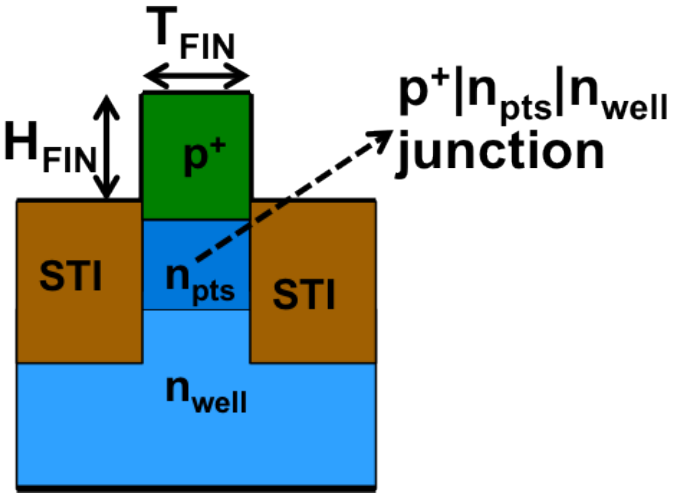


Figure 3.10: Punch-through stop implant below the fin laterally diffuses below the S/D junction region leading to $p^+|n_{pts}|n_{well}$ type of junction. Gate and oxide regions now shown.

controlled by the gate and more by the source-drain electric fields. In planar transistors halo implants below the source and drain regions were used to mitigate this current. With decreasing channel length we saw the need for increasing amount of halo implants that almost merged together from either sides. However with further scaling, ever higher halo implant dosage could not be used. The semiconductor industry moved away from planar transistors creating the need for multi-gate transistors where the 3-D nature of the channel lent better gate control over the channel. Today FinFETs on Bulk and SOI substrates have moved into mass production.

However FinFETs on bulk substrate have a region just below the fin that is less controlled by the gate. Source-Drain field would extend into these regions creating sub-surface leakage (although lesser in magnitude compared to planar transistors at the same channel length). In order to weaken these fields some amount of punch-through stop (PTS) implant (a.k.a. ground plane doping) is employed for bulk FinFETs in production [47]. Fig.3.9 illustrates a cross-section of a p-FinFET showing the presence of this implant just below the actual fin region. This implant present below the intrinsic fin region will laterally diffuse under the source/drain junction region. The use of high dose well implant will lead to an increase of junction tunneling current leakage component. Thus the magnitude of doping in this region will have to be less than or equal to the well doping. In Fig.3.10 the cross-section of a p-FinFET under the source/drain region is illustrated leading to creation of a double junction with two different n-type doping $p^+|n_{pts}|n_{well}$ as a general case. When the reverse bias applied to this junction is increased (for ex: through increased positive drain voltage), the depletion region edge will traverse through the n_{pts} region and could enter the n_{well} region too. This leads to a deviation in the behavior of junction capacitance from that of an ideal uniformly doped $p^+|n$ kind of step junction diode observed in planar bulk MOSFETs.

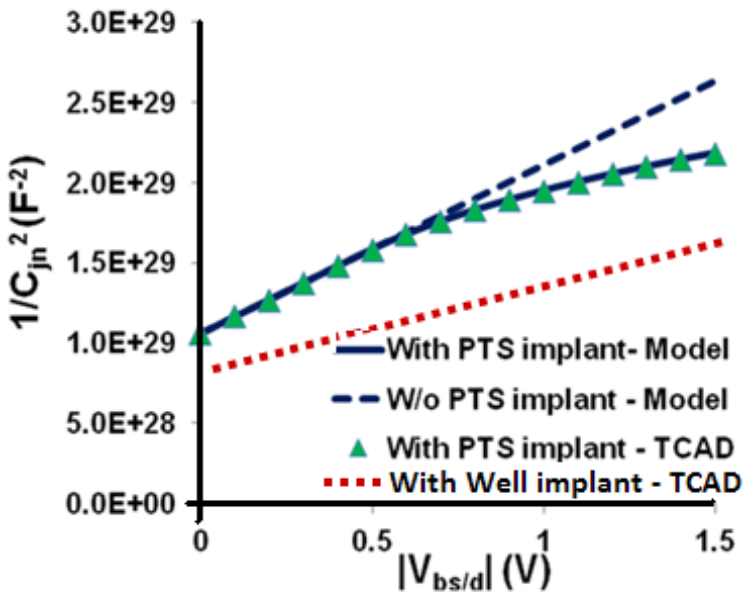


Figure 3.11: The $p^+|n_{pts}|n_{well}$ junction exhibits two slopes in a $1/C_{jn}^2$ vs V_{jn} plot in comparison to single slope of the ideal step junction. Values of doping used were $p^+ = 3\times10^{20}\,\mathrm{cm}^{-3}$, $n_{pts} = 10^{18}\,\mathrm{cm}^{-3}$, $n_{well} = 3\times10^{18}\,\mathrm{cm}^{-3}$ (Lines-Model, Symbols-TCAD)

Fig.3.11 shows that $1/C_{jn}^2$ vs. V_{jn} curve for FinFET S/D junctions with PTS deviates from the linear behavior of without PTS (ideal step junction). The slope of this curve is inversely proportional to the n-type doping at the edge of the depletion region. FinFET S/D junctions tend to show two different slopes and a higher junction capacitance when a PTS implant is used. In this section a new junction capacitance model has been developed to capture this process induced subtlety in the bulk FinFET junction region.

3.2.1 Reverse Bias Model

We will derive a junction charge model based on the well-known simple case of a single junction $p|n$ diode capacitance model for reverse bias [3]. The reverse bias depletion charge is modeled as follows,

$$Q_{jn,rev} = \begin{cases} C_{j01}\phi_{b1}\dfrac{1-\left(1-\frac{V_{bs/d}}{\phi_{b1}}\right)^{1-m_1}}{1-m_1} & 0 < V_{bs/d} < V_{bc} \\ C_{j01}\phi_{b1}\dfrac{1-\left(1-\frac{V_{bc}}{\phi_{b1}}\right)^{1-m_1}}{1-m_1} + C_{j02}\phi_{b2}\dfrac{1-\left(1-\frac{V_{bs/d}-V_{bc}}{\phi_{b2}}\right)^{1-m_2}}{1-m_2} & V_{bc} < V_{bs/d} \end{cases} \tag{3.13}$$

where $V_{bs/d}$ is the voltage across the junction, $C_{j01,2}$ are the zero bias capacitance values and $\phi_{b1,2}$ is the barrier height of the $p^+|n_{well}$ and $p^+|n_{pts}$ junctions. $m_{1,2}$ represent the gradient of the and $p^+|n_{pts}$ the $n_{pts}|n_{well}$ junctions. We can observe that the first term in Eqn.(3.13) is

the same as that for a single junction diode. Eqn.(3.13) maintains charge continuity at the cross-over voltage $V_{bs/d} = V_{bc}$. The continuity of the first and second derivatives of charge also needs to be ascertained for accuracy in the prediction of higher harmonic power content in the output of a transistor in Analog/RF circuit simulations. The continuity of the first derivative of charge in Eqn.(3.13) (which is also the junction capacitance) at $V_{bs/d} = V_{bc}$ yields,

$$C_{j01}\left(1-\frac{V_{bc}}{\phi_{b1}}\right)^{-m_1} = C_{j02} \tag{3.14}$$

Ensuring continuity of the second derivative of the charge (first derivative of capacitance), at $V_{bs/d} = V_{bc}$ gives rise to the below condition.

$$C_{j01}m_1\frac{\left(1-\frac{V_{bc}}{\phi_{b1}}\right)^{-1-m_1}}{\phi_{b1}} = \frac{C_{j02}m_2}{\phi_{b2}} \tag{3.15}$$

These conditions, Eqns.(3.14) and (3.15) are factored into the parameter extraction process. In the junction capacitance curve, $1/C_{jn}^2 - V_{bs/d}$, Fig.3.11 the first slope region corresponding to depletion edge traversing the PTS implant region, is used to extract the values for parameters C_{j01}, ϕ_{b1} and m_1 in a similar way as that for a single junction diode. Among the remaining four parameters (V_{bc}, C_{j02}, ϕ_{b2} and m_2) that correspond to the n_{well} region (the second slope region), conditions, Eqns.(3.14) and (3.15) allow us the flexibility to choose only two of them. We chose parameters C_{j02} and ϕ_{b2} that signify the depth of the PTS implant n_{well} region boundary and the n_{well} region doping concentration. The parameters V_{bc} and m_2 will now be determined by simultaneously solving Eqns.(3.14) and (3.15) using the values chosen for C_{j02} and ϕ_{b2}. The junction capacitance is then given by the derivative $dQ_{jn,rev}/dV_{bs/d}$ as follows,

$$C_{jn,rev} = \begin{cases} C_{j01}\left(1-\frac{V_{bs/d}}{\phi_{b1}}\right)^{-m_1} & 0 < V_{bs/d} < V_{bc} \\ C_{j02}\left(1-\frac{V_{bs/d}-V_{bc}}{\phi_{b2}}\right)^{-m_2} & V_{bc} < V_{bs/d} \end{cases} \tag{3.16}$$

To validate the model the structure in Fig.3.10 was simulated using TCAD [34]. A source/drain region p^+ doping was chosen as $3 \times 10^{20}\,\mathrm{cm}^{-3}$. The doping for the n_{PTS} and n_{well} region were chosen to be $10^{18}\,\mathrm{cm}^{-3}$ and $3 \times 10^{18}\,\mathrm{cm}^{-3}$ respectively. The two terminal capacitance across the source/drain and substrate were extracted. From Fig.3.11 we can observe that the cross-over voltage, V_{bc} is around 0.6 V of reverse bias junction voltage where the slope of the $1/C_{jn}^2$ plot changes reflecting the doping at the edge of the depletion region. The derived model in Eqn.(3.16) shows excellent agreement (after parameter tuning) with TCAD simulations for such a junction, Fig.3.11.

3.2.2 Forward Bias Model

The charge/capacitance model in Eqns. (3.13),(3.16) is not valid for voltages $V_{bs/d}$ approaching $-\phi_{b1}$, i.e. when the diode is forward biased. In a realistic diode, series resistances of the

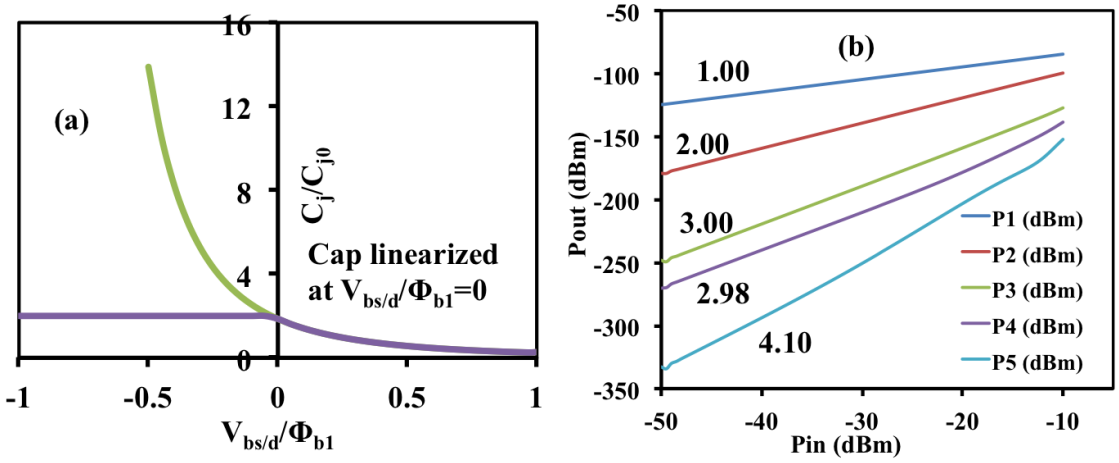


Figure 3.12: (a) The first derivative of junction capacitance using the linearization in Eqn. (3.17). (b) Results of harmonic balance test for a PMOS passive mixer configuration showing wrong slopes for the 4^{th} and 5^{th} harmonic output power.

quasi-neutral regions dominate and restrict the actual voltage drop across the junction (or rather across the depletion edges on either sides of the metallurgical junction). Eqn. (3.16) would numerically result in very large values making it unwieldy for implementation in a compact modeling framework. For this reason, industry standard compact models resort to an approximation for the forward bias region. For ex: BSIM4 resorts to using a quadratic equation to describe the junction charge for the forward bias (i.e. junction capacitance is made linear). For a p-FET the forward bias junction charge is essentially given by a Taylor series expansion of Eqn. (3.16) up to second order around $V_{bs/d} = 0\,\mathrm{V}$,

$$Q_{jn,fwd} = C_{j01} \cdot V_{bs/d} + \frac{C_{j01} m_1}{2\phi_{b1}} V_{bs/d}^2 \text{ for } V_{bs/d} < 0 \tag{3.17}$$

Eqn.(3.17) taken together with Eqn.(3.13) ensures the continuity of junction charge and its first derivative capacitance around $V_{bs/d} = 0\,\mathrm{V}$. However there is a discontinuity in the third derivative of the overall junction charge i.e. $d^3Q_{jn}/dV_{bs/d}^3$ at $V_{bs/d} = 0\,\mathrm{V}$. This discontinuity is not a cause of concern for convergence of a compact model. SPICE simulators require continuity of only up to second derivative of charge for convergence. However for the RF design community that is interested in accurate prediction of higher order inter-modulation (IM) distortion products to accurately predict out-of-band emission by wireless transmitters continuity of charge up to 6^{th}-order is desired. For ex: in the operation of a passive mixer, the FET operates in strict linear region with source-drain voltage, $V_{ds} = 0\,\mathrm{V}$. For this mixer, if the FET body is tied to ground, the voltage across the source/drain junction is also centered around 0 V during the mixing operation. The discontinuity when Eqns.(3.17) and (3.13) are put together will lead to incorrect predictions of higher order IM products. This can be verified using a simple harmonic-balance simulation set-up wherein a single-tone RF stimulation is supplied to the source of a FET with its gate voltage set to above the threshold

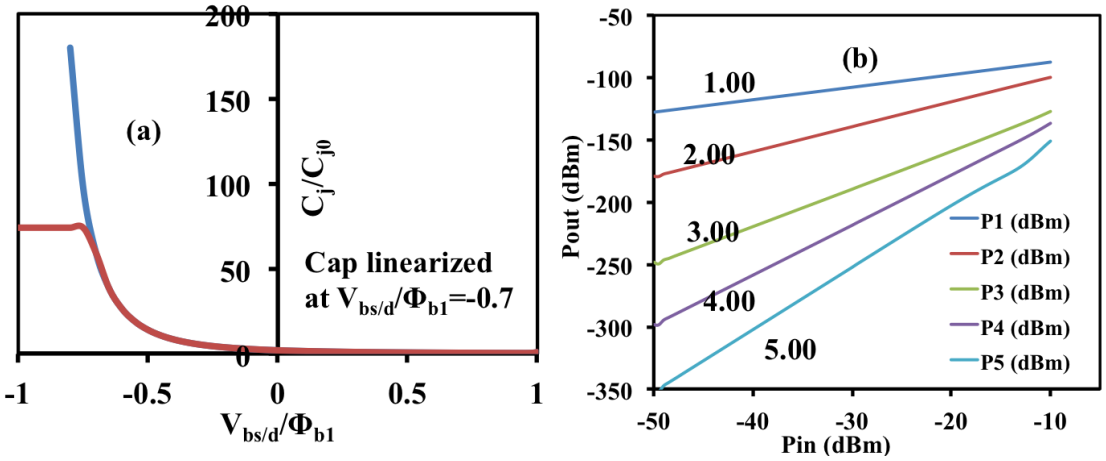


Figure 3.13: (a) The first derivative of junction capacitance using the linearization in Eqn.(3.18) with $k = 0.7$. (b) Results of harmonic balance test for a PMOS passive mixer configuration showing correct slopes for the 4^{th} and 5^{th} harmonic output power.

voltage. The power in the harmonic content of the drain current is observed. The slope of the n^{th}-harmonic component power *vs.* input power would be n at low power input. Any discontinuity or issues with model symmetry w.r.t. source and drain would lead to wrong slopes. In Fig.3.12 we report results from such a test for a model with a symmetric core but the above junction charge/capacitance model. As expected we see that deviation in the slope for the 4^{th} and 5^{th} harmonic output power. In order to rectify the same, we propose an alternate model as follows. Instead of a Taylor series expansion around $V_{bs/d} = 0\,\mathrm{V}$, pushing the transition point further into forward bias helps. We choose a quadratic Taylor series expansion around $V_{bs/d} = k\phi_{b1}$ for this as follows,

$$\begin{aligned} Q_{jn,fwd} = C_{j01}\phi_{b1}\frac{(1-k)^{1-m_1}}{1-m_1} + C_{j01}(1-k)^{-m_1}\cdot(V_{bs/d}+k\phi_{b1})... \\ +C_{j01}m_1\frac{(1-k)^{-1-m_1}}{2\phi_{b1}}\cdot(V_{bs/d}+k\phi_{b1})^2 \text{ for } V_{bs/d} < -k\phi_{b1} \end{aligned} \tag{3.18}$$

For implementation purposes in BSIM-CMG we have chosen $k = 0.9$. This change lends back accuracy to higher order IM products. This modified model was implemented for junction capacitance in BSIM-CMG and results from the harmonic balance simulations for a passive mixer shown in Fig.3.13. As expected we see that this model predicts accurate slopes up to 5^{th} harmonic content in the MOSFET. The persisting discontinuity at very high forward bias shouldn't be a problem. Given that the diode current behaves exponentially with bias voltage $V_{bs/d}$, all the voltage drop would be across the parasitic source-drain resistance or the substrate network resistance, restricting the diode junction from seeing such high voltages.

The overall junction charge model is as follows

$$Q_{jn} = \begin{cases} C_{j01}\phi_{b1}\frac{(1-k)^{1-m_1}}{1-m_1} + C_{j01}(1-k)^{-m_1}\cdot(V_{bs/d}+k\phi_{b1})... & \\ +C_{j01}m_1\frac{(1-k)^{-1-m_1}}{2\phi_{b1}}\cdot(V_{bs/d}+k\phi_{b1})^2 & V_{bs/d} < -k\phi_{b1} \\ C_{j01}\phi_{b1}\frac{1-\left(1-\frac{V_{bs/d}}{\phi_{b1}}\right)^{1-m_1}}{1-m_1} & -k\phi_{b1} < V_{bs/d} < V_{bc} \\ C_{j01}\phi_{b1}\frac{1-\left(1-\frac{V_{bc}}{\phi_{b1}}\right)^{1-m_1}}{1-m_1} + C_{j02}\phi_{b2}\frac{1-\left(1-\frac{V_{bs/d}-V_{bc}}{\phi_{b2}}\right)^{1-m_2}}{1-m_2} & V_{bc} < V_{bs/d} \end{cases} \tag{3.19}$$

The junction capacitance given by $dQ_{jn}/dV_{bs/d}$ is given by

$$C_{jn} = \begin{cases} \frac{C_{j01}}{(1-k)^{m_1}} + \frac{C_{j01}m_1}{\phi_{b1}(1-k)^{1+m_1}}\cdot(V_{bs/d}+k\phi_{b1}) & V_{bs/d} < -k\phi_{b1} \\ C_{j01}\left(1-\frac{V_{bs/d}}{\phi_{b1}}\right)^{-m_1} & -k\phi_{b1} < V_{bs/d} < V_{bc} \\ C_{j02}\left(1-\frac{V_{bs/d}-V_{bc}}{\phi_{b2}}\right)^{-m_2} & V_{bc} < V_{bs/d} \end{cases} \tag{3.20}$$

3.3 Summary

- A compact model for the quantum mechanical charge centroid behavior in multi-gate FETs has been presented. This model helps capture the non-linearity and reduction in gate capacitance in the moderate inversion region (which has increasingly become important for analog/RF circuit design) of the device. The developed phenomenological model maintains a balance of predictive nature (across geometry and bias) and flexibility to capture data of devices from different process technologies. The model has been validated with adequate TCAD simulations of both charge centroid and small signal device gate capacitance.

- A capacitance model to capture the double source / drain junction characteristics that arises in FinFETs on bulk substrate where a punch-through prevent implant is employed is discussed. The new capacitance model accurately captures the dual slope nature exhibited by the double junction as witnessed in a $1/C_{jn}^2 - V_{bs/d}$ curve. In addition we have also described a forward bias junction capacitance model that retains accuracy of RF inter-modulation linearity predictions up to 5^{th} order.

Chapter 4

Asymmetric Device Drain Current Modeling

An ideal MOSFET is inherently symmetric w.r.t. source and drain. However often times a derivative of the ideal MOSFET, an asymmetric one is sought for to take advantage of some particular property. For example, a laterally double-diffused MOS (LDMOS) FET routinely used in high-voltage applications has different amount of channel doping at source and drain end due to lateral diffusion of dopants from one end. A drain-extended MOS (DEMOS) has an extended drift region that is lightly doped at the drain where most of the voltage is dropped during high-voltage operation. V.P.Hu *et al.* have shown that an intentionally created asymmetric FinFET with source / drain underlap employed as access transistors in SRAMs leads to better read static noise margin while maintaining the same write static noise margin [48]. Among surround gate FET architectures, a vertical cylindrical gate (CG) FET is inherently asymmetric, Fig.4.1. As discussed in Chapter 1, a vertical CG-FET device is being sought after as the replacement for the current generation quasi-planar DRAM transistor. Also, a 3-D multi-transistor vertical NAND string could be made out of a single column of silicon for non-volatile flash memory applications [49]. While the devices in the middle of such a string are symmetric, the end devices are asymmetric w.r.t. top as source / drain operation. This asymmetry accentuates at high voltages that some devices are subjected to. Depending on the layout, one might want to use either the top as the drain or the source. While it might be pragmatic to visualize either configuration as separate devices for simulation purposes, it is of convenience to support both the configurations in a single SPICE model, thus requiring the model to be inherently asymmetric.

BSIM-CMG has been developed with an intention to cater to symmetric devices by default. Device asymmetry shall be treated as a mere extension of the symmetric model in a way that it does not significantly alter the structure of the existing model. In the so far developed model, we could introduce some amount of asymmetry in parameters that tend to differentiate the source and drain end. They are -

- Source / Drain resistances

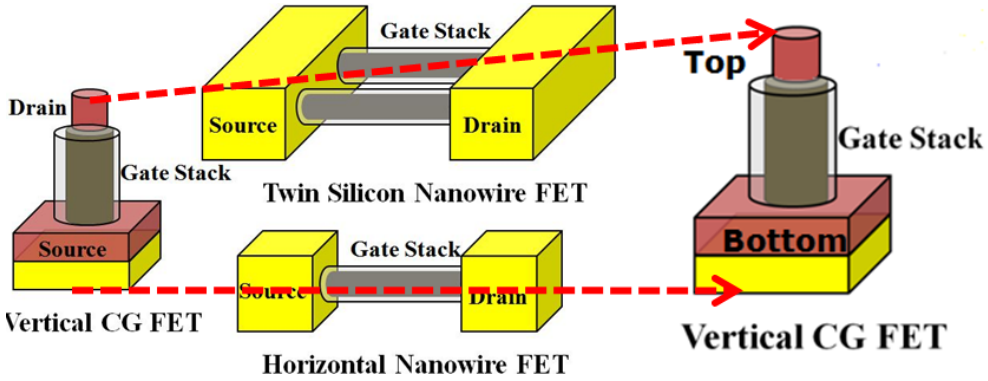


Figure 4.1: Among the viable device architectures for a gate-all-around transistor, the vertical cylindrical gate (CG) FET displays both structural asymmetry and process induced asymmetry w.r.t. Top electrode being used as the Drain or the Source of the transistor.

- Source / Drain junction current and capacitances
- Gate induced Source / Drain leakage current.

As we will see later these sub-models alone are insufficient to capture the device asymmetry in the vertical CG-FET transistor analyzed. In this chapter our goal will be to analyze the structural and process induced device asymmetry and capture them in BSIM-CMG. We shall translate this understanding by either identifying existing parameters that need asymmetric counterparts or augment existing device sub-models to capture the asymmetry. We will then validate the model using hardware data from a 2.2 V vertical CG-FET device.

4.1 Factors leading to Device Asymmetry

We identified three main factors leading to device asymmetry in a vertical CG-FET device explained as follows.

4.1.1 Graded Channel Doping

Graded channel doping in vertical MOSFET's have been studied before for it's benefits [50]. A device with higher doping at the source end is said to have better short-channel control (lower drain-induced barrier lowering (DIBL)) and improved hot-carrier performance (lower E-field at the drain end). However that was only half the story said about this device. Channel implant that sets the threshold voltage of the device is performed vertically as in a bulk-planar FET technology. However for a vertical CG-FET technology this happens to be along the channel from source to drain or vice-versa leading to a graded channel doping. In order to fully comprehend the action of such a device we perform TCAD simulations on an ideal CG-FET structure with about a decade of channel doping gradient from one end

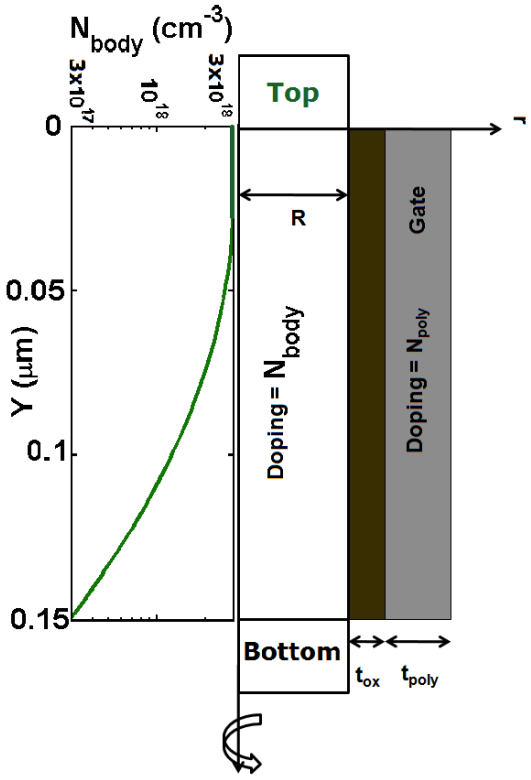


Figure 4.2: Vertical ion implantation induces a doping gradation along the channel from Top to Bottom. TCAD structure with a Gaussian doping profile and up to one decade gradation between source and drain.

of the channel to the other, Fig.4.2. As observed in a realistic technology, we shall call the channel end with higher doping to be the *top* electrode and the other the *bottom* electrode in the discussion that follows. This device could be configured in two ways - top as drain and bottom as drain.

Irrespective of which configuration, when operating in linear region (low V_{ds}) the drain current is symmetric, Fig.4.3. The threshold voltage is determined by the maximum barrier in the channel (or by the maximum doping in the channel). The inversion region (high V_g) current is determined by the sum of the source and drain resistances ($R_s + R_d$). However the drain current is asymmetric in saturation region (high V_{ds}), Fig.4.4. For high V_{ds}, the source side doping determines the threshold voltage of the device, i.e.,

$$V_{th,sat} \propto N_{channel}@\text{source end} \tag{4.1}$$

Hence the configuration where the source side has higher doping would see lower DIBL (better short channel control) as well. But the inversion region (high V_g) on-current is controlled by the doping at the drain end. The drain current is proportional to the drain saturation voltage, V_{dsat} which is set by pinch-off / velocity saturation occurring at the drain end of the

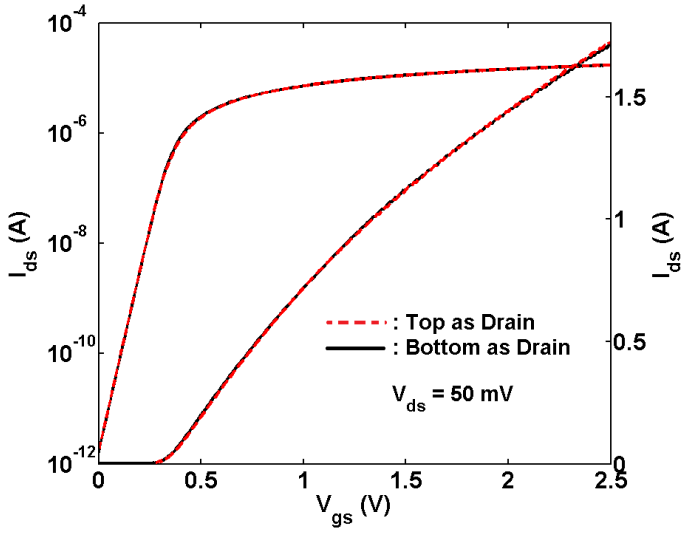


Figure 4.3: Linear region drain current is oblivious to Top being Drain or Source for non-uniform channel doping. Lines: TCAD Results

channel, i.e.

$$I_{on} \propto V_{dsat} = V_{gs} - V_{th,drain} (\propto N_{channel} @\text{drain end}) \tag{4.2}$$

Therefore, even though the top as drain configuration in Fig.4.2 would have a lower threshold voltage, this configuration has lower on-current as well. The on-current / saturation current is independent of V_{ds} and dependent on V_{gs} and hence is sensitive to the source side resistance R_s. This process induced asymmetric DIBL and asymmetric on-current can also be observed when we plot the drain current against V_{ds} for different gate voltages for both the configurations, Fig.4.5.

4.1.2 Bottom Electrode Resistance: Quasi-Saturation

By the virtue of its architecture the vertical CG-FET device requires a bottom electrode contact (either source or drain depending on the configuration) that is parallel to the channel and away from it. This leads to an inevitable drift region (n-doped region for a NMOS) between the bottom end of the channel and the bottom end of the metal electrode. Part of this drift region is either directly below the gate electrode or is under the influence of gate fringe field leading to a gated-resistance behavior. In device design to determine the distance between the channel and this electrode one needs to trade-off the drift region resistance with the gate-bottom electrode fringe capacitance. This region is similar in nature to that observed in a DEMOS used in high-voltage/power applications.

One well known phenomenon that arises in such a structure where the drift region is moderately doped is quasi-saturation. If the bottom electrode is used as the drain of the FET, at high current levels, the drift region is unable to furnish the required carriers leading to current saturation. This saturation occurs even before the regular FET channel saturation

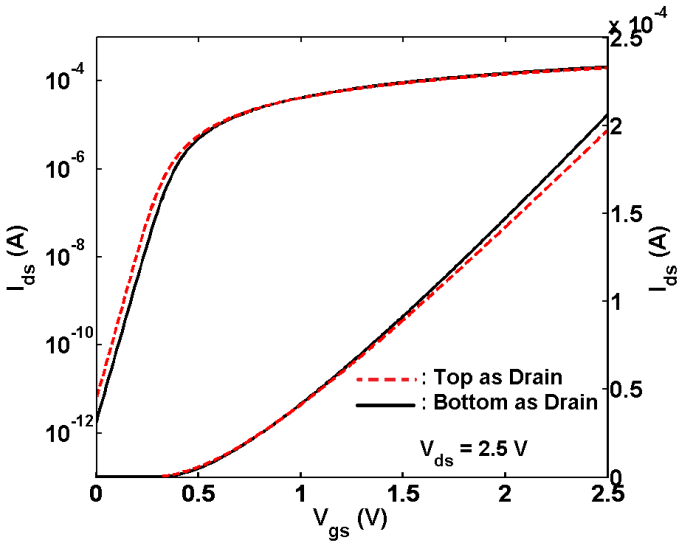


Figure 4.4: Top as Drain condition exhibits lower V_{th} (lower source-side doping) as well as lower I_{on} (higher drain side doping => lower V_{dsat}) in saturation region for a non-uniform channel doping. Lines: TCAD Results

at the drain end. This in turn leads to gate voltage de-sensitization leading to drain-current crowding as observed in a $I_d - V_d$ curve in the moderate inversion region for high gate / drain bias. The gate voltage de-sensitization could also be visualized as a negative feedback on the channel and can be captured with a bias-dependent resistance that is proportional to the applied voltage across it's ends as follows

$$R_{drift,d} = R_{contact,d} + R_{bulk,d}\frac{1 + RDDR \cdot (V_d - V_di)^{PRDDR}}{1 + PRWG \cdot V_{gdi}} \tag{4.3}$$

where $R_{contact}$ is the contact resistance, R_{bulk} is the bulk resistance of the drift region, $V_d - V_{di}$ is the voltage across the drift region, PRDDR is a tuning parameter to capture the quasi-saturation effect and PRWG captures the effect of gate voltage induced accumulation in the drift region. In order to verify the model, we perform TCAD simulations of a NMOS CG-FET structure with an extended bottom region and uniform channel doping. Floating metal contacts are inserted near the channel edges to gauge the voltage drop across the drift region and extract the resistance. The effect of gate voltage on the top source / drain region resistance is simulated by sweeping V_g. The gate fringe field in the n-doped region induces accumulation of electrons leading to a drop in the resistance of this region, Fig.4.6. The effect was found to be similar for the bottom source / drain region as well. The parameter PRWG in Eqn.(4.3) is used to capture this behavior. To assess the effect of quasi-saturation in the bottom drift region, the drain voltage, V_{ds} was swept maintaining the gate voltage, V_g at 2.5 V. A relatively low doping of $10^{18}\,\mathrm{cm}^{-3}$ was used in the drift region to illustrate the effect. Practical values might be higher than this value. When the bottom electrode is used as the drain of the device, we can observe that at high levels of current the drift region is

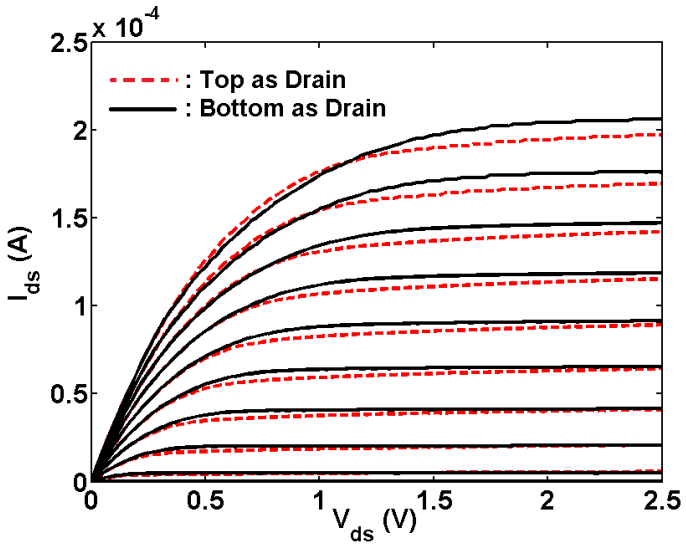


Figure 4.5: The asymmetric DIBL induced V_{th} shift also emphasizes itself as slightly different output conductance in $I_{ds} - Vds$ characteristics. Lines: TCAD Results

starved of carriers leading to quasi-saturation and an apparent increase in the resistance of the region, Fig.4.7. The parameters $V_{drfit,sat}$ and PRDDR are tuned to capture this effect.

However if the bottom electrode were to be used as the source of the device the drift region acts like a source resistance leading to V_{gs} de-biasing reducing both linear and saturation region drain current, Fig.4.7. When used as the source region, the drift region is able to supply the requisite carriers. This way a structural asymmetry establishes itself in the vertical CG-FET depending on which configuration it is being used.

4.1.3 Bottom Electrode Resistance: Junction modulated resistance

Another important aspect unique to a vertical CG-FET device built on bulk substrate is the fact that the bottom source / drain region has two metallurgical junctions. One between the bottom region and the channel region (which is present in a regular bulk MOSFET) and another separate one between the bottom region and the substrate. In all other MOSFET device architectures on bulk substrate there is just one metallurgical junction. This structure leads to some unique effects. There is an additional bias dependence due to the depletion region edge modulation of the bottom - substrate junction, Fig.4.8. In a device that is intended for high voltage operation, this drift region is moderately doped. Moderate doping would imply greater encroachment of the depletion region into the drift region from the bottom. The carriers carrying the current will have a reduced quasi-neutral region to traverse (X_{bot} in Fig.4.8) leading to a greater bulk resistance for this region. There is asymmetry in how this region behaves for the two possible configurations as well. In a typical circuit

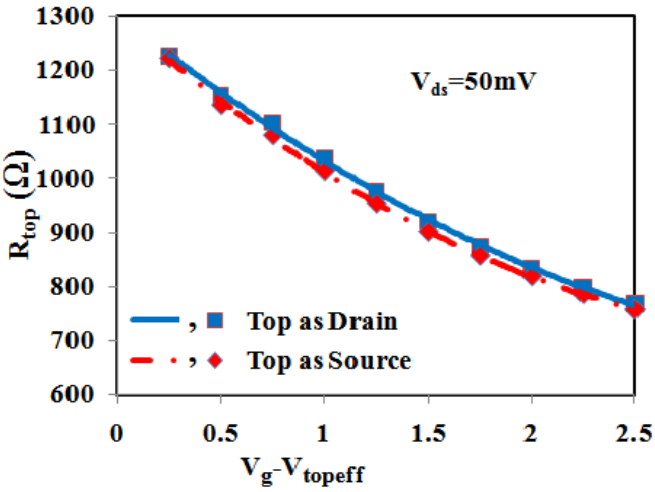


Figure 4.6: Gate voltage dependence of source / drain region resistance. Increasing gate voltage leads to increasing accumulation of electrons leading to decrease in resistance. Symbols: Extracted resistance from TCAD using a floating electrode, Lines: Asymmetry aware resistance model

configuration the source is held at a constant voltage (usually 0 V). If the bottom electrode were to be used as the source, we would see no modulation of the depletion region and hence X_{bot}. However if the bottom electrode were to be used as the drain, the bottom-substrate diode is reversed biased and X_{bot} reduces with increasing drain voltage leading to increased drift region resistance due to current constriction. This modulation of the drift region bulk resistance can be captured as follows,

$$R_{bulk} = \frac{R_{bulk0}}{1 - DRBJ\sqrt{V_{bi} + V_{bot,sub}}} \tag{4.4}$$

where DRBJ is a tuning parameter to represent the doping and the zero-bias width of the drift region, V_{bi} is the built-in voltage of the bottom-substrate junction and $V_{bot,sub}$ is the voltage across this junction. For the hardware device data we will be dealing with it was found that this effect is negligible.

4.2 Model Formulation

Although due to asymmetry the device might behave differently for the top as drain or top as source configuration there is no abrupt transition happening in the device. The transition from one mode to another is gradual and continuous through $V_{ds} = 0$ V. Some of the effects discussed above are subdued at low V_{ds}. While a rigorous derivation of the drain current under non-uniform graded channel doping has been achieved previously in [51], the

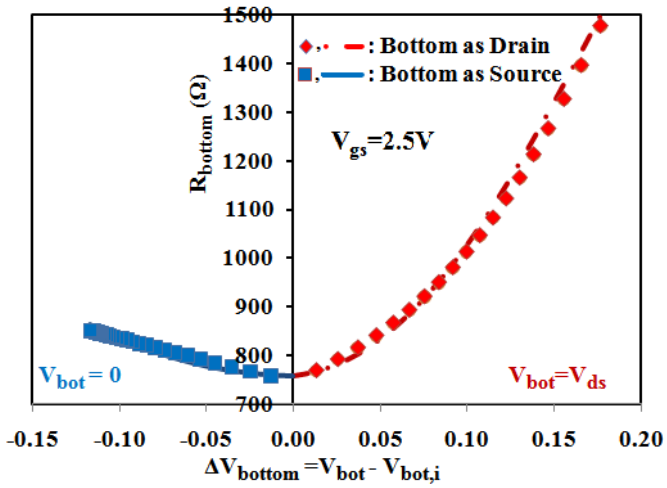


Figure 4.7: Resistance for the bottom electrode drift region for bottom as drain and bottom as source configuration. Quasi-saturation of the drift region leads to apparent increase in the resistace of the region when operating as the drain of the vertical CG-FET. Symbols: Extracted resistance from TCAD using a floating electrode, Lines: Asymmetry aware resistance model

asymmetry under source-drain swapping has not been addressed. That is, the derived model supports channel doping being high at the source and low at the drain and not vice the versa. This was possibly because the LDMOS devices are usually operated under one mode only to take advantage of the drift region as the drain for high voltage drop. On the other hand the vertical CG-FETs could be operated under both the cases wherein the drain end or the source end could have a higher doping. Also for properly designed vertical CG-FET, the gradient is not as much as that observed in LDMOS. X.Zhou *et al.* have attempted to introduce asymmetry in a generic compact model through different values for drain and source end saturation voltage for a MOSFET, [52]. This method was also shown to satisfy continuity across $V_{ds} = 0\,\mathrm{V}$. However this method is limited in scope and applicable only to saturation voltage asymmetry. We will explore an alternate simpler way to introduce asymmetry in a SPICE model through the use of smoothly transitioning asymmetric mathematical functions.

4.2.1 Choice of mathematical functions for asymmetry

Multiple mathematical smoothing functions can be created that would be infinitely continuous and smooth around $V_{ds} = 0\,\mathrm{V}$. However in order for use in a compact model we must be careful to investigate the shape of the higher order derivatives of such functions. We have chosen two functions to illustrate the same.

- **Hyperbolic smoothing function**

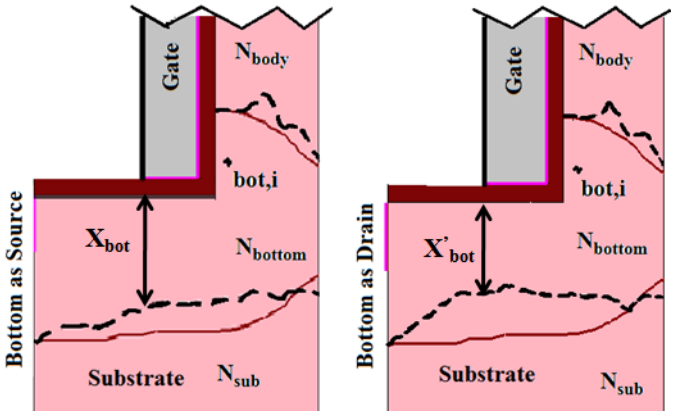


Figure 4.8: Structurally different top and bottom junctions exhibit different resistances (R). The bottom junction exhibits additional increase in R ($\propto X_{bot}^{-1}$) when used as drain due to reverse biased bottom-substrate junction depletion width fluctuation.

Hyperbolic functions are routinely used in a compact model framework to smoothly approach set minimum / maximum limits for parameters. Sum of two such hyperbolic functions each asymptotically approaching zero from positive and negative arguments could be used to create an asymmetric function as follows

$$W_f = \frac{1}{2} + \frac{1}{4}\left[\sqrt{\left(1 + \frac{V_{ds}}{2kT/q}\right)^2 + \delta} - \sqrt{\left(1 - \frac{V_{ds}}{2kT/q}\right)^2 + \delta}\right]$$

$$W_r = 1 - W_f$$

$$P_a = P_f \cdot W_f + P_r \cdot W_r \tag{4.5}$$

where P_a is the overall asymmetric parameter which takes the value of P_f for $V_{ds} > 0\,\mathrm{V}$ and P_r for $V_{ds} > 0\,\mathrm{V}$. δ is a smoothing parameter to tune the abruptness of the transition (typically ≈ 0.01 works).

- **Exponential smoothing function**

Another function that could be harnessed to smoothly approach zero with flexibility in it's smoothness is the exponential function. Tanh() function could be used to create a smooth asymmetric transition in the following way

$$W_f = 0.5 + 0.5 \cdot \tanh\left(k \cdot \frac{V_{ds}}{2kT/q}\right)$$

$$W_r = 0.5 - 0.5 \cdot \tanh\left(k \cdot \frac{V_{ds}}{2kT/q}\right)$$

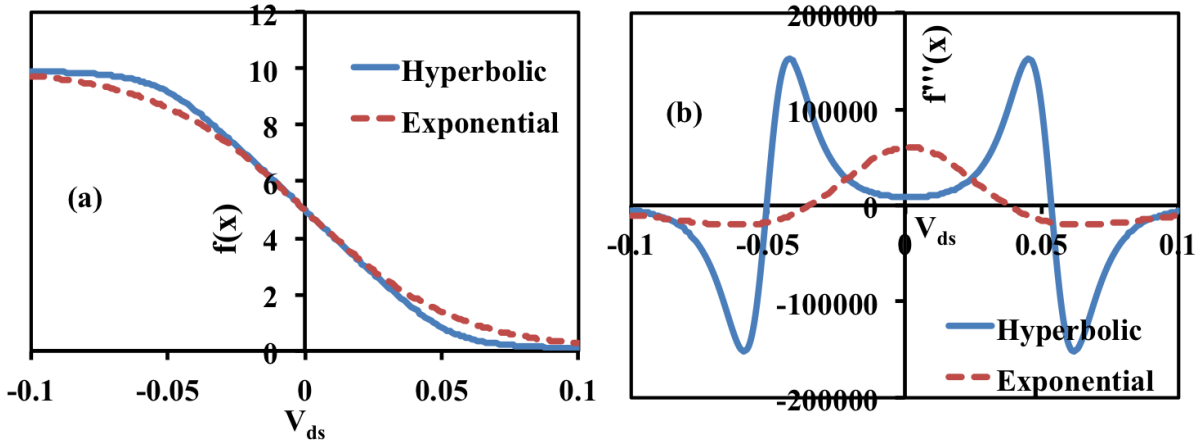


Figure 4.9: (a) Hyperbolic and Exponential functions used to create a smooth transition from 10 to 0 such that their first derivatives at $V_{ds} = 0\,\mathrm{V}$ are equal. (b) The third derivatives of the smoothing functions in (a) are markedly different.

$$P_a = P_f \cdot W_f + P_r \cdot W_r \tag{4.6}$$

where P_a, P_f and P_r take the same meaning as the previous case. k is a tuning parameter to change the abruptness of the function.

The choice of values for δ/k in the above are important. A very abrupt transition would show up as unexpected peaks and valleys in higher order derivatives of drain current (trans-conductance / output conductance etc.). Care should be used when setting the default values for these parameters in order not to produce nonphysical higher order derivatives. The accuracy of trans-conductance and output conductance is paramount in determining linearity related specification in RF CMOS design such as the inter-modulation products in a passive mixer. In order to assess the accuracy we equalize the transition rate from P_r to P_f by choosing appropriate values for $\delta\,(= 0.1)$ and $k\,(= 0.95)$ such that their first order derivative at $V_{ds} = 0\,\mathrm{V}$ are about the same, Fig.4.9(a). We then take a look at the higher order derivatives of the above functions. The third derivative is markedly different for both the functions Fig.4.9(b). The hyperbolic function seems to have two side lobes while the Tanh() function has a peak at $V_{ds} = 0\,\mathrm{V}$. With this observation we conclude that Tanh() function is a better choice by comparing it with third order curves from Gummel symmetry tests of core MOSFET model. The peak at $V_{ds} = 0\,\mathrm{V}$ for the Tanh() function aligns itself with that of the MOSFET itself and is probably less likely to alter the shape of output conductance and trans-conductance curves. Our experience in actual testing of the smoothing functions with BSIM-CMG has been the same.

Armed with a generic smoothing function with which we can induce asymmetry into a symmetric compact SPICE model framework, our goal boils down to identifying a select set

of parameters (physical / empirical) that would capture the requisite amount of asymmetry of the device. Graded channel doping led to asymmetric DIBL and asymmetric onset of saturation in a vertical CG-FET. To capture this phenomenon we choose parameters related to DIBL / output conductance and saturation voltage in the model to induce asymmetry in them. A list of such parameters with reference to BSIM-CMG (version 106.1.0) are below,

- ETA0: V_{th} shift parameter for DIBL
- PDIBL1: Output conductance parameter for DIBL
- VSAT1: Velocity saturation parameter for drain saturation voltage
- MEXP: Linear to Saturation region smoothing parameter
- PTWG: $G_{m,sat}$ degradation parameter

In order to capture the quasi-saturation effect in the drift region at high voltage Eqn.(4.3) was implemented in BSIM-CMG. Allowing for different values for $R_{bulk,d(s)}$, RD(S)DR and PRD(S)DR on the source and drain side (or rather top and bottom for vertical CG-FET) itself lent asymmetry to the device. In addition the following were also made asymmetric,

- RDDR: Drain side drift resistance coefficient
- RSDR: Source side drift resistance coefficient

To illustrate a model using the Tanh() function, the asymmetry in DIBL threshold voltage shift was captured by creating a reverse mode counterpart for ETA0 parameter as follows,

$$\Delta V_{th,DIBL} = -0.5 \cdot \left[\left\{ 1 + \tanh \frac{V_{ds}}{V_t} \right\} \cdot ETA0_f(L) + \left\{ 1 - \tanh \frac{V_{ds}}{V_t} \right\} \cdot ETA0_r(L) \right] \cdot V_{ds} \quad (4.7)$$

For small-signal capacitance we retain the previously formulated charge model using Ward-Dutton partition in Chapter 2, [35]. Strictly speaking, for laterally asymmetric device work by A.C. Aarts et al. and A. Roy et al. show that Ward-Dutton partition is not applicable [53, 54]. Additionally A. Roy et al. have proposed a new partition function that is dependent on the doping gradient function for the derivation of small-signal capacitance, [54]. However we found that for vertical CG-FETs designed for DRAM technologies do not have a high doping gradient and hence Ward-Dutton partition based terminal charge / capacitance model was still applicable.

4.3 Model Extraction

While extraction of MOSFET parameters itself could be iterative in some ways, the extraction of a high-voltage MOSFET with asymmetric source / drain bias dependent resistances could be even the more iterative. The iteration is in determining the correct set of resistance

values and intrinsic MOSFET parameters that are self consistent with each other. In this process it helps to have an almost equivalent symmetric test-structure at hand to determine the intrinsic MOSFET parameters. For ex: a device with high source / drain doping and negligible drift region adds to almost no resistance to the FET allowing for the characteristics to be close to that of the symmetric MOSFET. However not every time does one have the privilege of silicon area. In such cases the technique by A. Blaum *et al.* could be used to roughly estimate asymmetric source and drain resistance values which we will discuss here briefly, [55]. While direct extraction of the resistances individually is not possible (unless there special structures for direct probing of internal points are available) one could estimate the sum and difference of the resistances from terminal current and voltages as follows. This is based on the premise that linear region (low V_{ds}) current is influenced by both the source and drain side resistance but the saturation region current (high V_{ds}) is affected only by the source side resistance. In the linear region and strong inversion the resistance of the channel is negligible and hence the sum of the source and drain resistance is given as follows,

$$R_s + R_d = \frac{V_{ds}}{I_{ds}} \tag{4.8}$$

In the saturation region, the drain current is independent of drain to source voltage and depends only on the drain saturation voltage. The drain saturation voltage in turn depends on the gate to internal source voltage. The internal source voltage depends on the voltage drop across the source resistance. The drain saturation voltage is given by,

$$V_{d,sat} = V_{gs} - V_{TH} - I_d \cdot R_s \tag{4.9}$$

where V_{gsi} is the gate to source voltage drop, V_{TH} is the threshold voltage and I_d is the drain current. If the resistance at either end are different then at the same gate voltage the MOSFET should see different drain currents flowing through for the forward and reverse mode. Alternately, for the same drain current (same drain saturation voltage), the MOSFET would require different gate voltage (say V_{gsf} and V_{gsr}). Then the difference in the resistances can be given by,

$$R_d - R_s = \frac{V_{gsr} - V_{gsf}}{I_d \left(1 + \frac{\partial V_{TH}}{\partial V_{sb}}\right)} \tag{4.10}$$

The above equation assumes that the threshold voltage does not vary with V_{ds}, nevertheless a good approximation for difference of resistances. With the sum and difference we can estimate the value for source and drain resistances and use them as initial guess values for parameter extraction. This could be repeated for different gate voltages / drain current to roughly extract the voltage dependence of the same.

While the iterative process cannot be possibly outlined Fig.4.10 can be used as a guideline for parameter extraction. It illustrates the rough area of influence for some of the resistance and saturation parameters on a set of asymmetric $I_d - V_d$ curves.

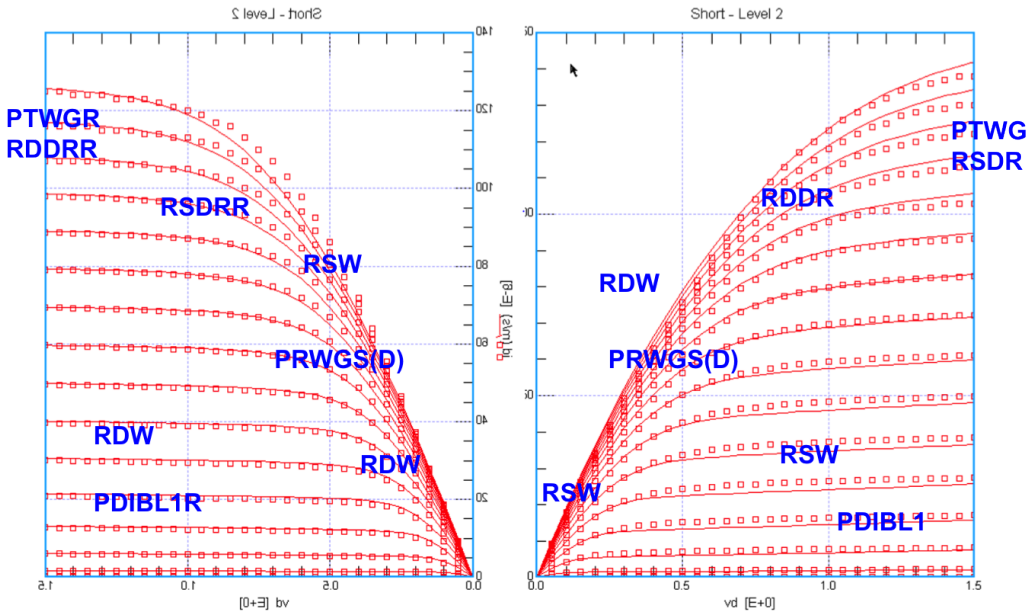
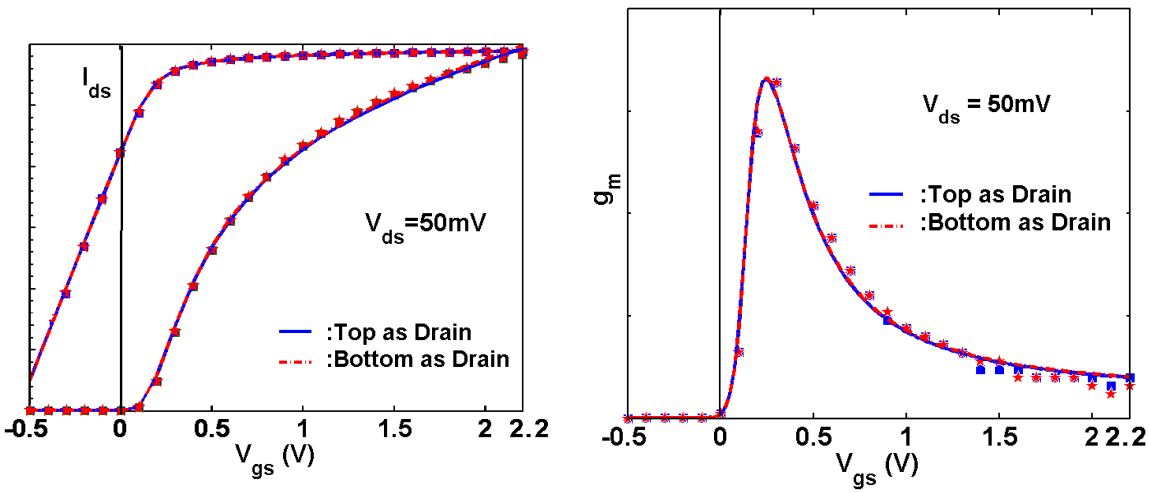


Figure 4.10: Region of influence of asymmetric drift region resistance and asymmetric velocity saturation parameters on drain current curves for both top as drain and top as source configuration during the process of parameter extraction.

4.4 Validation

After identifying the physical causes of asymmetry and choosing a select set of parameters in the existing model to attribute asymmetry to, we validate the developed model on a set of hardware data from a vertical CG-FET 2.2 V device. The vertical CG-FETs on bulk substrate were 150 nm tall (channel length), 60 nm channel diameter and 4.5 nm oxide thickness to sustain high voltages of operation. As expected from our discussion the top as drain and top as source configuration have almost identical current in the linear region ($V_{ds} = 50\,\text{mV}$), Fig.4.11a. The inherent mobility degradation and resistance models in BSIM-CMG help capture the drain current from sub-threshold to strong inversion and as well as the trans-conductance for both the configurations, Figs.4.11a and 4.11b. The asymmetry shows up for high V_{ds} operation, Fig.4.12. The developed model for capturing asymmetry in DIBL and onset of saturation help capture the saturation region drain current for both top as drain and bottom as drain configurations, Fig.4.12. The drain currents are not only different but we also observe a significant difference in the saturation region trans-conductance ($G_{m,sat}$), Fig.4.13. The degradation of $G_{m,sat}$ at higher gate voltages is a sign of quasi-saturation in the drift region. The bias dependent drift region resistance model not only helps capture the $G_{m,sat}$ degradation but also the gate voltage de-sensitization (current crowding) observed in $I_d - V_d$ curves, Fig.4.14. Separate parameters for the source and drain drift region and the asymmetry added in them help capture the difference in $G_{m,sat}$ and saturation region drain current for both the top as drain and bottom as drain configurations.



(a) (b)

Figure 4.11: (a) Linear region drain current for the top as drain and bottom as drain configurations shows no asymmetry. (b) Linear region transconductance, ($G_{m,lin}$) indicates that there is no asymmetry arising from mobility degradation in the channel. $G_{m,lin}$ also depends on sum of the source and drain resistances which is independent of the configuration. Symbols: Hardware Silicon Data, Lines: BSIM-CMG Model enhanced for asymmetry

In order to verify the continuity of the model across $V_{ds} = 0\,\mathrm{V}$, we sweep a symmetric drain to source voltage source across the FET in a passive mixer configuration (similar to Gummel symmetry test, [40]). The drain current and its derivatives are plotted at multiple gate voltages, Fig.4.15. We observe no discontinuities in the curve ascertaining the continuity of the model. Unlike the standard Gummel symmetry test the values for the drain current and its derivatives do not have to be symmetric across $V_{ds} = 0\,\mathrm{V}$ due to the inherent yet smooth asymmetry of the model.

4.5 Summary

A study of the primary causes of drain current asymmetry in vertical cylindrical gate (CG) FETs was undertaken. Graded channel doping from drain to source and asymmetric behavior of drift (extended) source / drain region have been identified as the leading causes of asymmetric drain current especially at high drain to source voltages of operation. This understanding was further translated into a SPICE model to be incorporated in the existing multi-gate model, BSIM-CMG. Developed model was validated with hardware data from an asymmetric vertical CG-FET technology operating up to 2.2 V. Excellent fitting results ($< 1\%$ RMS error) to hardware data were obtained.

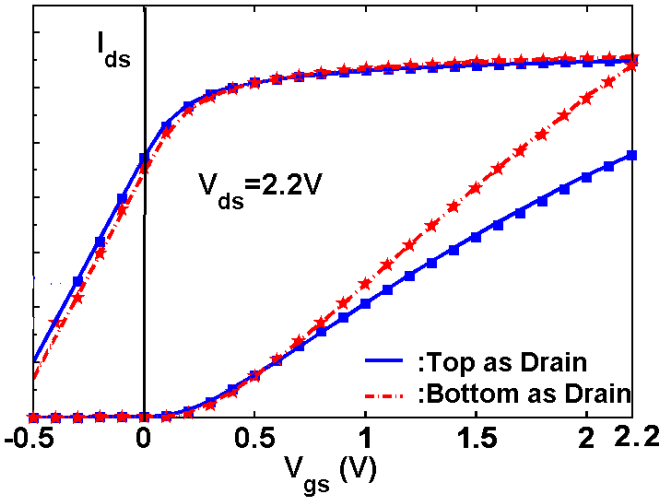


Figure 4.12: Saturation region drain current exhibits asymmetry in off-current (due to asymmetric DIBL) and asymmetry in on-current as well (due to asymmetric on-set of saturation) Symbols: Hardware Silicon Data, Lines: BSIM-CMG Model enhanced for asymmetry

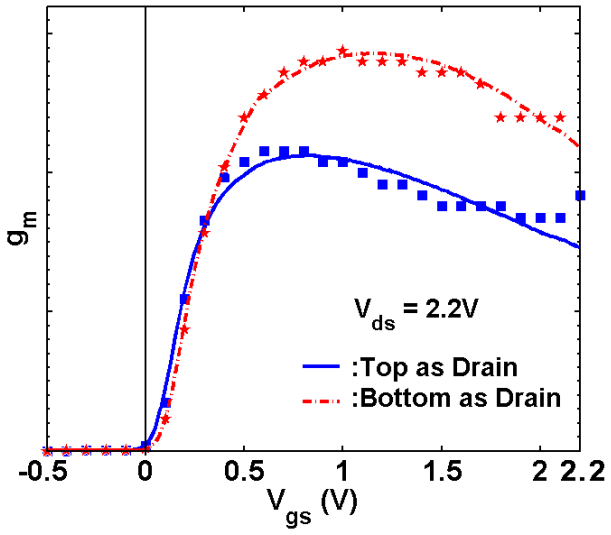


Figure 4.13: Transconductance in saturation region, $G_{m,sat}$, shows marked asymmetry in its characterisitics with different amount of degradation. Symbols: Hardware Silicon Data, Lines: BSIM-CMG Model enhanced for asymmetry

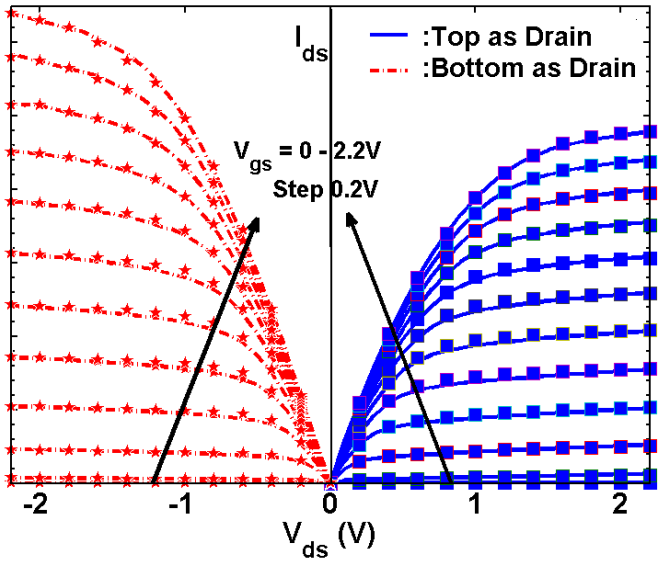


Figure 4.14: Drain curent vs. drain voltage for various gate voltage for both the top as drain and bottom as drain configurations. Negative V_{ds} is used to indicate the bottom as drain configuration to showcase both the configurations in one plot. Symbols: Hardware Silicon Data, Lines: BSIM-CMG Model enhanced for asymmetry

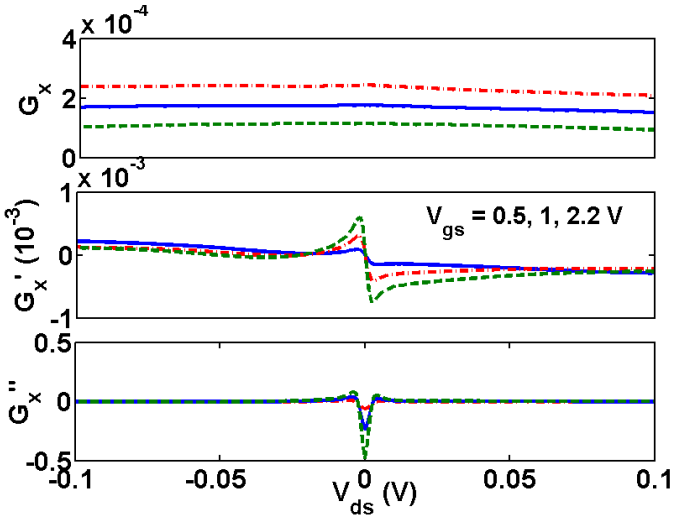


Figure 4.15: Results from a drain current continuity test. First, second and third derivatives of current in response to a symmetric drain-source voltage source are continuous and smooth around $V_{ds} = 0\,\mathrm{V}$ Symbols: Hardware Silicon Data, Lines: BSIM-CMG Model enhanced for asymmetry

Chapter 5

RF CMOS Modeling - Art of Parameter Extraction

RF performance of Si CMOS has improved greatly with geometry scaling. Cost sensitive commercial RF and mm-wave systems are increasingly benefiting from advances in silicon CMOS technology. However foundry provided SPICE models continue to be targeted towards digital/analog design flows. These models are extracted at low frequencies (over limited bias, wide range of length and width) and do not adequately cover higher frequencies. To make things even more difficult is the additional layer of complexity introduced by the varying styles of RF device layout and de-embedding procedures. There is no standard around RF device layout, the designer has a plethora of choices (like single or double side gate contact, dual side or all around substrate contacts) that trade of parasitic resistance and capacitance with layout complexity. Also, different RF applications require model accuracy in different frequency ranges. RF blocks like low noise amplifiers (LNA) work around a quiescent point (bias point, small signal operation). For such blocks the accuracy around this bias point suffices. On the other hand, RF blocks like power amplifiers (PA) and passive mixers are subject to large signal variation and must therefore be accurate across all bias voltages.

BSIM6 is an enabling update to the industry-standard BSIM4 model for RF circuit designers, [18]. BSIM6 uses a physical charge based core model to accurately predict the drain current and terminal charges of a MOSFET [20]. The model also captures a range of real device effects such as short-channel effects, mobility degradation, velocity saturation, drain-induced barrier lowering (DIBL), and channel length modulation. BSIM6 also contains models for series resistance, parasitic overlap and fringe capacitance (including their bias dependent behavior) which are very important for accurate prediction of RF device performance metrics. The gate resistance, substrate network and junction capacitance that affect the device behavior considerably at high frequencies are also accounted for in the model. Multiple options that trade off simulation speed and accuracy are provided for capturing the non-quasi static (NQS) effects (frequencies $> F_t$, the unity current gain cut off frequency). The model borrows many of the parameter names and conventions from BSIM4 thus aiding easy transfer of parametric knowledge from BSIM4 to BSIM6. BSIM6 meets stringent

compact model requirements for RF design which we will discuss later.

Any SPICE model (BSIM6 or advanced multi-gate models) is not complete without an associated extraction methodology. Many procedures have been developed in the past. Some of them include altering existing SPICE models by augmenting them with additional sub-circuit elements, [56]. Analytic models have been developed assuming a MOSFET small-signal approximate circuit schematic whose element values were then manually extracted, [57]. However this method is restricted in terms of scalability across device parameters, model portability across simulators and applications. In recent times software automation tools have been developed for parameter extraction and model creation, [58],[59]. However this has been used inefficiently wherein parameter optimization is performed on all measured data (both DC and small-signal S-parameter) in one shot, [60].

Having observed the need for a new RF model parameter extraction procedure for MOSFET compact models, in this chapter we will first highlight and discuss the RF relevant models in BSIM6. We will then describe the theory behind the proposed extraction procedure. We aim to create an approach that takes advantage of availability of parameter extraction tools yet is efficient. We elucidate a set of simple steps to be followed that covers both bias and broad frequency range. This intuitive methodology can be easily integrated directly into any RF design flow. This procedure in our experience takes only few hours to get a RF ready device modelcard. Finally we will discuss multiple use-case strategies while validating the procedure on S-parameter hardware data from multiple CMOS technologies for both long and short channel length transistors.

5.1 RF relevant models in BSIM6

In order to be able to capture the high frequency effects in a MOSFET, the drain current and terminal capacitance need to be more accurate than what is required of a model serving the purpose for digital and analog circuit design. The MOSFET device is a storehouse of parasitic resistance and capacitance, Fig.5.1. A number of additional models to capture these parasitics ranging from gate-resistance, substrate resistance, and junction diodes etc. need to be included. In this chapter we will restrict ourselves to extraction of the above discussed components only. Device noise is yet another important RF phenomenon that needs to captured in a compact model framework. Channel noise, correlated induced gate noise, shot noise and resistance thermal noise etc. are all captured in BSIM6. For a detailed discussion and extraction of channel noise and induced gate noise the reader should refer to [17]. Let us now indulge in a brief discussion of each of the models to comprehend their role in a RF model accuracy.

5.1.1 Drain Current and its derivatives

BSIM6 at its core contains a single piece smooth analytic equation for long channel drain current that is derived from a physical charge-based model. This charge-based, single piece

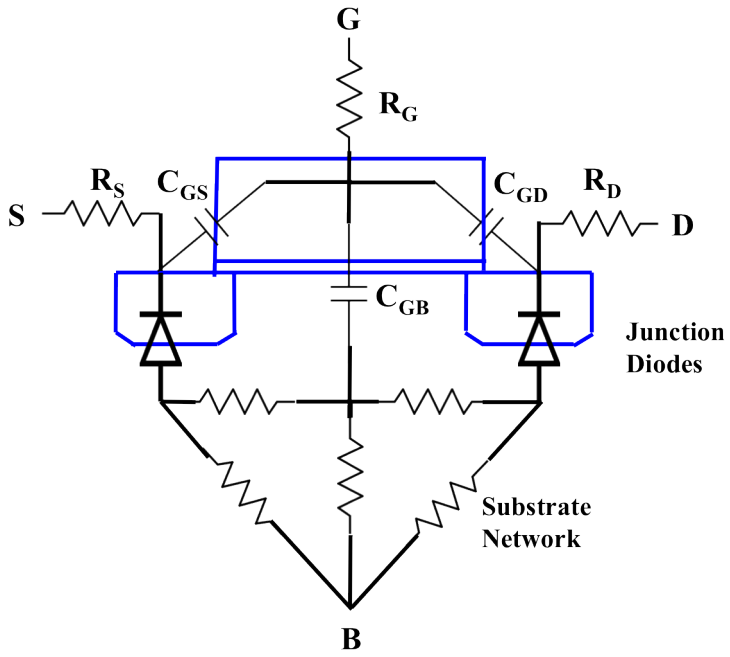


Figure 5.1: Overlay of intrinsic MOSFET and it's high frequency relevant model components.

nature as opposed to BSIM4's threshold voltage based two-piece (sub-threshold and strong inversion put together with a smoothing function) model renders it relatively more physical in behavior. This coupled with sub-models for various second order effects like mobility degradation, velocity saturation, and drain induced barrier lowering (DIBL) give it accuracy in predicting the derivatives of drain current *viz.* trans-conductance and output conductance. Similar to BSIM-CMG care has been taken to ensure symmetry and continuity of the BSIM6 model around $V_{ds} = 0\,\mathrm{V}$, [17]. As a result BSIM6 passes the traditional Gummel symmetry test as well, Fig.5.2 .

The accuracy of the higher order derivatives of drain current is important in predicting various RF linearity metrics like input referred third harmonic intercept (IIP3), harmonic and inter-modulation distortion products (HD_2, HD_3, IM_2, IM_3 etc.) and adjacent channel power (ACPR) etc. Although these metrics are also known to depend on source and load impedance, a simple model for IIP3 is given by

$$IIP3 = \frac{1}{6R_s}\frac{(1+\omega C_{gs}R_s)^2}{G_m''/6G_m + \Delta} \tag{5.1}$$

where R_s is the source resistance, G_m'' is the third derivative of drain current w.r.t. gate voltage, G_m is the trans-conductance and Δ is a term that depends on a variety of partial derivatives of drain current w.r.t. gate and drain voltages, [61]. The sweet spot for high linearity biasing in RF applications is determined by the zero of the denominator of Eq.(5.1). Δ is a function of bias whose accuracy is dependent on output conductance (DIBL), current

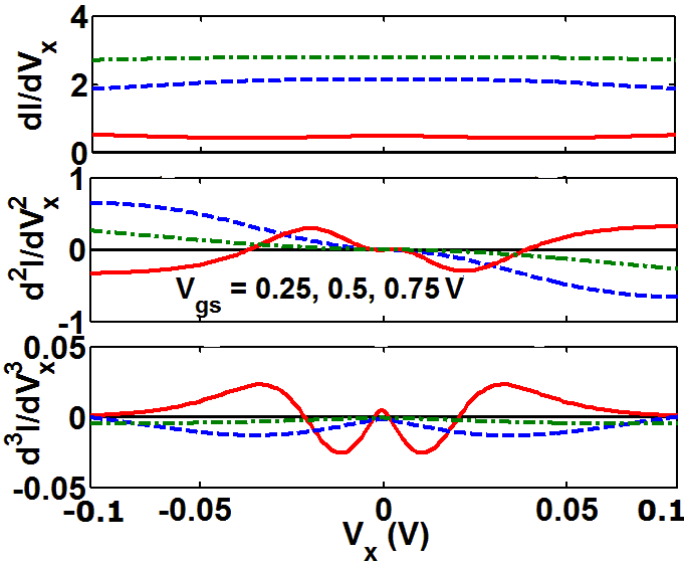


Figure 5.2: Results of Gummel Symmetry Test verify the symmetry and continuity of the BSIM6 drain current and its higher order derivatives, [40]

saturation and mobility degradation sub-models. Hence having very accurate sub-models in a compact SPICE model environment is paramount. Owing to its inherent symmetry and higher order continuity, BSIM6 also satisfies the harmonic-balance test where the harmonic distortion in response to a small-signal single tone excitation (at the source terminal) is observed at the drain for a constant gate voltage, [19] . On a plot of output *vs.* input power the slope of the n^{th} harmonic content should be n, Fig.5.3

5.1.2 MOS Intrinsic and Parasitic Capacitance

Along with the drain current, BSIM6 also consists of device intrinsic capacitance and trans-capacitance described through a charge based model. While in a digital design environment one is required to capture the average capacitance behavior over bias in order to predict accurate logic delay, RF design poses stringent requirements in terms of getting the overall shape of the capacitance right. Thus unlike the long channel capacitance model derived in Chapter 2, one requires to incorporate short channel effects into the charge model as well. In a long channel MOSFET, the gate capacitance decreases from its maximum value of WLC_{ox} to $2/3 \cdot WLC_{ox}$ with increasing drain-source voltage (V_{ds}) as the channel pinches off at the drain while the current saturates. However in short channel length MOSFETs the gate capacitance settles down at a value higher than $2/3 \cdot WLC_{ox}$ owing to channel length modulation, wherein the pinched off channel region becomes an increasing component of the actual channel length . The transition to this value from the maximum value happens around the drain saturation voltage, $V_{ds} = V_{dsat}$ that is determined by onset of velocity saturation rather than the channel pinch off. A detailed discussion of how these effects are

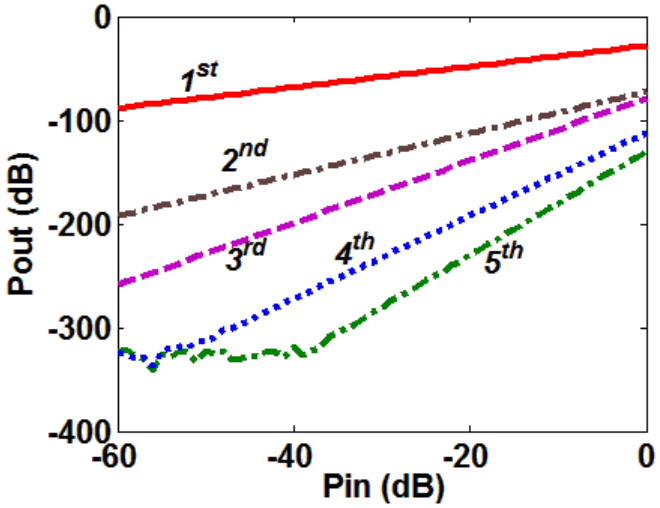


Figure 5.3: BSIM6 satisifies the Harmonic Balance test as well. The slope of the n^{th} harmonic power at the drain terminal is n when the MOSFET is excited at the source terminal with a single tone.

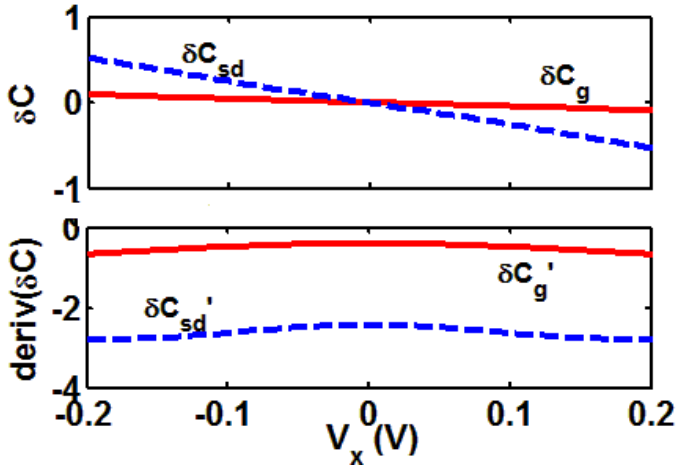


Figure 5.4: Results of AC Symmetry Test verify the symmetry and continuity of the BSIM6 capacitance, trans-capacitance and its higher order derivatives, [40]

incorporated in the charge model can be found in Appendix B. This along with quantum mechanical charge centroid model (which has been shown to affect the moderate inversion region of the gate capacitance in Chapter 3) render accuracy to the bias dependence of imaginary component of input admittance (Y_{gg}).

Apart from the intrinsic capacitance BSIM6 also includes overlap and fringe parasitic capacitance. The gate-drain/gate-source overlap capacitance arises from the gate electrode

overlap over the source/drain region along the length of the device. This capacitance is bias dependent due to the fact that the source/drain region traverses from depletion to accumulation as the gate-source/drain voltage ($V_{gs/d}$) becomes more positive. It is also possible that this region could be inverted for large negative $V_{gs/d}$, for ex: at the drain end of a NMOS in an inverter. This together with the external fringe capacitance influences the imaginary component of gate-drain trans-admittance (Y_{gd}). The small inner fringe capacitance through the channel that cannot be isolated through measurements is usually neglected in compact models.

Similar to DC current, the charge based model ensures symmetry in capacitance as well. BSIM6 capacitance model passes the AC symmetry test when subjected to a symmetric voltage across source and drain terminals. Fig.5.4, [40].

5.1.3 Junction Current and Capacitance

The source/drain - body junctions (S/D junctions) form an integral component of a RF MOSFET. Though for majority of applications the junction is typically in reverse-bias mode (to avoid high junction currents), it is not unusual in some large signal RF circuits (such as power amplifiers and switches) where the junction might get forward biased as well. Hence model accuracy is important for even forward bias region. In a MOSFET compact model three diode components are used to represent the metallurgical junctions on three different sides namely - bottom (area component), isolation side (perimeter component) and gate / channel side. The legacy BSIM4 three-component junction current model covering both forward and reverse bias currents and accounting for junction tunneling leakage currents was chosen for BSIM6, [38].

A good junction capacitance model lends accuracy to the imaginary component of the output admittance (Y_{dd}). At frequencies $> 10GHz$, this junction capacitance behaves almost as a short. The real component of Y_{dd} which is dominated by the MOSFET intrinsic output conductance (G_{ds}) is now influenced by the substrate network conductance. Accuracy in junction capacitance is important to capture this transition in the real component of Y_{dd} as well (usually around $1 - 10\,\text{GHz}$ in bulk MOSFETs). The junction capacitance is usually made linear on and after a certain bias value to contain the capacitance numerically for model stability. This approximation introduces a second order discontinuity. As discussed in Chapter 3 for BSIM-CMG, in BSIM6 as well the position of this discontinuity has been engineered away from $V_{bs/d} = 0\,\text{V}$ point (as opposed to in BSIM4) in order for accurate higher order distortion predictions of a non-linear MOSFET.

Often a doubt is cast regarding when to involve high-frequency effects of a diode, if any. For frequencies approaching the cut-off frequency (transit time) of the diode or when the diode is subject to high speed transients, one needs to take into account the charge storage in the quasi-neutral region of the diode [3]. This charge storage time is set by the minority carrier storage time. The minority carrier storage time is in turn inversely dependent on the

amount of impurities / doping / traps in the junction, (N_T)

$$\tau = \frac{1}{\sigma v_{th} N_T} \tag{5.2}$$

where σ is the carrier capture cross-section and v_{th} is the thermal velocity. These impurities / doping / traps are a result of the rapid thermal annealing step after junction dopants implantation. Measurements of minority carrier lifetime in well-annealed n-type silicon have shown that for doping concentration approaching $10^{20}\,\mathrm{cm}^{-3}$ and higher the lifetime approaches $1/10^{th}$ of a nano-second, [62]. It is safe to presume that in contemporary MOSFET junction the lifetimes are one to two orders of magnitude lower pushing the diode cut-off frequency to $0.5-1\,\mathrm{THz}$, which is well above the normal operating frequencies of interest today. However with ever increasing cut-off frequency of the intrinsic MOSFET and advent of CMOS wireless applications in research at 300 GHz, this assumption stands to be verified. For high voltage MOSFETs like LDMOS and DEMOS where a lowly doped source / drain region is used to drop the voltage, the minority carrier lifetime is indeed lower in this case and thus junction diode models for these devices need to consider non quasi-static effects like reverse recovery.

5.1.4 Gate Electrode Resistance

Gate Resistance influences the accuracy of real component of the input impedance (Y_{gg}). In a compact SPICE model it is usually captured with two components. One a physical gate electrode component and the other a virtual gate induced channel resistance component.

The physical gate electrode resistance represents the distributed resistance network of the poly-silicon or metal gate material over the channel. It is simply given by

$$R_{g,eltd} = k.\frac{W}{L}\frac{R_{shg}}{NF} \tag{5.3}$$

where R_{shg} is the sheet resistivity of the gate, W is the channel width per finger, L is the channel length and NF is the number of fingers in a device. Usually corrections are applied to W and L to account for any layout related discrepancies. The pre-factor k is a constant derived from a distributed R-C transmission line model. For a single side gate contact $k = 1/3$ and for double side gate contact $k = 1/12$. Caution must be exercised when activating this component. Some of today's process design kits seem to incorporate this component during parasitic extraction external to the SPICE model although with $k = 1/2$. In RF design it is usual practice to layout the device where the width is broken down into multiple fingers, NF in order to reduce the loses pertaining to this component of gate resistance. Increasing gate resistance hampers f_{max}, the maximum frequency for oscillation and the maximum stable gain (MSG) of the device. However having a very large NF to reduce this component is not advisable for the resistance of the external taper section of the layout contacting the gate would begin to dominate. Hence there is a trade-off between W and NF that must be dealt pragmatically by understanding the device

layout. Also while single side contact is preferred at mmWave frequencies where one tries to minimize the parasitic capacitance picked up between the gate leads running around and the source/drain/body traces, double side gate contact is preferred at RF frequencies where one has the liberty with the amount of extra parasitics that can be easily tuned out with on-chip inductors. If a very highly accurate and scalable gate electrode resistance model is required to capture a wide geometry of devices then splitting the gate resistance components further into bulk, contact and interface resistances with appropriate scaling is recommended, [63].

5.1.5 Non-Quasi Static Effects

Non-Quasi Static effects refers to the lag in response to a high-frequency small-signal or a high speed large signal transient applied to the MOSFET. For these signals the lumped model of the channel (as approximated by a quasi-static model) is no longer applicable and the distributed nature of the channel needs to be taken into consideration. The model so far developed (DC, small-signal charge/C-V etc.) was developed considering only the device electrostatics through the Poisson equation and ignoring the effect of the continuity equation along the channel. This implies that the terminal currents and charges are instantaneously available which indeed is not the case (this would mean we could obtain infinite gain from the device at higher frequencies). While more details and derivations for various non-quasi static effect models can be found in Appendix D, we will restrict to a brief overview of the options here.

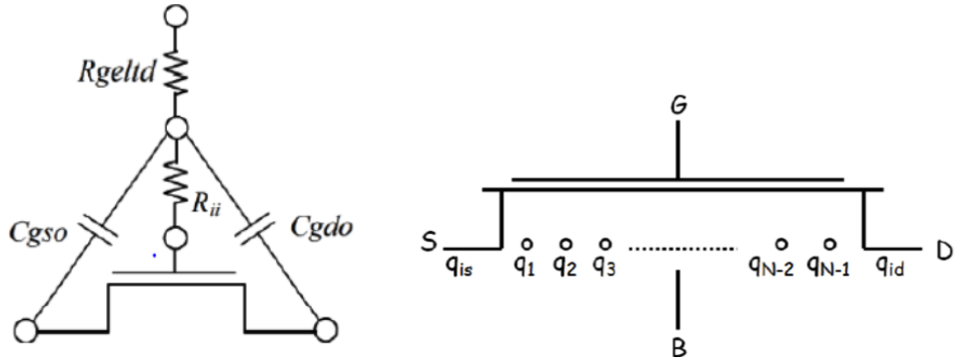
Similar to BSIM4 bulk planar model, BSIM6 offers both the channel induced gate resistance and the relaxation time sub-circuit approach. In the channel induced gate resistance approach, a virtual gate resistance in series with the physical gate resistance is added

$$R_{ii} = \left[X_1 NF \frac{\mu_{eff} W}{L} \left(Q_{ch} + X_2 \frac{kT}{q} C_{ox} \right) \right]^{-1} \tag{5.4}$$

where C_{ox} is the gate capacitance per unit area, μ_{eff} is the effective mobility (including both mobility degradation due to vertical field and velocity saturation due to lateral field) and Q_{ch} is the total channel charge (which is a function of gate and drain bias), Fig.5.5a,[64]. Parameters X_1 and X_2 are used to tune the bias dependence and recover from any approximation in the analytic expression for Q_{ch}.

Relaxation time approach captures the deficient or surplus charge in the channel relative to the steady state quasi-static charge using a R-C sub-circuit whose node voltage is the deficient charge, [65]. The relaxation time constant (R-C time constant), τ is the same as the the channel resistance in Eqn.(5.4) (assuming the capacitance is unity). This sub-circuit is solved self-consistently with the quasi-static model and the so calculated non quasi-static terminal charges replace the quasi static charges. When this sub-circuit approach is used, the channel induced gate resistance is no longer required and if used could lead to double-counting of the non-quasi static effect.

BSIM6 also offers a full-fledged charge segmentation based non-quasi static model. In this case the channel is broken down into multiple charge segments and current continuity



(a) Channel induced Gate resistance Model

(b) Channel segmentation Model

Figure 5.5: Model options in BSIM6 for capturing non-quasi static effects. For this work NQSMOD=1 refers to (a) and NQSMOD=2 refers to (b)

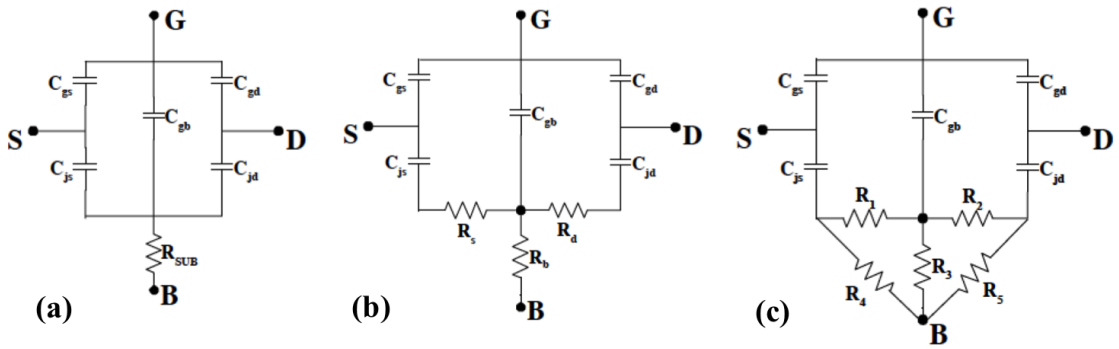


Figure 5.6: Some possible substrate / well-region network models - (a) one resistor (1R), (b) three resistor (3R) and (c) fice resistor (5R) networks

equations at the intermediate internal nodes are solved within the SPICE environment itself, Fig.5.5b. In this case the only input expected from the user is the choice for number of segments, N_{SEG}. Higher the number of channel segments, higher the accuracy but longer the simulation time. Similar to the relaxation time approach, using this option requires **no additional virtual gate resistance**, $R_{g,ch}$ at the gate.

5.1.6 Substrate Network

At high frequencies ($> 1GHz$) the impedance due to MOSFET intrinsic capacitance tend to become small. The behavior of the well region below the channel and junction now influences the 2-port parameters describing the MOSFET at these frequencies. The substrate / well-region is modeled as a five resistor network as in Fig.5.6(c) . The choice of the resistor network (one vs. three vs. five resistors) was made after a TCAD based study of the silicon

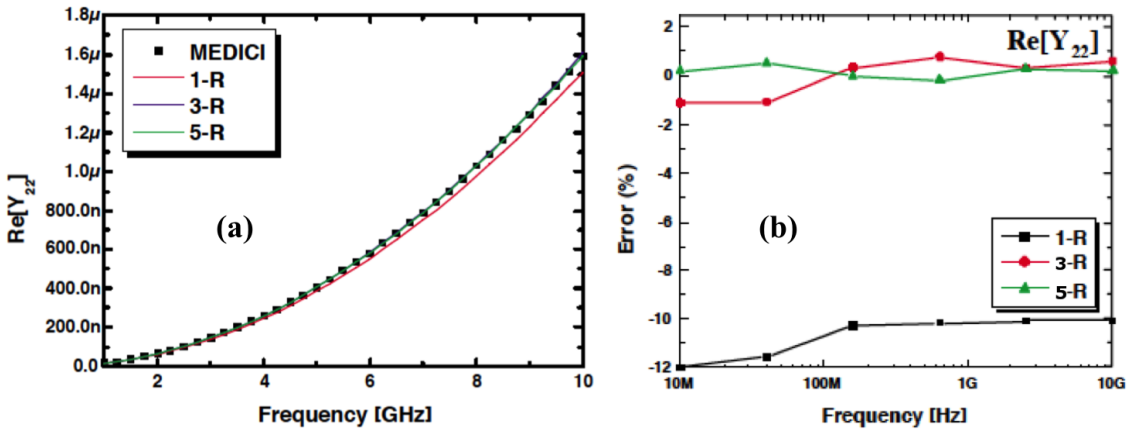


Figure 5.7: (a) Output conductance, $Re(Y_{dd}) = Re(Y_{22})$ plot surveying the accuracy of various substrate network models in comparison to TCAD data (b) Error in the resultant fit shows that the five resistance substrate network best describes the substrate for frequencies above 1 GHz, [66]

substrate with realistic well doping implant and multiple substrate contact configurations (two sided / all around the device etc.) previously in our group, [66]. As per the study in [66], the resistor values for substrate networks in Fig.5.6 were extracted from 2-port network small-signal simulations of a 180 nm NMOS device in off-state ($V_{gs} - V_{th} = -0.4V$) well below the threshold voltage of the device. A Gaussian doping profile was used in the substrate with a well-doping of about $10^{18}/cm^3$ at a depth of $0.5\mu m$. After a best-fit to the small-signal 2-port parameters was obtained, the error in the resultant fit was evaluated, Fig.5.7. $Re(Y_{dd})$ or $Re(Y_{22})$, the output impedance is most affected by the substrate network model accuracy. From Fig.5.7(b) we can easily deduce that the five resistor network is the model of choice for RF frequencies (up to $10GHz$) owing to its low error ($< 1\%$ rms) when capturing the TCAD results while the three resistor and one resistor models have up to 1% and 12% maximum error in the fit. The implemented model in BSIM6 also allows for generalized scaling of the parameters w.r.t. width, length and number of fingers of a FET.

The substrate like any other material is both a dielectric and a conductor, [67]. Hence the substrate should be represented as a resistor in parallel to a capacitor as normally used for passive device models on silicon substrate, [68]. However we can show that the capacitive effects of the substrate tend to dominate at very high frequencies when it is moderate - heavily doped (see Appendix C). Thus the five resistor network model should continue to be sufficient even for the mmWave frequencies but might need to be re-visited for THz range frequencies depending on the substrate / well doping.

5.2 Parameter extraction flow - Theory

From the previous section we can gather that BSIM6 is a RF ready model and has all the requisite components to predict a RF MOSFET. However, the values for the components (especially parasitic resistance and capacitance) need to be accurately extracted from measured data of the MOSFET. For this we will assume that we are given the following,

- DC measurement data for the FET spanning all terminal bias ranges of interest
- RF measurement data i.e. 2-port S-parameter data. The data should cover the whole frequency range and also the bias across gate and drain voltages.

The S-parameter data needs to be accurately de-embedded with appropriate procedures that maintain the accuracy up to and above the center frequency of interest [69]. Otherwise layout and lead parasitics could fold into the MOSFET model affecting the scaling predictions of the model. Some of these parasitics tend to heavily influence the high frequency behavior masking the actual MOSFET behavior leading to incorrect estimation of the values for the RF MOSFET components.

An ideal parameter extraction procedure should have the following features that enable accurate estimation of the component values,

- The targeted data set should easily obtained by direct measurements.
- A sub-set of this data should be identified that is predominantly influenced by a small set of parameters (that correspond to a physical effect).
- The complete extraction should be achievable in a single attempt without any iteration, i.e. extraction of parameters at step n should not affect (but just rely on the accuracy of) curve fit at step *n-1*.

For example, the extraction procedure outlined for MOSFET drain current in industry standard model for BSIM4 has the above outlined features, [38]. In what follows we will try to adhere to the same as much as possible.

Although it is standard practice to measure 2-port S-parameters (as Z and Y parameter measurements require ideal short or open circuits which cannot be obtained at high frequencies) we will discuss in terms of Y-parameters. Associating Y-parameters to device components is easier and helps easy interpretation of the data sub-sets. Although we will not rely on any particular small-signal model interpretation for the MOSFET, for extraction purposes it helps to visualize one to create approximate analytic expressions for the Y-parameters, Fig.**??**. A low frequency approximation for 2-port Y-parameters (port 1 - gate, (g) and port 2 - drain, (d) with source and body terminals tied to ground) is as follows

$$Y_{gg} = \omega^2(C_{gg}^2 R_g + C_{gd}^2 R_d + C_{gs}^2 R_s + Cgb^2 R_{sub,g}) + j\omega C_{gg} \tag{5.5}$$

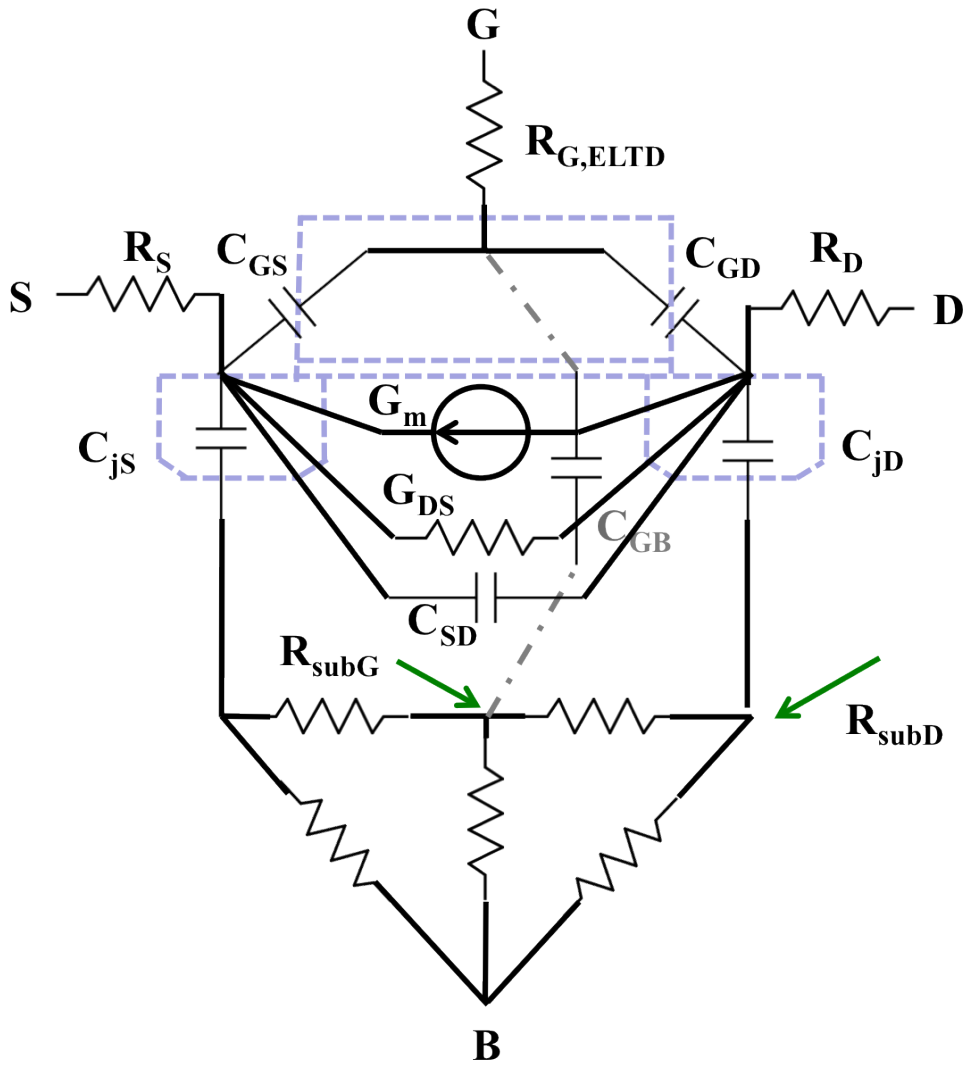


Figure 5.8: Small signal equivalent representation of a MOSFET for the purposes of parameter extraction

$$Y_{gd} = -\omega^2 C_{gd} C_{gg} R_g - j\omega C_{gd} \tag{5.6}$$

$$Y_{dg} = G_m - \omega^2 C_{gd} C_{gg} R_g - j\omega C_{dg} - j\omega G_m R_g C_{gg} \tag{5.7}$$

$$Y_{dd} = G_{ds} + \omega^2 C_{jd}^2 R_{sub,d} + \omega^2 C_{gd} C_{dg} R_g + j\omega C_{jd} + j\omega C_{sd} + j\omega C_{gd} + j\omega G_m R_g C_{gg} \tag{5.8}$$

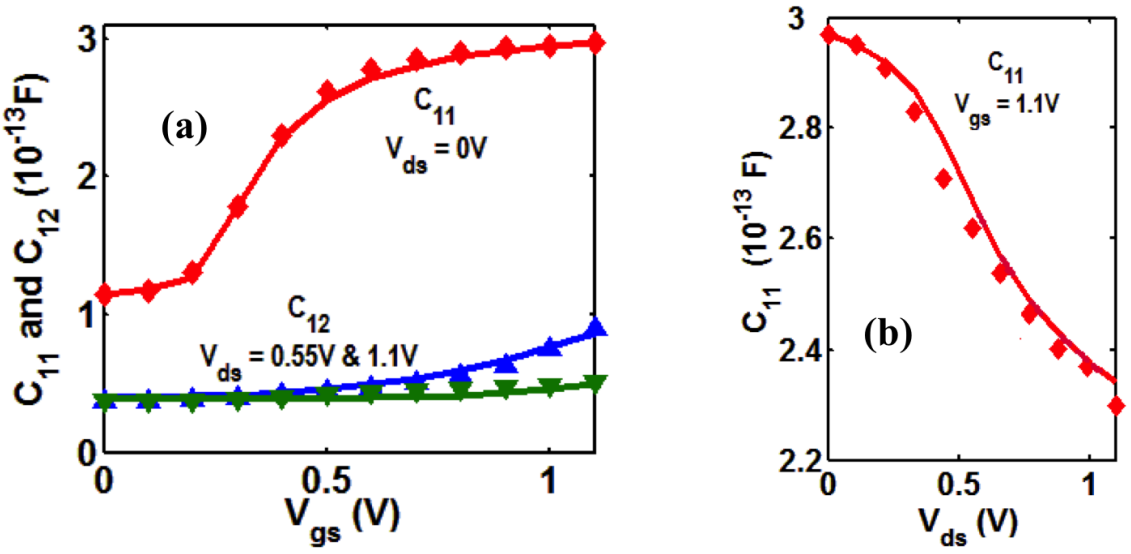


Figure 5.9: Process technology related parameters, overlap + fringe parasitic capacitance, short-channel capacitance parameters are extracted from $C_{gg} = C_{11}$ and $C_{gd} = C_{12}$. Symbols=Silicon Data, Lines=BSIM6 Model.

5.2.1 Intrinsic and Extrinsic Capacitance

From the derived Y-parameters we can clearly separate out the following intrinsic capacitance

$$C_{gg} = \frac{Im(Y_{gg})}{2\pi f} \tag{5.9}$$

$$C_{gd} = -\frac{Im(Y_{gd})}{2\pi f} \tag{5.10}$$

Process technology parameters such as oxide thickness, flat-band voltage, channel doping, and quantum mechanical confinement effects (leading to reduced gate capacitance) are extracted from C_{gg} ($V_{ds} = 0\,\mathrm{V}$), Fig.5.9(a). Both C_{gg} ($V_{ds} = 0\,\mathrm{V}$, near $V_{gs} = 0\,\mathrm{V}$) and C_{gd} in saturation region, $V_{ds} = V_{dd}$ are used to accurately extract the parasitic overlap and fringe capacitance over bias, Fig.5.9(a). Additionally C_{gg} vs. V_{ds} data under on-state (V_{gs}= V_{dd}) is used to capture short channel capacitance parameters corresponding to velocity saturation and channel length modulation effect, Fig.5.9(b) (see Appendix B). Even though equations derived above for $Im(Y_{gg})$ and $Im(Y_{gd})$ are approximate, we observed that they were predominantly influenced by the device capacitance. High amounts of access resistance (at source and drain) can influence the results in this step and might need an iteration after drain current extraction.

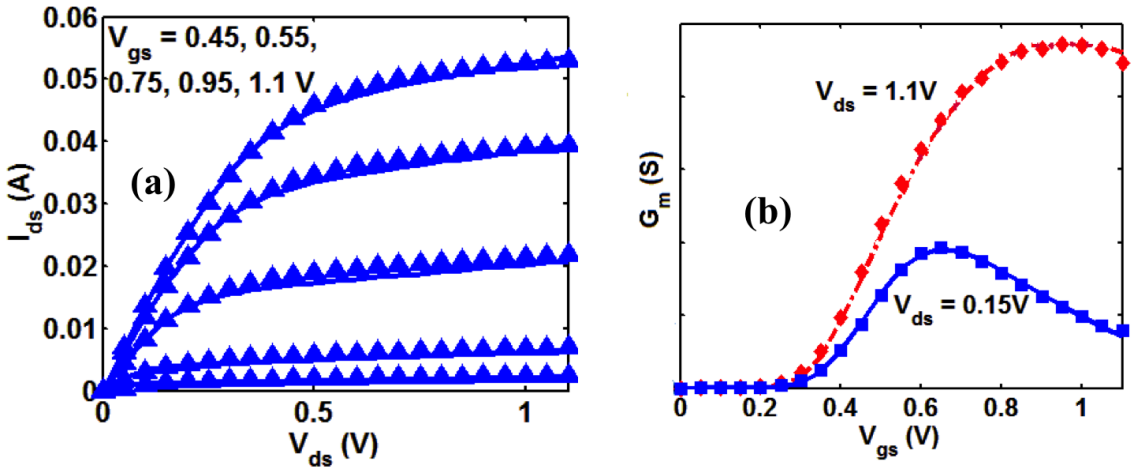


Figure 5.10: DC measurements from the same structure are used to extract parameters for mobility, series resistance, velocity saturation and output conductance. Accurate fit of not only the drain current but also the trans-conductance, output conductance (not shown here) and its derivatives are obtained in this step. Symbols=Silicon Data, Lines=BSIM6 Model.

5.2.2 Drain Current and Higher order Derivatives

The standard DC extraction procedure is applied in this step, [38]. Effects like series resistance, mobility degradation (low V_{ds}), velocity saturation (high V_{ds}) and output conductance parameters are extracted from the drain current, Fig.5.10(a). Additionally leakage currents like impact ionization current, gate-induced drain leakage (measured from body terminal) and gate tunneling current are also captured. This step ensures an accurate prediction of terminal currents and its derivatives like the trans-conductance and output conductance across gate, drain and body bias, Fig.5.10(b). Any asymmetry in the access resistance (to source and drain) will be extracted in this step. Linear region current is influenced by the sum of source and drain resistance, while the saturation region current is influenced by just the source resistance, allowing for accurate extraction with just drain current from forward mode of operation ($V_{ds} > 0\,\mathrm{V}$). The lead and external layout resistance (to be estimated from de-embedding structures) would also affect the measurements and thus needs to be included as a sub-circuit around the device for correct extraction. The extraction is performed not just on the DC currents but also giving equal weights to its higher order derivatives (the first and the second order derivative of the drain current). The importance of the derivatives for RF linearity prediction has been emphasized earlier. The accuracy in trans-conductance (G_m) and output conductance (G_ds) of the resultant model directly influence the low frequency accuracy of the small signal $Re(Y_{gd})$ and $Re(Y_{dd})$.

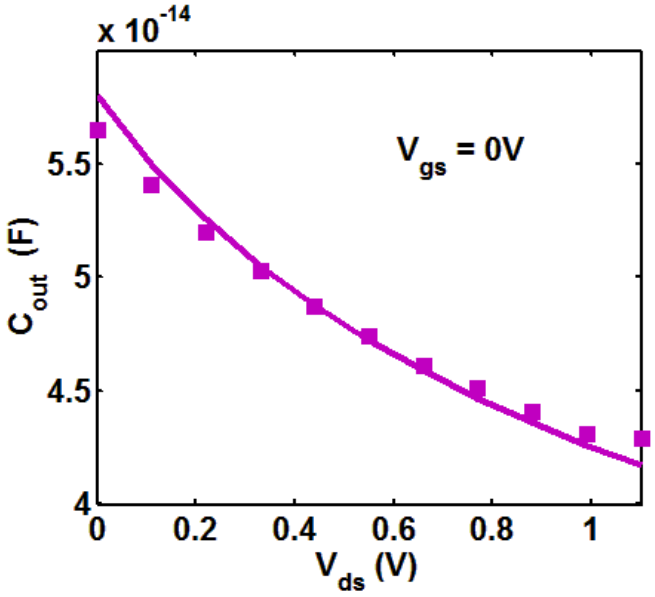


Figure 5.11: Junction capacitance parameters are extracted from $C_{jd} = C_{out}$ vs. V_{ds} curve at off state condition of the device ($V_{gs} = 0\,\mathrm{V}$). Symbols=Silicon Data, Lines=BSIM6 Model.

5.2.3 Junction Current and Capacitance

In this step the source-drain junction diode behavior is captured. Junction current measurements are either obtained from specialized diode structures or from the RF device and the relevant parameters are extracted. The junction capacitance can be obtained from the Y-parameters measured as follows,

$$C_{jd} = \frac{Im(Y_{dd} + Y_{gd})}{2\pi f} \tag{5.11}$$

C_{jd} is extracted from the off-state of the device ($V_{gs} = 0\,\mathrm{V}$), well below the threshold voltage. The MOSFET intrinsic capacitance (C_{gg}, C_{gd}, C_{dg}, C_{sd}) are very small compared to the junction capacitance in the off-state. Although we can observe from the derived Y_{dd} and Y_{gd} that C_{jd} is not exactly the junction capacitance (due to the approximations in the model) choosing off-state curves ensures that the extracted curve is predominantly influenced by the junction capacitance allowing for accurate extraction, Fig.5.11. As mentioned in the previous section, junction capacitance consists of three components (bottom, isolation side and channel side junctions) and hence more than three different devices are required for accurate estimation of the three components. The following matrix based estimation of the capacitance components could be useful for obtaining an initial guess for the respective

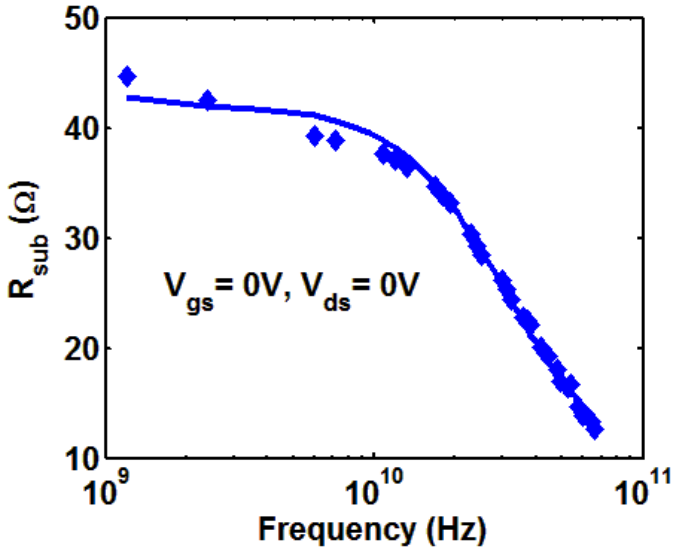


Figure 5.12: The resistance values of the substrate network are extracted from the $R_{sub,d} = R_{out}$ curve along frequency at off state condition of the device. Symbols=Hardware Data, Lines=BSIM6 Model.

capacitance values.

$$\begin{bmatrix} C_{jd1} \\ C_{jd2} \\ C_{jd3} \end{bmatrix} = \begin{bmatrix} A_1 & P_1 & W_1 \cdot NF_1 \\ A_2 & P_2 & W_2 \cdot NF_2 \\ A_3 & P_3 & W_3 \cdot NF_3 \end{bmatrix} \times \begin{bmatrix} C_J \\ C_{JSW} \\ C_{JSWG} \end{bmatrix} \tag{5.12}$$

where C_{jdi} is the extracted capacitance at zero bias ($V_{jn} = 0\,\mathrm{V}$), A_i is the junction area, P_i is the junction perimeter, W_i is the device width and NF_i is the number of fingers. Subscript $i = 1, 2, 3$ stands for the i^{th} device. C_J, C_{JSW} and C_{JSWG} stand for the zero bias junction model parameters for bottom , isolation side sidewall and channel/gate side sidewall capacitance. The devices should be so chosen such that the matrix is non-singular to obtain a unique solution for the capacitance components.

5.2.4 Substrate Effects

The effective substrate resistance looking into the drain, $R_{sub,d}$ (of say the five resistor network, Fig.5.6(c)) can be obtained from the measured Y-parameters as follows,

$$R_{sub,d} = \frac{Re(Y_{dd}) - G_{ds} - \omega^2 C_{gd} C_{dg} R_g}{(Im(Y_{dd}))^2} \tag{5.13}$$

Here output conductance G_{ds} can be obtained from our DC measurements / extraction that has been previously attempted on the same device. Once again similar to junction

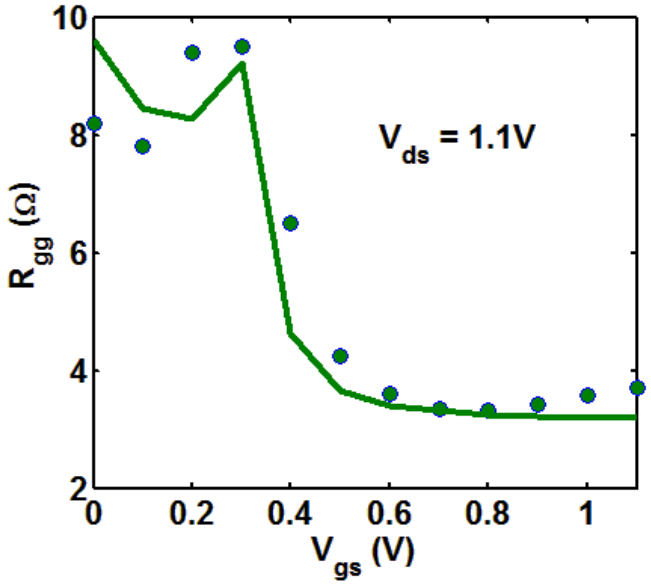


Figure 5.13: The gate electrode resistance and the channel-induced gate resistance parameters are extracted from the R_{gg} vs. V_{gs} curve. The channel-induced gate resistance component will already captured accurately if the segmentation based NQS model were used. Symbols=Hardware Data, Lines=BSIM6 Model.

capacitance, the extraction is performed from curves extracted from device in off-state ($V_{gs} = 0\,\mathrm{V}$) for all the remaining components (G_{ds}, C_{gd}, C_{dg}) are all small. The inaccuracy in yet to be extracted gate resistance $R_g g$ can be ignored as the associated capacitance components are small in off-state. Unlike the intrinsic and device capacitance, it is best to perform this extraction on the $R_{sub,d}$ along the frequency. The effect of substrate resistances show up somewhere in the $1 - 10\,\mathrm{GHz}$ region for regular bulk MOSFETs. In order to obtain a scalable model, it is best to perform simultaneous extraction of the $R_{sub,d}$ curves from multiple devices.

5.2.5 Gate Resistance and Non Quasi Static effects

Finally the effective gate resistance looking into the gate terminal is given as follows,

$$R_{gg} \approx \frac{Re(Y_{gg})}{(Im(Y_{gg}))^2} \tag{5.14}$$

This equation is approximate as the values of the MOSFET capacitance and the source, drain and substrate resistances all influence this curve as observed from the equation for Y_{gg}. However source and drain access resistance would have been estimated accurately from the drain currents extraction leaving the residual inaccuracy in R_{gg} to just the effective gate

resistance. Effective gate resistance as discussed before consists on two parts - the physical electrode resistance and the channel induced resistance component.

In a plot between extracted R_{gg} vs. the gate voltage (usually near the center frequency of the circuit application), the sub-threshold region ($V_{gs} < V_{th}$) is influenced by the effective substrate resistance as seen from the gate ($R_{sub,g}$) through the gate to body capacitance, C_{gb}, Fig.5.13. In order to accurately capture this region, some iterations between the substrate network in the previous step and gate resistance extraction might be required. The strong inversion region (on-state, $V_{gs} > V_{th}$) is influenced by the physical gate electrode resistance. The moderate inversion region is usually influenced by the channel induced gate resistance due to non-quasi static nature of the channel, Eqn.(5.4). The channel induced resistance component is dominant only at higher frequencies (approaching F_t of the device). Thus extraction at this step is best done by looking at R_{gg} as a function of both gate voltage and frequency. If the application of interest is well below F_t (both the center frequency and its harmonics of interest), the channel induced gate resistance or relaxation time approach is recommended. In this case the parameters X_1 and X_2 in Eqn.5.4 (or its equivalent) need to be extracted along with the physical gate electrode parameters. If the application is such that the center frequency or its harmonics of interest fall close to or beyond F_t, the channel segmentation model is recommended. For this model only the number of segments, N_{SEG} needs to be chosen. Accuracy of the model improves in a quadratic manner with the number of segments, i.e. in terms of the R_{gg},

$$R_{gg}(N_{SEG}) = R_{gg,NQS}\left(1 - \frac{1}{N_{SEG}^2}\right) \tag{5.15}$$

where $R_{gg,NQS}$ is the actual input impedance of the device, [70].

5.2.6 Miscellaneous Parasitic Elements

While we have extracted most of the device components, there might be some parasitics that are unaccounted. Most of these arise due to the selection of the plane of de-embedding. The most accurate test structure would be the one which has the de-embedding plane near the first metal layer. However as this is not always possible there will be remnant capacitance that need to be accounted. For example parasitic source to drain capacitance can be extracted from the following curve,

$$C_{sd} = \frac{Im(Y_{dd})}{2\pi f} - C_{gd} - C_{jd} \tag{5.16}$$

The so derived C_{sd} already contains the intrinsic device trans-capacitance between the source and drain. Thus any mismatch between the model and data at this stage is due to the layout capacitance. Depending on the frequency of operation and the length of the layout traces used the series inductance could also affect the small signal measurements at high frequency and needs to be accounted for in the model. Taking early cues from parasitic extraction tools on the final layout for these parasitics will be helpful. An initial guess closer to the final helps

accurate estimation of the rest of the model parameters. Else another iteration for accuracy would be necessary at this step. At times when a consistent value for all the components have not been found (seen as a bad fit in some Y-parameters but not the others), one could run the extraction routines on the full-fledged set of Y-parameters (both real and imaginary) involving key parasitic capacitance and resistances (diode and fringe capacitance, gate and substrate resistance etc.).

At this stage, one should have a model that has captured both the DC data across various bias conditions and small-signal S/Y-parameter data up to the frequency of interest. We will now continue to validate this approach developed in this section.

5.3 Validation

The above procedure has been validated using silicon data from three different technology nodes across different channel lengths (short, medium, long). Depending on the RF application, the designer might perform parameter extraction over a wide range of biases (global) or could just concentrate on a small range of biases (local). We will demonstrate both the strategies - a local fit for a long channel device and a global fit for a short channel device. In both the cases discussed, extraction was performed on DC curves and a small but comprehensive number of sub-sets of small-signal Y-parameters as discussed in the previous section. The plots shown in this section are a result of that procedure.

A 2 µm long channel device from a 90 nm process was used for validation purposes. The extracted unity current gain cut-off frequency,F_t and unity power gain cut-off frequency, F_{max} of the device at the bias point of interest were 850 MHz and 5 GHz respectively. For this case the target frequency range for fitting the data of interest far exceeded these values. For this purpose the charge segmentation based NQS model was switched on ahead of the extraction. The extraction procedure from the previous section was followed. We demonstrate that the BSIM6 RF model captures the small-signal data well for frequencies even greater than $10\times F_t$, Figs.5.14-5.18. Accurate values for gate resistance together with the NQS effect model helps capture the $Re(Y_{gg})$ curve, Fig.5.14. The intrinsic and extrinsic capacitance of the BSIM6 model affect the imaginary component of the small-signal Y-parameter, Figs.5.15,5.16. While the data shows that the trans-conductance, $G_m = |Y_{dg} - Y_{gd}|$ of the device drops at high frequencies, a SPICE model without NQS effects (i.e. a quasi-static model), predicts an increasing G_m, Fig.5.17. The channel induced gate resistance based first order NQS model captures the data up to F_t and is suitable for either shorter channel length devices (with high F_t) or for applications with lower frequencies of interest. Masons unilateral power gain, GU is considered an good metric for a RF device as it is invariant under lossless embedding around the device (used for matching or feedback). It thus serves as an important figure of merit for RF devices across technology and is also well captured by the model, Fig.5.18.

De-embedded silicon data for short to medium channel length devices from a 45 nm bulk technology was also used to validate the above procedure. For this application it was required to capture the data both across bias as well as across frequency (atleast up to F_t).

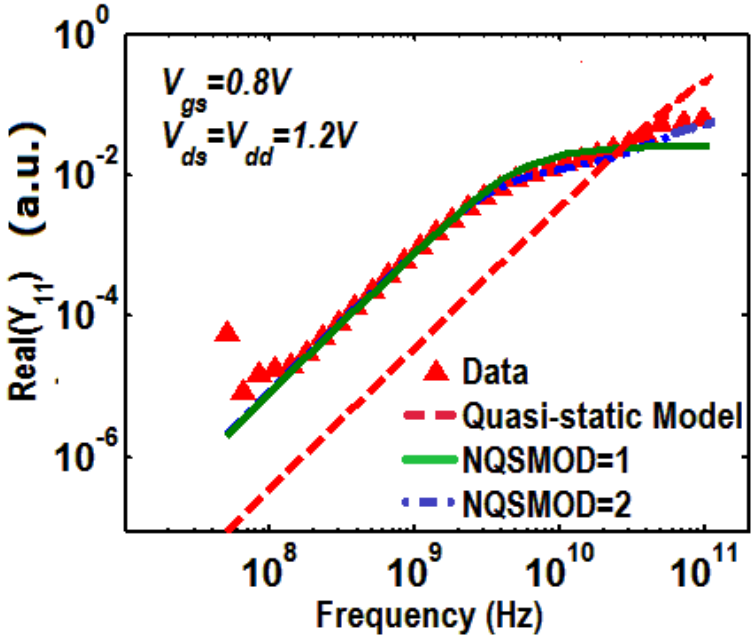


Figure 5.14: The NQS models in BSIM6 help capture the high frequency behavior of the input admittance ($Re(Y_{gg}) = Re(Y_{11})$) while the quasi-static model fails to capture the same. Symbols=Silicon Data, Lines=BSIM6 Model

The accuracy in extracted model's trans-conductance and capacitance helps predict the bias dependence of the cut-off frequency,F_t of the device, Fig.5.19. Here F_t was extracted as the 0 dB interception point of the extrapolated current gain, H_{dg} curve where H referes to 2-port hybrid parameters. The bias dependence as well as the frequency dependence of the gate resistance dominated $Re(Y_{gg})$ is modeled well, Fig.5.20. $Re(Y_{dd})$ is dominated by the substrate network in the device off state condition, and by the output conductance (drain induced barrier lowering, channel length modulation effects) of the intrinsic device in the on state / inversion condition, Fig.5.21. The trans-conductance, G_m is shown to be captured well by the model well even for frequencies approaching F_t of the device, Fig.5.22. The bias dependent gate resistance based NQS model used here was adequate enough for the frequency range of interest (highest frequency required was around F_t of the device). Given an overall excellent fit achieved using the BSIM6 model across bias and frequencies, the maximum power gain, G_{max} is well predicted by the model and matches the G_{max} extracted from the silicon data well.

In another exercise, BSIM6 was used to model short channel devices from a 65 nm CMOS bulk MOSFET technology using the outlined procedure. The same devices were incorporated in a low-noise amplifier operating at 12 GHz. The large signal response measurements from the amplifier matched the BSIM6 based simulations very well, Fig.5.24. This allows accurate prediction of RF linearity merits like P1dB (the output power at which the gain drops from its small-signal value by 1 dB) from the gain compression plot, Fig.5.25.

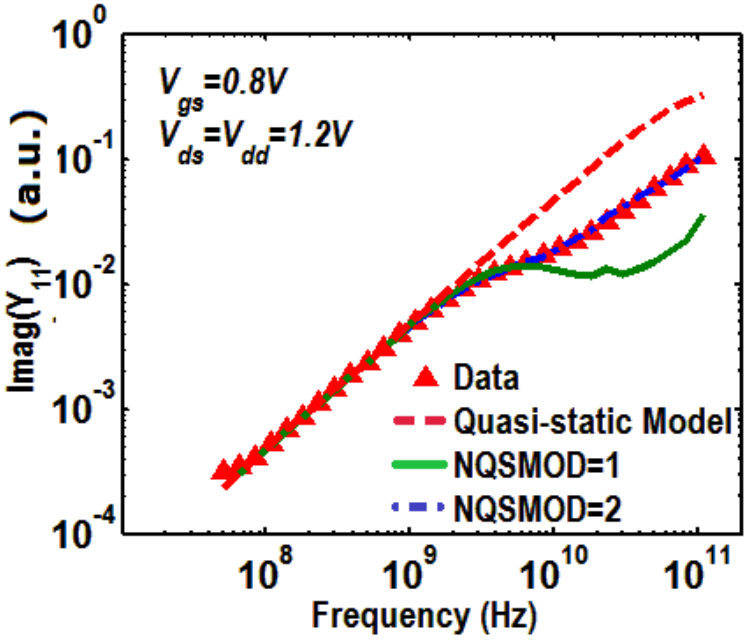


Figure 5.15: The intrinsic and extrinsic capacitance model determine the behavior of the $Im(Y_{gg}) = Im(Y_{11})$ curve. Symbols=Silicon Data, Lines=BSIM6 Model

5.4 Summary

A study was undertaken to examine the various components of a MOSFET SPICE model (such as in BSIM6) with regards to their impact on high frequency behavior of the MOSFET. We then developed a physical CMOS RF model creation procedure in a series of six steps. The developed procedure is small-signal model agnostic i.e. it doesn't rely on a particular schematic visualization of the MOSFET small-signal behavior. Thus this procedure can be used with any MOSFET compact model that have all the necessary RF components. A divide and conquer approach to hardware data was used that enables an efficient use of available parameter extraction automation tools. The extraction discussed takes advantage of all the components present in BSIM6 and required no additional sub-circuit elements. The developed procedure was validated on silicon data from multiple hardware bulk MOSFET technologies showing excellent fit to key RF device figure of merits. A sample representative Agilent ICCAP based file has been created to demonstrate the extraction procedure.

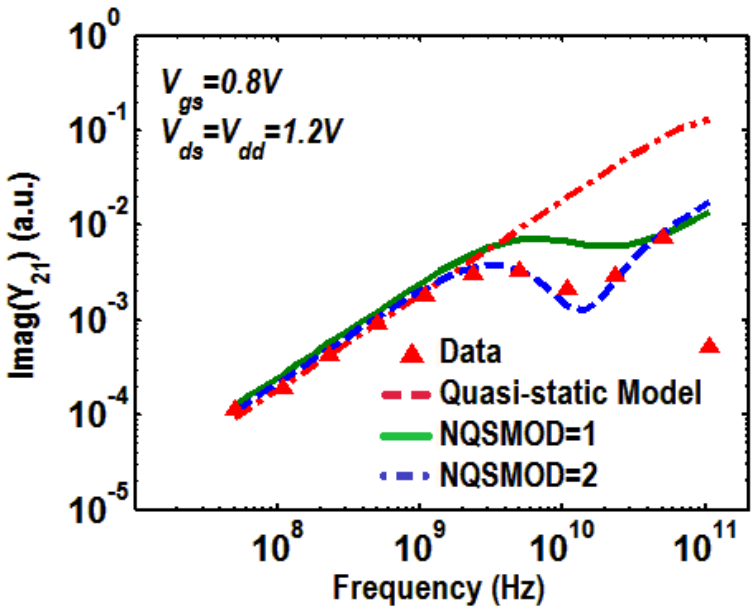


Figure 5.16: The model captures the behavior of $Im(Y_{dg}) = Im(Y_{21})$ up to 100 GHz accurately. Symbols=Silicon Data, Lines=BSIM6 Model

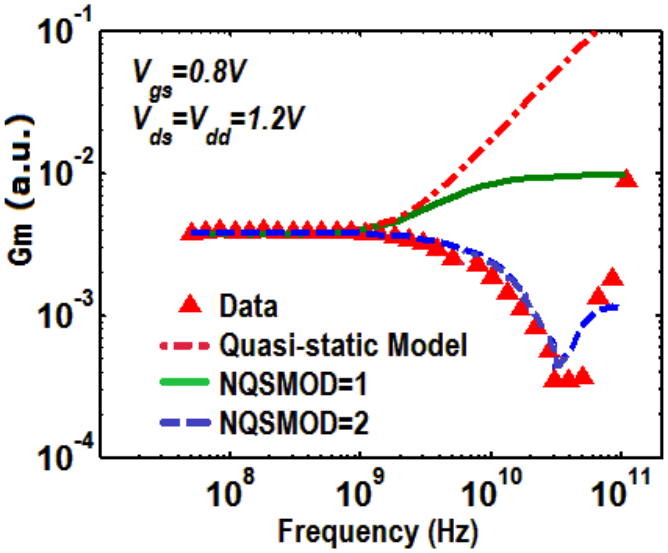


Figure 5.17: The channel segementation model predicts the trans-conductance, G_m roll off at high frequencies while the quasi-static model is in error while predicting the behavior of the device for frequencies approaching F_t. Symbols=Silicon Data, Lines=BSIM6 Model

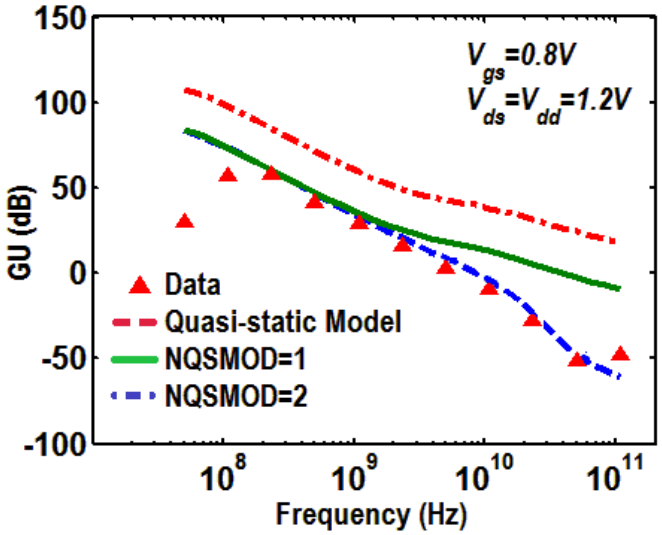


Figure 5.18: The resultant model predicts the Mason's Gain, GU of the transistor (that is invariant under addition of lossless components around the FET) accurately up to 100 GHz. Symbols=Silicon Data, Lines=BSIM6 Model

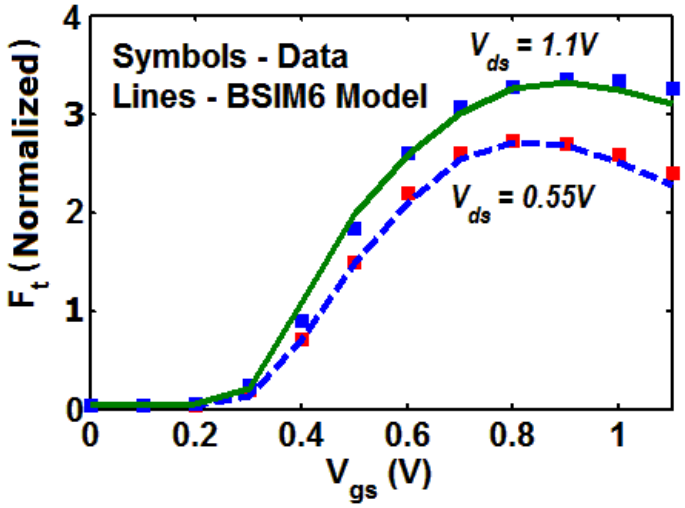


Figure 5.19: The bias dependence of the cut-off frequency, F_t relies on the model accuracy in predicting the trans-conductance, output conductance and device capacitance. Symbols=Silicon Data, Lines=BSIM6 Model

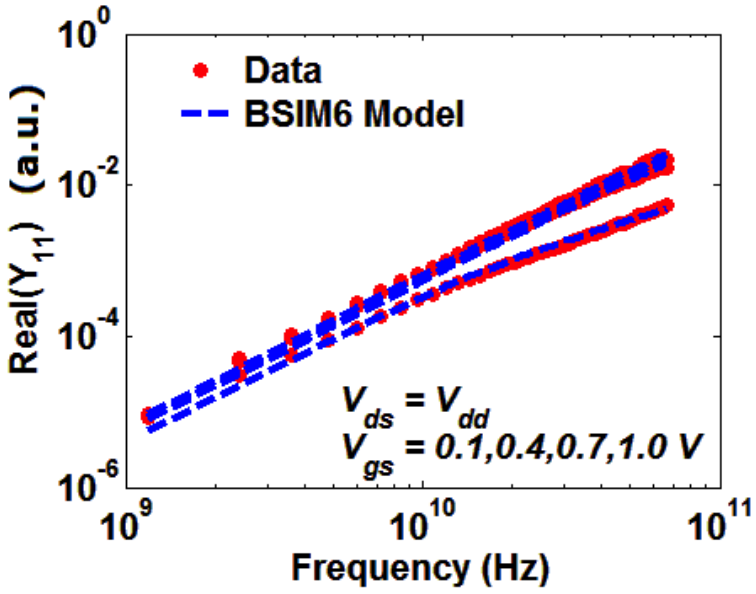


Figure 5.20: The gate electrode resistance and the channel induced gate resistance model help capture the bias and frequency dependence of the input admittance, $Re(Y_{gg}) = Re(Y_{11})$. Symbols=Silicon Data, Lines=BSIM6 Model

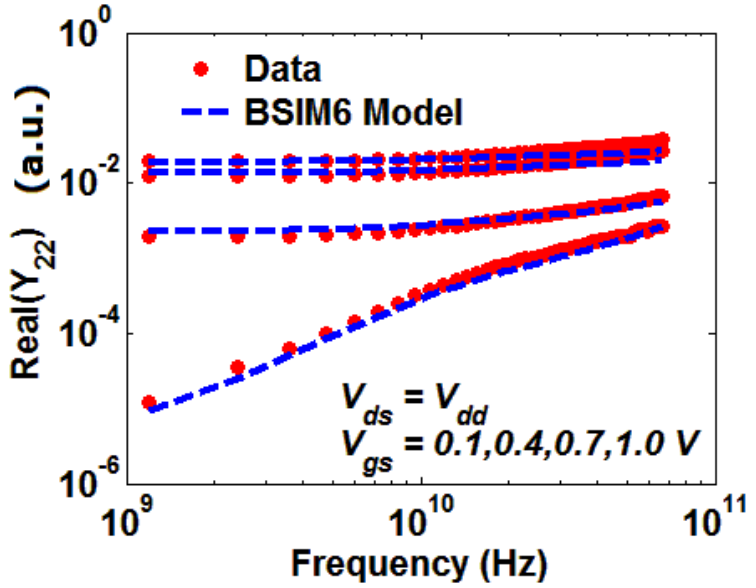


Figure 5.21: Output conductance model (for low frequency) and substrate network model (for high frequency) have been extracted accurately to predict the hardware output conductance ($Re(Y_{dd}) = Re(Y_{22})$) data. Symbols=Silicon Data, Lines=BSIM6 Model

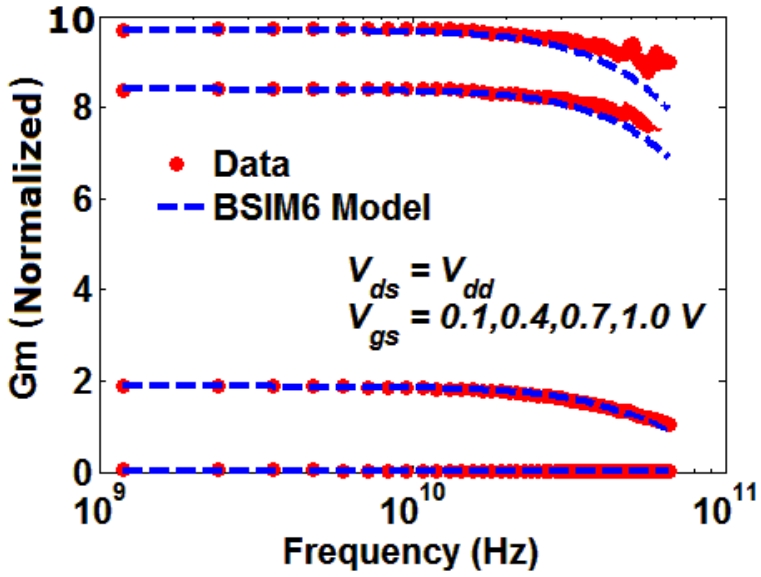


Figure 5.22: The resultant model with the gate induced resistance non-quasi static model predicts the low-frequency trans-conductance, G_m across bias and as well as the roll off when approaching frequencies close to the cut-off frequency, F_t of the device accurately. Symbols=Hardware Data, Lines=BSIM6 Model

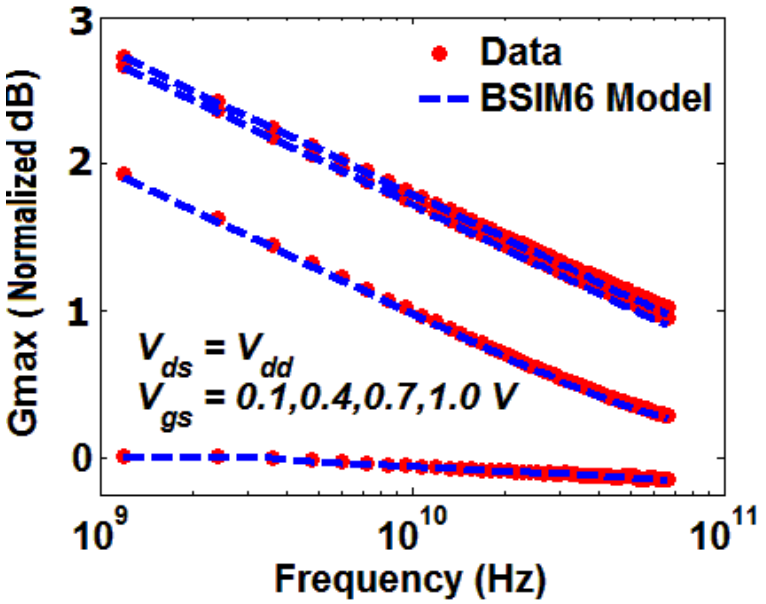


Figure 5.23: The model predicts the maximum gain, G_{max} of the device across bias up to F_t of the device very well. Symbols=Hardware Data, Lines=BSIM6 Model

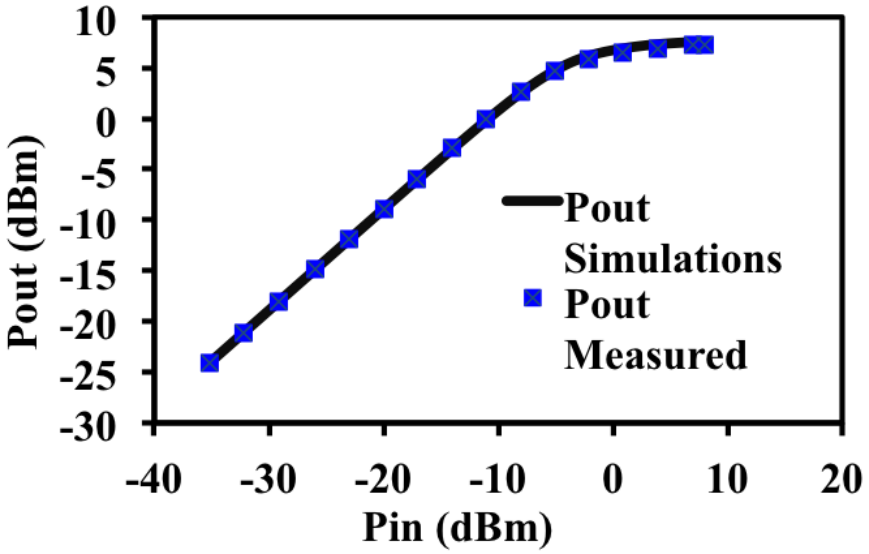


Figure 5.24: Large signal response of the low-noise amplifier at 12 GHz. Symbols=Hardware Data, Lines=BSIM6 Model

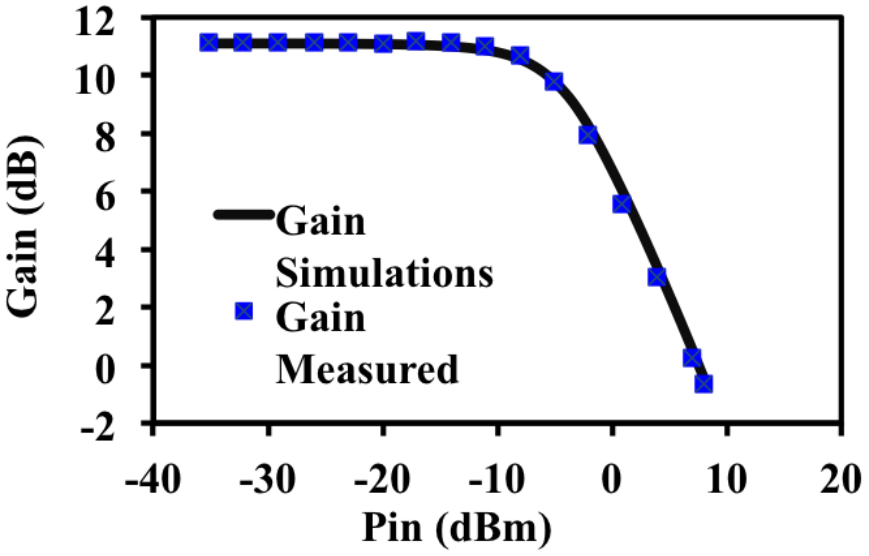


Figure 5.25: Gain (Pout-Pin) response of the low-noise amplifier at 12 GHz. Symbols=Hardware Data, Lines=BSIM6 Model

Chapter 6

Conclusions

Bulk-planar transistor has run out of steam unable to provide the requisite performance benefits with scaling. This has led to the introduction of a multitude of 3-dimensional (3-D) multi-gate transistor device architectures. FinFET (Intel TriGate) transistor device has already been brought into production as of 2012 at the 22 nm technology node with more foundries planning to bring them out at the 14/16 nm node. At the same time, vertical double gate or gate-all-around pillars with multi-layer transistors (24 and more) are expected to replace planar transistors at sub-$1X$ nm for NAND flash memory technology. Meanwhile vertical cylindrical gated transistors (CG-FET) are seen as a way to obtain $< 4F^2$ footprint per bit cell in a DRAM memory technology as opposed to being restricted to $6F^2$ per bit cell area with the use of a planar transistor. In order to enable all these technologies, accurate physics based compact SPICE models were required.

We developed a comprehensive physics based core model for a generic cylindrical gate MOSFET (CG-FET) for an ideal long channel device. The developed core model inherently supports FETs with intrinsic to high channel doping capturing (but assumes fully depleted channel) their effects in threshold voltage as well as moderate inversion region drain current. The core model also supports polysilicon gate depletion accurately predicting gate capacitance degradation at high gate voltages. This drain current and the terminal capacitance resulting from the core model were validated using numerical device simulations (TCAD) for CG-FETs across a range of channel diameter and doping values. Further, scaling length based short-channel model and various real device effect models were borrowed from the BSIM-CMG double gate model framework and added to augment the developed core model, [39].

A rule of thumb for keeping short channel effects under control for a cylindrical gate device is $L \approx 2R$, i.e. the channel length is roughly about the diameter of the channel. Thus, scaling the channel length for performance also implies scaling down the channel diameter. For diameters approaching < 20 nm, quantum mechanical confinement effects arise. We have developed a simple model that tracks the geometry and bias dependence of the channel mobile charge centroid. Using TCAD we validate this model by demonstrating that this model captures the gate capacitance degradation for moderate gate bias due to

increase in effective oxide thickness (gate oxide physical thickness plus the charge centroid from the channel interface), [71]. In addition the vertical CG-FET architecture proposed for DRAM memory bit cell shows asymmetric behavior depending on whether the top is used as the drain or the source of the device. In our study we found that vertical channel implantation creates a doping gradient from the top to bottom resulting in different doping in the channel at the top and bottom channel edges. Additionally structural differences in the top and bottom junctions render different resistance and bias dependence to each of them. These result in stark asymmetry in drain-induced barrier lowering based threshold voltage shift and the on-current for the top as the source or the top as the drain configurations. We have developed a mathematical framework and identified key physical parameters in the CG-FET compact model developed so far to capture the asymmetry.

All the above discussed, the CG-FET core model together with related real device effect improvements were implemented with the BSIM-CMG symmetric multi-gate FET model in Verilog-A. The so developed model was validated with two different hardware data sets. In Chapter 2 the cylindrical gate transistor model was first validated with a low voltage ($V_{dd} = 1.5\,\mathrm{V}$) vertical CG-FET technology showing excellent agreement to not only the bias dependence of the drain current (both the top contact as drain and bottom contact as drain) but also the derivatives of the drain current, [72]. The model was able the predict the temperature dependence of the drain current data as well. Asymmetry is more prominent at higher drain voltages. In Chapter 4, we validate the asymmetry model developed with silicon data from a high voltage ($V_{dd} = 2.2\,\mathrm{V}$) vertical CG-FET transistor, [73]. Our model enables the creation of one set of consistent model parameters to capture both the top contact as drain and bottom contact as drain configurations as against two separate models for the two modes. The transition from one mode to another happens in an infinitely smooth fashion around $V_{ds} = 0\,\mathrm{V}$, i.e. both the source-drain terminal currents and their higher order derivatives are continuous. The developed model satisfies the standard compact model quality tests such as Harmonic Balance, Gummel Symmetry and AC Symmetry, [40, 74]. One should note here that if the asymmetry model is also used then it suffices to ensure Gummel and AC continuity alone and not their symmetry around $V_{ds} = 0\,\mathrm{V}$.

In Chapter 3 the quantum mechanical charge centroid model was also extended for double-gate / FinFETs and validated with TCAD in a similar way as the CG-FETs. For this case the charge centroid is a function of the fin thickness and the channel doping in the fin region. For bulk FinFETs a ground plane implant below the fin region is used to prevent source-drain depletion region punch-through. In such devices we observed a unique double junction arising in the source/drain region due to abrupt transition in the doping from the ground-plane region to well region (for ex: $p^{+}|n_{punch-through}|n_{well}$ in a p-type FinFET). In Chapter 3 we developed a junction capacitance model for these kind of junctions where the depletion region could traverse through both the ground-plane and well doping regions and validated the developed model using TCAD, [75].

In Chapter 5 a RF CMOS model development procedure was proposed. First a critical study of all the RF relevant model components in a compact SPICE model is presented discussing their validity and impact on high frequency data. We observed that a model like

BSIM6 contains all the requisite components required to capture both DC and high frequency data. A short channel terminal charge model considering velocity saturation and channel length modulation effects was developed to augment the core model in BSIM6 (Appendix B). A charge segmentation based non-quasi static (NQS) model was derived for BSIM6 and short review of other NQS models was presented (Appendix D). Then a six step parameter extraction procedure was created that is formulated in a way conducive to current parameter extraction and optimization tools available. Then the developed procedure was validated on two different sets of bulk planar MOSFET hardware data, [76]. The developed procedure was performed on both a long and short channel device. While the intrinsic MOSFET behavior (trans-conductance, output conductance capacitance) dominates the low frequency behavior the sub-model components (junction and fringe capacitance, substrate network and gate resistance) help capture the high frequency ($> 1\,\text{GHz}$) data. Using the long channel device (which typically have lower cut-off frequency) we demonstrate the importance of the non-quasi static model to capture the MOSFET S-parameter data beyond its cut-off frequency. The developed procedure relies on no external sub-circuit elements and can be used with any general MOSFET compact model.

6.1 Recommended Future Work

Advances in transistor Compact Models and progress in transistor technology keep feeding into each other. Both near term and long term progress is required in the area of transistor Compact Models for SPICE.

Ideal FinFETs are best used when the channel is fully-depleted for near 60 mV per decade sub-threshold swing (and low off-current). This would also imply that the FinFETs are immune to body-bias. However FinFETs built on bulk substrate do have some amount of body bias based threshold voltage modulation tending to behave more like a bulk-planar MOSFETs. Current core models for double gate FinFETs discussed in Chapter 2 do not include this effect and needs to be captured.

When the current FinFET based silicon CMOS technology runs out of steam, product diversification can be obtained by including hybrid devices like an asymmetric double-gate transistor. As mentioned before in Chapter 1 these devices could be used in some unique ways such as to create single device RF mixers. A computation efficient model for asymmetric double-gate transistor supporting dual side inversion operation has not been achieved yet. While BSIM-IMG focuses on single side inversion treating the other gate as a threshold voltage knob, other known solutions ([22]) that try to address dual-side inversion tend to be computationally intensive. More research in this direction could enable such hybrid silicon based technologies.

As we continue to scale down the transistors for performance, advances in transistors made with newer materials will replace silicon based technology. Research has brought forward many choices such as carbon-nanotube FETs, graphene nano-ribbon FETs and MoS_2/WSe_2 based thin-film FETs. For sub-10 nm dimensions advances in atomistic quan-

tum numerical device simulators have revealed multitude of new effects such as source starvation and quantum capacitance etc. Classical understanding of effects like gate-tunneling stand to question, [77]. While material/ device specific compact models have been developed ([78]) a universal device compact model could aid in quick evaluation of the new technology under various circuit design scenarios. The universal device compact model would contain a generic transport model (an approximate representation of various transport mechanisms ballistic / quasi-ballistic/ drif-diffusion) and various current saturation mechanisms. Research in this direction is very much needed in the longer term.

Bibliography

[1] (2012). Process Integration, Devices, and Structures ITRS update 2012., [Online]. Available: `http://public.itrs.net`.

[2] K. Mistry, C. Allen, C. Auth, B. Beattie, et al., "A 45nm logic technology with high-k metal gate transistors, strained silicon, 9 cu interconnect layers, 193nm dry patterning, and 100% Pb-free packaging," in *IEEE International Electron Devices Meeting Technical Digest.*, 2007, pp. 247–250.

[3] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*. Wiley, 2006.

[4] T. Ghani, M. Armstrong, C. Auth, M. Bost, et al., "A 90nm high volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors," in *IEEE International Electron Devices Meeting Technical Digest.*, 2003, pp. 11.6.1–11.6.3.

[5] C. Wann, K. Noda, T. Tanaka, M. Yoshida, and C. Hu, "A comparative study of advanced MOSFET concepts," *IEEE Transactions on Electron Devices*, vol. 43, no. 10, pp. 1742–1753, 1996.

[6] D. Hisamoto, W.-C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T.-J. King, J. Bokor, and C. Hu, "FinFET-a self-aligned double-gate MOSFET scalable to 20 nm," *IEEE Transactions on Electron Devices*, vol. 47, no. 12, pp. 2320–2325, 2000.

[7] C. Auth, C. Allen, A. Blattner, D. Bergstrom, et al., "A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," in *Symposium on VLSI Technology (VLSIT)*, 2012, pp. 131–132.

[8] C. H Lin, R. Kambhampati, R. Miller, T. B. Hook, et al., "Channel doping impact on FinFETs for 22nm and beyond," in *Symposium on VLSI Technology (VLSIT)*, 2012, pp. 15–16.

[9] S. Y. Cha, *DRAM and Future of Commodity Memories*, Technology Short Course, Symposium on VLSI Technology (VLSIT), 2011.

[10] Y.-K. Choi, K. Asano, N. Lindert, V. Subramanian, T.-J. King, J. Bokor, and C. Hu, "Ultra-thin body SOI MOSFET for deep-sub-tenth micron era," in *IEEE International Electron Devices Meeting Technical Digest*, 1999, pp. 919–921.

[11] N. Planes, O. Weber, V. Barral, S. Haendler, et al., "28nm FDSOI technology platform for high-speed low-voltage digital applications," in *Symposium on VLSI Technology (VLSIT)*, 2012, pp. 133–134.

[12] A. Vladimirescu and S. Liu. (1980). The simulation of MOS integrated circuits using SPICE2, [Online]. Available: `http://www.eecs.berkeley.edu/Pubs/TechRpts/1980/9610.html`.

[13] B. Sheu, D. Scharfetter, C. Hu, and D. O. Pederson, "A compact IGFET charge model," *IEEE Transactions on Circuits and Systems*, vol. 31, no. 8, pp. 745–748, 1984.

[14] (2013). Berkeley SPICE IGFET Model, BSIM, [Online]. Available: `http://www-device.eecs.berkeley.edu/bsim/?page=about`.

[15] (2013). Compact Model Council, [Online]. Available: `https://www.si2.org/?page=1684`.

[16] M. V. Dunga, "Nanoscale CMOS Modeling," PhD thesis, EECS Department, University of California, Berkeley, Mar. 2008. [Online]. Available: `http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-20.html`.

[17] D. Lu, "Compact Models for Future Generation CMOS," PhD thesis, EECS Department, University of California, Berkeley, May 2011. [Online]. Available: `http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-69.pdf`.

[18] (2013). BSIM6.0.0 Technical Manual, [Online]. Available: `http://www-device.eecs.berkeley.edu/bsim/?page=BSIM6`.

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

[20] C. C. Enz and E. A. Vittoz, *Charge-Based MOS Transistor Modeling: The EKV Model for Low-Power and RF IC Design*. Wiley, 2006.

[21] S. Khandelwal, Y. Chauhan, D. Lu, S. Venugopalan, et al., "BSIM-IMG: A Compact Model for Ultrathin-Body SOI MOSFETs with Back-Gate Control," *IEEE Transactions on Electron Devices*, vol. 59, no. 8, pp. 2019–2026, 2012.

[22] S. Jandhyala and S. Mahapatra, "An efficient robust algorithm for the surface-potential calculation of independent DG MOSFET," *IEEE Transactions on Electron Devices*, vol. 58, no. 6, pp. 1663–1671, 2011.

[23] G. Dessai and G. Gildenblat, "Solution space for the independent-gate asymmetric DGFET," *Solid-State Electronics*, vol. 54, no. 4, pp. 382 –384, 2010.

[24] Y. Taur, X. Liang, W. Wang, and H. Lu, "A continuous, analytic drain-current model for DG MOSFETs," *IEEE Electron Device Letters*, vol. 25, no. 2, pp. 107–109, 2004.

[25] J.-M. Sallese, F. Krummenacher, F. Prgaldiny, C. Lallement, A. Roy, and C. Enz, "A design oriented charge-based current model for symmetric DG MOSFET and its correlation with the EKV formalism," *Solid-State Electronics*, vol. 49, no. 3, pp. 485 –489, 2005.

[26] F. Liu, J. He, J. Zhang, Y. Chen, and M. Chan, "A Non-Charge-Sheet Analytic Model for Symmetric Double-Gate MOSFETs with Smooth Transition between Partially and Fully Depleted operation modes," *IEEE Transactions on Electron Devices*, vol. 55, no. 12, pp. 3494–3502, 2008.

[27] D. Jimenez, B. Iniguez, J. Sune, L. Marsal, J. Pallares, J. Roig, and D. Flores, "Continuous analytic I-V model for Surrounding-Gate MOSFETs," *IEEE Electron Device Letters*, vol. 25, no. 8, pp. 571–573, Aug. 2004.

[28] F. Liu, J. He, L. Zhang, J. Zhang, J. Hu, C. Ma, and M. Chan, "A Charge-based model for long-channel Cylindrical Surrounding-Gate MOSFETs from intrinsic channel to heavily doped body," *IEEE Transactions on Electron Devices*, vol. 55, no. 8, pp. 2187–2194, Aug. 2008.

[29] G. Dessai, A. Dey, G. Gildenblat, and G. D. Smit, "Symmetric linearization method for Double-Gate and Surrounding-Gate MOSFET models," *Solid-State Electronics*, vol. 53, no. 5, pp. 548 –556, 2009.

[30] J. Duarte, S.-J. Choi, D.-I. Moon, J.-H. Ahn, J.-Y. Kim, S. Kim, and Y.-K. Choi, "A Universal Core Model for Multiple-Gate Field-Effect Transistors. Part I: Charge Model," *IEEE Transactions on Electron Devices*, vol. 60, no. 2, pp. 840–847, 2013.

[31] G. Dessai, "Compact Modeling of Multi-Gate Transistors," PhD thesis, Electrical Engineering, Arizona State University, Dec. 2012. [Online]. Available: `http://hdl.handle.net/2286/R.I.15862`.

[32] A. S. Householder, *The Numerical Treatment of a Single Nonlinear Equation*. McGraw-Hill, New York, 1970.

[33] B. Yu, H. Lu, M. Liu, and Y. Taur, "Explicit continuous models for Double-Gate and Surrounding-Gate MOSFETs," *IEEE Transactions on Electron Devices*, vol. 54, no. 10, pp. 2715–2722, Oct. 2007.

[34] (2013). Sentaurus Device, *Synopsys Inc.*, [Online]. Available: `http://www.synopsys.com/Tools/TCAD/DeviceSimulation/Pages/SentaurusDevice.aspx`.

[35] S.-Y. Oh, D. Ward, and R. Dutton, "Transient analysis of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 15, no. 4, pp. 636–643, Aug. 1980.

[36] M. Dunga, C.-H. Lin, D. Lu, W. Xiong, C. Cleavelin, P. Patruno, J.-R. Hwang, F.-L. Yang, A. Niknejad, and C. Hu, "BSIM-MG: a Versatile Multi-Gate FET model for mixed-signal design," in *Proceedings of IEEE Symposium on VLSI Technology*, Jun. 2007, pp. 60–61.

[37] C. Auth and J. D. Plummer, "Scaling theory for cylindrical, fully-depleted, surrounding-gate MOSFET's," *IEEE Electron Device Letters*, vol. 18, no. 2, pp. 74–76, 1997.

[38] (2013). BSIM4.7.0 Technical Manual, [Online]. Available: `http://www-device.eecs.berkeley.edu/bsim/?page=BSIM4`.

[39] (2013). BSIM-CMG106.1.0 Technical Manual, [Online]. Available: `http://www-device.eecs.berkeley.edu/bsim/?page=BSIMCMG`.

[40] C. McAndrew, "Validation of MOSFET model Source - Drain Symmetry," *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 2202–2206, 2006.

[41] V. P. Trivedi and J. G. Fossum, "Quantum-Mechanical Effects on the Threshold Voltage of undoped Double-Gate MOSFETs," *IEEE Transactions on Electron Devices*, vol. 26, no. 8, pp. 579–582, Aug. 2005.

[42] L. Ge and J. G. Fossum, "Analytical modeling of Quantization and Volume Inversion in thin Si-film DG MOSFETs," *IEEE Transactions on Electron Devices*, vol. 49, no. 2, pp. 287–294, Feb. 2002.

[43] Y.-S. Wu and P. Su, "Analytical Quantum-Confinement model for Short-Channel Gate-All-Around MOSFETs under Subthreshold region," *IEEE Transactions on Electron Devices*, vol. 56, no. 11, pp. 2720–2725, Nov. 2009.

[44] J. Roldan, A. Godoy, F. Gamiz, and M. Balaguer, "Modeling the Centroid and the Inversion Charge in Cylindrical Surrounding Gate MOSFETs, including Quantum Effects," *IEEE Transactions on Electron Devices*, vol. 55, no. 1, pp. 411–416, Jan. 2008.

[45] Y. Yuan, B. Yu, J. Song, and Y. Taur, "An analytic model for Threshold Voltage shift due to Quantum confinement in Surrounding Gate MOSFETs with anisotropic effective mass," *Solid-State Electronics*, vol. 53, no. 2, pp. 140 –144, 2009.

[46] C.-H. Lin, M. Dunga, A. Niknejad, and C. Hu, "A compact Quantum-Mechanical model for Double-Gate MOSFET," in *8th International Conference on Solid-State and Integrated Circuit Technology, (ICSICT)*, Oct. 2006, pp. 1272–1274.

[47] K. Okano, T. Izumida, H. Kawasaki, A. Kaneko, et al., "Process integration technology and Device characteristics of CMOS FinFET on bulk Silicon substrate with sub-10 nm Fin width and 20 nm gate length," in *International Electron Devices Meeting Technical Digest, (IEDM)*, 2005, pp. 721–724.

[48] V.-H. Hu, M.-L. Fan, P. Su, and C.-T. Chuang, "Evaluation of static noise margin and performance of 6T FinFET SRAM cells with asymmetric gate to source/drain underlap devices," in *IEEE International SOI Conference (SOI)*, 2010, pp. 1–2.

[49] S. Whang, K. Lee, D. Shin, B. Kim, et al., "Novel 3-Dimensional Dual Control-gate with Surrounding Floating-gate (DC-SF) NAND flash cell for 1Tb file storage application," in *IEEE International Electron Devices Meeting Technical Digest (IEDM)*, 2010, pp. 29.7.1–29.7.4.

[50] X. Chen, Q. Ouyang, G. Wang, and S. K. Banerjee, "Improved hot-carrier and short-channel performance in vertical nMOSFETs with graded channel doping," *IEEE Transactions on Electron Devices*, vol. 49, no. 11, pp. 1962–1968, 2002.

[51] Y. Singh Chauhan, F. Krummenacher, R. Gillon, B. Bakeroot, M. Declercq, and A.-M. Ionescu, "Compact Modeling of Lateral Nonuniform Doping in High-Voltage MOSFETs," *IEEE Transactions on Electron Devices*, vol. 54, no. 6, pp. 1527–1539, 2007.

[52] G. H. See, X. Zhou, K. Chandrasekaran, S. B. Chiah, Z. Zhu, C. Wei, S. Lin, G. Zhu, and G. H. Lim, "A Compact Model satisfying Gummel symmetry in higher order derivatives and applicable to asymmetric MOSFETs," *IEEE Transactions on Electron Devices*, vol. 55, no. 2, pp. 624–631, 2008.

[53] A. C. T. Aarts, R. Van Der Hout, J. Paasschens, A. Scholten, M. Willemsen, and D. B. M. Klaassen, "Capacitance modeling of laterally non-uniform MOS devices," in *IEEE International Electron Devices Meeting (IEDM) Technical Digest.*, 2004, pp. 751–754.

[54] A. Roy, Y. Chauhan, J. M Sallese, C. Enz, A.-M. Ionescu, and M. Declercq, "Partitioning scheme in lateral asymmetric MOST," in *Proceeding of the 36th European Solid-State Device Research Conference (ESSDERC)*, 2006, pp. 307–310.

[55] A. Blaum, J. Victory, and C. Mcandrew, "A simple physical extraction method for $R_D - R_S$ of Asymmetric MOSFETs," in *Proceedings of International Conference on Microelectronic Test Structures (ICMTS)*, 1999, pp. 141–146.

[56] Y. Cheng, M. Deen, and C.-H. Chen, "Mosfet modeling for RF IC design," *IEEE Transactions on Electron Devices*, vol. 52, no. 7, pp. 1286–1303, 2005.

[57] I. Kwon, M. Je, K. Lee, and H. Shin, "A simple and analytical parameter-extraction method of a microwave MOSFET," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 6, pp. 1503–1509, 2002.

[58] (2013). IC-CAP Device Modeling Software, *Agilent Inc.*, [Online]. Available: `http://www.home.agilent.com/en/pc-1297149/ic-cap-device-modeling-software-measurement-control-and-parameter-extraction?&cc=US&lc=eng`.

[59] (2013). BSIMProPplus, *ProPlus Design Solutions Inc.*, [Online]. Available: `http://www.proplussolutions.com/en/pro1/Advanced-SPICE-Modeling-Platform---BSIMProPlus.html`.

[60] S. Emami, C. Doan, A. Niknejad, and R. Brodersen, "Large-signal millimeter-wave CMOS modeling with BSIM3," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium, Digest of Papers*, 2004, pp. 163–166.

[61] X. Wei, G. Niu, Y. Li, M.-T. Yang, and S. Taylor, "Modeling and Characterization of Intermodulation Linearity on a 90-nm RF CMOS Technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, no. 4, pp. 965–971, 2009.

[62] J. Fossum, R. Mertens, D. Lee, and J. Nijs, "Carrier recombination and lifetime in highly doped silicon," *Solid-State Electronics*, vol. 26, no. 6, pp. 569 –576, 1983.

[63] B. Dormieu, P. Scheer, C. Charbuillet, H. Jaouen, and F. Danneville, "Revisited RF Compact Model of gate resistance suitable for high-k metal gate technology," *IEEE Transactions on Electron Devices*, vol. 60, no. 1, pp. 13–19, 2013.

[64] X. Jin, J.-J. Ou, C.-H. Chen, W. Liu, M. Deen, P. Gray, and C. Hu, "An effective gate resistance model for CMOS RF and noise modeling," in *International Electron Devices Meeting Technical Digest, (IEDM)*, 1998, pp. 961–964.

[65] M. Chan, K. Hui, C. Hu, and P.-K. Ko, "A robust and physical BSIM3 non-quasi-static transient and AC small-signal model for circuit simulation," *IEEE Transactions on Electron Devices*, vol. 45, no. 4, pp. 834–841, 1998.

[66] M. V. Dunga, "A scalable MOS Device substrate resistance model for RF and Microwave circuit simulation," Master's thesis, EECS Department, University of California, Berkeley, May 2004. [Online]. Available: `http://www.eecs.berkeley.edu/Pubs/Theses/Data/24800.pdf`.

[67] M. van Exter and D. Grischkowsky, "Carrier dynamics of electrons and holes in moderately doped silicon," *Phys. Rev. B*, vol. 41, pp. 12 140–12 149, 17 Jun. 1990.

[68] H. Hasegawa, M. Furukawa, and H. Yanai, "Properties of Microstrip Line on Si-SiO_2 system," *IEEE Transactions on Microwave Theory and Techniques*, vol. 19, no. 11, pp. 869–881, 1971.

[69] J. Bazzi, A. Curutchet, F. Pourchon, N. Derrier, D. Celi, and T. Zimmer, "Investigation of De-embedding procedures up to 110GHz," in *MOS-AK/GSA Workshop, Paris*, 2011. [Online]. Available: `http://www.mos-ak.org/paris/papers/P13_Bazzi_MOS-AK_Paris.pdf`.

[70] A. Scholten, R. van Langevelde, L. Tiemeijer, R. Havens, and D. Klaassen, "Compact MOS modelling for RF CMOS circuit simulation," in *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2001, pp. 197–201.

[71] S. Venugopalan, M. Karim, S. Salahuddin, A. Niknejad, and C. Hu, "Phenomenological Compact Model for QM Charge Centroid in Multigate FETs," *IEEE Transactions on Electron Devices*, vol. 60, no. 4, pp. 1480–1484, 2013.

[72] S. Venugopalan, D. D. Lu, Y. Kawakami, P. M. Lee, A. M. Niknejad, and C. Hu, "BSIM-CG: A compact model of cylindrical/surround gate MOSFET for circuit simulations," *Solid-State Electronics*, vol. 67, no. 1, pp. 79 –89, 2012.

[73] S. Venugopalan, Y. Chauhan, D. Lu, M. Karim, A. Niknejad, and C. Hu, "Modeling intrinsic and extrinsic asymmetry of 3D cylindrical gate/gate-all-around FETs for circuit simulations," in *11th IEEE Non-Volatile Memory Technology Symposium (NVMTS)*, 2011, pp. 1–4.

[74] G. Gildenblat, *Compact Modeling: Principles, Techniques and Applications*. Springer, 2010.

[75] S. Venugopalan, D. Lu, M. Karim, A. Niknejad, and C. Hu, "Compact Models for Real Device Effects in FinFETs," in *17th International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2012, pp. 292–295.

[76] S. Venugopalan, K. Dandu, S. Martin, R. Taylor, C. Cirba, X. Zhang, A. Niknejad, and C. Hu, "A non-iterative physical procedure for RF CMOS compact model extraction using BSIM6," in *IEEE Custom Integrated Circuits Conference (CICC)*, 2012, pp. 1–4.

[77] K. Ganapathi and S. Salahuddin, "Zener tunneling: congruence between semi-classical and quantum ballistic formalisms," *Journal of Applied Physics*, vol. 111, no. 12, p. 124 506, 2012.

[78] J. Deng and H.-S. Wong, "A compact SPICE model for Carbon-Nanotube field-effect transistors including nonidealities and its application - part I: Model of the intrinsic channel region," *IEEE Transactions on Electron Devices*, vol. 54, no. 12, pp. 3186–3194, 2007.

[79] J.-M. Sallese, M. Bucher, F. Krummenacher, and P. Fazan, "Inversion charge linearization in MOSFET modeling and rigorous derivation of the EKV compact model," *Solid-State Electronics*, vol. 47, no. 4, pp. 677 –683, 2003.

[80] G. Gildenblat, X. Li, W. Wu, H. Wang, A. Jha, R. van Langevelde, G. D. J. Smit, A. Scholten, and D. B. M. Klaassen, "PSP: An advanced surface-potential-based MOSFET model for circuit simulation," *Electron Devices, IEEE Transactions on*, vol. 53, no. 9, pp. 1979–1993, 2006.

[81] (2009). PSP 103.1 documentation, [Online]. Available: `http://pspmodel.asu.edu/psp_documentation.htm`.

[82] N. W. Ashcroft and D. . Mermin, *Solid State Physics*. Saunders College Publishing, 1976.

[83] (2009). Resistivity - Mobility calculator/graph for various doping concentrations in Silicon, [Online]. Available: `http://www.cleanroom.byu.edu/ResistivityCal.phtml`.

[84] J. Long and M. Copeland, "The modeling, characterization, and design of monolithic inductors for silicon RF IC's," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 3, pp. 357–369, 1997.

[85] Z. Zhu, G. Gildenblat, C. Mcandrew, and I.-S. Lim, "Accurate RTA-based nonquasi-static MOSFET model for RF and mixed-signal simulations," *IEEE Transactions on Electron Devices*, vol. 59, no. 5, pp. 1236–1244, 2012.

[86] A. Scholten, L. Tiemeijer, P. De Vreede, and D. B. M. Klaassen, "A large signal non-quasi-static MOS model for RF circuit simulation," in *International Electron Devices Meeting (IEDM) Technical Digest.*, 1999, pp. 163–166.

[87] M. Bucher and A. Bazigos, "An efficient channel segmentation approach for a large-signal NQS MOSFET model," *Solid-State Electronics*, vol. 52, no. 2, pp. 275 –281, 2008.

[88] H. Wang, X. Li, W. Wu, G. Gildenblat, R. van Langevelde, G. D. J. Smit, A. Scholten, and D. B. M. Klaassen, "A unified Nonquasi-static MOSFET Model for Large-Signal and Small-Signal simulations," *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 2035–2043, 2006.

[89] S. Sarkar, A. S. Roy, and S. Mahapatra, "Unified large and small signal non-quasi-static model for long channel symmetric DG MOSFET," *Solid-State Electronics*, vol. 54, no. 11, pp. 1421 –1429, 2010.

Appendix A

Further Reading for Real Device Effects Models

For more details on each of the real device effects implemented along with the cylindrical gate FET core model developed in this thesis (i.e. in BSIM-CMG) we recommend the following materials. References for both the exact implementation and the state of the art understanding are given. In some cases the state of the art is the model that is implemented, while in some cases well tested trade-offs are chosen in order to keep the overall computational expense of the model lower.

Mobility Degradation
"BSIM4.7.0 Technical Manual" (and references therein), Online:`http://www-device.eecs.berkeley.edu/bsim/Files/BSIM4/BSIM470/BSIM470_Manual.pdf`
R. van Langevelde, A.J. Scholten and D.B.M. Klaassen, "Physical Background of MOS Model 11" (and references therein), Online:`http://www.nxp.com/wcm_documents/models/mos-models/model-11/nl_tn2003_00239.pdf`

Access Resistance
"BSIM4.7.0 Technical Manual", Online:`http://www-device.eecs.berkeley.edu/bsim/Files/BSIM4/BSIM470/BSIM470_Manual.pdf`
R. van Langevelde, A.J. Scholten and D.B.M. Klaassen, "Physical Background of MOS Model 11", Online:`http://www.nxp.com/wcm_documents/models/mos-models/model-11/nl_tn2003_00239.pdf`
"PSP103.1 Technical Note", Online:`http://pspmodel.asu.edu/downloads/psp103p1_summary.pdf`

Velocity Saturation
V_{dsat} Model
"BSIM4.7.0 Technical Manual", Online:`http://www-device.eecs.berkeley.edu/bsim/Files/BSIM4/BSIM470/BSIM470_Manual.pdf`

Drain Current
"PSP103.1 Technical Note", Online:http://pspmodel.asu.edu/downloads/psp103p1_summary.pdf
Linear to Saturation V_{ds} smoothing
K. Joardar et al., "An improved MOSFET model for circuit simulation," IEEE Trans. on Electron Devices, vol.45, no.1, pp.134-148, Jan 1998.

Output Conductance
Channel Length Modulation and DIBL
"BSIM4.7.0 Technical Manual", Online:http://www-device.eecs.berkeley.edu/bsim/Files/BSIM4/BSIM470/BSIM470_Manual.pdf
Single piece model
"PSP103.1 Technical Note", Online:http://pspmodel.asu.edu/downloads/psp103p1_summary.pdf

Impact Ionization - Substrate Current
T.C. Ong, P.-K. Ko, C. Hu, "Modeling of substrate current in p-MOSFET's,", IEEE Electron Device Letters, vol.8, no.9, pp.413-16, Sep 1987.
For SOI MOSFETs
"BSIMSOIv4.4 Users Manual" (and the references therein), Online:http://www-device.eecs.berkeley.edu/bsim/Files/BSIMSOI/bsimsoi4p4/BSIMSOIv4.4_UsersManual.pdf

Gate Induced Drain Leakage Current
S.A. Parke, J.E. Moon, H.-J.C. Wann, P.-K. Ko, C. Hu, "Design for suppression of gate-induced drain leakage in LDD MOSFETs using a quasi-two-dimensional analytical model," IEEE Trans. on Electron Devices, vol.39, no.7, pp.1694-1703, Jul 1992.
R. van Langevelde, A.J. Scholten and D.B.M. Klaassen, "Physical Background of MOS Model 11", Online:http://www.nxp.com/wcm_documents/models/mos-models/model-11/nl_tn2003_00239.pdf

Gate Tunneling Current
Advanced Direct Tunneling Formalism
W.-C. Lee, C. Hu, "Modeling CMOS tunneling currents through ultrathin gate oxide due to conduction-and valence-band electron tunneling ," IEEE Trans. on Electron Devices, vol. 48, no. 7, pp. 1366-1373, July 2001.
Gate Current Partition in BSIM4
J. Hu, X Xi, A Niknejad, C Hu,"On gate leakage current partition for MOSFET compact model," Solid State Electronics, vol. 50, no. 11-12, pp. 1740-1743, Nov-Dec, 2006.
Gate Current Partition for a Charge/Surface Potential based model
R. van Langevelde, A.J. Scholten and D.B.M. Klaassen, "Physical Background of MOS Model 11", Online:http://www.nxp.com/wcm_documents/models/mos-models/model-11/nl_tn2003_00239.pdf

Source-Drain Junction Current and Capacitance

"BSIM4.7.0 Technical Manual", Online:http://www-device.eecs.berkeley.edu/bsim/Files/BSIM4/BSIM470/BSIM470_Manual.pdf

A. J. Scholten, G. D. J. Smit, R. van Langevelde, D. B. M. Klaassen, "JUNCAP",Online:http://www.nxp.com/models/simkit/other-models/juncap.html

Substrate Network

"BSIM4.7.0 Technical Manual", Online:http://www-device.eecs.berkeley.edu/bsim/Files/BSIM4/BSIM470/BSIM470_Manual.pdf

M. V. Dunga, "A scalable MOS Device substrate resistance model for RF and Microwave circuit simulation", Master's thesis, EECS Department, University of California, Berkeley, May 2004. Online: http://www.eecs.berkeley.edu/Pubs/Theses/Data/24800.pdf.

Noise Models

Thermal Noise

Darsen Lu, "Compact Models for Future Generation CMOS," PhD thesis, EECS Department, University of California, Berkeley, May 2011. Online:http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-69.pdf.

Flicker Noise

K.K. Hung, P.K. Ko, C. Hu and Y.C. Cheng, A Physics-Based MOSFET Noise Model for Circuit Simulators, IEEE Trans. Electron Devices, vol. 37, no. 5, pp. 1323-1333, 1990.

Other noise components such as thermal noise due to the substrate, electrode gate, and source/drain resistances and shot noise due to various gate tunneling components are trivial.

Gate Electrode Resistance

"BSIM4.7.0 Technical Manual", Online:http://www-device.eecs.berkeley.edu/bsim/Files/BSIM4/BSIM470/BSIM470_Manual.pdf

B. Dormieu et al., "Revisited RF Compact Model of Gate Resistance Suitable for High-k /Metal Gate Technology," IEEE Trans. on Electron Devices, vol.60, no.1, pp.13,19, Jan. 2013.

Non-Quasi Static Effects

Channel induced Gate Resistance Model

X. Jin, J.J. Ou, C.-H. Chen, W. Liu, J.M. Deen, P.R. Gray, C. Hu, "An effective gate resistance model for CMOS RF and noise modeling," International Electron Devices Meeting, (IEDM) Technical Digest, pp.961-64, 6-9 Dec. 1998.

Relaxation Time Approach

M. Chan, K.Y. Hui, C. Hu, P.-K. Ko, "A robust and physical BSIM3 non-quasi-static transient and AC small-signal model for circuit simulation," IEEE Trans. on Electron Devices, vol.45, no.4, pp.834-41, Apr 1998.

Surface Potential Spline Collocation Method

H. Wang, X. Li, W. Wu, G. Gildenblat et al., "A Unified Nonquasi-Static MOSFET Model for Large-Signal and Small-Signal Simulations," IEEE Trans. on Electron Devices, vol.53, no.9, pp.2035-43, Sept. 2006.
Spline Collocation Implementation and Coefficients
"PSP103.1 Technical Note", Online:`http://pspmodel.asu.edu/downloads/psp103p1_summary.pdf`

Self Heating
"BSIMSOIv4.4 Users Manual" (and the references therein), Online:`http://www-device.eecs.berkeley.edu/bsim/Files/BSIMSOI/bsimsoi4p4/BSIMSOIv4.4_UsersManual.pdf`

Short Channel Effects
General methodology
"BSIM4.7.0 Technical Manual" (and references therein), Online:`http://www-device.eecs.berkeley.edu/bsim/Files/BSIM4/BSIM470/BSIM470_Manual.pdf`
Scaling Lengths
Chung-Hsun Lin, "Compact Modeling of Nanoscale CMOS," PhD thesis, EECS Department, University of California, Berkeley, Dec 2007. Online:`http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-169.pdf`.
C. Auth and J. D. Plummer, "Scaling theory for cylindrical, fully-depleted, surrounding-gate MOSFET's," IEEE Electron Device Letters, vol. 18, no. 2, pp. 7476, 1997.

Temperature Dependence
"BSIM4.7.0 Technical Manual", Online:`http://www-device.eecs.berkeley.edu/bsim/Files/BSIM4/BSIM470/BSIM470_Manual.pdf`
Scaling Lengths
Although most often for parameters that do not have a direct physical interpretation the dependence is figured out by working on a diverse set of hardware data.

Appendix B

MOSFET Short Channel Capacitance Model

In this appendix we shall outline a short channel capacitance / charge model that incorporates velocity saturation and channel length modulation effects into a long channel capacitance model. We will discuss this keeping a charge based bulk MOSFET model like EKV/BSIM6 in perspective. This derivation is applicable for multi-gate MOSFET models as well (like BSIM-CMG) where in drain current and terminal capacitance are made functions source end and drain end channel charges. We will first briefly go through the derivation of a long and short channel drain current. We will then present the derivation of velocity saturation effect incorporated charge model in a unified fashion along with the regular long channel charge model. We will finally add channel length modulation / output conductance effects to the charge model as well.

B.1 Drain Current Derivation

The long channel drift-diffusion drain current equation with charge sheet approximation is given as follows,

$$I_d = \mu_{eff} W \left(-Q_i \frac{d\psi_s}{dy} + V_t \frac{dQ_i}{dy} \right) \tag{B.1}$$

where W is the width of the device, μ_{eff} is the long channel mobility with mobility degradation due to vertical field, Q_i is the channel charge, ψ_s is the surface potential and y is the direction along the channel with the source at $y = 0$ and the drain terminal at $y = L$ (length of the device). At this juncture to make the derivation tractable, a linear approximation relating the charge and surface potential is sought as follows,

$$dQ_i = n_q C_{ox} d\psi_s \tag{B.2}$$

where n_q is a bias dependent pre-factor and C_{ox} is the gate capacitance per unit area, [79]. With this approximation the drain current equation can be written as,

$$I_d = -K(2q+1)\frac{dq}{dy} \tag{B.3}$$

where,

$$K = 2n_q\mu_{eff}C_{ox}V_t^2$$

$$q = -\frac{Q_i}{2n_qC_{ox}V_t}$$

Eqn.(B.3) is integrated from source to drain to obtain the final expression for long channel drain current as follows,

$$I_d = K(q_s^2 - q_d^2 + q_s - q_d) = K(2q_a + 1)\Delta q \tag{B.4}$$

$$q_a = 0.5 * (q_s + q_d)$$

$$\Delta q = (q_s - q_d)$$

where q_s and q_d are normalized (to $-2n_qC_{ox}V_t$) source and drain side channel charge obtained from solving the implicit surface potential/charge based equation describing the electrostatics of the MOSFET (which were obtained from a solution for Poisson equation with appropriate boundary conditions). However for moderate to short channel length Eqn.(B.4) fails to capture the velocity saturation of the carriers due to high lateral electrical field especially near the drain. Velocity saturation mechanism is incorporated in the long channel drain current,Eqn.(B.3) using a lateral field dependent mobility model as follows,

$$I_d = -K\frac{2q+1}{\sqrt{1+\left(\frac{\mu_{eff}E}{v_{sat}}\right)^2}}\frac{dq}{dy} \tag{B.5}$$

where $E = d\psi_s/dx$ is the lateral electric field and v_{sat} is the saturation velocity of the carriers. This choice of velocity saturation model helps in retaining the symmetry and higher order continuity of the model w.r.t V_{ds}. Using Eqn.(B.2) again, we can re-write Eqn.(B.5) as follows

$$I_d = -K\frac{(2q+1)}{\sqrt{1+\left(\frac{\mu_{eff}}{v_{sat}}\frac{dq}{dy}\right)^2}}\frac{dq}{dy} \tag{B.6}$$

After separating out the derivative term this equation can be written as,

$$I_d = -K(2q+1)\frac{dq}{dy}\sqrt{1-\frac{I_d^2\left(\frac{\mu_{eff}}{v_{sat}}\right)^2}{K^2(2q+1)^2}} \tag{B.7}$$

The above equation is an implicit equation in I_d and might require iterative procedures to solve. But we resort to an approximation ($\sqrt{1-x} \approx 1 - x/2$) that retains the accuracy of the model,

$$I_d = -\frac{dq}{dy}\left(K(2q+1) - \frac{I_d^2 \left(\frac{\mu_{eff}}{v_{sat}}\right)^2}{K(2q+1)} \right) \tag{B.8}$$

Integrating this equation from source to drain one obtains,

$$I_d = \left[K(2q_a+1) - \frac{I_d^2 \left(\frac{\mu_{eff}}{v_{sat}}\right)^2}{2K(2q_a+1)} \right] \frac{\Delta q}{L} \tag{B.9}$$

Solving the quadratic equation, Eqn.(B.9) for I_d, we obtain the drain current with velocity saturation effect as follows,

$$I_d = \frac{K}{L} \frac{2q_a+1}{0.5\left[1+\sqrt{1+2\left(\frac{\mu_{eff}}{v_{sat}}\right)^2\left(\frac{\Delta q}{L}\right)^2}\right]} \Delta q = \frac{K}{L}\frac{2q_a+1}{D_{vsat}}\Delta q \tag{B.10}$$

Other than a modified mobility model, it is also common practice in Compact Models to introduce velocity saturation through a drain saturation voltage,V_{dsat} for model stability. V_{ds} is smoothly restricted to V_{dsat} resulting in a effective drain voltage,V_{dseff}. This V_{dseff} is used to determine the drain side charge now given by q_{deff} instead of q_d in Eqn.B.10. Output conductance effects like channel length modulation and drain induced barrier lowering are later added as a correction factor to Eqn.(B.10). For example, channel length modulation is added as follows,

$$I_d = \frac{K}{L}\frac{2q_a+1}{D_{vsat}}\Delta q \cdot M_{clm} \tag{B.11}$$

where

$$M_{clm} = 1 + \frac{\Delta L}{L} = 1 + P_{clm} * ln\left(1 + \frac{V_{ds} - V_{dseff}}{V_{dsat} + E_{sat}L}\frac{1}{P_{clm}}\right) \tag{B.12}$$

where ΔL is the length of the pinched-off / saturated region, $E_{sat} = 2v_{sat}/\mu_{eff}$ and P_{clm} is a tuning parameter. q_a is $0.5(q_{deff} + q_s)$ and Δq is $q_s - q_{deff}$.

B.2 Terminal Charge Derivation

In order to facilitate the derivation of terminal charge, let us first derive an explicit expression for channel charge along the channel, $q(y)$, [80]. For the long channel case, equating Eqn.(B.3) and Eqn.(B.4) we write,

$$\frac{dy}{dq} = -\frac{L}{\Delta q}\frac{2q+1}{2q_a+1} \tag{B.13}$$

Changing the variable from q to $s = q - q_a$ for convenience we get,

$$\frac{dy}{ds} = -\frac{L}{\Delta q}\frac{2s + G_L}{G_L} \tag{B.14}$$

where,

$$G_L = (2q_a + 1)$$

On similar lines equating Eqn.(B.8) and Eqn.(B.10) for short channel device and changing variable from q to s we get (use Eqn.(B.9)),

$$\frac{dy}{ds} = -\frac{L}{\Delta s}\frac{2s + G_S}{G_S} \tag{B.15}$$

where,

$$G_S = \frac{(2q_a + 1)}{D_{vsat}}$$

We observe the similarity in form for Eqn.(B.14) and Eqn.(B.15). This helps to unify the derivation for the terminal charge that we are about to discuss for both long and short channel device. In what follows we will just use G to represent both G_L and G_S. Integrating w.r.t. s (either Eqn.(B.14) or Eqn.(B.15)) we get,

$$y = -\frac{L}{\Delta q G}(s^2 + s \cdot G) + \frac{L}{2G}\left(\frac{\Delta q}{2} + G\right) \tag{B.16}$$

The total inversion charge in the channel is given by,

$$Q'_I = W\int_0^L Q_i dy = -2WC_{ox}n_q V_t \int_0^L q dy \tag{B.17}$$

Changing the variable of integration to s and replacing dy using Eqn.B.15 (the limits of the integral are given by $s = q_s - q_a = \Delta q/2$ at the source and $s = q_d - q_a = -\Delta q/2$ at the drain), we get

$$Q'_I = -2WC_{ox}n_q V_t \frac{L}{\Delta q}\int_{-\Delta q/2}^{\Delta q/2}(s + q_a)\frac{2s + G}{G} \cdot ds \tag{B.18}$$

Integrating the above yields,

$$Q'_I = -2WLC_{ox}n_q V_t \left(q_a + \frac{1}{6}\frac{\Delta q^2}{G}\right) \tag{B.19}$$

The source and drain terminal charges are calculated using Ward-Dutton partition, [35]. The drain terminal charge is given by,

$$Q'_D = W\int_0^L \frac{y}{L} Q_i dy = -2WC_{ox}n_q V_t \int_0^L \frac{y}{L} q dy \tag{B.20}$$

By changing the variable of integration to s and using Eqns.(B.15) and (B.16), the integral changes to,

$$Q'_D = -2WC_{ox}n_qV_t \int_{-\Delta q/2}^{\Delta q/2} \frac{1}{\Delta q}\frac{2s+G}{G}\left[\frac{L}{\Delta qG}(s^2+s\cdot G) - \frac{L}{2G}\left(\frac{\Delta q}{2}+G\right)\right](s+q_a)\cdot ds \tag{B.21}$$

Solving the above integral, the analytic expression for drain terminal charge is given by,

$$Q'_D = -WLC_{ox}n_qV_t\left[q_a - \frac{\Delta q}{6}\left(1-\frac{\Delta q}{G}-\frac{\Delta q^2}{5G^2}\right)\right] \tag{B.22}$$

Similarly, the source terminal charge is given by

$$Q'_S = W\int_0^L \left(1-\frac{y}{L}\right)Q_i dy = -2WC_{ox}n_qV_t\int_0^L\left(1-\frac{y}{L}\right)qdy \tag{B.23}$$

Alternately the source terminal charge is simply got by

$$Q'_S = Q'_I - Q'_D \tag{B.24}$$

The body charge per unit area at an arbitrary point in the channel is given by,

$$Q_b = -C_{ox}(V_{gs}-V_{fb}-\psi_p) - \left(1-\frac{1}{n_q}\right)Q_i \tag{B.25}$$

where ψ_p is the surface potential in the channel at $Q_i = 0$, [20]. The body terminal charge is thus given by,

$$Q'_B = \int_0^L W\cdot Q_b\cdot dy \tag{B.26}$$

Once again using Eqn.(B.15) we can integrate and write

$$Q'_B = -WLC_{ox}(V_{gs}-V_{fb}-\psi_p) - WLC_{ox}V_t(n_q-1)\left(q_a+\frac{1}{6}\frac{\Delta q^2}{G}\right) \tag{B.27}$$

The gate terminal charge can be obtained from the derived charges taking charge neutrality into consideration as follows,

$$Q'_G = -(Q'_D+Q'_S+Q'_B) \tag{B.28}$$

We obtain the long channel terminal charge equations if we replace G with G_L in Eqns.(B.19), (B.22) and (B.27). In order to incorporate velocity saturation effects in the same, we simply can replace G with G_S in all of them. In the above equations we have neglected the effect of poly-depletion. However it can be shown that including the effect in the derivation will not affect our arguments.

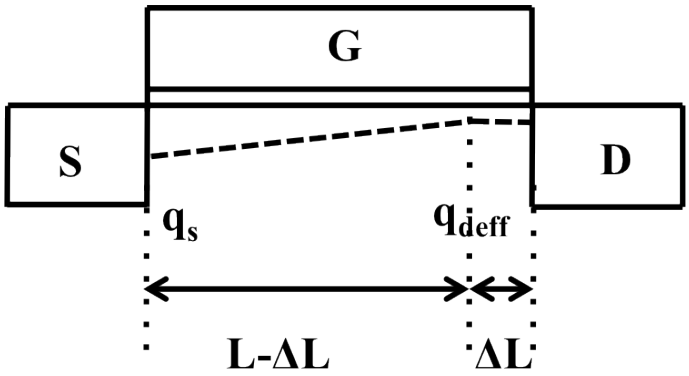


Figure B.1: MOSFET cross-section showing the channel charge (dotted lines) along the channel. Channel gets pinched off at higher V_{ds} at $L - \Delta L$ and the channel charge stays constant until the drain end.

In order to incorporate channel length modulation in the above, we will have to modify Eqn.(B.15) and Eqn.(B.16). Equating Eqn.(B.8) and Eqn.(B.11), Eqn.(B.15) can be re-written as,

$$\frac{dy}{ds} = -\frac{L}{M_{clm} \cdot \Delta s} \frac{2s + G_S}{G_S} \tag{B.29}$$

Integrating Eqn.(B.11), we obtain the modification for Eqn.(B.16) as

$$y = -\frac{L}{M_{clm} \cdot \Delta q \cdot G}(s^2 + s \cdot G) + \frac{L}{2M_{clm} \cdot G}\left(\frac{\Delta q}{2} + G\right) \tag{B.30}$$

With channel length modulation the channel saturates to a charge equal to q_{deff} beyond $y = L - \Delta L$ until the drain terminal at $y = L$. The channel behaves normally between the source terminal,$y = 0$ to $y = L - \Delta L$, Fig.B.1. We can then re-write the total inversion charge in the channel as follows,

$$Q_I = -2WC_{ox}n_qV_t \int_0^{L-\Delta L} q dy - 2WC_{ox}n_qV_t \int_{L-\Delta L}^{L} q_{deff} dy \tag{B.31}$$

Now using Eqn.(B.29) for just the first term and replacing $\Delta L/L$ with $M_{clm} - 1$, after integration of the second term we obtain,

$$Q_I = -\frac{2WLC_{ox}n_qV_t}{M_{clm}}\left(q_a + \frac{1}{6}\frac{\Delta q^2}{G}\right) - 2WLC_{ox}n_qV_t(M_{clm} - 1)q_{deff} \tag{B.32}$$

Similarly the drain terminal charge with channel length modulation is given as,

$$Q_D = -2WC_{ox}n_qV_t \int_0^{L-\Delta L} \frac{y}{L} q dy - 2WC_{ox}n_qV_t \int_{L-\Delta L}^{L} \frac{y}{L} q_{deff} dy \tag{B.33}$$

Using Eqns.(B.29) and (B.30), we get

$$Q_D = -\frac{WLC_{ox}n_qV_t}{M_{clm}^2}\left[q_a - \frac{\Delta q}{6}\left(1 - \frac{\Delta q}{G} - \frac{\Delta q^2}{5G^2}\right)\right] - WLC_{ox}n_qV_t\left(M_{clm} - \frac{1}{M_{clm}}\right)q_{deff} \tag{B.34}$$

In the above, for the second term we have used the approximation that $(L - \Delta L) \approx 1/M_{clm}$. The source terminal charge is given as,

$$Q_S = Q_I - Q_D \tag{B.35}$$

Finally the body terminal charge with channel length modulation taken into consideration is given by

$$Q_B = \int_0^{L-\Delta L} W \cdot Q_b \cdot dy + \int_{L-\Delta L}^{L} W \cdot Q_{beff} \cdot dy \tag{B.36}$$

where Q_{beff} is given by Eqn.(B.25) with $Q_i = 2n_qC_{ox}V_tq_{deff}$. Once again using Eqn.(B.29) we get,

$$Q_B = -\frac{WLC_{ox}}{M_{clm}}\left[(V_{gs} - V_{fb} - \psi_p) - (n_q - 1)V_t\left(q_a + \frac{1}{6}\frac{\Delta q^2}{G}\right)\right] + (M_{clm} - 1)Q_{beff} \tag{B.37}$$

The overall gate terminal charge can be obtained as follows,

$$Q_G = -(Q_D + Q_S + Q_B) \tag{B.38}$$

The terminal charges so derived in Eqns.(B.34), (B.35), (B.37) and (B.38) with $G = G_S$ represent the terminal charge model that is valid from long channel through short channel capturing both velocity saturation and channel length modulation effects in them. This final result is similar in nature to the one obtained for surface potential based models as well, [81].

The MOSFET capacitance / trans-capacitance are then given by,

$$C_{ij} = (2\delta_{ij} - 1)\frac{\partial Q_i}{\partial V_j} \tag{B.39}$$

where i, j stand for different terminals (D, G, S, B) and δ_{ij} is the Kronecker delta function. In a Compact Model environment in order to offer flexibility in tuning the model to fit data from various sources, the values of parameters for velocity saturation, v_{sat} and channel length modulation, P_{clm} are different for DC drain current models and AC capacitance models.

Appendix C

Revisiting Substrate Network Model for High Frequencies

In this Appendix we will present back of the envelope calculations to revisit the accuracy of the FET resistive substrate network model for RF, mm-wave and terahertz (THz) region frequencies.

According to the Drude model for free carriers in a conductor, the conductivity of a material decreases with increasing frequency from its low frequency (DC) value due to increasing carrier-carrier collisions within the material, [82]. At very high frequencies where conductivity drops off to very low values, the material tends to behave like a true dielectric (with dielectric constant ϵ_∞). The frequency dependent dielectric constant is given as follows,

$$\epsilon = \epsilon_\infty + \frac{j\sigma}{\omega\epsilon_0} \tag{C.1}$$

where ϵ_0 is the permittivity of free space and σ is the frequency dependent complex conductivity given by,

$$\sigma = \sigma_{dc}\frac{j\Gamma}{\omega + j\Gamma} \tag{C.2}$$

σ_{dc} is the DC conductivity given by $\sigma_{dc} = qN\mu$ and $\Gamma = 1/\tau$ is the damping rate where τ is the average collision time, [82]. N is the carrier density (or doping concentration) and one might recognize μ as the carrier mobility defined as $\mu = q\tau/m$ where m is the effective conductivity mass of the carriers.

Eq.(C.1) can now be simplified and written as follows,

$$\epsilon = \epsilon_{eff} + \frac{j\sigma_{eff}}{\omega\epsilon_0} \tag{C.3}$$

$$\epsilon_{eff} = \epsilon_\infty - \frac{\omega_p^2}{\omega^2 + \Gamma^2}$$

$$\sigma_{eff} = \frac{\epsilon_0\Gamma\omega_p^2}{\omega^2 + \Gamma^2}$$

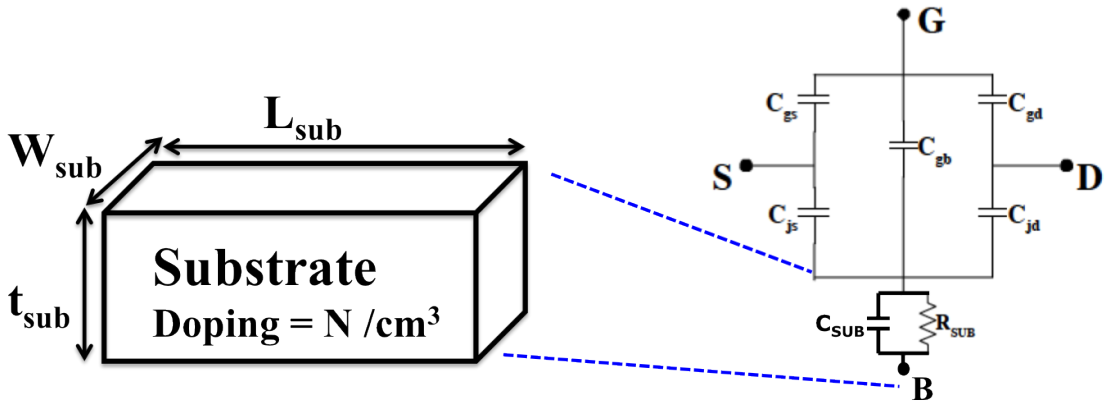


Figure C.1: A simplfied lumped model of the substrate below a MOSFET with doping of $N\,\mathrm{cm}^{-3}$. At high frequencies the substrate behaves both like a capacitor and a resistor.

In above the plasma angular frequency, ω_p is defined as

$$\omega_p = \sqrt{\frac{q^2 N}{\epsilon_0 m}} \tag{C.4}$$

Thus silicon substrate like any other material is both a dielectric and a conductor, [67]. This would imply that an effective lumped model for any material would be a resistor in parallel with a capacitor. Let us consider a simple substrate block of width W_{sub}, length L_{sub} and thickness t_{sub}, Fig.C.1. If the contacts were placed on either ends (across t_{sub} dimension) of this piece, the resistance and capacitance of this substrate is given by,

$$R_{sub} = \frac{1}{\sigma_{eff}} \frac{t_{sub}}{W_{sub} L_{sub}} \tag{C.5}$$

$$C_{sub} = \frac{\epsilon_0 \epsilon_{eff}}{t_{sub}} \cdot W_{sub} L_{sub}$$

The effective impedance of the substrate is then given by,

$$Z_{sub} = \frac{R_{sub}}{1 + j\frac{\omega}{\omega_{sub}}} \tag{C.6}$$

where ω_{sub} is the substrate corner angular frequency given by,

$$\omega_{sub} = 2\pi f_{sub} = \frac{1}{R_{sub} C_{sub}} = \frac{\sigma_{eff}}{\epsilon_0 \epsilon_{eff}} \tag{C.7}$$

The corner frequency f_{sub} could be interpreted as the frequency around which the capacitive effects of the substrate start to dominate over the resistive effects. Both σ_{eff} and ϵ_{eff} are

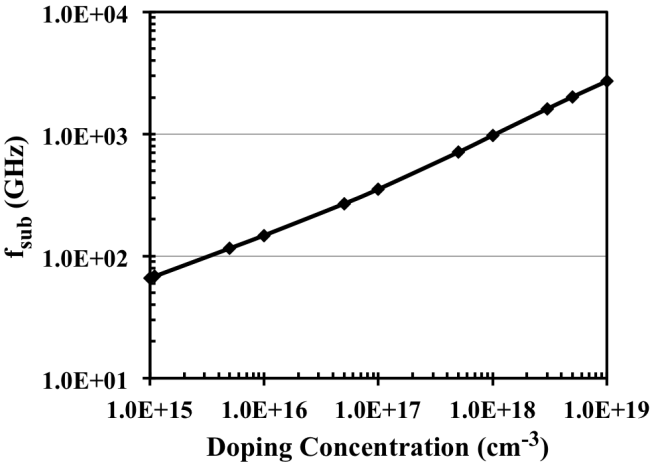


Figure C.2: Substrate cut-off frequency, f_{sub} around and beyond which capacitive effects of the substrate dominate as a function of p-type Boron doping for a Silicon substrate

functions of frequency and hence Eqn.(C.7) is required to be solved to obtain an analytic expression for f_{sub}. Using Eqns.(C.7) and (C.3)

$$\omega_{sub} = \frac{\Gamma}{\frac{\epsilon_\infty(\omega_{sub}^2+\Gamma^2)}{\omega_p^2} - 1} \tag{C.8}$$

Rearranging the above we obtain a cubic equation in f_{sub} as follows,

$$\epsilon_\infty \left(\frac{\Gamma}{\omega_p}\right)^2 \left(\frac{2\pi f_{sub}}{\Gamma}\right)^3 + \epsilon_\infty \left(\frac{\Gamma}{\omega_p}\right)^2 \left(\frac{2\pi f_{sub}}{\Gamma}\right) - 1 = 0 \tag{C.9}$$

The values for ω_p can be calculated for various doping levels using Eqn.(C.4). $\Gamma = 1/\tau$ can be obtained from known measured values for bulk mobility ($\mu = q\tau/m$) at various doping levels from literature, [3] or using online tools such as at [83]. With these values we solve for f_{sub} for various p-type (Boron) doping levels in a Silicon substrate. We then plot f_{sub} for various bulk doping levels for a p-type (Boron) Silicon, Fig.C.2.

For very low doping levels ($N = 10^{15}\,\mathrm{cm}^{-3}$) like those present below RF passive structures (a well-doping or threshold adjust implants are masked to retain the native substrate doping) the value for f_{sub} was found to be about 66 GHz. It rises to about 116 GHz for $N = 10^{16}\,\mathrm{cm}^{-3}$. For both RF ($0.3-30$ GHz) and mm-wave ($30-300$ GHz) applications the capacitance effects of the substrate are significant and rightfully so, circuit models for on-chip passives model the substrate as both a resistor and a capacitor, [84]. Due to low substrate cut-off frequency substrate with native low doping tend to behave more like a dielectric, minimizing Eddy current losses.

However for doping levels of $N = 10^{18}\,\mathrm{cm}^{-3}$ like those found below a typical MOSFET (well-doping) f_{sub} tends to approach a very high value of 0.97 THz. And for $N = 3\times10^{18}\,\mathrm{cm}^{-3}$

(observed in sub 45nm technologies that employ retrograde doping profiles), f_{sub} rises to even higher value of 1.6 THz. Clearly for applications below 500 GHz the substrate behaves resistively. The five resistor network lumped model previously discussed in Chapter 5 should be sufficient. However with scaling as the cut-off frequencies of the MOSFET approach 400 GHz and higher, the substrate model will need to be re-visited. The lumped first order model of R_{sub}-C_{sub} visualization for the substrate was used for an easy understanding. However given both the source/drain and substrate contacts in a MOSFET are placed on the same plane, a more distributed network should be considered. Distributed networks are known to be more broadband than a lumped network. Thus the actual f_{sub} value would be higher than the value predicted here.

Appendix D

Non-Quasi Static Effect Models

The operation of a MOSFET is described by the Poisson equation describing the electrostatics and the current-continuity equation describing the dynamics in a self-consistent manner given below [1], [3].

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{\rho}{\epsilon_{si}} \tag{D.1}$$

$$W \frac{\partial Q(x,t)}{\partial t} = \frac{\partial I(x,t)}{\partial x} \tag{D.2}$$

where x is the direction along the channel, ψ is the potential (usually taken at the channel-oxide interface), ρ is the charge density (in cm^{-3}) (both mobile and fixed charges), Q is the channel mobile charge (in cm^{-2}), t signifies time, and $I(x,t)$ is the current in the channel. Gradual channel approximation is assumed in the 1-dimensional Poisson equation, i.e. the vertical field dominates over the lateral field along the channel in the transistor. The continuity equation describes the fact that there is no build up of charges in a MOSFET along the channel.

In contemporary MOSFET compact models such as one developed in this thesis in Chapter 2 (or Appendix B), the quasi-static (QS) assumption is evoked (i.e. $\partial Q/\partial t = 0$) and a steady-state drain current expression is derived to describe the DC operation. The AC or the small-signal operation is described by the terminal charges. The quasi-static assumption implies that steady-state current and the channel charges establish instantaneously after terminal voltages are applied to the device. This is not true especially when the device is subject to high slew rate (high dV/dt) signals - either large voltage swings in a short time or very high frequency signals (approaching or higher than the cut-off frequency, F_t of the transistor) or both. There is an inherent delay in the response of the transistor's currents and charges to applied voltages (often visualized as a distributed resistance-capacitance network). As research keeps pushing the envelope for circuit applications today's circuit designs are either approaching F_t (such as tera-hertz CMOS or long channel length transistor driven by shorter

[1] In dimension scaled devices the quantum effects need to be taken in to consideration as well by adding in Schrodinger equation to the mixture, but we will ignore it in this discussion.

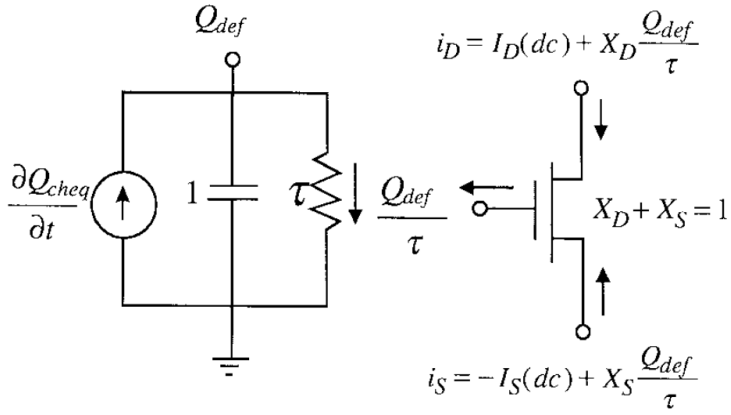


Figure D.1: A relaxation time approach for capturing non-quasi static effects. R-C sub-circuit representation of Eqn.(D.3) for SPICE implementation and the corresponding terminal charges.

channel transistor) or are being subject to high slew rate signals (such as in CMOS RF power amplifiers). One needs to note here that F_t is a function of terminal bias, especially gate voltage, Fig.5.19. Many circuit applications today are subjected to near sub-threshold gate voltages to lower operating power and hence will inherently have lower F_t than that reported as figure of merit for that device (which is usually at the highest operating voltage). For these cases today's compact models are required to support non-quasi static (NQS) mode of operation for being able to predict circuit behavior accurately.

D.1 Relaxation Time Approximation Model

A simple and elegant way to capture the non-quasi static behavior of the channel uses the relaxation time approach that tracks the deficient or surplus charge in the channel, [65].

$$\frac{dQ_{def}}{dt} = \frac{dQ_{cheq}}{dt} - \frac{Q_{def}}{\tau} \tag{D.3}$$

where $Q_{def} = Q_{nqs} - Q_{cheq}$ is the deficient/surplus channel, and Q_{nqs} is the channel charge considering non-quasi static effects. Q_{cheq} is the channel charge at steady state equilibrium or the quasi-static charge (same as Q'_I in Appendix B). τ is the relaxation time constant given as follows,

$$\tau = \left[X_1 \cdot NF\frac{\mu_{eff}W}{L}\left(Q_{ch} + X_2\frac{kT}{q}C_{ox}\right)\right]^{-1} \tag{D.4}$$

where Q_{ch} is the integrated channel charge. Within a SPICE model Eqn.(D.1) is implemented as a sub-circuit whose node-voltage tracks the deficient/surplus charge, Q_{def}, Fig.D.1. Parameters X_1 and X_2 are used for model flexibility. The source and drain terminal charges in

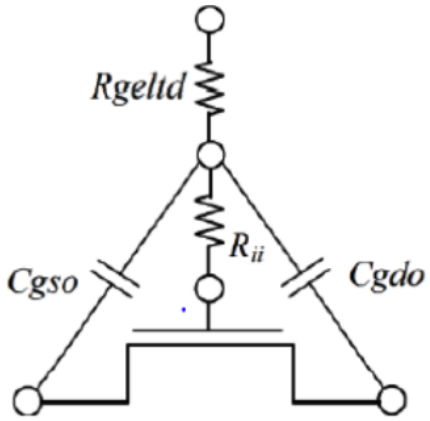


Figure D.2: A channel induced gate resistance, R_{ii} is added in series with the gate physical electrode resistance to capture the non-quasi static effect for gate terminal excited FETs.

this case are given by,

$$Q_{D,NQS} = X_{part}\frac{Q_{def}}{\tau} \tag{D.5}$$

$$Q_{S,NQS} = (1 - X_{part})\frac{Q_{def}}{\tau} \tag{D.6}$$

These terminal charges $Q_{D,NQS}$ and $Q_{S,NQS}$ replace their quasi-static equivalents Q'_D and Q'_S in Appendix B. The bias-dependent partition fraction X_{part} is approximated to its quasi-static equivalent and can be obtained as follows,

$$X_{part} = \frac{Q'_D}{Q'_I} \tag{D.7}$$

$Q_{B,NQS}$ is the same as Q'_B as the body terminal charge does not experience any non-quasi static effect (the holes traverse through a p-doped bulk region quickly). A more rigorous evaluation of this approach found it to be accurate up to $2F_t$ when compared with TCAD based simulations [85].

D.2 Channel Induced Gate Resistance Model

Another first-order model that helps capture non-quasi static effects in the channel is the channel induced gate resistance method, [64]. In this method a bias dependent gate resistance is added at the gate terminal whose value is proportional to the channel resistance, Fig.D.2 (**not to be confused with the physical gate electrode resistance**). The induced gate resistance used here is the same as the relaxation time constant used in the relaxation time

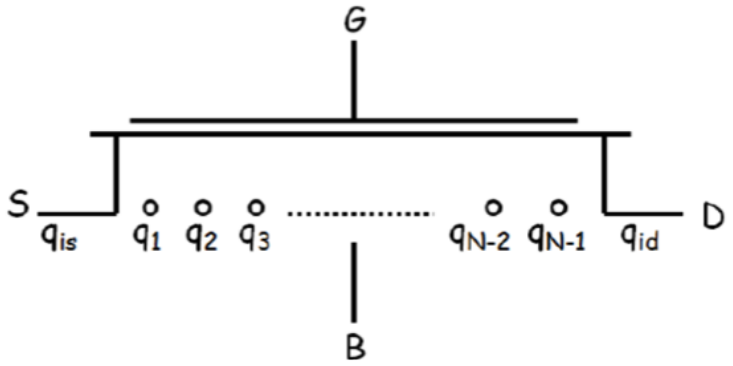


Figure D.3: $N_{SEG} = N$ charge segments in a MOSFET channel for non-quasi static effects simulation. q_i represents the channel charge at the i^{th} intermediate node.

approach (the capacitance in the previous case was assume unity).

$$R_{g,ch} = \left[X_1 NF \frac{\mu_{eff} W}{L} \left(Q_{ch} + X_2 \frac{kT}{q} C_{ox} \right) \right]^{-1} \tag{D.8}$$

Once again the parameters X_1 and X_2 have to be extracted from high-frequency data before model usage (see Chapter 5). The default value for $X_1 = 12$ that is similar to the result obtained for equivalent resistance of a distributed transmission line model with double side contact (here source and drain) and $X_2 = 1$. While this method performs similar to the relaxation time approach in terms of accuracy up to F_t of the device it is restricted in terms to applicability. This model should be used in cases where only the gate terminal is directly excited by an input signal. Non-quasi static effects for applications such as passive mixers or common-gate LNAs where the source terminal is excited cannot be captured by this model. The relaxation time constant approach is however valid for all terminal excitation as we deal with the terminal charges directly.

D.3 Charge Segmentation Model

As the channel length approaches infinitesimally small values, the carrier transit times through the channel tend to become small (large F_t). For this device the QS approximation is still valid. Using this concept we can then visualize a transistor as a series of connected shorter channel length transistors or as we shall refer to them here as 'charge segments' (owing to each smaller transistor carrying a part of the whole channel charge), Fig.D.3. Using Eqn.(B.10) in Eqn.(D.2), the continuity equation can be re-written as follows,

$$\frac{\partial q}{\partial t} + \mu_{eff} V_t \frac{\partial}{\partial x} \left[\frac{(2q+1)}{0.5 \left[1 + \sqrt{1 + 2 \left(\frac{\mu_{eff}}{v_{sat}} \frac{\partial q}{\partial x} \right)^2} \right]} \frac{\partial q}{\partial x} \right] = 0 \tag{D.9}$$

where the symbols follow the convention used in Appendix B. This equation is valid at any point along the channel and includes velocity saturation effects. The solution to this continuity equation along the channel captures the non-quasi static effects we are seeking. One way to accomplish this in a SPICE simulator environment is through a simple series connection of quasi-static transistors, commonly known as the channel segmentation approach [17, 86, 87]. If implemented within a compact model skillfully with care *by not including* short channel effects for each segment (whose channel length is L/N_{SEG}, where N_{SEG} is the number of segments) or by not adding series resistance and parasitic capacitance multiple times for each segment, this method captures all the important features of NQS effects. However often times this approach is ill-constrained (only first order continuity) leading to longer simulation times or non-convergence. A solution to the above continuity equation where the continuity constraint is imposed not just on the first order but even the second and third derivatives through finite difference coefficients is required. One such solution has been developed using spline-collocation method for surface potential based MOSFET model in [88] which we will adopt for a generic charge based model here.

The denominator for the drain current equation to include velocity saturation in Eqn.(D.9) has been so written to maintain consistency between the QS and the NQS models. Eqn.(D.9) can be further simplified into following form,

$$\frac{\partial q}{\partial t} + f\left(q, \frac{\partial q}{\partial x}, \frac{\partial^2 q}{\partial x^2}\right) = 0 \tag{D.10}$$

$$f\left(q, \frac{\partial q}{\partial x}, \frac{\partial^2 q}{\partial x^2}\right) = \frac{\mu_{eff} V_t}{D_v}\left[2\left(\frac{\partial q}{\partial x}\right)^2 - \frac{2q+1}{D_v}\left(\frac{\mu_{eff}}{v_{sat}}\right)^2\left(\frac{\partial q}{\partial x}\right)^2 \frac{\partial^2 q}{\partial x^2}\right]$$

$$D_v = \sqrt{1 + 2\left(\frac{\mu_{eff}}{v_{sat}}\right)^2\left(\frac{\partial q}{\partial x}\right)^2}$$

In order to solve this complex partial-differential equation in a SPICE environment one needs to convert it to (a set of) ordinary differential equations. In [88] the first order weighted residuals method was extended to assure current continuity up to third order. Following a similar approach, assuming the channel is broken down into N_{SEG} segments, the charge in each segment is assumed to be a cubic equation. For example the inversion charge in the n^{th} segment is given by,

$$q(x) = a_n x^3 + b_n x^2 + c_n x + d_n \qquad \frac{n-1}{N_{SEG}} L < x < \frac{n}{N_{SEG}} L \tag{D.11}$$

The boundary conditions for the charge at the source and drain are their respective quasi-static solutions i.e., $q(0) = q_s$ and $q(L) = q_d$. Applying the continuity conditions for q, $\partial q/\partial x$ and $\partial^2 q/\partial x^2$ at the node points in between any two charge segments and using the boundary conditions, one can derive a relation between the cubic equation coefficients (a_n,b_n, c_n, d_n

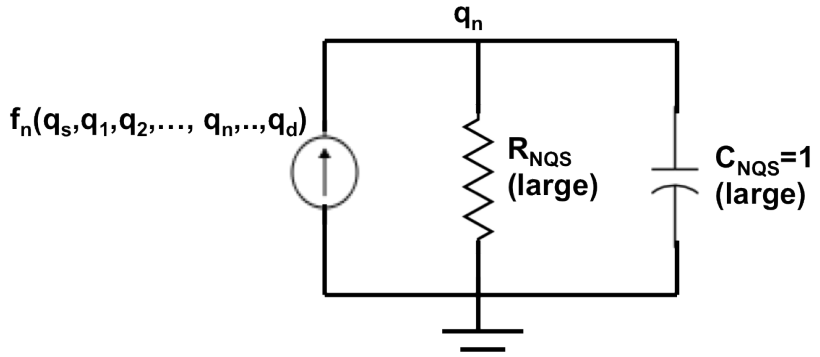


Figure D.4: R-C sub-circuit representation of Eqn.(D.15). The node voltage here represents the channel charge at the n^{th} intermedite node.

for $n = 1$ to $N_{SEG} - 1$) and the charges at these nodes (q_s,..., $q(x = nL/N_{SEG})$,..., q_d). For example, for $N_{SEG} = 3$ the continuity conditions to be imposed are as follows,

$$q\left(\frac{L}{3}\right)^{-} = q\left(\frac{L}{3}\right)^{-} \tag{D.12}$$

$$q\left(\frac{2L}{3}\right)^{-} = q\left(\frac{2L}{3}\right)^{-}$$

$$\left.\frac{\partial q}{\partial x}\right|_{x=L/3^{-}} = \left.\frac{\partial q}{\partial x}\right|_{x=L/3^{+}}$$

$$\left.\frac{\partial q}{\partial x}\right|_{x=2L/3^{-}} = \left.\frac{\partial q}{\partial x}\right|_{x=2L/3^{+}}$$

$$\left.\frac{\partial^2 q}{\partial x^2}\right|_{x=L/3^{-}} = \left.\frac{\partial^2 q}{\partial x^2}\right|_{x=L/3^{+}}$$

$$\left.\frac{\partial^2 q}{\partial x^2}\right|_{x=2L/3^{-}} = \left.\frac{\partial^2 q}{\partial x^2}\right|_{x=2L/3^{+}}$$

Additionally at the boundaries,

$$\left.\frac{\partial^2 q}{\partial x^2}\right|_{x=0} = \left.\frac{\partial^2 q}{\partial x^2}\right|_{x=L} = 0 \tag{D.13}$$

Solution for the $N_{SEG} = 3$ case can be found in [89] derived in the context of a double gate FET charge model. The value for the so derived coefficients remain the same. Only the function $f\left(q, \partial q/\partial x, \partial^2 q/\partial x^2\right)$ changes with FET architecture and assumptions that go into a short channel drain current equation.

The continuity equation at the n^{th} node can be written as

$$\frac{\partial q_n}{\partial t} + f_n\left(q_n, \frac{\partial q_n}{\partial x}, \frac{\partial^2 q_n}{\partial x^2}\right) = 0 \tag{D.14}$$

where q_n denotes the instantaneous channel charge at the n^{th} node. We note that the derivatives of the charge ($\partial q/\partial x$,$\partial^2 q/\partial x^2$) at the intermediate nodes can be expressed as a function of the node charges (since the coefficients a_n etc. are linear function of the node charges after imposing Eqns.(D.12),(D.13)), [89]. A matrix based evaluation of the coefficients of the node charges to express the charge derivatives for any N_{SEG} can be found in [81]. Thus the continuity equation at the n^{th} node changes to

$$\frac{\partial q_n}{\partial t} + f_n\left(q_s, q_1, q_2, ..., q_n, ..., q_d\right) = 0 \tag{D.15}$$

$N_{SEG} - 1$ such continuity equations at each of the nodes can be written in a similar way. Thus the partial differential equation, (D.10) has been converted to $N_{SEG} - 1$ ordinary differential equations (ODE). These $N_{SEG} - 1$ ODEs are coupled as the channel charge at the n^{th} node depends on the charge at all the other nodes through the source function, $f_n()$. These $N_{SEG} - 1$ ODEs can be represented as $N_{SEG} - 1$ R-C sub-circuits within a SPICE model, Fig.D.4. For SPICE implementation purposes the resistor R_{NQS} is chosen to be a large value ($1K\Omega$) to aid in convergence. The capacitance, C_{NQS} from Eqn.(D.15) is unity. To further aid in convergence the initial condition for the n^{th} node is set by the QS solution for the channel charge along the channel given by (using Eqn.(B.16)),

$$q_n(t=0) = q_a - 0.5G_S\left(1 - \sqrt{1 + 4\frac{\Delta q}{G_S}\left(\frac{n}{N_{SEG}} - y_m\right)}\right) \tag{D.16}$$

$$y_m = 0.5 + 0.25\frac{\Delta q}{G_S}$$

The terminal charges $Q_{D,NQS}$ and $Q_{S,NQS}$ at the drain and source end respectively can be obtained by integrating the cubic segments using the Ward-Dutton partition scheme, [35] (without channel length modulation effect).

$$Q'_{D,NQS} = -2WC_{ox}n_qV_t\int_0^L \frac{y}{L}q(y)dy \tag{D.17}$$

$$Q'_{D,NQS} = -2WC_{ox}n_qV_t\left(\frac{1}{90}q_s + \frac{1}{10}q\left(\frac{L}{3}\right) + \frac{4}{15}q\left(\frac{2L}{3}\right) + \frac{11}{90}q_d\right)$$

$$Q'_{S,NQS} = -2WC_{ox}n_qV_t\int_0^L \left(1 - \frac{y}{L}\right)q(y)dy \tag{D.18}$$

$$Q'_{S,NQS} = -2WC_{ox}n_qV_t\left(\frac{11}{90}q_s + \frac{4}{15}q\left(\frac{L}{3}\right) + \frac{1}{10}q\left(\frac{2L}{3}\right) + \frac{1}{90}q_d\right)$$

The total NQS inversion charge is given by,

$$Q'_{I,NQS} = Q'_{D,NQS} + Q'_{S,NQS} \tag{D.19}$$

The resultant expressions are a function of the converged solutions to the ODE at each of the internal node. A general equation for the above for any value of N_{SEG} can be found in [81]. Similar to QS terminal charge derivation in Appendix B, we add the channel length modulation effect to obtain the final set of terminal charges as follows,

$$Q_{I,NQS} = \frac{Q'_{I,NQS}}{M_{clm}} - 2WLC_{ox}n_qV_t(M_{clm}-1)q_{deff} \tag{D.20}$$

$$Q_{D,NQS} = \frac{Q'_{D,NQS}}{M^2_{clm}} - WLC_{ox}n_qV_t\left(M_{clm} - \frac{1}{M_{clm}}\right)q_{deff} \tag{D.21}$$

$$Q_{S,NQS} = Q_{I,NQS} - Q_{D,NQS} \tag{D.22}$$

The body terminal charge is given by,

$$Q_{B,NQS} = -\frac{WLC_{ox}}{M_{clm}}\left[(V_{gs} - V_{fb} - \psi_p)\right] - \frac{1}{2M_{clm}}\left(1 - \frac{1}{n_q}\right)Q'_{I,NQS} + (M_{clm}-1)Q_{beff} \tag{D.23}$$

The overall gate terminal charge can be obtained as follows,

$$Q_{G,NQS} = -(Q_{D,NQS} + Q_{S,NQS} + Q_{B,NQS}) \tag{D.24}$$

The above expressions for $Q_{D,NQS}$, $Q_{S,NQS}$, $Q_{B,NQS}$ and $Q_{G,NQS}$ replace their QS equivalent charges derived in Appendix B in order to obtain the NQS effects.

The main advantage of the charge segmentation model compared to the one implemented based on surface potential is a reduction in overall computation effort due to not having to convert between surface potential and channel charge at every node. The implemented charge segmentation method had similar speed advantages relative to a full-fledged segmentation approach as reported in [88] for a surface potential model i.e. with 3 segments the charge segmentation model was only 1.5 times slower and with 5 segments it was ≈ 2 times slower than the QS model. At 10 segments it was only 3.5 times slower than the QS model. With just 5 segments the implemented model was found to be accurate well beyond $10F_t$ of a bulk device (see Chapter 5).