COMPACT MODEL DEVELOPMENT FOR NANOSCALE FINFETS

Submitted in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

by

Venkatnarayan Hariharan
Roll Number: 04407603

Supervisors:

Prof. V. Ramgopal Rao and Prof. J. Vasi

Department of Electrical Engineering

INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY

2008

To

My Parents,

for being a source of motivation for me in my pursuit of science, and for convincing me that I can finish what I take up.

And To

My Wife and Children,

without whose unflinching support and many sacrifices I would not have been able to undertake this effort.

ॐ भूर्भुवः स्वः ।
तत् सवितुर्वरेण्यं ।
भर्गो देवस्य धीमहि ।
धियो यो नः प्रचोदयात् ॥

O God, Thou art the giver of life, the remover of pain and sorrow, the bestower of happiness; O Creator of the Universe, may we receive Thy supreme sin-destroying light; may Thou guide our intellect in the right direction.

(Gayatri Mantra)

APPROVAL SHEET

Thesis titled **Compact Model Development for Nanoscale FinFETs** by Venkatnarayan Hariharan is approved for the degree of Doctor of Philosophy.

Examiners

_____

_____

_____

Supervisors

_____

_____

Prof. V. Ramgopal Rao and Prof. J. Vasi

Chairman

_____

Date: _____

Place: _____

INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY, INDIA

CERTIFICATE OF COURSE WORK

This is to certify that Mr. Venkatnarayan Hariharan was admitted to the candidacy of the PhD Degree on 3rd Jan 2005, after successfully completing all the courses required for the PhD programme. The details of the course work done are given below.

S. No.	Course No.	Course Name	Credits
1.	EE 620	Physics of Transistors	6
2.	EE 724	Nanoelectronics	6
3.	EE 661	Physical Electronics	6
4.	EE 669	VLSI Technology	6
5.	EEs801	Seminar	4
6.	HS 699	Communication and Presentation Skills	4
7.	EE 618	CMOS Analog VLSI Design (Audit)	6

IIT Bombay Dy. Registrar (Academic)

Dated:

ABSTRACT

As the channel lengths of the conventional planar MOSFET are scaled down to shorter and shorter dimensions, severe short channel effects begin to show up that are getting increasingly difficult to suppress. As a result, several alternate devices have been studied. A FinFET is a variant of the conventional planar MOSFET that is more resilient to short channel effects, and is expected to be adopted at the 32 nm node or more likely at the 22 nm node. Besides technology integration challenges, there are other problems such as the availability of a good compact model that need to be addressed before it can be accepted by circuit designers as a viable technology alternative.

In this work, we have developed a completely closed-form DC model for the intrinsic portion (ie. not considering the parasitics) of a symmetric double gate FinFET. The focus was limited to the 2-dimensional aspect of the FinFET, thereby modeling a symmetrically driven, symmetric double-gate MOSFET (SDGFET). The key feature of the work is that we have redefined the boundary of the *core* model by including velocity saturation effects self-consistently into the model development, in a manner that Gummel symmetry compliance is met. Thus our model is expected to predict the device behavior more accurately when compared to existing approaches in the literature. The other salient feature of our model is that it is not based on the threshold voltage concept and is instead an inversion charge based model. The physical effects of velocity saturation (and the associated channel length modulation), low field mobility degradation, body doping and 2-dimensional field effects are included in the model. We have validated the model for devices as short as 20 nm.

The various approaches followed, and the assumptions and approximations made have been discussed in this report. Validation of the model was done by comparing with 2D numerical simulations.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

b	Structural parameter, given by (61)
β_1 (β_{1s}, β_{1d})	Intermediate constant (β_{1s} and β_{1d} are its values at the source and drain ends respectively)
β_2 (β_{2s}, β_{2d})	Intermediate constant (β_{2s} and β_{2d} are its values at the source and drain ends respectively)
C_{ox}	Gate oxide capacitance per unit area
c_s, η, θ, E_0	Low-field mobility related model parameters
ΔL	Extent of channel length modulation (CLM)
$\Delta\varphi$	Work function difference between the gate electrodes and intrinsic silicon
ε, ε_{ox}	Silicon permittivity, gate oxide permittivity
E_{sat}	Lateral electric field at the onset of velocity saturation
E_{xs}	Lateral electric field at the oxide-silicon interface
I_{DS}	Drain current
I_{DS0}	Drain current in the absence of any velocity saturation effects
J_n	Electron current density
L	Channel length
μ_0	Base mobility
N_a	Fin doping (in an n-channel device)
N_d	Source/drain doping (in an n-channel device)
n_i	Intrinsic electron carrier density
$\varphi_{fn}(x)$	Electron quasi fermi potential (= 0 at the source end)
Φ_t	Thermal voltage (kT/q)
$\Psi(x, y)$	Electrostatic potential (with respect to the φ_{fn} in the source end)
q	Electronic charge
Q_i	Inversion charge areal density
t_{ox}	Gate oxide thickness
V_{DSat}, I_{DSat}	Drain saturation voltage, current
v_{sat}	Saturation velocity
W_{fin}	Fin width (ie. distance between closest edges of the front and back gate oxides)

Chapter 1
Introduction

The holy grail of semiconductor technology has been to scale the conventional planar MOSFET to ever reducing channel lengths. This has the benefits of increased performance, increased functionality (due to the ability to squeeze in more transistors in the same die area), and reduced power consumption (on a per transistor basis). Using this scaling approach, the number of transistors on a chip has approximately doubled every 18 months, a trend known as the "Moore's Law" in the industry. The trend has set such a benchmark that industry informally sets it as a target to be met in their product plans.

Over the years, Dennard's revolutionary idea of simply scaling [1] has more or less worked as intended. But in the sub-100 nm regime, technology scaling has reached a point where any suggested work-arounds have some other problems linked to them. Reducing the channel length further severely increases short channel effects (SCE) like V_t roll-off which refers to the lowering of the threshold voltage V_t from the (long-channel) conventionally calculated value [2]. In order to keep power consumption within limits (which would also help successfully dissipate the heat from the die), the power supply voltage is also being scaled down. But that has its problems too. Lowering of supply voltage means lower gate overdrive, which means lower ON current I_{ON} and thereby slower performance, besides reduced noise

immunity. To retain the gate overdrive if one reduces V_t, then that increases sub-threshold conduction (increased OFF currents I_{OFF}). Moreover, in proceeding with the conventional scaling methodology, the gate oxide thickness has to be reduced to retain adequate gate-channel control. But that increases gate leakage. Further, doping concentrations need to be increased, which worsen the sub-threshold slope and also the substrate sensitivity. Increased doping concentrations in the body also mean greater random dopant fluctuation effects, and reduced carrier mobility. Efforts are on to have a thicker gate dielectric (but having a small effective oxide thickness EOT) in order to reduce gate leakage. This implies using a high-k gate dielectric. But this "high-k gate dielectric" search is not a simple effort. It has mostly yielded materials with poor thermal stability and/or a large number of interface traps when used with silicon. The poor thermal stability implies lower processing temperatures for subsequent steps. This in turn has meant that polysilicon gates cannot be used (lower temperature => insufficient activation => poly depletion), and therefore the efforts in trying to use a metal gate. But finding a metal with an arbitrarily tunable work function is not easy. While some high-k dielectrics such as Ta_2O_5, ZrO_2, etc have been found to have good thermal stability, they have other problems such as an undesirable band alignment with respect to silicon's bandgap, in a way that worsens the gate direct-tunneling current [3]. Finally, even if one has a good high-k gate dielectric (with a thin EOT), one ends up with unwanted effects like fringing fields that one could earlier neglect when one had a relatively thin physical oxide thickness. While there have been some commendable successes involving the conventional planar MOSFET such as the 45 nm technology recently introduced for commercial production with a physical gate length of 35 nm, a supply voltage of 1.1V, a hafnium based high-k gate dielectric with an EOT of 1 nm, and a metal gate electrode [4], it is felt to be a difficult task nevertheless to continue with the conventional planar MOSFET at future technology nodes.

As a result, various alternate device structures have been studied. The FinFET is one such emerging device which is considered to be a suitable successor to the conventional MOSFET winning-over many of the hurdles mentioned above, though it too is likely to be made using a high-k gate dielectric and a metal gate. The structure of a FinFET is shown in Fig. 1-1. It is so called because the thin channel region (body) stands vertically like the fin of a fish between the source and drain regions. The gate wraps around the body from three sides, and this is responsible for higher gate-channel control and therefore reduced SCE. In strong inversion, conduction predominantly occurs close to the sidewalls, whereas in sub-threshold it occurs along the fin center (ie. midway between the sidewalls).

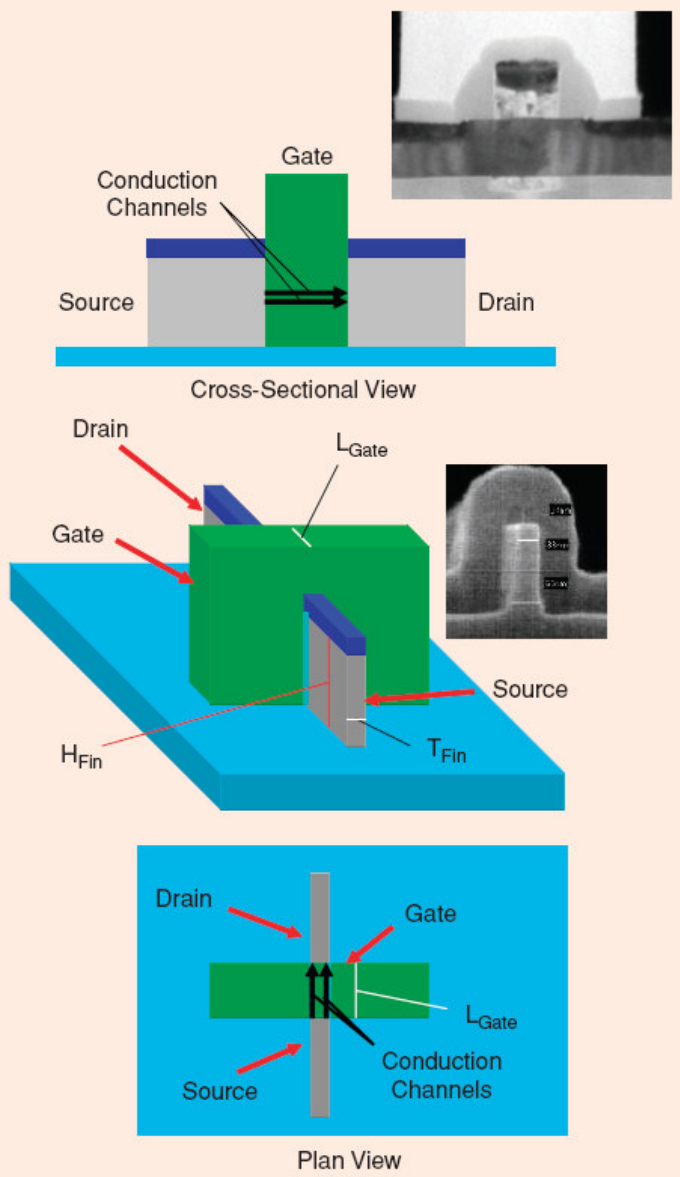


Figure 1-1: FinFET structure, with dimensions marked (from [5]).

The dimension T_{fin} shown in the figure is the *fin-width* and is alternately referred to as t_{si} or W_{fin} in the literature. We refer to it as W_{fin} in this report. H_{fin} is the height of the fin.

1.1 Problem definition

Due to its different structure, the electrostatics of a FinFET is different from those of a conventional planar MOSFET. As a result, existing compact models cannot be used *as is* to

simulate FinFET circuits. It may seem arguable that an existing compact model for a conventional planar MOSFET (such as BSIM [6] or PSP [7]) could be *forced* on a FinFET by relying on the parameter extraction steps to yield a good match. But such a strategy is risky and may not scale well (eg. may yield vastly differing extracted parameter values for slightly differing channel lengths, even when based on the same technology). A physically more correct and acceptable approach is to re-derive the model equations keeping the new structure in mind.

A reasonably structured FinFET is actually a device in which 3D effects play a non-negligible role (where by *reasonable*, we mean that the fin height is larger but not *significantly* larger than the fin width). Nevertheless, a 3D approach in compact model development is difficult. Hence it is customary to focus on the 2D aspect of the FinFET when developing a compact model. In other words, the fin height is assumed to be infinite. Thus all quantities derived in the model (such as currents, charges, etc) are on a per-unit-fin-height basis. The device being modeled is thus deemed as a double-gate MOSFET (DGFET).

1.2 Existing models

There have been many efforts to develop a core model for DGFETs. References [8,9] were based on charge-sheet-models. References [9-18] assumed a constant mobility. In [9], a perturbation technique was used to add support for body doping also. References [8,19] considered velocity saturation effects using the Caughey-Thomas model [20] or its variants, with exponent n=1. Using an exponent n=1 or any odd number in general is expected to yield a model that will not be Gummel symmetric [21]. In the PSP FinFET model [22], which uses the velocity saturation model as described in [7], the Caughey-Thomas model with exponent n=2 is used in an *over-simplified approach*. This is discussed in more detail in section 4.1.

Besides development of core models, which are usually long channel 1D models, there have also been efforts to model other effects individually that are then added to the core model. Regarding the modeling of 2D field effects, much of the existing work in the literature has focused on solving the 2D Poisson equation in the sub-threshold regime, with the intention of developing either a scaling theory [23-27] or models for threshold voltage roll-off, sub-

threshold slope and DIBL [23,24,28-31]. Some works have modeled 2D field effects in the above-threshold regime also [32,33], though their approaches are questionable. The above mentioned works on 2D field effects do not indicate how the models they develop can be incorporated into a closed-form compact model. These are discussed in more detail in chapter 7.

Some works have modeled quantum mechanical effects in DGFETs with ultra-thin fins. Reference [34] modeled inversion layer quantization effects, but was still a long channel 1D model. Reference [35] considered inversion layer quantization and also 2D field effects, but needed an iterative solution to a complex implicit equation. Reference [36] discussed quantum effects due to structural confinement, and its effect on the modified gate overdrive that is needed in order to maintain a certain level of carrier density in the fin. Reference [37] modeled structural confinement effects in strong inversion using a variational approach based on a trial eigenfunction. The UFDG model [38] is a physics/process based Poisson-Schroedinger solver that supports 2D field effects, low field mobility degradation, quasi-ballistic transport (velocity overshoot), parasitic BJT currents, gate induced drain leakage (GIDL) and other effects, and also has options to turn on quantum mechanical and self heating effects (for SOI-DGFETs). However the model evaluation needs iterative solutions, and is therefore considerably slower [38] than other (albeit empirical) models of similar capability.

1.3 Scope of the present work

In the present work, we have derived a closed-form inversion-charge based drain current model for a symmetrically driven symmetric DGFET (SDGFET), under the drift-diffusion transport formalism. We have improved over existing models for SDGFETs by redefining the boundary of the *core* model by including velocity saturation effects in a self-consistent manner as part of the *core* model development itself, in a manner that Gummel symmetry compliance is met. Futhermore, we have derived a closed form approximation for a key state variable, viz. the inversion charge areal density. Baselining this as the core model, we have then added into it support for channel length modulation, low field mobility degradation, body doping, and 2-dimensional field effects such as DIBL.

Besides incorporating 2D field effects into the core model, we have also developed models for threshold voltage roll-off, sub-threshold slope and the DIBL coefficient. Since our model is not threshold-voltage based, these models are not incorporated *as is* into our model. But they have nevertheless been developed in order to provide key insights into device design.

1.4 Organization of the report

This report is organized as follows. Chapter 2 describes some basic modeling concepts that are necessary to make the best use of the rest of this report, covering things such as an overview of FinFET operation, some compact modeling concepts and existing core models found in the literature. Chapters 3-7 describe the present work in detail. In chapter 3, we describe a non-conventional approach to solve the electrostatics of a double-gate FinFET. In chapter 4, we describe the core model formulation. Chapter 5 describes how support for body doping and low-field mobility degradation was added to the core model. Chapter 6 describes our development of an accurate closed form approximation for the inversion charge areal density. In chapter 7, we describe the incorporation of 2D field effects for ultra-short channel devices into the core model, besides presenting analytical models for threshold voltage, sub-threshold slope and DIBL. Lastly, we summarize and conclude in chapter 8, in which we also state the scope for future work.

Chapter 2
Basic Concepts

In this chapter, we present some basic background information that would help in following the rest of this report in the most useful manner. We first give a brief overview of DGFETs and FinFETs and their operation. We then cover some basic modeling concepts starting with what is meant by a compact model and then proceeding on to explain the concepts of velocity saturation, charge partitioning and Gummel symmetry. We end the chapter with an overview of the state of the art of existing core models found in the literature.

2.1 DGFETs and FinFETs

2.1.1 Double-gate MOSFETs (DGFETs)

The schematic of a generic n-channel DGFET is shown in Fig. 2-1.

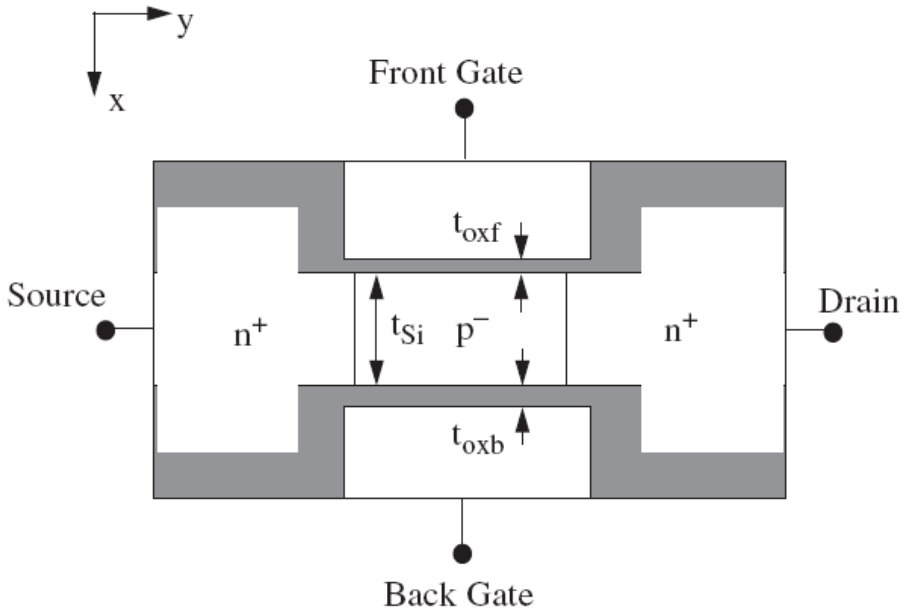


Figure 2-1: Schematic of an n-channel DGFET structure (from [39]).

The shaded region is the gate dielectric (and spacer). t_{oxf} and t_{oxb} are the thicknesses of the front and back gate dielectrics. In a DGFET the channel is under the control of two gates, viz. the front and back gates. This helps in reducing the effect of the drain field in reaching the source, and thus results in reduced SCE. A DGFET with identical material and thickness for the front and back gate electrodes and dielectric is called a symmetric DGFET (SDGFET). Conversely, a DGFET that is not a SDGFET is called an asymmetric DGFET (ADGFET).

Energy band diagrams across the channel at two gate biases for symmetric and asymmetric DGFETs are shown in Fig. 2-2 and 2-3. In a SDGFET with a mid-gap gate electrode, the bands in silicon are flat in the sub-threshold regime, implying a uniform (and low) carrier density across the fin. At high gate bias, the bands in silicon bend downwards near the sidewalls. When they are sufficiently bent so as to be near or above the quasi fermi level (which is spatially flat along the fin at V_{DS}=0), the carrier density near the sidewalls increases sharply, and the transistor is said to be strongly inverted. There are various definitions for the *exact* onset of strong inversion, and these will be discussed later in the chapter that describes threshold voltage modeling considering 2D field effects. The gate bias corresponding to the onset of strong inversion is called the threshold voltage V_t.

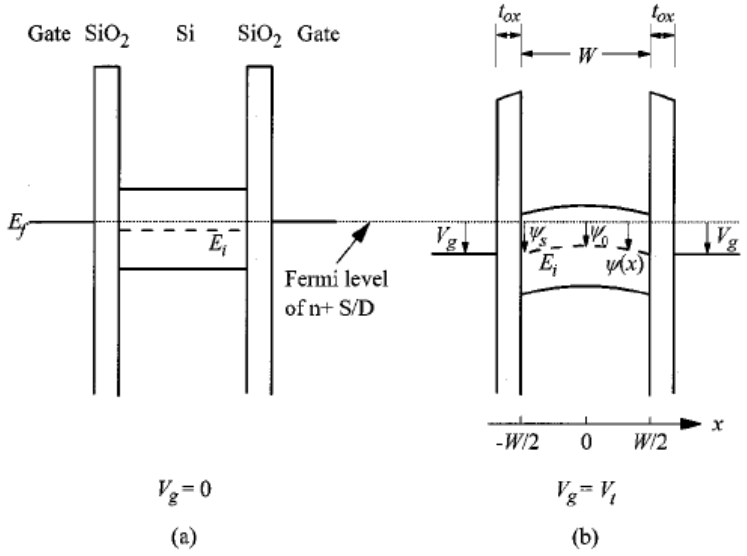


Figure 2-2: Energy band diagram for an SDGFET (from [40]).

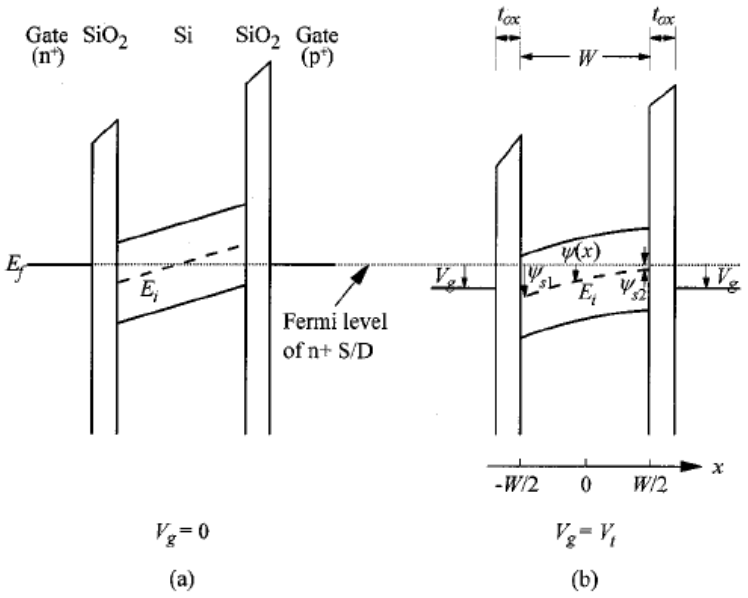


Figure 2-3: Energy band diagram for an ADGFET (from [40]).

Fig. 2-4 shows a vector plot of the typical current distribution in the intrinsic portion of a SDGFET for a 60 nm gate length device with a 20 nm fin width and 1.5 nm thick gate oxide, as seen from device simulations (with quantum effects disabled), at high gate and drain bias. As can be seen, the current is the largest near the sidewalls due to the largest inversion charge density there. In an actual device, the peak current actually occurs a little away from the sidewalls, due to quantum effects that shift the charge centroid away from the sidewalls.

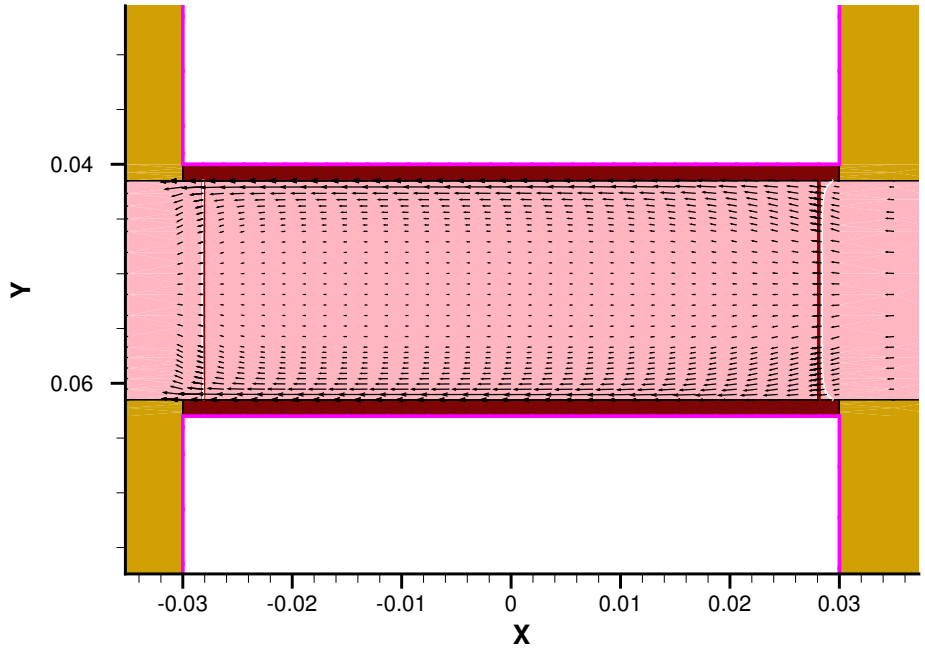


Figure 2-4: Vector plot of the current distribution in an SDGFET.

Notice the region at the edge of the intrinsic region, where it can be seen that the carriers tend to move away from the sidewalls and get uniformly distributed across the fin extension as they go into the extrinsic region.

2.1.2 FinFETs

Planar DGFETs are difficult to fabricate [5,41]. There are problems in aligning the top and bottom gates, as well as in building a low resistance contact to the bottom gate. A variant of the DGFET that can be relatively easily fabricated is the FinFET. The structure of a FinFET with a cut-plane view across the fin is shown in Fig. 2-5. To make a double-gate FinFET, the top oxide is made much thicker than the side oxides in order to effectively inactivate the top gate.

As is evident, the electrical width of a triple-gate FinFET is $W=2H_{fin}+W_{fin}$ (some papers use an alternate legend, referring to W_{fin} as t_{Si}). In many cases, W_{fin} is small in order to have acceptably small SCE. Moreover, in a DG-FinFET, the top gate is anyway ineffective. As a result, W is approximately $2H_{fin}$. As a result, the physics of a FinFET becomes largely similar to that of a DGFET. Thus, most of the literature that discuss compact model development for

DGFETs, can be applied to FinFETs with a minor parameter (H_{fin}) adjustment, as done in [22].

The long narrow portion of the fin that is not under the gate, is called the extension region. This is a region that is technologically unavoidable, because it is not possible to have a steep lateral doping gradient, starting from a highly doped source/drain, and ending with a lightly doped channel region (a lightly doped body is preferred because it helps reduce corner effects [42-44], random dopant fluctuations and mobility degradation effects). As a result, FinFETs typically have a relatively large parasitic series resistance.

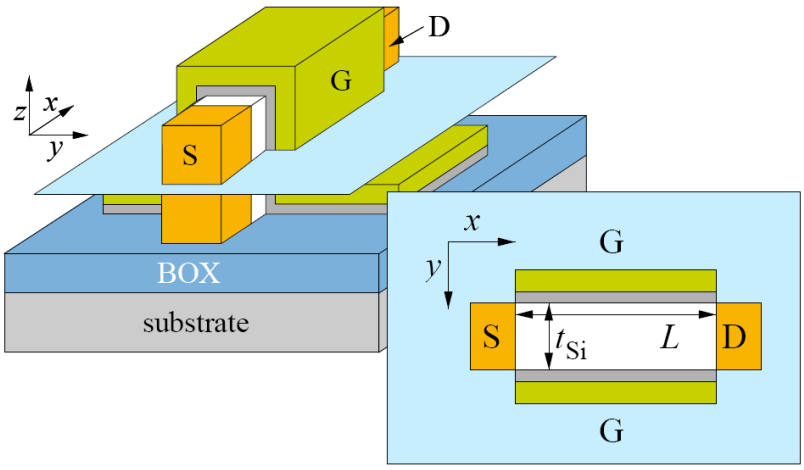


Figure 2-5: 3D view of a FinFET with a cut-plane across the fin (from the presentation slides of [22]).

2.2 Device models versus compact models

In their simplest form, device simulators self-consistently solve the Poisson equation, the continuity equations and the drift-diffusion equations. The models used by a *device* simulator are called *device models*. The Poisson equation relates the space derivatives of potential to the total charge density (mobile as well as fixed). The continuity equation is a statement of conservation of particles, and relates the time rate of change of carrier density to the current density and recombination-generation rates. The drift-diffusion equations relate the current density to the space derivatives of the potential (drift) and the carrier density (diffusion). Given adequate boundary conditions, device simulators solve these equations numerically.

They take into account low-field mobility degradation, high field effects like velocity saturation, impact ionization and many other effects by means of device models. Under certain conditions such as strong inversion in extremely scaled devices, the inclusion of the Schrödinger equation's self-consistent solution can also be turned on for more accurate prediction.

Solving these complex equations for one transistor alone takes a long time. Solving them for a complex circuit involving multiple transistors is simply impractical (in a device simulator). Hence there is a need to come up with *compact* models that can be implemented in a *circuit* simulator. A compact model for a device is a model that treats the device as a black box, and has simple, closed form expressions for the various external variables that are needed to simulate a circuit (involving many instances of that device). These are variables such as terminal currents, voltages, capacitances, charge, etc, and their inter-dependencies (eg. variation of capacitance with bias). If these expressions are simple and accurate enough, then the model is a good one. The simplicity comes at the cost of accuracy, and one of the challenges is to make the right trade-off in this regard [45]. Compact models have parameters, called model parameters. Physics-based compact models have fewer parameters than empirical compact models, and are therefore desirable. The specification of a compact model also states the means for extracting the parameters. Besides simplicity and accuracy, another figure of merit of a compact model is the ease and accuracy of extraction of the model parameters.

Whereas the conventional planar MOSFET has compact models built into commercial circuit simulators, that is not the case for FinFETs. Some of the common compact models for conventional MOSFETs are BSIM [6], PSP [7], HiSIM [46], EKV [47], etc. Recently the PSP group announced a compact model for FinFETs [22] that leverages the PSP model for conventional planar MOSFETs.

2.3 Velocity saturation

The velocity-field relationship is given [48,49] by $v = \mu E \sqrt{T_L / T_e}$, where T_L is the lattice temperature and T_e is the carrier temperature. At low electric fields, there is an approximate

thermal equilibrium with the carriers absorbing phonons from the lattice at nearly the same rate at which they emit to it. Hence $T_e \approx T_L$ and the velocity-field relationship is approximately linear with $v \approx \mu E$, implying a constant, field-independent mobility. As the electric field increases, the average energy of the carriers increases. Some of this energy is imparted to the lattice by optical phonon emission, at a rate higher than what it absorbs from the lattice. As a result, the lattice temperature T_L also increases. But the carrier temperature T_e nevertheless increases faster than T_L [48]. At sufficiently high fields, the most frequent scattering event is the emission of optical phonons, and the drift velocity saturates at a value called the *saturation velocity* v_{sat}. For silicon, v_{sat} is about $0.6\text{x}10^7$ to $1\text{x}10^7$ cm/s for electrons and about $0.4\text{x}10^7$ to $0.8\text{x}10^7$ cm/s for holes [50]. Velocity saturation causes the drain current to saturate at a lower drain voltage than that described by constant mobility models where current saturation is found to occur due to channel pinch-off. Indeed, for short channel devices, the carriers at the drain end saturate before the onset of channel pinch-off [7].

By considerations involving the balance between the rate of energy absorbed from the electric field and the net rate of energy lost to the lattice, a field-dependent mobility can be derived. The Caughey-Thomas model [20] is a semi-empirical model for expressing this, and is stated in (1).

$$\mu_{eff}(x) = \frac{\mu_0}{\left[1 + \left(\frac{\mu_0 |E_x|}{v_{sat}}\right)^n\right]^{1/n}} \tag{1}$$

The term E_x is commonly referred to as the driving electric field for velocity saturation [51]. Notice the presence of the *absolute-value* operator around E_x. This is done to make it applicable in that same form for normal as well as reverse mode operation. The exponent n is equal to 2 for electrons and 1 for holes [2].

2.4 Charge partitioning

For a compact model to be suitable for doing *transient* analysis, a drain current DC model is not enough. One also needs models for source and drain *terminal charges*. The source

terminal charge Q_S is defined as that quantity whose time derivative represents the *charging* current at the source end, and the drain terminal charge Q_D is defined as that quantity whose time derivative represents the *charging* current at the drain end. Specifically, Q_S and Q_D are defined as those quantities that when used in (2) and (3), yield a value for $I_S(t)$ and $I_D(t)$ that represent the *total instantaneous* current at the source and drain terminals respectively [52]. $I_0(t)$ is called the *transport* current, and represents the instantaneous current component that is the same at the source and drain ends at a given bias condition (the charging currents being the differential that needs to be added/subtracted [53] to/from it in order to make the total instantaneous current represent the exact instantaneous current at the respective terminal ends (which could be different at the source and drain ends).

$$I_S(t) = I_0(t) - \frac{dQ_S}{dt} \tag{2}$$

$$I_D(t) = I_0(t) + \frac{dQ_D}{dt} \tag{3}$$

For such a definition to hold, it was shown by Oh et al. in [52] that Q_S and Q_D must be given by (4) and (5) respectively, where $Q_i(y)$ is the inversion charge areal density at a distance y from the origin, L is the channel length, W is the width of the device (for a planar MOSFET), and the y-axis is deemed along the channel with the origin at the source end.

$$Q_S = -\frac{W}{L}\int_0^L (L-y)Q_i(y)dy \tag{4}$$

$$Q_D = -\frac{W}{L}\int_0^L yQ_i(y)dy \tag{5}$$

Equations (2) - (5) are derived [52] using a non-quasi-static analysis starting from the continuity equation using the drift-diffusion transport model, under conditions of constant mobility, and are referred to as the *Ward-Dutton charge partitioning scheme* in the literature. The point where quasi-static assumptions are usually introduced are hereafter, viz. in the modeling of $Q_i(y)$, in order to evaluate (4) and (5).

Equations (4) and (5) can very easily be adapted to a DGFET if W is set to 1 and the resulting terminal charges (and the currents) are interpreted as values per unit fin height.

2.5 Gummel symmetry

A model is said to be Gummel-symmetric [21] in currents if the first and second derivatives of I_{DS} with respect to V_{DS} are differentiable and continuous (with a value of 0) at V_{DS}=0, respectively.

Model symmetry around the V_{DS}=0 operating point is essential for simulating inter-modulation in RF circuits like passive mixers. It is well known [54,55] that asymmetry around V_{DS}=0 produces incorrect frequency dependence for the third harmonic (which is related to the 3rd derivative of I_{DS} with respect to V_{DS}) and amplitude. The third harmonic is the first major harmonic that one usually needs to worry about in circuits using balanced topologies, since even-order harmonics are reduced by about 40dB [56]. Other circuit simulations that are problematic from this standpoint are those of R-2R ladder circuits, and in general any application where the transistor is used as a gate-controlled resistor [56]. Hence model symmetry is very important especially for simulating analog circuits.

Models are usually [54] made symmetric by interchanging the source and drain internally for V_{DS}<0. However, if the model is inherently asymmetric (for example if it uses source-referenced threshold voltage), then such enforced "symmetrization" results in a singular behavior (ie. 2nd derivative d^2I_{DS}/dV_{DS}^2 doesn't exist at V_{DS}=0) regardless of the nature of the mobility or velocity saturation models. Indeed, this is a problem in BSIM3v3, and it was attempted to be fixed in BSIM4 by introducing a new effective V_{DS} through purely numerical means as $V_{DS,new} = \sqrt{\left(V_{DS}^2 + \delta^2\right) - \delta^2 \Big/ \sqrt{V_{DS}^2 + \delta^2}}$ where δ is sufficiently small (a model parameter), but this idea was later dropped due to some numerical issues [45]. Besides, it has been suggested [45] that such an approach would anyway not be able to solve the asymmetry in some of the capacitances at the V_{DS}=0 point.

On the other hand, for a more physical model, interchanging source and drain for V_{DS}<0 would yield the correct result provided that the underlying velocity saturation model used is

nonsingular. The root of this assertion goes back to the implicit use of the *ABS()* function around a term containing V_{DS} (on which the driving electric field E_x depends), in some velocity-field relationships (which form the basis of velocity saturation models). This has been one of the driving forces behind our choice of the velocity-field relationship in the present work.

A circuit that is used to test for Gummel symmetry is shown in Fig. 2-6.

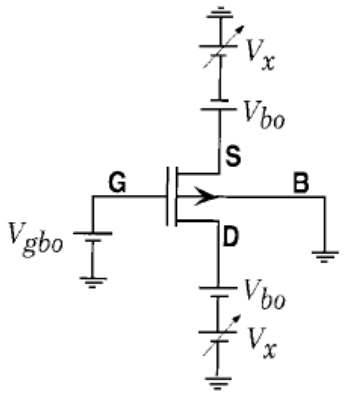


Figure 2-6: Circuit used to test for Gummel symmetry (from [21]).

In Fig. 2-6, V_{DS} always equal to $2V_x$. If V_x is swept from a negative value to an equal positive value, the transistor is driven symmetrically from reverse mode to forward mode.

2.6 Some existing core models

A good review of existing core models for undoped DGFETs is given in [57]. We discuss below some of the existing drain current core models in the literature and how they relate to our present work. Our discussion includes some works not covered in [57].

In the discussion below, the term *1D model* refers to a model developed under the premise of the gradual channel approximation (GCA), and is therefore a long channel model. Also, the term *core model* refers to the model that is developed consistently from fundamental principles, to distinguish it from the enlarged model that encompasses second-order effects

(such as 2D field effects, junction leakage, etc) that are usually added later on to the core model.

In [8], Pei et al. presented a core model that was a 1D piece-wise threshold voltage based closed-form model developed using constant mobility for an independently driven ADGFET. The strong inversion piece was developed using a simplified charge sheet model. The pieces were joined using smoothing functions. Physical effects added onto the core model were DIBL and source/drain series resistance, both of which were modeled empirically. Low-field mobility, velocity saturation and channel length modulation (CLM) were also added to the core model. Velocity saturation was modeled using a variant of the Caughey-Thomas model with exponent n=1 and using an averaged driving field V_{DS}/L.

In [9], Dunga et al. presented a core model that was a 1D closed-form model for the above-threshold regime developed using constant mobility for a symmetrically driven SDGFET, and was based on a charge sheet approximation. The core model included support for body doping developed using a perturbation technique. This actually forms the basis of the BSIM-MG model [58].

Reference [10] was one of the pioneering works by Prof Taur's group where the core model for a symmetrically driven SDGFET was very elegantly developed. It was a 1D single-piece model developed using constant mobility, and needed the solution of an implicit equation. All the 1D models developed after this were more or less similar in approach to that developed in [10].

In [11], He et al. presented a core model that was a 1D single-piece inversion-charge based model developed using constant mobility for a symmetrically driven SDGFET. It relied on the solution of the Lambert-W function [59] for calculation of the inversion charge, but did not present any closed form solution of the Lambert-W function.

In [12], the EKV group presented a core 1D single-piece inversion-charge based model developed using constant mobility for a symmetrically driven SDGFET. It differed from other similar approaches in that it made some approximations in calculating the inversion charge areal density, although it was still an implicit equation finally.

In [13], the EKV group added to their earlier work [12] by adding support for asymmetric DGFETs. They presented a core 1D single-piece inversion-charge based semi-empirical model for an ADGFET, by introducing the concept of common-mode and difference-mode gate voltages. The single-piece model was developed by joining three regimes using a smoothing function.

In [14], Prof. Taur's group enhanced their earlier work [10] by presenting a core 1D single-piece model developed using constant mobility for SDGFETs as well as ADGFETs. Expressions for terminal charges using the Ward-Dutton charge partitioning scheme were also presented, though they involved complex unevaluated integrals, and it was suggested that for a simulator implementation, the integrals could be pre-calculated in the expected range and be stored in memory before-hand, to be used in a table-lookup manner.

In [15], Ortiz-Conde et al. presented a core 1D single-piece surface potential based model for a symmetrically driven SDGFET that was developed using constant mobility. It relied on an evaluation of the surface potential using an expression developed earlier in [60], which was claimed to be a closed form expression, but in reality it was expressed in terms of the Lambert-W function [59], for which closed form expressions were not presented. Besides, [60] included some empirical parameters.

In [16], He et al. presented a core 1D single-piece inversion charge based model for a symmetrically driven SDGFET that was developed using constant mobility. It was very similar to their earlier work in [11] except that an intermediate implicit equation was retained as an implicit equation, unlike in [11] where they had transformed it approximately into a form whose solution can be expressed in terms of the Lambert-W function. Thus, the final model in [16] continued to need the solution of an implicit equation. With suitable variable substitutions, their core model was claimed to be essentially the same as that presented in [10] and [15].

In [17], Zhu et al. presented a core 1D single-piece surface potential based model for a symmetrically driven SDGFET that was developed using constant mobility. With suitable variable substitutions, it can be easily seen that it is essentially the same as the core model presented in [10]. It relied on the surface potential expressions presented in [18] (for

ADGFETs in general) which were claimed to be explicit, but in reality they were expressed in terms of the Lambert-W function [59], for which closed form expressions were not presented.

In [19], Wong and Shi presented a core 1D single-piece surface vertical electric field based model (or equivalently put, inversion charge based model) for a symmetrically driven SDGFET. The core model included *within* its boundary the physical effects of low-field mobility degradation and velocity saturation, although results were presented with one effect turned on at a time for the sake of simplicity. The model needed the solution of an implicit equation for the surface electric field at the source and drain ends. The velocity saturation model used was the Sodini-Ko-Moll model [61], which is a piece-wise variant of the Caughey-Thomas model with exponent n=1 (the variant differing in the way the critical electric field relates to v_{sat}), and therefore violates the Gummel symmetry condition in the final model, which has been verified by us.

The PSP FinFET model [22] presents a core 1D single-piece surface potential based model developed using constant mobility for a symmetrically driven SDGFET. Many other physical effects are then added on to the core model, such as low-field mobility degradation, velocity saturation, DIBL, etc all of which are based on the respective models for the conventional planar MOSFET as described in [7]. Thus, velocity saturation is modeled by replacing the constant mobility in the core model with an effective mobility that considers a *spatially averaged* lateral electric field as the driving field for velocity saturation.

To summarize, a vast majority of the core models in the literature have been developed using constant mobility. The only core model [19] for an SDGFET that considered velocity saturation within the boundary of the core model, was not Gummel-symmetry compliant. It is thus clear from this review that there exists a need for a *core* model that goes beyond constant mobility considerations and in a manner that Gummel-symmetry is preserved. The greater the number of physical effects considered in the core model (as opposed to adding them as second order effects onto a *constant mobility core* model), the greater is the physical accuracy (and complexity) of the model. The present work is motivated by such considerations, and is described in the subsequent chapters.

Chapter 3
Non-conventional Approach Using Functional Forms

As stated earlier in this report, there have been many works that do a 2D Poisson analysis in the sub-threshold regime, but none that correctly does a 2D analysis in strong inversion. In this chapter, we present an attempt made at doing a 2D analysis in the strong inversion region of operation (of a N-channel DGFET). The general approach was to assume a functional form for some quantity where it was felt acceptable to do so, and proceed with the analysis. We show how the Poisson, drift-diffusion and continuity equations were attempted to be solved analytically using the approximate functional forms.

The schematic of the SDGFET under consideration is shown in Fig. 3-1. All vector quantities are deemed positive if directed along the positive coordinate axes. *S* and *D* are the source and drain regions. G_f and G_b are the front and back gate electrodes, and are assumed to be driven symmetrically (ie. $V_{gf} = V_{gb}$). The source-body and drain-body junctions are assumed to be abrupt.

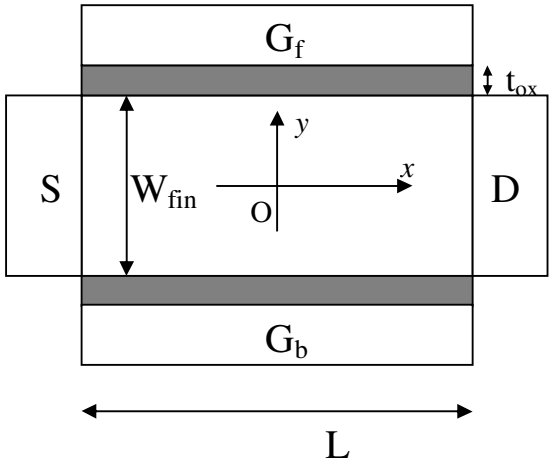


Figure 3-1: Schematic of a SDGFET, showing the coordinate axes and the dimensions labelled.

The structure, dimensions and coordinate axes shown in Fig. 3-1 are the ones considered in the rest of this report.

3.1 Analytical formulation

Neglecting recombination-generation, considering steady state conditions and focussing on electron currents alone, the 1D continuity equation is:

$$\nabla_x . J_{nx} = 0 \tag{6}$$

But

$$J_{nx} = -nq\mu_n \frac{\partial \phi_{fn}}{\partial x} \tag{7}$$

where

$$n = n_i e^{(\psi - \phi_{fn})/\phi_t} \tag{8}$$

and $\phi_t = kT/q$ is the thermal voltage, $E_{fn} = -q\phi_{fn}$ is the electron quasi fermi energy level, and $E_i = -q\psi$ is the intrinsic fermi level. The reference for ψ and ϕ_{fn} is the spatially flat ϕ_{fp} line (which is flat because hole currents are neglected). Thus deep in the source region:

$\phi_{fn} = 0$ (since $\phi_{fn} = \phi_{fp}$ there, due to zero inequilibrium), and

$\psi = \phi_t ln(N_d / n_i)$

Let the boundary conditions be given by:

$$\begin{aligned} &\psi(-L/2,0) = u_1 \\ &\psi(L/2,0) = u_1 + \Delta \\ &\phi_{fn}(-L/2) = p_1 \\ &\phi_{fn}(L/2) = p_1 + \Delta \end{aligned} \tag{9}$$

where $\Delta = V_{DS}$ - <*drop across extension region*> is the effective voltage across the active fin (where by *active* we mean the portion of the fin that is under the gate). The potential summation rule traversing from the gate electrode to the source can be written as:

$$V_{GS} - \Delta\phi = -\frac{\varepsilon E_{ys}(x)}{C_{ox}} + \psi(x, W_{fin}/2) \tag{10}$$

where $\Delta\varphi$ is the work function difference between the gate electrode and intrinsic silicon, and $E_{ys}(x)$ is the vertical electric field at the oxide-silicon interface. From (7) and (8) we get:

$$J_{nx} = -n_i q \mu_n \frac{\partial \phi_{fn}}{\partial x} e^{(\psi - \phi_{fn})/\phi_t} \tag{11}$$

From (6) and (11) we get

$$\frac{\partial \psi}{\partial x} - \frac{\partial \phi_{fn}}{\partial x} + \frac{\phi_t \dfrac{\partial^2 \phi_{fn}}{\partial x^2}}{\dfrac{\partial \phi_{fn}}{\partial x}} + \frac{\phi_t}{\mu_n}\frac{\partial \mu_n}{\partial x} = 0 \tag{12}$$

Based on observations of the spatial variation of ϕ_{fn} along the fin up to 3rd order derivatives, a functional form for it was assumed as:

$$\phi_{fn}(x) = \frac{\Delta}{2} + p_1 - ae^{-bc}cosh(bL/2) + \frac{x}{L}(\Delta - 2ae^{-bc}sinh(bL/2)) + ae^{b(x-c)} \tag{13}$$

where u_1, p_1, Δ, a, b and c are spatially independent, but bias-dependent constants. Basically this is a linear and exponential relation in x, with coefficients chosen so as to exactly match the boundary conditions (9). This fits very well with the φ_{fn} variation obtained from device simulations.

For carrier mobility, the Caughey-Thomas model for velocity saturation and the Lombardi model for low-field mobility degradation models the mobility as [51]:

$$\mu_n(x,y) = \frac{\mu_0}{\left[1 + \frac{\mu_0^2 E_x^2}{v_{sat}^2}\right]^{1/2} \left[1 + \mu_0 m E_y^2\right]} \tag{14}$$

where E_x and E_y are the x and y components of the electric fields, and m has a default value of about $2\text{x}10^{-15}$ s/V [51]. In [51], there was also a term $e^{-(y-W_{fin}/2)/l_{crit}}$ alongside m in (14) in order to model the damping of the degradation far from the interface, but we have not considered this for the sake of simplicity, considering the fact that far from the interface, the reduction of E_y would be sufficient to model this damping.

Next, the mobility model of (14) was simplified to:

$$\mu_n(x,y) = \frac{\mu_0 e^{a_0 - \left(a_1 x + a_2 x^2\right)}}{\left[1 + \mu_0 m E_y^2\right]} \tag{15}$$

The justification for this form of the expression is that exponentials are very convenient to handle. Besides, we were successfully able to fit (15) to corresponding device simulated mobilities along the fin center, in strong inversion at low as well as high V_{DS}. This is shown in Fig. 3-2.

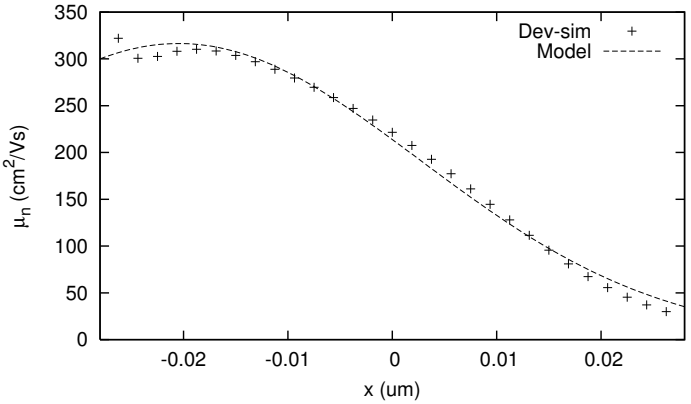


Figure 3-2: Goodness of fit for the mobility spatial functional form along the fin center, at V_{GS}=V_{DS}=1V.

Note that the constants a_0, a_1 and a_2 are all bias dependent. To use (15), we need an expression for the vertical electric field E_y. Using 1D electrostatics (and as will be shown in detail in the next chapter), one can easily show that:

$$\psi(x, y) = A_1(x) + 2\phi_t \ln\left(\sec(\beta y e^{\alpha(x)})\right) \tag{16}$$

where

$$\beta = \left[\frac{qn_i}{2\varepsilon\phi_t}\right]^{1/2} \tag{17}$$

$$\alpha(x) = \frac{A_1(x) - \phi_{fn}(x)}{2\phi_t} \tag{18}$$

$$E_y = -2\beta\phi_t e^{\alpha} tan(\beta y e^{\alpha}) \tag{19}$$

and $A_1(x) = \psi(x, 0)$ is the electrostatic potential along the fin center.

The 1D solution (16) is used only for computing E_y in the mobility relation. From (15), we have:

$$\frac{1}{\mu_n}\frac{\partial\mu_n}{\partial x} = -(a_1 + 2a_2 x) - \frac{2\mu_0 m E_y \partial E_y / \partial x}{1 + \mu_0 m E_y^2} \tag{20}$$

Based on the value of *m* as stated below (14), we can neglect the denominator in (20) and this approximates to:

$$\frac{1}{\mu_n}\frac{\partial \mu_n}{\partial x} = -\left[(a_1 + 2a_2 x) + 2\mu_0 m E_y \partial E_y / \partial x\right] \tag{21}$$

Per numerical simulations, it was found that the third term in the RHS of (21) is negative (since $|E_y|$ reduces with x), but is smaller in magnitude than the first two terms on the RHS put together (which are positive per curve fit extractions), and hence the overall expression for $\frac{1}{\mu_n}\frac{\partial \mu_n}{\partial x}$ is negative, which is in agreement with physical observations (see Fig. 3-2). From (19) we have:

$$\frac{\partial |E_y|}{\partial x} = 2\beta\phi_t e^{\alpha}[\beta y e^{\alpha} sec^2(\beta y e^{\alpha}) + tan(\beta y e^{\alpha})]\frac{\partial \alpha}{\partial x} \tag{22}$$

Substituting for $\partial\alpha / \partial x$ in (22) using (18) and (13), equation (22) becomes:

$$\frac{\partial |E_y|}{\partial x} = \beta e^{\alpha}\left[\beta y e^{\alpha} sec^2(\beta y e^{\alpha}) + tan(\beta y e^{\alpha})\right]\left[\frac{dA_1}{dx} - \frac{\Delta}{L} + \frac{2ae^{-bc}\sinh(bL/2)}{L} - abe^{b(x-c)}\right] \tag{23}$$

Equations (19), (21) and (23) give

$$\frac{1}{\mu_n}\frac{\partial \mu_n}{\partial x} = -\left[\begin{array}{l} a_1 + 2a_2 x + \\ 4\mu_0 m\beta^2\phi_t e^{2\alpha} tan(\beta y e^{\alpha})\left[\begin{array}{l}\beta y e^{\alpha} sec^2(\beta y e^{\alpha}) + \\ tan(\beta y e^{\alpha})\end{array}\right]\end{array}\right]\left[\frac{dA_1}{dx} - \frac{\Delta}{L} + \frac{2ae^{-bc}\sinh(bL/2)}{L} - abe^{b(x-c)}\right] \tag{24}$$

Equation (13) gives

$$\frac{\partial^2 \phi_{fn}}{\partial x^2} = ab^2 e^{b(x-c)} \tag{25}$$

Substituting (13), (24) and (25) in the continuity equation (12), we have:

$$\frac{\partial\psi(x,y)}{\partial x}=\frac{\Delta-2ae^{-bc}\sinh(bL/2)}{L}+abe^{b(x-c)}-\frac{\phi_t ab^2e^{b(x-c)}}{\frac{\Delta-2ae^{-bc}\sinh(bL/2)}{L}+abe^{b(x-c)}}+\phi_t\left(a_1+2a_2x\right)+4\mu_0m\beta^2\phi_t^2e^{2\alpha}tan(\beta ye^{\alpha})\begin{bmatrix}\beta ye^{\alpha}sec^2(\beta ye^{\alpha})+\\tan(\beta ye^{\alpha})\end{bmatrix}\left[\frac{dA_1}{dx}-\frac{\Delta}{L}+\frac{2ae^{-bc}\sinh(bL/2)}{L}-abe^{b(x-c)}\right] \tag{26}$$

From (26), dA_1/dx can be simply derived by evaluating $\partial\psi/\partial x$ at y=0. After finding dA_1/dx, we can easily integrate it to get:

$$A_1(x)=u_1+\frac{x+L/2}{L}\left[\Delta-2ae^{-bc}\sinh(bL/2)\right]+ae^{-bc}\left(e^{bx}-e^{-bL/2}\right)+\phi_t a_1\left(x+L/2\right)+\phi_t a_2\left(x^2-L^2/4\right)-\phi_t\ln\left[\frac{\frac{\Delta-2ae^{-bc}\sinh(bL/2)}{L}+abe^{b(x-c)}}{\frac{\Delta-2ae^{-bc}\sinh(bL/2)}{L}+abe^{b(-\frac{L}{2}-c)}}\right] \tag{27}$$

From (26) and (27), we get:

$$\frac{\partial\psi(x,y)}{\partial x}=\frac{\partial A_1(x)}{\partial x}+4\mu_0m\beta^2\phi_t^2e^{2\alpha}tan(\beta ye^{\alpha})\begin{bmatrix}\beta ye^{\alpha}sec^2(\beta ye^{\alpha})+\\tan(\beta ye^{\alpha})\end{bmatrix}\left[\frac{dA_1}{dx}-\frac{\Delta}{L}+\frac{2ae^{-bc}\sinh(bL/2)}{L}-abe^{b(x-c)}\right] \tag{28}$$

In theory, (28) can be integrated to get a first-order solution for the electrostatic potential, which can then be substituted into the 2D Poisson on the LHS, and the second-order potential solution $\psi(x,y)$ can then be derived from the RHS, which would be in terms of the seven bias-dependent constants a, b, c, a_0, a_1, a_2 and p_1. Boundary conditions (10) can then be used at

multiple values of x to get bias-dependent closed form expressions for the seven constants. Thus, six x-locations would need to be considered (viz. one less than seven, because 2 relations can be got from the ψ-related boundary conditions in (9), but one more constant would get introduced when integrating (28)). Constants u_1 and Δ are deemed external to the intrinsic device, and would come from the circuit simulator (taking pseudo nodes, and modeling parasitic resistance, for instance).

However, integrating (28) is not easy. Besides, other than the electron quasi fermi potential and the electrostatic potential along the fin center, the potentials elsewhere do not match very well with device simulated values, as is shown in the next section.

3.2 Comparison with device simulations

Device simulations were done using Synopsys Sentaurus Device [51], to compare the potential and electric field variations with the model predicted values. The 2D simulated structure is shown in Fig. 3-3 and zoomed into in Fig. 3-4.

The source-body and the drain-body junctions were abrupt. The default low-field (Lombardi) and high-field (Canali) mobility degradation models were turned on. Recombination-generation and quantum mechanical models were turned off. Table 3-1 lists the device dimensions and other details. The dimensions chosen were representative of realistic devices obtained from IMEC, Belgium. The doping profiles however were idealized.

Constants a, b, c, a_0, a_1, and a_2 were extracted from their introducing-equations (eg. (15) is the introducing equation for constants a_0, a_1, and a_2) by fitting to corresponding device simulated data. The extracted values were obtained for the V_{DS}=V_{GS}=1V case as:

a = 1.52 b = 122.4 c = 0.04

and

a_0 = -1.9 a_1 = 38.3 a_2 = 936

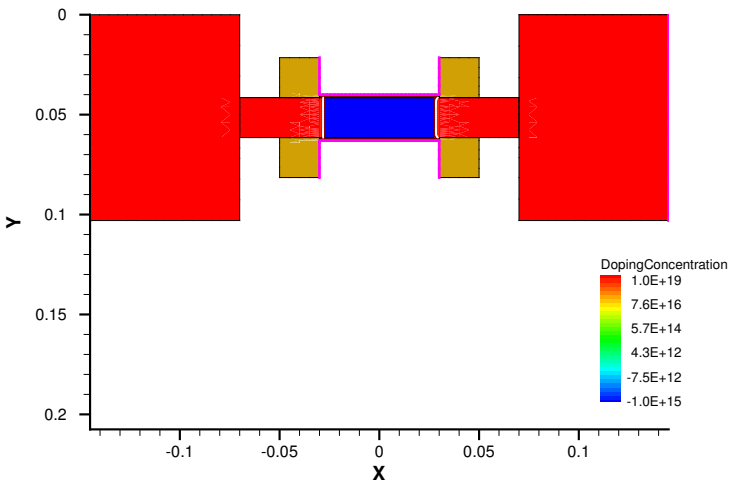


Figure 3-3: Simulated DGFET structure.

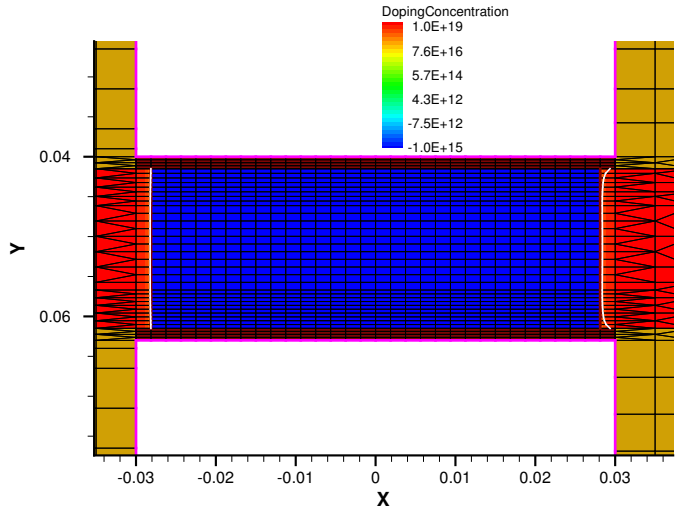


Figure 3-4: Simulated DGFET structure (zoomed into the fin region).

Table 3-1: Dimensions/profiles of the simulated structure.

Dimensions/Profiles	**Description**
L=60 nm	Gate length
L_{sde}=40 nm	S/D extension
L_{sd}=75 nm	S/D length

W_{sd}=103 nm	S/D width
W_{fin}=20 nm	Fin width
t_{ox}=1.5 nm (SiO_2)	Oxide thickness
N_a=10^{15} cm^{-3}	Fin doping (p-type)
N_d=10^{19} cm^{-3}	Source/drain doping (n-type)
Φ_m=4.6eV	Gate work function

For this bias condition, the other constants were measured from numerical simulations to be:

p_1 = 0.164V u_1 = 0.64V Δ = 0.488V

Using these constants, the φ_{fn} fitting curves are shown in Fig. 3-5, 3-6 and 3-7 for up to 2nd order derivatives. The matching curves for the predicted values of $A_1(x)$ and its x-derivative are shown in Fig. 3-8 and 3-9 respectively. As can be seen, the match is very good. However, the accuracy of the predicted spatial variation of the lateral fields (given by (28)) across the fin is quite poor, as can be seen in Fig. 3-10, 3-11 and 3-12 (at three different points along the channel).

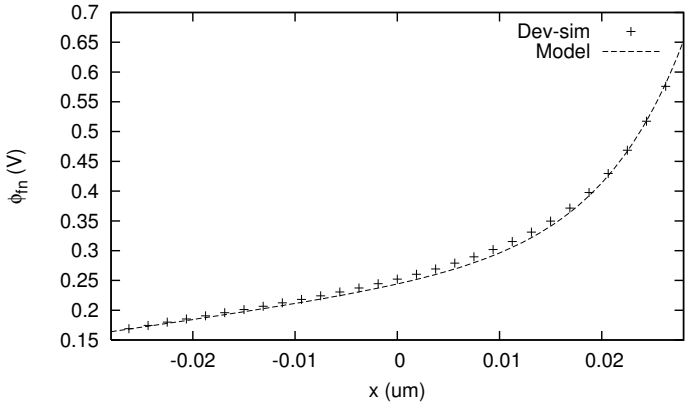


Figure 3-5: φ_{fn} along the channel.

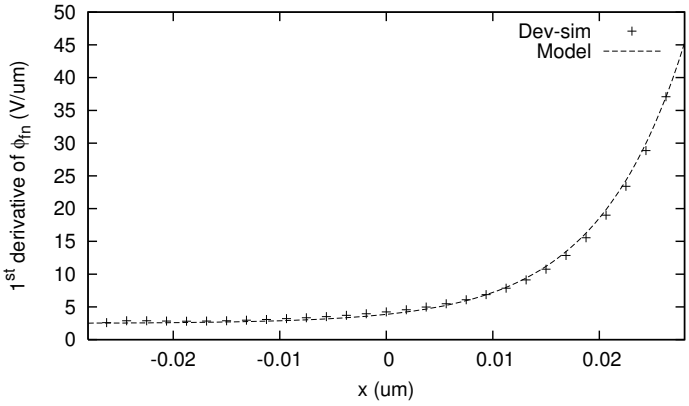


Figure 3-6: First derivative of φ_{fn} along the channel.

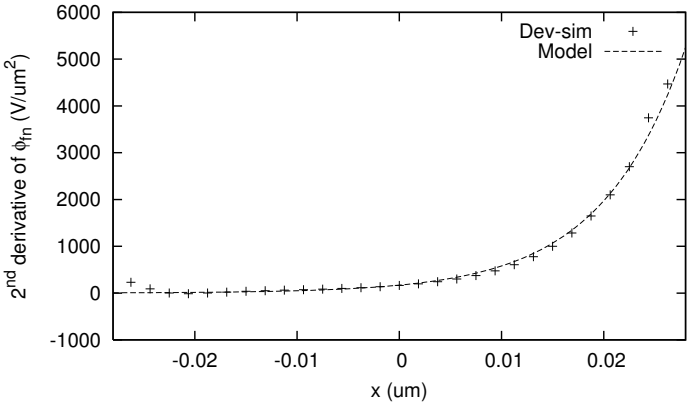


Figure 3-7: Second derivative of φ_{fn} along the channel.

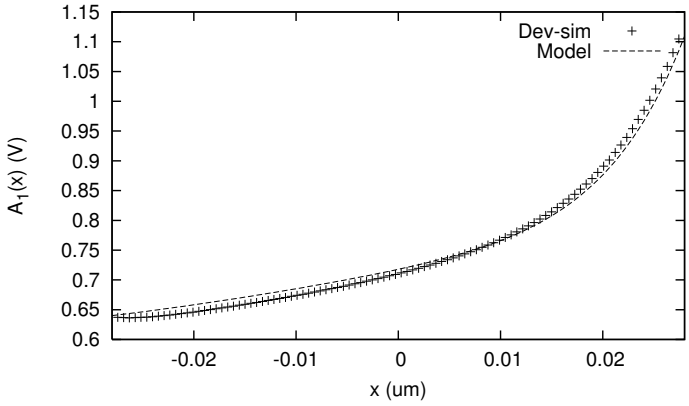


Figure 3-8: $A_1(x)$ along the channel.

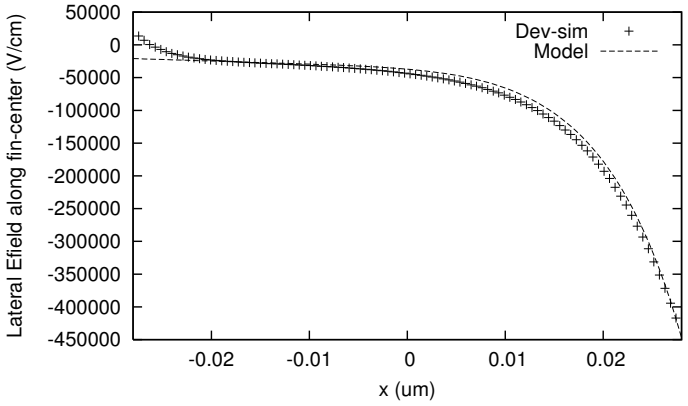


Figure 3-9: Lateral electric field along the channel midway between the 2 gates.

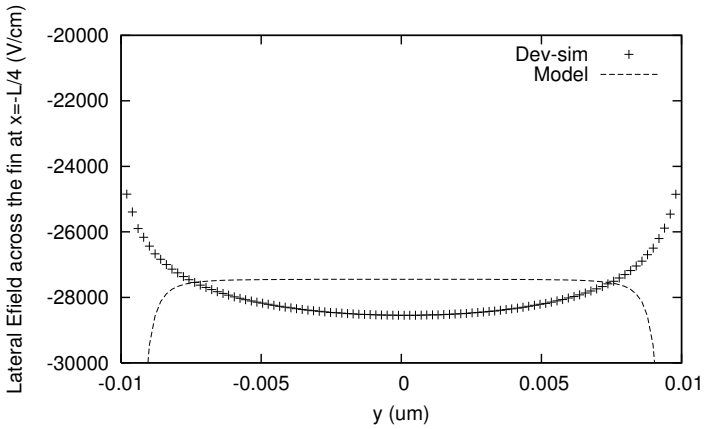


Figure 3-10: E_x across the fin at $x = -L/4$.

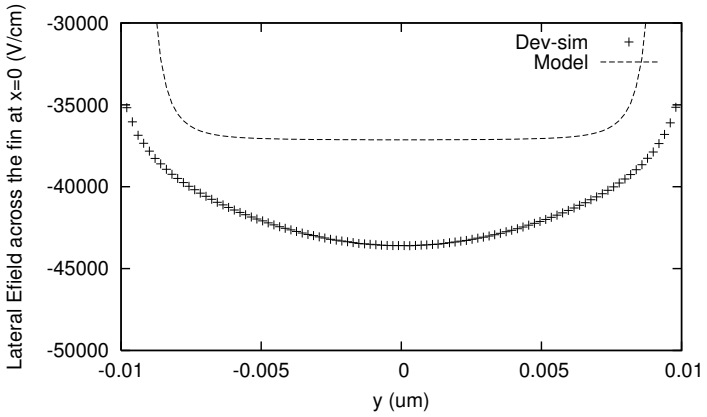


Figure 3-11: E_x across the fin at $x = 0$.

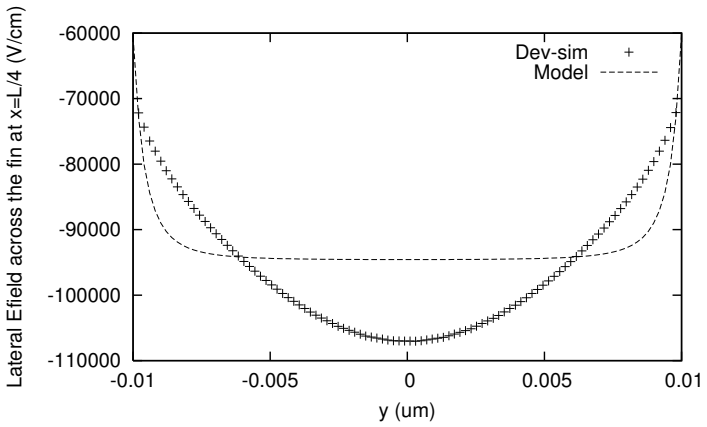


Figure 3-12: E_x across the fin at $x = L/4$.

The accuracy of the predicted spatial variation of the vertical field (given by (19) at the silicon-oxide interface along the channel is also poor, as can be seen in Fig. 3-13.

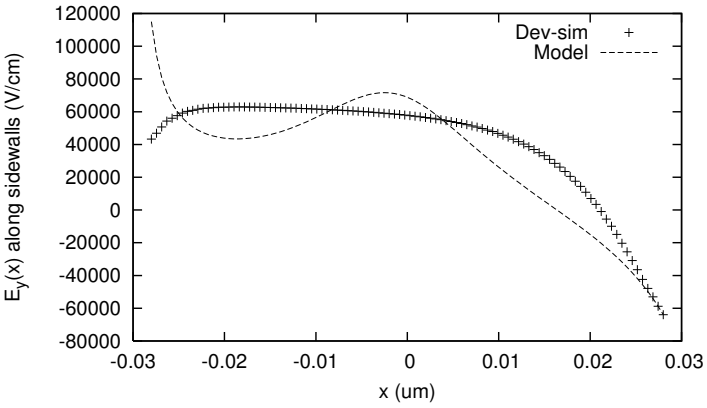


Figure 3-13: E_y as given by (19) from source to drain along the sidewalls ($y = W_{fin}/2$).

3.3 Summary

In this chapter, we presented a non-conventional 2D analysis of an SDGFET in the strong-inversion regime, using assumed functional forms. It was found that the model-predicted spatial variations of the mobility, the quasi-fermi potential and the electrostatic potential along the fin center match very well with 2D device simulations. But the fields (and therefore the potentials) elsewhere in the device do not match well. It is felt that the key reason for this is that the 2D Poisson equation had not been considered.

Considering that this approach did not yield an acceptable match for the core physical quantities of interest (potential, electric fields) in much of the device, and that neither the subsequent integration of (28) nor the determination of the bias-dependent constants seemed to be a tractable task, future work on this approach was abandoned. Instead, attention was hereafter focused on developing a 1D core model that self-consistently includes velocity saturation effects, with other effects added onto the core model as second-order effects. These are described in the subsequent chapters.

Chapter 4
Core Model

In the compact modeling arena, one usually starts with a basic long channel model (considering constant mobility) as the core model, and then adds one physical effect at a time. The more the number of physical effects considered *simultaneously* in the core model *itself* (the ideal desired objective) as opposed to adding one by one later on, the more accurate (although complex) the model becomes [53].

In this chapter, we describe the way in which we have improved over existing core models by including velocity saturation effects in a fundamental way as an integral part of the model derivation, and in a manner that Gummel symmetry compliance is met. We have focused on a FinFET with undoped or lightly doped body because FinFETs are expected to be undoped at the technology node where they will be introduced [5,30,62]. This is the reason many existing DGFET modeling works [10-17,19,30] also proceed with the undoped or lightly doped assumption.

4.1 Introduction

Amongst the models in the literature that considered velocity saturation effects, [8,19] had used the Caughey-Thomas model [20] or its variants with exponent n=1 (the variants, such as the Sodini-Ko-Moll model [61], differing in the way the critical electric field E_c relates to v_{sat}, but all of them nevertheless using an exponent n=1). References [7,22] considered velocity saturation effects using the Caughey-Thomas model with exponent n=2 but in an *over-simplified approach*, where by the term *over-simplified approach*, we refer to the approach of deriving a model using constant mobility and then simply replacing the constant mobility with an effective mobility that considers a *spatially averaged* lateral electric field as the driving field for velocity saturation. Prior to the present work, there has been no work done on modeling velocity saturation effects in DGFETs using the Caughey-Thomas model with *exponent* n=2 where the *spatial variation* of the lateral electric field driving the velocity saturation effect is *represented accurately*. We describe in this chapter a way in which velocity saturation effects are included as an integral part of the model derivation, as opposed to an *over-simplified* modeling. Hence, this approach is expected to be physically more accurate.

Using an exponent n=2, even though more complicated, has been found to yield a better match with experimental data for N-channel devices [2]. Further, it has been suggested [21] that using an exponent n=1 or any odd number in general, would yield a model that would fail the Gummel symmetry test at V_{DS}=0. That is the reason why models like PSP use n=2 for conventional MOSFETs; not just for N-channel devices, but also for P-channel devices [7,54] (PSP actually uses an adjusted form of the Scharfetter-Gummel model for velocity saturation in an *over-simplified* approach, which simplifies to the Caughey-Thomas model with n=2, except that the saturation velocity parameter v_{sat} becomes bias dependent in the case of P-channel devices [7]). Even though efforts after [20] such as the Canali model [63] have found a good experimental fit using *fractional* values for exponent n between 1 and 2, their work showed that the exponent n increases (towards 2) at temperatures higher than room temperature. Then, considering the fact that fractional exponents are hard to accommodate in a compact model derivation, and that the operating temperatures especially of high-speed devices are higher than room temperature, and that the Caughey-Thomas exponent is usually

not a temperature scaled parameter in compact models, this lends further justification for modeling velocity saturation using an exponent n=2 in a compact model.

Lastly, threshold voltage based models are not very physical [7], and charge sheet models are not valid in ultra-thin DGFETs as they fail to model phenomena such as volume inversion [10]. Hence we have developed an inversion-charge based drain current model. This is done by solving for the drain current (I_{DS}) of an undoped/lightly-doped SDGFET under the Gradual Channel Approximation (GCA), considering the intrinsic portion of the device. We have focused on mobility degradation due to velocity saturation; and other mobility degradation effects such as that due to the vertical field, have not been considered in the core model. This is not because vertical field mobility degradation is not important, but rather because it may be acceptable to incorporate an engineering model for vertical field mobility degradation (such as the one used in [7]) in an *over-simplified* manner (ie. replacing basal mobility with vertical field degraded mobility), since they have a stronger gate voltage dependence, and not so much of a spatial dependence. This is described in the next chapter.

4.2 Basic formulation

The schematic of the intrinsic portion of an N-channel SDGFET is shown in Fig. 3-1. Under the GCA and neglecting the body doping term, the 1D Poisson equation can be written as:

$$\frac{\partial^2 \psi}{\partial y^2} = \frac{q n_i}{\varepsilon} e^{\left(\psi - \phi_{fn}\right)/\phi_t} \tag{29}$$

where ψ is the electrostatic potential and φ_{fn} is the electron quasi fermi potential. The latter is assumed to be constant along y, in line with the assumption that the current is negligible along the y direction. All potentials (electrostatic as well as quasi-fermi) are referenced to the electron quasi fermi potential deep in the source region. Subject to the boundary condition that the electric field component E_y vanishes at $y=0$, the solution to (29) can very easily be obtained as:

$$\psi(x, y) = A_1(x) + 2\phi_t \ln\left(\sec\left(\beta y e^{\alpha}\right)\right) \tag{30}$$

$$E_y(x,y) = -2\beta\phi_t e^{\alpha} \tan\left(\beta y e^{\alpha}\right) \tag{31}$$

where:

$$\beta = \sqrt{\frac{qn_i}{2\varepsilon\phi_t}} \tag{32}$$

$$\alpha = \alpha(x) = \frac{A_1(x) - \phi_{fn}(x)}{2\phi_t} \tag{33}$$

$A_1(x)$ is the electrostatic potential at some point x along the center of the fin ($y=0$). $\alpha(x)$ is related to the electron density along the center of the fin. In terms of the surface potential $\psi_s(x)$, the normal component of the electric field at the oxide-silicon interface $E_{ys} = E_y(x, W_{fin}/2)$ is:

$$E_{ys}(\psi_s) = -2\beta\phi_t e^{\alpha}\sqrt{e^{\psi_s/\phi_t} - 1} \tag{34}$$

where:

$$\psi_s(x) = \psi(x, W_{fin}/2) - A_1(x) \tag{35}$$

The potential summation rule traversing from the gate electrode to the source can be written as:

$$V_{GS} - \Delta\phi = -\frac{\varepsilon E_{ys}(x)}{C_{ox}} + \psi(x, W_{fin}/2) \tag{36}$$

where $\Delta\varphi$ is the work function difference between the gate electrode and intrinsic silicon.

By Gauss law, the inversion charge areal density is given by:

$$Q_i(x) = 2\varepsilon E_{ys}(x) \tag{37}$$

The algebra gets easier if we define another intermediate variable β_1 as:

$$\beta_1 = \frac{\beta W_{fin} e^{\alpha}}{2} \tag{38}$$

Note that this β_1 is the same as β in [10]. Therefore,

$$2\alpha\phi_t = 2\phi_t \ln\left(\frac{2\beta_1}{\beta W_{fin}}\right) \tag{39}$$

Then, setting $y=W_{fin}/2$ in (30), and using (33) and (38) we get:

$$\psi(x, \frac{W_{fin}}{2}) = 2\phi_t \alpha(x) + \phi_{fn}(x) + 2\phi_t \ln\left(\sec\left(\beta_1\right)\right) \tag{40}$$

Setting $y=W_{fin}/2$ in (31), and using (38) we get:

$$E_{ys}(x) = \frac{-4\phi_t \beta_1 \tan\beta_1}{W_{fin}} \tag{41}$$

Substituting (41) and (40) in (36), and then using (39), we get:

$$f(\beta_1) = 0 = \frac{4\varepsilon\phi_t \beta_1 \tan(\beta_1)}{W_{fin} C_{ox}} + \phi_{fn} + 2\phi_t \ln\left(\frac{2\beta_1 \sec\left(\beta_1\right)}{\beta W_{fin}}\right) - (V_{GS} - \Delta\phi) \tag{42}$$

Equation (42) is the same as (4) in [10]. It is referred to as the *input voltage equation* in the literature [64]. β_1 is plotted versus $\left(V_{GS} - \Delta\phi - \phi_{fn}\right)$ in Fig. 4-1. Note that φ_{fn} is the smallest (= 0) at the source end and is the largest (= V_{DS}) at the drain end. It is thus clear, and as is also stated in [10], that $0 < \beta_1 < \pi/2$, in order for all the expressions encountered till now to be finite and physical. Thus, using (42), we can in principle determine β_1 at the source and drain ends by setting $\varphi_{fn}=0$ and $\varphi_{fn}=V_{DS}$ respectively. We refer to these as β_{1s} and β_{1d} respectively. An approximated form of (42) is [19]:

$$f(\beta_1) = 0 = \frac{4\varepsilon\phi_t\beta_1 \tan(\beta_1)}{W_{fin}C_{ox}} + \phi_{fn} + \phi_t \ln\left(\frac{4\left(\beta_1 \tan\left(\beta_1\right) + \beta_1^{\,2} \tan^2\left(\beta_1\right)\right)}{\beta^2 W_{fin}^{\,2}}\right) - (V_{GS} - \Delta\phi) \tag{43}$$

Recently there have also been [65,66] closed form approximate solutions to (42).

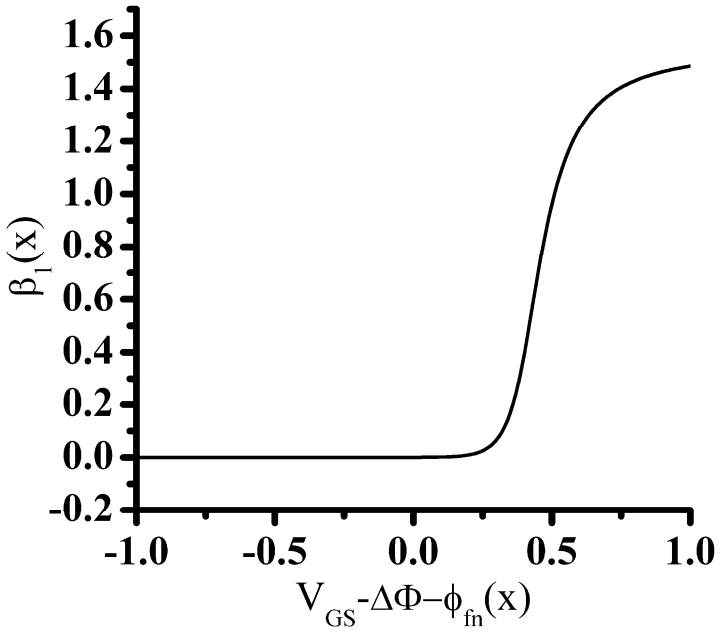


Figure 4-1: $\beta_1(x)$ versus (V_{GS}-$\Delta\varphi$-$\varphi_{fn}(x)$). Alternately, if x is pinned at $L/2$, then this plot can be interpreted as a plot of β_{1d} versus (V_{GS}-$\Delta\varphi$-V_{DS}).

Now, in the drift-diffusion model, the drain current *per unit fin height* is written as:

$$I_{DS} = -\mu_{eff}(x) Q_i(x) \frac{d\phi_{fn}}{dx} \tag{44}$$

We now model $\mu_{eff}(x)$ using the Caughey-Thomas model [20] with exponent n=2 as:

$$\mu_{eff}(x) = \frac{\mu_0}{\sqrt{1 + \frac{\mu_0^{\,2} E_{xs}^{\,2}}{v_{sat}^{\,2}}}} \tag{45}$$

where:

$$E_{xs}(x) = -\frac{\partial \psi(x, W_{fin}/2)}{\partial x} \tag{46}$$

In (46), we have simplistically chosen to model the driving field for velocity saturation as being the lateral field at the *oxide-silicon interface*. This is not unreasonable, since even though charge sheet models are invalid in DGFETs [10] and there is non-negligible current flowing even far from the oxide-silicon interface, the fact still remains that the current at the interface is dominant when compared to elsewhere as can also be seen in Fig. 2-4 (except in the sub-threshold regime [30] where the leakiest path is along the fin center. But as we will see, our model predicts the current quite well in the sub-threshold regime also).

To compute I_{DS}, our first effort is to compute E_{xs}, Q_i and $d\varphi_{fn}/dx$ in terms of the intermediate variable β_1. Substituting (40) and (39) in (46), we get:

$$E_{xs}(x) = -\left[\frac{d\phi_{fn}}{dx} + \frac{2\phi_t}{\beta_1}\left(1 + \beta_1 \tan(\beta_1)\right)\frac{d\beta_1}{dx}\right] \tag{47}$$

From (37) and (41) we get:

$$Q_i = \frac{-8\varepsilon\phi_t\beta_1 \tan(\beta_1)}{W_{fin}} \tag{48}$$

Differentiating (42) with respect to x and rearranging, we get:

$$\frac{d\phi_{fn}}{dx} = -2\phi_t\left[\frac{2\varepsilon}{W_{fin}C_{ox}}\left(\beta_1 \sec^2\beta_1 + \tan\beta_1\right) + \frac{1 + \beta_1 \tan\beta_1}{\beta_1}\right]\frac{d\beta_1}{dx} \tag{49}$$

Substituting (49) in (47), we get:

$$E_{xs}(\beta_1) = \frac{4\varepsilon\phi_t}{W_{fin}C_{ox}}\left(\beta_1 \sec^2\beta_1 + \tan\beta_1\right)\frac{d\beta_1}{dx} \tag{50}$$

where it is to be noted that $d\beta_1/dx$ is negative (because $\alpha(x)$ decreases towards the drain, and this is because the carrier density decreases towards the drain).

From (44) and (45) we get:

$$1+\frac{{\mu_0}^2 {E_{xs}}^2}{{v_{sat}}^2}=\frac{{\mu_0}^2 {Q_i}^2}{{I_{DS}}^2}\left(\frac{d\phi_{fn}}{dx}\right)^2 \tag{51}$$

Substituting (48), (49) and (50) in (51) and working out the algebra, we finally get an expression for $d\beta_1/dx$ in terms of β_1:

$$\frac{\partial\beta_1}{\partial x}=\frac{-W_{fin}}{4\mu_0\varepsilon\phi_t}\left\{\frac{16{\phi_t}^2{\beta_1}^2\tan^2\beta_1}{{I_{DS}}^2}\left[\frac{2\varepsilon}{W_{fin}C_{ox}}\left(\beta_1\sec^2\beta_1+\tan\beta_1\right)+\frac{1+\beta_1\tan\beta_1}{\beta_1}\right]^2-\frac{\left(\beta_1\sec^2\beta_1+\tan\beta_1\right)^2}{{C_{ox}}^2{v_{sat}}^2}\right\}^{-1/2} \tag{52}$$

Integrating (52) from the source end to some arbitrary point x gives:

$$\frac{-W_{fin}}{4\mu_0\varepsilon\phi_t}\left(x+\frac{L}{2}\right)=\int_{\beta_{1s}}^{\beta_1}\left\{\frac{16{\phi_t}^2{\beta_1}^2\tan^2\beta_1}{{I_{DS}}^2}\left[\frac{2\varepsilon}{W_{fin}C_{ox}}\left(\beta_1\sec^2\beta_1+\tan\beta_1\right)+\frac{1+\beta_1\tan\beta_1}{\beta_1}\right]^2-\frac{\left(\beta_1\sec^2\beta_1+\tan\beta_1\right)^2}{{C_{ox}}^2{v_{sat}}^2}\right\}^{1/2}d\beta_1 \tag{53}$$

The limiting case of (53) for the constant mobility case, got by setting $v_{sat}=\infty$, can be recognized as the exact same equation as the unnumbered equation between (5) and (6) in [10], which had solved it for the constant mobility case.

Equations (42) and (53) are the key equations in our approach.

In principle, an approach to find I_{DS} could be thus: From (42), we can find β_{1s} and β_{1d} for a given V_{GS} and V_{DS}. Then we can integrate (53) (assuming it is integrable) till the drain end (setting $x=L/2$ in the LHS and setting β_{1d} as the upper limit of the integral in the RHS), and the resulting equation would then only have I_{DS} as the unknown; thus we can find I_{DS}.

An I_d-V_d plot generated by numerically solving (53) and (42) in Scilab [67] is shown in Fig. 4-2 (see trace labeled *Exact*).

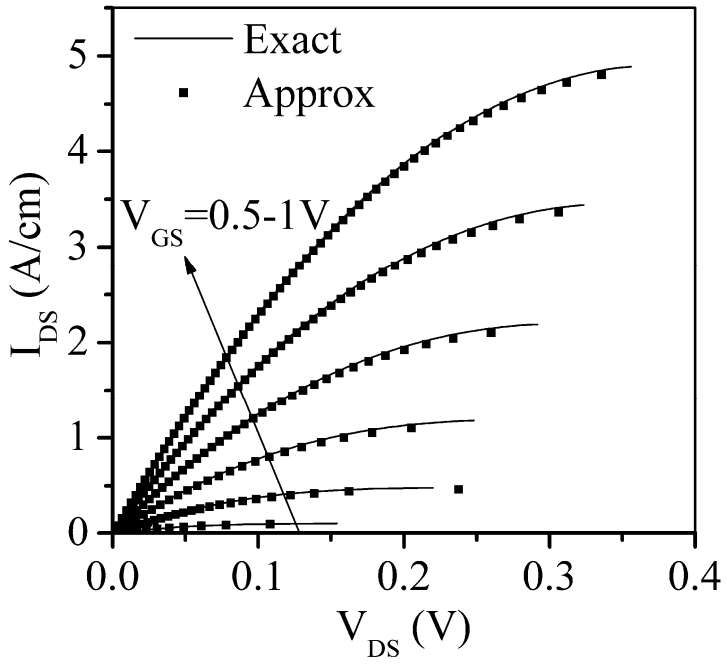


Figure 4-2: I_d-V_d plot by numerically solving (53) and (42) using a constant I_{DS} step-size.

Numerically solving (53) is tricky. It involves ramping I_{DS} for a given V_{GS} instead of ramping V_{DS} (as usually done), and finding β_{1d} using Newton-Raphson iterations that would make (53) hold. Once β_{1d} is found, then (42) can be used to find V_{DS}.

Now, (53) is not easily integrable, so we make some approximations in order to proceed. This is described in the next section.

4.3 Approximations

In (53), let us denote the $\frac{1+\beta_1 \tan\beta_1}{\beta_1}$ term by t_{12}. If this term t_{12} is multiplied by $1-\frac{\tan\beta_1-\beta_1}{\left(2+\beta_1\tan\beta_1\right)\tan\beta_1}$, then the analytics become simpler. Before proceeding with the simplified analytics, the origin and justification of this approximation is explained first.

4.3.1 Origin and justification of the approximation

In (53), it is clear that there is a maximum value of I_{DS} beyond which the integrand becomes imaginary. When I_{DS} is just large enough to cause the integrand to vanish, we can see that this situation corresponds to $d\beta_1/dx$ (from (52)) and therefore E_{xs} (from (50)) and $d\varphi_{fn}/dx$ (from (49)) all becoming infinite, and therefore μ_{eff} (from (45)) becoming 0. The model cannot handle these singularities, and so this extremum point must be considered as a limit of validity of the model. Physically, this means that near the drain end the carriers are traveling at their saturation velocity; and that between that point and the drain end, E_x and φ_{fn} are varying very rapidly. The limiting V_{DS} that causes this extremum is V_{DSat}. As stated in [2], if V_{DS} is increased beyond V_{DSat}, then the point where this extremum happens starts moving closer to the source. The electron quasi-fermi potential at that point continues to be pinned at V_{DSat}. It is only the region between the source end and that point that can be reasonably considered as the region where the GCA is valid.

Setting the integrand in (53) to 0, the limiting I_{DS} (I_{DSmax}) can be got as:

$$I_{DS\max} = 4\phi_t\beta_1 \tan\beta_1 C_{ox} v_{sat}\left[\frac{2\varepsilon}{W_{fin}C_{ox}} + \frac{1+\beta_1\tan\beta_1}{\beta_1\left(\beta_1\sec^2\beta_1 + \tan\beta_1\right)}\right] \tag{54}$$

The second term in (54) can be rewritten as:

$$f_1(\beta_1) = \frac{1+\beta_1\tan\beta_1}{{\beta_1}^2 + \left(1+\beta_1\tan\beta_1\right)\beta_1\tan\beta_1} \tag{55}$$

Compare this to:

$$f_2(\beta_1) = \frac{1+\beta_1\tan\beta_1}{\beta_1\tan\beta_1 + \left(1+\beta_1\tan\beta_1\right)\beta_1\tan\beta_1} \tag{56}$$

Equation (55) and (56) differ only in the first term in the denominator. They are both plotted in Fig. 4-3.

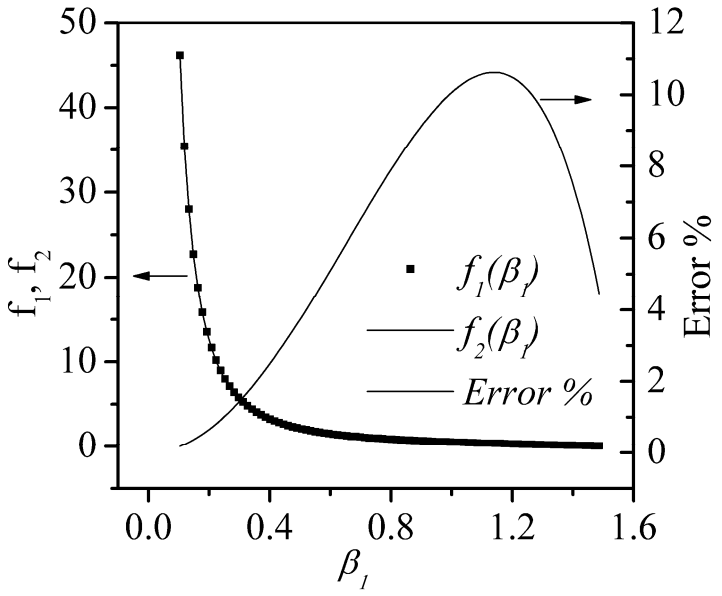


Figure 4-3: Comparison of $f_1(\beta_1)$ and $f_2(\beta_1)$.

We see a reasonably good match, with a maximum error of 10.62% and an average error of 5.82%. Approximating $f_1(\beta_1)$ as $f_2(\beta_1)$ translates to making the approximation that is stated in the first paragraph of section 4.3 As a further validation of this approximation, the I_d-V_d plots have been re-generated using Scilab by numerically solving (53) and (42) but this time using this approximation, and they are shown in Fig. 4-2 (see trace labeled *Approx*). One can clearly see a very good match.

4.3.2 Use of the approximation

Approximating (55) as (56), equation (53) can be simplified as:

$$\frac{-W_{fin}}{4\mu_0\varepsilon\phi_t}\left(x+\frac{L}{2}\right) \cong \int_{\beta_{2s}}^{\beta_2} \sqrt{\frac{16\phi_t{}^2\beta_2{}^2}{I_{DS}{}^2}\left[\frac{2\varepsilon}{W_{fin}C_{ox}}+\frac{1+\beta_2}{(2+\beta_2)\beta_2}\right]^2-\frac{1}{C_{ox}{}^2 v_{sat}{}^2}}\, d\beta_2 \qquad (57)$$

where we have changed from the state variable β_1 to β_2, which is given by:

$$\beta_2 = \beta_1 \tan\beta_1 \qquad (58)$$

Note: $\beta_{2s} = \beta_{1s} \tan \beta_{1s}$ and $\beta_{2d} = \beta_{1d} \tan \beta_{1d}$

Equation (57) is still not easily integrable and we need to make further approximations. The integrand in (57) is of the form $\sqrt{Af^2(x) - B}$, where the second term $B = \frac{1}{C_{ox}{}^2 v_{sat}{}^2}$. Noting that this term B will be small (limiting to 0 in the case of negligible velocity saturation effects), this expression can be approximated as: $\sqrt{A} f(x) - \frac{B}{2\sqrt{A} f(x)}$. Making this approximation in (57) and after solving some simple integrals and some algebra, we finally get:

$$I_{DS} = \frac{8 a_1(x) \phi_t}{\frac{W_{fin}(x + L/2)}{4\mu_0 \varepsilon \phi_t} + \sqrt{\left(\frac{W_{fin}(x + L/2)}{4\mu_0 \varepsilon \phi_t}\right)^2 + \frac{a_1(x) a_2(x)}{b \cdot C_{ox}{}^2 v_{sat}{}^2}}} \tag{59}$$

where:

$$a_1(x) = \frac{b}{2}\left(\beta_{2s}{}^2 - \beta_2(x)^2\right) + \left(\beta_{2s} - \beta_2(x)\right) - \ln\left(\frac{\beta_{2s} + 2}{\beta_2(x) + 2}\right)$$

$$a_2(x) = \frac{(2b-1)}{\sqrt{4b^2+1}} \ln\left\{\frac{\left[\frac{2b(\beta_{2s}+1)+1}{\sqrt{4b^2+1}} + 1\right].\left[\frac{2b(\beta_2(x)+1)+1}{\sqrt{4b^2+1}} - 1\right]}{\left[\frac{2b(\beta_{2s}+1)+1}{\sqrt{4b^2+1}} - 1\right].\left[\frac{2b(\beta_2(x)+1)+1}{\sqrt{4b^2+1}} + 1\right]}\right\} + \ln\left(\frac{b\beta_{2s}{}^2 + (2b+1)\beta_{2s} + 1}{b\beta_2(x)^2 + (2b+1)\beta_2(x) + 1}\right) \tag{60}$$

and:

$$b = \frac{2\varepsilon}{W_{fin} C_{ox}} \tag{61}$$

The drain current I_{DS} expression can then be derived by setting $x=L/2$ in (59) and $\beta_2 = \beta_{2d}$ in the expressions for a_1 and a_2 in (60) (and calling them a_{1d} and a_{2d} respectively). We get:

$$I_{DS} = \frac{2I_{DS0}}{1+\sqrt{1+\frac{8{\mu_0}^2\varepsilon{\phi_t}^2 a_{1d}a_{2d}}{W_{fin}L^2C_{ox}{v_{sat}}^2}}} \tag{62}$$

$$I_{DS0} = \frac{16\mu_0\varepsilon{\phi_t}^2 a_{1d}}{W_{fin}L} \tag{63}$$

where I_{DS0} is the current in the absence of velocity saturation (constant mobility current).

After determining I_{DS} using (62) and (63), (59) and (60) can in principle be used to determine the *spatial* variation of $\beta_2(x)$, though it is not in closed form. From that, we can determine the *spatial* variation of $Q_i(x)$ using (48) directly, *without* having to determine the intermediate variable $\beta_1(x)$ first using (58). Once we know the *spatial* variation of $Q_i(x)$, we can then use the Ward-Dutton charge partitioning scheme [52] to derive expressions for the source and drain terminal charges.

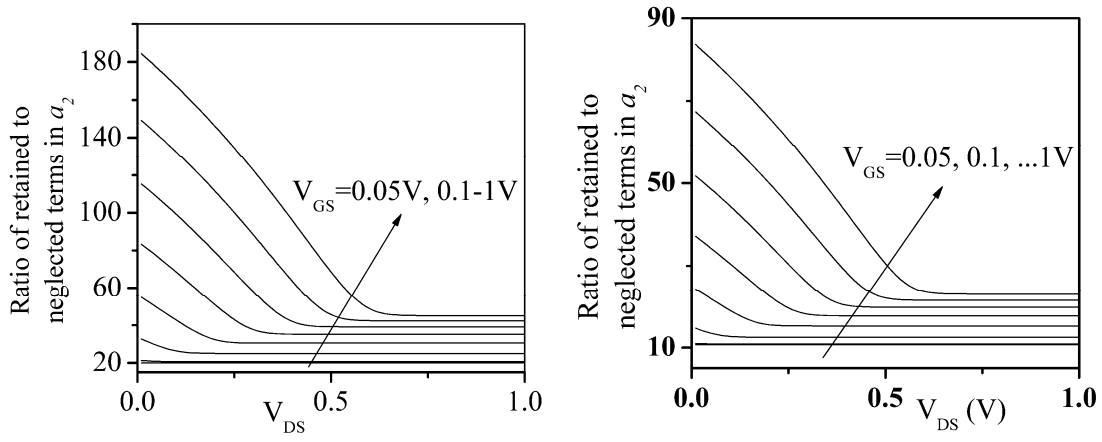
I_{DS} in (62) can be further simplified by considering that the second term in the square root is small (meaning large v_{sat}). We then get:

$$I_{DS} = \frac{I_{DS0}}{1+\frac{I_{DS0}\mu_0 a_{2d}}{8C_{ox}{v_{sat}}^2 L}} \tag{64}$$

Further, a plot of the terms comprising a_2 in (60) reveals that the first logarithm term is negligible. Thus a_2 can be approximated as:

$$\begin{aligned} a_{1d} &= \frac{b}{2}\left({\beta_{2s}}^2 - {\beta_{2d}}^2\right) + \left(\beta_{2s} - \beta_{2d}\right) - \ln\left(\frac{\beta_{2s}+2}{\beta_{2d}+2}\right) \\ a_{2d} &= \ln\left(\frac{b{\beta_{2s}}^2 + \beta_{2s}(2b+1)+1}{b{\beta_{2d}}^2 + \beta_{2d}(2b+1)+1}\right) \end{aligned} \tag{65}$$

The ratio of the dominant (retained) terms and the neglected terms (in approximating (60) as (65)) is plotted in Fig. 4-4. As can be clearly seen, the approximation in a_{2d} is quite valid.



(a) L=1 um, W_{fin}=20 nm, t_{ox}=1.5 nm (b) L=30 nm, W_{fin}=10 nm, t_{ox}= 1 nm

Figure 4-4: Ratio of the retained terms to the neglected terms in a_2 for long and short channel devices, for V_{GS}=0.05V, 0.1, 0.2, …1V.

Equations (63), (64) and (65) are the final drain current equations in our model.

4.4 The drain saturation voltage V_{DSat}

As explained earlier, the approach loses validity when I_{DS} becomes so high that the integrand in (53) becomes 0. This value of I_{DS} is I_{DSmax}, or I_{DSat}. To determine I_{DSat} and V_{DSat}, it is not immediately obvious whether the approach of obtaining them by setting the integrand in (53) to 0, is the same as the approach of obtaining them by setting the derivative of I_{DS} with respect to V_{DS} to 0. We make this clear below by a simple use of the chain rule and Leibniz rule [68].

From (42), we have the following facts:

$\beta_{1d}(V_{GS}, V_{DS})$ is a function of V_{GS} and V_{DS} **(Statement A1)**

Further, if $g(\beta_1, I_{DS})$ is the integrand in (53), and if $f(\beta_1, I_{DS})$ is its integral, then we have the following facts from (53):

$f(\beta_{1d}, I_{DS}) - f(\beta_{1s}, I_{DS}) = k$, a constant **(Statement A2)**

$$\int g(\beta_1, I_{DS})d\beta_1 = f(\beta_1, I_{DS}) \qquad \textbf{(Statement A3)}$$

From (Statement A3), we see that $\frac{\partial f}{\partial \beta_{1d}}$ equals $g(\beta_{1d}, I_{DS})$, which is the integrand in (53) evaluated at the drain end. But $\frac{\partial f}{\partial \beta_{1d}} = \frac{\partial f}{\partial I_{DS}} \frac{\partial I_{DS}}{\partial V_{DS}} \frac{\partial V_{DS}}{\partial \beta_{1d}}$. Thus if $\frac{\partial I_{DS}}{\partial V_{DS}}$ is 0, then $\frac{\partial f}{\partial \beta_{1d}}$ is also 0, meaning that the integrand in (53) evaluated at the drain end is also 0. Thus the two approaches/conditions for finding I_{DSat} and V_{DSat} are equivalent.

One could therefore argue that an approach to find V_{DSat} is to set the integrand in (53) to 0, as computed at the *drain* end (meaning that using β_{1d} in place of β_1 in (53)). The drain end is the end where the integrand would vanish first, because I_{DSmax} is smaller if β_1 is smaller, and β_1 is the smallest at the drain end as seen from Fig. 4-1. But this approach is not easy, because β_{1DSat} in (53) is not known (because V_{DSat} is not known). So we will instead set $\frac{\partial I_{DS}}{\partial V_{DS}} = 0$ in our final drain current equations (63), (64) and (65).

From (65), we can eliminate β_{2d} in the expressions for a_{1d} and a_{2d}, and thus write a_{2d} in terms of a_{1d}. Thus, we can have an expression for I_{DS} in terms of a_{1d} alone. Then, since β_{1d} is a function of V_{DS} from (42), and β_{2d} is a function of β_{1d} from (58), and a_{1d} is a function of β_{2d} from (65), then, setting $\frac{\partial I_{DS}}{\partial V_{DS}} = 0$ is equivalent to setting $\frac{\partial I_{DS}}{\partial a_{1d}} = 0$. This simplifies the algebra.

We describe next three methods that were employed to calculate V_{DSat} using this approach.

4.4.1 *Method 1*: 5th degree polynomial

Differentiating (64) with respect to a_{1d} and setting it to 0 and rearranging gives:

$$\frac{\partial I_{DS0}}{\partial a_{1d}} = \frac{\mu_0 {I_{DS0}}^2}{8 C_{ox} {v_{sat}}^2 L} \frac{\partial a_{2d}}{\partial a_{1d}} \qquad (66)$$

Now,

$$\frac{\partial a_{2d}}{\partial a_{1d}} = \frac{\dfrac{\partial a_{2d}}{\partial \beta_{2d}}}{\dfrac{\partial a_{1d}}{\partial \beta_{2d}}} \tag{67}$$

From (65), we can easily calculate $\frac{\partial a_{1d}}{\partial \beta_{2d}}$ and $\frac{\partial a_{2d}}{\partial \beta_{2d}}$, and then substituting them into (67) and simplifying, we finally get (68). Note however that in calculating $\frac{\partial a_{1d}}{\partial \beta_{2d}}$, we neglect the sole logarithm term in the expression (65) for a_{1d} in order for the algebra to be tractable. This approximation is justified because that logarithm term is negligible in above-threshold conditions, which is the regime velocity saturation effects are predominant in.

$$\frac{\partial a_{2d}}{\partial a_{1d}} = \frac{2b}{c} + \frac{(1-2b)\left(1+\dfrac{2b}{c}\right)}{c\beta_{2d} + 2c - (1-2b)} \tag{68}$$

where:

$$c = c(a_{1d}) = b^2 {\beta_{2s}}^2 + 2b\beta_{2s} + 1 - b\left(2a_{1d} + 1\right) \tag{69}$$

From (69) and from the expression for a_{1d} in (65) after neglecting the sole logarithm term therein, we get:

$$\beta_{2d} = \frac{\sqrt{b+c} - 1}{b} \tag{70}$$

Substituting (70) in (68), we get:

$$\frac{\partial a_{2d}}{\partial a_{1d}} = \frac{2b}{c} + \frac{(1-2b)\left(1+\dfrac{2b}{c}\right)}{\dfrac{c}{b}\left(\sqrt{b+c} - 1\right) + 2c - (1-2b)} \tag{71}$$

From (63) we have:

$$\frac{\partial I_{DS0}}{\partial a_{1d}} = \frac{16\mu_0 \varepsilon \phi_t^{\,2}}{W_{fin} L} \tag{72}$$

Substituting (72), (63) and (71) in (66) gives:

$$\frac{16\mu_0 \varepsilon \phi_t^{\,2}}{W_{fin} L} = \frac{\mu_0}{8 C_{ox} v_{sat}^{\,2} L} \left[\frac{16\mu_0 \varepsilon \phi_t^{\,2} a_{1d}}{W_{fin} L} \right]^2 \left[\frac{2b}{c} + \frac{(1-2b)\left(1+\frac{2b}{c}\right)}{\frac{c}{b}\left(\sqrt{b+c}-1\right)+2c-(1-2b)} \right] \tag{73}$$

Rearranging to get the $\sqrt{b+c}$ term isolated and then squaring, we get:

$$b+c = \left\{ \left[\frac{1-2b}{c} \right] \left[b+c+\frac{bk(c+2b)a_{1d}^{\,2}}{c-2bka_{1d}^{\,2}} \right] \right\}^2 \tag{74}$$

where:

$$k = \frac{2\mu_0^{\,2} \varepsilon \phi_t^{\,2}}{W_{fin} L^2 C_{ox} v_{sat}^{\,2}} \tag{75}$$

From (69), we have:

$$a_{1d} = \frac{n-c}{2b} \tag{76}$$

where:

$$n = \left(b\beta_{2s}+1\right)^2 - b \tag{77}$$

Substituting (76) in (74) and after some tedious algebra, we get a fifth degree polynomial equation of the form:

$$f(c) = \sum_0^5 m_i c^i = 0 \tag{78}$$

where:

$$
\begin{aligned}
m_0 &= 4bk^2n^4 - (4b^2 - kn^2)^2(1-2b)^2 \\
m_1 &= 4\left[k^2n^4 - 4bkn^2\left(b+kn\right)\right] - 4(4b^2 - kn^2)(2b+kn)(1-2b)^2 \\
m_2 &= 4\left[2b\left(k^2n^2 + 2\left(b+kn\right)^2\right) - 4kn^2\left(b+kn\right)\right] - 4(2b+kn)^2(1-2b)^2 + 2k(4b^2 - kn^2)(1-2b)^2 \\
m_3 &= 4\left[2\left(k^2n^2 + 2\left(b+kn\right)^2\right) - 4bk\left(b+kn\right)\right] + 4k(2b+kn)(1-2b)^2 \\
m_4 &= 4\left[bk^2 - 4k\left(b+kn\right)\right] - k^2(1-2b)^2 \\
m_5 &= 4k^2
\end{aligned}
\tag{79}
$$

where k and n are given by (75) and (77) respectively.

For a given V_{GS}: β_{1s}, β_{2s} and n can be calculated in order from (42), (58) and (77) respectively. Having calculated n for a given V_{GS}, and having calculated k, all the m_i coefficients can be calculated using (79), and (78) can then be solved for c. Being a 5th degree polynomial equation, it will have 5 roots. We should reject those roots for which the term being squared in the RHS of (74) is negative. This is because in arriving at (74), we have done a squaring operation, which artificially expands the root set. Having rejected such roots if any, amongst the remaining roots, we should then look for the largest, real root that would yield a value of β_{2d} (as calculated using (70)) that lies between 0 and β_{2s} (to be physically meaningful). This would correspond to the smallest V_{DS} at which I_{DS} saturates. If such a root does not exist, then it means that the slope never becomes 0, and the single expression (64) can be used as is in all regimes. This is indeed found to be the case for low V_{GS}.

Per the Abel-Ruffini theorem [69], polynomials of order 5 (ie. quintics) and higher cannot be solved analytically in general. They can only be solved numerically, using methods like the Jenkins-Traub method [69]. However it is found that by discarding m_5, the resulting 4th order polynomial (ie. quartic) gives a very close match to that obtained from the 5th order polynomial, with a maximum error of 0.06%. This is shown in Fig. 4-5. This could be advantageous because quartics and lower order polynomials have a closed form analytical solution [69], though it is still quite complex. In conjunction with the constraint that it should be the largest, real root, one could in principle solve (78) non-iteratively after approximating it to a quartic. However in the results discussed in section 4.7.1, we have solved the quintic numerically.

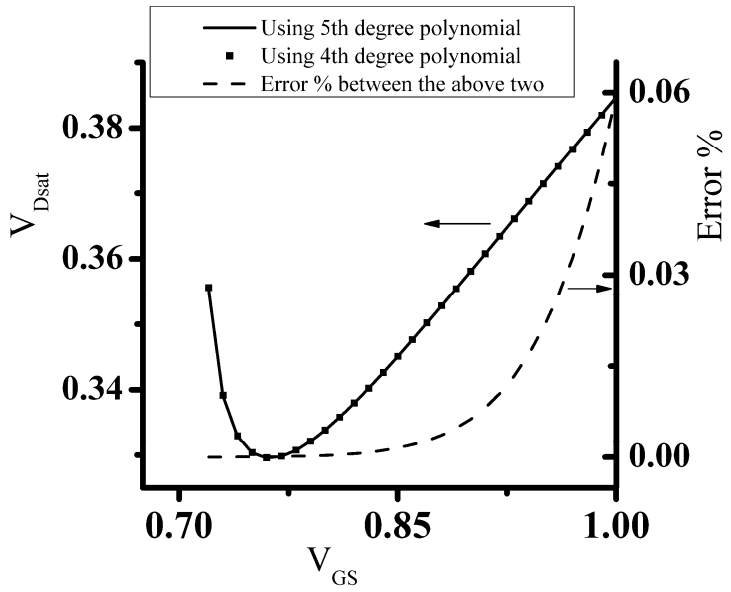


Figure 4-5: Comparison of polynomial solutions obtained by using the 5th and 4th degree versions.

Fig. 4-5 may need an explanation as to why (78) doesn't have a valid root at lower V_{GS}. The cause of this is our formulation accounting for drift and diffusion (using the product $\mu_{eff}.grad_x(\phi_{fn})$), as opposed to just drift (using the product $\mu_{eff}.grad_x(\psi)$). In the latter case, velocity saturation would ensure that the product saturates at v_{sat} when E_x becomes infinity.

From (49) and (50), we have:

$$\frac{grad_x(\phi_{fn})}{|E_{xs}|} = 1 + \frac{1+\beta_1 \tan\beta_1}{b\beta_1\left(\beta_1 \sec^2\beta_1 + \tan\beta_1\right)} \tag{80}$$

where b is given by (61).

At moderately low V_{GS} (and below) and moderately high V_{DS} (and above), it is clear from Fig. 4-1 that β_1 is very small ($\rightarrow 0$). Then, clearly the second term in the RHS of (80) is dominant. Thus the first term "1" can be neglected. Then, multiplying it by Q_i given by (48) gives:

$$Q_i.\frac{grad_x(\phi_{fn})}{E_{xs}} \cong \frac{\left(1+\beta_1 \tan\beta_1\right)\tan\beta_1}{b\left(\beta_1 \sec^2\beta_1 + \tan\beta_1\right)}.\frac{8\varepsilon\phi_t}{W_{fin}} \tag{81}$$

It is clear that this term tends to $4\varepsilon\phi_t/bW_{fin}$ for very small β_1. Thus, $\mu_{eff}.Q_i.grad_x(\phi_{fn}) = \mu_{eff}.E_{xs}.Q_i.\frac{grad_x(\phi_{fn})}{E_{xs}}$ tends to $\frac{4\mu_{eff}.E_{xs}\varepsilon\phi_t}{bW_{fin}}$, which *monotonically* increases towards $4v_{sat}\varepsilon\phi_t/bW_{fin}$ as V_{DS} is increased. This explains why there is no 0 slope in the I_d-V_d curve at lower V_{GS}.

On the other hand, at higher V_{GS}, the converse is true (β_1 tends towards $\pi/2$ as can be seen from Fig. 4-1) and the RHS of (80) tends towards 1 (the second term can be neglected; this physically means that the ratio of diffusion to drift component is negligible, which is a known situation at moderately high V_{GS}). Then, $\mu_{eff}.Q_i.grad_x(\phi_{fn}) = \mu_{eff}.E_{xs}.Q_i.\frac{grad_x(\phi_{fn})}{E_{xs}}$ tends towards $\mu_{eff}.E_{xs}.Q_i$. Then, since $\mu_{eff}E_{xs}$ monotonically tends towards v_{sat} starting from *0*, whereas Q_i monotonically decreases as V_{DS} is increased, there is a point where their product peaks (ie. has a *0* slope) in the I_d-V_d curve.

This situation was also verified by numerically solving (53) exactly (ie. by not using the approximation stated in section 4.3), where it was seen that at lower V_{GS}, a zero slope doesn't occur for *any* V_{DS}, but it does occur when solved for higher V_{GS} (to detect this, the I_{DS} step size was progressively reduced at higher V_{DS}). This is in agreement with our analytical V_{DSat}-determining procedure, which has a valid root for the polynomial (78) only for large V_{GS}, as seen in Fig. 4-5.

Having calculated a physically meaningful root for c as explained above, β_{2d}, β_{1d} and V_{DS} (=V_{DSat}) can be calculated in order from (70), (58) and (42) respectively. Finding β_{1d} using (58) also involves a numerical solution.

Thus, we can find V_{DSat} for a given V_{GS}. All in all, it involves solving a polynomial equation (iteratively or non-iteratively as described above), and 2 more iterative (Newton-Raphson like) numerical computations (one for calculating β_{1s} for a given V_{GS} using (42), and one for calculating β_{1d} using (58) for a given β_{2d}).

4.4.2 *Method 2*: Introducing a new model parameter V_{GSC}

Solving a fifth or even a fourth order polynomial given by (78) in a compact model implementation is a daunting task. Even if one approximates it as a fourth order polynomial and relies on the closed form exact solutions (as that given by the Ferrari method [69], for instance), there are implementation problems such as which of the four multiple roots to choose. Switching amongst the multiple roots based on root values, essentially means implementing IF statements whose conditional involves expressions that are bias dependent. This is in violation of recommended ways [70] to implement compact models in high-level languages like Verilog-A, because it could result in a discontinuity in the derivatives and hence cause convergence problems. Hence, an alternate method was explored which did not involve solving a high-order polynomial. This is described in this section.

The two terms in the RHS of (71) were examined. A plot of the ratio of the first term to the second term was analyzed for various geometries. It turns out that the second term is much smaller, as can be seen in Fig. 4-6 which was plotted for a L=1 um, W_{fin}=20 nm, T_{ox}=1.5 nm device. It was also tried for much shorter channel length devices (20 nm) and a similar dominance was noticed. We can see from the figure that the first term is very dominant. Hence we proceed by approximating the second term by a greatly simplified term, viz. $\frac{1-2b}{cP_3}$, where P_3 is a sufficiently large parameter. The right axis in Fig. 4-6 also shows the exact V_{DSat} (using the 5th degree polynomial solution), and the approximate V_{DSat} calculation done assuming $P_3=\infty$. We can see a reasonably close match.

Thus, the approximation to (71) that we propose is:

$$\frac{\partial a_{2d}}{\partial a_{1d}} \cong \frac{2b}{c} + \frac{1-2b}{cP_3} \tag{82}$$

The form of the 2nd term in the RHS of this approximation is chosen so that it results in a quadratic equation in c, and also so that the sign of the 2nd term is in agreement with that of the corresponding term in the exact equation. It can be easily seen that a necessary condition (but not sufficient) for the 2nd term in the RHS of (71) to be positive is that (*1-2b*) be positive.

Such a design on the sign is not a required design; we devised it such simply so that a sign constraint can be imposed on the P_3 model parameter that it be positive.

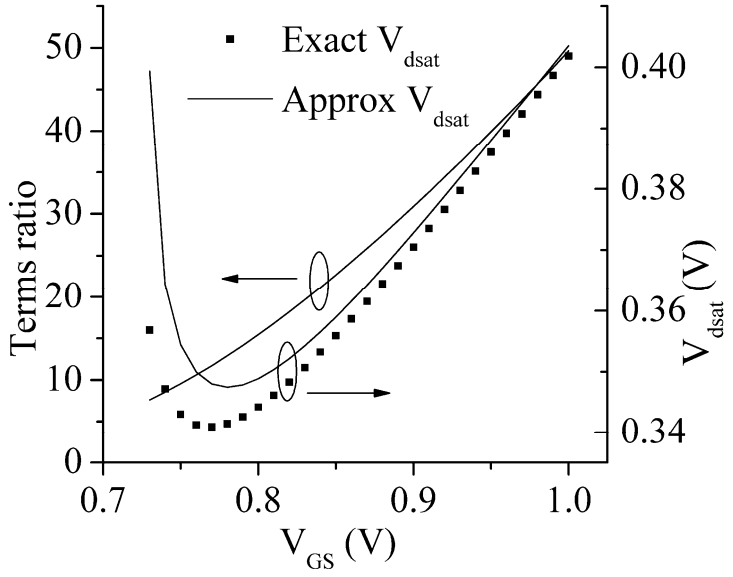


Figure 4-6: V_{DSat} approximation *Method 2*.

Using this approximation, we get a quadratic equation that finally yields a closed-form solution for c as given by (as a replacement to solving the quintic polynomial (78)):

$$c = n - \frac{2b\left[\sqrt{b^2 + kn\left(2b + \frac{1-2b}{P_3}\right)} - b\right]}{k\left(2b + \frac{1-2b}{P_3}\right)} \tag{83}$$

Having found c, the rest of the procedure remains the same to calculate V_{DSat}.

One disadvantage of this method is that there is a minimum V_{GS} below which a physically valid solution to (83) does not exist (corresponding to the constraint that $\beta_{2d} > 0$ to be physically meaningful, where β_{2d} is calculated using (70) after calculating c using (83)). We call this minimum V_{GS} as V_{GSC}. However it should be noted that this conclusion is in agreement with the physical arguments given in the discussion of (80) for this situation (viz. the fact that drift as well as diffusion is considered).

It is not possible to derive a closed-form exact expression for V_{GSC}. So instead we deem V_{GSC} as a model parameter and extract it.

4.4.3 *Method 3*: Considering only drift component

The approach in *Method 2* is not as robust as we would like. For instance, it was found that the validity of the underlying approximations (in the terms analysis) get questionable if oxide thicknesses even slightly away from realistic ones are considered. Besides, it introduced a model parameter V_{GSC}, whose extraction is not very straightforward. Hence an alternate approach was developed which we describe in this section.

To find V_{DSat}, we limit our focus to the drift component. For this, we follow the same approach described earlier except that we only consider the drift component I_{DS}^{drift} (as also done in MOS Model 11 [56] for conventional MOSFETs), and we make the approximation (in the equation equivalent to (53)) that I_{DS}^{drift} is spatially constant, which is a valid approximation in strong inversion since majority of the current is then due to drift. We get an equation similar to (64), but with modified expressions for a_{1d} and a_{2d} as shown below (compare with (65)):

$$\begin{aligned} a_{1d}^{drift} &= \frac{b}{2}\left(\beta_{2s}{}^2 - \beta_{2d}{}^2\right) \\ a_{2d}^{drift} &= \ln\left(\frac{\beta_{2s}{}^2}{\beta_{2d}{}^2}\right) \end{aligned} \tag{84}$$

We then set $\frac{\partial I_{DS}^{drift}}{\partial V_{DS}} = 0$. By doing so, we get:

$$\beta_{2dsat} = \frac{\sqrt{1+2bk\beta_{2s}{}^2}-1}{\sqrt{2bk}} \tag{85}$$

It is clear that β_{2dsat} as calculated using (85) always satisfies the condition $0<\beta_{2dsat}<\beta_{2s}$, implying that V_{DSat} exists for all V_{GS} thus obviating the need for a parameter like V_{GSC}, and also being in agreement with the physical arguments given earlier for this situation (viz. the

fact that only drift is considered). For a given V_{GS}, the quantities β_{1s} and β_{2s} can be calculated in order using (42) and (58) (or using the closed form approximation described later), and (85) can then be solved in closed form for β_{2dsat}, from which V_{DSat} can be calculated in closed form using (43) and (58).

Amongst the three methods, *Method 3* was chosen to be used in the final version of the model, because of its simplicity.

4.5 V_{DSeff} smoothing function

Having found V_{DSat}, a V_{DSeff} was defined as in [54], in order to smoothly vary between the transition regions and limit V_{DS} at V_{DSat} when it exceeds V_{DSat}:

$$V_{DSeff} = V_{DS}\left(1+\left(\frac{V_{DS}}{V_{DSat}}\right)^{A_X}\right)^{-1/A_X} \tag{86}$$

where A_X is a smoothening model parameter that controls the sharpness of the transition of V_{DSeff} between the values of V_{DS} and V_{DSat}.

V_{DS} is then replaced by V_{DSeff} wherever it occurs in the model equations developed until now, so that those equations remain applicable even if V_{DS} exceeds V_{DSat}.

4.6 Channel Length Modulation (CLM)

We next describe the model for CLM in the post-velocity saturation regime for DGFETs. It is on the same lines as the analysis done by Ko et al. in [71] as described in [2], except that we have adapted it to a double-gate structure.

Refer to Fig. 4-7. It shows the velocity saturated region near the drain. Notice that the (x_1, y) coordinate axis has its origin at the start of the velocity saturated region.

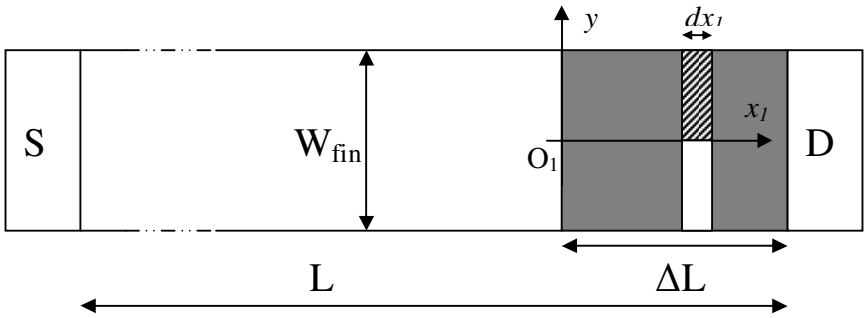


Figure 4-7: Velocity saturated region (shown shaded) in a SDGFET. An elemental vertical strip is shown hatched.

Once $V_{DS} > V_{DSat}$, the carriers are moving at their saturation velocity in a small region near the drain end. The GCA is violated in this region as E_x and E_y are both sizeable and rapidly changing. Consider a small vertical strip (shown hatched) of width dx_1 at a distance x_1 from the origin O_1. Let us make the quasi-2D approximation that E_x is constant along y, and is equal to the x-gradient of the potential at the oxide-silicon interface. That is, let us approximate that:

$$E_x(x_1) = -\frac{\partial \psi(x_1, W_{fin}/2)}{\partial x_1} \tag{87}$$

Applying Gauss Law to the hatched strip, we have:

$$\frac{W_{fin}}{2} E_x(x_1 + dx_1) - \frac{W_{fin}}{2} E_x(x_1) + E_{ys}(x_1) dx_1 = \frac{Q_i(x_1) dx_1}{2\varepsilon} \tag{88}$$

This translates to:

$$W_{fin} \frac{dE_x(x_1)}{dx_1} + 2E_{ys}(x_1) = \frac{Q_i(x_1)}{\varepsilon} \tag{89}$$

This is similar to (24) in [19].

Now, in the velocity saturated region where all the carriers are moving at v_{sat}, Q_i has to be constant with x_1, because $I_{DS}=Q_i(x_1)v_{sat}$ which is the current per unit fin height, is constant. Therefore:

$$Q_i(x_1) = Q_i(0) \tag{90}$$

Now at x_1=0, the GCA is still valid, and we can apply the 1D Gauss Law, which gives:

$$E_{ys}(0) = \frac{Q_i(0)}{2\varepsilon} \tag{91}$$

Using (90) and (91) in (89), we get:

$$W_{fin}\frac{dE_x(x_1)}{dx_1} = 2\left[E_{ys}(0) - E_{ys}(x_1)\right] \tag{92}$$

Using (36), which is valid even if the GCA is not, (92) becomes:

$$W_{fin}\frac{dE_x(x_1)}{dx_1} = \frac{2C_{ox}}{\varepsilon}\left[\psi(0, W_{fin}/2) - \psi(x_1, W_{fin}/2)\right] \tag{93}$$

In (93), $\psi(0,W_{fin}/2)$ is in the O_1 coordinate system. In other words, in the O coordinate system, it is equal to $\psi(L-\Delta L, W_{fin}/2)$. Let us call this ψ_{sat}. It can be got once V_{DSat} is obtained as per the procedure described in the previous section. In that section, we found β_{2d} when finding V_{DSat}. Once β_{2d} is known, then we can use (58), (41) and (36) to find:

$$\psi_{sat} = \psi(L-\Delta L, W_{fin}/2) = V_{GS} - \Delta\phi - \frac{4\varepsilon\phi_t\beta_{2dsat}}{W_{fin}C_{ox}} \tag{94}$$

An alternate approximation for ψ_{sat} is:

$$\psi_{sat} = V_{dsat} + \psi_0 \tag{95}$$

where ψ_0 is the amount by which the electrostatic potential ψ is above the electron quasi fermi potential φ_{fn} in the source end. This is based on the approximation that the electron density is approximately constant with x between O and $O1$ (meaning that $\psi(x, y)$ and $\varphi_{fn}(x)$ are approximately parallel for all y for $0<x<L-\Delta L$).

We have used (95) in the present work. From (93) and (87), we have:

$$\frac{d^2\psi(x_1, W_{fin}/2)}{dx_1^2} = \frac{2C_{ox}}{\varepsilon W_{fin}}\left[\psi(x_1, W_{fin}/2) - \psi_{sat}\right] \tag{96}$$

The boundary conditions for this differential equation are:

$$\begin{aligned} &\psi(0, W_{fin}/2) = \psi_{sat} \\ &\frac{d\psi(0, W_{fin}/2)}{dx_1} = -E_{sat} \end{aligned} \tag{97}$$

where E_{sat} is a model parameter.

With these boundary conditions, the solution to (96) is:

$$\psi(x_1, W_{fin}/2) = \psi_{sat} + lE_{sat}\sinh(x_1/l) \tag{98}$$

$$l = \sqrt{\frac{\varepsilon W_{fin}}{2C_{ox}}} \tag{99}$$

Then, ΔL is got by setting $\psi(\Delta L, W_{fin}/2) = V_{DS} + \psi_0$. We get:

$$\Delta L = l \cdot \ln\left[\frac{V_{DS} - (\psi_{sat} - \psi_0)}{l \cdot E_{sat}} + \sqrt{\left(\frac{V_{DS} - (\psi_{sat} - \psi_0)}{l \cdot E_{sat}}\right)^2 + 1}\right] \tag{100}$$

Using (95) in (100), this simplifies to:

$$\Delta L = l \cdot \ln\left[\frac{V_{DS} - V_{DSat}}{l \cdot E_{sat}} + \sqrt{\left(\frac{V_{DS} - V_{DSat}}{l \cdot E_{sat}}\right)^2 + 1}\right] \tag{101}$$

This is exactly the same as (3.97) in [2], the only difference being the expression (99) for l (compare with (3.95) in [2]) where we see that $W_{fin}/2$ is taking the place of the junction depth term x_j.

In our model implementation, in (101), we replaced the V_{DSat} term with V_{DSeff} as defined in (86), in order to have a non-zero ΔL only when $V_{DS} > V_{DSat}$.

4.7 Model validation

2D device simulations of an N-channel SDGFET were done using Synopsys Sentaurus Device [51]. The device structure was created with abrupt source-body and drain-body junctions. The body was lightly doped at $1x10^{15}$ cm^{-3} p-type, and the source and drain regions were kept short in length and were doped $1x10^{19}$ cm^{-3} n-type. In order to focus on just the mobility degradation due to the lateral field, other models were disabled, such as vertical field mobility degradation, doping dependant mobility, etc. Long channel devices were simulated in order to help do a reasonable comparison with the analytical model, because the analytical model described thus far is not geared to handle effects such as DIBL, etc. For the same reason, recombination-generation models and quantum mechanical models, etc were also turned off.

In the device simulations, the driving field for the Canali model [63] was specified as the gradient of the quasi-fermi-potential instead of the lateral electric field, due to numerical stability issues. Though this is different from what our analytical model does (the analytical model uses the lateral electric field as the driving field), either specification is supposed to give nearly the same result in the device simulations (page 222 in [51]).

The results obtained from these device simulations were then compared with the model-predicted values, as described below.

4.7.1 Using *Method 1* for V_{DSat} calculation

To validate the drain current equations as developed using *Method 1* for V_{DSat} calculation, the device simulations specifications included the following in addition to the ones stated in section 4.7.

A mid-gap gate work function of 4.6eV was used for the gate electrode. The electron mobility and the saturation velocity used by the device simulator were 1430 cm^2/Vs and 1.07x10^7 cm/s respectively. The default exponent of 1.11 (for electrons) was used for the Canali model in the device simulator. Simulations were done for 3 geometries, viz. *(i)* L=1 um, W_{fin}=20 nm, T_{ox}=1.5 nm; *(ii)* L=800 nm, W_{fin}=20 nm, T_{ox}=1.4 nm; and *(iii)* L=600 nm, W_{fin}=10 nm, T_{ox}=1.3 nm. I_d-V_d characteristics for all the three devices and a sampling of the I_d-V_g characteristics for each device at some biases are shown in Fig. 4-8 - 4-13. The analytical model parameters were *extracted* using a Heirarchical Particle Swarm Optimization (HPSO) algorithm based parameter extractor developed at IIT Bombay [72], using the I_d-V_d device simulation data. The 1 um device and the 800 nm device had the same analytical model parameters values extracted, viz. μ_0=1375 cm^2/Vs, $\Delta\varphi$= -0.004V, v_{sat}=0.64x10^7 cm/s, E_{sat}=5.8x10^4 V/cm. The 600 nm device had slightly different parameters extracted, viz. μ_0=1319 cm^2/Vs, $\Delta\varphi$= -0.004V, v_{sat}=0.72x10^7 cm/s, E_{sat}=15x10^4 V/cm.

As can be seen, the analytical versus device-simulation matching is very good.

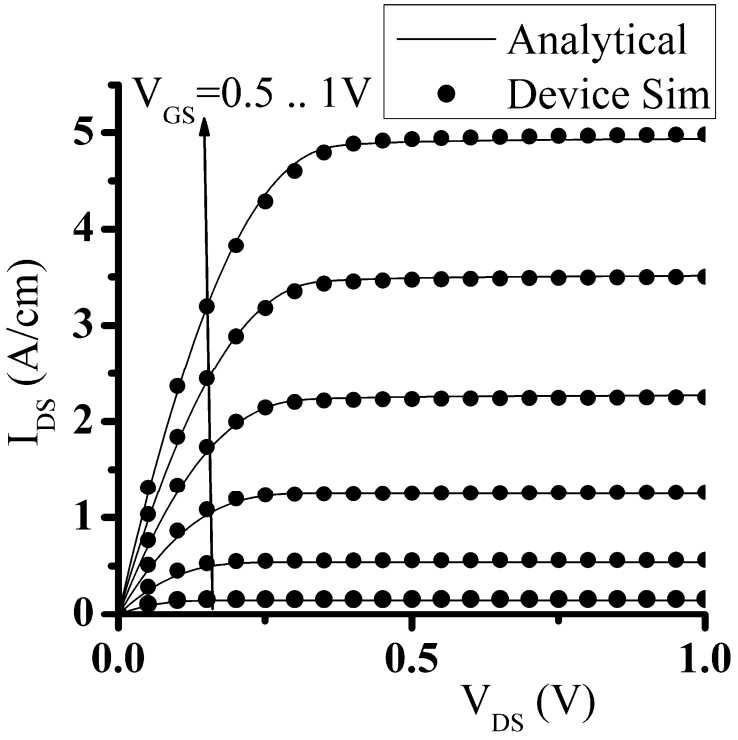


Figure 4-8: Output characteristics for the 1 um device (using *Method 1* for V_{DSat} calculation).

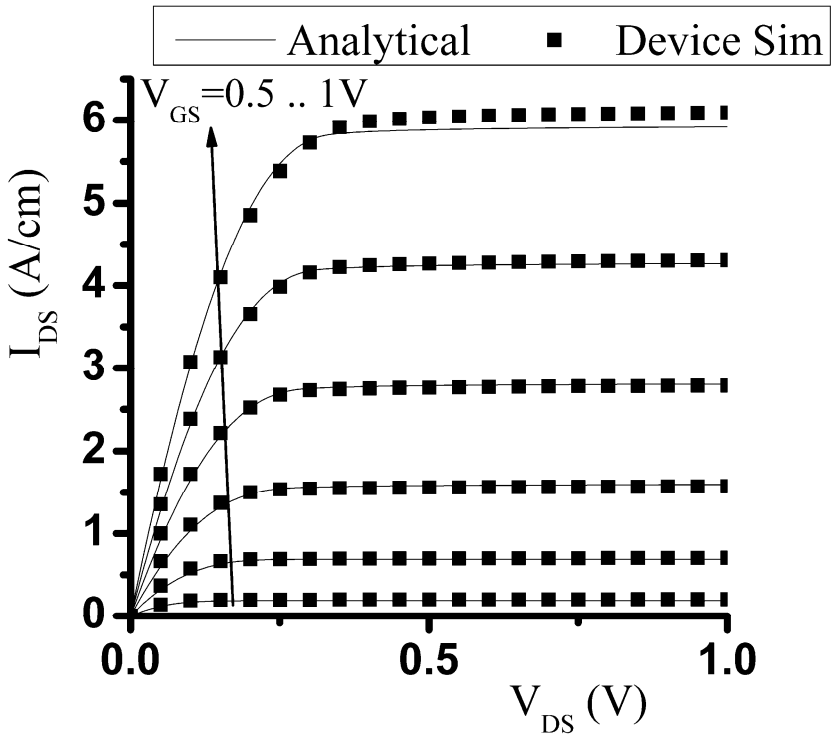


Figure 4-9: Output characteristics for the 0.8 um device (using *Method 1* for V_{DSat} calculation).

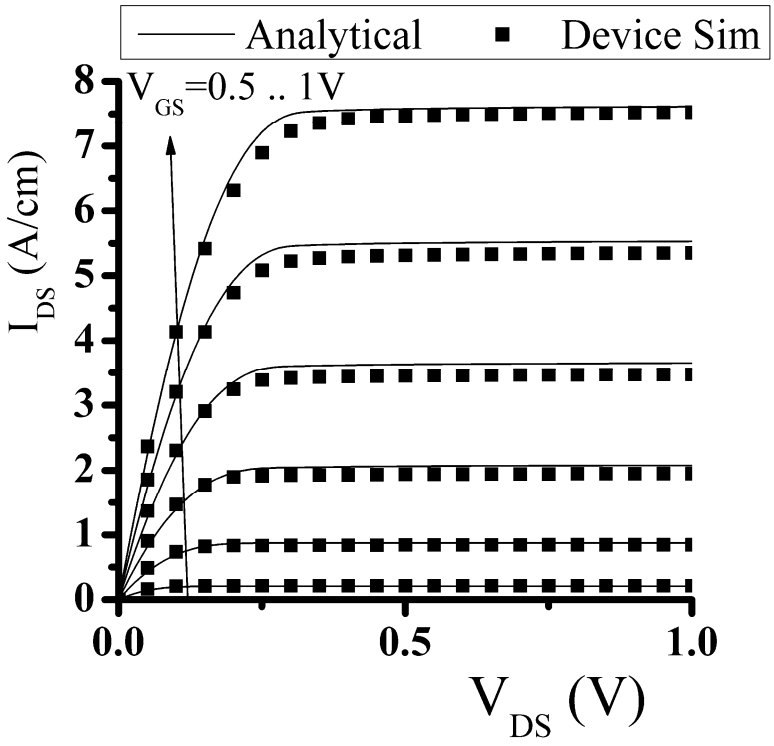


Figure 4-10: Output characteristics for the 0.6 um device (using *Method 1* for V_{DSat} calculation).

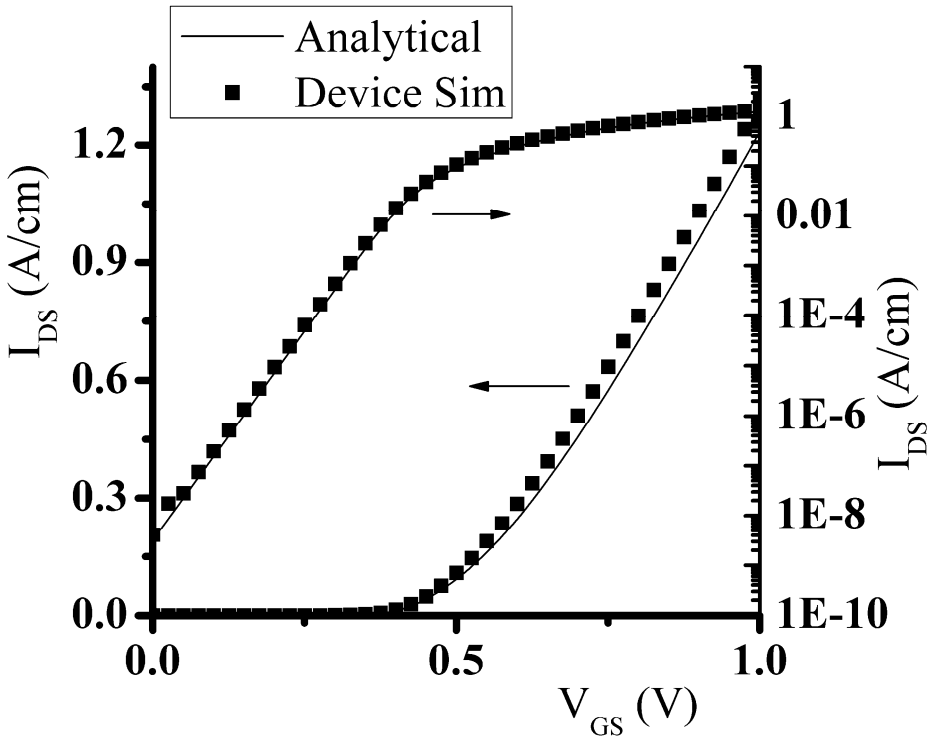


Figure 4-11: Transfer characteristics for the 1 um device at V_{DS}=50mV (using *Method 1* for V_{DSat} calculation).

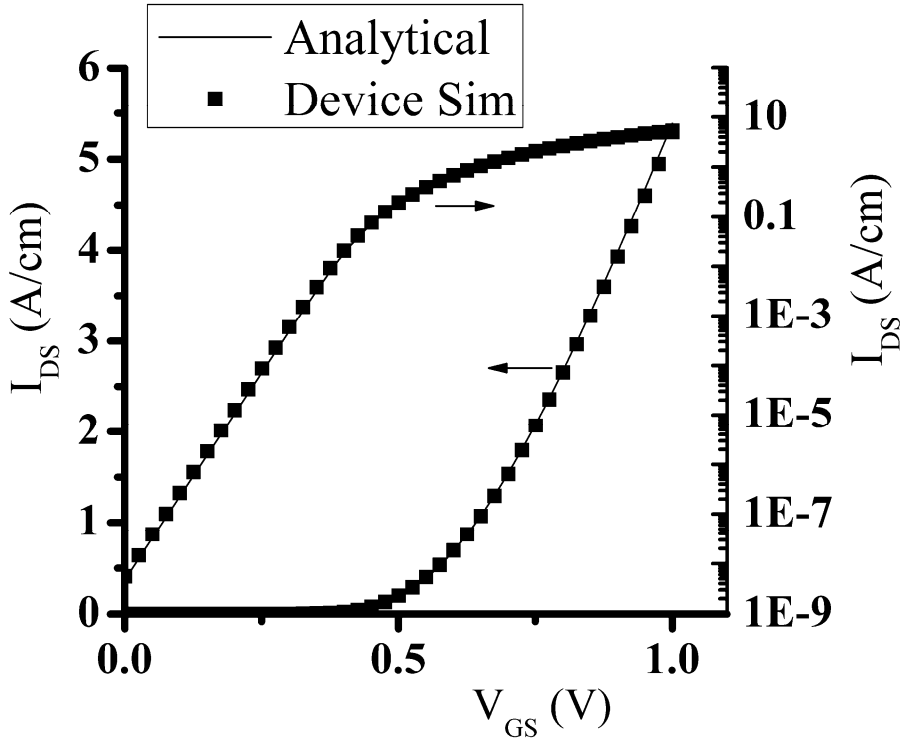


Figure 4-12: Transfer characteristics for the 0.8 um device at V_{DS}=0.24 V (using *Method 1* for V_{DSat} calculation).

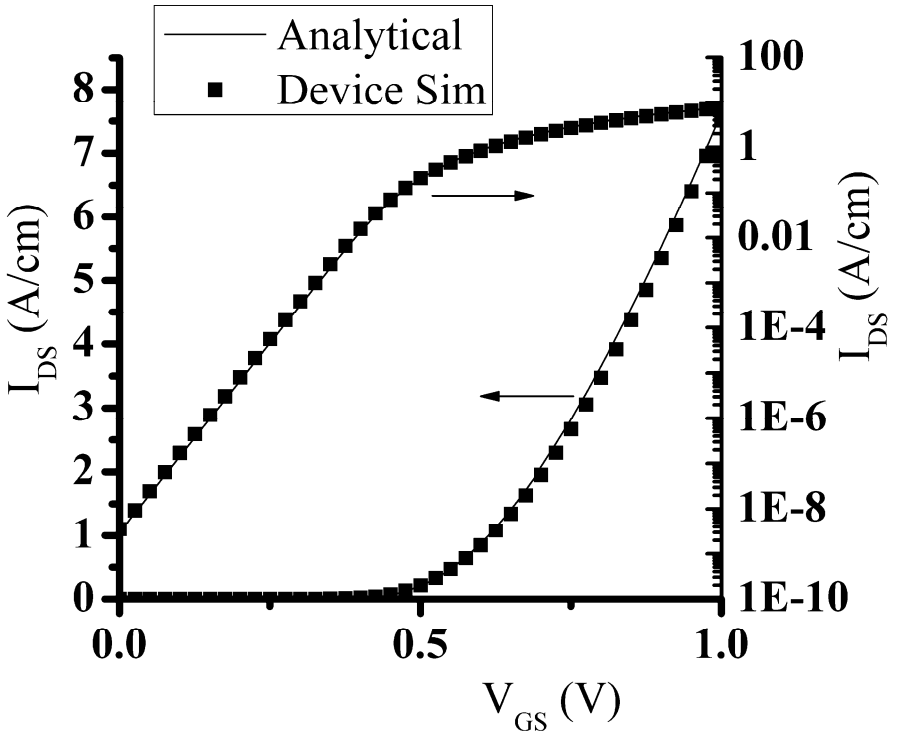


Figure 4-13: Transfer characteristics for the 0.6 um device at V_{DS}=1 V (using *Method 1* for V_{DSat} calculation).

4.7.2 Using *Method 2* for V_{DSat} calculation

The validation of the drain current equations as developed using *Method 2* for V_{DSat} calculation is described in the next chapter which discusses the support for body doping and low-field mobility degradation. This is because the V_{GSC}-based V_{DSat} calculation scheme was developed at a time *after* support for body doping and low-field mobility degradation was added to the core model.

4.7.3 Using *Method 3* for V_{DSat} calculation

To validate the drain current equations as developed using *Method 3* for V_{DSat} calculation, the device simulations specifications included the following in addition to the ones stated in section 4.7.

A mid-gap gate work function with a zero barrier (with respect to intrinsic silicon) was used for the gate electrodes. The electron basal mobility was downgraded to 300 cm^2/Vs in order to emulate realistic vertical-field-degraded mobilities. Default values were used for everything else. Thus the saturation velocity and the Canali exponent used by the device simulator were 1.07×10^7 cm/s and 1.11 respectively.

Simulations were done for 2 channel lengths, viz. *(i)* L=100 nm, W_{fin}=10 nm, T_{ox}=1 nm; and *(ii)* L=200 nm, W_{fin}=10 nm, T_{ox}=1 nm, and the results were compared with the analytical model. A sampling of I_d-V_d, g_{ds}-V_d, I_d-V_g and g_m-V_g characteristics for each device is shown in Fig. 4-14 - 4-17. As can be seen, the model matches very well with TCAD results.

The parameters μ_0, v_{sat} , E_{sat} and A_X were extracted from the corresponding device simulation data using a parameter extraction program developed at IIT Bombay [72]. The extracted values were μ_0=270 cm^2/Vs, $v_{sat}=0.71 \times 10^7$ cm/s, $E_{sat}=4.3 \times 10^6$ V/cm and A_X=2 for the 100 nm device and μ_0=287 cm^2/Vs, $v_{sat}=0.592 \times 10^7$ cm/s, $E_{sat}=9.7 \times 10^6$ V/cm and A_X=2.51 for the 200 nm device.

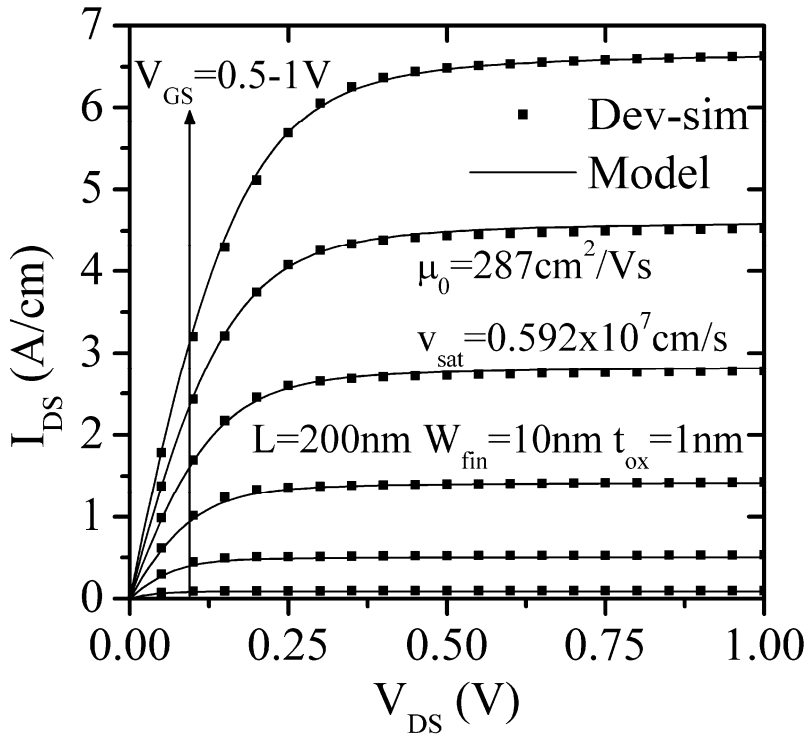


Figure 4-14: Output characteristics for the 200 nm device (using *Method 3* for V_{DSat} calculation).

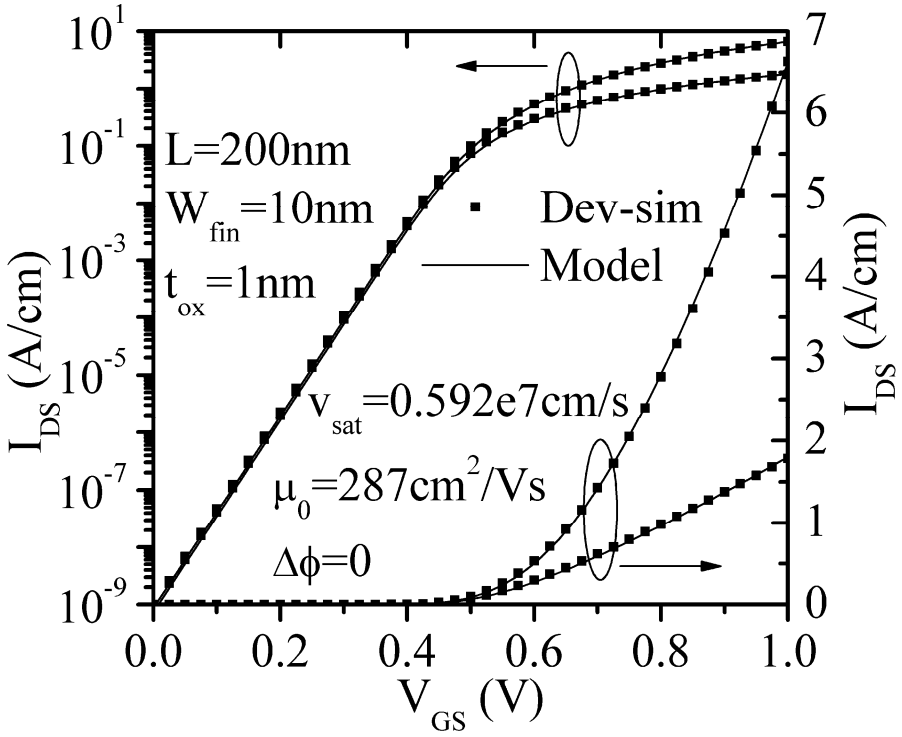


Figure 4-15: Transfer characteristics for the 200 nm device at V_{DS}=50 mV and 1 V (using *Method 3* for V_{DSat} calculation).

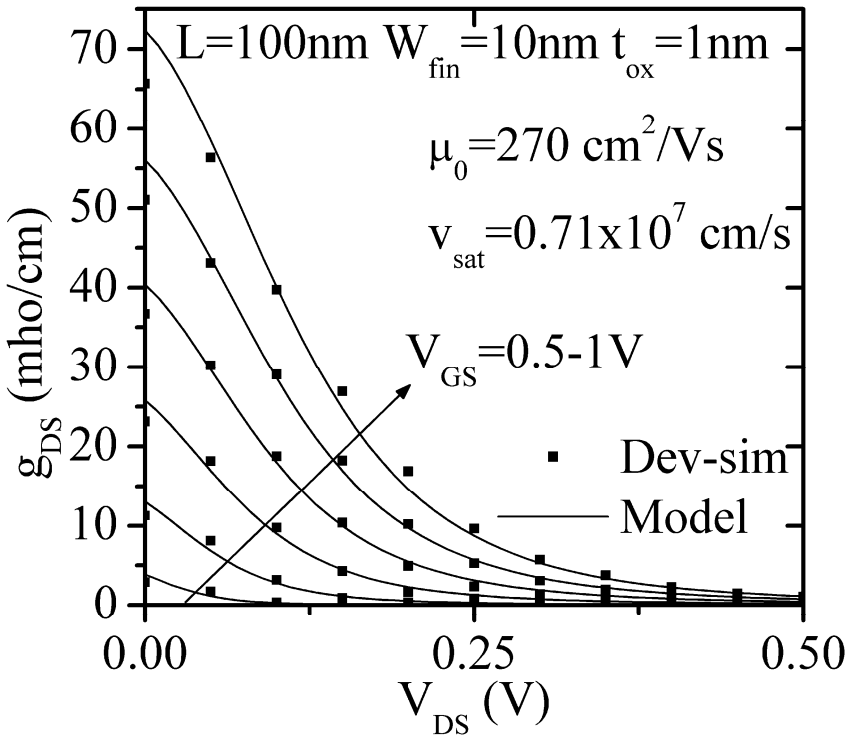


Figure 4-16: Output conductance of the 100 nm device (using *Method 3* for V_{DSat} calculation).

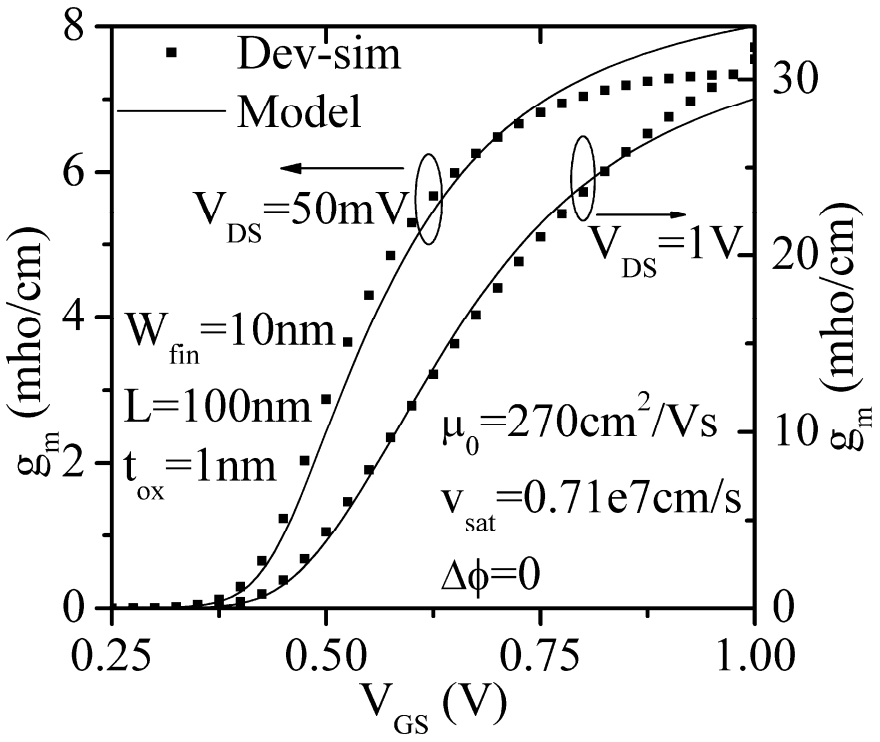


Figure 4-17: Trans-conductance of the 100 nm device at V_{DS}=50 mV and 1 V (using *Method 3* for V_{DSat} calculation).

The extracted values for μ_0, v_{sat} and A_X are used in the respective analytical model curves shown in Fig. 4-14 - 4-17. For the remaining parameters, the analytical model uses a fixed $\Delta\varphi$=0V (same as that used in the device simulations) and a fixed value for the CLM parameter E_{sat}=4.3x10^6 V/cm (viz. the one extracted for the 100 nm device) for both channel length devices. The extracted basal mobilities are thus not far from the value of 300 cm^2/Vs used in the device simulator.

4.7.4 Model benchmarking and Gummel symmetry

The core model developed in this chapter has also been compared with some other existing SDGFET models, and the result is shown in Fig. 4-18. As can be seen, our model improves on other models. In doing this comparison, V_{DS} was *not* limited at V_{DSat} using (86), in order to avoid ambiguities related to the proper value of the parameter A_X to use, when drawing conclusions from the comparison.

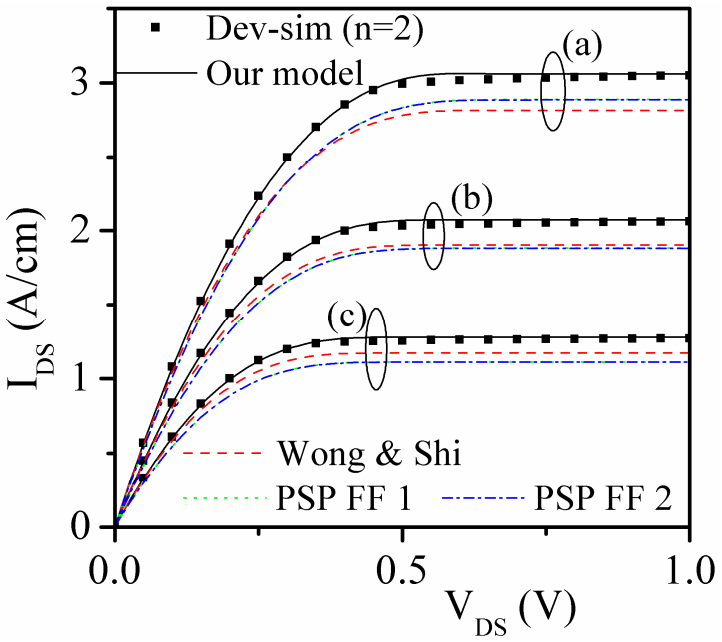


Figure 4-18: Comparison of various models for a L=1 um device.

Also, in the device simulation data shown in Fig. 4-18, the Canali exponent was set to 2 in order to test the accuracy of our model. Most importantly, fitting/extracted parameters were *not* used in the corresponding analytical model traces. Instead the parameter values used in the analytical model were *exactly the same* as those used in the device simulation (v_{sat}, μ_0, etc). As can be seen, the current predicted by our model matches very well with the 2D device simulation results, thus showing that our formulation is in agreement with the stated premise of n=2.

In Fig. 4-18, the curve-groups (a), (b) and (c) are for V_{GS}=1V, 0.9V and 0.8V respectively. The curves labeled Wong & Shi are for the model (equation (20)) from [19]. The curves labeled *PSP FF 1* and *PSP FF 2* use the drain current equation from the PSP FinFET model (equation (28) in [22]). However, the curves labeled *PSP FF 1* use a uniform driving field for velocity saturation (equation (3.44) in [52]), whereas the curves labeled *PSP FF 2* use a linearly varying driving field for velocity saturation (equations (3.45) and (3.46) in [52]). A plot using G_{vat} as defined in PSP 102.2 (equations (4.144) and (4.145) in [73]) was also examined. It is not shown here because it very closely coincides with the curves labeled *PSP FF 1* and *PSP FF 2*.

One of the objectives of our effort was also to ensure that Gummel symmetry is met. A model is said to be Gummel-symmetric in currents if the first and second derivatives of I_{DS} with respect to V_{DS} are differentiable and continuous (with a value of 0) at V_{DS}=0, respectively. It is an important requirement for models used in RF and analog circuit simulations especially for transistors that operate near the V_{DS}=0 point, such as passive mixers and R-2R ladders. We have tested our model for Gummel symmetry using the circuit shown in Fig. 2-6 and the results are shown in Fig. 4-19, where $V_X=V_{DS}/2$ and $I_X=I_{DS}$. As can be seen, our model is compliant with Gummel symmetry requirements.

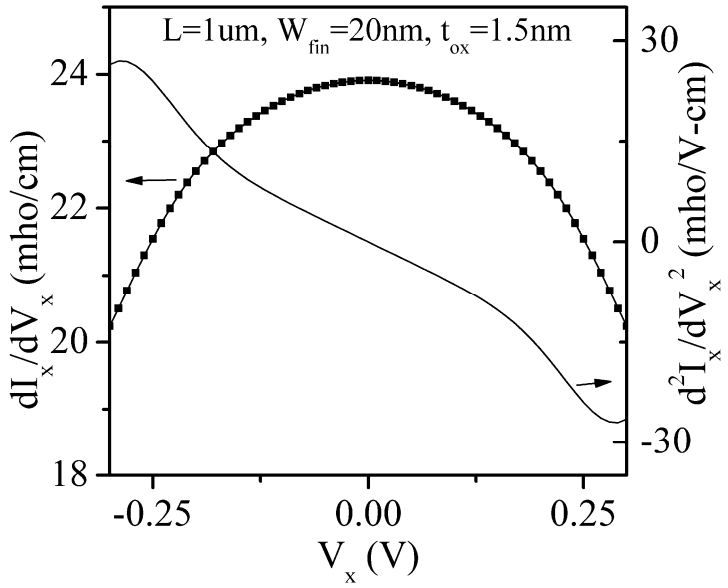


Figure 4-19: Gummel symmetry tests [19] show model symmetry with respect to V_{DS}=0. The symbols are a flipped version of the line.

4.8 Summary

In this chapter, we presented the development of an enhanced core model for the drain current of an SDGFET, that considers velocity saturation effects in a self-consistent manner. We also showed three methods to calculate the drain saturation voltage V_{DSat}. Inclusion of channel length modulation (CLM) effects was also described. Finally, the validation of the model with 2D device simulations, Gummel symmetry compliance and benchmarking with other models of similar capability were presented, where the superiority of our model over other existing models was clearly demonstrated.

Chapter 5
Empirical Corrections, Support for Body Doping and Low-field Mobility Degradation

In this chapter, we describe the way in which we have enhanced the core model by adding support for (uniform) body doping and mobility degradation due to the vertical field (ie. low field mobility degradation), as second order effects. We also describe how a minor empirical correction was added to the core model in order to get a better fit with device simulation results.

5.1 Summary of the core model

The core drain current model described till now considers velocity saturation effects in a self-consistent manner. To summarize, the drain current is given by (102), (103) and (104).

$$I_{DS} = \frac{I_{DS0}}{1 + \frac{I_{DS0}\mu_{eff}a_{2d}}{8C_{ox}v_{sat}{}^{2}L}} \tag{102}$$

$$I_{DS0} = \frac{16\mu_{eff}\varepsilon{\phi_t}^2 a_{1d}}{W_{fin}L} \tag{103}$$

$$\begin{aligned} a_{1d} &= \frac{b}{2}\left({\beta_{2s}}^2 - {\beta_{2d}}^2\right) + \left(\beta_{2s} - \beta_{2d}\right) - \ln\left(\frac{\beta_{2s}+2}{\beta_{2d}+2}\right) \\ a_{2d} &= \ln\left(\frac{b{\beta_{2s}}^2 + \beta_{2s}(2b+1)+1}{b{\beta_{2d}}^2 + \beta_{2d}(2b+1)+1}\right) \end{aligned} \tag{104}$$

where I_{DS0} is the current in the absence of velocity saturation. In the core model, only high field mobility (ie. velocity saturation effects) was considered. Thus I_{DS0} in the core model was described as the current under constant mobility conditions. When *enhancing* the core model to add support for low-field mobility degradation as a *second order correction*, one should interpret I_{DS0} as the drain current in the presence of only low-field mobility degradation. By *second order correction*, we mean a correction that will enable us to use the same functional form for the drain current expression in the *enhanced* model as that in the core model except that some term(s), viz. mobility, is suitably replaced. Accordingly, the term μ_0 as discussed in chapter 4 is replaced by μ_{eff} as can be seen in (102) and (103), where μ_{eff} is the low-field mobility.

Chapter 4 also described closed form expressions for V_{DSat} and channel length modulation effects, which we use as is here except that we replace μ_0 by μ_{eff}.

5.2 Empirical correction

The above described core model was seen to match well for devices up to about L=100 nm. For devices shorter than that, in order to get a good match in I_d-V_d characteristics with TCAD results, we introduce a bias dependence in the saturation velocity v_{sat} by introducing a model parameter θ_{SatG} on the lines of [73], suitably adapted from the realm of surface potential to that of inversion-charge. It is observed that this helps in yielding a better match with TCAD I_d-V_d data. The physical origins of this parameter are not clear, and hence we deem this as an empirical correction.

We define a *normalized* (spatial) average inversion charge areal density q_{im} as $q_{im}=|avg(Q_i)|/(4\Phi_t C_{ox})$, where $|avg(Q_i)|$ is given by (114) and (118). The quantity q_{im} is thus a dimensionless measure of the extent of inversion, and increases with higher gate bias (for an n-channel device). Then, following on the lines of [73] (equation 4.142 in [73]), we define w_{sat} as the value of q_{im} subject to a maximum of 100, which is implemented as $w_{sat} = 100 \cdot q_{im}/(100+q_{im})$. This up-clamping is believed to be done for reasons of numerical stability (else w_{sat} could increase indefinitely as the gate bias is increased in the internal iterations of the simulator). The modified saturation velocity v_{sat2} is then defined as $v_{sat}/(1+ \theta_{SatG} \cdot w_{sat})$ or $v_{sat} \cdot (1- \theta_{SatG} \cdot w_{sat})$, depending on whether θ_{SatG} is positive or negative, respectively. In the core model, v_{sat} is then replaced by v_{sat2}.

5.3 Body doping

Body doping is expected to be low in FinFETs, and is therefore expected to make a difference mainly in the sub-threshold regime. When a body doping of N_a is considered in sub-threshold, (29) in sub-threshold becomes (neglecting the mobile charge term):

$$\frac{\partial^2 \psi^l}{\partial y^2} = \frac{qN_a}{\varepsilon} \tag{105}$$

where the term ψ^l denotes that it pertains to the low V_{GS} regime (sub-threshold). Following the same solution procedure as in chapter 4, we find the potential and electric field to be:

$$\psi^l(x, W_{fin}/2) = A_1(x) + \frac{\beta^2 W_{fin}^2 \phi_t N_a}{4 n_i} \tag{106}$$

$$E_{ys}^l(x) = E_y^l(x, W_{fin}/2) = \frac{-\beta^2 W_{fin} \phi_t N_a}{n_i} \tag{107}$$

By comparing (106) with (30) and comparing (107) with (41), we can approximate a *merged* model for the electrostatic potential and electric field (*merged* in the sense that it is valid in sub-threshold as well as above-threshold) as:

$$\psi(x, W_{fin}/2) = A_1(x) + 2\phi_t \ln\left(\sec\beta_1\right) + \frac{\beta^2 W_{fin}^2 \phi_t N_a}{4n_i} \tag{108}$$

$$E_{ys}(x) = -\frac{4\phi_t \beta_2}{W_{fin}} - \frac{\beta^2 W_{fin} \phi_t N_a}{n_i} \tag{109}$$

As can be seen, these collapse to the correct respective expressions in the extreme cases of sub-threshold and above-threshold. They also collapse to the corresponding expressions developed in chapter 4 if N_a=0. The other piece that needs to be corrected in the model described in chapter 4 to account for body doping is the Gauss Law (37) which becomes:

$$-\frac{Q_i(x)}{2} + \frac{qN_a W_{fin}}{2} = -\varepsilon E_{ys}(x) \tag{110}$$

It is interesting to note that when (110) is evaluated considering (109), the expression (48) for Q_i remains unchanged.

Proceeding then the same way as done in chapter 4, the potential summation rule (36) finally simplifies to a modified *input voltage equation*:

$$f(\beta_1) = 0 = \frac{4\varepsilon\phi_t\beta_1 \tan(\beta_1)}{W_{fin}C_{ox}} + \phi_{fn} + 2\phi_t \ln\left(\frac{2\beta_1 \sec\left(\beta_1\right)}{\beta W_{fin}}\right) + \frac{qN_a W_{fin}}{2}\left(\frac{1}{C_{ox}} + \frac{W_{fin}}{4\varepsilon}\right) - (V_{GS} - \Delta\phi) \tag{111}$$

Compare this with (42) and the effect of N_a is seen as an addition of a *spatially constant* term on the RHS.

Because it is spatially constant, all subsequent development in chapter 4 (such as the drain current expression) remains unchanged. *The net result is that the final drain current equations are unchanged, and only the calculation of β_1 (and therefore β_2) is affected.*

5.4 Low-field mobility degradation

Vertical-field mobility degradation [2,56] (also called low-field mobility degradation, where by the term low-field it refers to low *lateral* field) is due to surface roughness scattering (dominant at high V_{GS}), surface acoustic phonon scattering (dominant at medium V_{GS}), and coulomb scattering (dominant at low V_{GS}). For modeling this we use the same engineering model as that used in the PSP model [7] for conventional MOSFETs but modified suitably for a DGFET for the depletion charge term (which is spatially constant in a UTB device such as a DGFET, and is equal to $qN_aW_{fin}/2$ in half the device), viz.

$$\mu_{eff} = \frac{\mu_0}{1+\left(\frac{\left|E_{eff}\right|}{E_0}\right)^{\theta}+\frac{c_s}{\left(1+\frac{\left|\overline{Q_i}\right|}{qN_aW_{fin}}\right)^2}} \tag{112}$$

where:

$$\left|E_{eff}\right| = \frac{\eta\left|\overline{Q_i}\right|+qN_aW_{fin}}{2\varepsilon} \tag{113}$$

In doing so, we have introduced the following model parameters: μ_0, θ, η, E_0 and c_s.

Note however that in (112) and (113), we have a term $|avg(Q_i)|$. This is a spatial average of Q_i taken along the source-drain direction. The rationale for taking a spatial average is to keep the formulation simple. From (48), we know that:

$$\left|\overline{Q_i}\right| = \frac{8\varepsilon\phi_t\overline{\beta_2}}{W_{fin}} \tag{114}$$

where $avg(\beta_2)$ is given by (118) and is discussed in the next section.

Equations (112), (113), (114) and (118) describe our low-field mobility model.

Such an approach makes μ_{eff} dependent on V_{DS}, albeit weakly (via the calculation of the spatial average of Q_i, due to the presence of the β_{2d} term). If we are to use the V_{DSat} expression derived in the core model as is (regardless of whether via *Method 1*, *2* or *3*), then that would make V_{DSat} depend on V_{DS}, which is very unphysical. A more correct approach is to derive a modified expression for V_{DSat} using the newly introduced dependence in μ_{eff} on V_{DS}. But that makes it difficult to get a closed form expression for V_{DSat}. Therefore, as a simplification, we assume a constant (parameterized) V_{DS} when calculating $|avg(Q_i)|$ for calculating μ_{eff} for the purpose of calculating V_{DSat} (only). This is achieved by defining:

$$\beta_{2d}^{VDSat} = P_1 \beta_{2s} \tag{115}$$

where $\beta_{2d}{}^{VDSat}$ is the value used in place of β_{2d} in (118) when calculating $|avg(Q_i)|$ for calculating μ_{eff} for the purpose of calculating V_{DSat}. Note that P_1 is a model parameter (between 0 and 1) that we have introduced. For all other purposes, when μ_{eff} needs to be calculated (to calculate I_{DS0} for instance), the actual β_{2d} is used.

It is interesting to note that the V_{DSat} calculation in MOS Model 11 [56] seems to have a dependence on V_{DS} (via the $\Delta\psi$ term in (H.18) and (3.44) in [56]), which is very unphysical.

We describe next the development of an expression for $|avg(Q_i)|$.

5.4.1 Spatial average of β_2 (and therefore Q_i)

We derive in this section an expression for the spatial average of β_2, which is needed for use in (114)). We assume constant mobility for purposes of this calculation, and we will justify this approach in a non-rigorous manner at the end of this section. Eliminating $a_1(x)$ in (59) and (60) (after neglecting the sole logarithm term in the expression for $a_1(x)$ in (59)) and considering constant mobility (meaning $v_{sat} = \infty$), we get:

$$\beta_2(x) = \frac{1}{b}\left[\sqrt{\left(b\beta_{2s}+1\right)^2 - \frac{bW_{fin}I_{DS0}(x+L/2)}{8\mu_0\varepsilon\phi_t{}^2}} - 1\right] \tag{116}$$

We then define the spatial average of β_2 as:

$$\overline{\beta_2} = \frac{1}{L}\int_{-L/2}^{L/2} \beta_2(x)\cdot dx \tag{117}$$

Equation (117) can easily be evaluated using (116) and the expressions (63) and (65) for I_{DS0}, and we finally get an expression for the spatial average of β_2 as (*with no approximations involved*):

$$\overline{\beta_2} = \frac{2b\left(\beta_{2s}^2 + \beta_{2d}^2 + \beta_{2s}\beta_{2d}\right) + 3\left(\beta_{2s} + \beta_{2d}\right)}{3\left[b\left(\beta_{2s} + \beta_{2d}\right) + 2\right]} \tag{118}$$

It can be easily seen this spatial average becomes equal to β_{2s}, when $\beta_{2s} = \beta_{2d}$. We also see that a mobility term does not appear in (118), thus lending credibility to the assumption of constant mobility we made earlier in this section.

Equation (118) is used in the calculation of effective mobility due to vertical field mobility degradation.

5.5 Model validation

2D device simulations were done using Synopsis Sentaurus Device [51] assuming abrupt source-body and drain-body junctions. In order to enable Coulomb scattering in the mobility calculation, the University of Bologna mobility model [74] was used instead of the default Lombardi model. Simulations were done for 2 geometries, viz. *(i)* L=0.8 um, W_{fin}=20 nm, T_{ox}=1.4 nm; and *(ii)* L=0.4 um, W_{fin}=10 nm, T_{ox}=1 nm. For the 0.8 um device, simulations were done for 4 different body dopings, viz. $1x10^{16}$, $1x10^{17}$, $1x10^{18}$ and $2x10^{18}$ cm^{-3}. The device simulation results were used as virtual experimental data, and model parameters were extracted from it using a parameter extraction program developed at IIT Bombay [72]. Fig. 5-1 - 5-7 show a sampling of the matching plots obtained by using the model enhancements stated in this chapter. These plots were generated using *Method 2* for the closed form calculation of V_{DSat} described in chapter 4.

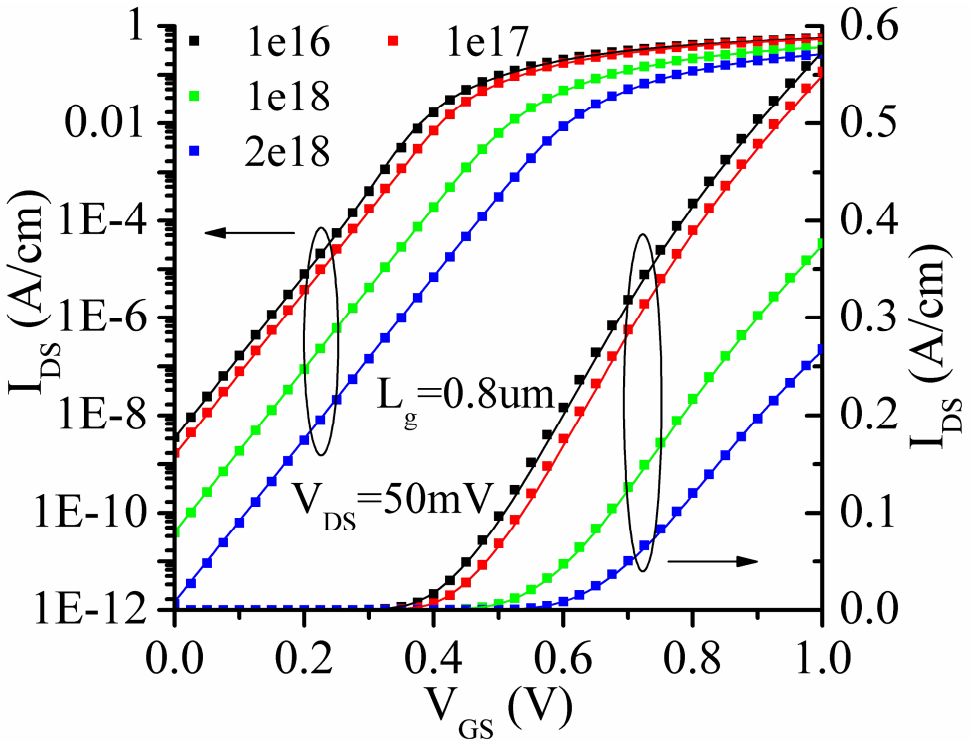


Figure 5-1: I_dV_g curves for the 0.8 um device for various body dopings. The dots are device simulation data, the lines of the corresponding color are model data.

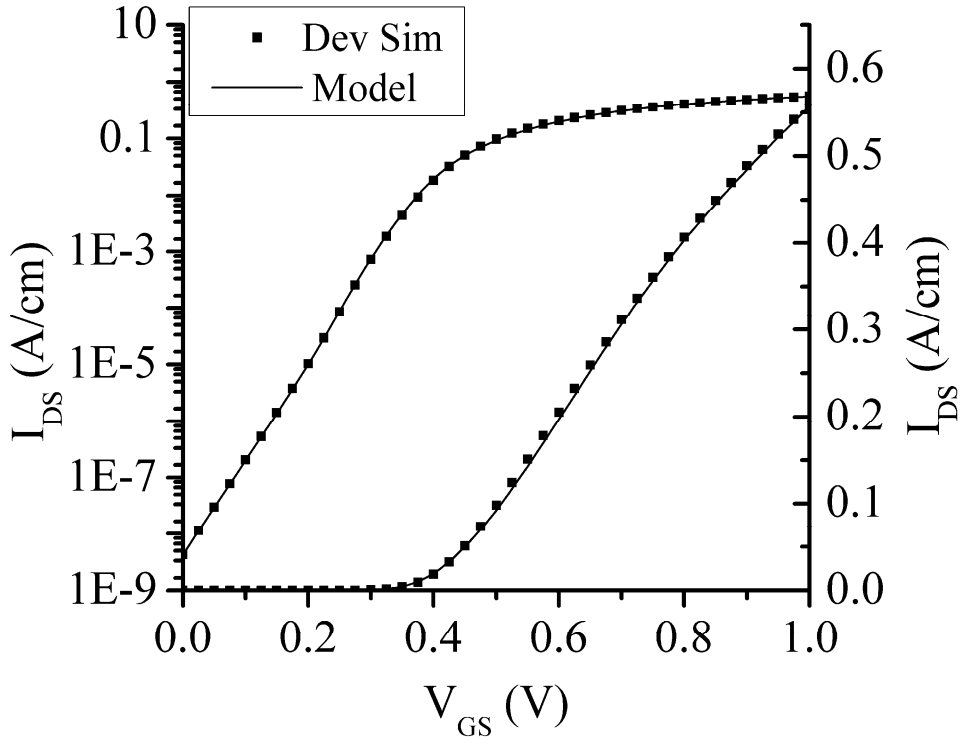


Figure 5-2: Transfer characteristics for the 0.8 um device (N_a=1x10^{15} cm^{-3}) at V_{DS}=50 mV (using *Method 2* for V_{DSat} calculation).

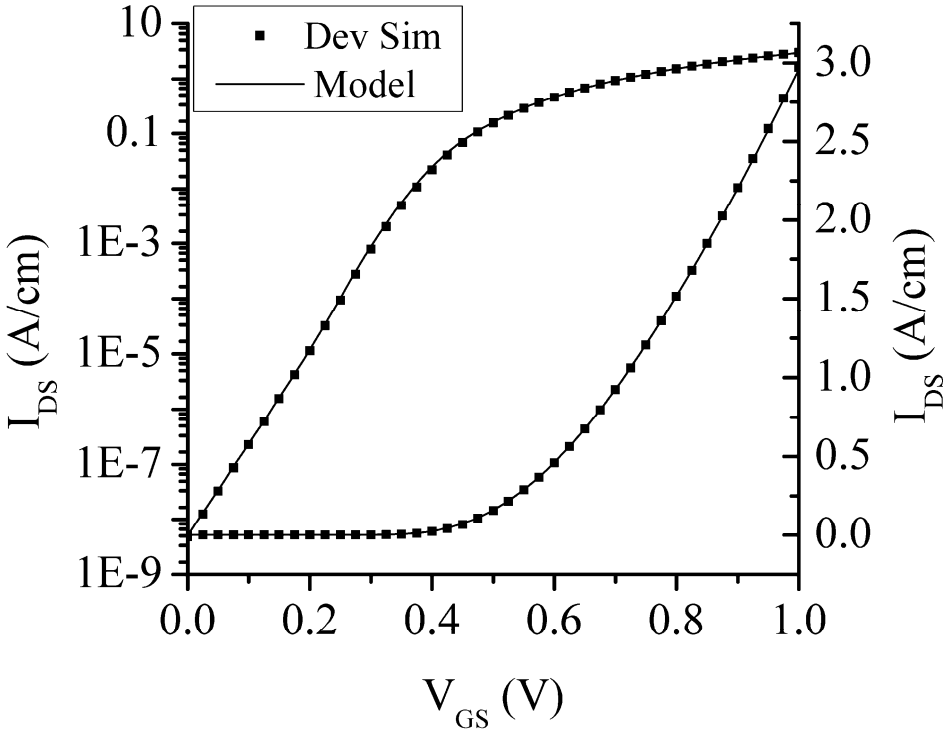


Figure 5-3: Transfer characteristics for the 0.8 um device (N_a=1x10^{15} cm^{-3}) at V_{DS}=1 V (using *Method 2* for V_{DSat} calculation).

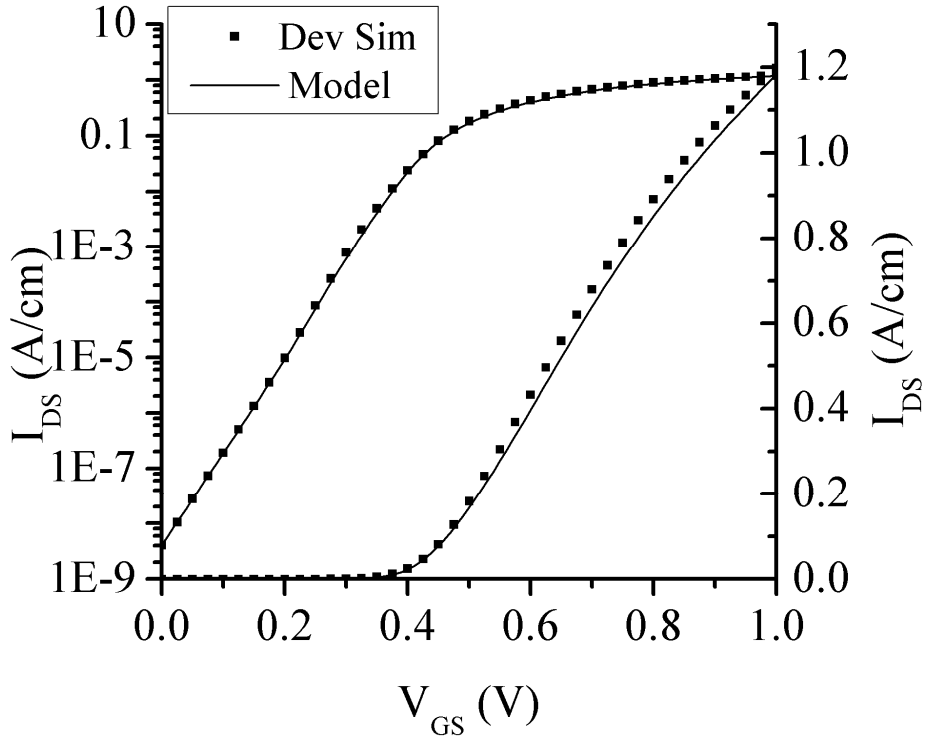


Figure 5-4: Transfer characteristics for the 0.4 um device (N_a=1x10^{15} cm^{-3}) at V_{DS}=50 mV (using *Method 2* for V_{DSat} calculation).

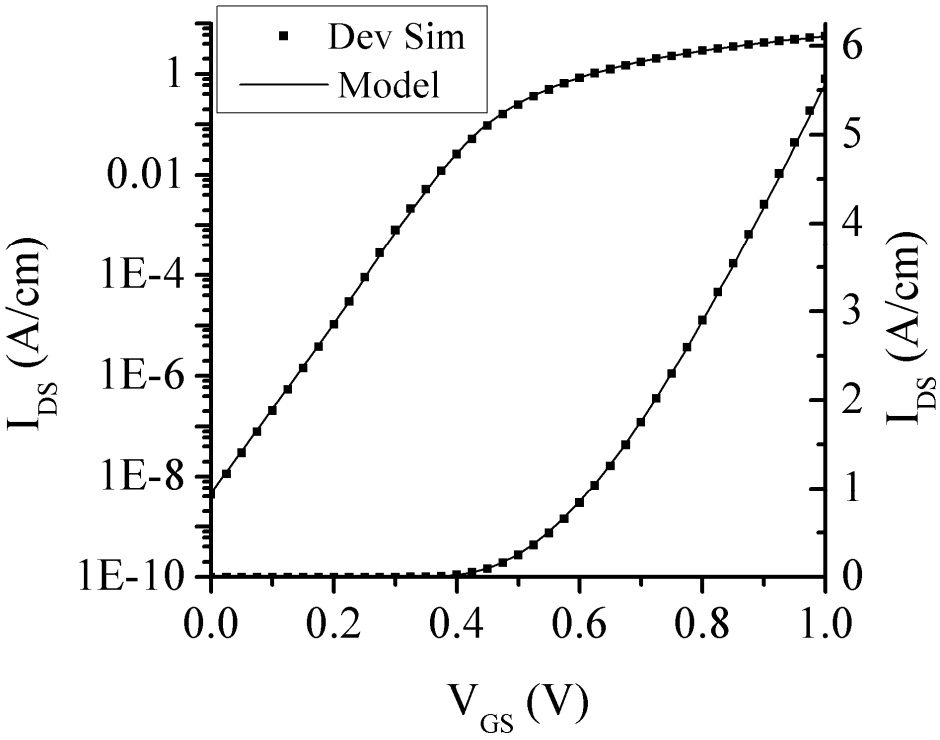


Figure 5-5: Transfer characteristics for the 0.4 um device (N_a=1x10^{15} cm^{-3}) at V_{DS}=1 V (using *Method 2* for V_{DSat} calculation).

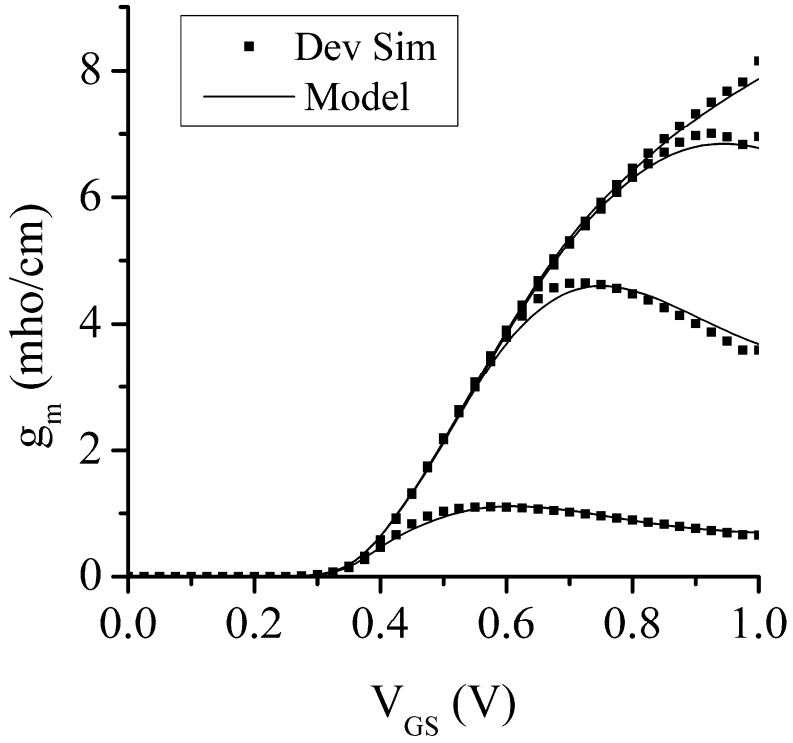


Figure 5-6: Trans-conductance of the 0.8 um device (N_a=1x10^{15} cm^{-3}) at V_{DS}=50 mV, 0.24 V, 0.43V and 1 V (using *Method 2* for V_{DSat} calculation).

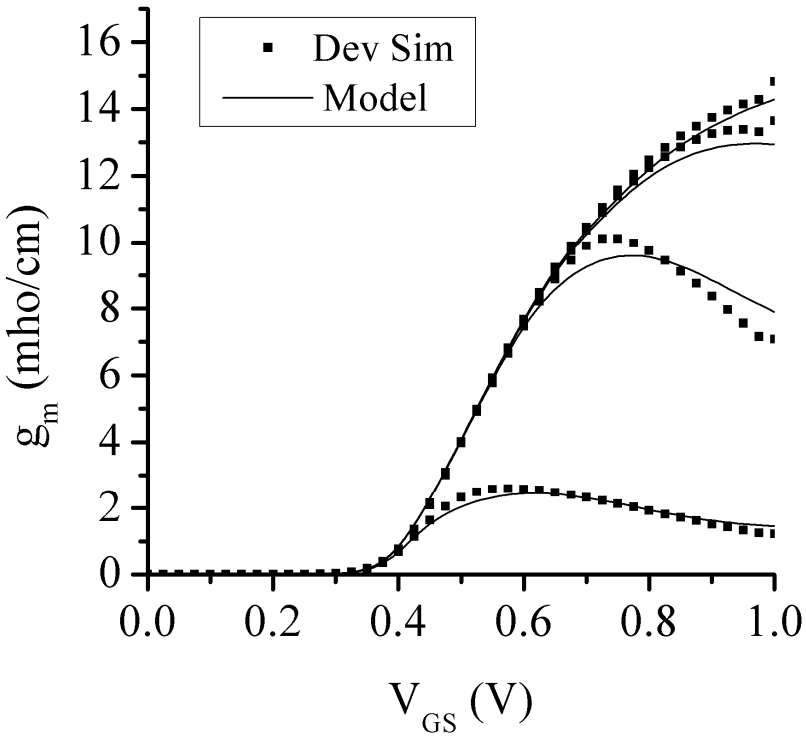


Figure 5-7: Trans-conductance of the 0.4 um device (N_a=1x10^{15} cm^{-3}) at V_{DS}=50 mV, 0.24 V, 0.43V and 1 V (using *Method 2* for V_{DSat} calculation).

Amongst the heretofore 15 model parameters, 4 were set to known values (N_a, W_{fin}, t_{ox} and θ_{SatG}) and the others were *extracted* in a 3-step fashion from I_D-V_G (including g_m-V_G) and I_D-V_D TCAD data for generating the model data shown in Fig. 5-1 - 5-7. The parameter θ_{SatG} was set to 0, since these are relatively longer channel length devices. As can be seen, the match between TCAD predicted values and model predicted values are very good. Also, as physically expected, I_{ON} decreases and V_t increases, as the body doping increases. The g_m-V_g curves exhibit a downward trend at high V_{GS} that is more severe at low V_{DS}, which is typical of low-field mobility degradation.

5.6 Summary

In this chapter, we described how our core model presented in chapter 4 was enhanced in order to support body doping and low-field mobility degradation. It was shown that supporting body doping results in only a change to the *input voltage equation* and that the

drain current expression remains unchanged. A minor empirical correction was also added in order to get a better match with I_d-V_d TCAD device simulation data for short channel devices.

Finally, validation of the model with 2D device simulations was also presented, for two different device geometries and for four different body doping values.

Chapter 6
Closed-form Approximate Solution to the *Input Voltage Equation*

The core model presented until now features a transcendental *input voltage equation* (42) relating the value of an intermediate variable β_1 to the gate and drain voltages. This equation can only be solved numerically by means of iterative techniques. But in order to implement the model in a circuit simulator, it is desirable to have closed form expressions. In this chapter, we present a closed form approximation for the inversion charge areal density. In [65,66], the same goal was addressed, albeit differently; viz. a closed form approximation for β_1 was developed using which one can then calculate $\beta_1\tan(\beta_1)$, which is proportional to the inversion charge areal density. In the present work, we bypass solving for β_1, and instead directly solve for the inversion charge areal density.

6.1 Drawbacks of existing approximations

The approximations described in [65,66] as well as our proposed approximation described in the next section were implemented in Scilab [67], and investigated for accuracy, numerical robustness, and computation expense in terms of CPU Time. The investigation was carried out on an AMD Opteron 2.4 GHz Linux-based PC featuring 16GB RAM. The CPU Time of each approximation was determined using the *timer()* function in Scilab. The free variable $(V_{GD}-\Delta\varphi)$ (see (120) in the next section) was swept from -0.34V to 4V in steps of 0.01V, and the candidate approximation procedures were executed at each bias point (implying 435 executions). The odd value of -0.34V was chosen because that is the lowest bias point at which the approximations discussed in both [65] as well as [66] work (see the column titled *Numerical Robustness* in Table 6-1). Table 6-1 summarizes the results.

Table 6-1: Benchmarking of the various approximations.

Approximation	**CPU Time (a.u.)**	**Numerical robustness: Does not work if $(V_{GD}-\Delta\varphi)$ is less than**	**Accuracy (Peak, Average error% in inversion charge areal density when compared to numerically calculated values)**
Morris et al. [65]	0.62	-0.34V	9.36%, 1.7%
Yu et al. (Taur Group) [66]	0.24	-1.34V	10^{-10}%, 10^{-12}%
Our β_2 approx	0.14	Works at all biases	4.56%, 0.36%
Taur-Modified	0.26	Works at all biases	5×10^{-5}%, 2×10^{-5}% (If the bias range is limited to 1V, then the peak/average errors are 10^{-9}% and 10^{-11}% respectively)

For bias ranges below -0.34V, the approximation described in [65] gives a *divide-by-zero* error. The root cause of this problem was diagnosed to be the computation of ζ in (13) in [65], which gets calculated as 0 when z is very small (as small as 10^{-15}), which happens when $(V_{GD}-\Delta\varphi)$ is less than about -0.34V. As a result, a divide-by-zero error occurs in the computation of $\varphi(z)$ in (13) in [65]. Another drawback of the approximation described in [65] is the piece-wise computation of the Lambert-W function [59], where one needs to switch between equations (14) and (15) in [65], depending on the bias point. Such bias-dependent break-points of a piece-wise computation are not desirable [70] in a compact model.

Similarly, for bias ranges below -1.34V, the approximation described in [66] gives a *singularity of log function* error. The root cause of this problem was diagnosed to be the computation of η_0 in (14) in [66]. When $(V_{GD}-\Delta\varphi)$ is less than about -1.34V, the second term in the square root in (13) in [66] gets neglected in comparison to the first term, and z_1 gets computed as 0. As a result, the computation of η_0 in (14) in [66] ends up evaluating $ln(0)$, thereby giving rise to the singularity.

The numerical robustness was investigated for many geometries (W_{fin} and t_{ox} combinations), and the limit-points of -0.34V and -1.34V of the approximations described in [65] and [66] respectively, are found to be quite typical for various geometries (deviating just marginally from -1.34V to -1.35V at the most for a different geometry).

The limit-point of $(V_{GD}-\Delta\varphi)$ = -1.34V of the approximation described in [66] could be a problem, for instance, for a mid-gap gate electrode with $\Delta\varphi$=0, V_{GS}=0V and V_{DS}=1.4V, which is just a bit above operating voltages in most (though not all) applications, and can very well be encountered in the simulator iterations even if the operating voltages are less than 1V. If the gate electrode work function is closer to the conduction band edge, then $\Delta\varphi$ increases and that further limits the maximum allowed V_{DS} to values less than 1V.

Using similar considerations, it can easily be seen that the limit-point of $(V_{GD}-\Delta\varphi)$ = -0.34V of the approximation described in [65] presents a far greater problem.

Our proposed approximation (labeled *Our β_2 approx* in Table 6-1, and described in detail in the next section) doesn't have any such numerical issues, and as can be seen from Table 6-1, is the most computationally efficient procedure. Its accuracy (when compared to values

calculated by numerically solving the original transcendental equation) is not the best, but as we will show in this chapter, it is as acceptable as the other candidate approximations when one considers its accuracy with respect to values obtained from TCAD device simulations. Besides, our proposed approximation is suited for use in inversion-charge based compact models, where the accuracy requirements are not as severe as those in surface-potential based compact models [64] because quantities of interest like drain current, etc are linearly or quadratically dependant on the key state variable (inversion-charge density) in the former whereas they are exponentially dependant on the key state variable (surface-potential) in the latter (the approximations described in [65] and [66] are inherently targeted for use in surface-potential based compact models).

In Table 6-1, the approximation labeled *Taur-Modified* is our suggested enhancement to the approximation described in [66] to overcome its numerical robustness limitation, if the accuracy of our core approximation is not acceptable in some cases. It is described at the end of the next section.

6.2 The proposed approximation

As described earlier, when one solves the 1D Poisson equation under the Gradual Channel Approximation and writes the potential summation rule traversing from the gate electrode to the source electrode, one gets the transcendental equation shown below as equation (119). This is the equation that this chapter is focused on solving using approximation techniques:

$$f(\beta_1) = 0 = \frac{4\varepsilon\phi_t\beta_1\tan(\beta_1)}{W_{fin}C_{ox}} + \phi_{fn} + 2\phi_t \ln\left(\frac{2\beta_1\sec(\beta_1)}{\beta W_{fin}}\right) - (V_{GS} - \Delta\phi) \qquad (119)$$

Equation (119) can also be written as:

$$\frac{4\varepsilon\phi_t\beta_1\tan(\beta_1)}{W_{fin}C_{ox}} + 2\phi_t \ln\left(\frac{2\sqrt{\beta_1^2 + \beta_1^2\tan^2(\beta_1)}}{\beta W_{fin}}\right) = V_{GS} - \Delta\phi - \phi_{fn} \qquad (120)$$

Using (58), (120) can be written as:

$$\frac{4\varepsilon\phi_t\beta_2}{W_{fin}C_{ox}} + 2\phi_t \ln\left(\frac{2\sqrt{\beta_1^2+\beta_2^2}}{\beta W_{fin}}\right) = V_{GS} - \Delta\phi - \phi_{fn} \tag{121}$$

Also, the inversion charge areal density expression (48) becomes:

$$Q_i = \frac{-8\varepsilon\phi_t\beta_2}{W_{fin}} \tag{122}$$

It is to be noted that in (121), the first term in the LHS dominates the second term to some extent at large values of β_1, and the second term (i.e. the logarithm term) dominates the first term overwhelmingly at small values of β_1. This is seen in Fig. 6-1, where we numerically calculate and plot the (absolute value of the) ratio of the first to the second term in the LHS of (121), versus β_1.

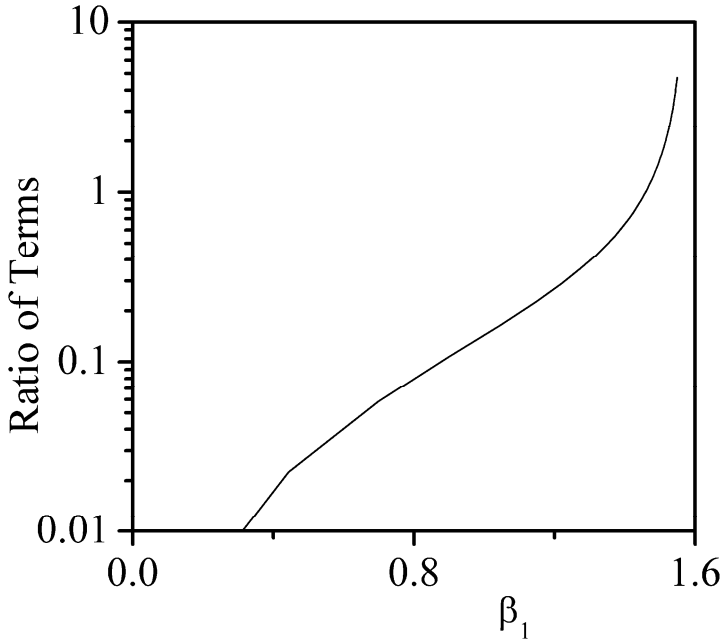


Figure 6-1: Absolute value of the ratio of the first to the second term in the LHS of (121), versus β_1.

But at small values of β_1, we have $\beta_1^2 \cong \beta_1 \tan\beta_1 = \beta_2$. Using this approximation, (121) becomes:

$$\frac{4\varepsilon\phi_t\beta_2}{W_{fin}C_{ox}} + 2\phi_t \ln\left(\frac{2\sqrt{\beta_2+\beta_2^{\,2}}}{\beta W_{fin}}\right) = V_{GS} - \Delta\phi - \phi_{fn} \tag{123}$$

Note that (123) is the same as (43), and is equivalent to (9) in [19]. This simplifies to:

$$\beta_2^{\,2} + \beta_2 = u^2 e^{-2b\beta_2} \tag{124}$$

if we define u as:

$$u = \frac{\beta W_{fin}}{2} e^{(V_{GS}-\Delta\phi-\phi_{fn})/2\phi_t} \tag{125}$$

and b is given by (61).

Now, at *small* values of β_2, (124) can be approximated as (by neglecting the $\beta_2^{\,2}$ term in the LHS and approximating the exponential term in the RHS as 1):

$$\beta_{2\_sm} = u^2 \tag{126}$$

where the *_sm* postfix in the subscript of β_2 indicates that it corresponds to *small* values of β_2.

Similarly, at *large* values of β_2, (124) can be approximated as (by neglecting the β_2 term in the LHS):

$$\beta_{2\_lrg}{}^2 = u^2 e^{-2b\beta_{2\_lrg}} \tag{127}$$

where the *_lrg* postfix in the subscript of β_2 indicates that it corresponds to *large* values of β_2. Equation (127) can be simplified as follows:

$$\beta_{2\_lrg} = ue^{-b\beta_{2\_lrg}}$$
$$b\beta_{2\_lrg}e^{b\beta_{2\_lrg}} = bu$$
$$b\beta_{2\_lrg} = W(bu)$$

$$\beta_{2\_lrg} = \frac{W(bu)}{b} \tag{128}$$

where $W(x)$ is the Lambert-W function [59] defined as the solution of $We^{W} = x$. Since the argument bu in (128) is real and positive, we are interested in the principal branch of the Lambert-W function, in the range where $x>0$.

We now utilize the following closed form approximation [75] for the Lambert-W function that is valid as long as the argument is not too large:

$$W(x) \cong 0.04 + 0.665\left[1 + 0.0195 \cdot \ln(1+x)\right] \cdot \ln(1+x) \tag{129}$$

This has been shown [75] to be valid for $x<500$. Using (129) in (128), we get:

$$\beta_{2\_md} \cong \frac{0.04 + 0.665\left[1 + 0.0195 \cdot \ln(1+bu)\right] \cdot \ln(1+bu)}{b} \tag{130}$$

One of the approximations used in arriving at (130) (viz. neglecting β_2 in comparison to β_2^2) is actually more compellingly valid for large values of β_2 (as opposed to only medium values), but we still choose to label this as being applicable to medium values of β_2, because the approximation that we have used for the Lambert-W function (viz. equation (129)) is not applicable for very large values of its argument.

It should be noted that the notion of *small*, *medium* and *large* used in this work is only to suggest their relative magnitudes. Our proposed approximation does *not* suggest using a piece-wise computation where the pieces are bias dependent. Indeed, the same, entire computation procedure is followed for all biases.

A plot of β_{2_sm} and β_{2_md} shows that they differ in orders of magnitude, with $\beta_{2\_sm} < \beta_{2\_md}$ for small values of β_2, and $\beta_{2\_md} < \beta_{2\_sm}$ for medium values of β_2. Hence, in order to be valid in a wider range of β_2 (viz. small as well as medium values), they can be conveniently combined as (similar to a parallel combination of resistors):

$$\beta_{2\_sm\_md} = \frac{\beta_{2\_sm} \cdot \beta_{2\_md}}{\beta_{2\_sm} + \beta_{2\_md}} \tag{131}$$

This is a computationally efficient smoothing function when uniting two quantities of vastly differing orders of magnitude. This gives a single-piece approximation for β_2 that is valid for small *and* medium values of *bu*, but not for large values of *bu*.

We shall now focus on large values of *bu*. The approximation for the Lambert-W function in (129) is not valid for large arguments, so we will use a different approximation for it for large values of *bu*.

Now, the Newton-Raphson iteration equation for the Lambert-W function can easily be derived as:

$$W_{k+1} = \frac{W_k{}^2 + xe^{-W_k}}{W_k + 1}$$

It is known [76] that for arguments larger than 10, an initial guess of $W_0 = \ln(x) - \ln\left[\ln(x)\right]$ gives a good approximation for the Lambert-W function in very few iterations. Indeed, in our approach, we use only one iteration. Using the suggested initial guess W_0 and a single iteration, an approximation for the Lambert-W function for large values of the argument x comes out to be:

$$W_{lrg}(x) \cong \frac{\left\{\ln(x) - \ln\left[\ln(x)\right]\right\}^2 + \ln(x)}{1 + \ln(x) - \ln\left[\ln(x)\right]} \tag{132}$$

Using (128) and (132), we then get

$$\beta_{2\_lrg} = \frac{\left\{\ln(x) - \ln\left[\ln(x)\right]\right\}^2 + \ln(x)}{b \cdot \left\{1 + \ln(x) - \ln\left[\ln(x)\right]\right\}} \tag{133}$$

where x is the same as *bu* except that it is down-clamped at a value of 10. For this, we can define x as:

$$x = bu\left[1+\left(\frac{10}{bu}\right)^{n_1}\right]^{(1/n_1)} \tag{134}$$

where n_1 is a sufficiently large number. The form of (134) (and (135)) is motivated by the way V_{DSeff} is up-clamped at V_{DSat} in the SP model [54]. This smoothing function is different from the one used in (131) because here we need to unite two quantities that could be comparable in magnitude. A simple plot of β_{2_lrg} as obtained from (133) and (134), and of $\beta_{2\_sm\_md}$ as obtained from (131) for large values of bu, reveals that $\beta_{2\_lrg} < \beta_{2\_sm\_md}$ for large values of bu but they are comparable in the order of magnitude (this comparison is essentially equivalent to comparing $W_{lrg}(x)$ and $W(x)$ as given by (132) and (129) respectively for large values of x, thus making the comparison independent of device geometry). Then, since we know that β_{2_lrg} is the more accurate approximation of β_2 than $\beta_{2\_sm\_md}$ for large values of bu, we want to design the combined approximation for β_2 such that it approximates towards β_{2_lrg} for large values of bu. We do this by designing the *global* β_2 as being equal to $min(\beta_{2\_sm\_md}, \beta_{2\_lrg})$, where by the term *global* β_2, we mean a *single-equation* approximation for β_2 as opposed to *piece-wise* approximations. Designating the global β_2 as β_{2x}, we can accomplish this by defining β_{2x} as:

$$\beta_{2x} = \beta_{2\_lrg}\left[1+\left(\frac{\beta_{2\_lrg}}{\beta_{2\_sm\_md}}\right)^{n_2}\right]^{(-1/n_2)} \tag{135}$$

where n_2 is a sufficiently large number.

β_{2x} as given by (135) is a rough approximation for a single-equation expression for β_2. However, its accuracy can be greatly increased by using it as an initial guess for 3 Newton-Raphson iterations of the modified transcendental equation (124). The Newton-Raphson formula for (124) can easily be derived as:

$$\beta_{2_{k+1}} = \beta_{2_k} - \frac{\beta_{2_k}{}^2 + \beta_{2_k} - t_k}{2\beta_{2_k} + 1 + 2bt_k} \tag{136}$$

where :

$$t_k = u^2 e^{-2b\beta_{2k}} \tag{137}$$

It turns out that in (134), a value of n_1=1 is adequate. Any larger value does not make a noticeable difference in the final value while unnecessarily increasing computation expense. The larger the value, the tighter is the down-clamping of bu at 10, which however is not needed to be a very tightly clamped value (i.e. it is acceptable if bu is down-clamped at a value slightly larger than 10). Hence a value of n_1=1 is adequate. Similarly in (135), a value of n_2=8 was found to be adequate. Any larger value does not make a noticeable difference in the final value but does increase computation expense, whereas a lower value degrades the accuracy at large V_{GS}. These are the values used for n_1 and n_2 in our model.

We re-emphasize that the procedure described above is *not* a piece-wise procedure (in piece-wise procedure, the various steps are executed for specific bias ranges). Indeed, all the steps are executed for *all* biases in our procedure, and the process of combining them and executing the 3 Newton-Raphson iterations at the end ensures that any errors introduced in intermediate steps are later reduced.

By carrying out the procedure twice, once by setting φ_{fn}=0 and once by setting φ_{fn}=V_{DS}, we get β_{2s} and β_{2d} respectively, which are the values of β_2 at the source and drain ends respectively. Having got those, (122) can be used to get the inversion charge areal density at the source and drain ends.

The entire procedure is summarized in Appendix-A.

In the following sub-section, we describe an enhancement to the approximation described in [66] to overcome its numerical robustness limitation.

6.2.1 Enhancement to the approximation described in [66]

Considering the root cause analysis (discussed in section 6.1) of the limited numerical robustness of the approximation described in [66], we suggest a different starting guess for z_1

(which equals tan(β_1) in [66]) than that used in (12) and (13) in [66]. If we use the following starting guess (where $\beta_{2\_sm\_md}$ is given by (131)):

$$z_1 = \sqrt{\beta_{2\_sm\_md}} \tag{138}$$

then we overcome the numerical robustness limitation of the approximation described in [66], remain about as computationally efficient as the approximation originally described in [66], and have an accuracy (in β_2) that is mid-way between the approximation originally described in [66] and our core approximation (and actually be almost as accurate as the approximation originally described in [66] if the bias range is limited up to 1V, as stated in Table 6-1). Having chosen the starting guess z_1 as in (138), the rest of the procedure remains the same as in [66]. In this chapter, we refer to this enhancement as the *Taur-Modified* approximation.

The justification for choosing z_1 as in (138) is that at small values of β_1, $\beta_2 = \beta_1 \tan(\beta_1) \approx \tan^2(\beta_1)$. But at small and medium values of β_1, we have already shown that $\beta_{2\_sm\_md}$ is a good approximation for β_2. Extending this assertion beyond small values of β_1 since this is only a starting guess, we can thus say that $\beta_{2\_sm\_md} \approx \tan^2(\beta_1)$. Therefore z_1, which equals tan(β_1) in [66], equals $\sqrt{\beta_{2\_sm\_md}}$.

6.3 Results

For a range of values of $(V_{GS}\text{-}\Delta\varphi\text{-}V_{DS})$, the original transcendental equation (119) was solved *numerically* for β_1, and then (58) was used to calculate β_2. This is our benchmark result, and we refer to it as the *numerical* solution. This term is not to be confused with other quantities like currents obtained from TCAD simulations, which we specifically refer to as *device simulation* data. We then used the analytical procedure summarized in Appendix-A to calculate β_2. We refer to this as the *analytical* solution. We have plotted the numerical and analytical solutions in Fig. 6-2 where results from using the existing approximations in the literature are also shown. Fig. 6-3 shows the absolute error percentage in the analytical solution *with respect to the numerical solution*, for the same candidate approximations that are plotted in Fig. 6-2.

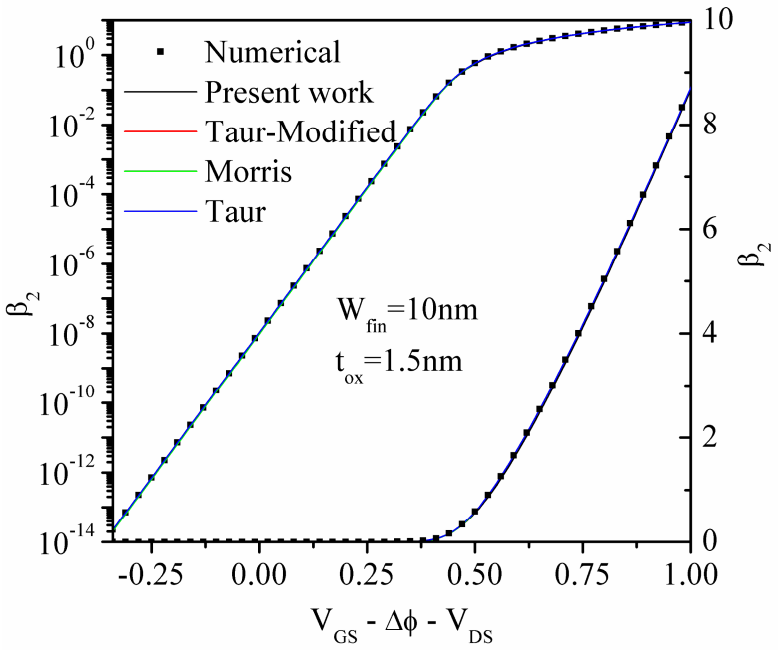


Figure 6-2: Numerical and analytical solutions for β_2 using the candidate approximations, when $(V_{GS}-\Delta\varphi-V_{DS})$ is swept from -0.34V to 1V. All the traces coincide with each other very closely.

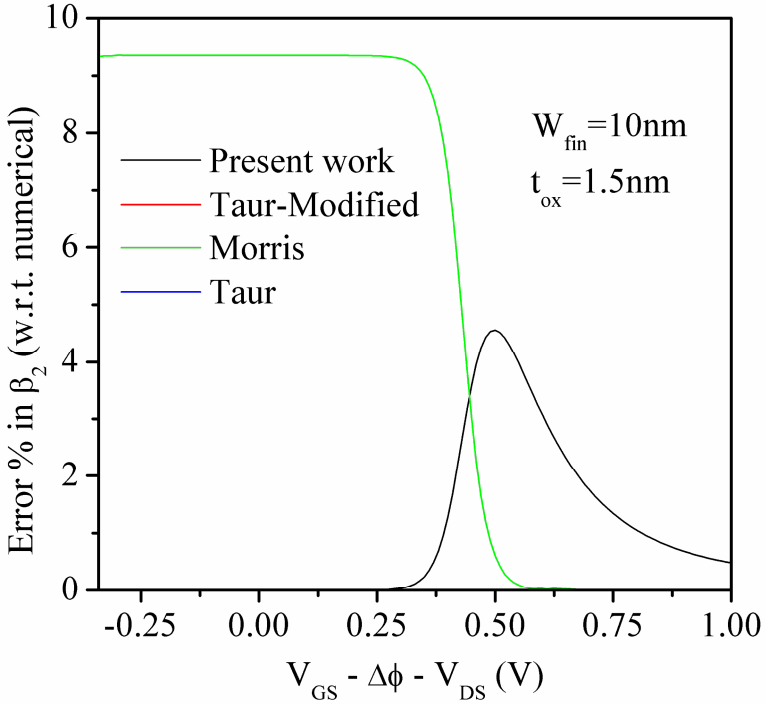


Figure 6-3: Error % in β_2 with respect to the numerically calculated β_2, using the candidate approximations. The maximum and average error percentages in our proposed approximation are 4.6% and 0.4% respectively. The curves labelled *Taur* and *Taur-Modified* are not visible because their error percentages are very low.

It is seen that the maximum and average error percentages in our proposed approximation are 4.6% and 0.4% respectively. It is felt that the main cause for the maximum error being as high as about 4.6% (which occurs at a small range of values of $(V_{GS}-\Delta\varphi-V_{DS})$), is the error involved in approximating (121) as (123). This is further confirmed when one notices a similar error profile in Fig. 3b in [19]. But for practical use, a maximum error of 4.6% is acceptable, as we will soon show. Fig. 6-4 shows the same curves as in Fig. 6-3, but plotted over a wider bias range of -4V to 4V, and limited to only our core approximation and the *Taur-Modified* approximation (the latter is our proposed enhancement of the approximation described in [66]). Even though the voltages are not expected to be as high as 4V at the geometries considered in most applications, it is possible that in the various internal iterations during the solution process in a circuit simulator, higher than normal voltages are encountered [53]. More importantly, I/O devices use higher voltages, as do also some of the reliability studies involving circuits. It is therefore imperative that these bias ranges are also reasonably handled.

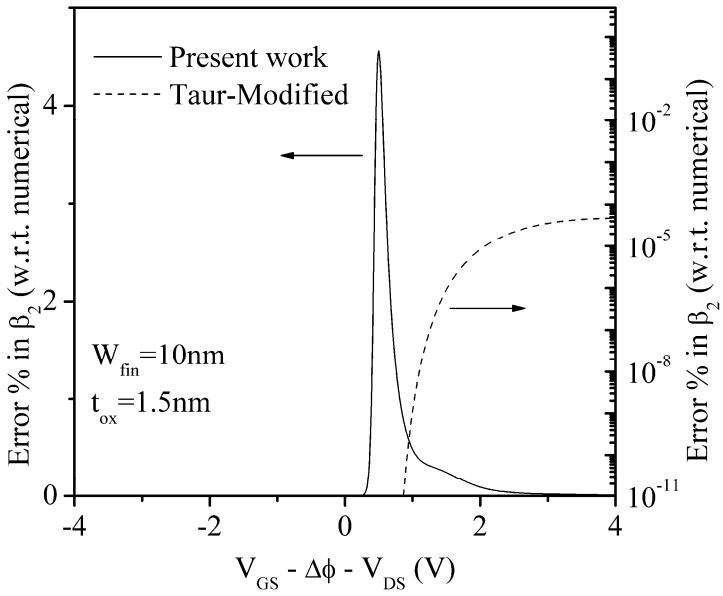
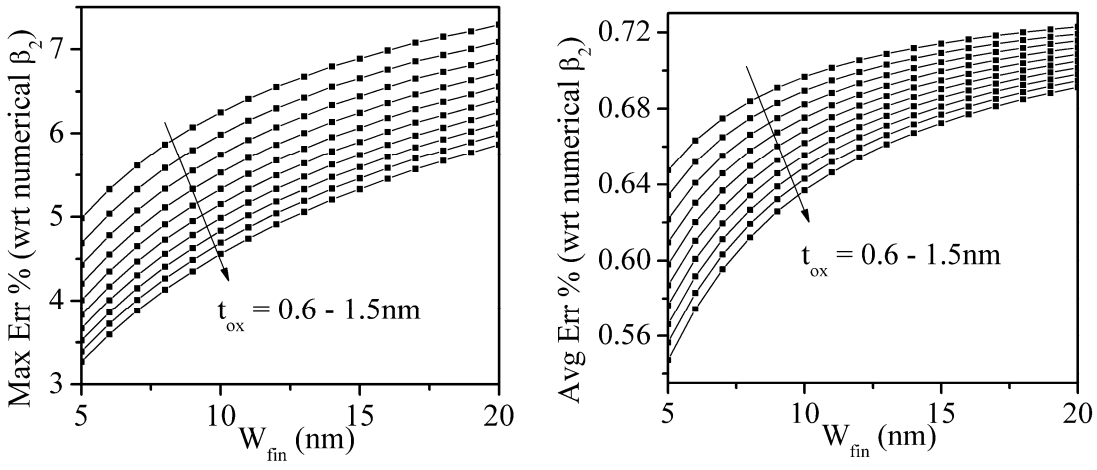


Figure 6-4: Same curves as in Fig. 6-3, but plotted over a wider bias range of -4V to 4V, and limited to only our proposed approximation and our proposed enhancement of [66].

To show that the approximation is valid for a range of geometries, we show in Fig. 6-5 the variation of the *maximum* (and *average*) absolute error percentage versus W_{fin} (i.e. the maximum error between numerical and analytical solutions when $(V_{GS}-\Delta\varphi-V_{DS})$ is swept from -1V to 1V), for various t_{ox}.



(a) Maximum error %. (b) Average error %.

Figure 6-5: (a) Maximum and (b) average absolute error percentages versus W_{fin} for various t_{ox} (i.e. the maximum (average) error between numerical and analytical solutions when $(V_{GS}-\Delta\varphi-V_{DS})$ is swept from -1V to 1V).

Fig. 6-6 - 6-9 shows drain current and trans-conductance plots where all the candidate approximations are compared with 2D device simulation results using Synopsys Sentaurus Device. The device geometry considered was t_{ox}=1.5 nm and W_{fin}=10 nm; and the channel length L, work function difference $\Delta\varphi$ and mobility μ used was 1 um, 0V and 300 cm^2/Vs respectively. These are the same values as those considered in [66]. Note however that the approximation itself is independent of the channel length, because the original transcendental equation is independent of the channel length. The analytical models used the same parameter values as those used in the device-simulations. As can be seen, our approximation results in a very good match in currents as well as conductances. Table 6-2 explains what the curves labelled (a)-(j) mean.

Figures 6-10 and 6-11 show the error in the drain current and trans-conductance at V_{DS}=50mV and 1V respectively, where the error is with respect to the corresponding device simulation values. In Fig. 6-10 and 6-11, it is clear that the curves labeled (c)-(e) overlap very closely, so also do the ones labeled (h)-(j), substantiating the validity of our suggested enhancement to the approximation described in [66]. Further, it is also clear that the approximation described in [66] (even though very highly accurate when compared with *numerically solving the original transcendental equation*), is only marginally better in accuracy (when *compared with*

device simulation data) than our proposed approximation which is numerically more robust and computationally more efficient.

Table 6-2: Meaning of the curve labels in Fig. 6-6 to 6-11.

Curve label	**Key state variable, and calculated using**	**Drain current model used**
(a) and (f)	β_2, calculated using our proposed approximation described in section 6.2	Sallese et al. [12]
(b) and (g)	β_1, calculated using [66]	Sallese et al. [12]
(c) and (h)	β_1, calculated using [66]	Taur et al. [10]
(d) and (i)	β_1, calculated using our suggested enhancement to [66] described in section 6.2.1	Taur et al. [10]
(e) and (j)	β_1, calculated numerically	Taur et al. [10]

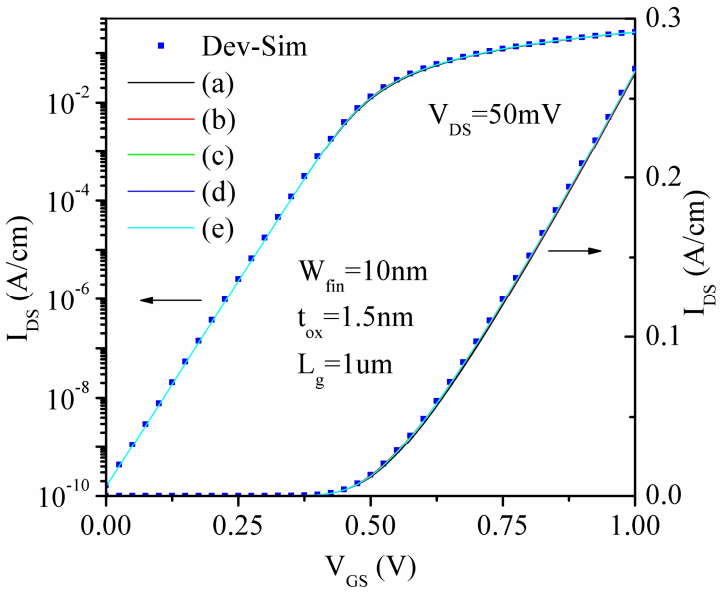


Figure 6-6: I_d-V_g plot at V_{DS}=50 mV generated using the candidate approximations, and compared with 2D device simulation results. The curve labels are explained in Table 6-2.

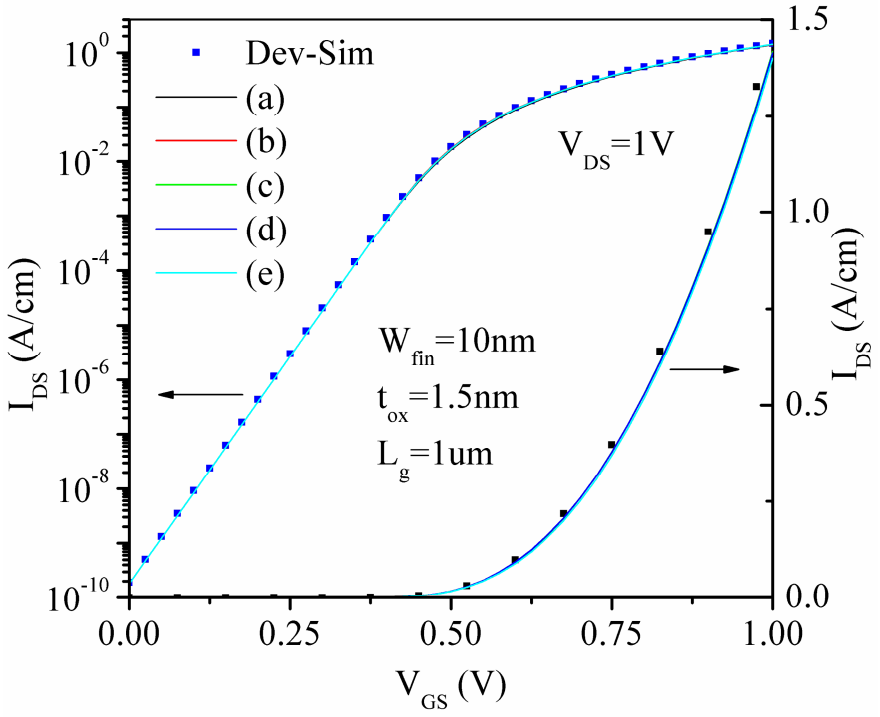


Figure 6-7: I_d-V_g plot at V_{DS}=1 V generated using the candidate approximations, and compared with 2D device simulation results. The curve labels are explained in Table 6-2.

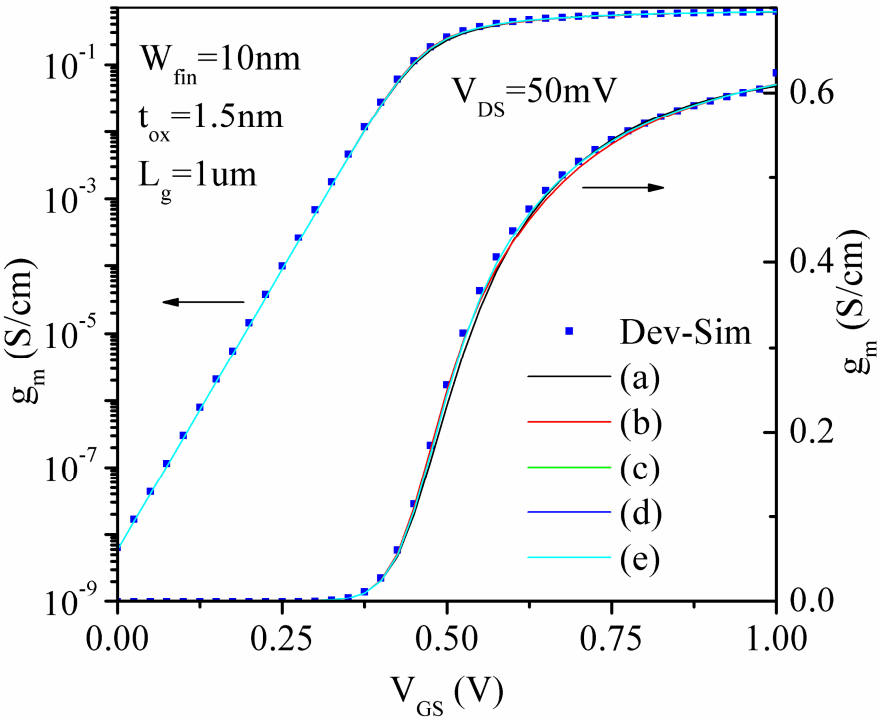


Figure 6-8: g_m-V_g plot at V_{DS}=50 mV generated using the candidate approximations, and compared with 2D device simulation results. The curve labels are explained in Table 6-2.

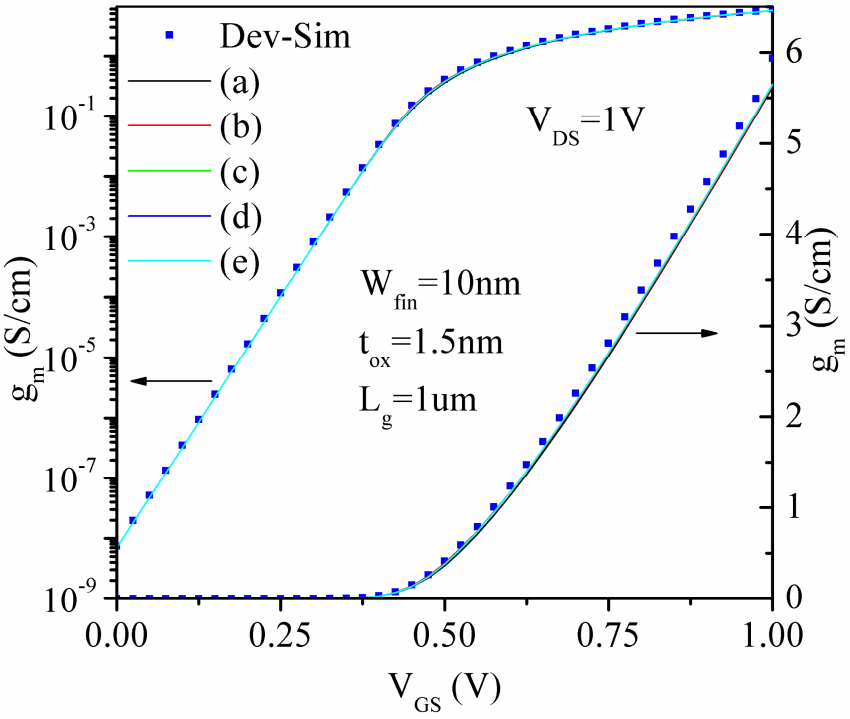


Figure 6-9: g_m-V_g plot at V_{DS}=1 V generated using the candidate approximations, and compared with 2D device simulation results. The curve labels are explained in Table 6-2.

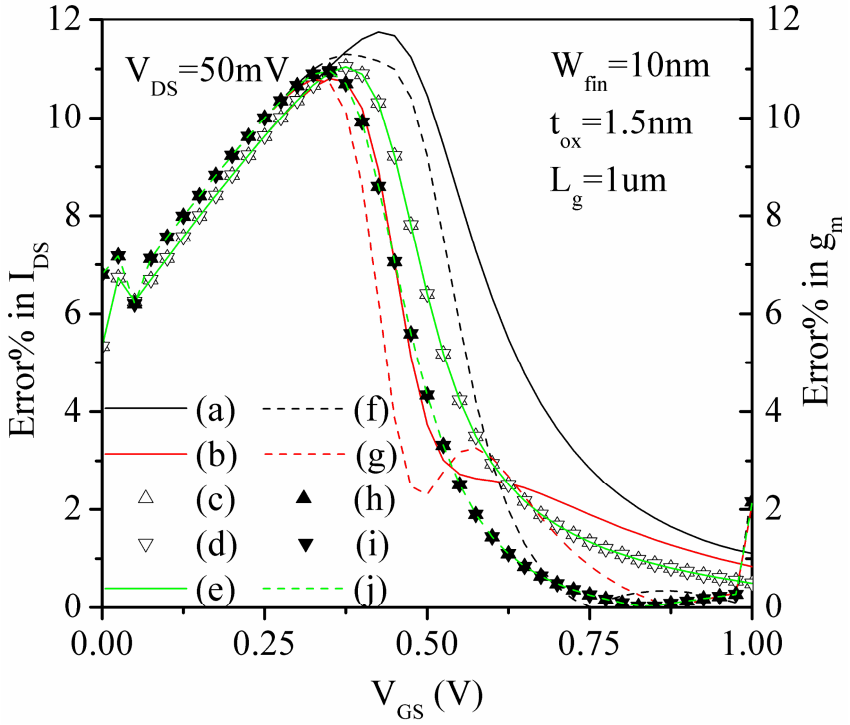


Figure 6-10: Error% in I_{DS} (left axis) and g_m (right axis) at V_{DS}=50 mV where the error is with respect to the corresponding device simulation values. The curve labels are explained in Table 6-2. The traces (a)-(e) denote the error in I_{DS} (left axis), and the traces (f)-(j) denote the error in g_m (right axis).

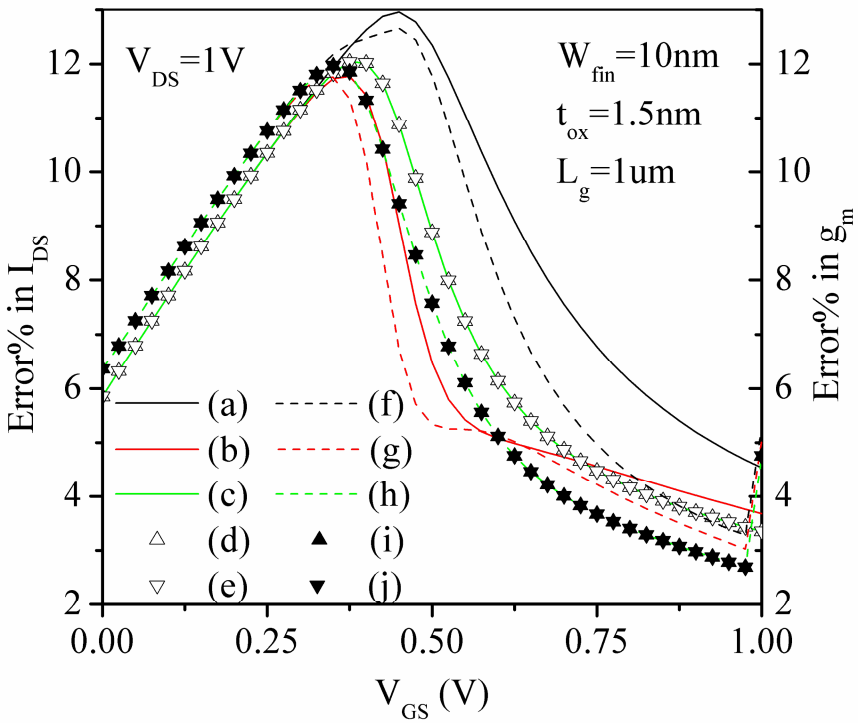


Figure 6-11: Error% in I_{DS} (left axis) and g_m (right axis) at V_{DS}=1 V where the error is with respect to the corresponding device simulation values. The curve labels are explained in Table 6-2. The traces (a)-(e) denote the error in I_{DS} (left axis), and the traces (f)-(j) denote the error in g_m (right axis).

6.4 Summary

In this chapter, we presented a closed form approximation for the inversion charge areal density that is computationally more efficient and numerically more robust than other similar approximations in the literature. Its accuracy, while lesser than other approximations, was shown to be acceptable for compact modeling purposes. We also presented an enhancement to an existing approximation in the literature, that helps in increasing its numerical robustness at a minimal loss of accuracy.

Chapter 7
Two-Dimensional Field Effects

The model presented until now considers only two short channel effects, viz. low-field mobility degradation and velocity saturation. These have been developed under the premise of the Gradual Channel Approximation (GCA), which assumes that the variation of the electric field *along* the fin is negligible in comparison to that *across* the fin. Thus our approach until now has been a 1-D approach rather than a full 2-D approach. This assumption is invalid in short channel devices especially at high drain voltages, where the lateral field penetrates till the source and affects the source-bulk energy barrier. This effect is called Drain Induced Barrier Lowering (DIBL). It manifests itself in the form of a higher current than that predicted by a model developed using the GCA. This effect is modeled by V_t-based models as a reduction in V_t.

In this chapter we describe the enhancement of our core model to add support for DIBL. Even though our core model is not a V_t-based model, we have developed a V_t, S and DIBL model. This is because from a device/circuit designer point of view, these three quantities are important metrics, and so it is always useful if one can model these quantities. Our V_t, S and DIBL model is described in section 7.2.

Indeed, even though we have developed a V_t, S and DIBL model, it is not incorporated in our core model *as is*. For incorporating DIBL effects into our model, we use a different technique, which is described in section 7.3.

7.1 Two-dimensional analysis

Breaking away from the premise of the GCA means that one needs to solve the 2D Poisson equation (139). As is commonly done when handling 2D field effects, we neglect body doping and assume constant mobility in this analysis. This is in line with the general approach in compact modeling that one handles one physical effect at a time [16,53].

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = \frac{q n_i}{\varepsilon} e^{(\psi - \phi_{fn})/\phi_t} \tag{139}$$

Being a *non-linear* partial differential equation (PDE), equation (139) is difficult to solve analytically. Hence (139) is usually analyzed in the sub-threshold regime, where the RHS is negligible and is therefore neglected. Any conclusions drawn (such as the extent of DIBL) are then applied to the above-threshold regime also, even though this is strictly not correct. In sub-threshold, (139) reduces to (140).

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = 0 \tag{140}$$

It is worthwhile mentioning at this time that some works [28,32,33] analyze (139) without limiting the form of the RHS to one that is valid only in the sub-threshold regime. Moreover, they get rid of one variable φ_{fn} from the RHS of (139) by applying an assumption that they show (via device simulations) to be valid only in sub-threshold. This assumption is that φ_{fn} is largely constant along the fin, and that much of the gradient in it occurs very near the drain end. This is shown in Fig. 7-1. And since much of the device sub-threshold behavior is governed by the potential at the point of minimum potential (also called the *virtual cathode* [33]), which occurs approximately halfway between the source and the drain (and moves towards the source end as the drain bias is increased), they assume that φ_{fn} = 0 all over the

device spatial domain in (139). The resulting equation is then solved using the superposition principle [77].

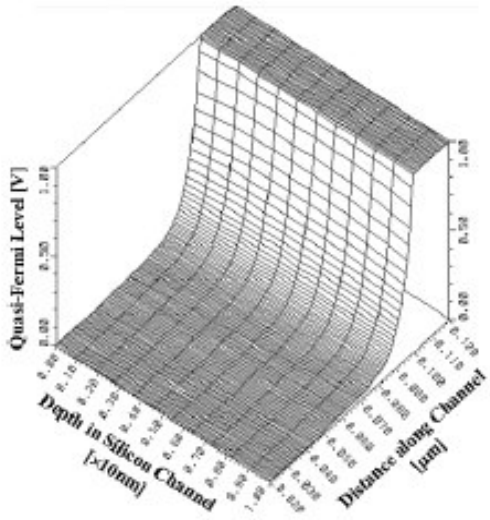


Figure 7-1: Spatial variation of φ_{fn} (Medici simulations) at threshold for a L=60 nm, W_{fin}=10 nm, t_{ox}=1.5 nm device at V_{DS}=1 V (from [33]).

The above technique has two features that are not technically sound:

- The superposition principle is not applicable for non-linear PDE's (which (139) continues to remain even if $\varphi_{fn} = 0$), and

- If an approximation (viz. that $\varphi_{fn} = 0$) that is valid only in sub-threshold is used, then any solutions found would strictly not be applicable to above-threshold conditions; as such, retaining a non-zero RHS seems pointless.

The goal of all these works is generally to develop a scaling theory by introducing the concept of a scale length. The extent by which the channel length exceeds the scale length is a measure of the robustness to SCE.

Some other approaches [23,25,26], all of which were based on the approach developed in [78], do a 2D analysis by assuming a parabolic potential variation *across* the fin. These result in a scale length proportional to the geometric mean of t_{ox} and W_{fin}, and large errors occur when one of the thicknesses is much larger than the other [29]. Large errors are also shown to occur when high-k gate dielectrics are used [27]. In another approach, the gate-oxide regions

are replaced by a silicon region whose thickness is three times that of the actual gate-oxide thickness, thus eliminating the boundary conditions at the silicon/oxide interface [30]. While the perpendicular fields are scaled properly in this one-region model, the lateral fields are not. This is the same approach as used in the original work by Nguyen in solving the 2-D problem for bulk MOSFETs [79]. It is a good approximation only when the gate-oxide is much thinner than W_{fin}. In [28,32,33], the oxide and silicon regions are separate, but the vertical field in the oxide was assumed to be uniform when formulating the boundary conditions at the oxide-silicon interface, a very commonly used assumption but nevertheless questionable when high-k gate dielectrics are used.

The general prediction of these scaling theories is that the channel length should be atleast 2 times W_{fin}, in order to have acceptable SCE. Then, given that FinFETs are expected to be introduced at the 32 nm or more likely the 22 nm node, it is expected that W_{fin} will be about 10 nm, and EOT will be about 1 nm. At these advanced nodes, a high-k gate dielectric is also expected to be used. In a simplistic manner, a high-k dielectric constant of 12 would mean a physical oxide thickness of about 3.1 nm in order to give an EOT of 1 nm (neglecting inversion layer quantization and the charge-centroid shifting effects). A gate-oxide thickness of 3.1 nm (especially of high-k material, and therefore conducive to increased S-D fringing fields through it) cannot be deemed a negligible fraction of W_{fin} = 10 nm. Hence it is our belief that these scaling theories will be inaccurate.

Therefore, we chose to build our 2D model by building on top of the 2D analysis presented by Liang and Taur in [29], which is claimed to be valid for a wide range of t_{ox}-W_{fin} combinations. The work in [29] is based on the concept of the generalized scale length introduced in [27]. We briefly describe next the 2D analysis done in [29].

7.1.1 Sub-threshold analysis as done in [29]

In order to aid in the readability of the rest of this chapter, a brief overview of the approach followed in [29] is described here. It should be noted that [29] had shown the solution for both symmetric as well as asymmetric DGFETs, and had used a different coordinate system. We

summarize below the portions relevant to symmetric DGFETs, and as adapted to our coordinate system.

In [29], focus was limited to the sub-threshold region in a lightly doped/undoped DGFET. Therefore (140) was solved instead of (139). For a SDGFET, the various boundary conditions are:

$$\begin{aligned}
&\psi(x,|t_{ox}+W_{fin}/2|)=V_{GS}-\Delta\phi \\
&\psi(-L/2,y)=E_g/2q \qquad \text{for y in Silicon} \\
&\psi(L/2,y)=V_{DS}+E_g/2q \qquad \text{for y in Silicon} \\
&\psi(-L/2,y)=(E_g/2q)\left[\frac{W_{fin}/2+tox-|y|}{t_{ox}}\right]-(V_{GS}-\Delta\phi)\left[\frac{W_{fin}/2-|y|}{t_{ox}}\right] \qquad \text{for y in oxide} \\
&\psi(L/2,y)=(V_{DS}+E_g/2q)\left[\frac{W_{fin}/2+tox-|y|}{t_{ox}}\right]-(V_{GS}-\Delta\phi)\left[\frac{W_{fin}/2-|y|}{t_{ox}}\right] \text{for y in oxide}
\end{aligned} \tag{141}$$

The oxide boundary conditions are based on an assumption of linear variation of potential at the side boundary (but not in the oxide interiors, as assumed by other works). In addition to this, continuity of $\psi(x,y)$ and its y-gradient is enforced at the oxide-silicon interface (using the actual y-gradient of the $\psi(x,y)$ solution in the oxide region in place of E_{oxy}, as opposed to approximating E_{oxy} as V_{ox}/t_{ox} as other works have done). The superposition principle [77] was used and the electrostatic potential was written as:

$$\psi(x,y)=v(y)+u_L(x,y)+u_R(x,y) \tag{142}$$

where $v(y)$ is the 1D solution of (140) and satisfies the front and back gate boundary conditions, $u_L(x,y)$ and $u_R(x,y)$ are the 2D solutions of (140) that satisfy the source side and the drain side boundary conditions respectively (only). The individual superposed 2D pieces $u_L(x,y)$ and $u_L(x,y)$ were assumed to be of the form $X(x)Y(y)$, and the usual separation of variables [77] technique was used, followed by application of the boundary conditions, to generate a general solution for the electrostatic potential (in the different regions) as an infinite series, given by (142), (143), (144) and (145).

$$v(y)=V_{GS}-\Delta\phi \tag{143}$$

$$u_L(x,y) =$$

$$\sum_{i=1}^{\infty}(-1)^{i-1}b_{2i-1}\frac{\sinh\left[\frac{\pi}{\lambda_{2i-1}}\left(\frac{L}{2}-x\right)\right]}{\sinh\left(\pi L/\lambda_{2i-1}\right)}\cos\left(\pi y/\lambda_{2i-1}\right) \quad \text{(in Silicon)}$$

$$\sum_{i=1}^{\infty}(-1)^{i-1}b_{2i-1}\frac{\cos\left(\frac{\pi W_{fin}}{2\lambda_{2i-1}}\right)}{\sin\left(\pi t_{ox}/\lambda_{2i-1}\right)}\frac{\sinh\left[\frac{\pi}{\lambda_{2i-1}}\left(\frac{L}{2}-x\right)\right]}{\sinh\left(\pi L/\lambda_{2i-1}\right)}\sin\left[\frac{\pi}{\lambda_{2i-1}}\left(\frac{W_{fin}}{2}+t_{ox}-|y|\right)\right] \quad \text{(in Oxide)} \tag{144}$$

$$u_R(x,y) =$$

$$\sum_{i=1}^{\infty}(-1)^{i-1}c_{2i-1}\frac{\sinh\left[\frac{\pi}{\lambda_{2i-1}}\left(\frac{L}{2}+x\right)\right]}{\sinh\left(\pi L/\lambda_{2i-1}\right)}\cos\left(\pi y/\lambda_{2i-1}\right) \quad \text{(in Silicon)}$$

$$\sum_{i=1}^{\infty}(-1)^{i-1}c_{2i-1}\frac{\cos\left(\frac{\pi W_{fin}}{2\lambda_{2i-1}}\right)}{\sin\left(\pi t_{ox}/\lambda_{2i-1}\right)}\frac{\sinh\left[\frac{\pi}{\lambda_{2i-1}}\left(\frac{L}{2}+x\right)\right]}{\sinh\left(\pi L/\lambda_{2i-1}\right)}\sin\left[\frac{\pi}{\lambda_{2i-1}}\left(\frac{W_{fin}}{2}+t_{ox}-|y|\right)\right] \quad \text{(in Oxide)} \tag{145}$$

We refer to each term in the series summation in (144) and (145) as $u_{Ln}(x,y)$ and $u_{Rn}(x,y)$.

The terms λ_n are called the *generalized scale lengths* [27], with λ_1 being the most dominant, followed by λ_3, λ_5, etc (only odd terms are needed for SDGFETs). In general, they decrease in magnitude as $1/n$ [29]. Thus λ_3 is approximately one-third of λ_1, etc. By applying the dielectric boundary conditions, the geometry-dependent relation for λ_n was shown to be:

$$\tan\left(\pi t_{ox}/\lambda_n\right)\tan\left(\frac{\pi W_{fin}}{2\lambda_n}\right)=\frac{\varepsilon_{ox}}{\varepsilon} \tag{146}$$

In [80], a closed form approximation for the dominant term λ_1 was proposed as $\lambda_1=W_{fin}(1+b)$ when $t_{ox} << W_{fin}$, where $b=2\varepsilon/(W_{fin}C_{ox})$.

What clearly comes across from (146) is that λ_n are not simply dependent on $(\varepsilon_{ox}/t_{ox})$, thus showing that this approach is more accurate than other approaches that have simply scaled the oxide thickness by three and replaced it with silicon, when doing the 2D analysis.

Reference [29] then goes on to derive expressions for coefficients b_1 and c_1 in the infinite series solution using a conjugate function $g_1(y)$. Such a conjugate function was needed to be derived because the *remnant sin()* and *cos()* functions in the infinite series for $\psi(x,y)$ (*remnant* in the sense that when x is pinned down at any constant value such as $\pm L/2$ for example) are not orthogonal to each other. By *requiring* that these functions be orthogonal to $g_1(y)$ in the domain $0 \le |y| \le W_{fin}/2+t_{ox}$, they derived region-wise expressions for $g_1(y)$. These come out to be a scalar multiple of $u_{L1}(-L/2,y)$. Also as a result of the orthogonality condition, $g_1(y)$ ends up having a discontinuity at the oxide-silicon interface, but is continuous in its derivative. This is shown in Fig. 7-2.

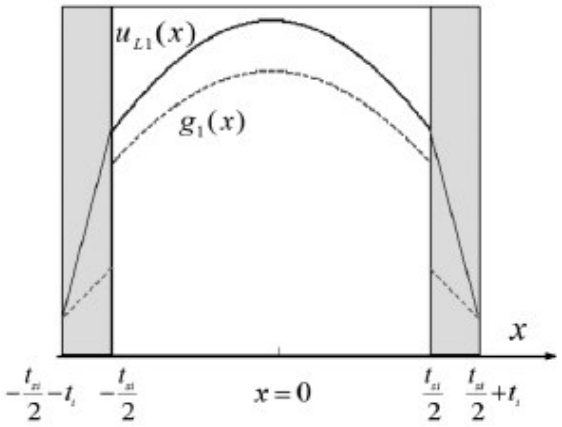


Figure 7-2: Plot of the lowest order eigen function u_{L1} and its conjugate function $g_1(x)$. Notice the different coordinate system (from [29]).

Finally, [29] stated an integral equation for the drain current under the drift-diffusion transport mechanism using conditions of constant mobility as given by (147). Retaining only the first order terms in (142)-(145), they then *numerically* evaluated (147) to get I_d-V_g characteristics in sub-threshold. They showed a good match with TCAD device simulations up to about L=40 nm for a W_{fin}=10 nm, t_{ox}=1.5 nm device (for which λ_1=17.73 nm). They did not get a very good match for a L=24 nm device, and the reason claimed was that such a short device fails the $L>1.5\lambda_1$ criterion, which is necessary in order to justify retaining only the first order terms in the infinite series.

$$I_{DS} = \frac{qn_i\mu_0\phi_t\left(1-e^{-V_{DS}/\phi_t}\right)}{\int_{-L/2}^{L/2}\frac{dx}{\int_{-W_{fin}/2}^{W_{fin}/2}e^{\psi(x,y)/\phi_t}dy}} \tag{147}$$

7.1.2 Higher order terms

In [29], only the first order terms in the infinite series for the electrostatic potential were retained when numerically evaluating the drain current, thus making the model inapplicable to extremely short channel devices. In order to be applicable to extremely short channel devices, we modified the approach in [29] by retaining one more term in the infinite series for the electrostatic potential. For that we have to derive expressions for coefficients b_3 and c_3, for which we need to find the conjugate function $g_3(y)$. Taking a cue from the $g_1(y)$ expression presented in [29], we assumed that $g_3(y)$ is also a scalar multiple of $u_{L3}(-L/2,y)$. Requiring then that the integral I as given by (148) depend only on b_3 and λ_3 (ie. should not depend on other order terms like b_1, λ_1, b_5, λ_5, … etc, nor on any of the c-series coefficients (even if of the same order) c_1, c_3, c_5, …etc), we derived an expression for $g_3(y)$ as given by (149). Note: The orthogonality condition can be useful in determining b_3 only if this requirement is met.

$$I = \int_{-W_{fin}/2-tox}^{Wfin/2+tox} \left[\psi(-L/2, y) - (Vg - \Delta\phi)\right] g_3(y)dy \tag{148}$$

$$g_3(y) = \begin{cases} \cos(\pi y / \lambda_3) & \text{for y in silicon} \\ \dfrac{\sin(\pi W_{fin}/(2\lambda_3))}{\cos(\pi t_{ox}/\lambda_3)} \sin\left[\dfrac{\pi}{\lambda_3}\left(W_{fin}/2 + t_{ox} - |\,y\,|\right)\right] & \text{for y in oxide} \end{cases} \tag{149}$$

Then, evaluating (148) once by using the boundary condition (141) for the potential and then once again by using the infinite series solution (142)-(145), and equating the two, we derived an expression for b_3. Notice that (148) has $x = -L/2$, which is the source end. Repeating this effort by setting $x = L/2$ (but using the same $g_3(y)$), we derived an expression for coefficient c_3. Generalizing this approach, we state in (150) the expressions for b_n and c_n that we derived (of any order in general). Equations (142)-(145) and (150) give the complete electrostatic potential solution. It should be noted that the b_n and c_n coefficients have dimensions of potential.

$$b_{2n+1} = (-1)^n B_{2n+1}\left[\frac{E_g}{2q} - (V_{GS} - \Delta\phi)\right]$$

$$c_{2n+1} = (-1)^n B_{2n+1}\left[\frac{E_g}{2q} + V_{DS} - (V_{GS} - \Delta\phi)\right] \qquad (150)$$

$$B_{2n+1} = \frac{2\lambda_{2n+1}^2 \tan\left(\pi t_{ox}/\lambda_{2n+1}\right)\sin\left(\frac{\pi W_{fin}}{2\lambda_{2n+1}}\right)}{\pi^2 t_{ox}\left[\frac{W_{fin}}{2} + \frac{t_{ox}\sin\left(\pi W_{fin}/\lambda_{2n+1}\right)}{\sin\left(2\pi t_{ox}/\lambda_{2n+1}\right)}\right]} \qquad \text{n=0, 1, 2, ...}$$

7.1.3 Point of minimum potential (x_c)

As pointed out in [29], the electrostatic potential variation *across* the fin in the sub-threshold regime has a maximum that occurs at the fin centre (ie. $y_c = 0$). In the sub-threshold regime, the points of weakest gate control are the points along the fin centre ($y = 0$). This is where the maximum leakage current flows. Therefore the $y = 0$ line is an important line to analyze as far as the variation of electrostatic potential in sub-threshold is concerned. Furthermore, the variation of the potential at $y = 0$ *along* the fin has a minimum that occurs at some point which we label as x_c. It is the point of minimum potential along the line of crucial interest. It is an important point because many of the useful metrics of the device (such as V_t, etc) can be derived on the basis of this point. Some works [33] refer to this point as the *virtual cathode*. From this point, one sees a potential barrier when looking towards the source, and also towards the drain. The barrier seen when looking towards the source (which is modulated by V_{DS} for short channel devices) is the main barrier that dictates the useful metrics. We derive in this section an expression for x_c.

To derive the point of minimum potential, we set:

$$\left.\frac{\partial\psi(x,0)}{\partial x}\right|_{x=x_c} = 0 \qquad \text{in silicon} \qquad (151)$$

Using (142)-(145), (150) and (151) and retaining only two orders of coefficients, viz. the first order (b_1, c_1) and the third order (b_3, c_3), and making approximations like $L >> \lambda_3$ (but not

necessarily >> λ_1) which helps us make approximations like $\cosh(\pi L/2\lambda_3) \approx (1/2)\exp(\pi L/2\lambda_3)$, we get an expression for x_c as given by (152).

$$x_c = \frac{\lambda_1}{\pi}\ln\left[\frac{\sqrt{4\left(b^2c_1 - b_1\right)\left(b^2b_1 - c_1\right) + b^2c^2} - bc}{2\left(b^2c_1 - b_1\right)}\right]$$

$$\text{where } c = \frac{2\lambda_1\left(b_3 - c_3\right)}{\lambda_3}\sinh\left(\pi L/\lambda_1\right)e^{-\pi L/2\lambda_3} \qquad (152)$$

$$\text{and } b = e^{\pi L/2\lambda_1}$$

Equation (152) for x_c matches well with TCAD device simulations. For a L = 20 nm, W_{fin} = 10 nm, t_{ox} = 1 nm device, (152) predicts x_c = -0.22 nm at V_{GS} = 0, V_{DS} = 50 mV and x_c= -2.62 nm at V_{GS} = 0, V_{DS} = 1V. The corresponding TCAD measured x_c is -0.3 nm and -2.5 nm.

The general behavior of x_c is that it is 0 (ie. halfway between the source and the drain) at V_{DS} = *0*, and moves towards the source end as V_{DS} is increased.

7.1.4 Drain current model in sub-threshold

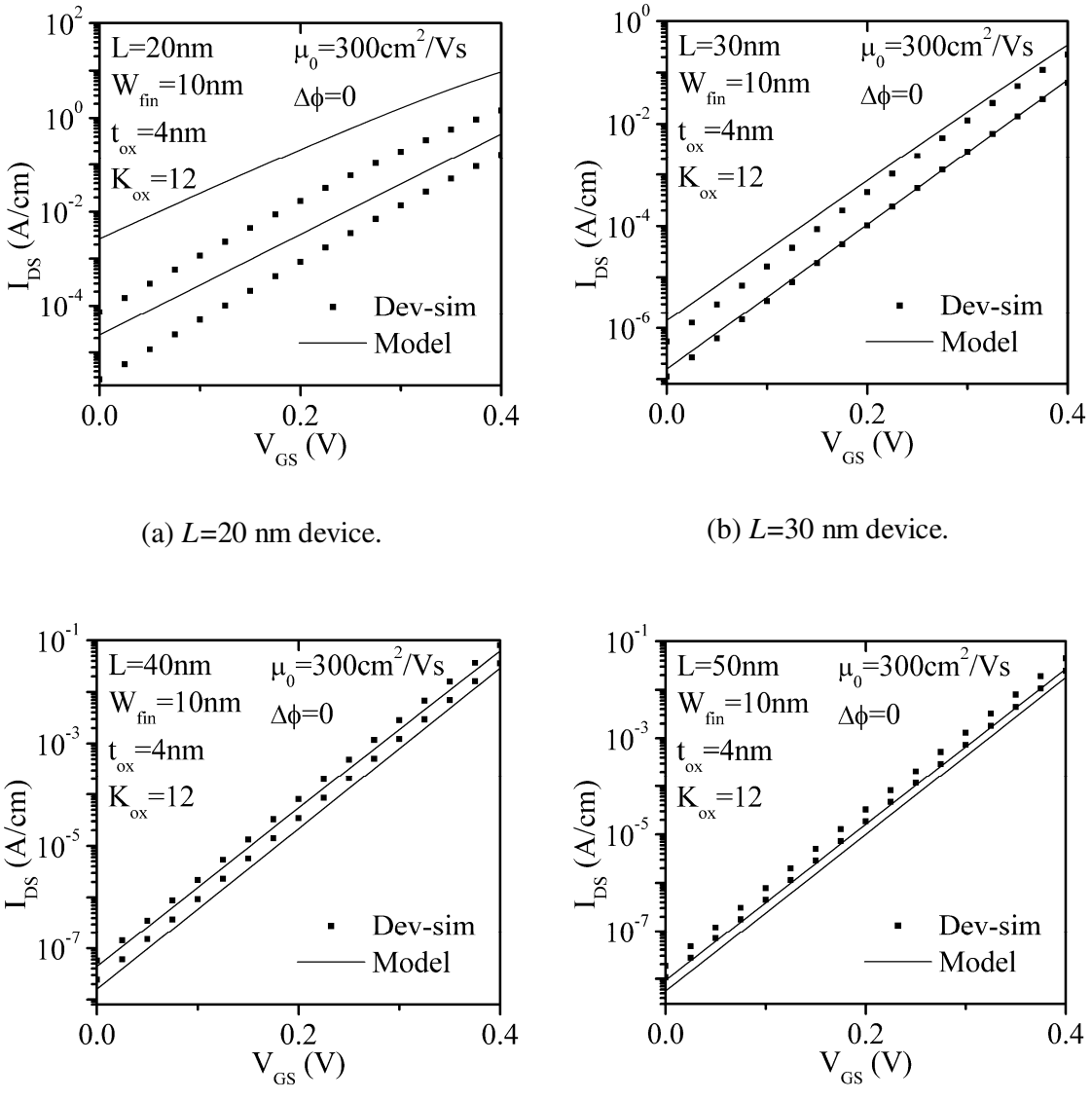
Equation (147) has two integrals. As done in [28,33], we approximate the inner integral *across* the fin as the fin width multiplied by the value of the integrand at $y = y_1$, where y_1 is mid-way between y = 0 and the oxide-silicon interface. Thus, $y_1 = W_{fin}/4$. For the outer integral *along* the fin, we approximate it as the channel length multiplied by the value of the integrand at $x=x_c$, which we have stated to be the defining point for many of the device's metrics. Using this technique and using (142)-(145) and (150) and retaining only two orders of coefficients, we get an expression for the drain current as given by (153).

$$I_{DS} = \frac{q\mu_0 n_i W_{fin}\phi_t}{L}e^{(V_{GS}-\Delta\phi)/\phi_t}\left(1 - e^{-V_{DS}/\phi_t}\right)\cdot H \quad \text{where} \quad y_1 = W_{fin}/4 \quad \text{and}$$

$$H = \exp\left\{\left[\frac{1}{\phi_t}\right]\left[\frac{b_1\sinh\left[\frac{\pi}{\lambda_1}\left(\frac{L}{2} - x_c\right)\right] + c_1\sinh\left[\frac{\pi}{\lambda_1}\left(\frac{L}{2} + x_c\right)\right]}{\sec\left(\pi y_1/\lambda_1\right)\sinh\left(\pi L/\lambda_1\right)} - \frac{\left(b_3 + c_3\right)e^{-\pi L/2\lambda_3}}{\sec\left(\pi y_1/\lambda_3\right)}\right]\right\} \qquad (153)$$

From (153), it is clear that $H \geq 1$, and tends towards 1 as L/λ_1 increases.

Fig. 7-3 shows a plot of the analytical model versus TCAD device simulated currents in sub-threshold, for a device with a high-k gate dielectric of K_{ox} = 12, physical t_{ox} = 4 nm, and W_{fin} = 10 nm at V_{DS} = 50 mV and 1 V, for four channel length devices. The model uses the exact same parameter values as those used in the device simulations.



(a) L=20 nm device.

(b) L=30 nm device.

(c) L=40 nm device.

(d) L=50 nm device.

Figure 7-3: Drain current in sub-threshold using the analytical model given by (153), compared with TCAD device simulations for various channel lengths, without using any fitting parameters. Each plot shows traces at V_{DS}=50 mV and 1 V.

It can be seen that there is a mismatch between the model and the TCAD data, especially for the 20 nm device. This is attributed largely to the approximations made in evaluating the two nested integrals in (147).

7.1.5 Third order coefficients

The impact of retaining the third order coefficients was examined (for example, by examining the magnitude of the first and second terms in the expression for H in (153), the magnitude of c in (152), etc). It was found that they are always 2 orders of magnitude lower (at high V_{DS}) or even lesser, even for channel lengths as small as 10 nm. The primary reason for this is that the generalized scale lengths decrease as $1/n$, and so while $exp(-L/\lambda_1)$ may not be negligible for short channel devices, $exp(-L/\lambda_3)$ is about e^3 times smaller. This combined with the fact that the b_n vary as the square of λ_n in (150), makes it clear that terms like $b_3 exp(-L/\lambda_3)$ will be about $(3^2 \cdot e^3) \approx 180$ times smaller than $b_1 exp(-L/\lambda_1)$. Thus, it is a very good approximation to altogether neglect the third order coefficients as has been done in the literature.

Nevertheless in the following section on V_t, S and DIBL modeling, we have retained them where possible.

7.2 V_t, S and DIBL model

We describe next how we used the expressions developed until now in order to derive models for the threshold voltage V_t, sub-threshold slope S, and the DIBL coefficient.

7.2.1 Threshold voltage V_t

For lightly doped devices, the classical definition for V_t based on the bands bending by $2\varphi_b$ (where φ_b is the separation between the intrinsic level and the fermi potential in the silicon) does not hold [81]. We have therefore derived expressions for V_t using three methods that we describe next, none of which rely on the classical $2\varphi_b$-based definition.

7.2.1.1 Q_{th} *method*

We follow here the definition of V_t as used in [33] and later used in [32]. We refer to this method as the Q_{th} *method*.

Reference [33] defined $Q_{th}(x_c)$, the inversion carrier areal density at $x=x_c$, as:

$$Q_{th} = \int_{-W_{fin}/2}^{W_{fin}/2} n_i e^{\psi_{\min}(y)/\phi_t} dy \cong W_{fin} n_i e^{\psi'_{\min}/\phi_t} \tag{154}$$

where $\psi_{min}(y) = \psi(x_c, y)$; and $\psi_{min}' = \psi(x_c, y_1)$ where y_1 is mid-way between $y = 0$ and the oxide-silicon interface. Thus, $y_1 = W_{fin}/4$. Q_{th} is thus the inversion *carrier* areal density (not inversion *charge* areal density; we are following here the convention as used in [33] and [32]) at $x = x_c$. It should be noted that in (154), the electron quasi fermi level is assumed to be close to 0. As stated earlier, this has been shown to be valid in sub-threshold.

The idea is to use a pre-calibrated threshold value for Q_{th}, to deem it as the onset of inversion.

From (154) we have:

$$\psi'_{\min} = \phi_t \ln\left(\frac{Q_{th}}{W_{fin} n_i}\right) \tag{155}$$

The RHS of (155) can therefore be equated with the electrostatic potential as given by (142)-(145) and (150) evaluated at $x = x_c$, $y = y_1$ and $V_{GS} = V_t$ and retaining only two orders of coefficients, viz. the first order (b_1, c_1) and the third order (b_3, c_3). After some tedious algebra, we get (156). It should be noted that (156) depends on x_c, which in turn depends on V_{GS}. Setting V_{GS} to V_t in the x_c expression and then using it in (156) may seem to be the more correct way to compute V_t, but that approach gives an implicit equation for V_t. So as an approximate technique, we set $V_{GS} = 0$ when calculating x_c *for use in (156)*.

The long channel expression for V_t can easily be obtained from (156) by setting $L >> \lambda_n$, and we get (157).

$$V_t = \frac{H_3}{H_4}$$

$$H_3 = \Delta\phi + \phi_t \ln\left(\frac{Q_{th}}{W_{fin} n_i}\right) - B_1\left(\frac{E_g}{2q} + \Delta\phi\right)\cos\left(\frac{\pi y_1}{\lambda_1}\right)\frac{\cosh\left(\frac{\pi x_c}{\lambda_1}\right)}{\cosh\left(\frac{\pi L}{2\lambda_1}\right)} -$$

$$B_1 V_{DS}\cos\left(\frac{\pi y_1}{\lambda_1}\right)\frac{\sinh\left[\frac{\pi}{\lambda_1}\left(\frac{L}{2} + x_c\right)\right]}{\sinh\left(\frac{\pi L}{\lambda_1}\right)} - B_3\left(\frac{E_g}{q} + V_{DS} + 2\Delta\phi\right)\cos\left(\frac{\pi y_1}{\lambda_3}\right)e^{-\pi L/2\lambda_3} \quad (156)$$

$$H_4 = 1 + B_1\cos\left(\frac{\pi y_1}{\lambda_1}\right)\frac{\cosh\left(\frac{\pi x_c}{\lambda_1}\right)}{\cosh\left(\frac{\pi L}{2\lambda_1}\right)} + 2B_3\cos\left(\frac{\pi y_1}{\lambda_3}\right)e^{-\pi L/2\lambda_3}$$

$$V_{tL} = \Delta\phi + \phi_t \ln\left(\frac{Q_{th}}{W_{fin} n_i}\right) \quad (157)$$

Reference [33] suggested that (157) can be used for a long channel device to calibrate Q_{th}. Having once done that, (156) can be used as a V_t model.

7.2.1.2 g_m-C_{ox} method

We follow here the definition of V_t as used in [28,82]. We refer to this method as the g_m-C_{ox} *method*. Per this method, V_t is defined as the gate voltage at which equation (158) holds.

$$\left.\frac{dQ_i(x_c)}{dV_g}\right|_{V_g = V_t} = \frac{C_{ox}}{2} \quad (158)$$

where

$$Q_i = \int_{-W_{fin}/2}^{W_{fin}/2} q n_i e^{\psi_{\min}(y)/\phi_t} dy \cong W_{fin} q n_i e^{\psi'_{\min}/\phi_t} \quad (159)$$

In (147), using (159) for the inner integral *across* the fin and making the same approximation for the outer integral *along* the fin as that discussed above (153), we can easily see that the threshold condition (158) is equivalent to that given by (160).

$$\left.\frac{dI_{DS}}{dV_g}\right|_{V_g=V_t} = \frac{\mu_0 \phi_t C_{ox}}{2L}\left(1-e^{-V_{DS}/\phi_t}\right) \tag{160}$$

Using (153) in the LHS of (160), and assuming $x_c \approx 0$ (else the expressions become intractable), we get the expression for V_t as given by (161) and (162).

$$V_t = \Delta\phi + \phi_t \ln\left(\frac{\phi_t C_{ox}}{2qn_i W_{fin} G}\right) \tag{161}$$

where

$$G = H\left[1-2B_1\cos\left(\frac{\pi y_1}{\lambda_1}\right)\frac{\sinh\left(\pi L/2\lambda_1\right)}{\sinh\left(\pi L/\lambda_1\right)} - 2B_3\cos\left(\frac{\pi y_1}{\lambda_3}\right)e^{-\pi L/2\lambda_3}\right] \tag{162}$$

It should be noted that H (given by (153)) in (162) depends on V_{GS} (even after assuming $x_c \approx 0$) via the b_n and c_n coefficients. Setting V_{GS} to V_t in the H expression and then using it in (162) may seem to be the more correct way to compute V_t, but that approach gives an implicit equation for V_t. So as an approximate technique, we set $V_{GS} = 0$ when calculating *H for use in (162)*.

The long channel V_t is obtained from (161) by simply setting $G = 1$.

7.2.1.3 Constant current method

This is our own definition and is motivated by the constant current method of experimentally extracting V_t. In the constant current method, one measures the gate voltage at which the current per unit square reaches a pre-defined threshold current I_0. Different works have used different values for this threshold current I_0. Reference [23] had used an I_0 of 1 uA per unit (H_{fin}/L), whereas [83] suggests an I_0 of 0.1 uA per unit (H_{fin}/L). We follow the definition of

[83]. Translating to units of A/cm (current per unit fin height), this condition is equivalently stated in (163) where L_{nm} is in units of nanometers.

$$I_0 = \frac{1}{L_{nm}} \; A/cm \tag{163}$$

Equating (163) and (153), we get an expression for V_t as stated in (164). As before, in order to avoid implicit equations, we set V_{GS} = 0 when calculating *H for use in (164)*. The long channel V_t is obtained from (164) by simply setting H=1.

$$V_t = \Delta\phi + \phi_t \ln\left(\frac{10^{-7}}{q\mu_0 n_i W_{fin} \phi_t H(1-e^{-V_{DS}/\phi_t})}\right) \tag{164}$$

In (164), the denominator of the logarithm term has units of current. When using (164), care should be taken that the units of the various terms *comprising* the denominator of the logarithm term are such that their product is in units of Amperes (not uA or mA or some other variant), in order for the validity of (164) to hold as stated.

The key distinguishing factor about this definition is that the mobility appears in this definition, unlike the other two definitions.

7.2.2 Sub-threshold slope *S*

Sub-threshold slope S is defined [2] as:

$$S = \frac{2.3 I_{DS}}{\partial I_{DS} / \partial V_{GS}} \tag{165}$$

Using (153) and (165) and assuming $x_c \approx 0$ and neglecting the b_3 and c_3 terms, we get:

$$S = \frac{2.3\phi_t}{1 - B_1 \dfrac{\cos(\pi y_1 / \lambda_1)}{\cosh(\pi L / 2\lambda_1)}} \tag{166}$$

Thus, we see that S is greater than 60 mV/decade, and that it decreases towards 60 mV/decade as L/λ_1 is increased.

7.2.3 DIBL coefficient

The DIBL coefficient is defined [32] as:

$$DIBLC = -\frac{V_t\big|_{V_{DS}=1V} - V_t\big|_{V_{DS}=50mV}}{1-0.05} \tag{167}$$

As an illustration, we use (167) and the g_m-C_{ox} based V_t definition given by (161) to calculate this. The calculation involves calculating x_c at the two drain biases of 50 mV and 1 V. For simplicity, we assume that x_c is weakly dependent on V_{DS}. Specifically, we assume that x_c at 50 mV and that at 1 V are approximately equal to that at the approximate midpoint of V_{DS} = 0.5V. We finally get:

$$DIBLC = \frac{B_1 \cos\left(\pi y_1/\lambda_1\right)}{\sinh\left(\pi L/\lambda_1\right)} \sinh\left[\frac{\pi}{\lambda_1}\left(\frac{L}{2} + x_c^{avg}\right)\right] \tag{168}$$

where x_c^{avg} is the value of x_c at V_{DS} = 0.5V.

7.2.4 Comparison with TCAD device simulations

All the three analytical V_t definitions were used to predict the V_t rolloff with channel length, and the analytical V_t was compared with the V_t extracted from TCAD device simulations. The device simulations were done under similar conditions (a constant mobility of 300 cm^2/Vs, and the same geometry). Moreover, for the analytical model, the values of λ_1 and λ_3 were calculated for the given $(W_{fin}, t_{ox}, K_{ox})$ combination by numerically solving (146).

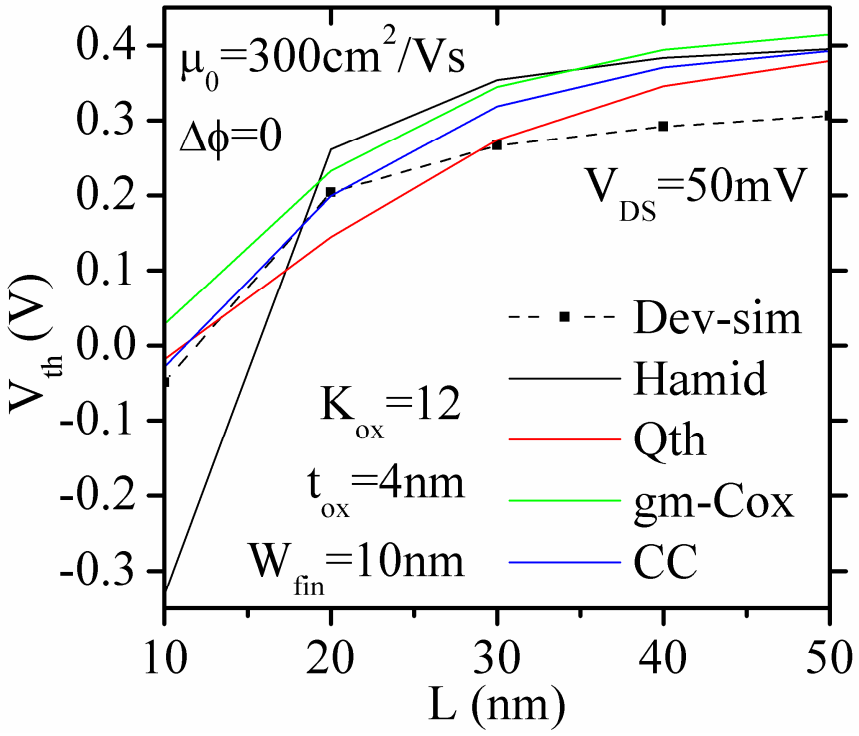


Figure 7-4: V_t roll-off for a device with W_{fin}=10 nm, a high-k gate dielectric of K_{ox}=12 and physical t_{ox}=4 nm at V_{DS}=50 mV.

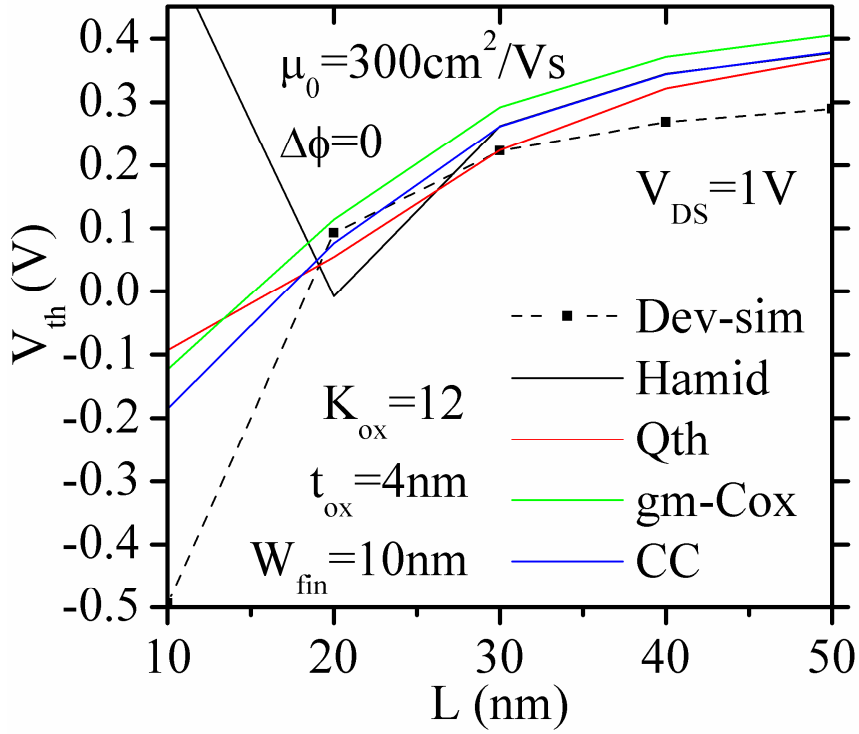


Figure 7-5: V_t roll-off for a device with W_{fin}=10 nm, a high-k gate dielectric of K_{ox}=12 and physical t_{ox}=4 nm at V_{DS}=1 V.

Fig. 7-4 and 7-5 show a plot of the V_t roll-off behaviour for a device with a high-k gate dielectric of K_{ox} = 12 and physical t_{ox} = 4 nm, at two different drain biases. For this figure, Q_{th} was calibrated from a L = 1 um device. The V_t from TCAD device simulations were extracted using a constant current definition of 0.1 uA/(H_{fin}/L). The curves labelled *Hamid* correspond to the model in [32]. The curves labelled *CC* correspond to the constant current analytical V_t model given by (164). The other curves are self-explanatory.

It is seen that all the three definitions used by us do show a V_t-rolloff behaviour, however none of the analytical models match very well with TCAD extracted V_t. We believe that the reasons for this are: *(a)* the approximations discussed above (153) that are made in deriving the V_t expression, and *(b)* the uncertainty as to how well the constant current V_t extraction method (from simulations) matches the various analytical V_t definitions. The constant current analytical V_t model given by (164) comes closest to it, especially at shorter channel lengths.

It should also be noted that the V_t model given in [32] shows an unphysical rollup at L=10 nm at V_{DS}=1 V.

Fig. 7-6 and 7-7 show the roll-up behaviour in DIBL and sub-threshold slope when plotted versus channel length L. Values extracted from 2D device simulations are also shown in the figures. The model-predicted DIBL values match well with TCAD-predicted values till about L=20 nm, whereas the model-predicted SS values match well till about L=40 nm. It should be noted that the L=10 nm device is a badly scaled device since W_{fin}=L in this case, and the approximations we have made are very questionable for this device. That explains the severe mismatch between model-predicted and TCAD-predicted values for this device.

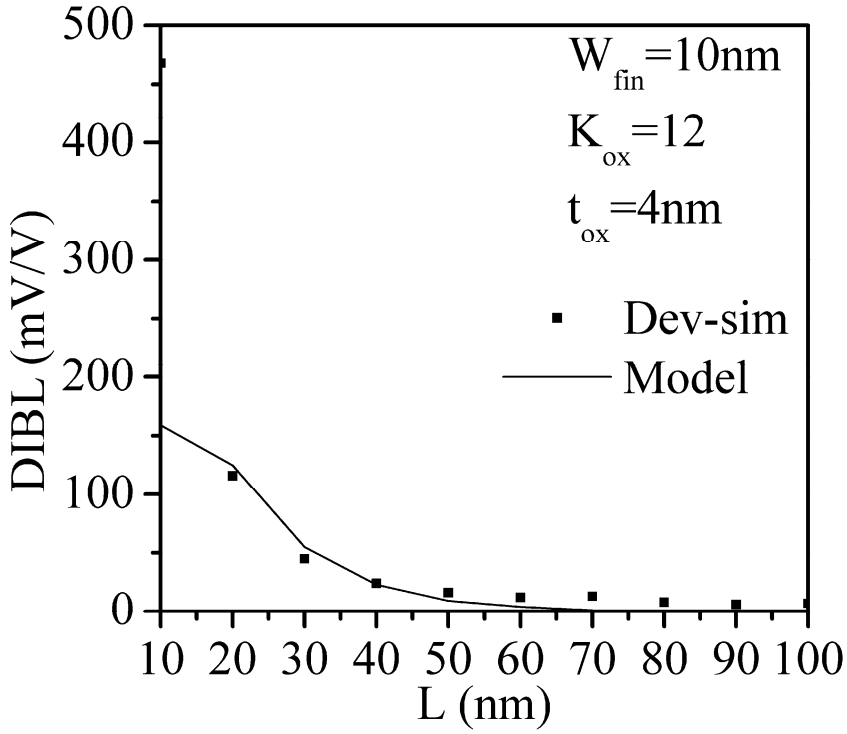


Figure 7-6: DIBL roll-up: analytical model compared with TCAD extracted values.

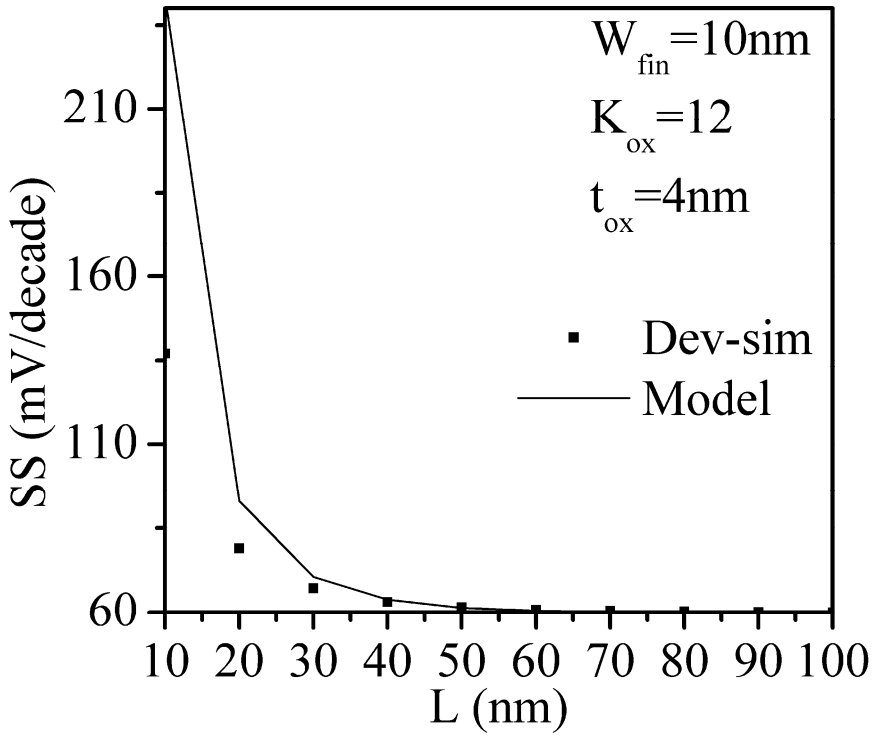


Figure 7-7: Sub-threshold slope roll-up: analytical model compared with TCAD extracted values.

7.3 Incorporation of DIBL effect into the core model

The conventional way to add 2D field effects into a core model is via a second order correction using fitting parameters [84]. But the introduction of fitting parameters using arbitrary trial and error methods would give rise to future problems not easily foreseeable. For instance, one problem could be that *local* fitting parameters may not scale well with channel length in a consistent manner. Hence as far as possible, any introduction of fitting parameters should arise out of an effort in *physically* modeling the 2D field effects. In order to incorporate 2D field effects into our core model, we leverage our effort in developing the physical 2D effects model described in the previous section.

The V_t model described till now cannot be integrated into our core model as is, because our core model is not threshold voltage based, but is rather an inversion charge based model. Hence we adopt a different technique to incorporate the 2D field effects into our core model.

7.3.1 Fin width modulation

It is well known [10] that due to volume inversion characteristics, the sub-threshold current is proportional to W_{fin}. It is also known that in strong inversion, the on current is weakly dependent on W_{fin}. Hence we adopt the technique of artificially modifying W_{fin} to a value such that the off current predicted by the core model matches that predicted by the 2D model.

Compact models need to be very simple models for computational efficiency reasons. Considering then the findings summarized in section 7.1.5, and making the approximation that $x_c \approx 0$, and approximating $2\sinh(\pi L/2\lambda_1) \approx \exp(\pi L/2\lambda_1)$, we can simplify (153) as (169). A simple calculation shows that the error in approximating $2\sinh(\pi L/2\lambda_1) \approx \exp(\pi L/2\lambda_1)$ is greater than 1% only when $(L/\lambda_1) < 1.5$, and that the error is greater than 5% only when $L < \lambda_1$. At these ultra short channel lengths, one would anyway have unacceptably bad short channel behavior [29] such that the device is not useful any more. Thus, for practical devices, these approximations are quite valid. Making such approximations in the drain current expression (153), we get:

$$I_{DS}^{2D} = \frac{q\mu_0 n_i W_{fin}\phi_t}{L} e^{(V_{GS}-\Delta\phi+(b_1+c_1)\cos(\pi y_1/\lambda_1)e^{-\pi L/2\lambda_1})/\phi_t}\left(1-e^{-V_{DS}/\phi_t}\right) \tag{169}$$

The superscript *2D* in the LHS denotes that this is a current derived from a 2D model.

Next let us consider the sub-threshold current as predicted by the 1D core model developed in the earlier chapters. We first translate from the inversion charge domain to the terminal voltage domain, as far as the state variable is concerned. In sub-threshold, we have already shown that an approximate solution of the input voltage equation is:

$$\beta_2 \approx \beta_{2sm} = u^2 = \frac{\beta^2 W_{fin}^2}{4} e^{(V_{GS}-\Delta\phi-\phi_{fn})/\phi_t} \tag{170}$$

In sub-threshold, we can neglect the $\beta_{2s}^2 - \beta_{2d}^2$ term in the a_{1d} expression (65) in the core model. Also, we make the approximation that $ln(1+x) \approx x$ for small values of x (x representing the difference in β_2 between the source and drain ends, both of which are very small in sub-threshold) in the a_{1d} expression (65) in the core model. Using (170) for the remaining terms in the expression for a_{1d}, we get the sub-threshold current predicted by the 1D core model (63) as:

$$I_{DS}^{1D} = \frac{q\mu_0 n_i W_{fin}^{equiv1D}\phi_t}{L} e^{(V_{GS}-\Delta\phi)/\phi_t}\left(1-e^{-V_{DS}/\phi_t}\right) \tag{171}$$

To get the same off current I_{off}, we equate (169) and (171) at $V_{GS} = 0$. We denote the values of b_1 and c_1 at $V_{GS} = 0$ as b_{10} and c_{10}. We get an expression for $W_{fin}^{equiv1D}$ as:

$$W_{fin}^{equiv1D} = W_{fin}\exp\left\{\frac{\left[\left(b_{10}+c_{10}\right)\cos\left(\pi y_1/\lambda_1\right)e^{-\pi L/2\lambda_1}\right]}{\phi_t}\right\} \tag{172}$$

where expressions for b_{10} and c_{10} can be got from (150) as:

$$b_{10} = B_1\left[\frac{E_g}{2q} + \Delta\phi\right]$$
$$c_{10} = B_1\left[\frac{E_g}{2q} + V_{DS} + \Delta\phi\right] \qquad (173)$$
$$B_1 = \frac{2\lambda_1^2 \tan\left(\pi t_{ox}/\lambda_1\right)\sin\left(\frac{\pi W_{fin}}{2\lambda_1}\right)}{\pi^2 t_{ox}\left[\frac{W_{fin}}{2} + \frac{t_{ox}\sin\left(\pi W_{fin}/\lambda_1\right)}{\sin\left(2\pi t_{ox}/\lambda_1\right)}\right]}$$

In our model, we deem λ_1 as a model parameter in order to aid in yielding a better prediction of the drain current, even though it is physically given by the relation (146).

Next, as we have already seen, there is a mismatch in the analytical versus TCAD device simulation currents, arising out of the approximations made in evaluating the nested integrals. To account for this mismatch, we add two fitting parameters s_{t1} and s_{t2} as follows:

- We replace the V_{DS} term in (173) by $s_{t1}.V_{DS}$
- We replace the $(b_{10}+c_{10})$ term in (172) by $(b_{10}+c_{10})\ s_{t2}$

During parameter extraction, if the extracted values of the s_{t1} and s_{t2} parameters are close to 1, then that indicates that the 2D model is a valid model.

Even though the on-current in strong inversion is only weakly dependent on W_{fin}, it is observed that allowing the fin width to be artificially modified under all bias regimes including strong inversion yields a poor match when trying to extract the model parameters. Hence we use a smoothing function to modulate W_{fin} only at low V_{GS}, but smoothly revert it to the original W_{fin} at high V_{GS}. For this, we define a smoothing function s_m as given by (174).

$$s_m = P_2\beta_{2s}\left(1+\left(P_2\beta_{2s}\right)^{A_{X2}}\right)^{(-1/A_{X2})} \qquad (174)$$

Here P_2 and A_{X2} are model parameters. Parameter P_2 controls the V_{GS} bias point at which the scheme for incorporating DIBL effects stops using a fin-width-modulation technique, and A_{X2}

controls the sharpness of this transition. β_{2s} is a dimensionless quantity and is indicative of the gate bias and so is ideally suited to this task; it is very small in sub-threshold, and can be large in strong inversion. Our technique results in s_m behaving like a unit step function. We then define the modulated fin width as given by (175):

$$W_{fin2} = W_{fin}^{equiv1D} + \left(W_{fin} - W_{fin}^{equiv1D}\right) \cdot s_m \tag{175}$$

This effectively ensures that the fin width modulation is negligible in strong inversion where β_{2s} is large. W_{fin2} is the modified fin width that is used in all expressions in place of W_{fin} in the core model.

7.3.2 Sub-threshold slope ideality factor

Using (171) and (165), it is clear that the sub-threshold slope in the core model is the ideal $2.3\varphi_t$ = 60 mV/decade. However, if the φ_t term in the input voltage equation is replaced by $n\varphi_t$, then it can be easily seen that the sub-threshold slope becomes $2.3n\varphi_t$ mV/decade. We call this the *modified 1D sub-threshold slope*.

From (166) which describes the sub-threshold slope as predicted by the 2D model, we can see that it equals the *modified 1D sub-threshold slope* if n is given by

$$n = \frac{1}{1 - B_1 \dfrac{\cos\left(\pi y_1 / \lambda_1\right)}{\cosh\left(\pi L / 2\lambda_1\right)}} \approx \frac{1}{1 - 2B_1 \cos\left(\pi y_1 / \lambda_1\right) e^{-\left(\pi L / 2\lambda_1\right)}} \tag{176}$$

Having found n, we simply replace the φ_t term in the input voltage equation (and therefore in its closed form approximate solution) by $n\varphi_t$.

However, as we have already seen, there is a mismatch in the analytical versus TCAD device simulation currents, arising out of the approximations made in evaluating the nested integrals. To account for this mismatch, we add a fitting parameter n_{if} related to the ideality factor as

follows: Instead of replacing the φ_t term in the input voltage equation (and therefore in its closed form approximate solution) simply by $n\varphi_t$, we replace it by $n\phi_t \cdot n_{if}$.

During parameter extraction, if the extracted value of the n_{if} parameter is close to 1, then that indicates that the 2D model is a valid model.

Again, in order to limit this modulation of φ_t only in sub-threshold, we use a technique similar to (175), as stated below:

$$\phi_{t2} = n\phi_t \cdot n_{if} + \left(\phi_t - n\phi_t \cdot n_{if}\right) \cdot s_m \tag{177}$$

Φ_{t2} is the modified thermal voltage that is used in all expressions in place of Φ_t in the core model.

7.3.3 DIBL effect in the above-threshold regime

The fin width modulation and sub-threshold slope modulation techniques described above would work only in the sub-threshold regime. While it is true that the DIBL effect is the maximum in the sub-threshold regime, it is nonetheless non-negligible even in the above-threshold regime, more so for shorter channel length devices. In order to incorporate DIBL effects into our core model in the above-threshold regime, we borrow an empirical technique from the PSP model [73], where the DIBL effect is simply modeled as a reduction ($\Delta\Delta\varphi$) in $\Delta\varphi$ by a quantity proportional to the absolute value of V_{DS}.

$$\Delta\Delta\phi = -c_f \cdot V_{dsx} \tag{178}$$

V_{dsx} [73] in (178) is simply $\sqrt{V_{DS}^2 + 0.01} - 0.1$, and is thus approximately equal to $|V_{DS}|$; and c_f is a model parameter.

7.4 Validation of the complete model including all effects

The complete model including support for all the effects developed till now, was then validated versus TCAD device simulation results, for ultra-short channel devices.

2D device simulations were done for five devices with channel lengths ranging from L=20 nm to 50 nm, all having a SiO_2 gate oxide with t_{ox}=1 nm and fin width W_{fin}=10 nm. A mid-gap work function with zero barrier (with respect to intrinsic silicon) was used for the gate electrode. The basal mobility was the default 1417 cm^2/Vs. The default Lombardi model was turned on for low-field mobility degradation and the default Canali model was turned on for velocity saturation.

The model parameters were extracted from the TCAD device simulation data using the HPSO algorithm [72]. The extracted parameter values for each channel length device are summarized in Table 7-1. Ideally, the model parameters should be the same for all channel lengths (if they are global parameters), or should show a consistent trend with channel length (if they are local parameters). Since the (W_{fin} - t_{ox}) combination is the same for all the 4 devices, ideally the first order generalized scale length λ_1 should also have been the same for all four devices. Even though this is not so, a consistent trend is observed in that it decreases as channel length decreases. We attribute the inconsistency in the trend of some parameters to a non-optimal parameter extraction strategy, such as the lack of usage of the most reasonable default values in some extraction steps. Furthermore, we believe that in some cases the extraction/optimization problem could perhaps be an under-constrained one, resulting in multiple combinations that yield a good fit. This could be fixed by fixing some parameters as extracted from long channel devices (such as basal mobility). We leave these and other such geometrical scaling problems as open for future work.

Fig. 7-8 - 7-23 plot various characteristics for all the devices considered, showing the match between the model-predicted values and TCAD results. The curves corresponding to the analytical model in Fig. 7-8 - 7-23 use the extracted parameters. As can be seen, the match is very good for such short channel devices in sub- as well as above-threshold. The DIBL effect can be clearly seen in the I_d-V_g curves when one notices the implicit threshold voltage reduction at high V_{DS}. The g_m-V_g curves exhibit a downward trend at high V_{GS} that is more

severe at low V_{DS}, which is typical of low-field mobility degradation. A slight kink is seen in the g_m-V_g curve at high V_{DS}, and this is attributed to an unusually large value extracted for the A_{X2} smoothening parameter, which we believe can be fixed by suitable adjustments in the parameter extraction strategy. The I_d-V_d and g_{ds}-V_d curves show a low but nevertheless non-zero output conductance at high V_{DS}, which is a result of incorporating channel length modulation and DIBL effects in the model. The g_{DS}-V_d curves show the largest mismatch at low V_{DS} and moderately low V_{GS}, and this could be attributed to the fact that we had used *Method 3* to calculate V_{DSat}, which assumed that the drift component is dominant and which is a questionable assumption under those bias conditions. Many of the model parameters extracted such as the work function difference, basal mobility, saturation velocity, scale length, and some DIBL-related fitting parameters, were close to their *ideal* value. The *ideal* value of a *physical* parameter is the value as used in the TCAD tool, and the *ideal* value of a *fitting* parameter is a value that effectively implies non-usage of that fitting parameter; some parameters like E_{sat} , A_X, etc do not correspond to a similar entity in the TCAD tool and they therefore do not have any designated *ideal* value.

Table 7-1: Extracted values of the model parameters.

	L=20 nm	**L=30 nm**	**L=40 nm**	**L=50 nm**	***Ideal value***
WFIN (W_{fin}) (nm)	10	10	10	10	10
TOX (t_{ox}) (nm)	1	1	1	1	1
NA (N_a) (cm^{-3})	10^{15}	10^{15}	10^{15}	10^{15}	10^{15}
L1 (λ_1) (nm)	13.37	15.85	18	18.5	15.47
DELTA_PHI ($\Delta\varphi$) (V)	-0.0012	0.018	0.008	-0.007	0
MU0 (μ_0) (cm^2/Vs)	1413	1232	1195	1286	1417

CS (c_s)	1.2	0.9	0.812	0.95	-
ETA_MU (η)	0.65	0.53	0.59	0.53	-
E0_MU (E_0) (x 10^4 V/cm)	9.6	9.4	11	9	-
THETA_MU (θ)	0.98	0.93	0.91	0.9	-
VSAT (v_{sat}) (x 10^7 cm/s)	1.12	0.97	0.89	0.82	1.07
P1 (P_1)	0.99	0.68	0.56	0.76	1
AX (A_X)	1.51	1.56	1.55	1.53	-
ESAT (E_{sat}) (V/cm)	$7.7x10^8$	$8.1x10^6$	$9.9x10^9$	$9.9x10^9$	-
THESATG (θ_{SatG})	0.074	0.041	0.016	0.005	0
SUBT1 (s_{t1})	0.4	0.31	0.27	0.16	1
SUBT2 (s_{t2})	0.89	1	1	1	1
CF (c_f)	0.043	0.00012	$4.1x10^{-6}$	0.0011	0
AX2 (A_{X2})	0.47	1.51	1.77	0.41	-
P2 (P_2)	3.41	5	5	5	-
NIF (n_{if})	1.002	1.005	1.018	1.046	1

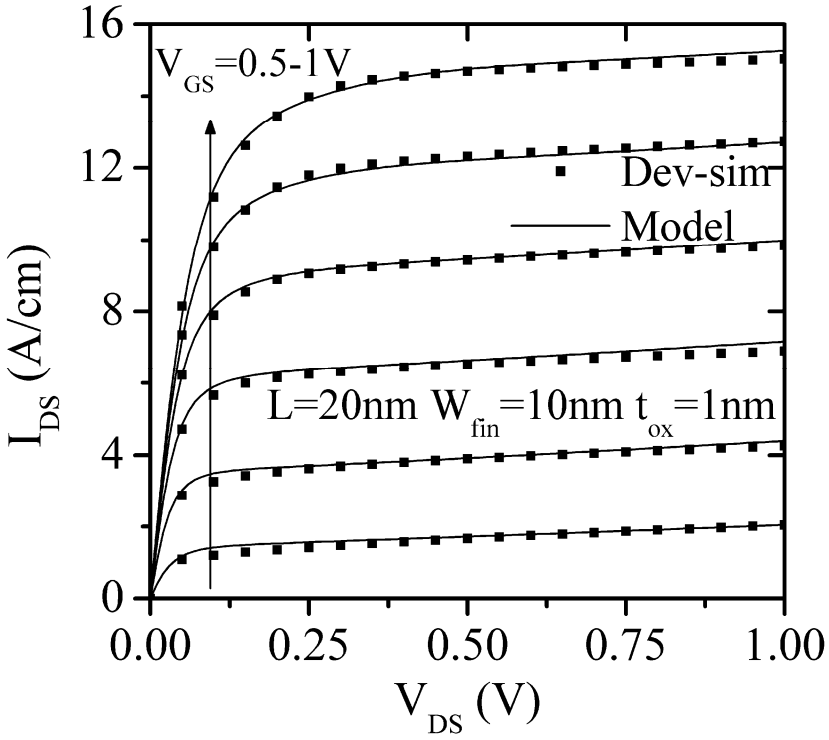


Figure 7-8: I_d-V_d curves for a L=20 nm device using the final model.

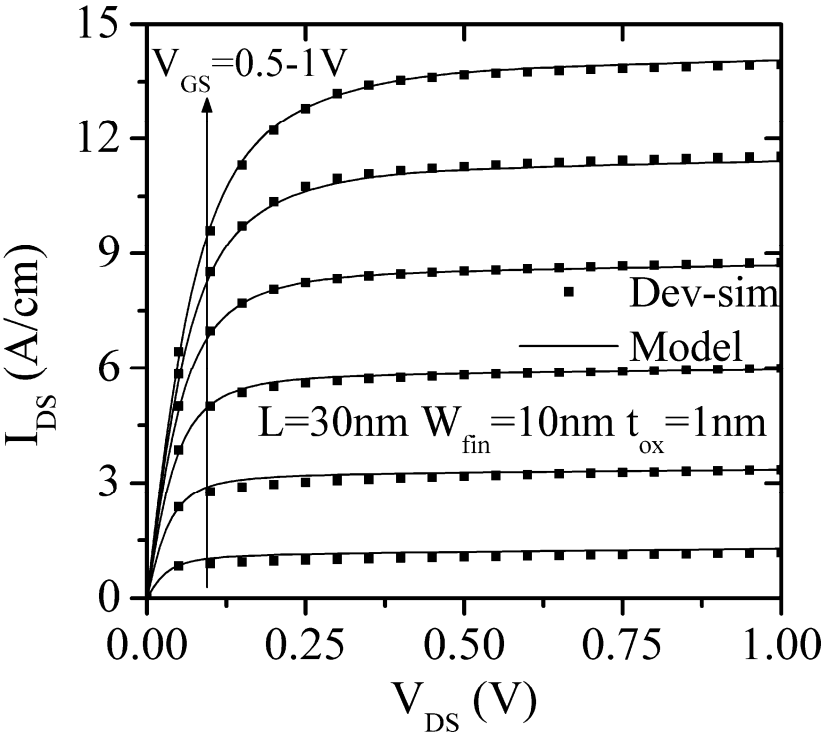


Figure 7-9: I_d-V_d curves for a L=30 nm device using the final model.

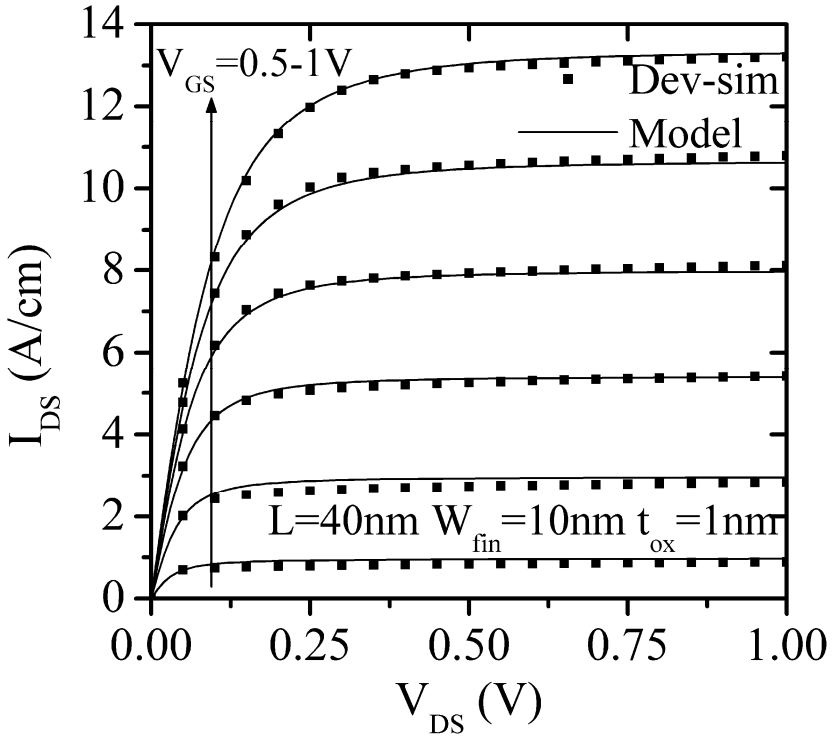


Figure 7-10: I_d-V_d curves for a L=40 nm device using the final model.

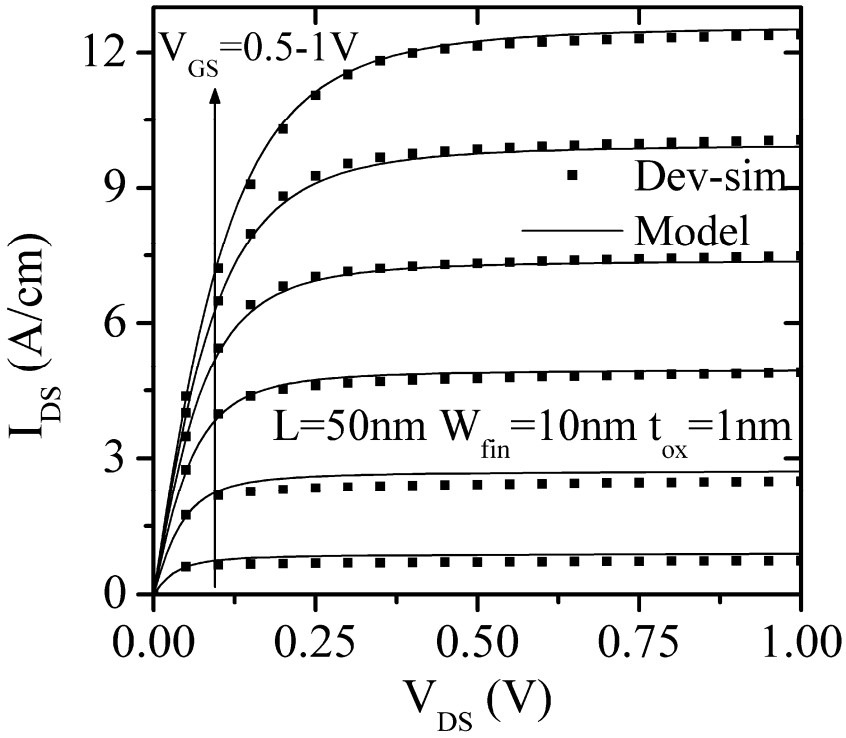


Figure 7-11: I_d-V_d curves for a L=50 nm device using the final model.

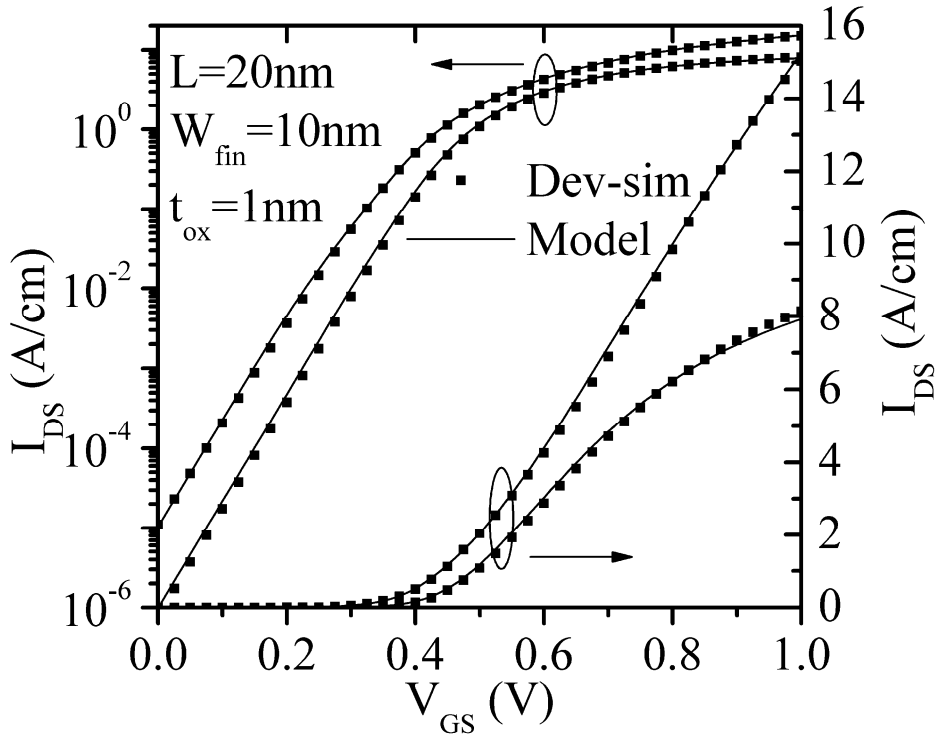


Figure 7-12: I_d-V_g curves for a L=20 nm device using the final model.

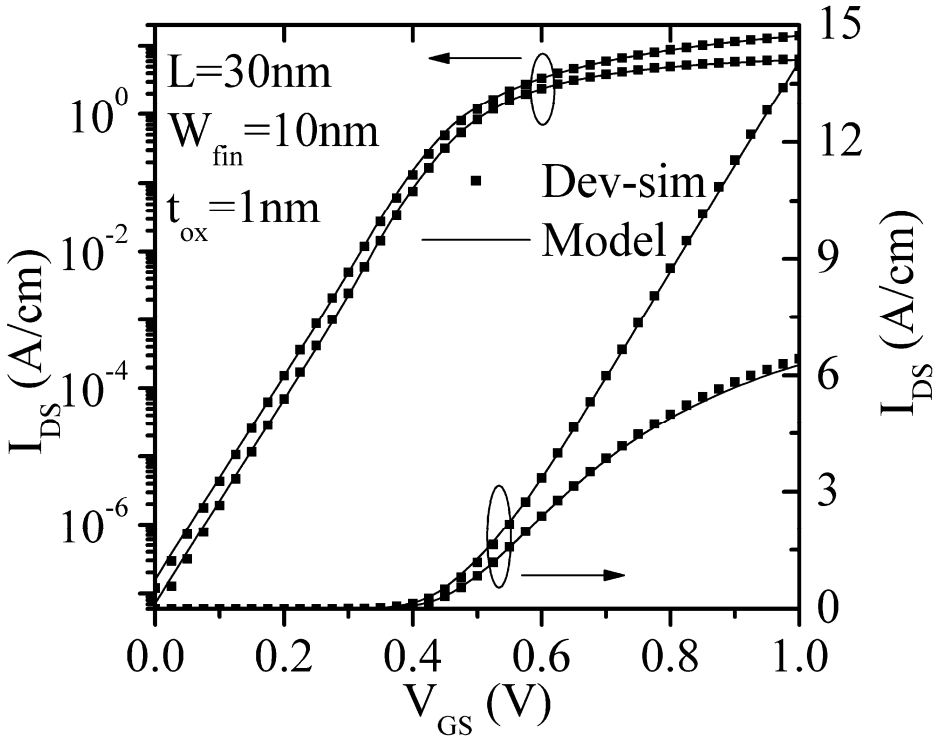


Figure 7-13: I_d-V_g curves for a L=30 nm device using the final model.

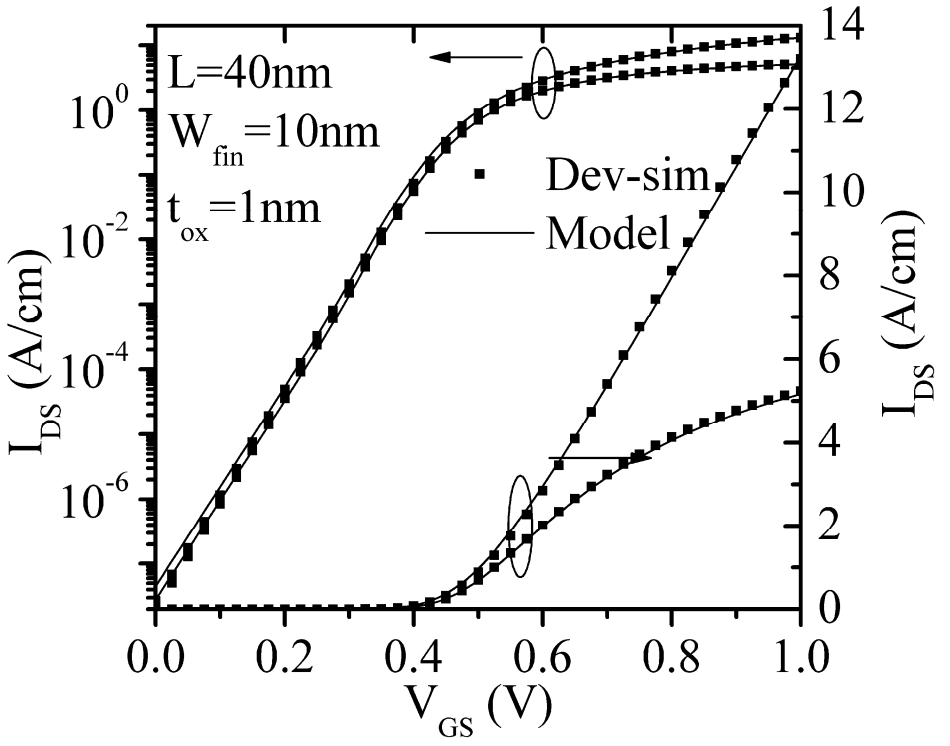


Figure 7-14: I_d-V_g curves for a L=40 nm device using the final model.

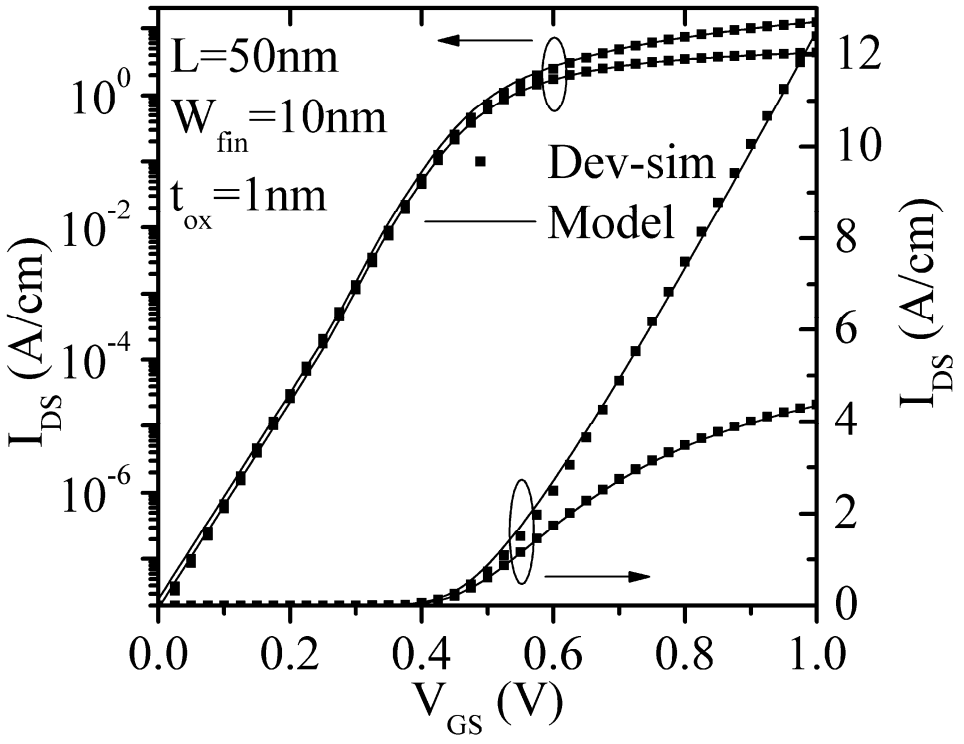


Figure 7-15: I_d-V_g curves for a L=50 nm device using the final model.

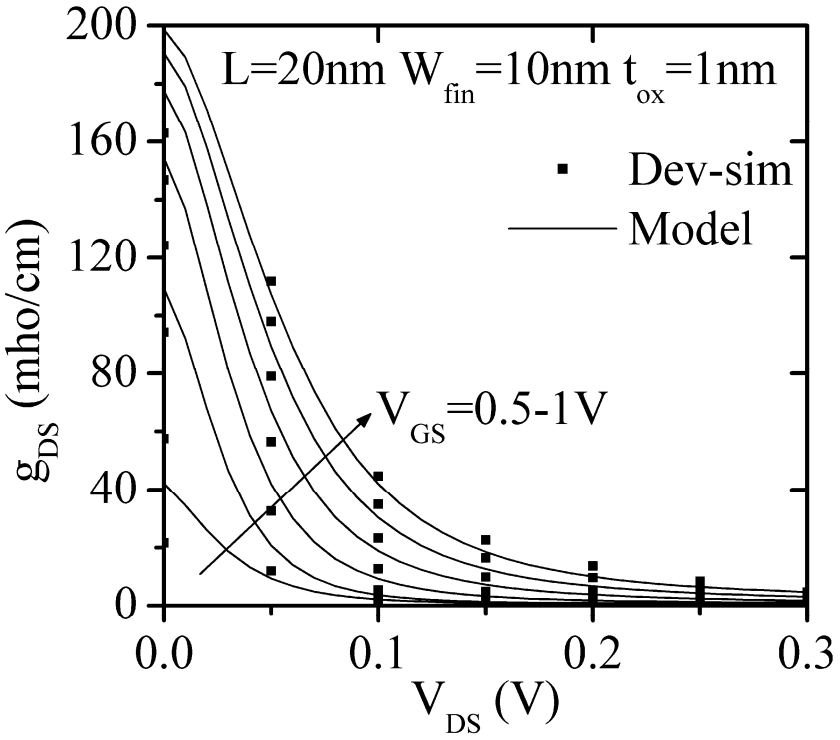


Figure 7-16: g_{DS}-V_d curves for a L=20 nm device using the final model.

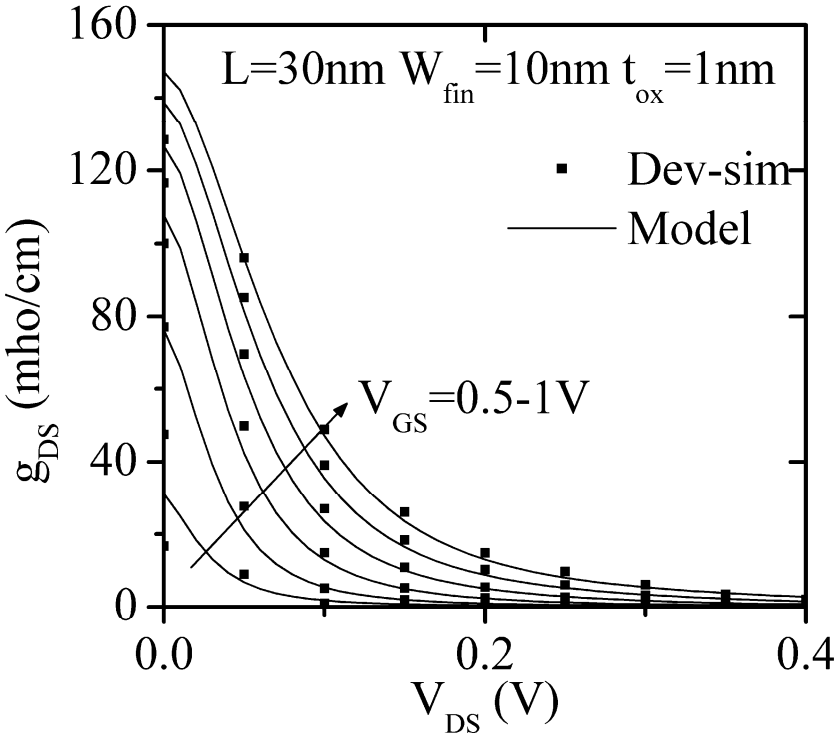


Figure 7-17: g_{DS}-V_d curves for a L=30 nm device using the final model.

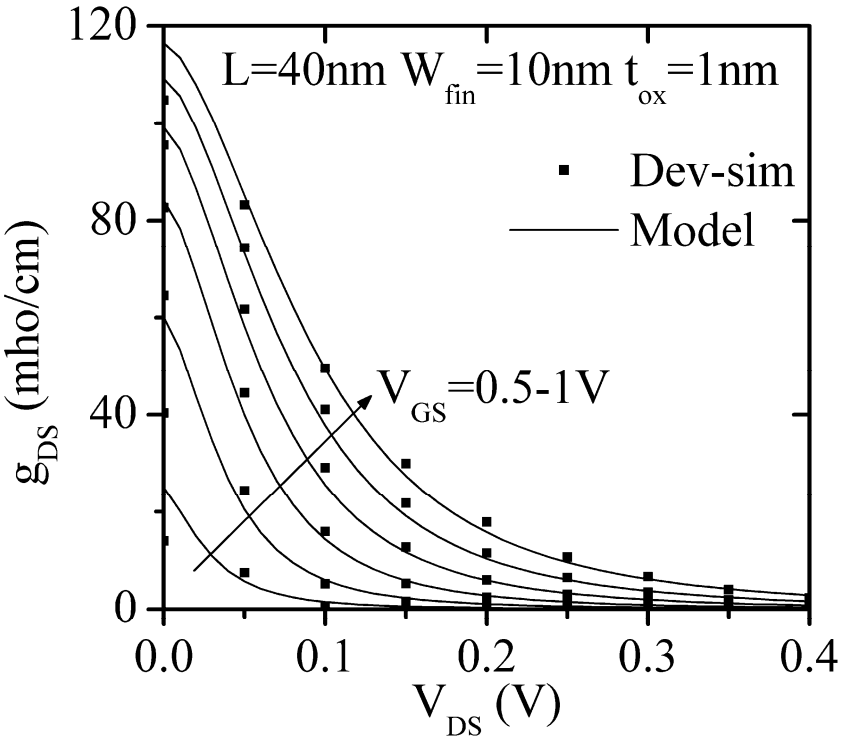


Figure 7-18: g_{DS}-V_d curves for a L=40 nm device using the final model.

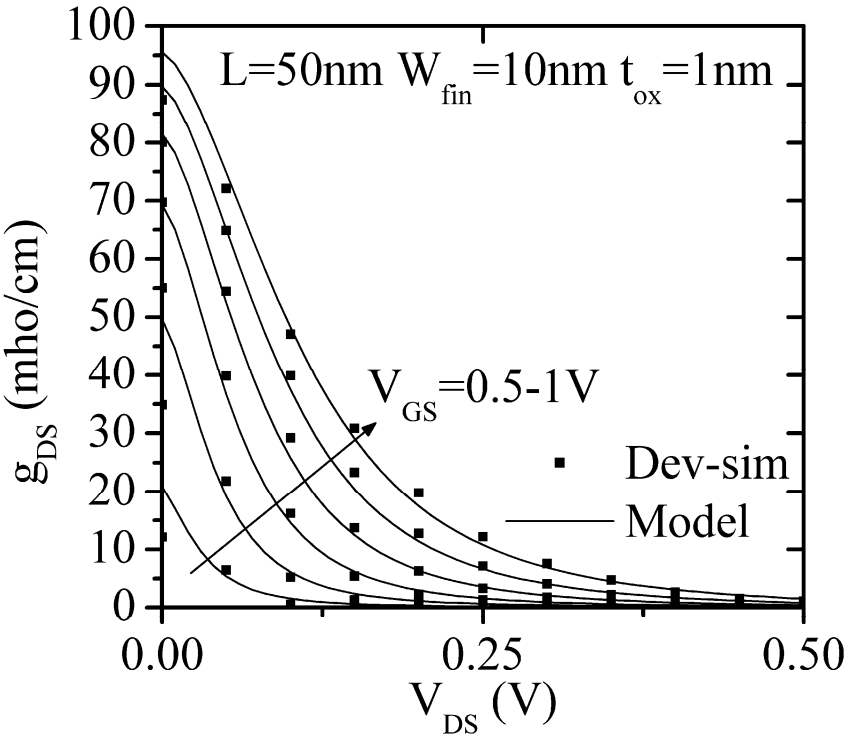


Figure 7-19: g_{DS}-V_d curves for a L=50 nm device using the final model.

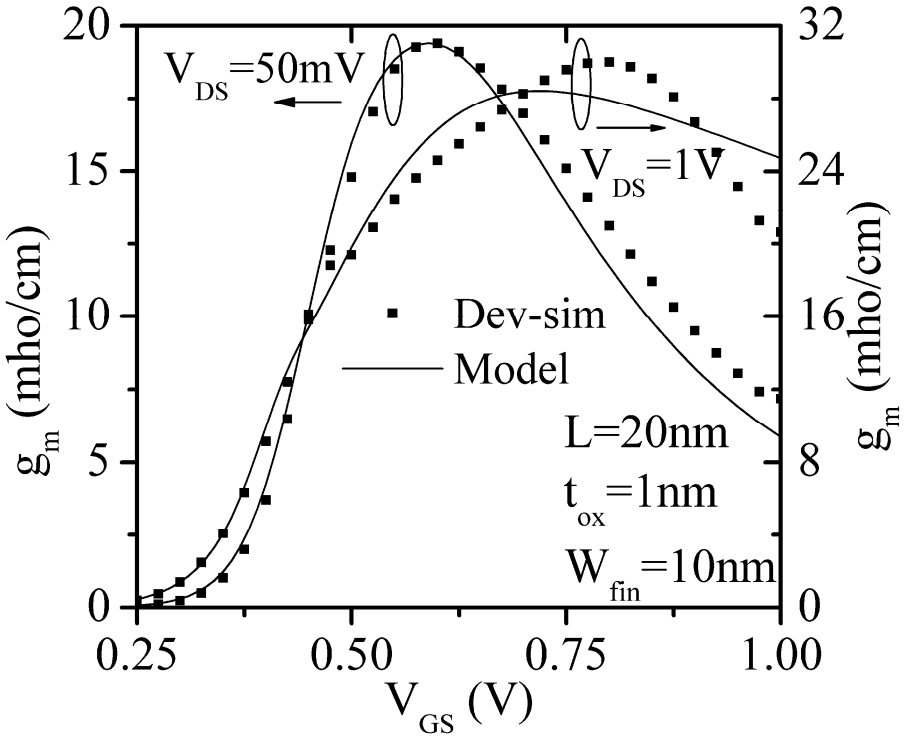


Figure 7-20: g_m-V_g curves for a L=20 nm device using the final model.

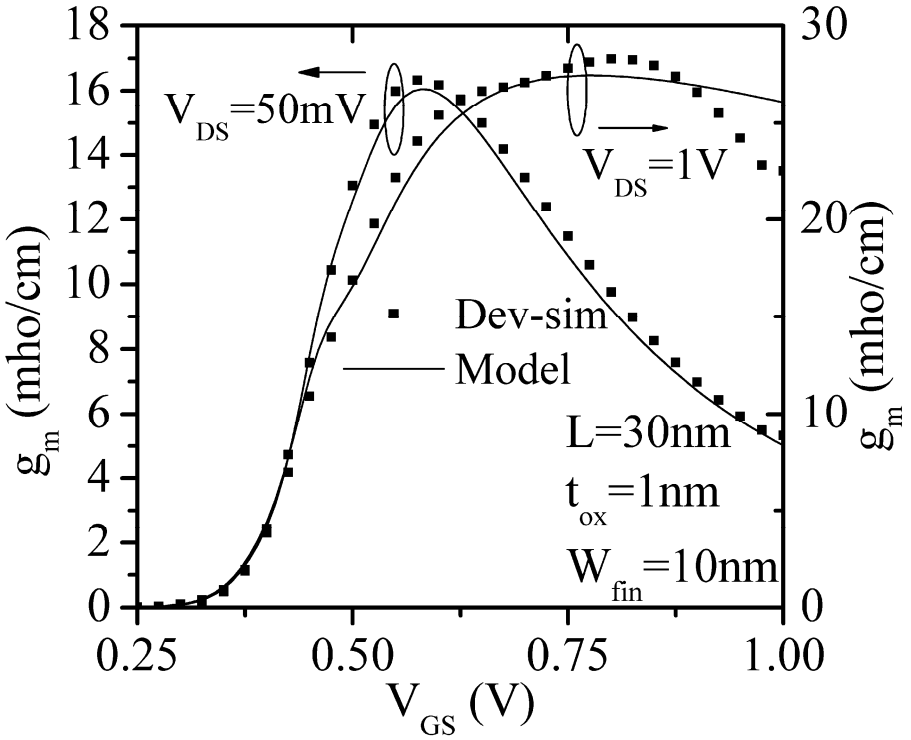


Figure 7-21: g_m-V_g curves for a L=30 nm device using the final model.

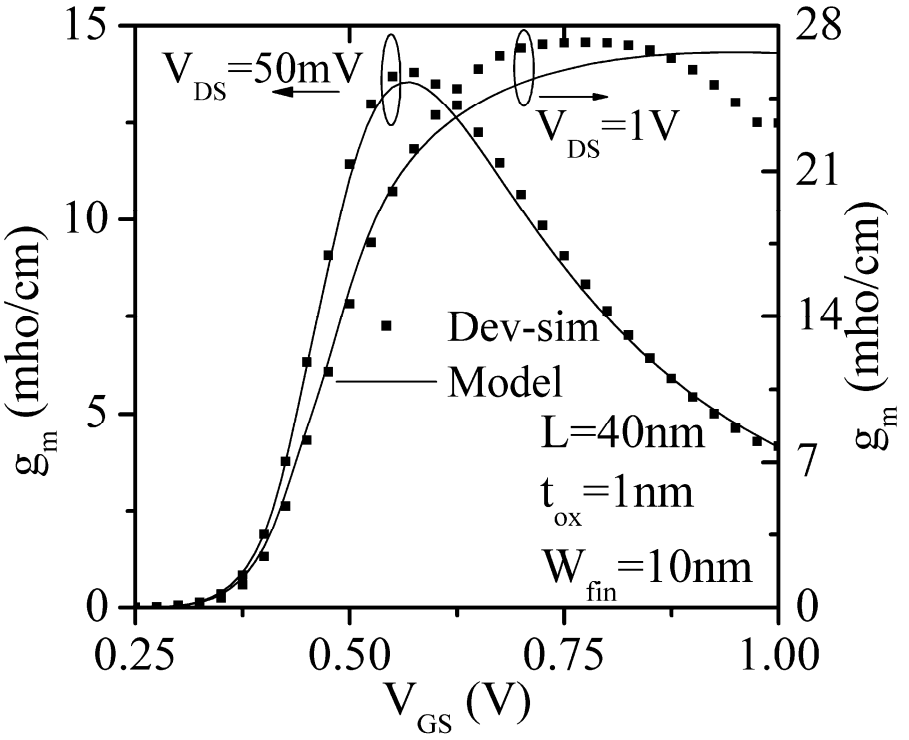


Figure 7-22: g_m-V_g curves for a L=40 nm device using the final model.

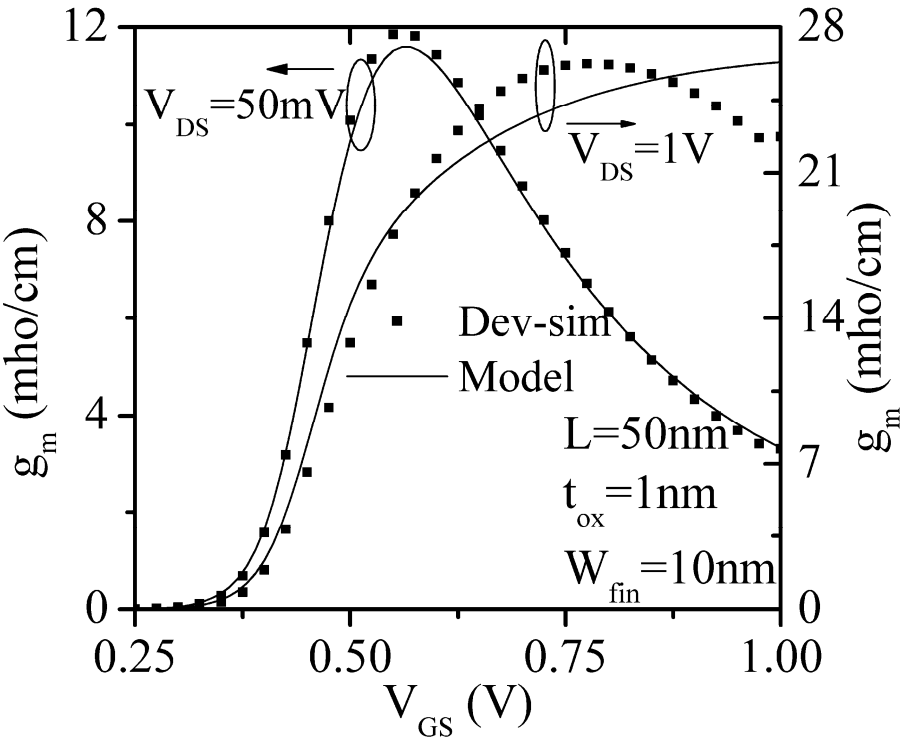


Figure 7-23: g_m-V_g curves for a L=50 nm device using the final model.

7.4.1 Parameter extraction details

The model parameters were extracted from the TCAD device simulation data using the HPSO algorithm [72] in four steps. The model has 21 parameters, of which 3 were set to known values (**WFIN**, **TOX** and **NA**) and the others were extracted in four steps from I_d-V_g (including g_m-V_g) and I_d-V_d TCAD data.

In the first step, focus was limited to the sub-threshold regime (V_{GS} < 0.4V) at low V_{DS}=50mV and the work function difference $\Delta\varphi$, first order scale length λ_1, s_{t2} and n_{if} were extracted from I_d-V_g and g_m-V_g data. Reasonable default values were used for the other model parameters in this step.

In the second step, focus was extended to the entire V_{GS} range though still at low V_{DS}=50mV and the mobility parameters and the smoothing function parameters P_2 and A_{X2} were extracted from I_d-V_g and g_m-V_g data, considering the parameters extracted up until the prior step. Reasonable default values were used for the other as-yet-unextracted model parameters.

In the third step, focus was limited to the sub-threshold regime (V_{GS} < 0.4V) at high V_{DS}=1V and the DIBL parameters s_{t1} and c_f were extracted from I_d-V_g and g_m-V_g data, considering the parameters extracted up until the prior step. Reasonable default values were used for the other as-yet-unextracted model parameters.

In the fourth step, the velocity saturation and CLM parameters were extracted from I_d-V_d data considering the parameters extracted up until the prior steps. This step focused on I_d-V_d data for V_{GS} >= 0.5V.

Some details about the usage of the parameter extraction tool are discussed in Appendix-B.

7.5 Summary

In this chapter, an analytical model for V_t, S and DIBL for ultra short channel SDGFETs was presented considering 2D electrostatics. In the later part of the chapter, it was shown how this 2D modeling effort was leveraged to incorporate 2D field effects into the core model.

The final model is predominantly a physics based model, with a few fitting parameters added in order to aid in a better match with device simulation results. A good match was shown between model predicted values and 2D device simulations for devices as small as L=20 nm.

Chapter 8
Conclusions and Future Scope

We have presented a physics-based, completely closed-form, single-piece, inversion-charge-based drain current model for an ultra short-channel, symmetrically driven symmetric double-gate MOSFET that includes support for low-field mobility degradation, velocity saturation, channel length modulation, body doping and DIBL effects. It improves over existing models by including velocity saturation effects self-consistently and in a manner that Gummel-symmetry compliance is met. The superiority of our core model over other existing models that consider velocity saturation has also been shown.

For modeling 2D field effects like DIBL in nanoscale DGFETs, we have built up on an existing 2D analysis that solved for the 2D electrostatic potential, after evaluating the merits of that 2D analysis over other similar analyses. Using this approach, we have developed models for threshold voltage roll-off, sub-threshold slope and the DIBL coefficient, which are useful metrics from a device design perspective. We have also leveraged this 2D modeling work to incorporate the 2D field effects into our core model.

Finally, a good match has been shown between model predicted values and 2D device simulation results for devices as short as 20 nm, which is shorter than the shortest channel length SDGFET for which validation of a closed-form model has ever been shown in the literature. Gummel symmetry compliance was also demonstrated.

Some of the points that could be taken up for future work are:

1. The model developed till now is a DC model. In order to be able to do transient and AC analysis in a circuit simulator, expressions for terminal charges also need to be developed. The discussion below (63) lays the groundwork for how this can be done using the Ward-Dutton charge partitioning scheme [52] once we have the *spatial* variation of $Q_i(x)$ using (48). The evaluation of the Ward-Dutton integral is complicated under conditions of constant mobility itself. Evaluating it for a redefined core model which includes a lateral-field dependent mobility is harder still. Then, considering that the Ward-Dutton charge partitioning scheme was derived under conditions of constant mobility (and it would therefore amount to stretching it beyond its applicability-domain if one were to apply it to conditions of variable mobility), one could add expressions for terminal charges to the model presently developed by simply borrowing the simplified terminal charge expressions as presented in the PSP-FinFET model [22], which were developed under constant mobility conditions.

2. Other realistic effects need to be accounted for, such as quantum mechanical effects (gate leakage, inversion layer quantization, and quantization due to structural confinement in case of ultra-thin fins), parasitic series resistance, junction leakage, gate induced drain leakage (GIDL), 3D parasitic effects that are especially severe in FinFETs, gate resistance models for accurate RF simulations, velocity overshoot, as well as improved mobility models (that consider remote coulomb scattering, degraded mobility with high-k gate dielectric interfaces, anisotropic mobility depending on the orientation of conduction-surface, etc).

Each of these physical effects can be added as second order effects by adapting expressions from more mature models that model that physical effect adequately enough.

REFERENCES

[1] R. H. Dennard, F. H. Gaensslen, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE J. Solid-State Circuits*, vol. 9, no. 5, pp. 256-268, 1974.

[2] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices* Cambridge University Press, 1998.

[3] H. S. P. Wong, "Beyond the conventional transistor," *IBM Journal of Research and Development*, vol. 46, no. 2-3, pp. 133-168, 2002.

[4] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, and R. Chau, "A 45nm Logic Technology with High-k+ Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging," *IEDM Tech. Dig.*, pp. 247-250, 2007.

[5] E. J. Nowak, I. Aller, T. Ludwig, K. Kim, R. V. Joshi, C. T. Chuang, K. Bernstein, and R. Puri, "Turning silicon on its edge- double gate CMOS/FinFET technology," *IEEE Circuits and Devices Magazine*, vol. 20, no. 1, pp. 20-31, 2004.

[6] X. Xi, "BSIM 4.5.0 MOSFET Model Users Manual," Univ.of California, Berkeley, 2004. Available: http://www-device.eecs.berkeley.edu/bsim3/BSIM4/BSIM450/doc/BSIM450_Manu.tar.

[7] G. Gildenblat, X. Li, W. Wu, H. Wang, A. Jha, R. van Langevelde, G. D. J. Smit, A. J. Scholten, and D. B. M. Klaassen, "PSP: An Advanced Surface-Potential-Based MOSFET Model for Circuit Simulation," *IEEE Trans. Electron Devices*, vol. 53, no. 9, pp. 1979-1993, 2006.

[8] G. Pei, W. Ni, A. V. Kammula, B. A. Minch, and E. C. C. Kan, "A physical compact model of DG MOSFET for mixed-signal circuit applications-part I: model description," *IEEE Trans. Electron Devices*, vol. 50, no. 10, pp. 2135-2143, 2003.

[9] M. V. Dunga, C. H. Lin, X. Xi, D. D. Lu, A. M. Niknejad, and C. Hu, "Modeling Advanced FET Technology in a Compact Model," *IEEE Trans. Electron Devices*, vol. 53, no. 9, p. 1971, 2006.

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

[11] J. He, X. Xuemei, M. Chan, C. H. Lin, A. Niknejad, and C. Hu, "A non-charge-sheet based analytical model of undoped symmetric double-gate MOSFETs using SPP approach," *Proc. International Symp. Quality Electronic Design*, pp. 45-50, 2004.

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

[13] A. S. Roy, J. M. Sallese, and C. C. Enz, "A closed-form charge-based expression for drain current in symmetric and asymmetric double gate MOSFET," *Solid State Electron.*, vol. 50, no. 4, pp. 687-693, 2006.

[14] H. Lu and Y. Taur, "An analytic potential model for symmetric and asymmetric DG MOSFETs," *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 1161-1168, 2006.

[15] A. Ortiz-Conde, F. J. Garcia Sanchez, and J. Muci, "Rigorous analytic solution for the drain current of undoped symmetric dual-gate MOSFETs," *Solid State Electron.*, vol. 49, no. 4, pp. 640-647, 2005.

[16] J. He, F. Liu, J. Zhang, J. Feng, J. Hu, S. Yang, and M. Chan, "A carrier-based approach for compact modeling of the long-channel undoped symmetric double-gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, no. 5, pp. 1203-1209, 2007.

[17] Z. Zhu, X. Zhou, S. C. Rustagi, G. H. See, S. Lin, G. Zhu, C. Wei, and J. Zhang, "Analytic and explicit current model of undoped double-gate MOSFETs," *Electronics Letters*, vol. 43, no. 25, pp. 1464-1466, Dec.2007.

[18] Z. Zhu, X. Zhou, K. Chandrasekaran, S. C. Rustagi, and G. H. See, "Explicit Compact Surface-Potential and Drain-Current Models for Generic Asymmetric Double-Gate Metal-Oxide-Semiconductor Field-Effect Transistors," *Japanese Journal of Applied Physics*, vol. 46, no. 4B, pp. 2067-2072, 2007.

[19] M. Wong and X. Shi, "Analytical IV relationship incorporating field-dependent mobility for a symmetrical DG MOSFET with an undoped body," *IEEE Trans. Electron Devices*, vol. 53, no. 6, pp. 1389-1397, 2006.

[20] D. M. Caughey and R. E. Thomas, "Carrier mobilities in silicon empirically related to doping and field," *Proc. IEEE*, vol. 55, no. 12, pp. 2192-2193, 1967.

[21] K. Joardar, K. K. Gullapalli, C. C. McAndrew, M. E. Burnham, and A. Wild, "An improved MOSFET model for circuit simulation," *IEEE Trans. Electron Devices*, vol. 45, no. 1, pp. 134-148, 1998.

[22] G. D. J. Smit, A. J. Scholten, N. Serra, R. M. T. Pijper, R. van Langevelde, A. Mercha, G. Gildenblat, and D. B. M. Klaassen, "PSP-based scalable compact FinFET model," *Proc. Workshop Compact Model. , NSTI-Nanotech*, pp. 520-525, 2007.

[23] K. Suzuki, Y. Tosaka, and T. Sugii, "Analytical Threshold Voltage Model for Short Channel n^+-p^+ Double-Gate SOI MOSFET's," *IEEE Trans. Electron Devices*, vol. 43, pp. 732-738, May1996.

[24] K. Suzuki and T. Sugii, "Analytical models for n+ - p+ double-gate SOI MOSFET's," *IEEE Trans. Electron Devices*, vol. 42, no. 11, pp. 1940-1948, 1995.

[25] K. Suzuki, T. Tanaka, Y. Tosaka , H. Horie , and Y. Arimoto, "Scaling theory for double-gate SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 40, pp. 2326-2329, Dec.1993.

[26] R.-H. Yan, A. Ourmazd, and K. F. Lee, "Scaling the SOI MOSFET: From Bulk to SOI to Bulk," *IEEE Trans. Electron Devices*, vol. 39, pp. 1704-1710, July1992.

[27] D. J. Frank, Y. Taur, and H. S. P. Wong, "Generalized scale length for two-dimensional effects in MOSFETs," *IEEE Electron Device Lett.*, vol. 19, no. 10, pp. 385-387, 1998.

[28] A. Tsormpatzoglou, C. A. Dimitriadis, R. Clerc, Q. Rafhay, G. Pananakakis , and G. Ghibaudo , "Semi-Analytical Modeling of Short-Channel Effects in Si and Ge

Symmetrical Double-Gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, pp. 1943-1952, Aug.2007.

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

[30] G. Pei, J. Kedzierski, P. Oldiges, M. Ieong, and E. C. C. Kan, "FinFET design considerations based on 3-D simulation and analytical modeling," *IEEE Trans. Electron Devices*, vol. 49, no. 8, pp. 1411-1419, 2002.

[31] J. C. S. Woo, K. W. Terrill, and P. K. Vasudev, "Two-dimensional analytic modeling of very thin SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 37, no. 9, pp. 1999-2006, 1990.

[32] H. A. E. Hamid, J. R. Guitart, and B. Iñíguez, "Two-Dimensional Analytical Threshold Voltage and Subthreshold Swing Models of Undoped Symmetric Double-Gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, pp. 1402-1408, June2007.

[33] Q. Chen, E. M. Harrell, and J. D. Meindl, "A Physical Short-Channel Threshold Voltage Model for Undoped Symmetric Double-Gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 50, pp. 1631-1637, July2003.

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

[35] D. Munteanu, J. L. Autran, X. Loussier, S. Harrison, R. Cerutti, and T. Skotnicki, "Quantum short-channel compact modeling of drain-current in double-gate MOSFET," *Proc. ESSDERC Conf.*, pp. 137-140, 2005.

[36] G. Pei, V. Narayanan, Z. Liu, and E. C. Kan, "3D analytical subthreshold and quantum mechanical analyses of double-gate MOSFET," *IEDM Tech. Dig.*, pp. 5-3, 2001.

[37] G. Lixin and J. G. Fossum, "Analytical modeling of quantization and volume inversion in thin Si-film DG MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, no. 2, pp. 287-294, 2002.

[38] J. G. Fossum, L. Ge, M.-H. Chiang, V. P. Trivedi, M. M. Chowdhury, L. Mathew, G. O. Workman, and B.-Y. Nguyen, "A process/physics-based compact model for nonclassical CMOS device and circuit design," *Solid State Electron.*, vol. 48, no. 6, pp. 919-926, June2004.

[39] K. Kim, J. G. Fossum, and C. T. Chuang, "Process/physics-based threshold voltage model for nano-scaled double-gate devices," *International Journal of Electronics*, vol. 91, no. 3, pp. 139-148, 2004.

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

[41] P. M. Solomon, K. W. Guarini, Y. Zhang, K. Chan, E. C. Jones, G. M. Cohen, A. Krasnoperova, M. Ronay, O. Dokumaci, and H. J. Hovel, "Two gates are better than one," *IEEE Circuits and Devices Magazine*, vol. 19, no. 1, pp. 48-62, 2003.

[42] J. G. Fossum, J. W. Yang, and V. P. Trivedi, "Suppression of corner effects in triple-gate MOSFETs," *IEEE Electron Device Lett.*, vol. 24, no. 12, pp. 745-747, 2003.

[43] M. Stadele, R. J. Luyken, M. Roosz, M. Specht, W. Rosner, L. Dreeskornfeld, J. Hartwich, F. Hofmann, J. Kretz, and E. Landgraf, "A comprehensive study of corner effects in tri-gate transistors," *Proc. ESSDERC Conf.*, pp. 165-168, 2004.

[44] W. Xiong, J. W. Park, and J. P. Colinge, "Corner effect in multiple-gate SOI MOSFETs," *Proc. IEEE Int. SOI Conf.*, pp. 111-113, 2003.

[45] W. Liu, *MOSFET Models for Spice Simulation, Including BSIM3v3 and BSIM4* John Wiley & Sons, Inc. New York, 2001.

[46] M. Miura-Mattausch, H. Ueno, M. Tanaka, H. J. Mattausch, S. Kumashiro, T. Yamaguchi, K. Yamashita, and N. Nakayama, "HiSIM: a MOSFET model for circuit simulation connecting circuit performance with technology," *IEDM Tech. Dig.*, pp. 109-112, 2002.

[47] M. Bucher, C. Lallement, C. Enz, F. Thqodoloz, and F. Krummenacher, "The EPFL-EKV MOSFET Model Equations for Simulation," *EPFL Lausanne Switzerland, Technical Report, July*, 1998.

[48] S. M. Sze, "Physics of semiconductor devices," *A Wiley-Interscience Publication, New York: Wiley, 1981, 2nd ed.*, 1981.

[49] K. Hess, *Advanced theory of semiconductor devices* IEEE Press New York, 2000.

[50] C. Y. Chang and S. M. Sze, *ULSI Devices* Wiley, New York, 2000.

[51] *Synopsys Sentaurus Device Manual, version Y-2006.06,* Synopsys, Inc., 2006.

[52] S. Y. Oh, D. E. Ward, and R. W. Dutton, "Transient analysis of MOS transistors," *IEEE Trans. Electron Devices*, vol. 27, no. 8, pp. 1571-1578, 1980.

[53] Y. Tsividis, *Operation and modeling of the MOS transistor*, 2 ed, WCB McGraw-Hill, 1999.

[54] G. Gildenblat, H. Wang, T. L. Chen, X. Gu, and X. Cai, "SP: an advanced surface-potential-based compact MOSFET model," *IEEE J. Solid-State Circuits*, vol. 39, no. 9, pp. 1394-1406, Sept.2004.

[55] G. Gildenblat, X. Li, H. Wang, W. Wu, R. van Langevelde, A. J. Scholten, G. D. J. Smit, and D. B. M. Klaassen, "Introduction to PSP MOSFET model," *Proc. Workshop Compact Model. , NSTI-Nanotech*, pp. 19-24, 2005.

[56] R. van Langevelde, A. J. Scholten, and D. B. M. Klaassen, "Physical Background of MOS Model 11, Level 1101," Nat.Lab.Unclassified Report 2003/00239, Apr.2003. Available: http://www.semiconductors.philips.com/Philips_Models/.

[57] A. Ortiz-Conde, F. J. Garcia-Sanchez, J. Muci, S. Malobabic, and J. J. Liou, "A Review of Core Compact Models for Undoped Double-Gate SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, no. 1, pp. 131-140, 2007.

[58] M. V. Dunga, C. H. Lin, D. D. Lu, W. Xiong, C. R. Cleavelin, P. Patruno, J. R. Hwang, F. L. Yang, A. M. Niknejad, and C. Hu, "BSIM-MG: A Versatile Multi-

Gate FET Model for Mixed-Signal Design," *IEEE Symp. VLSI Tech.*, pp. 60-61, 2007.

[59] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Advances in Computational Mathematics*, vol. 5, pp. 329-359, 1996.

[60] S. Malobabic, A. Ortiz-Conde, and F. J. G. Sanchez, "Modeling the undoped-body symmetric dual-gate MOSFET," *Proc. IEEE Int. Conf. Devices, Circuits and Systems*, vol. 1 2004.

[61] C. G. Sodini, P. K. Ko, and J. L. Moll, "The effect of high fields on MOS device and circuit performance," *IEEE Trans. Electron Devices*, vol. 31, no. 10, pp. 1386-1393, 1984.

[62] L. Chang, Y. Choi, D. Ha, P. Ranade, S. Xiong, J. Bokor, C. Hu, and T. J. King, "Extremely scaled silicon nano-CMOS devices," *Proc. IEEE*, vol. 91, no. 11, pp. 1860-1873, 2003.

[63] C. Canali, G. Majni, R. Mindler, and G. Ottaviani, "Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature," *IEEE Trans. Electron Devices*, vol. 22, pp. 1045-1047, 1975.

[64] J. Watts, C. McAndrew, C. Enz, C. Galup-Montoro, G. Gildenblat, C. Hu, R. van Langevelde, M. Miura-Mattausch, R. Rios, and C. T. Sah, "Advanced Compact Models for MOSFETs," *Proc. Workshop Compact Model. , NSTI-Nanotech*, 2005.

[65] H. Morris, E. Cumberbatch, H. Abebe, and V. Tyree, "Compact Models for the IV Characteristics of Double Gate and Surround Gate MOSFETs," *Proc. IEEE UGIM Symp.*, pp. 119-123, 2006.

[66] B. Yu, H. Lu, M. Liu, and Y. Taur, "Explicit Continuous Models for Double-Gate and Surrounding-Gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, no. 10, pp. 2715-2722, 2007.

[67] "Scilab 4.x," Scilab Consortium, INRIA, 2006. Available: http://www.scilab.org.

[68] K. F. Riley, M. P. Hobson, and S. J. Bence, *Mathematical Methods for Physics and Engineering* Cambridge University Press, 1998.

[69] J. V. Uspensky, *Theory of Equations* McGraw-Hill, New York, 1948.

[70] G. J. Coram, "How to (and how not to) write a compact model in Verilog-A," *Proc. Behavioral Modeling and Simulation Conf.*, pp. 97-106, 2004.

[71] P. K. Ko, R. S. Muller, and C. Hu, "A unified model for hot-electron currents in MOSFET's," *IEDM Tech. Dig.*, vol. 27 1981.

[72] *R. Thakker, Dept. of Electrical Engg., IIT Bombay, private communication* 2007.

[73] G. D. J. Smit, A. J. Scholten, D. B. M. Klaassen, R. van Langevelde, G. Gildenblat, X. Li, and W. Wu, "PSP 102.2," Arizona State University and NXP Semiconductors Research, Oct.2007. Available: http://pspmodel.asu.edu/downloads/psp1022_summary.pdf.

[74] S. Reggiani, M. Valdinoci, L. Colalongo, and G. Baccarani, "A Unified Analytical Model for Bulk and Surface Mobility in Si n-and p-Channel MOSFET's," *Proc. ESSDERC Conf.*, vol. 1, pp. 240-243, Sept.1999.

[75] A. Ringwald and F. Schrempp, "QCDINS 2.0 - A Monte Carlo generator for instanton-induced processes in deep-inelastic scattering," *Computer Physics Communications*, vol. 132, no. 3, pp. 267-305, 2000.

[76] R. Johnson, "Computing LambertW," West Hills Institute of Mathematics, West Hills, CA, USA, 2007. Available: http://www.whim.org/nebula/math/lambertw.html.

[77] E. Kreyszig, *Advanced Engineering Mathematics*, 5th ed, John Wiley & Sons, 1983.

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

[79] T. N. Nguyen, "Small-Geometry MOS transistors: Physics and modeling of surface- and buried-channel MOSFETs." PhD Dissertation, Stanford Univ., 1984.

[80] S. H. Oh, D. Monroe, and J. M. Hergenrother, "Analytic description of short-channel effects in fully-depleted double-gate and cylindrical, surrounding-gate MOSFETs," *IEEE Electron Device Lett.*, vol. 21, no. 9, pp. 445-447, 2000.

[81] B. Mazhari and D. E. Ioannou, "Surface potential at threshold in thin-film SOI MOSFET's," *IEEE Trans. Electron Devices*, vol. 40, no. 6, pp. 1129-1133, 1993.

[82] J. B. McKitterick and A. L. Caviglia, "An analytic model for thin SOI transistors," *IEEE Trans. Electron Devices*, vol. 36, no. 6, pp. 1133-1138, 1989.

[83] S. Cristoloveanu and S. S. Li, *Electrical Characterization of SOI Materials and Devices* Kluwer Academic Publishers, 1995.

[84] B. Iniguez, T. A. Fjeldly, A. Lazaro, F. Danneville, and M. J. Deen, "Compact-Modeling Solutions For Nanoscale Double-Gate and Gate-All-Around MOSFETs," *IEEE Trans. Electron Devices*, vol. 53, no. 9, pp. 2128-2142, 2006.

APPENDIX – A: CLOSED-FORM CALCULATION OF THE INVERSION CHARGE AREAL DENSITY

The complete closed-form procedure to get an accurate approximation for β_2 (which is proportional to Q_i, per (48) and (58)) is summarized below. The details are described in chapter 6.

$$\beta = \sqrt{\frac{qn_i}{2\varepsilon\phi_t}} \qquad b = \frac{2\varepsilon}{W_{fin}C_{ox}} \qquad u = \frac{\beta W_{fin}}{2} e^{(V_{GS}-\Delta\phi-\phi_{fn})/2\phi_t}$$

$$\beta_{2\_sm} = u^2 \qquad \beta_{2\_md} \cong \frac{0.04 + 0.665\left[1 + 0.0195\cdot\ln(1+bu)\right]\cdot\ln(1+bu)}{b}$$

$$\beta_{2\_sm\_md} = \frac{\beta_{2\_sm}\cdot\beta_{2\_md}}{\beta_{2\_sm}+\beta_{2\_md}}$$

$$n_1 = 1 \qquad x = bu\left[1+\left(\frac{10}{bu}\right)^{n_1}\right]^{(1/n_1)} \qquad \beta_{2\_lrg} = \frac{\left\{\ln(x)-\ln\left[\ln(x)\right]\right\}^2 + \ln(x)}{b\cdot\left\{1+\ln(x)-\ln\left[\ln(x)\right]\right\}}$$

$$n_2 = 8 \qquad \beta_{2x} = \beta_{2\_lrg}\left[1+\left(\frac{\beta_{2\_lrg}}{\beta_{2\_sm\_md}}\right)^{n_2}\right]^{(-1/n_2)}$$

$$t_1 = u^2 e^{-2b\beta_{2x}} \qquad \beta_{2y} = \beta_{2x} - \frac{{\beta_{2x}}^2 + \beta_{2x} - t_1}{2\beta_{2x} + 1 + 2bt_1}$$

$$t_2 = u^2 e^{-2b\beta_{2y}} \qquad \beta_{2z} = \beta_{2y} - \frac{{\beta_{2y}}^2 + \beta_{2y} - t_2}{2\beta_{2y} + 1 + 2bt_2}$$

$$t_3 = u^2 e^{-2b\beta_{2z}} \qquad \beta_2 = \beta_{2z} - \frac{{\beta_{2z}}^2 + \beta_{2z} - t_3}{2\beta_{2z} + 1 + 2bt_3}$$

APPENDIX – B: USAGE OF THE PARAMETER EXTRACTION TOOL

We briefly describe here the usage of the parameter extraction tool developed at IIT Bombay by Rajesh Thakker et al. [72].

The first step is to implement the model code whose parameters need to be extracted in the C language in one or more C source files (called *model source files*) using a pre-specified interface. For the sake of brevity, the details of this interface are not explained here and can be obtained from [72]. The interface is designed on a separation-of-competence basis; in other words, it is designed to minimize the overhead on the model-implementer thereby allowing him to focus on his area of competence – viz. model implementation – when integrating with the parameter extractor. Other than a starting interface-related code in the entry-point C source file, the rest of the code in the *model source files* is specific to the model, and the parameter extraction program does not need to have visibility to it. The *model source files* are then compiled and linked with the main optimizer engine object file, resulting in a final executable. We refer to this executable as the *extractor program.*

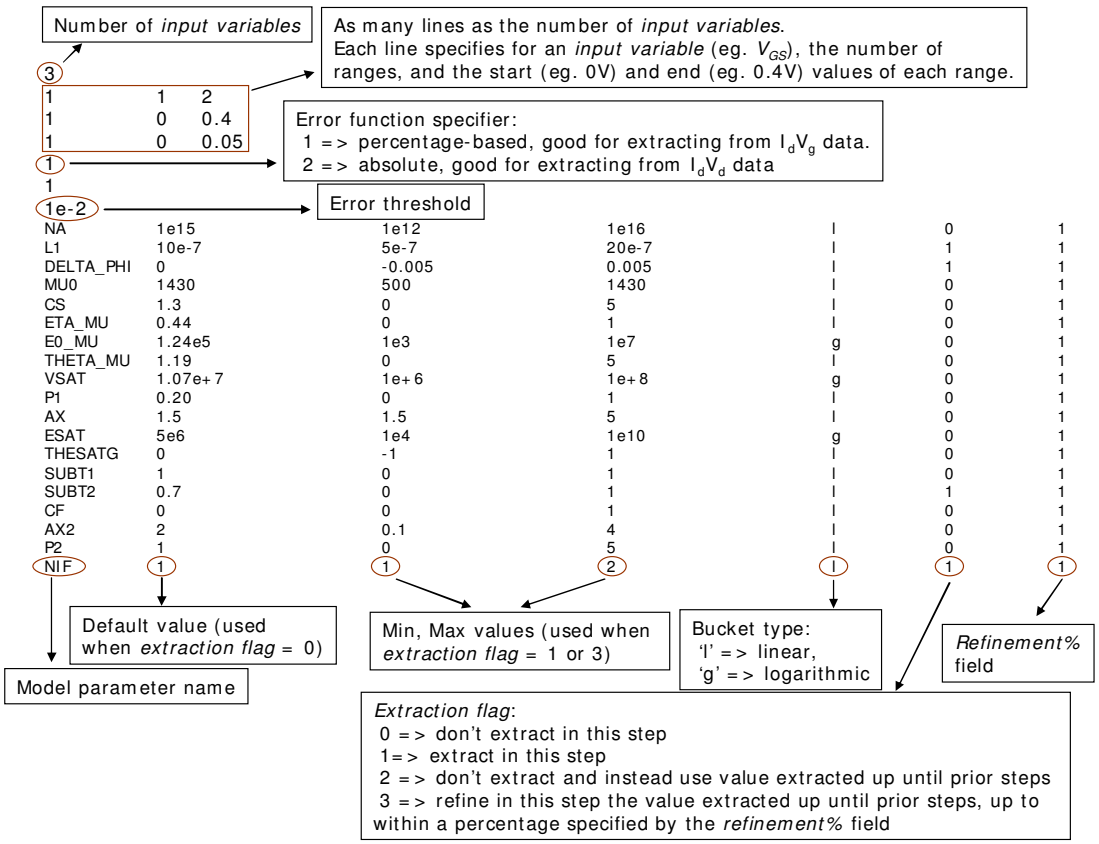
The *extractor program* takes two file names as inputs. The first one is referred to as the *algorithm-controller* file, and the second one is referred to as the *file-information* file.

The *algorithm-controller* file has many parameters that govern the internal working of the *extractor program*, a vast majority of which don't need to be modified during the various runs of the *extractor program* (neither during the evolution of the model, nor during various runs for one version of the model). The two parameters that need to be modified most often are *number_of_para* and *opt_steps*. The former needs to be set to the total number of model parameters that are to be extracted. The latter needs to be set to the total number of parameter extraction steps within one run of the *extractor program*.

The *file-information* file has as many lines as there are steps in the extraction process. Thus for an extraction strategy with four steps (as in the latest version of our model), there would be four lines in the *file-information* file. Each line contains information specific to an extraction step, and lists the name of the *parameter driver* file, the *target data* file, the *output*

file, and the *optimized model data* file. The *parameter driver* file is a step-specific input file and specifies information about the various model parameters (their valid range, whether to extract them in that step, etc), and its format is explained below in more detail. The *target data* file is a step-specific input file and has the target data values to which the model predicted values need to be matched with, by suitable optimization/extraction of the model parameters. Each line in it has the *input variable* values (eg. gate and drain biases), the target data value (eg. drain current) and a weighting factor, all separated by white space. The *output* file is a step-specific output log file, and lists the final extracted parameters at the end of that step. The *optimized model data* file is a step-specific output file in the same format as the *target data* file and contains the model predicted values when calculated using the final extracted parameter values in that step.

The step-specific *parameter driver* file has the following format (contents shown actually correspond to that used in step 1 of our extraction strategy for the final version of our model. In our implementation, the values of the **WFIN** and **TOX** parameters were hard-coded within the *model source files*):



LIST OF PUBLICATIONS

Accepted:

1. V. Hariharan, J. Vasi and V. R. Rao, "Drain Current Model for Undoped Symmetric Double-Gate FETs using a Velocity Saturation Model with Exponent *n=2*", *Proc. ISDRS*, Dec 2007, pp. 1-2.
2. V. Hariharan, R. Thakker, M. B. Patil, J. Vasi and V. R. Rao, "Closed Form Current and Conductance Model for Symmetric Double-Gate MOSFETs using Field-dependent Mobility", *Proc. Workshop Compact Model., NSTI-Nanotech*, June 2008, pp. 857-860.
3. V. Hariharan, J. Vasi and V. R. Rao, "Drain Current Model Including Velocity Saturation for Symmetric Double-Gate MOSFETs", *IEEE Trans. Electron Devices*, vol. 55, no. 8, pp. 2173-2180, 2008.
4. V. Hariharan, J. Vasi and V. R. Rao, "A CAD-Compatible Closed-form Approximation for the Inversion Charge Areal Density in Double-Gate MOSFETs", *Elsevier Solid-State Electronics* (Accepted).
5. V. Hariharan, J. Vasi and V. R. Rao, "An Improvement to the Numerical Robustness of the Surface Potential Approximation for Double-Gate MOSFETs", *IEEE Trans. Electron Devices (Brief)* (Accepted, to appear in Mar 2009 issue).

Submitted:

1. V. Hariharan, R. Thakker, K. Singh, A. B. Sachid, M. B. Patil, J. Vasi and V. R. Rao, "Drain Current Model for Nanoscale Double-Gate MOSFETs", submitted to *Elsevier Solid-State Electronics*.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my guides Prof. V. Ramgopal Rao and Prof. J. Vasi, for their invaluable help and guidance during the course of my work. I am very grateful to them for encouraging me throughout the duration of my PhD. I can never forget the thoroughness with which Prof. Rao would review my manuscripts, regardless of which part of the world he may be at that time. I am thankful for the time and attention that he devoted to my work inspite of having numerous other obligations. His attitude towards work has been a source of inspiration for me. I have also learned a lot from him by means of other activities that he would involve me in, such as reviewing manuscripts submitted for journals and conferences of which he was the editor, teaching assistantship duties, being involved in organizing conferences held on campus, etc.

Nor can I forget the individual attention that Prof. Vasi gave to my work inspite of occupying a busy office at the institute. His comments on my efforts at writing manuscripts, especially in the early stage when I was relatively new to such activities, helped me immensely in writing future manuscripts and any technical documentation in general. Indeed, he was instrumental in proof-reading my thesis in great detail and pointing out the errors. I am also thankful to him and Prof. S. V. Kulkarni for arranging my tests and interviews when I was applying for the PhD program at IIT Bombay. It was Prof. Vasi's courses on device physics during my undergraduate years at IIT Bombay that drew me back to my alma mater for my PhD. I will be ever indebted to him for this.

I would also like to thank my RPC members Prof. M. B. Patil and Prof. K. G. Anil. Besides their effort in reviewing my annual progress seminars, they have also been available for helpful discussions involving circuit simulations and realistic device engineering concepts. I am also thankful to Prof. D. K. Sharma for his insightful inputs in the numerous *Technology-Circuit Interactions* meetings. I would also like to extend my gratitude to IITB professors who have taught me various courses during my PhD, such as Prof. Souvik Mahapatra, Prof. A. N. Chandorkar and Prof. Rakesh Lal.

I would also like to thank my friends C. R. Manoj, Rajesh Thakker, Angada Sachid, Sudhakar Mande, Ramesh Navan, S. P. Tiwari, Brajesh Pande and many other friends, for the valuable

discussions that I have had with them. I especially want to thank Rajesh Thakker for helping me use his parameter extraction tool. I would also like to mention our lab system administrator Yogesh Save for his effort in keeping the systems up and running. I will never forget the days I spent with Debabrata Maji during my internship in Taiwan. Special mention must also be made about my seniors in the lab Neeraj Jha, Bharath Kumar, K. Narsimhulu, D. Vinay Kumar, Deleep Nair and B. Anand.

I take this opportunity to thank the Department of Electrical Engineering, IIT Bombay for providing me various facilities for my research work, besides giving me the travel grants to attend conferences. I would like to thank Ms. Tanvi Shelatkar, Ms. Jyoti, Ms. Madhu, Ms. Vaishali, Ms. Shashi Anchan, and other EE Dept., IRCC and administrative staff for their numerous and timely help in many official matters.

I want to thank all my family members and well wishers for their constant encouragement and loving support for all these years, without which I could not have completed this work.

I would also like to thank Dr Ke-Wei Su, Dr Sally Liu, Dr J. R. Shih, Dr Kenneth Woo and Dr Alan Liang of TSMC for letting me spend 3 months in their Spice Modeling Group on a summer internship project. I am very grateful to Dr. Thao N. Nguyen for taking the effort to send me a copy of his PhD thesis by postal mail when I contacted him about my inability to access it from Stanford University's library. I also wish to thank Dr. Ronald van Langevelde for the time he took to answer my queries via email.

Finally, I want to thank Puneet Sarna for preparing the Standard Thesis Template in Microsoft Word.

Venkatnarayan Hariharan

July 2008

Indian Institute of Technology Bombay