MODELING, DESIGN, AND PERFORMANCE OF NANOSCALE DOUBLE-GATE CMOS

By

SIDDHARTH CHOUKSEY

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2009

To my family, friends, and the almighty

ACKNOWLEDGEMENTS

Being where I am today is literally a dream come true. I believe success happens when talent meets an opportunity. Numerous opportunities came my way, many times indirectly and disguised, that made all this possible. I do not know what made those opportunities come my way, if not, the supreme force guiding our lives - God. Therefore, first of all I would like to thank God for making this Ph.D. a part of his plan. I do not think its within my capability to fully express my sense of awe and gratitude toward my advisor, Dr. Jerry G. Fossum. Working with him has been a tremendous learning experience from me. He taught me not only through the classes, feedbacks, and discussions, but many times when he was not even intending! I have learned a lot just by watching him work; his enthusiasm for work, patience, and analytical skills are exemplary. I thank Dr. Fossum for his support and guidance throughout my work. I would like to thank Dr. Scott Thompson, Dr. Jing Guo, and Dr. Selman Hershfield for agreeing to be on my committee as well as for helpful discussions during my research work. I would like to acknowledge Freescale Semiconductor for their support, and thank Intel Corp. for giving me opportunity to intern with them twice. I would like to thank my mother for her undiminished confidence in me, and for all her prayers. I will not be surprised if this Ph.D. means more to her than what it means to me. I would like to thank my father for his support, and my brother, sister-in-law, sister, brother-in-law, my niece and nephew for their encouragement. I was fortunate to have Murshed Chowdhury, Shishir Agrawal, Zhichao Lu, Zhenming Zhou, Weimin Zhang, Sueng-Hwan Kim, and Dabraj Sarkar as my colleagues. I would like to thank them for all the help they extended to me throughout my work. I would like to thank my numerous friends, in and out of Gainesville, for their companionship through these years.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor
CMOS	Complementary MOS
SOI	Silicon-on-Insulator
FD	Fully Depleted
PD	Partially Depleted
UTB	Ultra-Thin Body
UFDG	University of Florida Double-Gate
SCE	Short-Channel Effect
DIBL	Drain-Induced Barrier Lowering
DICE	Drain-Induced Charge Enhancement
QM	Quantum-Mechanical
SDE	Source/Drain Extension
SG	Single Gate
S	Subthreshold Slope
RF	Radio Frequency
TBOX	Thin Back Oxide
GP	Ground Plane
LP	Low Power
HP	High Performance

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

MODELING, DESIGN, AND PERFORMANCE OF NANOSCALE DOUBLE-GATE CMOS

By

Siddharth Chouksey

December 2009

Chair: Jerry G. Fossum
Major: Electrical and Computer Engineering

This dissertation seeks to understand the unique physics of, to explore non-conventional ways of designing, and to gain insights on the performance of nanoscale double-gate (DG) MOSFETs, particularly the quasi-planar FinFET structure. Our work includes modeling of drain-induced charge enhancement, suggesting a novel way of adjusting the threshold voltage of nanoscale DG MOSFETs via limited source/drain dopants in the channel, comparing the analog/RF performance of DG FinFETs and bulk-silicon MOSFETs, and studying and designing ultra-thin-BOX FD/SOI MOSFETs with comparisons to DG FinFETs.

Drain-induced charge enhancement (DICE) is a short-channel effect which is unique to nanoscale DG MOSFETs with undoped bodies because of their significantly high carrier mobility. We model this effect, and study its effect on the current, charge, capacitance, and transcapacitance of DG MOSFETs. We find that DICE is a beneficial effect because it increases current without significantly affecting gate capacitance.

Adjusting the threshold voltage of DG MOSFETs with undoped bodies for low-power and high-performance applications is a challenging task. We propose a design approach in which limited densities of source/drain dopants in the channel can be used to effect an adjustment of

threshold voltage in DG MOSFETs, while maintaining low sensitivities to random-doping fluctuations.

Most of the current literature on the analog/RF performance of DG MOSFETs is based on experimental results, with little physics-based explanation of the results. We give physical insights on the design and performance of DG MOSFETs for analog/RF applications, and compare them with bulk-silicon MOSFETs. We find that like for digital applications, DG MOSFETs have superior analog performance than that of conventional planar bulk-silicon MOSFETs.

Recently, there has emerged a considerable interest in planar FD/SOI MOSFETs with ultra-thin BOX. We give our physical insights, based on device simulations, on the design and performance of ultra-thin-BOX FD/SOI MOSFETs, and check their scalability toward the end of the CMOS roadmap compared with DG FinFETs.

CHAPTER 1
INTRODUCTION

Scaling conventional MOSFETs, i.e., bulk-silicon and partially depleted (PD)/SOI MOSFETs, to gate lengths (L_g) < ~40nm has become increasingly challenging because the channel-doping density required has become very high. Conventional MOSFETs rely on channel-doping density to control threshold voltage (V_t) and short-channel effects (SCEs). As these devices are scaled, the latter are controlled by reducing the depletion width by increasing channel-doping density. For L_g < ~40nm, the required channel-doping density has become so high that the variation in V_t due to random variations in the channel-doping density has become a serious issue. At such short L_g, reliable control of channel doping density is virtually impossible. Under this scenario, double-gate (DG) MOSFETs, e.g. FinFETs, have emerged as a most promising candidate to replace the bulk-Si MOSFET [1]. The primary advantage of the FinFET is the excellent control of SCEs [2] without relying on channel doping, which makes it potentially scalable to the end of the SIA ITRS roadmap [1]. Since FinFETs rely on undoped ultra-thin bodies (UTB) to control SCEs, random variations in threshold voltage (V_t) and other device characteristics due to process variations can be greatly reduced [3]. In Fig. 1.1 we show the basic structure of a FinFET. The gate is wrapped over the thin vertical fin, forming two sidewall gates. The top of the fin could be gated, forming a triple-gate device [4], but we focus on the DG structure which is more pragmatic [2].

DG-FinFET technology has not yet received complete acceptance by the integrated-circuits manufacturing companies because of some of the challenges associated with the DG-FinFET technology like higher cost of SOI wafers, control of the nanoscale fin, and lack of reliable ways of engineering the source/drain doping profile. Higher cost of SOI wafers has led to some interest in the bulk-Si FinFETs [5], [6]. However, due to significantly high, controlled

substrate-doping density required to suppress source/drain-leakage current, and need to precisely match the depth of source and drain regions to the substrate doping, the viability of bulk-Si FinFETs is doubtful [7]. In order to scale DG FinFETs to the end of the roadmap, reliable ways of engineering the source/drain profile will have to be developed. This task is particularly challenging because the UTB thickness tends to be about 5nm near the scaling limit of L_g. The diffusion of source/drain dopants through such thin bodies is not well understood. As we will show through comprehensive studies in this dissertation, DG FinFETs tend to have significantly better performance for both digital and analog/RF applications. But, in order for DG FinFETs to replace bulk-Si MOSFETs, the technological challenges like those mentioned here will have to be overcome.

The physics of DG MOSFETs, with coupled gates and UTB, is significantly different from bulk-Si MOSFETs. Also, designing them for various applications is significantly different from designing bulk-Si MOSFETs. Designing circuits with independently controlled bias on the two gates of DG MOSFETs [8], using the space between two fins on the Si substrate to enhance the drive current (ITFET [9]), and engineering the S/D extension region to realize a bias-dependent effective channel length (L_{eff}) [2] are some of the unique possibilities that exist with DG MOSFETs. Therefore, needless to say, in order to realize the full potential of DG MOSFETs, their physics has to be better understood, and non-conventional ways of designing them will have to be explored. In this dissertation we discuss our contributions towards these goals. Since DG FinFETs are most prospective among all the contemporary DG MOSFETs, we use them as the representative DG MOSFET structure in our studies, although most of the discussion in this dissertation is generically applicable to all kinds of DG MOSFET structures.

In FinFETs, the undoped UTB between the two (connected) gates results in unusually low transverse electric field and quite high carrier mobility (μ_{eff}) [10], [11]. This produces significant

saturation-region effects [12] that are not prevalent in the conventional devices, e.g., carrier-velocity overshoot [13] and near-ballistic transport [11]. “Drain-induced charge enhancement” (DICE), which is a strong-inversion counterpart to the drain-induced barrier lowering (DIBL) in weak inversion, is such an effect that heretofore has not been generally acknowledged. In Chapter 2, the significance of DICE in DG MOSFETs is revealed via analytical modeling, implemented in our physics-based compact model UFDG [14], [15], [16] and supported by numerical device simulations.

Because of the undoped body, there is no pragmatic way of tuning V_t of DG MOSFETs for low-power (LP) and high-performance (HP) applications. However, the undoped body enables the design of DG MOSFETs with G-S/D underlap. It has been shown that incorporating G-S/D underlap in the DG MOSFET design yields L_{eff} that decreases with increasing gate bias (V_{GS}), and that can hence be used to effect a design tradeoff between SCE control, or off-state current (I_{off}) for LP applications, and S/D series resistance ($R_{S/D}$), or on-state current (I_{on}) for HP applications [17]. In Chapter 3, we show how this I_{off}-I_{on} design tradeoff can be extended by allowing limited densities of S/D dopants to diffuse into the FinFET channel for direct V_t adjustment in HP versus LP applications.

While CMOS scaling is predominantly driven by digital applications, analog performance can benefit from scaling [1]. For bulk-Si MOSFETs, for example, the transconductance (g_m) increases with scaling. And, scaling is the primary way of improving the cut-off frequency [$f_T = g_m/(2\pi C_G)$ where C_G is the total gate capacitance] of analog devices. However, these benefits come with a cost of enhanced SCEs that undermine the output conductance (g_{DS}) and voltage gain ($A_{vo} = g_m/g_{DS}$) of the bulk-Si MOSFETs. Therefore, there is considerable interest in exploring the possibility of using FinFETs, which have better L_g scalability than bulk-Si

MOSFETs, for RF applications. In Chapter 4 we use our physical insights, supplemented with published numerical simulation and experimental results, to check the analog figures-of-merit (FOMs) of DG FinFETs, and compare them with those of bulk-Si MOSFETs.

Recently, there has been a noticeable interest in the planar ultra-thin-box fully depleted (FD)/SOI MOSFET [18] as an alternative to the conventional bulk-Si MOSFET, and in lieu of the quasi-planar FinFET. The motivation for this kind of MOSFET architecture is the better control of SOI UTB thickness as compared to that of the fin-UTB thickness, in addition to the close similarities of FD/SOI and PD/SOI processing. In Chapter 5, we do a thorough, simulation-based evalutation of ultra-thin-box FD/SOI CMOS, projecting LP and HP scaling limits and noting needed process complexities, with comparisons to FinFET CMOS.

In Chapter 6 we conclude this dissertation with a summary and suggestions for future work.

In Appendix A we describe the UFDG model refinements done for more reliable predictions of nanoscale DG MOSFET characteristics, as discussed herein.

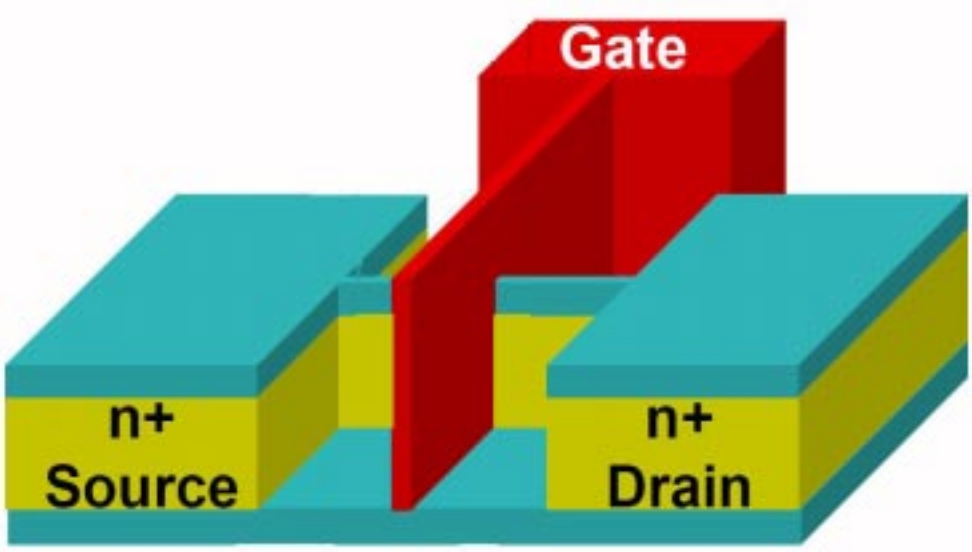


Figure 1-1. Structure of a FinFET with gate wrapped over the vertical fin, forming two sidewall gates.

CHAPTER 2
DICE: A BENEFICIAL SHORT-CHANNEL EFFECT IN DOUBLE-GATE MOSFETs

2-1 Introduction

Drain-induced barrier lowering (DIBL) [19] is a well-publicized SCE in scaled MOSFETs. In accord with the 2-D Poisson equation in the weakly inverted channel/body, high drain bias (V_{DS}) lowers the potential barrier height at the virtual source, thereby allowing increased carrier diffusion from the source to the channel, causing higher I_{off}, and lowering the (saturation) V_t. Drain-induced charge enhancement (DICE) is the counterpart effect in strong inversion, which was analyzed for classical SOI MOSFETs about 20 years ago [20] but which heretofore has been insignificant. Work based on numerical simulations of bulk-Si MOSFETs published in 1980 [21] revealed the effect, also showing it to be insignificant. However, we show in this chapter, and explain why, this is not the case for nanoscale DG MOSFETs, in which DICE can significantly benefit the drive, or saturation-region current ($I_{DS(sat)}$) and CMOS speed, with little effect on the gate capacitance.

2-2 Modeling

Calibrations of UFDG to fabricated nanoscale FinFETs, with undoped UTBs, have tended to give results like the I_{DS}-V_{GS} characteristics of an L_g = 60nm device [22] shown in Fig. 2.1. Note that the UFDG predictions agree well with the measured data, except for high V_{DS}, i.e., in the saturation region where UFDG underpredicts $I_{DS(sat)}$ (by about 7% for this 60nm FinFET). We have inferred from such results, complemented by numerical device simulations, that DICE can significantly affect the saturation-region current of nanoscale DG MOSFETs.

The saturation region of the nanoscale MOSFET is characterized, as illustrated in Fig. 2.2, by a UTB/channel that can be divided into a “gradual” [19] portion ($0 < y < L_{gch}$) adjacent to the source and a high-electric field (i.e., $|E_y|$) portion ($L_{gch} < y < L_g$) adjacent to the drain in which

carrier velocity is saturated. (Below saturation, for $V_{DS} < V_{DS(sat)}$, $L_{gch} = L_g$.) Actually, when velocity (v) overshoot occurs, $v \cong v_{sat(eff)} > v_{sat}$ in the high-$|E_y|$ region, where $v_{sat(eff)}$ depends on the MOSFET bias ($V_{GS} > V_t$ and $V_{DS} > V_{DS(sat)}$) [13]. DICE is literally an enhancement of the channel inversion-charge density (Q_{ch}), which increases $I_{DS(sat)}$ as we show herein. Furthermore, it is significant in defining L_{gch} and the MOSFET terminal charges, although C_G is not affected much. Without DICE, the classical (2-D Gauss law-based) analysis [19], [20] of the high-$|E_y|$ portion of the channel is erroneous, predicting a too short L_{gch} as we will show.

We model DICE via an approximate solution of the 2-D Poisson equation in the rectangular UTB/gradual channel (see Fig. 2.2) of a generic undoped DG MOSFET. Previously, such an analysis was done [20] for single-gate devices based on the inversion charge-sheet approximation, neglecting bulk inversion. Here we account for two gates and bulk inversion, which is quite significant in DG MOSFETs with undoped UTBs [23], making no assumption about the inversion-charge distribution in the body. Letting $\phi_0(x,y)$ be the potential in the undoped UTB/gradual channel for $V_{DS} = 0$ under a strong-inversion condition ($V_{GS} > V_t$), we note from Poisson's equation that $V_{DS} > 0$ defines $\phi(x,y) = \phi_0(x,y) + \Delta\phi(x,y)$, where the perturbation $\Delta\phi(x,y)$ is related to the V_{DS}-induced change in Q_{ch}:

$$\Delta Q_{ch}(y) = -\varepsilon_{Si}\int_0^{t_{Si}}\left(\frac{\partial^2}{\partial x^2}\Delta\phi(x,y)\right)dx - \varepsilon_{Si}\int_0^{t_{Si}}\left(\frac{\partial^2}{\partial y^2}\Delta\phi(x,y)\right)dx \qquad (2.1)$$

where the integrations are done over the UTB thickness t_{Si}. The first integral (<0 for an nMOSFET) in (2.1) reflects the pinch-off tendency, and, in the classical gradual-channel approximation [19], the second integral is assumed to be negligible. However, for nanoscale DG MOSFETs this assumption can be invalid, and the second integral can be significant DICE ($\Delta Q_{ch}^{DICE} < 0$ for an nMOSFET), as we characterize herein.

Numerical. simulations of DG MOSFETs done with Medici [24] reveal that

$$\frac{\partial^2}{\partial y^2}\Delta\phi(x,y) \cong \eta \tag{2.2}$$

in the gradual-channel portion of the UTB, where η is a spatial constant; that is, the V_{DS}-induced perturbation in the electric field along the channel is nearly linear in y. This approximation obtains because, for high μ_{eff}, the electron velocity is close to being saturated, and hence not varying much in y, and so the added charge density, associated with $\varepsilon_{Si}\partial(\Delta E_y)/\partial y$, is nearly constant (for continuous current). Integrating (2.2) twice along L_{gch}, with boundary conditions $\Delta\phi(x, 0) = 0$ (due to strong inversion) and $\Delta\phi(x, L_{gch}) = V_{DS(eff)}$, yields $\eta \cong 2V_{DS(eff)}/L_{gch}^2$, where $V_{DS(eff)}$ ($\sim V_{DS(sat)}$) is the effective bias at the end of the gradual channel. Using (2.2) in (2.1) then yields

$$\Delta Q_{ch}^{DICE} \cong -\varepsilon_{Si} t_{Si} \frac{2V_{DS(eff)}}{L_{gch}^2}. \tag{2.3}$$

This simple, but physical model shows that DICE manifests as a nearly uniform enhancement of the inversion-charge density everywhere along the gradual channel, and in the high-$|E_y|$ portion as well because of the velocity saturation. Both (2.2) and (2.3) are consistent with the empirical result derived from strong-inversion numerical simulations in [21] of a V_t shift linearly proportional to V_{DS}. Extending the basic DG MOSFET analysis in UFDG [25] accordingly, we use (2.1) and (2.3) to express the total inversion-charge density along the channel as

$$Q_{ch}(y) = Q_{ch0} + C_{oxf}\Delta\phi(0,y) + C_{oxb}\Delta\phi(t_{si},y) + \Delta Q_{ch}^{DICE}, \tag{2.4}$$

where Q_{ch0} is the $V_{DS} = 0$ charge density; $C_{oxf} = \varepsilon_{ox}/t_{oxf}$ and $C_{oxb} = \varepsilon_{ox}/t_{oxb}$ are the front- and back-gate oxide capacitances.

Prior to saturation, i.e., for $V_{DS} < V_{DS(sat)}$ ($=V_{DS(eff)}$ at the onset of saturation), (2.3) and (2.4) still apply, but with $V_{DS(eff)}$ and L_{gch} replaced with V_{DS} and L_g (assumed here to be the

channel length as indicated in Fig. 2.2), respectively. For the UFDG DICE upgrade, affecting channel current and terminal charges, the existing formalism [25] is revised, in accord with (2.4), by simply replacing Q_{ch} with ($Q_{ch} + \Delta Q_{ch}^{DICE}$). (In the model code, because of uncertainty in L_g and t_{Si}, we actually use (*DICE* x ΔQ_{ch}^{DICE}), where *DICE* is a tuning parameter that is typically ≅1.) Note that this revision alters the model characterization of the bias-dependent L_{gch}, which influences the terminal charges (and device capacitances and transcapacitances) as well as the channel current.

The UFDG model for channel current [25] is upgraded for DICE directly, using (2.4) and the altered characterization of L_{gch}. The impact of DICE on the terminal charge modeling (discussed in Appendix A-2) is a bit more involved since ΔQ_{ch}^{DICE} and the new L_{gch} characterization affect the charge partitioning. The basic 2-D analysis [19], [25] of the high-$|E_y|$ region that defines L_{gch} is extended to include the DICE charge. Application of the 2-D Gauss law yields a second-order differential equation for the V_{DS}-induced perturbation of potential:

$$\frac{d^2}{dy^2}\Delta\phi(y) = \frac{\Delta\phi(y) - V_{DS(eff)}}{l_c^2} + \eta\,, \tag{2.5}$$

from which we get

$$\Delta\phi(y) - V_{DS(eff)} + \eta l_c^2 = l_c E_{sat}\sinh\left(\frac{y - L_{gch}}{l_c}\right) + \eta l_c^2\cosh\left(\frac{y - L_{gch}}{l_c}\right) \tag{2.6}$$

where η conveys the DICE effect; $l_c = [\varepsilon_{Si}t_{Si}/(C_{oxf} + C_{oxb})]^{1/2}$ and $E_{sat} = 2v_{sat(eff)}/\mu_{eff}$ is $|E_y(L_{gch})|$ [25]. Evaluating (2.6) at $y = L_g$, where $\Delta\phi = V_{DS}$, yields following expression of L_{gch}.

$$V_{DS} - V_{DS(eff)} + \eta l_c^2 = l_c E_{sat}\sinh\left(\frac{L_g - L_{gch}}{l_c}\right) + \eta l_c^2\cosh\left(\frac{L_g - L_{gch}}{l_c}\right). \tag{2.7}$$

Now, the component of front-gate charge defined by (2.6) is

$$Q_{Gf}^{sat} = W_g C_{oxf} \int_{L_{gch}}^{L_g} [V_{GfS} - \Phi_{GfB} - \phi_0(0, y) - \Delta\phi(y)] dy \tag{2.8}$$

Φ_{GfB} is the gate-body work-function difference and W_g is the device width. The back-gate charge is defined similarly, and for a symmetrical DG device, $Q_{Gb} = Q_{Gf}$. Note that $E_y(L_g)$, given by the derivative of (2.6) evaluated at $y = L_g$, defines, via Gauss's law, a depletion-region component of the drain charge:

$$Q_{D(dep)} = W l_c^2 (C_{oxf} + C_{oxb}) \left[E_{sat} \cosh\left(\frac{L_g - L_{gch}}{l_c} \right) + \eta l_c \sinh\left(\frac{L_g - L_{gch}}{l_c} \right) \right] \tag{2.9}$$

which includes charge linked to the gradual channel as well as the high-$|E_y|$ portion. The upgraded charge model reflects directly the charge neutrality,

$$Q_{Gf} + Q_{Gb} + Q_S + Q_D = 0 \tag{2.10}$$

(with the undoped-body charge $Q_B = 0$), where the source (Q_S) and drain (Q_D) charges include partitioned components of Q_{ch} in (2.4) integrated along the channel.

All the terminal charge components, and the channel current, depend on L_{gch}, implied by (2.6), and on $V_{DS(eff)}$. The latter is obtained from the current analysis of the gradual channel [19], [25], upgraded with DICE, i.e., with Q_{ch} including ΔQ_{ch}^{DICE} as in (2.4):

$$V_{DS(eff)} = \frac{Q_{ch}(0) E_{sat} L_{gch}}{Q_{ch}(0) - (C_{of} + C_{ob}) E_{sat} L_{gch}}. \tag{2.11}$$

UFDG solves (2.11) and the nonlinear expression for L_{gch} from (2.7) iteratively via Newton-Raphson, and then the current and the terminal charges are evaluated. Figure 2.3 shows L_{gch} vs. V_{DS} predicted by UFDG for an 18nm DG nMOSFET, with and without the DICE upgrade. Note that without DICE, L_{gch} is substantively underpredicted for high V_{DS}, which results in erroneous current and terminal-charge predictions.

The modified 2-D analysis of the high-$|E_y|$ region, with ΔQ_{ch}^{DICE}, changes all the terminal charges without DICE [20], [25] due to the perturbations in $V_{DS(eff)}$ and L_{gch}. These changes are only loosely coupled to the gate because $V_{DS(eff)}$ in (2.11) and L_{gch} given by (2.7) are weakly dependent on V_{GS}. This implies that the gate capacitance ($C_G = dQ_G/dV_{GS}$) is nearly independent of DICE. Our analysis shows that a portion of drain-depletion charge is imaged in the gate via an inner fringe field, and that the rest is imaged in the channel as the DICE charge ($\Delta Q_{ch}^{DICE} W_g L_g$). Hence, the DICE charge in the channel is supported mainly by the drain. Indeed, DICE is very significant in defining all the terminal charges of nanoscale DG MOSFETs, and is therefore important in predicting the various device capacitances and transcapacitances, as well as channel current. We demonstrate this significance in the next section.

Our analysis also explains the emerging significance of DICE in nanoscale DG MOSFETs, and why it is negligible in conventional devices. The high mobility in the undoped channel of DG MOSFETs, while lowering $V_{DS(eff)}$ in (2.3) and (2.11), decreases L_{gch} as given by (2.7). The latter effect is predominant in (2.3), making ΔQ_{ch}^{DICE} in nanoscale DG MOSFETs significant. Further, the high μ_{eff} tends to yield significant velocity overshoot ($v_{sat(eff)} > v_{sat}$), which renders DICE even more significant by increasing $V_{DS(sat)}$, and thereby increasing $V_{DS(eff)}$ for a given L_{gch}. Our analysis also shows that L_{gch} scales faster than $V_{DS(eff)}$, and for well-tempered L_g scaling, L_{gch} scales faster than L_g, and hence faster than t_{Si} and t_{ox}. Thus, ΔQ_{ch}^{DICE}, as modeled by (2.3), becomes more significant with scaling.

2-3 Model Corroboration and Predicted DICE Impacts

We include in Fig. 2.1 the UFDG-predicted current for the 60nm pFinFET with our DICE modeling incorporated. Note now the excellent match with the high-V_{DS} measured data. To further verify the DICE model, which we have also noted to be consistent with [21], and to corroborate the significance of DICE in nanoscale DG devices, we simulate a simplified L_g =

18nm symmetrical DG nMOSFET with Medici [24], and compare the DICE predicted by it to that predicted by UFDG. Series resistance was kept low to avoid any discrepancy in the effective gate and drain biases due to possible disagreement in currents. And, since the physical modeling in Medici can differ from that in UFDG, additional simplifications were made. In particular, we fixed the low-$|E_y|$ μ_{eff} to be 300cm^2/V-s (which is comparable to on-state electron mobilities measured in DG FinFETs [10], [11]) in both the simulations, and we turned off the velocity-overshoot models, i.e., we set $v_{sat(eff)} = 10^7$cm/s ($\cong v_{sat}$). Further, we initially let t_{Si} = 12nm (which is not thin enough for L_g = 18nm to adequately suppress the unwanted SCEs [26]) to avoid anomalously high n(x) for thin t_{Si} that Medici predicts. Later we check devices with thinner t_{Si}. Figure 2.4 shows the Medici-predicted variation of inversion-electron density across the UTB (t_{Si} = 12nm) at the virtual source (i.e., where ϕ is minimum along y) for low (50mV) and high (1.0V) drain biases, with V_{GS} = 1.0V. Note the strong bulk inversion (which would be enhanced by quantization [27]). The V_{DS}-induced enhancement in the areal density of inversion charge at the virtual source, i.e., DICE, reflected by Fig. 2.4 is about 20%. Figure 2.5 shows the enhancement in the V_{GS} = 1.0V current due to DICE in the same device as predicted by UFDG. Because of the high μ_{eff}, the high-V_{DS} current is restrained by the ballistic limit [11]; that is, the current (per W_g) is nearly $Q_{ch}(0)$ times the thermal injection velocity at the virtual source. Hence, the enhancement in the current in Fig. 2.5 virtually reflects ΔQ_{ch}^{DICE}, which, indeed, is close to that predicted by Medici in Fig. 2.4 at V_{DS} = 1.0V. To further show the near equality of these simulations, we show in Fig. 2.6 the current predicted by UFDG for the same device as in Fig. 2.5, but with the ballistic-current limit turned off, and compare it with that predicted by MEDICI, which does not account for the ballistic limit. Note that the predicted currents match very well.

In Fig. 2.7 we show UFDG-predicted I_{DS} vs. V_{DS}, at V_{GS} = 1.0V, for the symmetrical 18nm DG nMOSFET (t_{Si} = 12nm, $t_{oxf} = t_{oxb}$ = 1.2nm, midgap gate), but now with the proper series

resistance, and mobility [10], [11], velocity-overshoot [13], and quantization [27] modeling, with and without DICE. The current at V_{DS} = 1.0V is enhanced 24% by DICE (more than the ≅20% in Fig. 2.5 mainly because of the velocity overshoot), which is quite substantial because the current is at the ballistic limit and thus directly reflects ΔQ_{ch}^{DICE}. We stress that the current enhancement due to DICE is smaller for devices in which the current is not ballistically limited (like the pFinFET in Fig. 2.1) because it would be undermined some by the increase in L_{gch} caused by DICE (see Fig. 2.3). Note also in Fig. 2.7 that DICE is much less significant for $V_{DS} < V_{DS(sat)} \cong$ 0.2V.

As we noted previously, the 18nm DG MOSFET with t_{Si} = 12nm is not well-tempered with regard to SCEs. Indeed, as shown in Table 2.1, the UFDG-predicted DIBL for this device is unacceptable at 260mV/V. However, the additional UFDG predictions of DIBL and DICE in Table 2.1, for thinner t_{Si}, support our claim that DICE can be a beneficial short-channel effect, even when unwanted SCEs are adequately suppressed. For t_{Si} = 8nm, DICE enhances the on-state current by 15%, while DIBL is only 110mV/V. The DICE benefit is even more dramatic when a G-S/D underlap is incorporated, as has been suggested for optimally designed nanoscale DG MOSFETs [26]. The underlap increases the weak-inversion effective channel length that defines the SCEs (e.g., DIBL), without significantly affecting the strong-inversion channel length (and hence DICE). The UFDG-predicted values of DIBL in Table 2.1 for the 18nm DG MOSFET with a 2nm underlap reveal this. For the now optimal t_{Si} = 10nm, DICE yields a 19% current enhancement while DIBL is limited to 100mV/V.

We include in Fig. 2.7 the predicted drain conductance ($g_{DS} = dI_{DS}/dV_{DS}$), with and without DICE. Since the current is ballistic ($\propto Q_{ch}(y=0)$ given by (2.4), with V_{DS} dependence only in ΔQ_{ch}^{DICE}), g_{DS} without DICE is zero at high V_{DS}. With DICE, however, it becomes finite, defined directly by ΔQ_{ch}^{DICE} in (2.4), as characterized in (2.3). The V_{DS} dependence of L_{gch} in

(2.3), implied by (2.7) and illustrated in Fig. 2.3, is the main reason for the finite g_{DS} due to DICE. We note that for lower V_{GS}, g_{DS} is about the same as that (with DICE) in Fig. 2.7 since ΔQ_{ch}^{DICE} in (2.3) is only weakly dependent on V_{GS}. However, for moderate inversion (e.g., at typical operating points for low-power RF transistors today), g_{DS} tends to be lower, controlled by channel-length modulation, as DIBL tends to preempt DICE.

As discussed in Sec. 2.2, the gate capacitance is nearly independent of DICE. This is shown by the UFDG-predicted $C_G(V_{GS})$ curves in Fig. 2.8 for the same 18nm DG nMOSFET at high and low V_{DS}, with and without DICE. (The UFDG-predicted $C_G(V_{GS})$ is in good accord with the Medici-predicted characteristic at low V_{DS}. However, at high V_{DS} there is some discrepancy, which we believe is due to non-physical dependences of carrier transport on high $|E_y|$ used in Medici for weak inversion.) So, since DICE gives enhanced current without increased gate capacitance, should it yield faster CMOS? The answer is yes, as shown by the UFDG/Spice3-predicted CMOS inverter-chain delay revealed in Fig. 2.9. With 18nm DG MOSFETs, like that in Figs. 2.7 and 2.8 (with the proper hole mobility [10] assumed for the pMOSFET), the average propagation delay per stage is reduced by about 18% by DICE (less than the 24% I_{on} enhancement in Fig. 2.7 because of the V_{DS} dependence of the DICE benefit, and because of second-order increases in drain capacitance and gate-drain transcapacitance due to DICE).

2-4 Summary

Drain-induced charge enhancement, the strong-inversion counterpart to DIBL, has been modeled analytically, with numerical support, and shown to be a significant short-channel effect in nanoscale DG MOSFETs. Further, it becomes more significant with well-tempered scaling. DICE substantially increases the inversion-charge density, without affecting the gate capacitance significantly since the added charge is supported mainly by the drain. It therefore is beneficial to digital CMOS speed, as we demonstrated via UFDG/Spice3 simulations. However, it increases

drain conductance, and hence could be problematic in analog applications, although it does subside in moderate inversion where today's low-power RF transistors operate. Our analysis further demonstrates that physical compact models for nanoscale DG MOSFETs should account for DICE to ensure valid terminal charge (and capacitance and transcapacitance) as well as current modeling.

Table 2-1. UFDG-predicted DICE and DIBL in an L_g = 18nm nMOSFET with varying UTB thickness, with and without a 2nm gate-source/drain underlap.

t_{Si} (nm)	DICE (%)	DIBL (mV/V)	DIBL (mV/V) w/ underlap
12	24	260	150
10	19	180	100
8	15	110	60

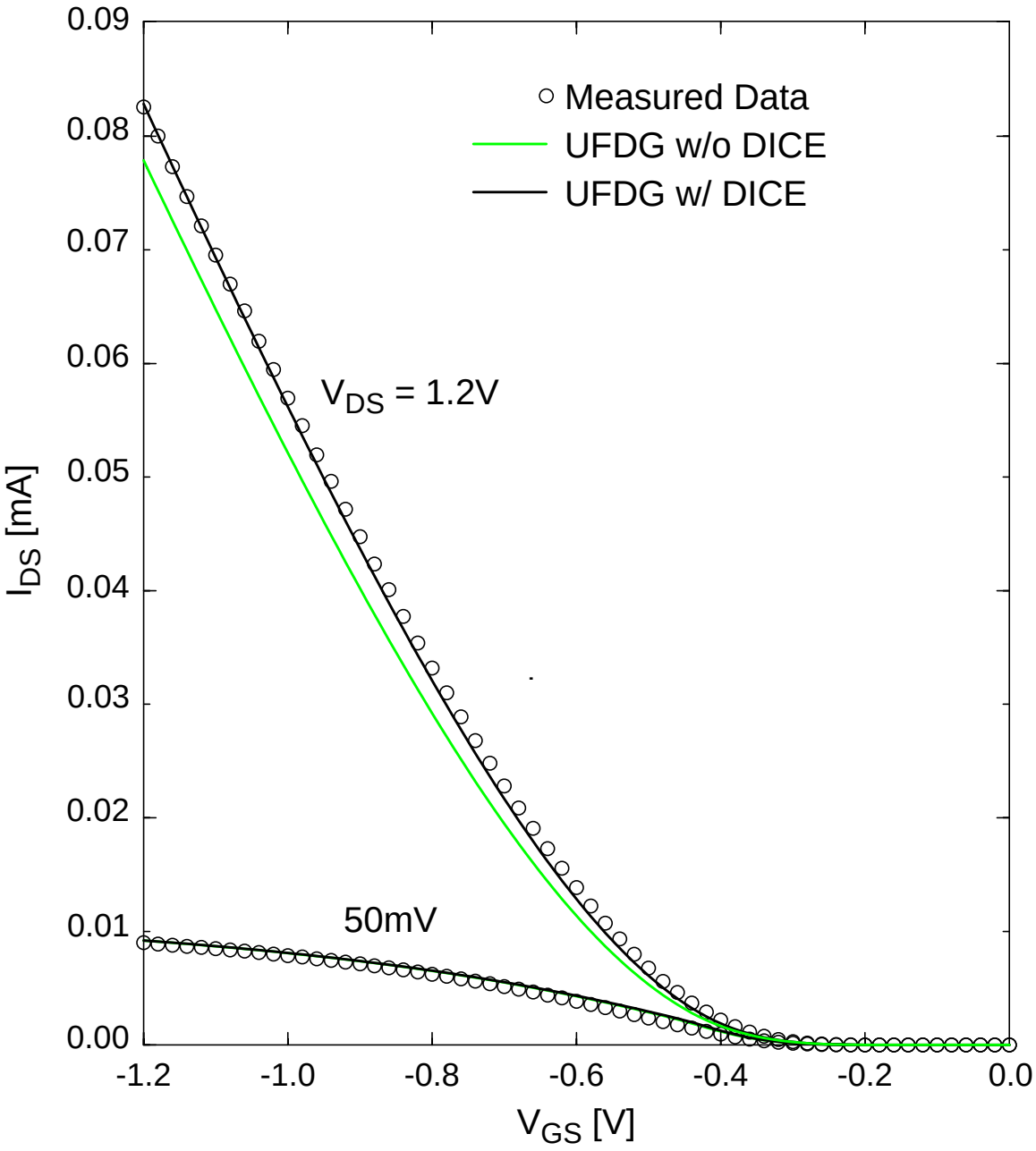


Figure 2-1. Results of calibrating UFDG to an undoped L_g = 60nm DG pFinFET (fin aspect ratio h_{Si}/t_{Si} = 100nm/17nm) [22], with and without DICE.

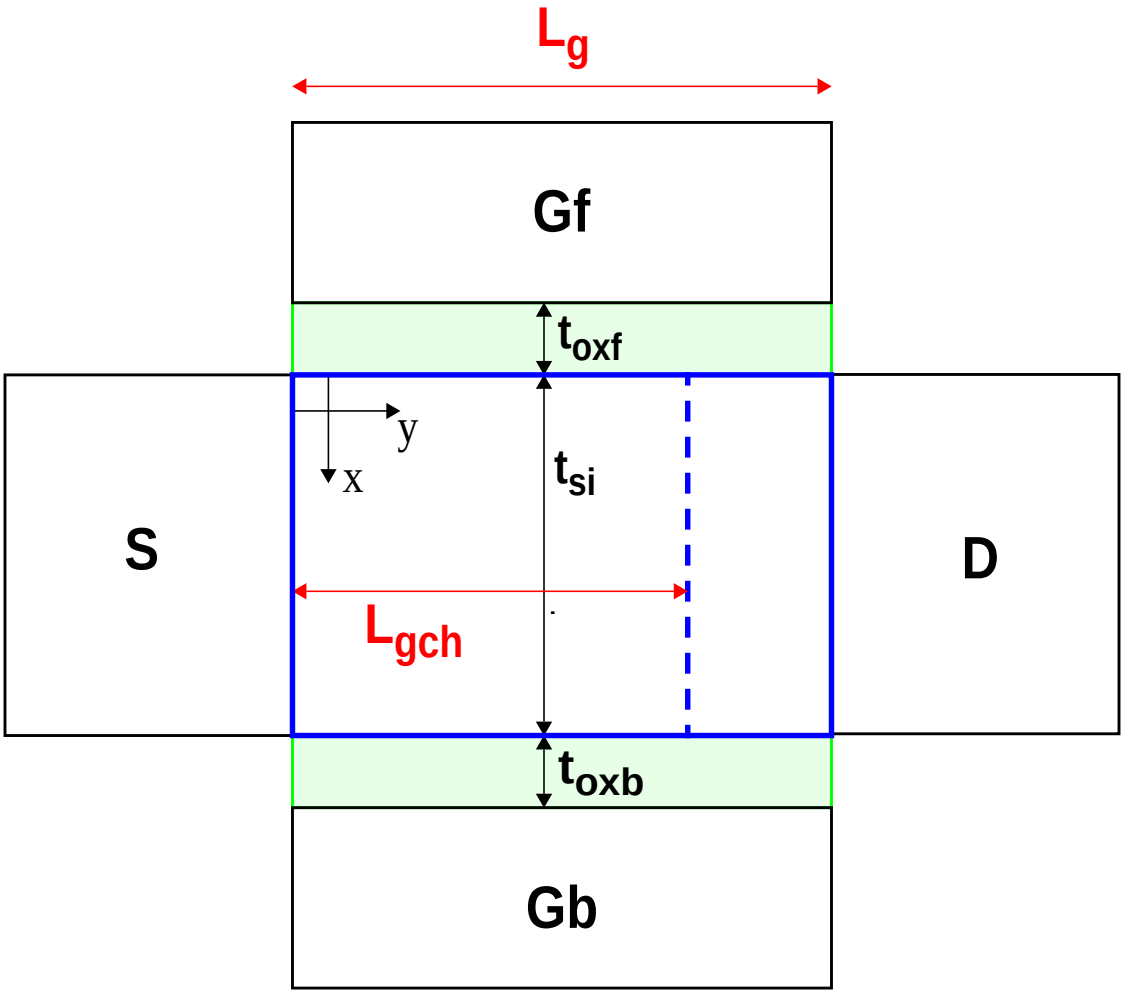


Figure 2-2. Illustration of a DG MOSFET biased in the saturation region, showing the body/channel divided into a gradual channel (L_{gch}) and a high-field ($|E_y|$) portion (L_g - L_{gch}). The effective channel length is assumed to equal the gate length here.

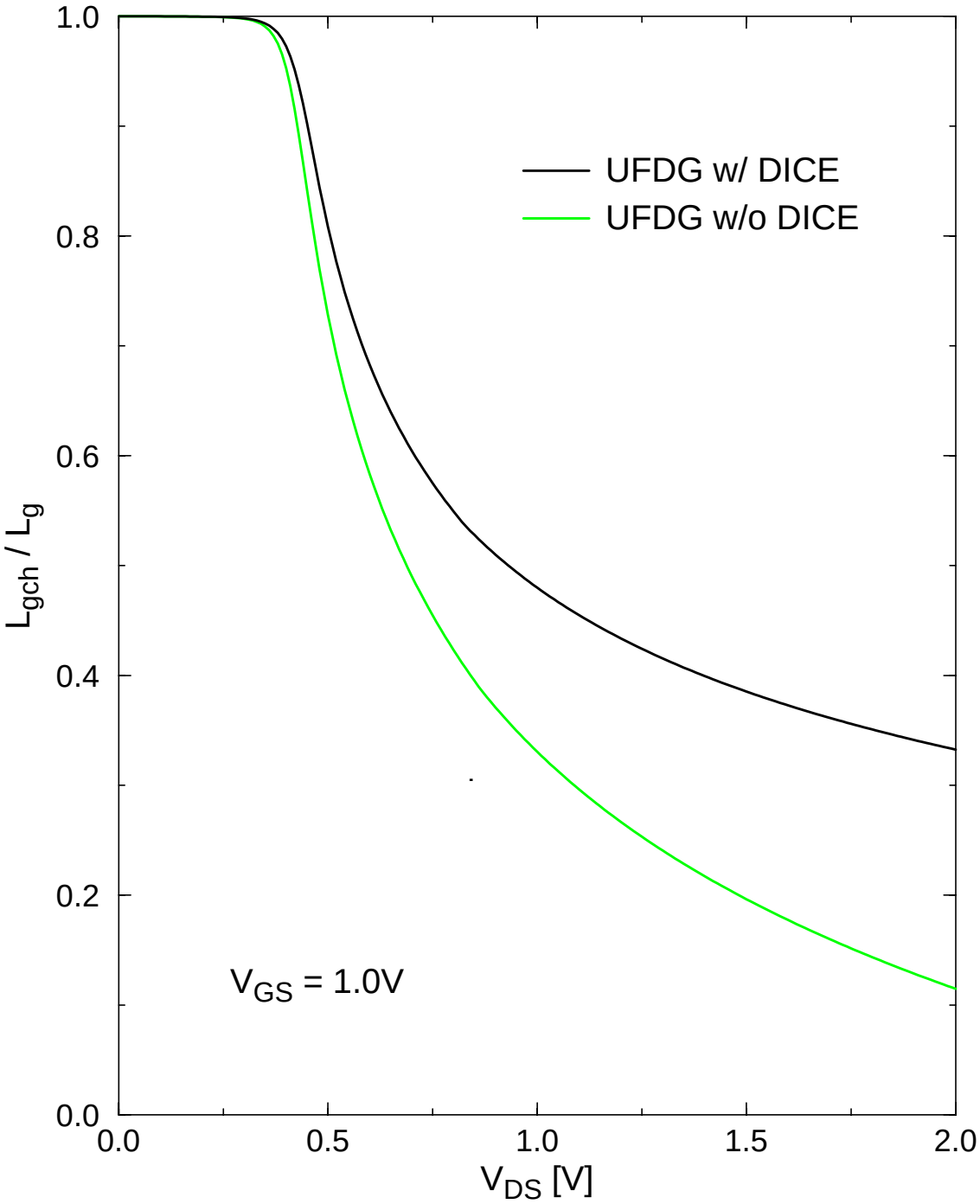


Figure 2-3. UFDG-predicted gradual-channel length, relative to L_g = 18nm, versus drain voltage for a DG nMOSFET, with and without DICE; t_{Si} = 12nm, $t_{oxf} = t_{oxb}$ = 1.2nm, midgap gate.

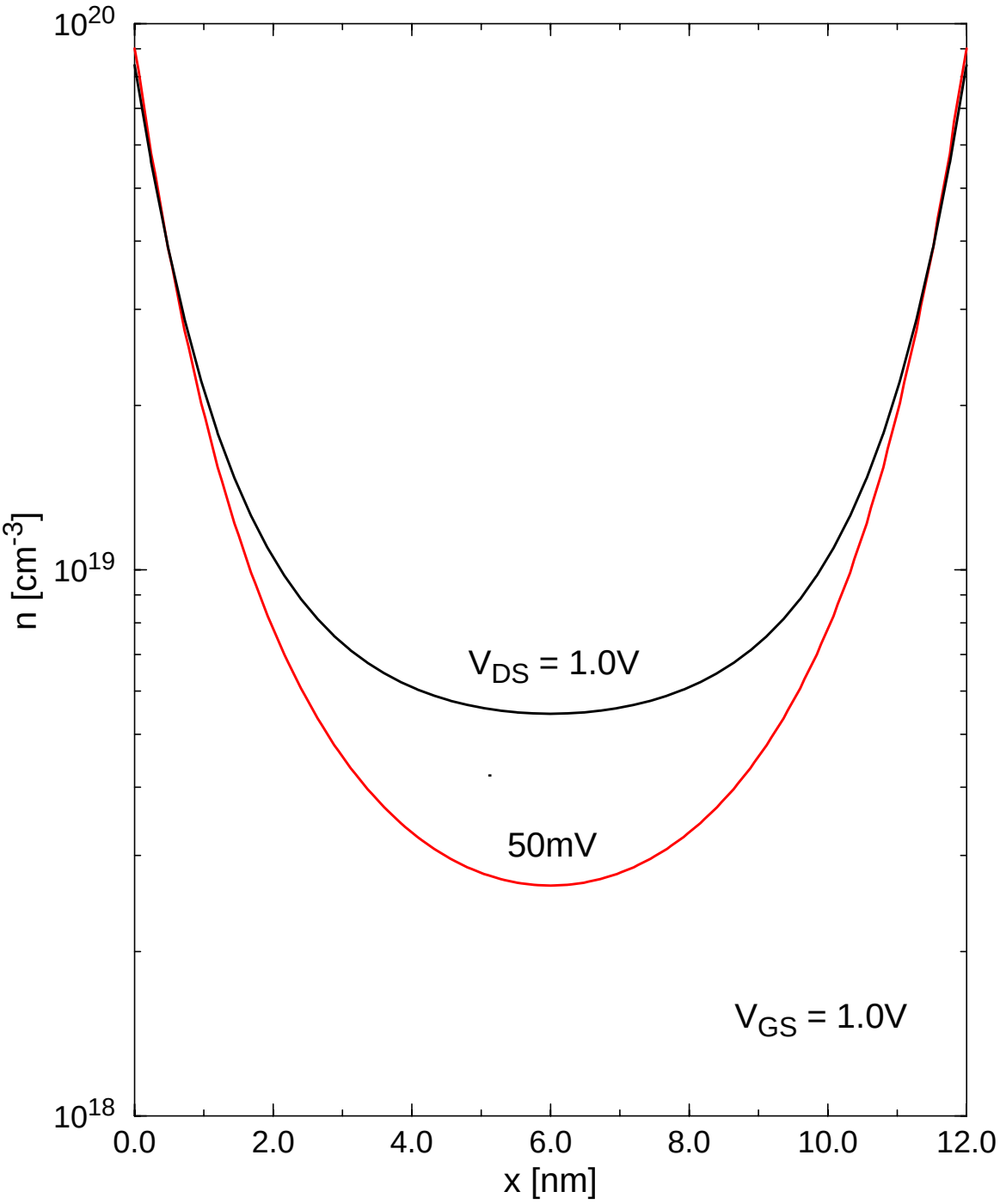


Figure 2-4. Medici-predicted inversion-electron density across the UTB at the virtual source of a simple 18nm DG nMOSFET for low and high drain voltages; t_{Si} = 12nm, $t_{oxf} = t_{oxb}$ = 1.2nm, midgap gate. The high V_{DS} increases the integrated electron density (in the bulk) by about 20%.

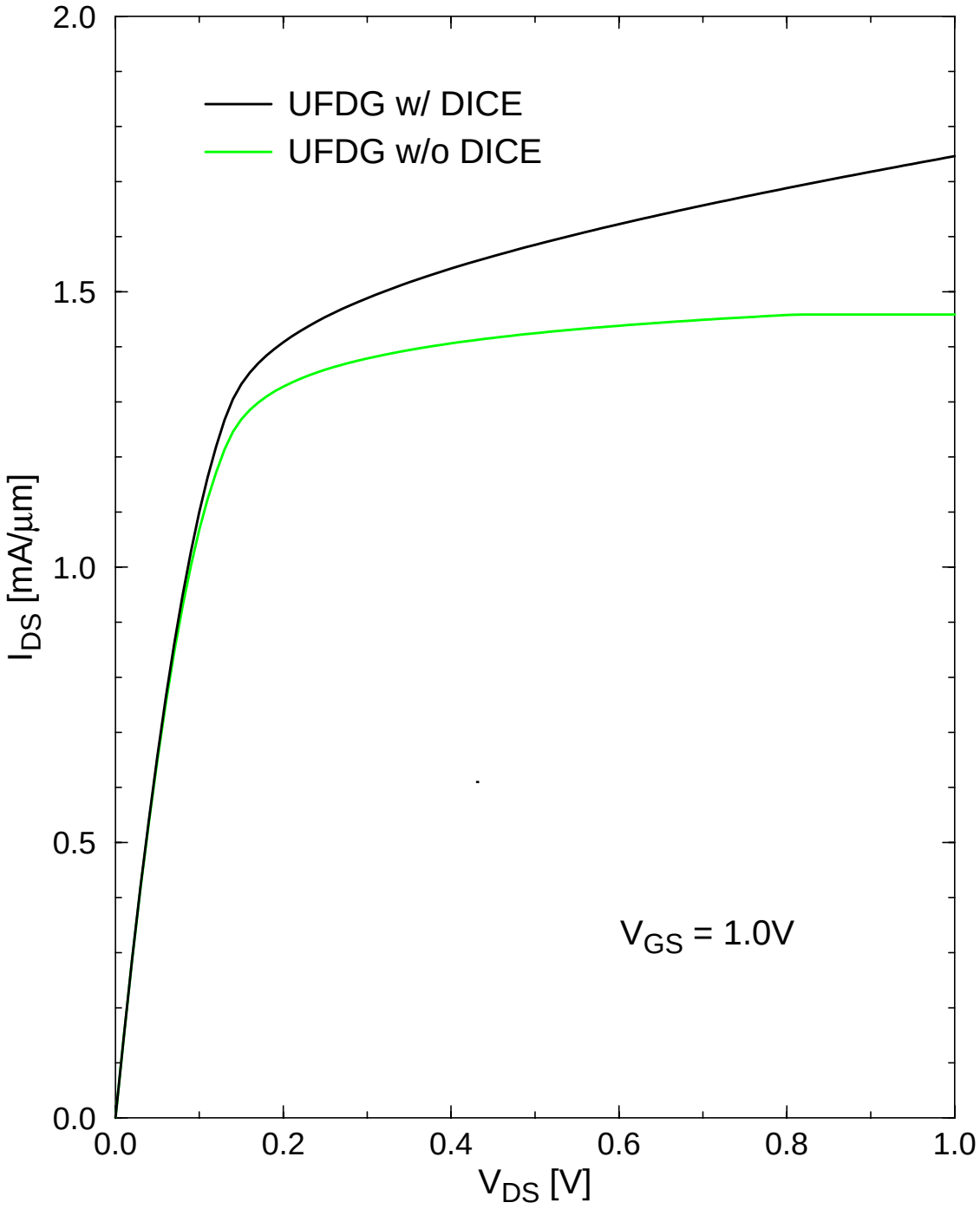


Figure 2-5. UFDG-predicted drain current versus voltage characteristics, with and without DICE, of an 18nm DG nMOSFET simplified to correspond to the Medici simulation of Fig. 2.4. At V_{DS} = 1.0V, DICE increases the near-ballistic current by about 20%, in accord with the inversion-charge enhancement indicated in Fig. 2.4.

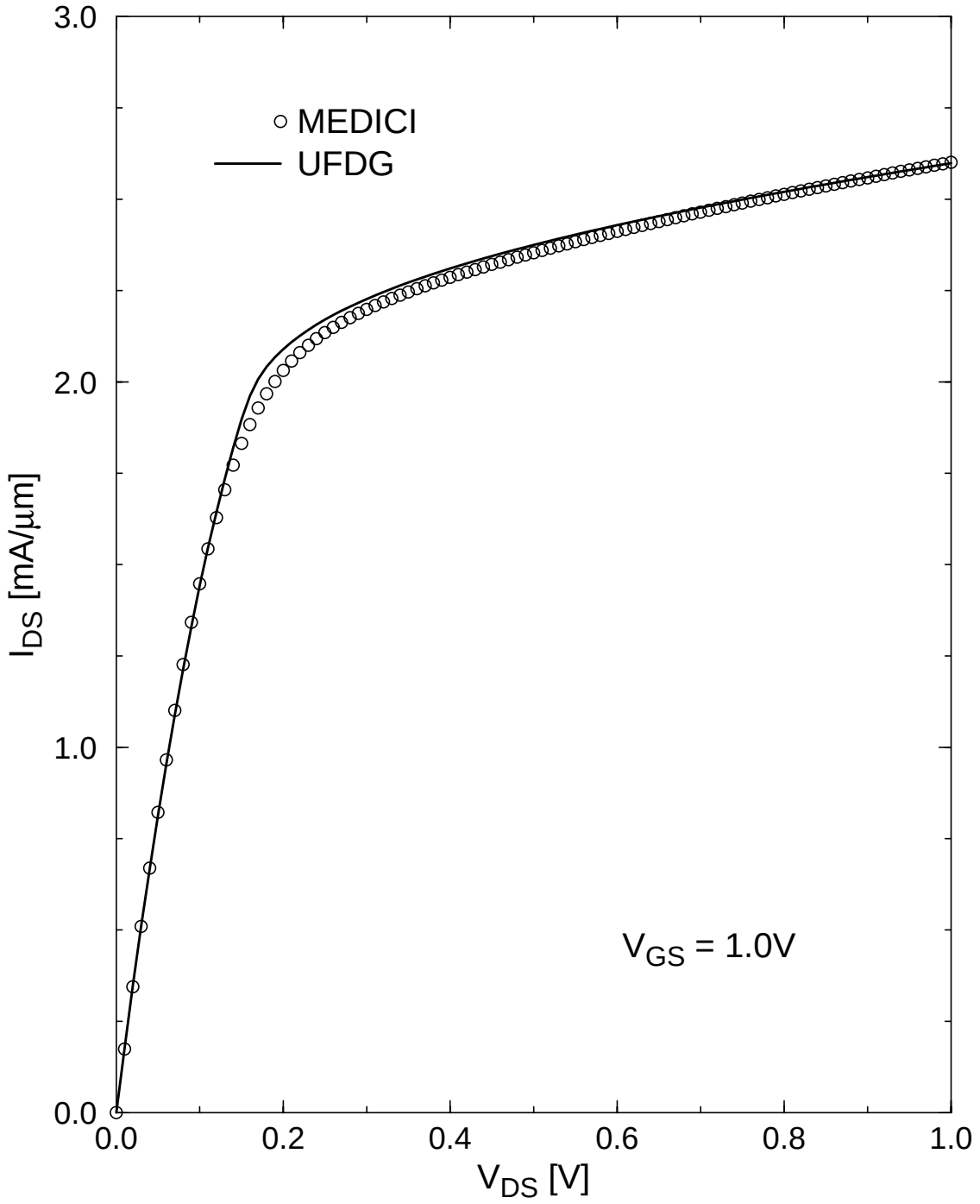


Figure 2-6. UFDG- and MEDICI-predicted drain current versus voltage characteristics of the 18nm DG nMOSFET of Figs. 2.4 and 2.5. The ballistic current limit was turned off in UFDG to correspond to MEDICI, which does not account for it.

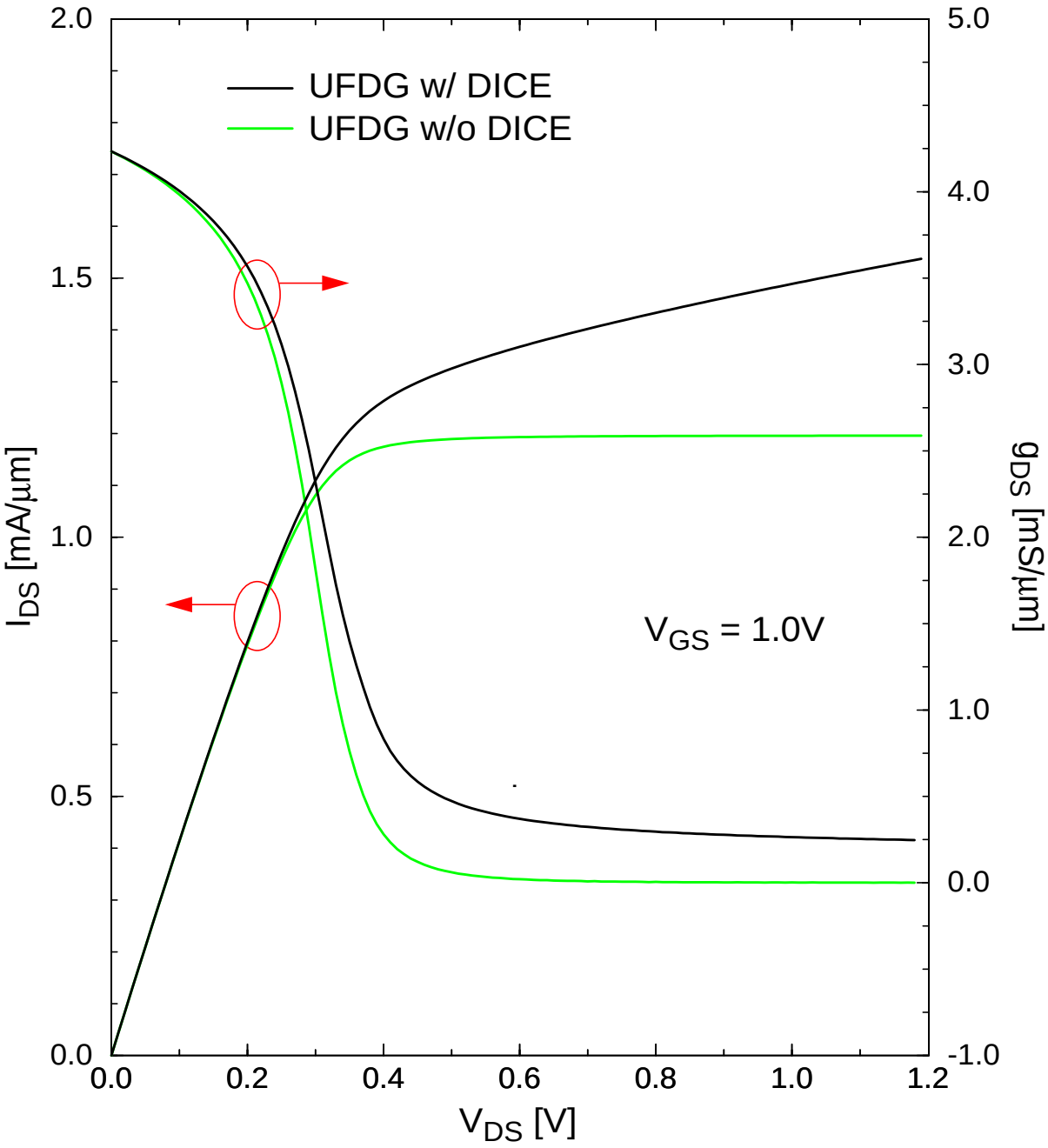


Figure 2-7. UFDG-predicted drain current versus voltage characteristics, with and without DICE, for the 18nm DG nMOSFET of Fig. 2.5, but with the proper series resistance, mobility, and velocity-overshoot modeling; t_{Si} = 12nm, $t_{oxf} = t_{oxb}$ = 1.2nm, midgap gate. At V_{DS} = 1.0V, with V_{GS} = 1.0V, DICE increases the near-ballistic I_{DS} by 24%. The predicted drain conductance, which is increased by DICE, is also shown.

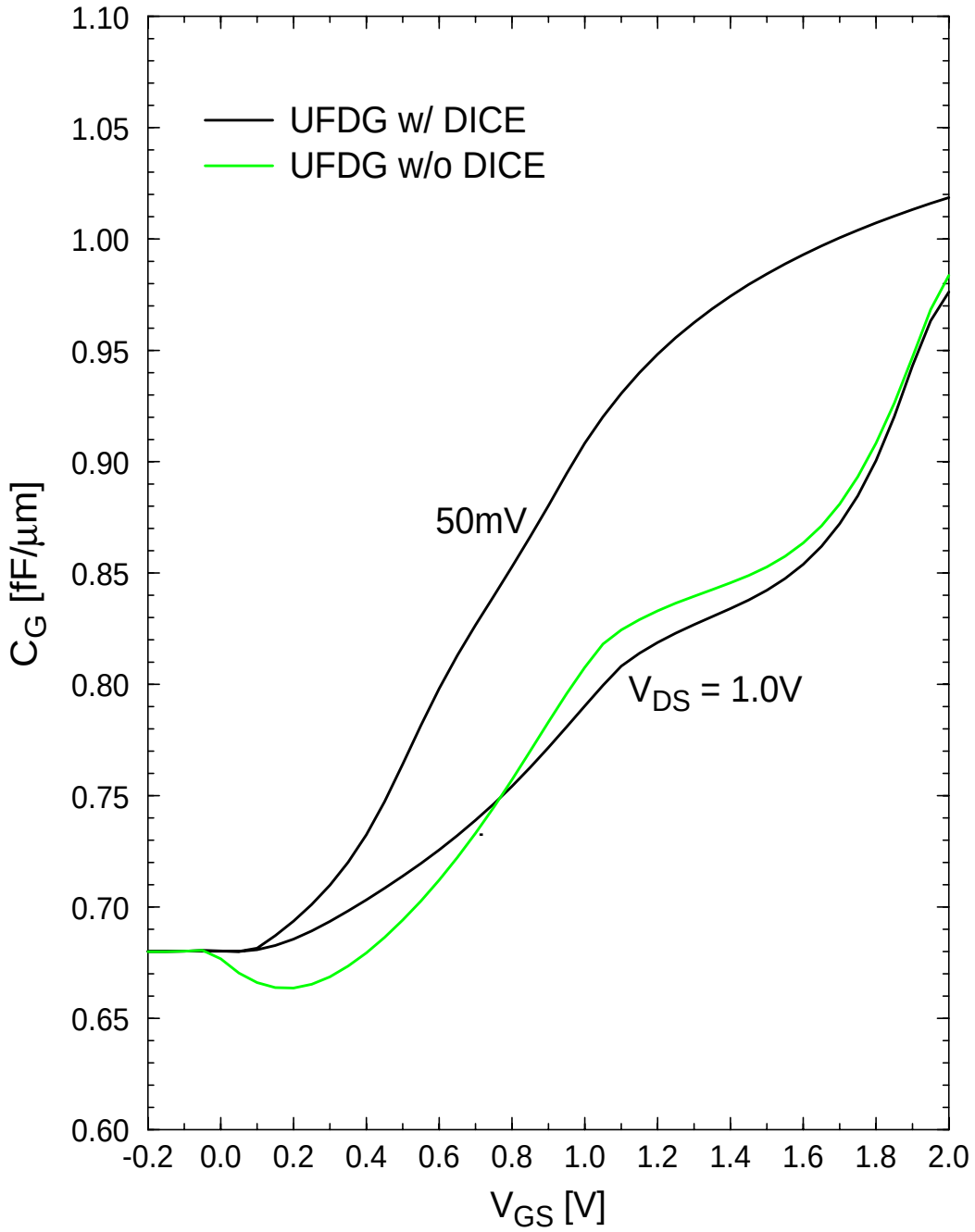


Figure 2-8. UFDG-predicted gate capacitance versus voltage characteristics of the 18nm DG nMOSFET of Fig. 2.7 at low and high drain voltages, with and without DICE. Note that not accounting for DICE results in a non-physical dip ($dC_G/dV_{GS} < 0$) in moderate inversion. DICE effectively removes this dip by increasing the strong-inversion gate charge (because of longer L_{gch}) without affecting C_G ($=dQ_G/dV_{GS}$) significantly.

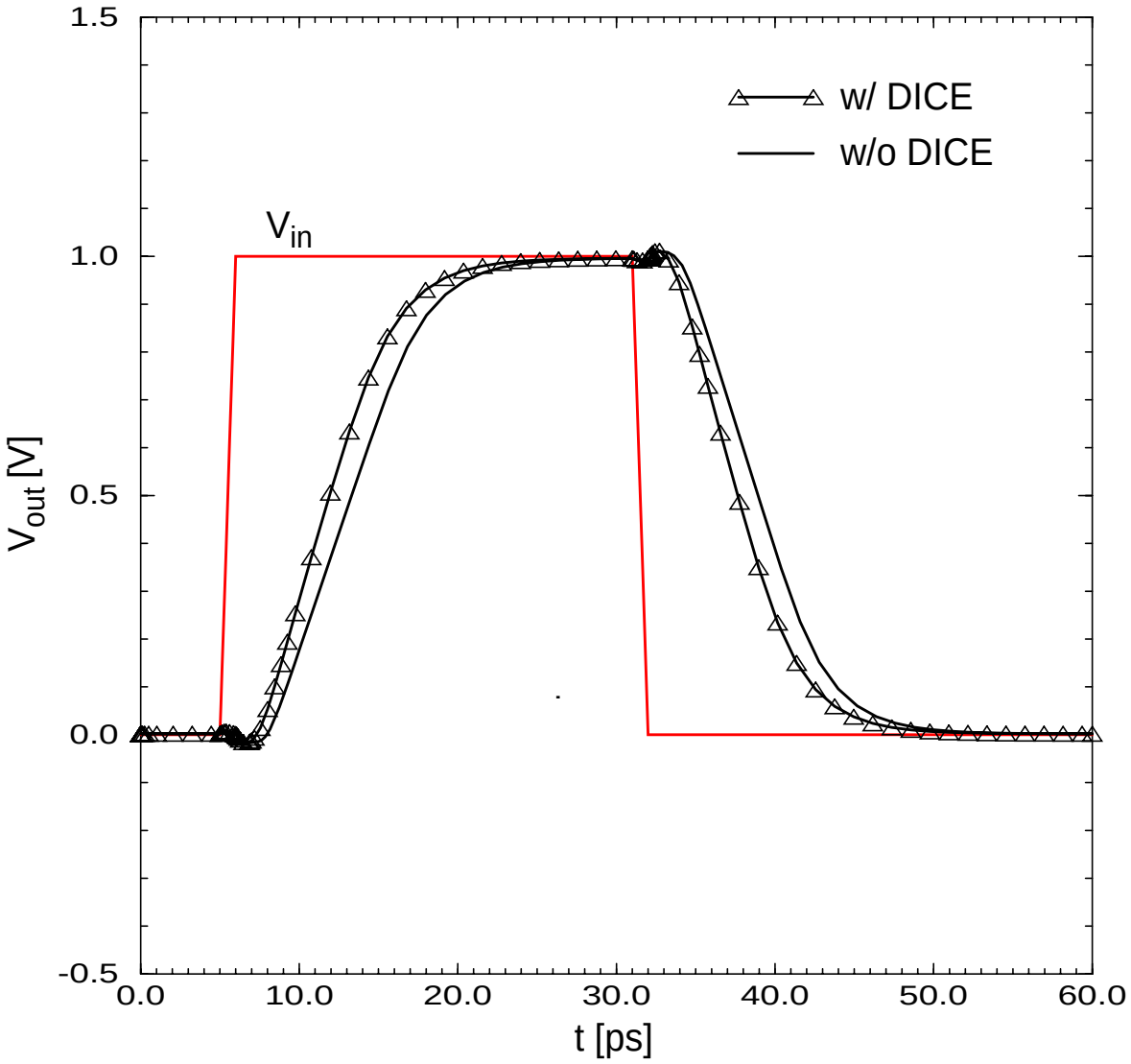


Figure 2-9. UFDG/Spice3-predicted output voltage transient of a two-stage 18nm DG CMOS inverter chain, with the input voltage pulsing shown to reveal the (pull-down plus pull-up) propagation delay; V_{DD} = 1.0V. The DG MOSFETs were designed like that in Figs. 2.7 and 2.8; the output was loaded with a capacitor with value comparable to the gate capacitance of the MOSFETs.

CHAPTER 3
THRESHOLD VOLTAGE ADJUSTMENT IN NANOSCALE DG FINFETS VIA LIMITED SOURCE/DRAIN DOPANTS IN THE CHANNEL

3-1 Introduction

As mentioned in Chapter 1, for nanoscale FinFETs, the fin-body/channel must be ultra-thin, and hence must be left undoped to avoid random-doping effects on V_t. Adjusting V_t for different CMOS applications, i.e., LP and HP, is thus a design problem; varying V_t via near-midgap-gate work-function engineering has not been shown to be a viable option. In this chapter, we propose a design approach in which S/D dopants are allowed to diffuse into the FinFET channel for V_t adjustment in HP versus LP applications. We demonstrate and verify, via simulations and measurement results, that such engineering of the lateral S/D doping profile [$N_{SD}(y)$] can be controlled to reliably lower V_t in strong inversion (for higher I_{on}) while not affecting it significantly in weak inversion (for lower I_{off} with negligible process-induced variations). We further show, via rigorous random-doping analyses, that the effects of random-dopant fluctuations (RDF) are acceptable in such FinFET design.

We have previously put forth the idea of *pragmatic* FinFET-CMOS design [2], which uses only one, near-midgap gate metal for the nMOS and pMOS devices and retains the (relatively thick) SiON gate dielectric. In this context, we showed that incorporating a G-S/D underlap can be used to effect an I_{off}-I_{on} tradeoff via a L_{eff} that decreases with increasing V_{GS} [26]. We demonstrated this idea for nanoscale-FinFET SRAM design [17]. The new design approach proposed herein augments the utility of S/D engineering for G-S/D underlap [26], enabling wider V_t adjustment for different CMOS applications, and, unlike gate work-function engineering, allowing independent adjustment of V_t for I_{off} and I_{on}.

3-2 S/D Doping-Dependent V_t

The viability of our proposed design for V_t adjustment stems from the dependence of V_t on the *distribution* of S/D dopants in the UTB/channel, which we first analyze. Possible doping profiles, assumed to be gaussian {$N_{SD}(y) = N_0 \exp[-(y+L_{ext})^2/\sigma_L^2]$ where N_0 is the density at the S/D end of the S/D extension and σ_L is the lateral straggle}, are illustrated in Fig. 3.1. For weak inversion, the current is determined by diffusion of carriers in a portion of the channel over which the electric potential is relatively invariant in y, and the threshold voltage (V_{tw}) is defined at the point of minimum potential [19] within this region. Because of negligible longitudinal (in y) electric field in this region, V_{tw} can be expressed based on a 1-D (in x) analysis. For an undoped (negligible acceptor dopants) n-channel DG FinFET with a midgap gate,

$$V_{tw} \cong \phi_c - \frac{qN_{SD}(L_g/2)t_{Si}}{2C_{ox}} - \Delta V_t^{DIBL}, \tag{3.1}$$

where ϕ_c is the characteristic surface potential at threshold (≅0.4V) [3] and ΔV_t^{DIBL} represents the reduction in V_{tw} due to DIBL [19]; t_{Si} is the silicon-fin thickness and $C_{ox} = \varepsilon_{ox}/t_{ox}$. The N_{SD} term in (3.1) defines the reduction in V_{tw} due to S/D-donor depletion charge (assumed to be uniform in x) in the channel. This term is defined at $y = L_g/2$ because, in nanoscale FinFETs, the noted diffusion-current region is short and located near the center of the channel [19]. Note then that $N_{SD}(L_g/2)$ is most critical in defining V_{tw}, and I_{off}. For nanoscale FinFETs, we note that the depletion-charge term will typically be negligible unless $N_{SD}(L_g/2) > \sim 10^{18}cm^{-3}$ [3]. We also note that N_{SD} away from the center of the channel can also affect I_{off} by perturbing the solution of the 2-D Poisson equation in the UTB, which defines ΔV_t^{DIBL} [19]. However, this perturbation, which increases ΔV_t^{DIBL} due to more encroachment of the S/D electric field, is typically small when $N_{SD}(L_g/2) < 10^{18}cm^{-3}$. We have neglected the V_{tw} increase due to quantization; this increase is small in symmetrical undoped DG MOSFETs for $t_{Si} > 4nm$ [28].

For strong inversion, the current is drift and is determined by the average conductivity of the (gradual) channel. The (aerial) inversion-charge density (Q_{ch} is magnitude) in the channel, with $N_{SD}(y)$, can be expressed as [19]

$$Q_{ch}(y) = Q_{ch0} - 2C_{ox}\Delta\phi(y) + qN_{SD}(y)t_{Si} + \Delta Q_{ch}^{DICE} \tag{3.2}$$

where $Q_{ch0} = 2C_{ox}(V_{GS} - \phi_s)$ is the inversion-charge density at the (virtual) source for $N_{SD} = 0$, with ϕ_s being the surface potential ($>\phi_c$ due to finite inversion-layer capacitance) there, $\Delta\phi$ is the V_{DS}-induced perturbation in the potential along the channel, and ΔQ_{ch}^{DICE} is the enhancement in the inversion-charge density due to DICE. Note that Q_{ch} increases linearly with N_{SD}. The current is defined by integrating (3.2) along the gradual channel (L_{ch}) [19], and it is increased by N_{SD} accordingly. Based on the integration, Q_{ch0} is effectively increased to ($Q_{ch0} + q\overline{N}_{SD}t_{Si}$), where

$$\overline{N}_{SD} = \frac{1}{L_{ch}} \int_0^{L_{ch}} N_{SD}(y)dy \tag{3.3}$$

is the average density of S/D dopants in the gradual channel. The (extrapolated) threshold voltage (V_{ts}) can hence be defined by

$$Q_{ch0} + q\overline{N}_{SD}t_{Si} \cong 2C_{ox}(V_{GS} - V_{ts}), \tag{3.4}$$

or

$$V_{ts} \cong \phi_s - \frac{q\overline{N}_{SD}t_{Si}}{2C_{ox}} - \Delta V_t^{DICE}, \tag{3.5}$$

where ΔV_t^{DICE} represents the reduction in V_{ts} due to DICE. Note that L_{ch} depends on both V_{DS} and V_{GS}, and hence so do $\overline{N}_{SD}$ and V_{ts}.

Note how V_{tw} in (3.1) [where $I_{off} \propto \exp(-qV_{tw}/kT)$] and V_{ts} in (3.5) [where $I_{on} \propto (V_{GS} - V_{ts})$] differ. For weak inversion, ΔV_t^{DIBL} is significant, and V_{tw} is lowered by the depletion-

charge term near the center of the channel where the carrier transport is predominantly diffusion [19]. Thus, $N_{SD}(y \sim L_g/2)$ must be designed to be less than ~$10^{18}cm^{-3}$ to avoid excessive random-doping effects on I_{off}. For strong inversion, ΔV_t^{DIBL} is small because of the large transverse electric field [2], however ΔV_t^{DICE} could be significant, and V_{ts} is lowered by the integrated depletion charge. Thus, if $\overline{N}_{SD}$ is greater than ~$10^{18}cm^{-3}$, and $N_{SD}(y \sim L_g/2)$ is negligible as noted, then I_{on} can be enhanced by a lowered V_{ts}, and I_{off} and V_{tw} will not be affected significantly, provided the noted $\Delta V_t^{DIBL}(N_{SD})$ increase is limited (which may require an I_{on}-I_{off} design tradeoff). The $N_{SD}(y)$ labeled HP in Fig. 3.1 is such a doping profile. The $N_{SD}(y)$ labeled LP in Fig. 3.1 shows negligible N_{SD} everywhere in the channel, but yields a significantly beneficial G-S/D underlap. The underlap yields $L_{eff} > L_g$ in weak inversion, and thereby reduces the SCEs [26] and parasitic capacitance as well [29]. The essence of our S/D engineering-based design, including the extension for V_t adjustment, is thereby defined. We show in Fig. 3.2 measured I_{DS}-V_{GS} characteristics of two L_g = 70nm DG FinFETs having different $N_{SD}(y)$. From our UFDG [16]-calibration results, we find that the doping profiles in devices 2J and 1G are like HP and LP profiles, respectively, albeit nonoptimal since these devices were fabricated [22] without emphasis on S/D engineering. The difference of the strong-inversion V_{ts} between the two devices shown in Fig. 3.2(a), which is due to different $\overline{N}_{SD}$ in the channels, is significant. (Note that V_{ts} can decrease with increasing V_{GS} due to the previously noted bias dependence of N_{SD} in (3.3).) The V_{tw} variation between the two devices in Fig. 3.2(b) (≅200mV) is due largely (≅100mV as inferred from the data) to different SCEs as defined by L_{eff}. Note that the corresponding values of I_{on} and I_{off}, interpreted based on our physical insights from (3.1)-(3.5), experimentally confirm the efficacy of the proposed V_t design approach for relatively long (70nm) DG FinFETs. For shorter devices, the approach should still be valid as long as L_g is not short enough (<20nm as implied by UFDG projections [11]) such that I_{on} is limited by ballistic

transport. For such short devices, I_{off} will still be defined by V_{tw} in (3.1), and I_{on} will be defined by $Q_{ch}(0)$ in (3.2) which will increase with $N_{SD}(0)$. Therefore, our proposed design approach is generally useful for future LP/HP applications of nanoscale DG FinFETs. (We would like to thank Shishir Agrawal, a Ph.D. student at the University Florida, for calibrating measured characteristics with UFDG.)

3-3 Demonstration and Verification of Design Approach

We further check the proposed S/D engineering for LP (low I_{off}) and HP (high I_{on}) FinFET-CMOS applications using our process/physics-based compact model for DG MOSFETs, UFDG (with physical modeling of carrier mobility, velocity overshoot, quasi-ballistic transport, and quantization) [16], linked [17] to 2-D numerical device simulations done with Medici [24]. For L_g = 18nm (HP45nm node [1]), with a midgap gate and t_{ox} = 1.0nm, we assume L_{ext} = 12nm, and an undoped UTB with t_{Si} = 10nm for adequate SCE control. We define $N_{SD}(y)$ with different straggle, as in Fig. 3.1, for an LP (σ_L = 5.5nm, which yields weak-inversion L_{eff} = 24nm and negligible N_{SD} in the channel) and an HP (σ_L = 8.5nm, which yields L_{eff} = 18nm, $\overline{N}_{SD}$ = $4x10^{18}cm^{-3}$, and $N_{SD}(y \sim L_g/2) = 5x10^{17}cm^{-3}$) application. We assume a nominal (doable [22]) S/D series resistance R_{SD} = 100Ω-μm, and, for the LP design, increase it by ΔR_{SD} = 25Ω-μm due to the underlap [17].

Fig. 3.3(a) shows UFDG-predicted subthreshold I_{DS}-V_{GS} characteristics of the two FinFETs, which reflect a V_{tw} difference of ≅180mV (at 100nA/μm/L_g). We stress that the lower V_{tw} of the HP device is due mainly to the shorter L_{eff} (=L_g) and larger SCEs. (A ~20mV reduction in the HP V_{tw} caused by N_{SD} in the channel was ignored.) So, the longer L_{eff} for the LP device yields the (1/300)x decrease in I_{off}, relative to the HP device, evident in Fig. 3.3(a). (For the HP design, we did not account for the reduction in subthreshold mobility [~x(1/2.5)] due to Coulomb scattering by the S/D dopants [10]; I_{off} would thus be lowered accordingly, tending to offset the

small increase due to the noted 20mV-lower V_{tw}.) UFDG-predicted strong-inversion I_{DS}-V_{GS} characteristics (per fin height), at $V_{DS} = V_{DD} = 1.0V$, of the LP and HP DG nFinFETs are plotted in Fig. 3.3(b). Note that V_{ts} of the HP device is considerably lower than that of the LP device [by ≅100mV, as defined by $\overline{N}_{SD} = 4x10^{18}cm^{-3}$ in (3.5)]. The HP I_{on} in Fig. 3.3(b) is 29% higher than that of the LP device; about 20% of this increase is due to the lower V_{ts}, and rest is due to the lower R_{SD}. (For the HP device, we evaluated $\overline{N}_{SD}$ at low V_{DS}, but we note that its value at V_{DS} = 1.0V could be lower due to reduction in L_{ch} with increasing V_{DS}. Due to uncertainty in L_{ch} at $V_{DS} = 1.0V$, $\overline{N}_{SD}$ is difficult to evaluate accurately. However, we estimate that the worst-case enhancement in I_{on} due to the V_{ts} lowering is about 15%.) To solidify the $N_{SD}(y)$-based design approach, we also show in Fig. 3.3 the UFDG-predicted I_{DS}-V_{GS} characteristics of the LP device with L_g increased to 28nm, as projected (for LSTP) in the SIA roadmap [1] for the 45nm node. The longer L_g yields an acceptable I_{off} ≅ 10pA/μm, more than four-orders of magnitude lower than that of the HP device. UFDG/Spice3-predicted RO delays for this LP and the HP design are 3.6ps and 1.9ps, respectively, at $V_{DD} = 1.0V$. The delay for the HP device at the same bias without V_{ts} reduction is 2.4ps; the limited S/D doping in the channel reduces the HP delay by more than 20%.

3-4 Sensitivity and RDF Analyses

To absolutely confirm the viability of our proposed V_t-adjustment design, we must consider sensitivities to process variations and RDF effects. For the former, we did a Medici/UFDG-based analysis, results of which are shown in Table 3.1. We find from the simulations that letting σ_L vary by +/-18% in both designs of Figs. 3.1 and 3.3 does not cause prohibitive variations in I_{off}: 1.5x/(1/1.3)x for LP and 80x/(1/20)x for HP. And, sensitivity to variation in the local number of dopants is acceptable as well. The HP profile in Fig. 3.1 shows $N_{SD}(L_g/2) = 5x10^{17}cm^{-3}$, which implies only ~1 dopant near the center of the channel. This number could randomly increase to 3,

and the lowered V_{tw}, given by (3.1) with the implied $N_{SD}(y \sim L_g/2)$, would not prohibitively increase I_{off} (<100x). (These weak-inversion results are valid, but we have found that for undoped UTB devices with G-S/D underlap, Medici-predicted strong-inversion currents are not reliable due to inadequate UTB-transport modeling.) In Table 3.1 we also show the UFDG-predicted variations in I_{on}. The variations in I_{on} are much lower than those in I_{off} because I_{on} varies linearly with N_{SD} as implied by (3.2), whereas I_{off} is exponentially dependent on N_{SD}. Also, I_{on} depends on $\overline{N}_{SD}$ and not directly on N_{SD}, and so the exact position of dopants in the channel is not important for I_{on}.

The main concern about the RDFs is the effect on V_{tw} [31] and I_{off}, as implied by Table 3.1. Therefore, we did a more detailed Medici-based study of the variations in I_{off} due to RDFs. As illustrated in Fig. 3.4, we divided the S/D-extension and the channel/UTB regions via grids with equal spacings, and assumed for this analysis that the height of the fin (width of the FinFET) is equal to the grid spacing (2nm as in [31]). For each lattice site in every cube thus formed, the probability of having a S/D dopant at that location is calculated based on the $N_{SD}(y)$-defined doping density at that location, and then a silicon atom or dopant atom is randomly placed at that location, following the method described in [32]. The number of dopants randomly placed in the cube, divided by the volume of the cube, defines the doping density assigned to the corresponding region in the Medici domain. Once the doping density is specified every where in the S/D-extension and channel regions, I_{off} and V_{tw} are extracted from the Medici-predicted current-voltage characteristics. The process is repeated a sufficiently large number of times, and the variations are noted. Because of the large number of simulations used, the RDF along the actual height of the FinFET is implicitly accounted for [31], thereby enabling use of the noted (unrealistically) small fin height and 2-D Medici simulations for computational efficiency. (We would like to thank Askhan Behnam, a Ph.D. student at the University of Florida, for helping with

the above simulations.)

Table 3.2 shows the predicted standard deviation of V_{tw} [$\sigma(V_{tw})$], and device yield for 18nm LP and HP DG FinFETs based on acceptable variations in I_{off} (<100x). For the HP device, the 95% yield indicates the viability of our design approach. The yield increases, and $\sigma(V_{tw})$ decreases, as we limit the S/D dopants in the channel by decreasing σ_L. Insignificantly small $\sigma(V_{tw})$ in the LP device implies negligible variations in I_{off}, which is crucial for LP applications. Note that since we considered the random variations of dopants in the S/D-extension regions, our simulation results in Table 3.2 account for the effect of variation in L_{eff} as well. Based on our results in Tables 3.1 and 3.2 then, we conclude that RDF effects are virtually nonexistent in the LP design, and are adequately controlled in HP design when $N_{SD}(y \sim L_g/2)$ is limited as noted.

3-5 Summary

An extended approach to nanoscale DG FinFET design for LP and HP nanoscale-CMOS applications via S/D engineering [i.e., control of $N_{SD}(y)$ for G-S/D underlap and V_t adjustment] was proposed, and demonstrated to be viable by device measurements and simulations, including sensitivity and RDF-effects studies. The approach exploits the idea of allowing limited S/D dopants properly distributed in the channel for HP-V_t design. We demonstrated the design approach at the 45nm node. Scaling L_g to <10nm, as projected at the end of the SIA roadmap [1], will require the lateral straggle of $N_{SD}(y)$ to be reduced by about a factor of two, which appears feasible with acceptable sensitivities via new processing such as laser annealing. However, additional analyses including quantization effects, mobility degradation, and ballistic transport are called for. Our work extends the utility of S/D engineering in the design of nanoscale FinFETs, now for V_t adjustment as well as G-S/D underlap. But, additional work on S/D processing for control of $N_{SD}(y)$ in the thin-fin extensions and channel is needed to effect this utility.

Table 3-1. Medici-predicted sensitivity of I_{off} to variations in σ_L (+/-18% of the nominal σ_{L0} noted) of $N_{SD}(y)$ in the L_g = 18nm LP and HP DG FinFETs.

σ_L	+18%	-18%
LP (σ_{L0}=5.5nm)	1.5x	(1/1.3)x
HP (σ_{L0}=8.5nm)	80x	(1/20)x

Table 3-2. Medici-predicted standard deviation of V_{tw} and associated I_{off}-based yield due to the RDF of $N_{SD}(x,y)$ in the L_g = 18nm LP and HP DG FinFETs.

Application	$\sigma(V_{tw})$	Yield
HP (σ_L=8.5nm)	100mV	95%
LP (σ_L=5.5nm)	18mV	100%

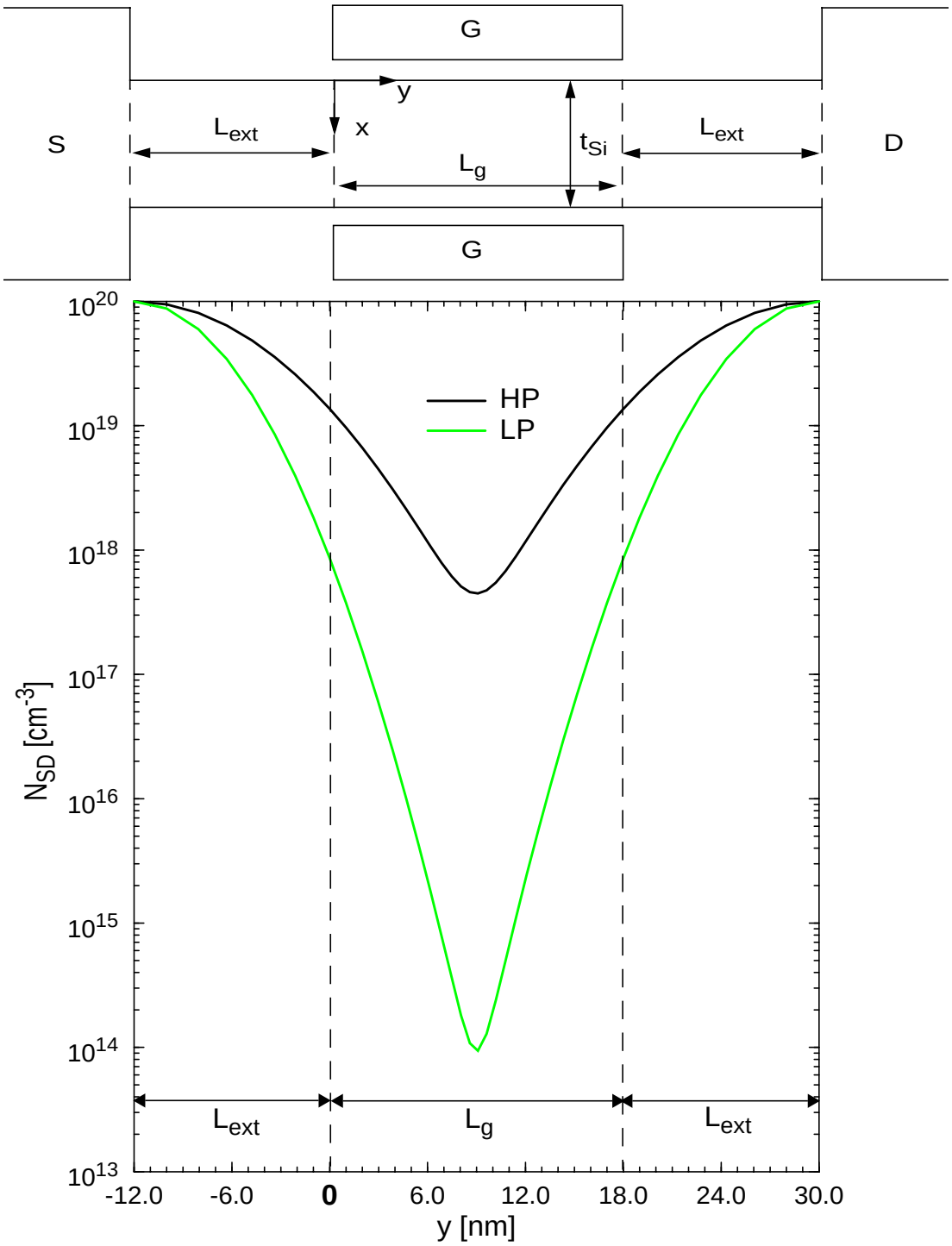


Figure 3-1. S/D-extension lateral doping profiles in an undoped DG FinFET, showing variable encroachment into the channel. The HP profile exemplifies one which lowers V_{ts} via S/D dopants in the channel, but does not affect V_{tw}; the LP profile does not affect V_{tw} nor V_{ts}, but does yield an effective G-S/D underlap. The FinFET structure corresponding to the doping profiles is indicated.

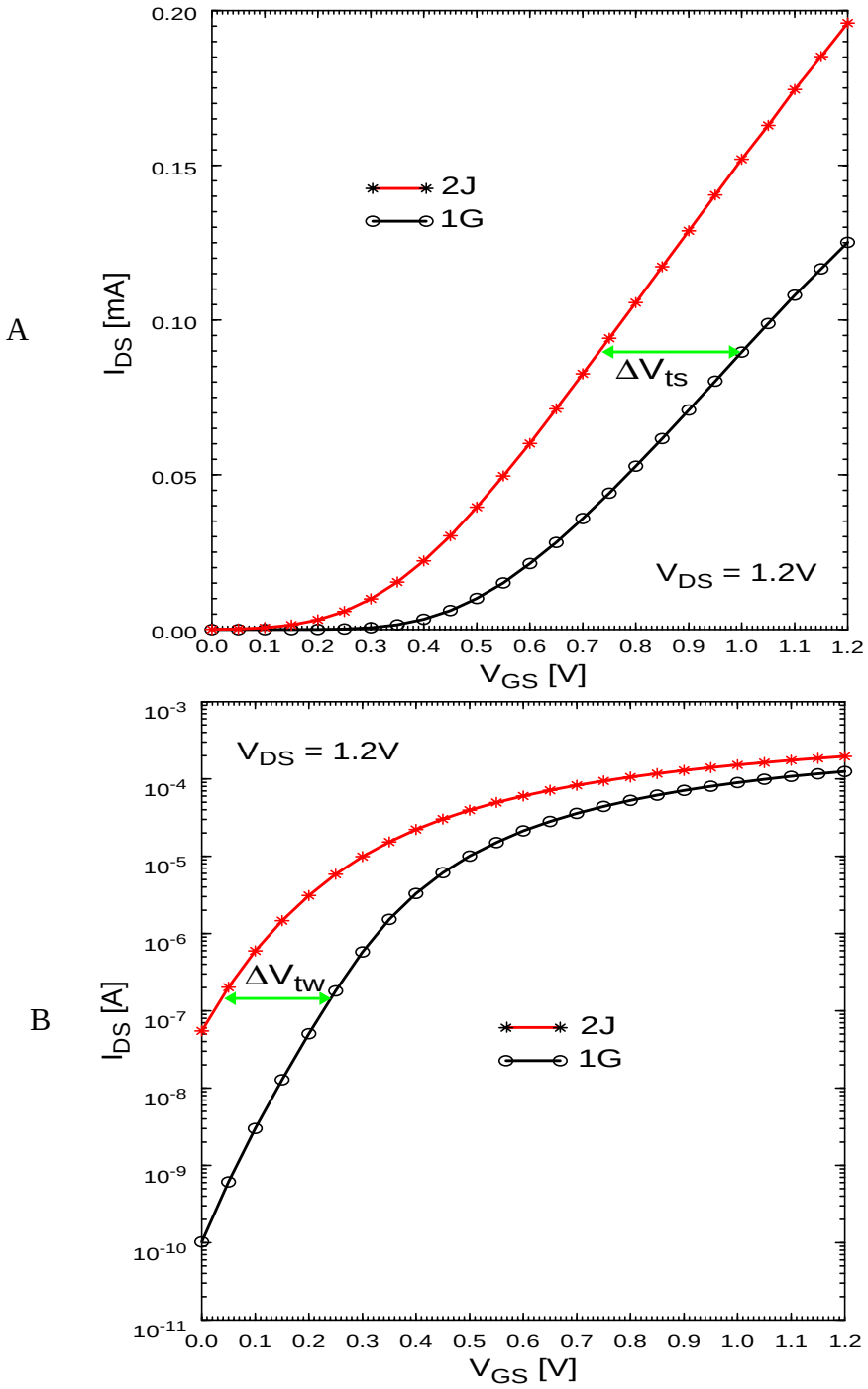
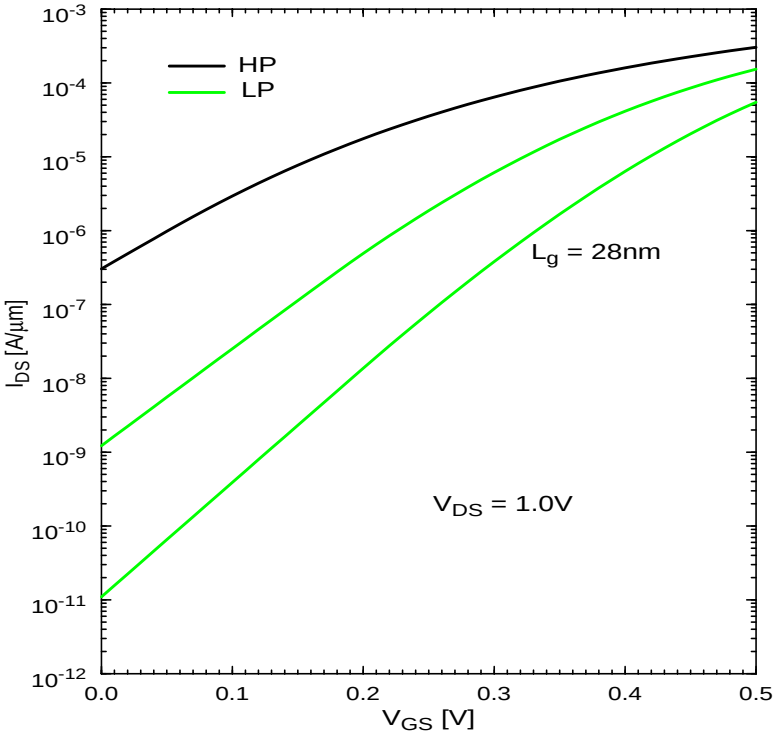


Figure 3-2. Measured current-voltage characteristics of two L_g = 70nm undoped DG nFinFETs [30], which have different $N_{SD}(y)$ due to variations in the S/D processing. The ΔV_{ts} and ΔV_{tw} indicated reflect LP (~device 1G) and HP (~device 2J) features of the respective devices, governed by $N_{SD}(y)$ as defined by (3.1)-(3.5). A) Strong-inversion characteristics. Note the strong $\Delta V_{ts}(V_{GS})$ dependence, which indicates overly excessive densities of S/D dopants in the channel of device 2J (i.e., G-S/D overlap). B) Subthreshold characteristics.

A



B

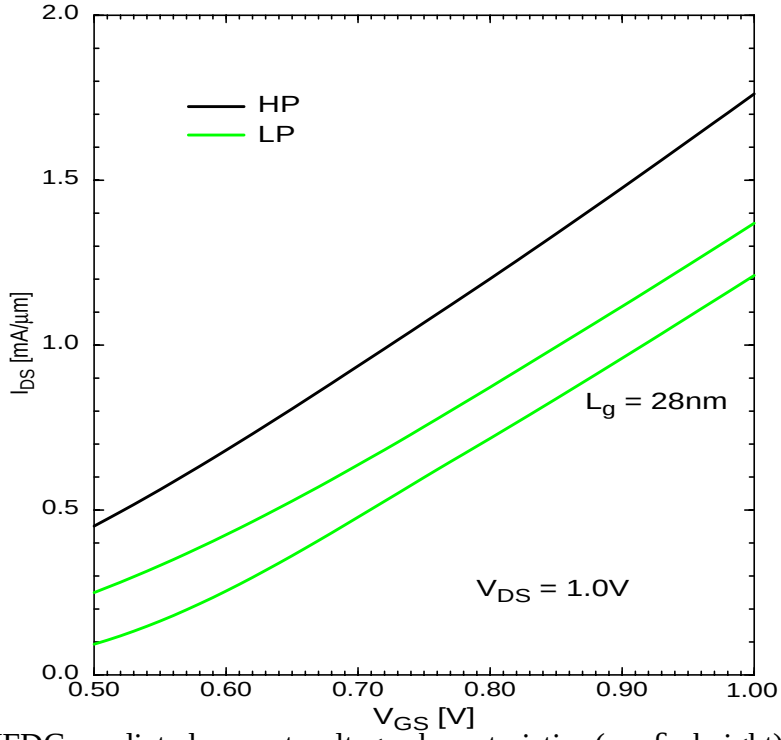


Figure 3-3. UFDG-predicted current-voltage characteristics (per fin height) of the 18nm LP and HP DG FinFETs, and of the LP device with longer L_g = 28nm as projected in the SIA roadmap [37]. A) Subthreshold Characteristics. B) Strong-inversion Characteristics. The strong-inversion characteristic of the HP device is approximate; the gate work function was decreased to account for the lowered V_{ts} due to S/D dopants in the channel.

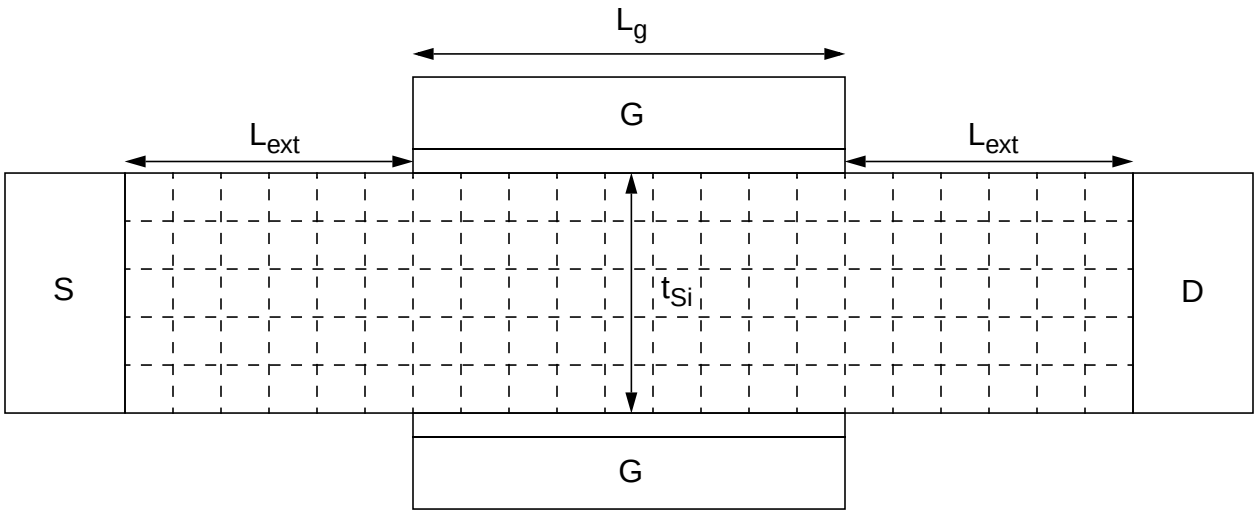


Figure 3-4. The DG FinFET structure showing how the S/D-extension and channel regions were partitioned (into 2nm cubes) to account for the RDF of $N_{SD}(x,y)$ in the (2-D) Medici domain.

CHAPTER 4
PHYSICAL INSIGHTS ON ANALOG/RF PERFORMANCE OF DOUBLE-GATE FINFETS, WITH COMPARISION TO BULK-SILICON MOSFETS

4-1 Introduction

DG MOSFETs have been shown to be much better suited than bulk-Si MOSFETs for nanoscale digital applications [33], primarily because of their inherently better immunity to SCEs and higher μ_{eff}. The high μ_{eff} [10], due mainly to low transverse electric field in the undoped ultra-thin fin-body (UTB), tends to produce significant saturation-region effects that are not observed in conventional devices, for example, carrier-velocity overshoot [13], quasi-ballistic transport [10], and drain-induced charge enhancement (DICE) [12] which can affect analog performance. In this chapter, we use our physical understanding of nanoscale FinFETs, and UFDG [14], [15], [16], supplemented with published numerical-simulation and experimental data, to compare the analog/RF performance of undoped DG FinFETs with that of conventional bulk-Si MOSFETs, for both low-power and high-frequency RF applications. Further, we give insights on the optimal design of FinFETs for RF applications, and discuss the effects of scaling on analog FOMs of FinFETs.

In order to make a fair comparison of the two devices, we consider FinFETs and planar bulk-Si MOSFETs that occupy nearly equal layout areas. For FinFETs, we assume L_g is about P/4, where P is the pitch of the technology. For proper control of SCEs in DG FinFETs, t_{Si} must be about $L_g/2 \sim P/8$. Technologically, an aspect ratio of fin height (h_{Si}) to t_{Si} of 4 is reliably achievable. Therefore, $h_{Si} \sim P/2$. So, for comparison to bulk-Si MOSFETs with gate width (W) equal to P, we assume one-fin (per pitch) FinFETs with effective width $2h_{Si} = W$, which is consistent with the noted structure for SCE control.

4-2 Device Characteristics

For qualitative insight, we begin by writing simple yet physical expressions for I_{DS}, and deriving expressions for g_m and f_T. Based on these expressions we check and compare the RF performances of the DG FinFETs and bulk-Si MOSFETs. In order to simplify the discussion, we derive the expressions assuming low V_{DS}, and later we discuss their utility at high V_{DS}. We also assume strong inversion, or moderate inversion near the onset of strong inversion.

For both devices, we can write a generic expression for I_{DS} in terms of an effective width of the device (W_{eff}), equal to W for bulk-Si MOSFETs and $2h_{Si}$ for FinFETs, as

$$I_{DS} = W_{eff} Q_{ch} \upsilon \tag{4.1}$$

where υ represents the average carrier velocity in the channel and Q_{ch} here is the (aerial) inversion-charge density for bulk-Si MOSFETs, or half of it for (symmetrical) DG FinFETs. From the basic MOS equation [19], Q_{ch} (magnitude) can be expressed as

$$Q_{ch} \cong C_{ox}(V_{GS} - \Phi_{MS} - \phi_s), \tag{4.2}$$

where Φ_{MS} is the gate-body work-function difference. Note, for low V_{DS}, Q_{ch} [$\cong Q_{G(int)}$, where $Q_{G(int)}$ is the magnitude of the intrinsic charge density on the gate) will be nearly constant along the channel. From (4.1), we derive

$$g_m = W_{eff} C_{G(int)} \upsilon \tag{4.3}$$

where the intrinsic (per unit area, excluding parasitics) gate capacitance $C_{G(int)}$ is $dQ_{G(int)}/dV_{GS}$, which from (4.2) can be expressed as

$$\frac{1}{C_{G(int)}} = \frac{1}{C_{ox}} + \frac{1}{C_{inv}}, \tag{4.4}$$

in terms of the inversion-layer capacitance $C_{inv} = dQ_{ch}/d\phi_s$. [The body capacitance present in bulk devices, which is absent in FinFETs due to the two coupled gates, is neglected in (4.2) and (4.4)].

If C_p is the parasitic gate capacitance per unit width, then the total gate capacitance is $C_G = C_{G(int)}(W_{eff}L_g) + C_pW_{eff}$, and hence, with (4.3), f_T can be expressed as

$$f_T = \frac{\upsilon}{2\pi(L_g + C_p/C_{G(int)})}. \quad (4.5)$$

Note that at high V_{DS}, Q_{ch} will be a function of position in the channel, and will be different from Q_{ch} in (4.2). Typically, this V_{DS}-induced perturbation is only loosely coupled to the gate. Therefore, (4.3) and (4.5) capture the predominant dependences, even at high V_{DS}, of the two defined analog FOMs, and are hence useful for general comparison of DG FinFETs and bulk-silicon MOSFETs. Note in (4.5) how scaling L_g tends to increase f_T, as mentioned earlier, but also note the dependences on υ, $C_{G(int)}$ [or on C_{inv} in (4.4)], and C_p, which underlie performance differences between DG FinFETs and bulk-Si MOSFETs.

A. Low-Power RF Applications

CMOS devices for low-power RF applications are generally biased in the moderate-inversion region [34], where both drift and diffusion current components are important. In FinFETs, because of the high μ_{eff}, significant velocity overshoot [13] is typically observed, which leads to substantially higher carrier drift velocity. Also, because of high μ_{eff}, carriers diffuse faster in FinFETs. Therefore, for low-power RF applications, υ is significantly higher for FinFETs than for bulk-Si MOSFETs.

The inversion-charge centroid in FinFETs is farther away from the surface than in bulk-Si MOSFETs because of bulk inversion [23] in the undoped UTB, especially for moderate inversion. This leads to lower C_{inv} in FinFETs, and hence lower $C_{G(int)}$ in (4.4). The parasitic (inner and outer) fringe capacitance (C_f) and the overlap capacitance (C_{ov}) will be about the same in the two devices for equal W_{eff} (if the FinFET does not have G-S/D underlap [26], [29]), thus rendering C_p in both devices comparable. Hence, C_G in FinFETs (even without underlap) will be lower than in

bulk-Si MOSFETs. However, for FinFETs with G-S/D underlap [26], [29] (which cannot be incorporated into bulk-Si MOSFETs with doped channels), there is no C_{ov} and C_f is reduced, implying an added advantage as we discuss later.

Since the inversion-layer thickness in FinFETs, with the noted bulk inversion [23], is ~t_{ox}, and (electron) μ_{eff} is ~3x-higher than in bulk-Si MOSFETs [10], (4.3) suggests that the FinFET g_m is comparable to that of the bulk-Si MOSFET. Significant velocity overshoot [13] in FinFETs will tend to benefit g_m, however. The lower $C_{G(int)}$, with comparable g_m, implies higher f_T in (4.5) for FinFETs than for bulk-Si MOSFETs. Further, a big advantage of FinFETs over bulk-Si MOSFETs is their much lower g_{DS}. In bulk-Si MOSFETs, g_{DS} is much higher than in undoped FinFETs [35] mainly because of the presence of halo (pocket) implants in the bulk devices, which introduce V_{DS}-dependent barriers at the source and the drain [36]. For bias in the moderate-inversion region, we note that drain-induced barrier lowering (DIBL), a SCE, is important in defining g_{DS}. Therefore, even in the absence of halo implants, the g_{DS} in bulk-Si MOSFETs will tend to be higher than that in FinFETs because of better SCE control in FinFETs. The lower g_{DS}, with comparable g_m, makes A_{vo} in FinFETs larger than in bulk-Si MOSFETs.

The qualitative comparisons made above, which suggest FinFET superiority for low-power RF applications, are generally consistent with the experimental results in [35], although comparable f_T was measured for the FinFETs and bulk-Si MOSFETs examined. This inconsistency could very well be due to higher C_{ov} in the FinFETs. Since they were fabricated with undoped UTBs, but without G-S/D underlap, excessive overlap is quite likely.

We show now that the FinFET superiority can be enhanced by incorporating G-S/D underlap in the design, which is not a feasible option for bulk-Si MOSFETs [26]. As noted before, DIBL is important in defining g_{DS} in the moderate-inversion region. Therefore, an underlap, which has been shown to be effective in controlling SCEs [26], will improve the analog

performance. Figure 4.1 shows the UFDG-predicted impact of an underlap on g_{DS} at V_{GS} = 0.4V for a 28nm DG FinFET, with V_t = 0.16V at low V_{DS}. For this device, UFDG predicts that the underlap reduces DIBL from 120mV/V to 50mV/V, which translates to 75% reduction in g_{DS} at V_{DS} = 1.2V.

A G-S/D underlap can also minimize C_p significantly, and thus increase f_T in (4.5). The underlap eliminates C_{ov}, and reduces both the inner and outer C_f components [29]. Thus, even with the doubling of C_f in the DG device, C_p is lower. The underlap will increase parasitic source/drain resistance ($R_{S/D}$), but that will not be overly important in the moderate-inversion region because of the low current levels. Therefore, underlap increases f_T by reducing C_G without significantly affecting g_m. Indeed, for the FinFET in Fig. 4.1, UFDG predicts, as shown in Fig. 4.2, that the underlap reduces C_G by 28% at V_{GS} = 0.4V, while it predicts negligible deterioration in g_m. The net result is a 40% enhancement in f_T. More elaborate discussion on the effect of underlap on f_T is given in [34]. As mentioned before, experimental data presented in [35] show comparable values of f_T for sub-optimally designed FinFETs (i.e., without underlap) and bulk-Si MOSFETs. However, our results and those in [34] clearly show that a well designed G-S/D underlap in undoped DG FinFETs can lead to much higher f_T than in bulk-Si MOSFETs.

B. High-Frequency RF Applications

The relative performances for high-frequency RF applications can be assessed by comparing the FinFETs and bulk-Si MOSFETs biased where g_m is maximum. This bias point is generally in the strong-inversion region, where, unlike in moderate inversion, $R_{S/D}$ is significant. For example, it tends to reduce g_m because of reduced effective gate bias, and hence lower f_T. With similar S/D doping density, FinFETs tend to have higher $R_{S/D}$ than bulk-Si MOSFETs, even without G-S/D underlap which tends to increase it. In general, the resistance of a S/D extension region can be expressed as

$$R_{ext} = \rho_{ext}\frac{L_{ext}}{A_{ext}} \tag{4.6}$$

where ρ_{ext} is the resistivity of the extension region, and A_{ext} is its cross-sectional area. For FinFETs, $A_{ext} = h_{Si}t_{Si}$ (unless the fin is flared), and is typically less than $A_{ext} = Wx_j$ of bulk-Si MOSFETs, where x_j is the junction depth. For $t_{Si} \cong x_j$, which is reasonable, $W = 2h_{Si}$ implies that R_{ext} in FinFETs is about a factor-of-two higher than that in bulk-Si MOSFETs. Indeed, the experimental data presented in [35] show lower f_T in the FinFETs than in bulk-Si MOSFETs, which was attributed to higher $R_{S/D}$. However, the FinFET $R_{S/D}$ reported in [35] (~1000Ω-μm) is abnormally high; in fact, $R_{S/D} \cong 100$Ω-μm has been recently achieved [22].

As noted previously, FinFETs tend to have higher υ and lower $C_{G(int)}$ than bulk-Si MOSFETs, which imply higher f_T. However, for high-frequency RF applications with the devices biased in strong inversion, the higher $R_{S/D}$ of the FinFETs will undermine f_T some. We assess this effect of $R_{S/D}$ using UFDG to predict how much f_T would be increased if $R_{S/D}$ were reduced from a nominal value of 100Ω-μm to half of this value. For the 28nm DG FinFET of Figs. 4.1 and 4.2, UFDG predicts that f_T increases by 14%. We can infer then that the inherently higher $R_{S/D}$ in nanoscale FinFETs will cause only about a 15% reduction in f_T relative to that in bulk-Si MOSFETs (which is much lower than that reported in [35]). And, the reduction will be even less for longer L_g. This small effect of $R_{S/D}$ on f_T is explained by the fact that while it decreases g_m by lowering the effective gate bias, it also decreases C_G.

The effect of higher FinFET $R_{S/D}$ on A_{vo} is also not overly substantive. For the 28nm DG FinFET of Figs. 4.1 and 4.2, UFDG predicts that A_{vo} increases by about 20% when $R_{S/D}$ is reduced from its nominal value of 100Ω-μm to half of this value. As noted previously, based on [36], g_{DS} in FinFETs tends to be lower than in bulk-Si MOSFETs, by as much as an order of magnitude [35]. Therefore, even with twice as much $R_{S/D}$, A_{vo} in FinFETs should be substantially

higher than in bulk-Si MOSFETs. Unlike in the moderate-inversion region, G-S/D underlap is ineffective in reducing C_p in the strong-inversion region [29]. Also, G-S/D underlap is insignificant in defining g_{DS} because DIBL tends to subside in strong inversion. However, we noted that G-S/D underlap will tend to reduce C_G by eliminating C_{ov}. Therefore, it could be beneficial for high-frequency RF applications, even though it would increase $R_{S/D}$. As shown in Table 4.1, the 28nm DG FinFET with a G-S/D overlap of 2.8nm (10% of L_g) has a lower UFDG-predicted f_T than that of the FinFET with a G-S/D underlap optimized for low-power RF applications in accord with [34]. This G-S/D underlap introduces an extra $R_{S/D}$ of 60Ω-μm. However, elimination of C_{ov} has a more significant effect on f_T, leading to a higher value for the FinFET designed with G-S/D underlap. This demonstrates our design insight that G-S/D underlap could be beneficial for high-frequency RF applications. The design of G-S/D underlap for high-frequency RF applications will involve trading off $R_{S/D}$ for C_{ov}. The optimal design will be the one in which C_{ov} is eliminated with minimum extra $R_{S/D}$. Therefore, the design challenge will be to maximize the doping concentration in the S/D-extension region near the gate edge, without diffusing significant dopants in the UTB. This implies that the optimal G-S/D underlap for high-frequency RF applications will be smaller than that for low-power RF applications. This is shown in Table. 4.1, where the UFDG-predicted f_T of a 28nm DG FinFET designed with a short G-S/D underlap (~2nm) is higher than that of the FinFET with a G-S/D underlap (~4.5nm), which is optimal for low-power RF applications

4-3 RF FinFET Scaling

In Section 4.2 we concluded that analog/RF performance of FinFETs is superior to that of bulk-Si MOSFETs. Therefore, in this section we focus only on FinFETs, and discuss the effect of their scaling on analog FOMs. The scaling effect was discussed in [35], with FinFETs scaled from 1μm to 60nm. For FinFETs with such large gate lengths, I_{DS} is defined by Q_{ch} at the position in

channel where the velocity saturates, times the saturation velocity (v_{sat}), or times a higher effective velocity ($v_{sat(eff)} > v_{sat}$) due to overshoot [13]. But, as L_g is scaled, $v_{sat(eff)}$ increases for a given V_{DS} because of more overshoot, and this quasi-ballistic transport becomes quite significant. Ultimately then, I_{DS} becomes so high that it is limited by the maximum thermal velocity (v_{inj}) with which carriers can enter the channel from the source [10], and the transport appears ballistic, with I_{DS} defined by Q_{ch} at the (virtual) source times v_{inj}. Therefore, as we show in this section, the effect of scaling on analog FOMs of FinFETs as they are scaled to the gate lengths (<30nm) where quasi-ballistic transport becomes significant could differ substantially from that discussed in [35]. We discuss the effect of scaling on analog FOMs of FinFETs as they are scaled from 65nm to 18nm, with proper accounting for quasi-ballistic transport. UFDG predicts that the limitation of the current due to quasi-ballistic transport will onset at $L_g \sim 30$nm. Since the effect of quasi-ballistic transport is most noticeable in the strong-inversion region, we look at the analog FOMs in this bias region. We assume undoped DG FinFETs, with $t_{Si} = L_g/2$ and a pragmatic [2] SiON $t_{ox} = 1.3$nm.

UFDG predictions of the FinFET g_m versus L_g are plotted in Fig. 4.3. We see a steady increase of g_m down to $L_g \cong 30$nm, below which g_m tends to saturate. For $L_g > 30$nm, I_{DS} is not limited by v_{inj}, and the quasi-ballistic transport improves with decreasing L_g due to increasing $v_{sat(eff)}$; g_m, dependent on υ in (4.3), increases with decreasing L_g. For $L_g < 30$nm, υ at the source approaches v_{inj}, which is virtually bias-independent, and hence g_m ultimately saturates. This is not good for RF.

The UFDG predictions plotted in Fig. 4.4 show the effect of scaling L_g on the FinFET g_{DS}, with and without the ballistic current limit in the model. In general, g_{DS} increases with decreasing L_g. For increasing V_{DS}, the magnitude of Q_{ch} is reduced because of the pinch-off tendency, but this reduction is compensated by DICE [12], with the charge enhancement being comparable to

the pinch-off charge loss. Therefore, g_{DS} is predominantly determined by the increase in $v_{sat(eff)}$ with V_{DS}, and hence increases with decreasing L_g until I_{DS} reaches the ballistic limit. When I_{DS} is limited by v_{inj}, g_{DS} tends become independent of L_g; it does increase some for very short L_g because of DICE and virtually no pinch-off, however. Note in Fig. 4.4 how much g_{DS} is reduced by the effect of v_{inj}. The bias-independent nature of v_{inj} is the main reason for this reduction. This is beneficial for RF.

UFDG predictions of the FinFET A_{vo} versus scaled L_g are plotted in Fig. 4.5. These results reflect the g_m and g_{DS} trends in Figs. 4.3 and 4.4. For longer L_g where the ballistic current limit is not so important, A_{vo} decreases significantly with scaling. However, when ballistic limit sets in, because of the saturation tendencies of g_m and g_{DS}, A_{vo} virtually saturates. This is also beneficial for RF.

Finally, UFDG predictions of the effect of scaling on f_T are plotted in Fig. 4.6. In general, f_T increases with scaling, which is good for RF. When the ballistic current limit is not important, f_T increases with decreasing L_g because of both the L_g and υ dependences in (4.5), with the latter increasing with $v_{sat(eff)}$. When v_{inj} begins to limit I_{DS}, f_T increases only because of decreasing L_g, and this increase can be quite significant if the $C_p/C_{G(int)}$ ratio in (4.5) is limited. Based on our simulations, we can project that f_T > 500GHz is attainable for L_g = 18nm FinFETs.

The UFDG-predicted results in Fig. 4.3-4.6 suggest that DG FinFETs in RF applications can indeed be scaled to at least 18nm gate lengths, with very good performance projected. Below 18nm, their RF performance tends to be undermined by the quasi-ballistic transport, which also tends to limit their digital performance. We note however that the effect of quasi-ballistic transport on analog FOMs in the moderate-inversion region, for low-power RF applications, will not be as significant as in the strong-inversion region discussed herein.

4-4 Summary

Most of the literature discussing the analog/RF performance of DG FinFETs and bulk-Si MOSFETs is based on numerical-simulation and experimental results, with little emphasis on physical explanations. In this chapter, analog/RF performances of FinFETs and bulk-Si MOSFETs have been insightfully compared, and FinFETs have been shown to be superior to bulk-Si MOSFETs. We showed and discussed the beneficial effect of G-S/D underlap on analog/RF performance of FinFETs, for both low-power and high-frequency applications. We also gave insights on optimal design of G-S/D underlap. Our analysis also showed that the effect of higher $R_{S/D}$ in FinFETs, as compared to that in bulk-Si MOSFETs, does not significantly affect the analog FOMs. Finally, we showed that the undermining effect on analog FOMs of scaling FinFETs to gate lengths below ~30nm, where UFDG predicts that quasi-ballistic transport will limit the current. This trend is significantly different from that for FinFETs with longer gate lengths, for which we showed how they are superior to bulk-Si MOSFETs in RF applications. We note, therefore, that simulation and analysis of nanoscale FinFETs in RF applications must account for the quasi-ballistic transport.

Table 4-1. Comparison of UFDG-predicted f_T of an L_g = 28nm DG FinFET having a G-S/D overlap, with that of FinFETs having G-S/D underlap optimized for low-power and high-frequency RF applications.

Design	$R_{S/D}$ (Ω-μm)	f_T (GHz)
G-S/D Overlap	100	322
G-S/D Underlap for Low-Power RF	160	332
G-S/D Underlap for High-Frequency RF	120	363

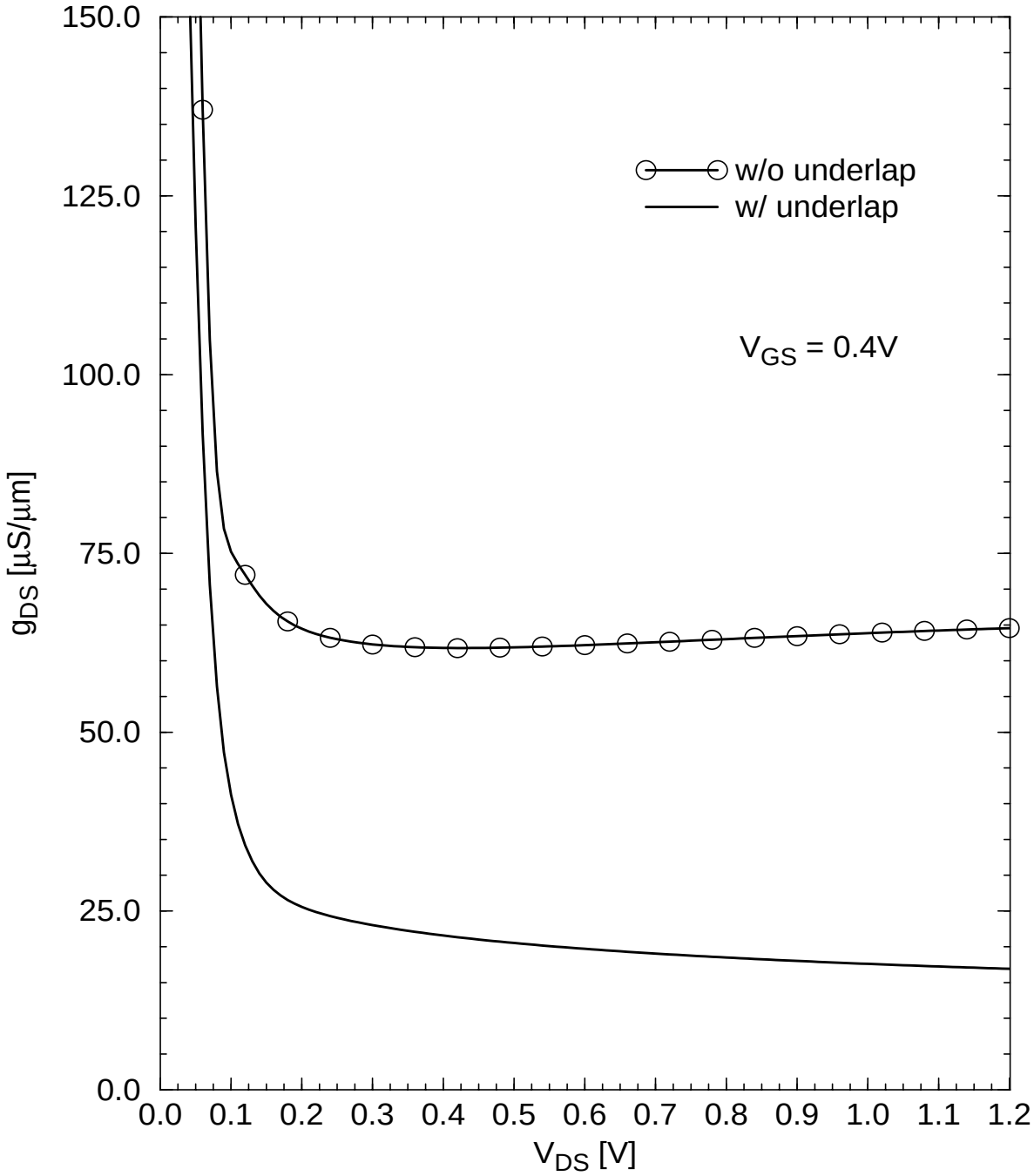


Figure 4-1. UFDG-predicted g_{DS} of a 28nm DG FinFET with t_{Si} = 14nm and t_{ox} = 1.3nm. The values have been normalized to h_{Si}. For the device with G-S/D underlap, a guassian doping profile with straggle of 11.4nm was used, and $L_{ext} = L_g$. This doping profile leads to a G-S/D underlap of about 4.5nm, which is optimal for low-power RF applications [34].

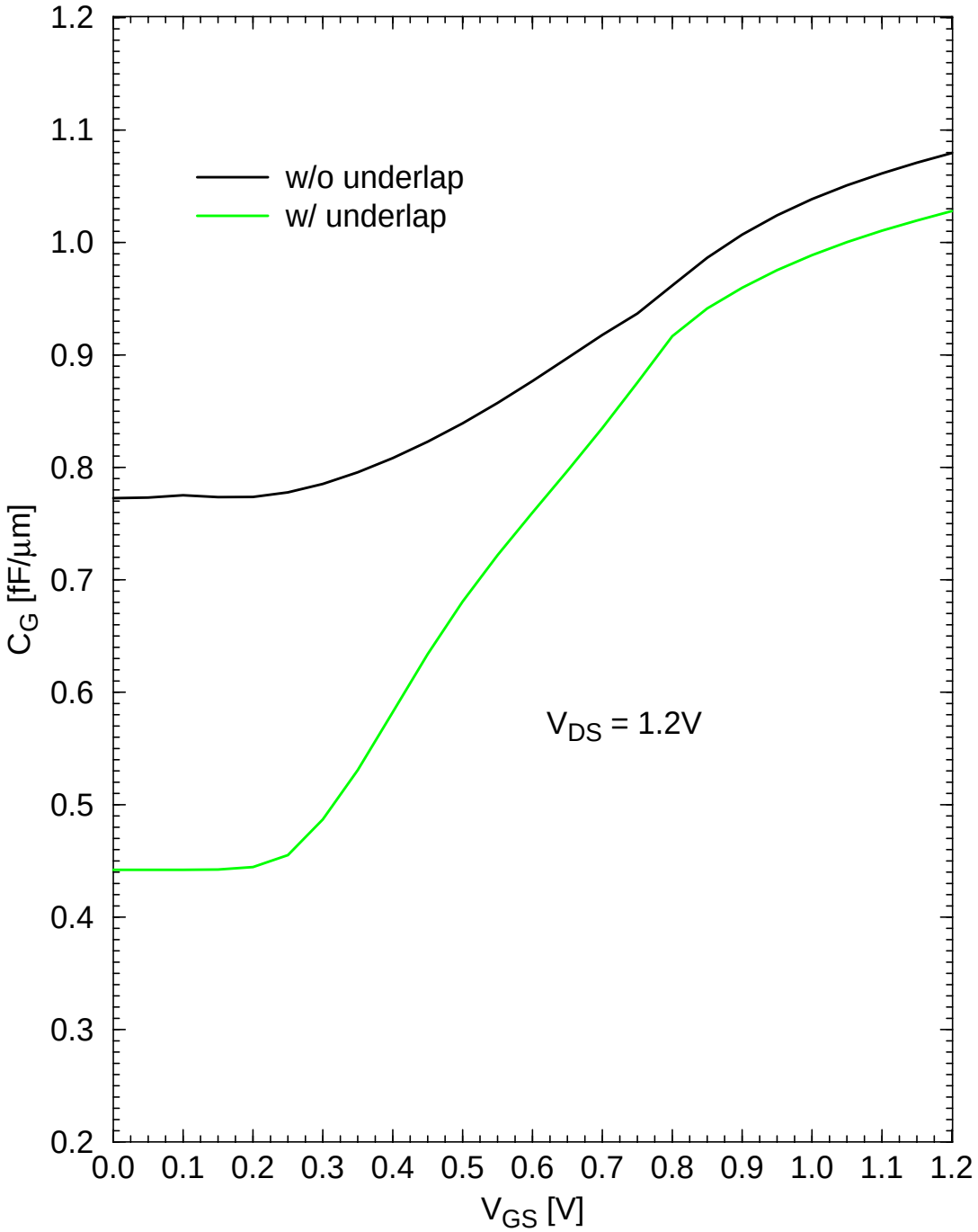


Figure 4-2. UFDG-predicted C_G (normalized to h_{Si}) versus V_{GS} at V_{DS} = 1.2V, with and without G-S/D underlap. The S/D doping profile is same as for Fig. 4.1. This S/D doping profile introduces an extra $R_{S/D}$ of about 60Ω-μm. Note the decreasing benefit of G-S/D underlap with increasing V_{GS}. The difference in the two plots at high V_{GS} is mainly due to the difference in $R_{S/D}$.

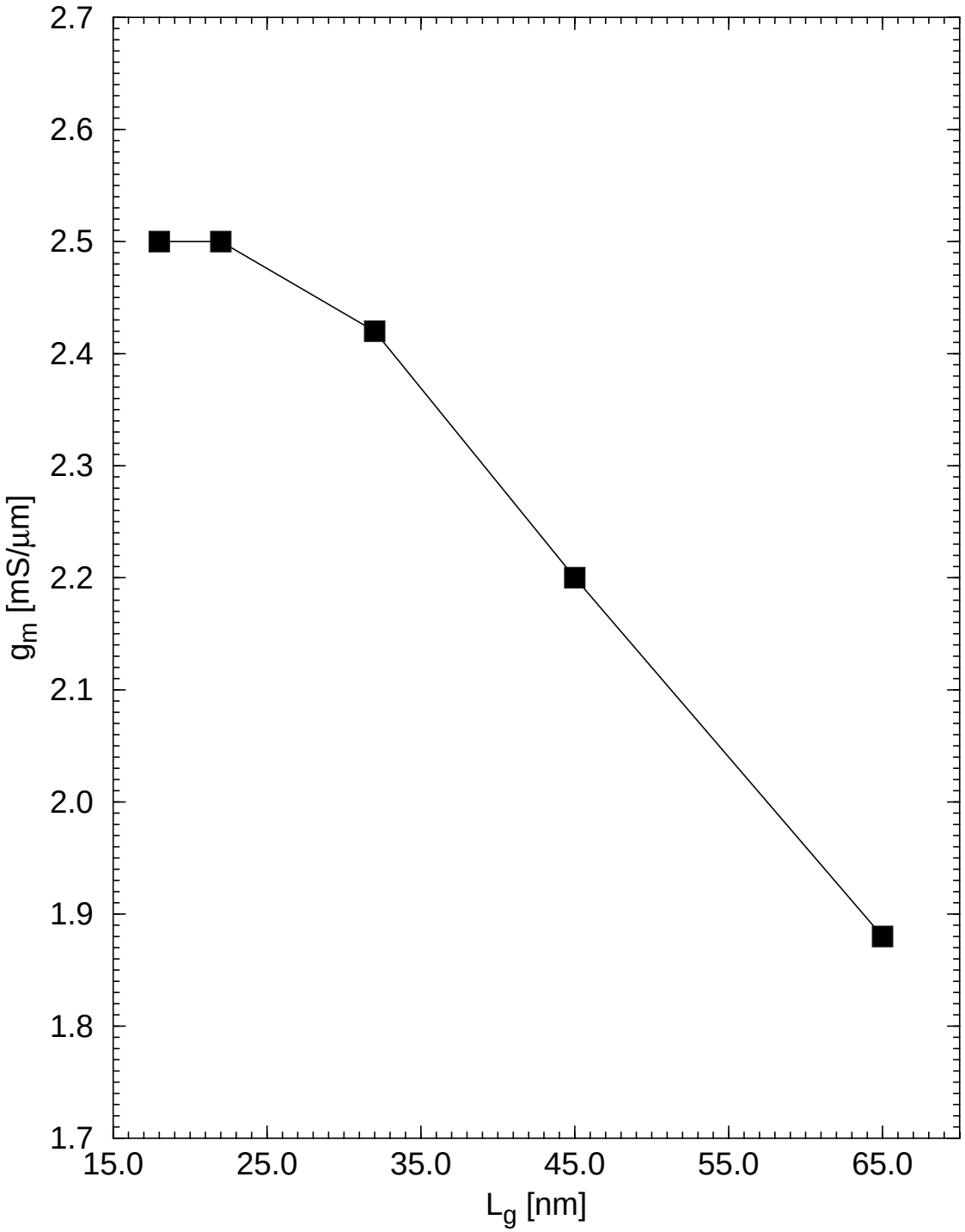


Figure 4-3. UFDG-predicted g_m versus L_g of DG FinFETs with t_{Si} = $0.5L_g$, t_{ox} = 1.3nm, and abrupt S/D doping profiles (no underlap); V_{GS} = 1.0V and V_{DS} = 1.2V.

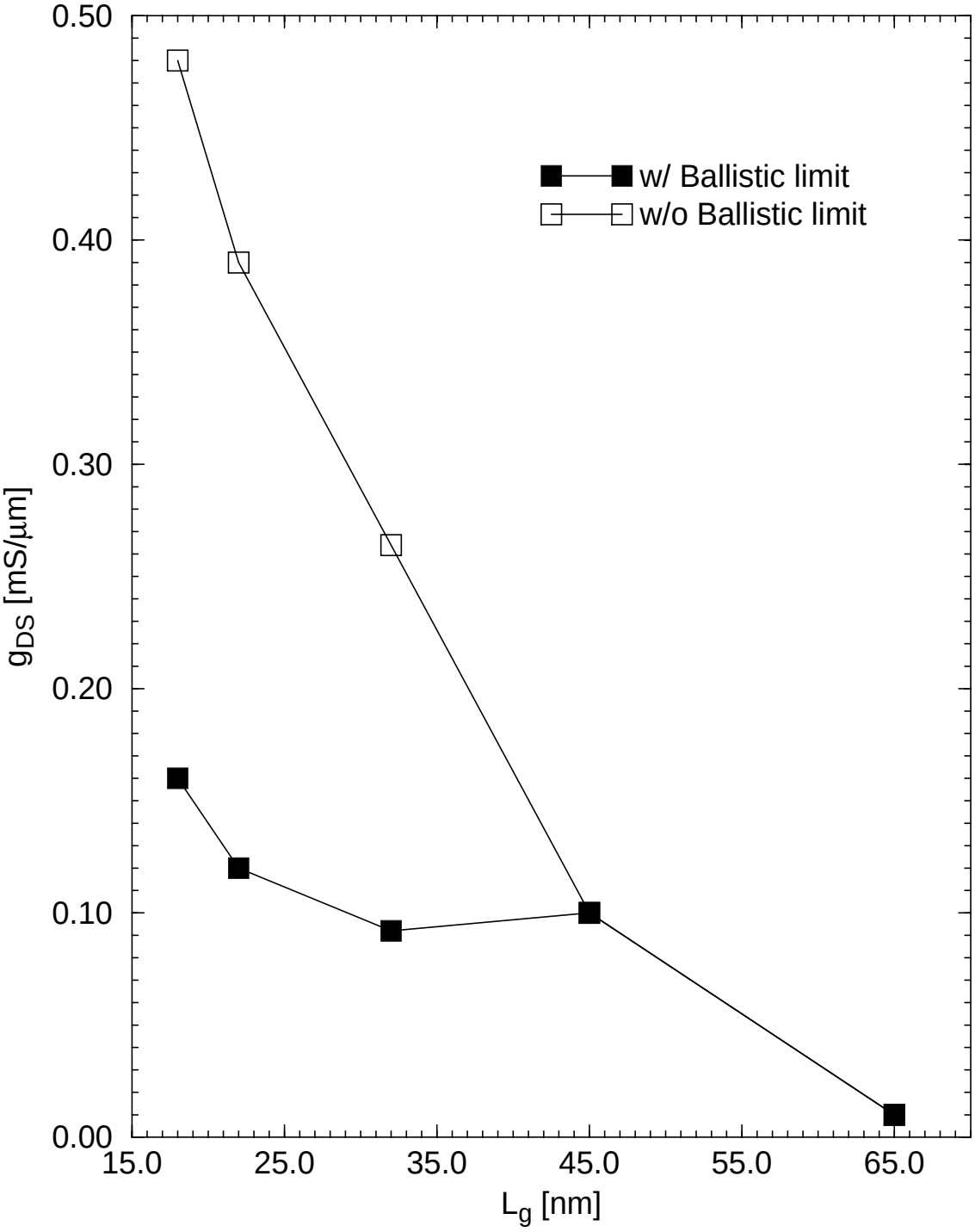


Figure 4-4. UFDF-predicted g_{DS} versus L_g for the FinFETs of Fig. 4.3, with and without consideration of quasi-ballistic limit; V_{GS} = 1.0V and V_{DS} = 1.2V.

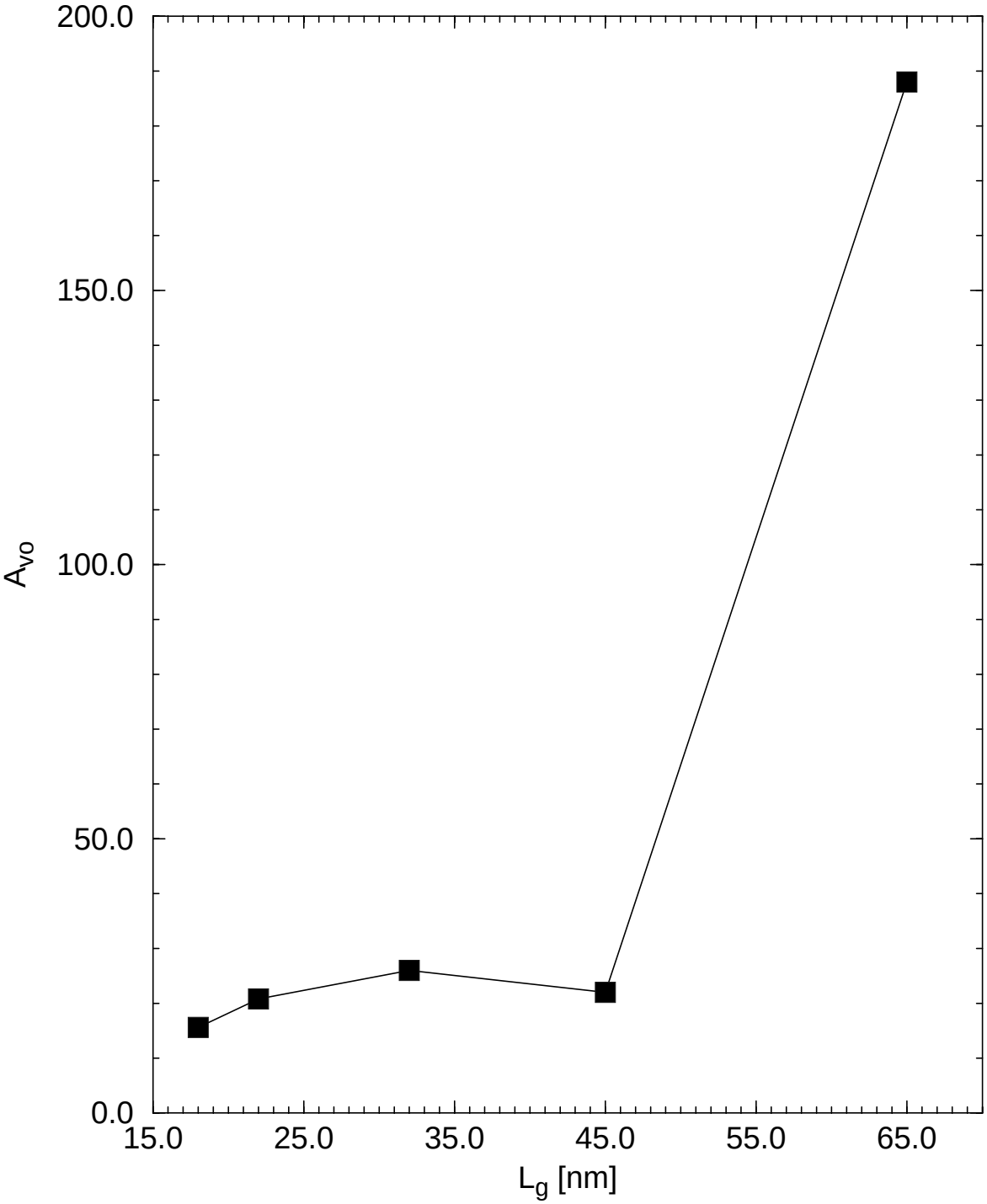


Figure 4-5. UFDG-predicted A_{vo} versus L_g for the FinFETs of Fig. 4.3; V_{GS} = 1.0V and V_{DS} = 1.2V.

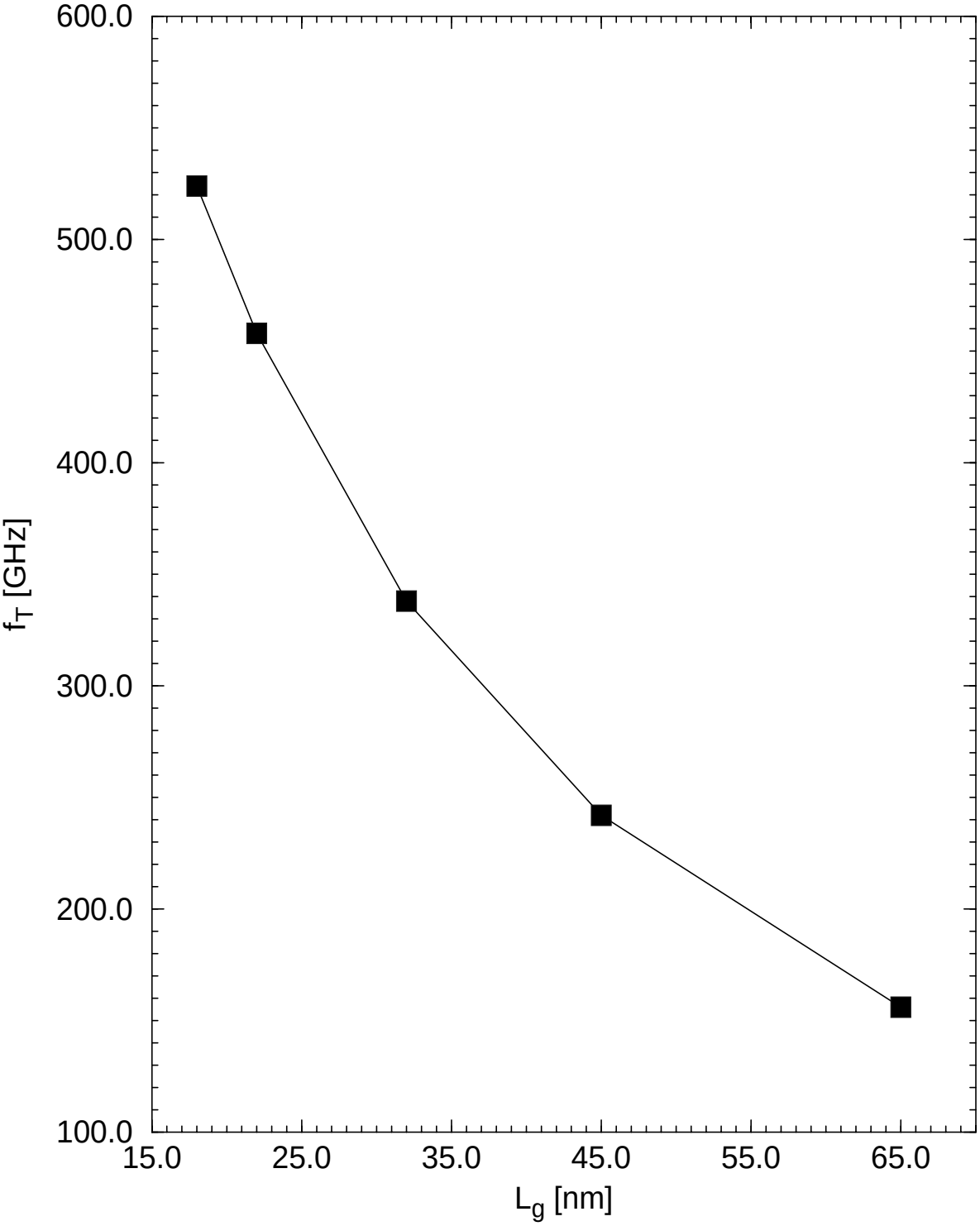


Figure 4-6. UFDG-predicted f_T versus L_g for the FinFETs of Fig. 4.3; V_{GS} = 1.0V and V_{DS} = 1.2V.

CHAPTER 5
INSIGHTS ON DESIGN AND SCALABILITY OF THIN-BOX FD/SOI CMOS

5-1 Introduction

MOSFETs with undoped, fully depleted (FD) ultra-thin bodies (UTBs) will have to replace the conventional (bulk-Si and partially depleted (PD) SOI) devices if CMOS technology is to be scaled to the end of the SIA roadmap [37] where gate lengths (L_g) are projected to be less than 10nm. No intentional doping in the UTB/channel eliminates the problem of threshold-voltage (V_t) variation, which plagues conventional nanoscale MOSFETs. Candidate FD devices include quasi-planar double-gate (DG) FinFETs [38], which we have addressed in previous chapters, and planar single-gate FD/SOI MOSFETs, which have recently evolved to thin-BOX (TBOX) structures with heavy ground-plane (GP) doping in the substrate under the device [39]-[42] as illustrated in Fig. 5.1. The GP, with proper doping (p^+ for nMOS and n^+ for pMOS), provides a back-"gate" work function which, with thin BOX, tends to accumulate the back surface of the UTB, thereby increasing the transverse electric field (E_x) and ameliorating the short-channel effects (SCEs). Also, the GP can be properly biased (negatively for nMOS and positively for pMOS) to enhance the amelioration as well as adjust V_t. Note, though, the added CMOS process and layout complexity that the selective GPs imply. Unlike conventional devices, where the channel doping density simultaneously affects SCEs and V_t of the device, it is believed that TBOX devices can be designed with independent control of SCEs and V_t; UTB thickness defines SCEs, and GP bias can be used to tune V_t [43]. However, as we will show in this chapter, this in not the case, and significantly complex processing is required to scale TBOX device toward the end of the CMOS roadmap.

An obvious design issue of the TBOX MOSFET is the high V_t implied by the increased E_x. From the 1-D Gauss's law at the front surface of an undoped-body TBOX MOSFET, we get [3]

$$V_{GfS} = \Phi_{GfS} + \left(1 + \frac{\varepsilon_{Si} t_{oxf}}{\varepsilon_{oxf} t_{Si}}\right)\phi_{sf} - \frac{\varepsilon_{Si} t_{oxf}}{\varepsilon_{oxf} t_{Si}}\phi_{sb}, \quad (5.1)$$

where ϕ_{sf} and ϕ_{sb} are the potential at the front and back surfaces, respectively, and Φ_{GfS} is the work-function difference between the front gate and the undoped body. Without SCEs (or, for long L_g) and with mid-gap gate (as commonly presumed for undoped UTBs), we can, for (hole) accumulation (in the nMOSFET) at the back surface, derive from (5.1)

$$V_{t(long)} \cong \left(1 + 6\frac{t_{oxf}}{t_{Si}}\right)\phi_c \ , \quad (5.2)$$

where $\phi_c \sim 0.4$V is the characteristic surface potential at (inversion and accumulation) threshold in an undoped body, and t_{Si} and t_{oxf} are the UTB and front-oxide thicknesses, respectively. The $V_{t(long)}$ expression in (5.2) is an approximation because we assumed $\varepsilon_{Si}/\varepsilon_{ox} \cong 3$ and $\phi_{sb} \cong -\phi_c$ for hole accumulation at the back surface, and ignored finite Fermi potential due to natural doping in the body. For $t_{Si} \sim 10$nm and $t_{oxf} \sim 1$nm, (5.2) gives $V_{t(long)} \cong 0.6$V. And, $V_{t(long)}$ will increase significantly with the level of accumulation (until it is very strong) and with decreasing t_{Si}; for $t_{Si} = 5$nm, (5.2) gives $V_{t(long)} \cong 0.9$V! The SCEs will lower V_t, but nonetheless it tends to be too high for HP applications; LP applications have thus been the recent focus of TBOX FD/SOI CMOS studies [41], [42]. We therefore first design and examine nanoscale TBOX devices for LP, i.e., low I_{off}, and then consider the possibility of HP applications. We use Taurus [44] for 2-D numerical simulations, and physically interpret results, with support from our physics-based compact model UFDG [16], to gain insights on design and scalability.

5-2 LP Devices

For later comparisons, we first design and simulate a TBOX FD/SOI nMOSFET that can be related to recent measured and simulated LP-device results in [41] and [42]. For $L_g = 25$nm (=L_{eff}

due to assumed abrupt source/drain junctions without gate underlap/overlap), we assume t_{oxf} = 1.2nm (=EOT), t_{BOX} = 10nm, unbiased (grounded) p^+ GP, midgap gate, and undoped UTB with thickness (t_{Si}) tuned to get adequate SCE control, or DIBL ≅ 100mV/V for V_{DD} = 1.0V. Our Taurus simulations show that t_{Si} = 6nm (=L_g/4.2) is needed, which is significantly more aggressive UTB scaling than that (~L_g/3) specified in [41] and [42]. (Our assumption of abrupt S/D junctions leads to worst-case SCEs for a given L_{eff}, but it is not the reason for this inconsistency.) Predicted device characteristics are given in Table 5.1, where they are compared with results for the same nMOSFET without the GP (DIBL = 120mV/V) and with thick BOX (DIBL = 220mV/V). Note the benefit of the thin BOX even without the GP. The significant reduction in DIBL yielded by the thin BOX alone is due to, first, better control of the electric potential in the UTB by the (grounded) substrate, or back "gate," second, the enhanced E_x, which tends to move the inversion-charge centroid towards the (front) gate [3], [45], and, third, amelioration of the electric field-fringing effect in the thin BOX [45], [46]. The reduced DIBL lowers I_{off}, but the dramatic decrease of I_{off} shown in Table 5.1 reflects a significant increase of $V_{t(long)}$ in (5.2) due to the thin BOX as well.

(In our Taurus simulations, we are neglecting quantization, the effect of which on V_t is, for thick BOX and reasonable t_{Si} > 4nm, indeed negligible [28] due to low E_x. For TBOX devices with (unbiased or biased) GP, E_x is higher for long L_g, but for nanoscale L_g, E_x is reduced by 2-D effects, as shown by Taurus, such that the quantization effect on V_t is still small.)

As mentioned, the thin t_{BOX} = 10nm ameliorates the field fringing in the BOX, but it does not completely suppress its effects. In Fig. 5.2 we show Taurus- and UFDG-predicted weak-inversion current-voltage characteristics of an L_g = 30nm TBOX nMOSFET, with t_{BOX} = 10nm and t_{Si} = 6nm. The UFDG results, with BOX field-fringing modeling [46] tuned to the Taurus results, and without the field-fringing modeling activated, clearly show significant degradation of

the subthreshold characteristics, or SCEs, due to the fringing even for 10nm BOX. (We thank Shishir Agrawal, a post-doctoral student at the University of Florida, for doing the simulation in Fig. 5.2.)

Contrary to expectation [39]-[42], our simulation results in Table 5.1 (for the nMOSFET) show that the addition of the unbiased GP only slightly reduces DIBL and I_{off} further. The p^+-GP benefit is small because it only increases the back-gate work function by ~200mV (for a typical p-substrate), which does not increase E_x much for t_{BOX} = 10nm. However, for the TBOX pMOSFET, the n^+ GP will decrease the back-gate workfunction by ~700mV, which will impact the device characteristics much more. Simulations show that for the TBOX pMOS counterpart to the nMOSFET in Table 5.1, the GP reduces DIBL from 140mV/V to 120mV/V, but decreases I_{off} by a factor of ~100. We note that the effect of the higher E_x is negligible with regard to DIBL because of the thin t_{Si}, but is dramatic with regard to I_{off} via the much higher $V_{t(long)}$. To illustrate this point further, we show in Fig. 5.3 Taurus-predicted weak-inversion current-voltage characteristics of TBOX CMOS devices, with and without GPs. We thank Shishir Agrawal for doing this simulation. Note that the GP effect (especially the increased V_t) is dramatic for the pMOSFET, but much less significant for the nMOSFET. We thus might question whether the (unbiased) GP, with its process complexity, is worthwhile for the TBOX nMOSFET. (We show later that it is needed, with bias, for scalability, however.) Even with it, the predicted I_{off} ~ 1000pA/μm is much too high for LP, even though the SCEs are controlled; Ironically; V_t is too low.

Our simulation results in Table 5.1 are in general accord with measured DIBL and I_{off} vs. L_g and t_{Si} for thick-BOX nanoscale FD/SOI MOSFETs in [47]. The inconsistency of our results with those in [41] and [42], regarding SCE control and LP design, could be due to G-S/D underlap [26] in the devices examined in [41] and [42], which renders $L_{eff} > L_g$ and yields better SCE

control (and which is essential for nanoscale UTB-device design [48]). To check, we redesign our L_g = 25nm thin-BOX/GP nMOSFET with underlap; we increase t_{Si} = 8nm (~L_g/3 in line with [41] and [42]), and now tune the underlap (L_{eSD}) to get DIBL = 100mV/V. Our simulations reveal that L_{eSD} = 2.5nm is needed, meaning $L_{eff} = L_g + 2L_{eSD}$ = 30nm. Part of the inconsistency is possibly explained, but our predicted V_t is still too low and I_{off} is too high for LP.

To get an acceptable L_g = 25nm LP design, we must decrease I_{off} to ~10pA/μm, irrespective of DIBL. Because of the lesser effect of the GP in the nMOSFET, the design of this device is more demanding, and hence we focus on it. We include the GP because biasing it is one way of decreasing I_{off}. Another way is thinning t_{Si}, as implied by (5.2), which we check first. We begin with the device having the 2.5nm underlap, and we now tune t_{Si} to get acceptable V_t and I_{off} at V_{DD} = 1.0V. We find that t_{Si} = 6nm (=L_g/4.2 = L_{eff}/5) is needed, and we get $I_{off} \cong$ 30pA/μm. We note that the predominant effect of thinning t_{Si} here is to increase $V_{t(long)}$. Interestingly, we get DIBL = 60mV/V for this device, meaning that DIBL alone is not an acceptable design criterion for LP thin-BOX FD/SOI MOSFETs. This new insight is reflected in Table 5.2, where we give Taurus-predicted I_{off} and DIBL for our LP nMOSFET, compared with predictions for variable t_{Si}. Note that thicker t_{Si} = 7nm, which yields $I_{off} \cong$ 100pA/μm and DIBL = 80mV/V, could be an acceptable LP device as well. But also note that the designs of these LP TBOX/GP MOSFETs are much more stringent (i.e., much lower t_{Si}/L_g ~ 1/4 is needed) than suggested in [41] and [42].

Perhaps a GP bias (for which the selective p^+ and n^+ GPs are necessary in the CMOS technology), which increases E_x, can loosen the design criteria (i.e., enable use of thicker t_{Si}) and give a more optimistic scaling outlook, albeit with added technology and layout complexity. Focusing on I_{off} as the main LP design criterion, we give simulation results in Table 5.3 derived by tuning t_{Si} to get I_{off} ~ 10pA/μm for GP bias (V_{GP}) ranging from 0V (as in Table 5.2) to -5.0V, which ensures strong back-surface accumulation and thus maximizes E_x. Indeed, the t_{Si}

requirement can be relaxed significantly with GP bias. For the strong-accumulation case (which requires a sizeable V_{GP} magnitude), a good LP design is achieved with t_{Si} = 15nm (=L_g/1.7 = L_{eff}/2 with the 2.5nm underlap), as opposed to 6nm for no GP bias. Interestingly, note in Table 5.3 the anti-correlation of I_{off} and DIBL for decreasing V_{GP}; as we relax the t_{Si} requirement for I_{off} control with decreasing V_{GP}, DIBL increases whereas I_{off} decreases. This surprising result is due to the fact that $V_{t(long)}$ is being increased, while the SCEs worsen, with the former effect being predominant with respect to I_{off}. Note that decreasing V_{GP} will tend to increase I_{off} by increasing junction tunneling at the back surface, but the effect of increase in $V_{t(long)}$ due increase in E_x predominates, and therefore, there is a net decrease in I_{off} with decreasing V_{GP}.

The encouraging results in Table 5.3 concerning use of GP bias suggest that, with it, TBOX FD/SOI CMOS may be scaled considerably. We check the scalability, in terms of L_{eff} for devices with underlap, by trying to tune t_{Si} for I_{off} ~ 10pA/μm, with acceptable DIBL, as L_{eff} is shortened (from 30nm as in Table 5.3) and V_{GP} is set to ensure strong accumulation. (This V_{GP} will tend to become more negative with decreasing L_{eff} due to increase in E_x, as defined by decreasing t_{Si} required for SCE control.) We still let t_{oxf} = 1.2nm and t_{BOX} = 10nm. The simulation results are given in Table 5.4. In line with the above note about the importance of $V_{t(long)}$ vs. t_{Si}, we find that V_t becomes too high when DIBL ≅ 100mV/V, leading to extremely low I_{off} and reflecting poor I_{on}. Gate work-function tuning to decrease V_t is thus called for, which adds even more complexity to the fabrication process. We include in Table 5.4 the decrease in the work function ($\Delta\Phi_{Gf}$ below midgap) needed to increase I_{off} to ~10pA/μm, and thereby get acceptable I_{on} for LP performance. Further, we stop the scaling when t_{Si} reaches ~5nm, which thus defines the scaling limit of L_{eff} ≅ 18nm for LP thin-BOX/GP (with sizeable V_{GP}) CMOS; the needed gate work function is near conduction band-edge for the n-channel device. The assumed minimum SOI thickness is based on the quantization effect [28], which becomes prohibitively severe for thinner t_{Si}, as well as the SOI

technology. Additional simulations reveal that without GP bias, the LP scaling limit, set by t_{Si} = 5nm with a midgap gate, is L_{eff} = 28nm.

From (5.1), the V_t of a TBOX MOSFET can be expressed as

$$V_t \cong \Phi_{GfS} + \phi_{sf}^{t} + 3\frac{t_{oxf}}{t_{Si}}\left(\phi_{sf}^{t} - \phi_{sb}^{t}\right), \tag{5.3}$$

where ϕ_{sf}^{t} and ϕ_{sb}^{t} are the potential at the front and back surfaces at threshold, respectively. Unlike in the derivation of the simple expression for $V_{t(long)}$ in (5.2), here we do a more thorough analysis by considering the finite Fermi potential (ϕ_F) due to natural p-type doping (~10^{15}cm^{-3}) in the unintentionally doped UTB [3]. We assume for the nMOSFET with strong back-surface accumulation, $\phi_{sf}^{t} = \phi_c + \phi_F$ and $\phi_{sb}^{t} = -\phi_a + \phi_F$, where $\phi_a \cong \phi_c$ is the strong-accumulation counterpart to ϕ_c. Similarly for the pMOSFET with strong back-surface accumulation, we assume $\phi_{sf}^{t} = -\phi_c + \phi_F$ and $\phi_{sb}^{t} = \phi_a + \phi_F$. With the above ϕ_{sf}^{t} and ϕ_{sb}^{t} for nMOSFET and pMOSFET, and noting that $\Phi_{GfS} = -\phi_F$ for a midgap gate, we find that the finite ϕ_F does not affect V_t when the back surface is accumulated, and so the CMOS devices are symmetric in this regard. Therefore, the TBOX pMOSFET for LP applications can be designed and scaled in a manner similar to that of the TBOX nMOSFET.

5-3 HP Devices

To check the HP application, and its scalability, we first presume that V_{GP} must be negative enough for strong accumulation to ensure SCE control. We begin with the L_{eff} = 30nm nMOSFET in Table 5.4, and tune $\Delta\Phi_{Gf}$ to increase I_{off} to an acceptable ~100nA/μm, which implies acceptable I_{on} (if the external S/D series resistance in the UTB device is adequately limited [47]). We then scale L_{eff}, tuning t_{Si} and $\Delta\Phi_{Gf}$ to get DIBL ≅ 100mV/V as well as I_{off} ~ 100nA/μm. The simulation results, in Table 5.5, are discouraging. They show, for a specific L_{eff}, that the $\Delta\Phi_{Gf}$ required to tune I_{off} for HP is much larger than that required for LP, being near conduction band-

edge for L_{eff} = 25nm and perhaps being impossible for L_{eff} = 18nm. The required $\Delta\Phi_{Gf}$ increases with decreasing L_{eff} due to the increase in $V_{t(long)}$ in (5.2) defined by the thinner t_{Si} needed for SCE control. Thus, as indicated in Table 5.5, the HP scaling limit could be longer than the L_{eff} = 18nm defined by t_{Si} = 5nm, where an unacceptably high [49] $\Delta\Phi_{Gf}$ = 850mV would be necessary. Without GP bias (V_{GP} = 0V), the HP scalability defined by t_{Si} = 5nm is L_{eff} = 25nm, with a required $\Delta\Phi_{Gf}$ = 200mV. These results then suggest that, for the nMOSFET, the realistic scaling limit is reachable even without a GP bias, and thus even without the GP, as we have intimated, but is not close to the end of the ITRS roadmap.

The same analysis of the TBOX pMOSFET for HP applications leads to a similar no-GP design. Its SCEs will be virtually the same as in the no-GP nMOSFET, and hence its scaling limit is also $L_{eff} \cong$ 25nm. In a TBOX without GP bias (and no GP) and with depleted back surface, from the 1-D analysis in [3] we get

$$\phi_{sb}^{t} = -\left(\frac{C_b}{C_b + C_{oxb}}\right)\phi_{sf}^{t}, \tag{5.4}$$

where $C_b = \varepsilon_{Si}/t_{Si}$ and $C_{oxb} = \varepsilon_{ox}/t_{oxb}$. Assuming that $\phi_{sf}^{t} = \phi_c + \phi_F$ for the nMOSFET, and $\phi_{sf}^{t} = -\phi_c + \phi_F$ for the pMOSFET, with ϕ_{sb}^{t} defined by (5.4), we get from (5.3)

$$V_{tp} = -V_{tn} + 2r\phi_F, \tag{5.5}$$

where V_{tn} and V_{tp} are the threshold voltages of the TBOX HP nMOSFET and pMOSFET, respectively, and r is the body factor [3]. Therefore, TBOX CMOS for the HP application is not symmetric. For the TBOX pMOSFET, the needed $\Delta\Phi_{Gf}$ due to the ϕ_F term in (5.5) is negative (above midgap), and its magnitude is about 50mV (in worst case, as implied by (5.5)) less than that for the nMOSFET.

5-4 Comparisons with FinFETs

In this section we compare potential performances of nanoscale thin-BOX FD/SOI MOSFETs and DG FinFETs in order to gain more insights on the viability of the former. Since, as discussed, the performance and design challenges of the TBOX nMOSFET and pMOSFET are similar, we focus on the n-channel device. In Table 5.6, we summarize our projected scaling limits (in terms of $L_{eff} = L_g + 2L_{eSD}$) of thin-BOX/GP nMOSFETs for LP and HP CMOS as defined by t_{Si} = 5nm, and compare them with Taurus predictions for DG nFinFETs with the same t_{Si}, or fin thickness. The scalability of the thin-BOX/GP device with no V_{GP} and midgap gate is worst of all. Negative V_{GP} for strong back accumulation tends to improve it, but also brings in the requirement of tuned $\Delta\Phi_{Gf}$. For LP applications, we project a scaling limit L_{eff} = 18nm, which, with G-S/D underlap, implies $L_g \sim$ 10nm. The projected HP scaling limit of $L_{eff} \cong$ 18nm in Table 5.6 is questionable because of the very large $\Delta\Phi_{Gf}$ required to get acceptable I_{off} and I_{on}, as noted before. A scaling limit of $L_{eff} \cong$ 25nm, with t_{Si} = 10nm, is probably more realistic; as shown in Table 5.5, it requires a near conduction band-edge gate work function. Coincidently, as shown in Table 5.6, the projected HP scaling limit without V_{GP} for t_{Si} = 5nm is also 25nm. Therefore, for HP thin-BOX FD/SOI CMOS, V_{GP} seems unnecessary, and thus so do the GPs, but the scalability is not good.

DG FinFETs yield the most relaxed t_{Si} requirement for SCE control due to the two gates [48]. Therefore, for HP applications where the scalability is limited by SCE requirements, the FinFET, with midgap gate, is most scalable. The $L_{eff} \cong$ 15nm limit noted in Table 5.6 implies, with G-S/D underlap [26], an L_g scaling limit near the end of the ITRS. Note that our simulation results in Table 5.6 indicate that near the scaling limit, for adequate SCE control in the DG FinFET, $t_{Si}/L_{eff} \cong 1/3$ will be needed, which is smaller than generally presumed $t_{Si}/L_{eff} \cong 1/2$ [50]. This is because t_{oxf} is not being scaled due to gate-leakage considerations. For LP applications,

the scalability of the DG FinFET, with midgap gate, is limited by the I_{off} requirement. Therefore, it is less scalable than the thin-BOX/GP MOSFET with V_{GP}, for which the noted work-function tuning is necessary to get acceptable I_{off}. Of course, such tuning could be used for the FinFET as well. With a near valence band-edge gate, as shown in Table 5.6, the DG nFinFET could be scaled to $L_{eff} \cong 15nm$ ($L_g < 10nm$), limited by the SCE requirements, rendering it most scalable for LP applications also.

The comparison of the two FD devices must involve more. The negative V_{GP} applied to the thin-BOX/GP nMOSFET increases the confinement of electrons towards the front surface, leading to higher inversion-layer capacitance (C_i) than in the DG FinFET, which is subject to significant bulk inversion [48]. This difference is reflected in Fig. 5.4, where we show a comparison of Taurus-predicted low-V_{DS} strong-inversion currents in the L_{eff} = 30nm thin-BOX/GP nMOSFET in Table 5.5 and in an L_{eff} = 30nm DG nFinFET with t_{Si} (=15nm) tuned to get similar DIBL and $\Delta\Phi_{Gf}$ (=60mV) tuned to get similar I_{off}. The simulation results in Fig. 5.4 imply, at V_{GS} = 1.0V, that higher C_i in the thin-BOX/GP MOSFET leads to 17% higher inversion charge density ($Q_i = qN_{inv}$) than that supported by each gate in the DG FinFET. This translates to less than 2x (1.7x for low V_{DS}) current in the FinFET relative to that in the FD/SOI MOSFET. (The external S/D series resistance was kept low in the simulations in order to avoid any discrepancy in effective biases due to different current levels, and a constant μ_{eff} model was used so that the difference in currents reflected directly the different Q_i and C_i.)

The higher C_i due to E_x is a beneficial effect for the thin-BOX/GP MOSFET. However, E_x also implies lower μ_{eff} due to more surface-roughness scattering, which tends to negate the higher Q_i with regard to I_{on}. For a typical N_{inv} in strong inversion, the surface electric field in the thin-BOX/GP MOSFET is higher than that in the DG FinFET not only because of the GP, but also because the entire inversion charge is supported by one gate, as opposed to two gates in the

FinFET. Thus, strong-inversion carrier mobility can be higher in the FinFET by a factor of two, or more [48]. Further, the high E_x also underlies an enhanced quantization effect in strong inversion, which lowers N_{inv} at a given V_{GS}. Due to thin box, source/drain-substrate capacitance could be significant in TBOX CMOS, and, unlike in DG FinFETs, there is a non-zero (effective) body capacitance ($C_{b(eff)}$) associated with TBOX MOSFETs [3], which tends to increase the subthreshold slope (S) and I_{off} (although the SCEs usually define S), as well as undermine CMOS speed. Also, heavily doped GP may lead to significant depletion in the S/D extension regions which would increase the series resistance and accentuate SCEs [51]. However, TBOX MOSFETs may have some advantage over FinFETs because of their planar structure; e.g., controlling UTB thickness in the TBOX device is easier than in FinFETs, which implies that yield of TBOX-based SRAM is higher than that of FinFET-based SRAM [52]-[53].

5-5 Conclusions

We have used numerical device simulations, with physics-based interpretations of results, to design and study thin-BOX FD/SOI CMOS, and to assess its scalability, for both LP and HP applications, relative to that of DG-FinFET CMOS. For LP, we found that both I_{off} and DIBL must be considered as design criteria, and that they are not necessarily correlated. We found that, with the complex processing and layout (due to selective, sizable GP biasing and tunable gate work function for different L_g), the LP thin-BOX/GP MOSFET scalability, with G-S/D underlap ($L_{eff} > L_g$), is $L_g \sim 10$nm; whereas, with gate work-function tuning and G-S/D underlap, the LP DG FinFET can be scaled to the end of the SIA roadmap where $L_g < 10$nm is projected. For HP applications, the DG FinFET is clearly the winner in terms of scalability. Our simulation results show that without any required work-function tuning, i.e., with a midgap gate, the HP DG FinFET can also be scaled to the end of the roadmap; whereas, because of implausible work-function engineering required, the scalability of the HP thin-BOX MOSFET is severely limited, although

the GP can be eliminated and the processing significantly simplified.

Thin-BOX FD/SOI MOSFETs have the advantage of being planar and, therefore, similar to conventional PD/SOI devices in terms of processing. However, the added process complexities noted herein, with less potential HP scalability than DG FinFETs, make them less attractive. Hence, thin-BOX FD/SOI CMOS could be a viable interim technology, bridging conventional CMOS and DG-FinFET CMOS. The latter technology, which can be pragmatically designed [48], is potentially scalable to the end of the SIA roadmap for both HP and LP applications.

Table 5-1. Taurus-predicted characteristics of L_g = 25nm (= L_{eff}) FD/SOI nMOSFETs with midgap gate and t_{Si} = 6nm.

Design	DIBL (mV/V)	I_{off} (pA/μm)
Thin BOX w/ GP	100	10^3
Thin BOX	120	3×10^3
Thick BOX (200nm)	220	10^6

Table 5-2. Taurus-predicted characteristics, vs. t_{Si}, of L_g = 25nm TBOX/GP nMOSFETs with 2.5nm G-S/D underlap and midgap gate.

t_{Si} (nm)	DIBL (mV/V)	I_{off} (pA/μm)
6	60	30
7	80	100
8	100	1000

Table 5-3. Taurus-predicted characteristics, vs. V_{GP}, of L_g = 25nm TBOX GP nMOSFETs with 2.5nm G-S/D underlap and midgap gate.

V_{GP} (V)	t_{Si} (nm)	DIBL (mV/V)	I_{off} (pA/μm)
0	6	60	30
-1.0	9	90	20
-5.0	15	100	10

Table 5-4. Taurus-predicted characteristics, vs. L_{eff}, of TBOX/GP nMOSFETs with V_{GP} for strong accumulation and controlled DIBL. The work-function reduction below midgap required to increase I_{off} as shown to ~10pA/μm for feasible LP is given.

L_{eff} (nm)	t_{Si} (nm)	I_{off} (pA/μm)	$\Delta\Phi_{Gf}$ (mV)
30	15	10	0
25	10	1	100
l8	5	10^{-3}	450

Table 5-5. Taurus-predicted characteristics, vs. L_{eff}, of TBOX/GP nMOSFETs with V_{GP} for strong accumulation. The work-function reduction below midgap required to increase I_{off} to ~100nA/μm for acceptable I_{on} and viable HP is given.

L_{eff} (nm)	t_{Si} (nm)	DIBL (mV/V)	$\Delta\Phi_{Gf}$ (mV)
30	15	100	380
25	10	100	500
l8	5	110	850

Table 5-6. Taurus-predicted LP and HP scaling limits (L_{eff}, which, with G-S/D underlap, can be 5-10nm longer than L_g), defined by t_{Si} = 5nm, for thin-BOX/GP nMOSFETs (t_{BOX} = 10nm) and DG nFinFETs (all with t_{oxf} = 1.2nm). The devices have been designed for I_{off} ~ 10pA/μm and ~100nA/μm for LP and HP applications, respectively, with DIBL ≤ 100mV/V. The 18nm limit for the HP thin-BOX/GP device with V_{GP} (for strong accumulation) is questionable due to the very large $\Delta\Phi_{Gf}$ needed; the scaling limit of 25nm without V_{GP} (and without GP) is more realistic and pragmatic.

LP	Thin-BOX/GP w/o V_{GP}	Thin-BOX/GP w/ V_{GP}	DG FinFET
L_{eff} (nm)	28	18	25/15
$\Delta\Phi_{Gf}$ (mV)	0	450	0/-450
HP	**Thin-BOX/GP w/o V_{GP}**	**Thin-BOX/GP w/ V_{GP}**	**DG FinFET**
L_{eff} (nm)	25	18?	15
$\Delta\Phi_{Gf}$ (mV)	200	850	0

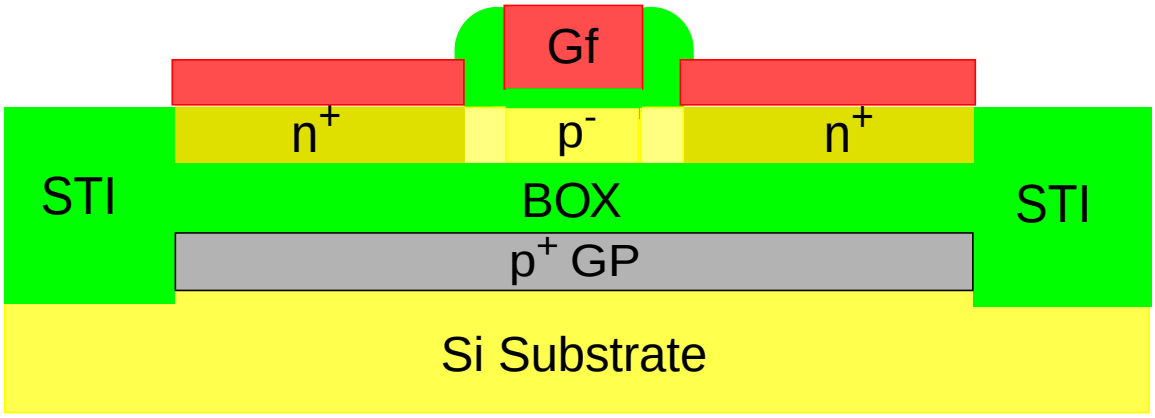


Figure 5-1. Basic thin-BOX FD/SOI nMOSFET structure, with P^+ GP.

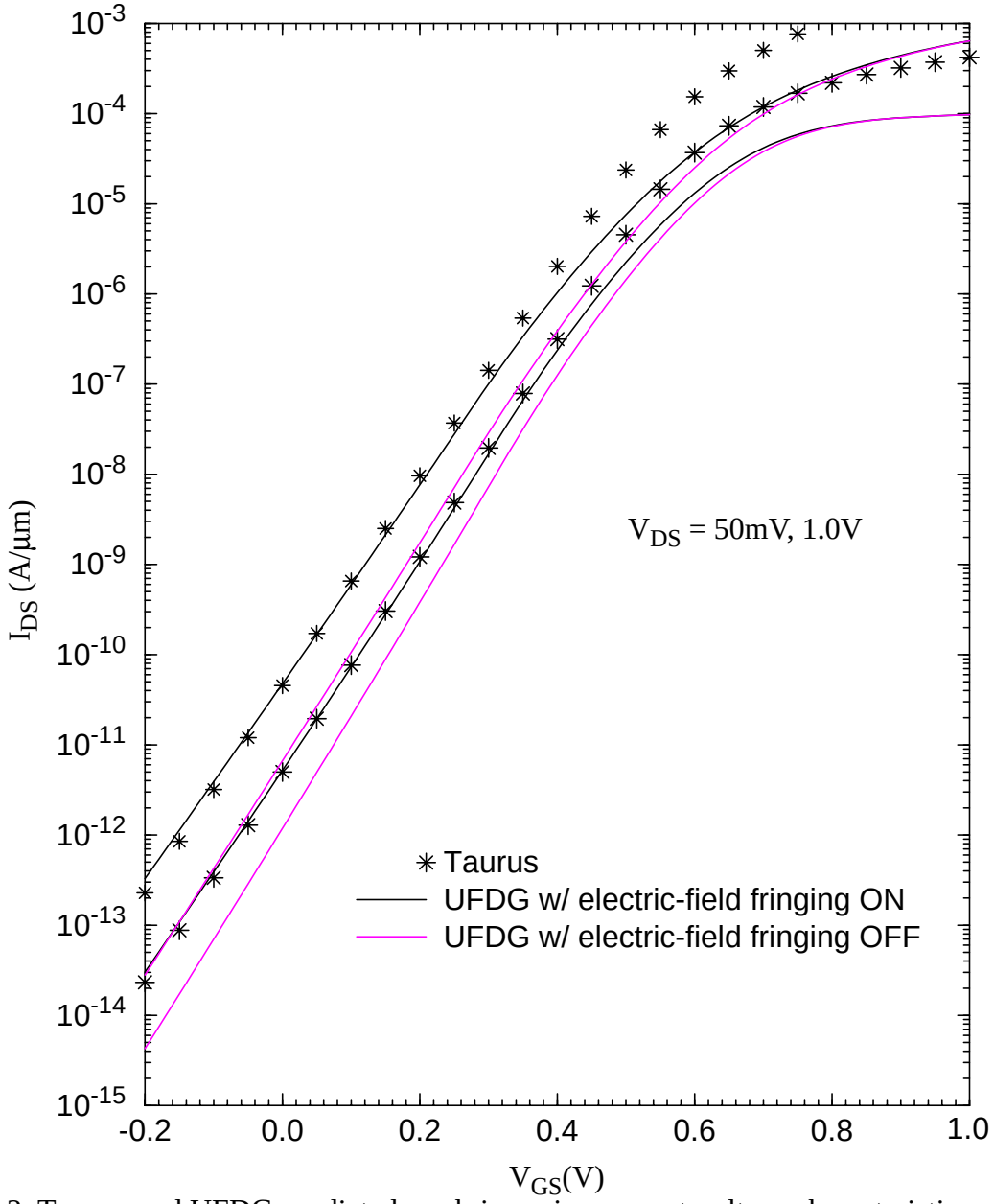
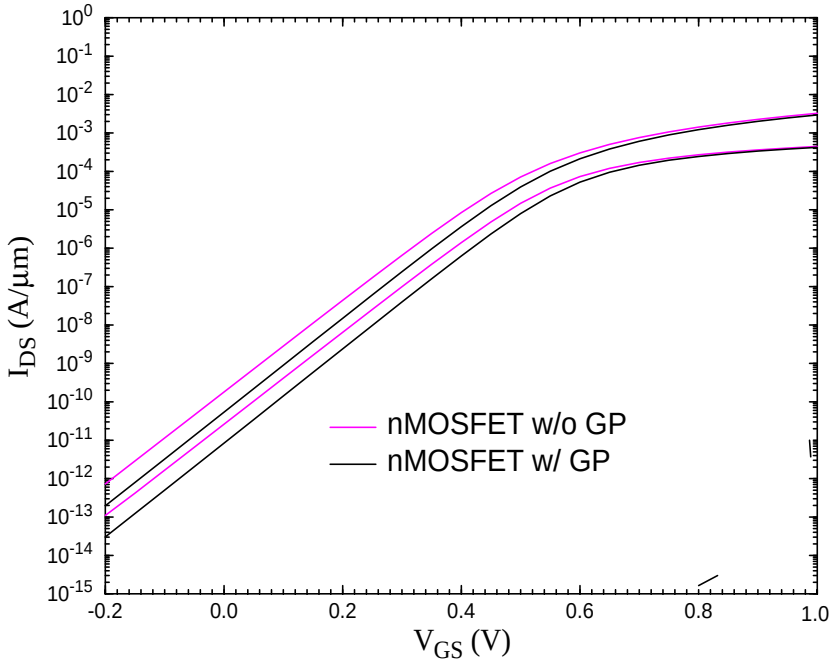


Figure 5-2. Taurus- and UFDG-predicted weak-inversion current-voltage characteristics of a TBOX FD/SOI nMOSFET (L_g = 30nm, t_{Si} = 6nm, t_{BOX} = 10nm, no GP). The UFDG prediction, with the electric-field fringing parameters tuned to fit Taurus data, is contrasted to a UFDG prediction with the field fringing modeling deactivated. A constant electron mobility (μ_n=300cm^2/Vs) was assumed in both the Taurus and UFDG simulations. The Taurus-UFDG discrepancies in strong inversion are due to no significant series resistance in the Taurus domain.

A



B

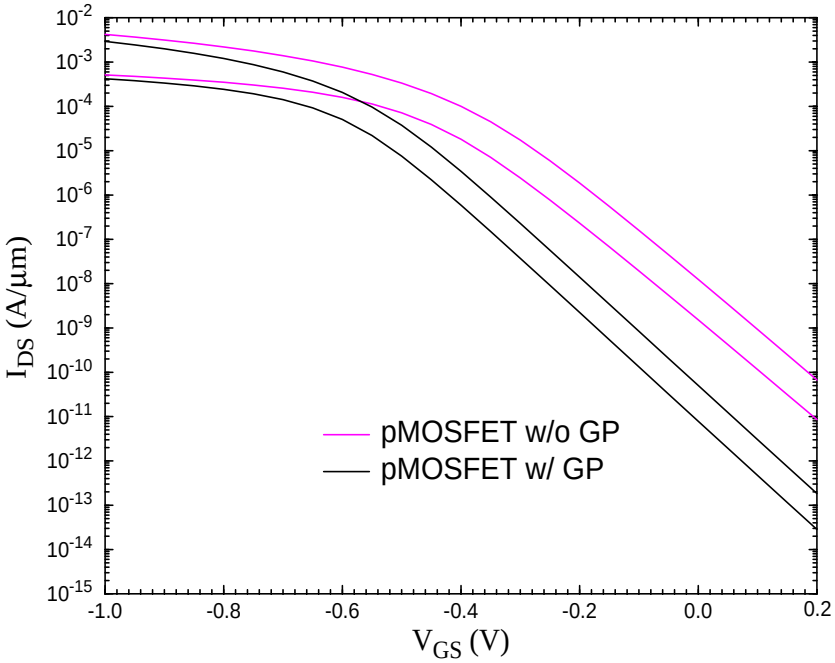
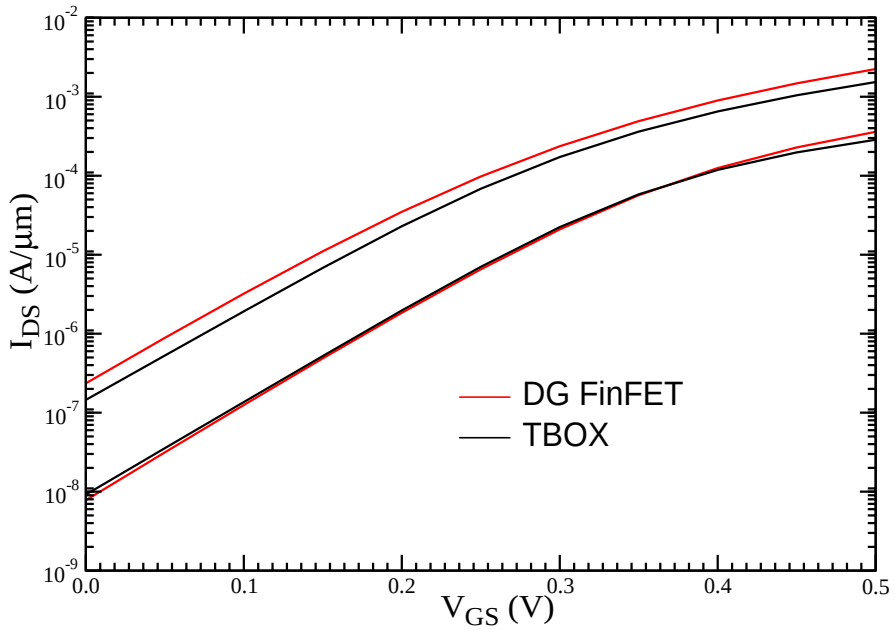


Figure 5-3. Comparison of Taurus-predicted current-voltage characteristics of TBOX FD/SOI CMOS devices, with and without GPs. The devices were designed with L_g = 25nm, t_{Si} = 5.5nm, S/D spacer width = 15nm, and Gaussian S/D doping profile, which gives L_{eff} = 30nm, i.e., 2.5nm G-S/D underlap. The substrate for both the nMOS and pMOS devices was assumed to be p^--type, which underlies the dramatic GP impact in the latter and lesser impact in the former. A) nMOSFET. B) pMOSFET.

A



B

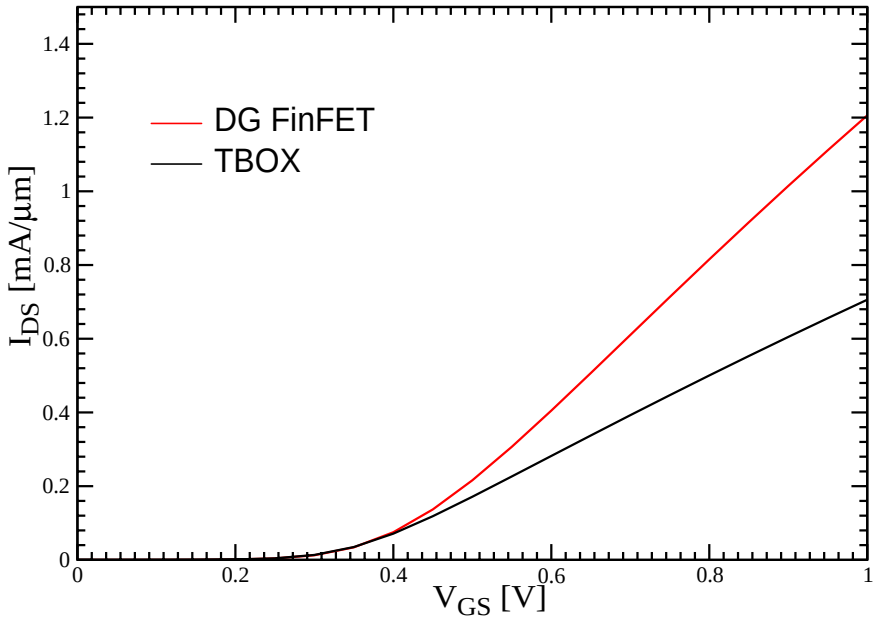


Figure 5-4. Comparison of Taurus-predicted current-voltage characteristics of a TBOX nMOSFET with that of a DG nFinFET counterpart. The TBOX device is the same as L_{eff} = 30nm TBOX device in Table 5.5; the DG FinFET was designed with L_{eff} = 30nm, t_{Si} = 15nm, and t_{oxf} = 1.2nm. A) Comparison of weak-inversion characteristics of the two devices; $\Delta\Phi_{Gf}$ and t_{Si} of the FinFET were tuned to get I_{off} and DIBL close to that of the TBOX device. B) Comparison of strong-inversion characteristics of the two devices at low V_{DS} (50mV), showing less than 2x current in FinFET relative to that of the TBOX MOSFET.

CHAPTER 6
SUMMARY AND FUTURE WORK

6-1 Summary

This dissertation is focused on physics-based modeling, optimal design, and performance of nanoscale DG FinFET CMOS for both digital and analog/RF applications. The major contributions of the research are summarized as follows.

In Chapter 2, we modeled DICE in nanoscale double-gate MOSFETs, and implemented it in our process/physics-based compact model UFDG. The effect of DICE on the transistor current, terminal charge, capacitance, and transcapacitance was studied, and verified via numerical and UFDG simulations. We found that DICE is a beneficial effect because it significantly increases inversion-charge density, and hence current, without significantly affecting gate capacitance. This is because the enhanced inversion charge is primarily supported by the drain. Its beneficial effect was demonstrated by the speed enhancement of DG FinFET CMOS via UFDG/Spice3 simulations. We showed that the current enhancement due to DICE is significantly larger for devices in which quasi-ballistic transport is important. For such devices, the current enhancement due to DICE is directly proportional to the enhancement in inversion-charge density, whereas, for devices in which quasi-ballistic transport is not important, the enhancement in current due to DICE is undermined by the increase in the gradual-channel length caused by DICE.

In Chapter 3, we proposed a design approach in which S/D engineering can be used to adjust the V_t of nanoscale DG MOSFETs for low-power and high-performance digital applications via limited densities of S/D dopants in the channel. The design approach was demonstrated and verified via numerical and UFDG simulations, and experimental results. It was shown that the proposed design approach augments the utility of the G-S/D underlap. The issue of random-doping effects was duly addressed by comprehensive Medici-based numerical

simulations. Our simulations considered the random-doping effects in the ultra-thin body, as well as in the S/D extension regions. Therefore, our sensitivity results include random variations in V_t, and also the random variations in L_{eff}. It was argued that the design approach will be viable and beneficial even in sub-10nm regime. It was noted, however, that at such short L_g, additional analyses of quantization effects, mobility degradation, and ballistic transport will be called for. Also, source/drain engineering techniques to support this design approach will have to be developed.

In Chapter 4, we compared the analog/RF performance of double-gate FinFETs with that of conventional planar bulk MOSFETs, and showed that the FinFET performance is better. Optimal design of FinFETs with G-S/D underlap, and the effect of higher $R_{S/D}$ in FinFETs on performance were discussed. We showed that G-S/D underlap significantly benefits the low-power analog/RF performance of DG FinFETs by reducing DIBL and G-S/D parasitic capacitances. Since low-power devices are biased in the moderate-inversion region, where the g_m/I_D ratio tends to be maximum, the increase in $R_{S/D}$ is not important due to low operating-current levels. Devices for high-frequency applications, on the other hand, are biased in the strong-inversion region, where g_m is maximum. Therefore, for high-frequency applications, increase in $R_{S/D}$ due to underlap could be important due to high operating-current levels, and could be detrimental to the device performance. However, it was shown that small G-S/D underlap would still be desirable because it will benefit the performance by eliminating G-S/D overlap capacitance, the effect of which was found to be more significant than that of the associated increase in $R_{S/D}$. And finally, the effect of scaling on the analog/RF performance of FinFETs, as L_g is scaled below ~30nm where quasi-ballistic transport becomes important, was discussed.

In Chapter 5, we studied the design of thin-BOX FD/SOI MOSFETs for low-power and high-performance digital applications, and compared their scalability to that of FinFETs. For low-

power applications we found that, with complex processing and layout, thin-box FD/SOI MOSFETs can be scaled to L_g ~ 10nm; whereas FinFETs can be scaled to L_g < 10nm with significantly simpler processing. For high-performance applications, the scalability of FinFETs is significantly better than that of thin-BOX FD/SOI MOSFETs. We noted that thin-BOX FD/SOI is similar to conventional planar MOSFETs in terms of the processing, therefore it could be a viable interim technology, bridging conventional CMOS and DG-FinFET CMOS.

6-2 Future Work

In Chapter 3, we proposed, and demonstrated using Medici/UFDG simulations, a design approach in which limited densities of source/drain dopants in the channel can be used for V_t adjustment in nanoscale DG MOSFETs. This design approach will be viable only if reliable source/drain engineering techniques are successfully realized. In order to realize these techniques, the diffusion of dopants through thin films should be modeled and experimentally verified. Also, the proposed design approach should be corroborated by additional analyses of quantization effect, mobility degradation, and ballistic transport for L_g ~ 10nm.

In Chapter 4, we discussed the optimal design of DG FinFETs for low-power and high-frequency analog/RF applications. This study was done at the device level with the emphasis on optimizing basic FOMs. The study should be further corroborated by demonstrative circuit simulations and analyses.

In Chapter 5, we studied the design of thin-BOX FD/SOI MOSFETs for low-power and high-performance applications. The study is primarily based on weak-inversion region analysis. In the strong-inversion region, UFDG is incapable of predicting the characteristics of thin-BOX devices with back-gate bias because of carrier accumulation. For a more comprehensive study, strong inversion analyses of UFDG should be upgraded to render the model fully applicable to thin-BOX FD/SOI devices with back-gate bias.

APPENDIX A
UFDG MODEL REFINEMENTS

A-1 Introduction

UFDG [16] is a process/physics based generic compact model which is applicable to diverse SOI-based DG MOSFETs. UFDG very effectively captures the unique physics of DG MOSFETs with UTB via robust modeling of terminal charge and carrier transport. UFDG has been verified by numerical device simulations and measured experimental results, and has been shown to reliably predict the characteristics of DG MOSFETs with L_g as short as ~30nm. For shorter L_g, due to lack of experimental results and inadequate modeling of carrier transport in the commercially available device simulators, there is no direct way of verifying the accuracy of UFDG, for example, that of its quasi-ballistic/ballistic carrier transport modeling. More than likely, future UFDG upgrades will be needed as CMOS technology is further scaled. Such model upgrading is exemplified by this appendix, in which we discuss the refinements we did in UFDG to make it more robust for reliable prediction of nanoscale DG MOSFETs studied in this dissertation.

A-2 Strong-Inversion Intrinsic Charge Modeling

The voltage-dependent terminal charges of the DG MOSFET is characterized in order to model the charge dynamics for large-signal transient simulations. The terminal charges are assumed quasi-static, and are individually integrated based on spatial dependence in the MOSFET which follow from the analyses in [25], modified by the DICE analyses discussed in Chapter 2. Charge neutrality, which is important for the stability and convergence of a compact model, is satisfied due to physics-based treatment of all the terminal charges. The charge modeling in UFDG is divided into two parts: triode and saturation regions, which are eventually merged for a continuous model.

In the triode region, front-gate charge (Q_{Gf}) is characterized by the integral:

$$Q_{Gf} = W_g C_{oxf} \int_0^{L_g} [V_{GfS} - \Phi_{GfB} - \phi_0(0, y) - \Delta\phi(y)] dy . \quad \text{(A-1)}$$

Back-gate charge (Q_{Gb}) is characterized by a similar integral at the back surface. Next, the integrated charge in the channel is evaluated by the integral:

$$Q_{ch} = W_g \int_0^{L_g} Q_{ch}(y) dy , \quad \text{(A-2)}$$

where $Q_{ch}(y)$ is given by (2.4), which includes the DICE charge in the channel. The integrated inversion charge is then partitioned between the source and the drain. The portion of the inversion charge assigned to the drain is evaluated by the integral:

$$Q_{D(ch)} = W_g \int_0^{L_g} \frac{y}{L_g} Q_{ch}(y) dy , \quad \text{(A-3)}$$

and rest is assigned to the source,

$$Q_{S(ch)} = Q_{ch} - Q_{D(ch)} . \quad \text{(A-4)}$$

Finally, body-depletion charge (Q_B) is defined as,

$$Q_B = -W_g L_g q N_A t_{Si} . \quad \text{(A-5)}$$

Typically Q_B is negligible because of undoped body. Note that in the previous [25] charge modeling, the drain-depletion charge in the triode region was ignored. However, as we discussed in Chapter 2, DICE charge is significant in nanoscale DG MOSFETs, and is primarily supported by the drain. Therefore, a corresponding depletion charge is assigned to the drain in order to maintain the charge neutrality.

In saturation region, the charges defined for the triode region will still be valid with L_g replaced by L_{gch}, and V_{DS} replaced by $V_{DS(eff)}$, and they will be augmented by the charges

associated with the high-E_y region. With L_{gch} and $V_{DS(eff)}$ solved as discussed in Chapter 2, we define the saturation-region component of the front-gate charge (Q_{Gf}^{sat}) by the integral:

$$Q_{Gf}^{sat} = W_g C_{oxf} \int_{L_{gch}}^{L_g} [V_{GfS} - \Phi_{GfB} - \phi_0(0, y) - \Delta\phi(y)] dy, \tag{A-6}$$

and the saturation-region component of the back-gate charge (Q_{Gb}^{sat}) will be defined by a similar integral at the back surface.

Since the carrier velocity is saturated in the high-E_y region, based on the current continuity, the charge density will be nearly uniform along y, and channel charge is defined by simple integral:

$$Q_{ch}^{sat} = W_g \int_0^{L_g} Q_{ch}(y) dy \cong W_g (L_g - L_{gch}) Q_{ch}(L_{gch}). \tag{A-7}$$

Analogous to the scheme in the triode region, Q_{ch}^{sat} is partitioned between the source ($Q_{S(ch)}^{sat}$) and the drain ($Q_{D(ch)}^{sat}$). As defined in Chapter 2, the drain-depletion charge is assigned to the drain, which includes charge components of both triode region and saturation region, hence thereby charge neutrality is followed.

All evaluated charge components associated with the saturation region is added to their respective triode-region components to give the total charge associated with the terminal, as follows:

$$Q_{Gf} = Q_{Gf}(L_g, V_{DS(eff)}) + Q_{Gf}^{sat}, \tag{A-8}$$

$$Q_{S(ch)} = Q_{S(ch)}(L_g, V_{DS(eff)}) + Q_{S(ch)}^{sat}, \tag{A-9}$$

$$Q_{D(ch)} = Q_{D(ch)}(L_g, V_{DS(eff)}) + Q_{D(ch)}^{sat} + Q_{D(dep)}, \tag{A-10}$$

$$Q_{Gb} = -(Q_{Gf} + Q_{S(ch)} + Q_{D(ch)} + Q_B). \tag{A-11}$$

A-3 Weak-Inversion Inner Fringe Charge Model

UFDG's basic fringe-capacitance model [29] is based on the solution of Laplace equation between two conducting plates separated by an angle. For example, UFDG calculates the inner-fringe capacitance between source and the gate by assuming source and gate to be two conducting plates, and solving the Laplace equation between the two in the weak-inversion region. However, in nanoscale DG MOSFETs, the electric-fields originating from the depletion charge in the drain can significantly influence the solution of Laplace's equation between the source and the gate. This inadequacy in the inner-fringe capacitance model was identified by the V_{DS}-dependent non-physical "dip" in the UFDG-predicted C_G-V_{GS} characteristics in the moderate-inversion region. UFDG fixes this inadequacy by quasi-physically assigning a depletion charge to the drain, and imaging it in the gate. This drain-depletion charge is modeled as,

$$Q_{D(dep)if} = \frac{Q_{if} W_g \varepsilon_{Si} t_{Si} V_{DS}}{L_{eff}}. \qquad \text{(A-12)}$$

The drain-depletion charge in (A-12) is imaged in the front and back gate equally. Q_{if} is a tuning parameter, and L_{eff} is the effective channel length of the MOSFET [26] in the weak-inversion region. This formalism effectively removes the dip in the C_G-V_{GS} characteristics and significantly improves the spline linking the weak- and strong-inversion region formalism.

A-4 DIBL-Dependent Definition of V_{TW}

In UFDG, a novel cubic spline in terms of both the front- and back-gate bias is used to link the rigorous weak- and strong-inversion formalism. For physics-based compact modeling, the MOSFET I_{DS}-V_{GS} characteristics are generally partitioned into four regions of operation: accumulation, weak inversion, moderate inversion, and strong inversion. This division of the I_{DS}-V_{GS} characteristics facilitates developing rigorous analytical models based on approximations valid only in the given region. Because both the 2-D effects and the inversion carrier density are

important in the moderate-inversion region, analytical modeling of this region is formidable. Hence, in general, it is modeled via smoothing functions or a polynomial spline that link the rigorous weak- and strong-inversion region formalisms. For a polynomial spline, one needs to know the gate voltages at the boundaries of the moderate-inversion region, and I_{DS} and g_m at those gate voltages.

For classical devices, Tsividis quantitatively characterized the noted boundaries of the moderate-inversion region in terms of the inversion capacitance, depletion capacitance, and the gate-oxide capacitance. It is based on the exponential dependence of the inversion charge on the gate bias for the weak-inversion region, and on the linear dependence for the strong-inversion region. In UFDG we modify this characterization for the nonclassical devices.

The gate bias at the boundary corresponding to the weak inversion (V_{TW}) and strong inversion (V_{TS}), must always be evaluated first in order to determine which region of operation a given gate bias corresponds to. In UFDG we model boundaries of the moderate-inversion region by equating two expression of the inversion-region charge density at the boundaries. The first one is defined by the (1-D) Gauss's law in the intrinsic body, and the second expression is an approximation based on the linear potential distribution given by the spatially constant transverse electric field for the undoped body/channel. However, we have discovered that for nanoscale DG MOSFETs, 1-D analysis does not adequately characterizes the inversion-charge density in the channel, and DIBL must be accounted for in the model.

To model DIBL [45], we write $\phi(x, y) = \phi_0(x, y) + \Delta\phi(x, y)$, where $\phi_0(x, y)$ is the solution of the potential with $V_{DS} = 0$, and $\Delta\phi(x, y)$ is the perturbation in the potential due to drain bias, which, for weak inversion, with undoped body, satisfies

$$\frac{\partial^2}{\partial x^2}\Delta\phi(x,y) + \frac{\partial^2}{\partial y^2}\Delta\phi(x,y) = 0. \qquad \text{(A-13)}$$

We separate the two partial derivatives assuming,

$$\frac{\partial^2}{\partial x^2}\Delta\phi(x,y) = \frac{\partial^2}{\partial y^2}\Delta\phi(x,y) = \eta_0, \tag{A-14}$$

where η_0 is a spatial constant. Then, integrating twice along the channel, with the boundary conditions $\Delta\phi(y = 0) = 0$ and $\Delta\phi(y = L_{eff}) = V_{DS}$ yields $\eta_0 = 2V_{DS}/L_{eff}^2$, where L_{eff} is the effective channel length of the device that governs the 2-D effects in the UTB, and that is defined by the source/drain doping profile. Then by integrating (A-14) twice along the film we obtain,

$$\Delta\phi_{(sb)} = \left(\frac{C_b}{C_{ob} + C_b}\right)\Delta\phi_{(sf)} + \left(\frac{1}{C_{ob} + C_b}\right)\frac{\varepsilon_{Si} t_{Si} \eta_0}{2}, \tag{A-15}$$

and,

$$\Delta\phi_{(sf)} = \left[\frac{C_{ob} + 2C_b}{C_b\left(C_{of} + C_{ob}\right) + C_{ob}C_{of}}\right]\frac{\varepsilon_{Si} t_{Si} \eta_0}{2}, \tag{A-16}$$

where $\Delta\phi_{(sf)}$ and $\Delta\phi_{(sb)}$ are the perturbations of the minimum surface potentials (in y) at the front and back surfaces. In UFDG we account for DIBL in the definition of V_{TW} by perturbing $\Delta\phi(0)$ by (A-15) and (A-16). Note that SCEs tend to subside at the moderate/strong-inversion boundary due to strong transverse electric field. Therefore, model for evaluation of V_{TS} does not needs to be altered.

LIST OF REFERENCES

[1] International Technology Roadmap for Semiconductors, Semiconductor Industry Association, Austin, TX, 2005.

[2] J. G. Fossum, L. Q. Wang, J-W. Yang, S.-H. Kim, and V. P. Trivedi, "Pragmatic design of nanoscale multi-gate CMOS," *IEDM Tech. Dig.*, pp. 613-616, Dec. 2004.

[3] V. P. Trivedi, J. G. Fossum, and W. Zhang, "Threshold voltage and bulk inversion effects in nonclassical CMOS devices with undoped ultra-thin bodies," *Solid-State Electronics*, pp. 170-178, Jan 2007.

[4] D. Boyle *et al.*, "Tri-Gate Fully-Depleted CMOS Transistor: Fabrication, Design, Layout," *Symp. VLSI Tech. Dig.*, pp. 133-134, 2003.

[5] T. -S. Park, E. Yoon, and J. -H. Lee. "A 40nm body-tied FinFET (OMEGA MOSFET) using bulk Si wafer." *Physica E*, vol 19, no. 1, pp. 6-12, Jul 2003.

[6] T. -S. Park *et al.*, "Fabrication of body-tied FinFETs (OMEGA MOSFET) using bulk Si wafer." in *VLSI Symp. Tech. Dig.*, pp. 135-136, 2003.

[7] M. Poljak, V. Jovanovic, and T. Suligoj, "Technological constraintes of bulk FinFET structure in comparision with SOI FinFET", *ISDRS Tech. Dig.*, pp. 1-2, Dec 2007.

[8] W. Zhang, J. G. Fossum, L. Mathew, and D. Yang, "Physical Insights Regarding Desing and Performance of Independent-Gate FinFETs," *IEEE Trans. Electron Devices*, vol. 52, pp. 2198-2206, Oct 2005.

[9] W. Zhang, J. G. Fossum, and L. Mathew, "The ITFET: A Noverl FinFET-Based Hybrid Device," *IEEE Trans. Electron Devices*, vol. 53, pp. 2335-2343, Sept 2006.

[10] M. M. Chowdhury and J. G. Fossum, "Physical Insights on Electron Mobility in Contemporary FinFETs," *IEEE Electron Device Lett.*, vol. 27, pp. 482-485, June 2006.

[11] M. M. Chowdhury, V. P. Trivedi, J. G. Fossum, and L. Mathew, "Carrier mobility/ transport in undoped-UTB DG FinFETs," *IEEE Trans. Electron Devices*, vol. 54, pp. 1125-1131, May 2007.

[12] J. G. Fossum and S. Chouksey, "Modeling of Saturation-Region Characteristics of Nanoscale Double-Gate MOSFETs," *Tech. Proc. 2007 Nanotechnology Conf. (Vol. 3 - WCM)*, pp. 510-511, May 2007.

[13] L. Ge, J. G. Fossum and B. Liu, "Physical compact modeling and analysis of velocity overshoot in extremely scaled CMOS devices and circuits," *IEEE Trans. Electron Devices*, vol. 48, pp. 2074-2080, Sept. 2001.

[14] J. G. Fossum *et al.*, "A Process/Physics-Based Compact Model for Nonclassical CMOS Device and Circuit Design," *Solid-State Electronics.*, vol. 48, pp. 919-926, June 2004.

[15] J. G. Fossum *et al.*, "Recent Upgrades and Applications of UFDG," *Tech. Proc. 2006 Nanotechnology Conf. (WCM)*, pp. 674-679, May 2006.

[16] J. G. Fossum, "UFDG MOSFET MODEL (Ver. 3.7) User's Guide," SOI Group, Univ. Florida, Gainesville, July 2007.

[17] S. -H. Kim, and J. G. Fossum, "Design Optimization and Performance Projections of Double-Gate FinFETs With Gate-Source/Drain Underlap for SRAM Application," *IEEE Trans. Electron Devices*, vol. 54, pp. 1934-1942, Aug 2007.

[18] H. -Yu Chen *et al.*, "Novel 20nm Hybrid SOI/Bulk CMOS Technology with 0.183mm2 6T-SRAM Cell by Immersion Lithography," *Symp. VLSI Tech. Dig.*, pp. 16-17, 2005.

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

[20] S. Veeraraghavan and J. G. Fossum, "A Physical Short-Channel Model for the Thin-Film SOI MOSFET Applicable to Device and Circuit CAD," *IEEE Trans. Electron Devices*, vol. 35, pp. 1866-1875, Nov. 1988.

[21] F. M. Klaassen and W. C. J. de Groot, "Modelling of Scaled-Down MOS Transistors," *Solid-State Electronics*, vol. 23, pp. 237-242, Mar. 1980.

[22] L. Mathew (Freescale Semiconductor), private communication, Dec. 2006.

[23] S.-H Kim, J. G. Fossum, and V. P. Trivedi, "Bulk inversion in FinFETs and implied insights on effective gate width," *IEEE Trans. Electron Devices*, vol. 52, pp. 1993-1997, Sept 2005.

[24] *Medici-4.0 User's Manual*, Synopsys, Inc., Durham, NC, 2004.

[25] M.-H. Chiang, *Process-Based Compact Modeling and Analysis of Silicon-on-Insulator CMOS Devices and Circuits, Including Double-Gate MOSFETs*. Ph.D. Dissertation, Univ. Florida, Gainesville, 2001.

[26] V. P. Trivedi and J. G. Fossum, "Nanoscale FinFETs with gate-source/drain underlap," *IEEE Trans. Electron Devices*, vol. 52, pp. 56-62, Jan 2005.

[27] L. Ge and J. G. Fossum, "Analytical modeling of quantization and volume inversion in thin Si-film DG MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, pp. 287-294, Feb. 2002.

[28] V. P. Trivedi and J. G. Fossum, "Quantum-mechanical effects on threshold voltage of undobed double-gate MOSFETs", *IEEE Electron Device Lett.*, vol. 26, p. 579-582, Aug. 2005.

[29] S.-H. Kim, J. G. Fossum, and J. Yang, "Modeling and significance of fringe capacitance in nonclassical CMOS devices with gate-source/drain underlap," *IEEE Trans. Electron Devices*, vol. 53, pp. 2143-2150, Sept. 2006.

[30] L. Mathew *et al.*, "Double-Gate CMOS Technology with Sub-Lithographic (<20nm), 100nm Tall, Undoped Channel, TiN+$Hf_xZr_{1-x}O_2$ Gate, Multiple Silicided Source/ Drain with Record PMOS I_{on}/I_{off}," *Proc. IEEE Si Nanoelectronics Workshop*, June 2007, pp. 100-101.

[31] M.-H. Chiang, J.-N. Lin, K. Kim, and C.-T. Chuang, "Random Dopant Fluctuation in Limited-Width FinFET Technologies," *IEEE Trans. Electron Devices*, vol. 54, pp. 2055-2060, Aug 2007.

[32] D. J. Frank, Y. Taur, M. Ieong, and H.-S. P. Wong, "Monte Carlo Modeling of Threshold Variation Due to Dopant Fluctuations," *Symp. VLSI Tech. Dig.*, pp. 169-170, June 1999.

[33] J. G. Fossum, L. Ge, and M.-H Chiang, "Speed Superiority of Scaled Double-Gate CMOS," *IEEE Trans. Electron Devices*, vol. 49, pp. 808-811, May. 2002.

[34] T. C. Lim, and G. A. Armstrong, "Scaling issues for analogue circuits using Double Gate SOI transistors," *Solid-State Electronics.*, vol. 51, pp. 320-327, Feb 2007.

[35] V. Subramaniam *et al.*, "Planar Bulk MOSFETs Versus FinFETs: An Analog/RF Perspective," *IEEE Trans. Electron Devices*, vol. 53, pp. 3071-3079, Dec 2006.

[36] A. Chatterjee *et al.*, "Transistor Design Issues in Integrating Analog Functions with High performance Digital CMOS," Proc. *VLSI Symp. Tech.*, pp. 147-148, June 1999.

[37] *Internat. Tech. Roadmap for Semiconductors*, Semiconductor Industry Assoc., http://public.itrs.net, 2008.

[38] J. Kedzierski *et al.*, "High-performance symmetric-gate and CMOS-compatible V_t assymteric-gate FinFET devices", *IEDM Tech. Dig.*,, p. 437-440, Dec. 2001.

[39] R. Tsuchiya *et al.*, "Silicon on Thin Box: a new paradigm of the CMOSFET for low-power high-performance applications featuring wide-range back-bias control", *IEDM Tech. Dig.*, p. 631-634, Dec. 2004.

[40] O. Weber *et al.*, "High immunity to threshold voltage variability in undopes ultra-thin FDSOI MOSFETs and its physical understandins", *IEDM Tech. Dig.*, p. 245-248, Dec. 2008.

[41] S. Monfray *et al.*, *Proc. EUROSOI Conf.*, p. 107-108, Jan. 2009.

[42] P. Scheiblin *et al.*, *Proc. EUROSOI Conf.*, p. 109-110, Jan. 2009.

[43] G. G. Shahidi, "FDSOI for Low Power CMOS", *IEEE SOI Conf. Tech. Dig.*, p. 1-2, Oct. 2009.

[44] *Taurus-2006 User's Manual*, Synopsys, Inc., Durham, NC, 2006.

[45] V. P. Trivedi and J. G. Fossum, "Scaling fully depleted SOI CMOS", *IEEE Trans. Electron Devices*, vol. 50, p. 2095-2103, Oct. 2003.

[46] P. C. Yeh and J. G. Fossum, “Physical subthreshold modeling applied to viable design deep submicrometer fully depleted SOI low-voltage CMOS technology”, *IEEE Trans. Electron Devices*, vol. 42, p. 1605-1613, Sept. 1995.

[47] A. Majumdar *et al.*, “Gate Length and Performace Scaling of Undoped-Body Extremely Thin SOI MOSFETS”, *IEEE Electron Device Lett.*, vol. 30, p. 413-415, Apr. 2009.

[48] J. G. Fossum, “Physical Insights on nanosclae multe-gate CMOS design”, *Solid-State Electronics*, vol. 51, p. 188, Feb. 2007.

[49] J. Kedzierski *et al.*, “Multi-gate FinFET and fully-depleted SOI devices using total gate silicidation”, *IEDM Tech. Dig.*, p. 247, Dec. 2002.

[50] J. -W Yang and J. G. Fossum,. “On the Feasibility of Nanoscale Triple-Gate CMOS Transistors”, *IEEE Trans. Electron Devices*, vol. 52, pp. 1159-1164, Jun 2005.

[51] R. Yan *et al.*, “LDD Depletion Effects in Thin-BOX FDSOI Devices with a Ground Plane” *IEEE SOI Conf. Tech. Dig.*, pp. 53-54, 2003.

[52] T. -J. K. Liu *et al.*, “SRAM Cell Design Considerations for SOI Technology” *IEEE SOI Conf. Tech. Dig.*, pp. 27-33, 2003.

[53] C. Shin *et al.*, “SRAM Yield enhancement with Thin-BOX FD-SOI” *IEEE SOI Conf. Tech. Dig.*, pp. 37-38, 2003.

BIOGRAPHICAL SKETCH

Siddharth Chouksey was born in Betul, India in 1983. He received the Bachelor of Technology in information and communication technology from Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar, India in 2005. He did his Doctor of Philosophy in electrical and computer engineering at the University of Florida, Gainesville. His research interest concerns device theory, modeling and design of non-classical silicon-on-insulator (SOI) and multi-gate MOSFET, for both digital and analog applications. In summer of 2008 and 2009, he interned in the Components Research Group at Intel, Hillsboro, OR, where he was actively involved in the design of novel nanoscale floating-body memory cells.