Scaling Opportunities for Bulk Accumulation and Inversion MOSFETs for Gigascale Integration

A Thesis
Presented to
The Academic Faculty

by

Raghunath Murali

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Electrical and Computer Engineering
Georgia Institute of Technology
February 2004

Scaling Opportunities for Bulk Accumulation and Inversion MOSFETs for Gigascale Integration

Approved by:

Professor James D. Meindl, Advisor

Professor Jeffrey A. Davis

Professor Phillip E. Allen

Professor Dennis W. Hess
(School of Chemical Engineering)

Professor John D. Cressler

Date Approved: Feb. 6, 2004

To my parents

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor, Dr. Meindl, for his guidance, motivation, and inspiration. I thank my committee members Dr. Phil Allen, Dr. John Cressler, Dr. Jeffrey Davis, and Dr. Dennis Hess for their time and effort. I thank all the GSI group members, in particular, Azad, Chen, Femi, Hiren, Jim, Kaveh, Lihui, Muhannad, Reza, and Tony for many motivating and spirited discussions. I thank former GSI group members Azeez Bhavnagarwala, Keith Bowman, and Ragu Venkatesan, for many inspiring discussions. I thank Blanca Austin for helping me ramp up on my research by providing me the code for all her models, and for proof-reading many of my publications. I thank Mike Dennen, Bill Richards, and Drew Vinal of Thunderbird Technologies for providing critical comments on many parts of my work. I deeply appreciate the help of the GSI group program manager, Jennifer Tatham (aka "GSI Mom"), for her constant support and encouragement. I thank my parents for their never ending love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The objective of this research is to comprehensively compare bulk accumulation and inversion MOSFETs, and find application areas where each is superior. Short channel effect (SCE) models for accumulation and inversion MOSFETs are derived that accurately predict threshold voltage, subthreshold swing, and subthreshold current. A source/drain junction depth dependent characteristic length is derived that can be used to rapidly assess the impact of junction depth scaling on minimum channel length. A fast circuit simulation methodology is developed that uses physically based I-V models to simulate inversion and accumulation MOSFET inverter chains, and is found to be accurate over a wide range of supply voltages. The simulation methodology can be used for rapid technology optimization, and performance prediction. Design guidelines are proposed for accumulation MOSFET design; the guidelines result in a low process sensitivity, low SCE, and a subthreshold current less than the allowable limit. The relative performance advantage of accumulation/inversion MOSFETs is gate-technology dependent. In critical comparisons, on-current is evaluated by means of a full band Monte Carlo device simulation. Gate-leakage, and band-to-band tunneling leakage at the drain-substrate region are included in the performance analysis. For mid-bandgap metal gate, accumulation MOSFETs perform better than inversion MOSFETs for hi-performance (HiP) and low-operating power (LOP) applications. For tunable metal gate technology, inversion MOSFETs always perform better than accumulation MOSFETs. For dual poly technology, accumulation MOSFETs perform better than inversion MOSFETs for low standby power (LSTP) applications. A comprehensive scaling analysis has been performed on accumulation and inversion MOSFETs using both SCE models and 2-D simulations. Results show that accumulation MOSFETs can scale better than inversion MOSFETs for mid-bandgap metal gate HiP, and LOP applications; and poly gate LSTP applications.

CHAPTER I

INTRODUCTION

1.1 Introduction

Device miniaturization reduces the physical dimensions of devices, thereby increasing the density of devices on a die. Miniaturization not only results in increased functionality on a chip but also improved performance. The International Technology Roadmap for Semiconductors (ITRS) [1] predicts feature sizes as small as 13nm by 2016. ITRS groups application areas into low standby power (LSTP), low operating power (LOP) and high-performance (HiP). The on-current (I_{ON}) and off-current (I_{OFF}) requirements of these application areas are shown in Table 1. Device scaling results in a number of effects, some classical and others quantum mechanical. Proper modeling of these effects is quintessential in achieving optimal performance for future technology.

Important classical effects arising from device scaling include threshold voltage rolloff, drain induced barrier lowering (DIBL), and subthreshold swing (S) rollup. These effects considerably degrade device performance and limit device scaling. Quantum mechanical (QM) effects arising as a result of scaling include quantization of electron energy in the channel; and tunneling leakage at the gate-channel, gate-overlap, and drain-substrate regions. These effects tend to degrade the device performance. Non-local carrier transport becomes important in short channel length MOSFETs, resulting in performance improvement by means of velocity overshoot [2].

Table 1: ITRS application areas and their I_{ON}, and I_{OFF} requirements.

application	I_{ON} $(\mu A/\mu m)$	I_{OFF} $(nA/\mu m)$
Low standby power (LSTP)	400	0.001
Low operating power (LOP)	600	0.3-0.7
High-performance (HiP)	900	300-1000

The goal of the proposed research is to comprehensively compare bulk accumulation and inversion devices in order to assess their scalability for gigascale integration. Both physical models and numerical simulations are used to compare the devices. Physically based device models allow physical insight into device scaling phenomena, rapid optimization, and circuit simulation. Numerical device simulations allow inclusion of non-uniform doping, QM effects, and non-local carrier transport.

1.2 Background

Bulk MOSFETs can be divided into two broad categories: accumulation type, where the current flow arises from majority carriers, and inversion type, where the current flow arises from minority carriers. The various bulk MOSFET structures are shown in Fig. 1. Based on the vertical doping profile, inversion FETs can be further classified into uniformly doped surface channel inversion (UD-SCI) FETs, and retrograde doped surface channel inversion (RD-SCI) FETs. The RD-SCI FET is also known as low impurity channel transistor (LICT) [3], atomic layer doped (ALD) transistor [4], super steep retrograde (SSR) doped FET [5], or epi-MOSFET [6]. The main difference between UD-SCI and RD-SCI FETs is the existence of a lightly doped layer in the RD-SCI device.

The accumulation FET can be further classified into surface channel accumulation (SCA) FET, and buried channel accumulation (BCA) FET. The main difference between BCA and SCA devices is the location where the channel first opens. BCA FET conduction starts in a neutral buried channel, and with increasing applied gate bias, subsequently extends to an accumulated surface layer. On the other hand, SCA FET conduction is limited to a surface accumulation layer. BCA devices normally have highly doped/deep implant layers that prevent the extension of the bulk depletion region all the way to the surface; SCA devices are characterized by lightly doped/shallow implant layers that allow the extension of the bulk depletion region all the way and beyond the $\mathsf{Si/SiO_2}$ interface. As a result, the BCA device can be designed to achieve a low threshold voltage (V_T), while the SCA device's lower V_T boundary is the flat-band voltage ($\mathsf{V_{FB}}$). The channel (or tub) and well (or substrate) doping types of various FET types are shown in Table 2.

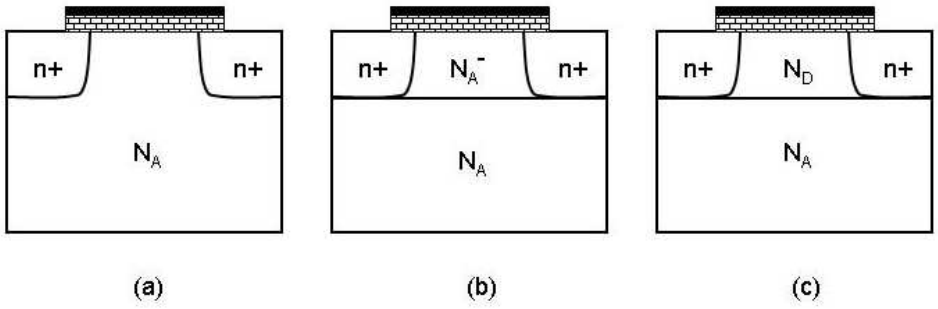


Figure 1: Vertical profile of bulk MOS devices. (a) Uniformly-Doped Surface Channel Inversion (UD-SCI). (b) Retrograde-Doped Surface Channel Inversion (RD-SCI). (c) Buried Channel Accumulation (BCA), Surface Channel Accumulation (SCA), and Fermi-Threshold FET (FF).

Table 2: Channel and well doping for different n-FET types.

device	*channel/tub-doping*	*substrate/well doping*
UD-SCI	p-type (N_A)	p-type (N_A)
RD-SCI	p-type (N_A^-)	p-type (N_A)
BCA/SCA	n-type (N_D)	p-type (N_A)

The SCI device is the conventional MOSFET [7] commonly used since the inception of MOS technology. BCA/SCA FETs too date back to the beginning of MOS technology [8],[9]. BCA/SCA FETs on insulating substrates [9] were introduced at about the same time as SCI silicon on insulator (SOI) FETs. In single-poly CMOS technology, the n-channel device is generally an SCI device, while the p-channel device is BCA/SCA type. To avoid buried channel operation and the resultant SCE increase [10], dual poly technology was introduced to make both n-MOS and p-MOS devices to be SCI type. But p+ poly gates (used for SCI p-MOSFETs) have problems of Boron penetration [11]; and dual poly technology results in increased process complexity, and reduced circuit density compared to single-poly technology. So accumulation FETs are still used in poly gate CMOS technology [12]-[16]. Accumulation FETs are also used in mid-bandgap metal gate CMOS [17], where the same metal gate is used for both n-MOS and p-MOS FETs.

The lightly doped layer in the RD-SCI FET introduces extra degrees of freedom in device optimization. For a given threshold voltage (V_T), the RD-SCI FET is believed to exhibit

lower short channel effects (SCE) compared to the UD-SCI FET [18]. The RD-SCI FET is also believed to result in better current drive capability as a result of low channel doping [19]. However, other studies show that the RD-SCI FET has only a marginal advantage over the UD-SCI FET in terms of drive current [20].

Many modifications have been proposed to the original SCI MOSFET structure in order to overcome SCE. Reducing the source/drain (S/D) junction depth [21] is an important way of minimizing charge sharing and thus SCE. However, ultra-shallow source/drain junctions result in increased source/drain resistance (r_{sd}) [22] and thus degrade I_{ON}. In order to achieve both ultra-shallow junctions and low r_{sd}, the grooved-gate MOSFET [23], electrically induced source/drain extension MOSFET [24],[25], and raised source/drain MOSFET [22], [26] have been introduced. In order to reduce the source resistance (R_s, which has a greater impact on I_{ON} than the drain resistance, R_d [27]), asymmetric source/drain structures have been introduced [28]. Another method of reducing SCE in structures with deep S/D junctions is to use dielectric pockets [29]. SiGe was used in the S/D regions to fabricate sub-100nm SCI p-MOSFETs with ultra-shallow S/D extensions and good I_{ON} [30]. All the modifications mentioned above can be easily applied to accumulation FETs. S/D engineering is not considered in this work since it has a similar impact on both accumulation and inversion FETs.

Another way of controlling SCE in inversion FETs is to introduce halos [31], [32] below the S/D region. Halos are high-concentration implants that reduce charge sharing resulting from S/D regions. In an accumulation FET too, halos reduce SCE [33]. Because of the complexity of analytically modeling non-uniform channel doping in the lateral direction, halo-doped MOSFETs are analyzed using 2D numerical simulations in this work. Asymmetric halos have been proposed to take advantage of velocity overshoot [28] in scaled MOSFETs and to improve analog performance [34].

Another way of to improve MOSFET performance is to operate it at a low temperature. The important effects of reducing the operating temperature are a reduction in S and an increase in channel mobility (μ). There is also an increase in saturation velocity (v_{sat}) and more velocity overshoot. The net result of this is an increase in I_{ON} and a reduction in

I_{OFF} [35].

1.3 Motivation

There are reportedly many advantages of accumulation FETs. The low vertical field in the channel region of the accumulation device when operating in buried channel mode results in increased mobility with respect to its SCI counterparts [36]. In addition, the low implant doping concentration in the SCA device will also result in high mobility. The Fermi FET, a device that attempts to take advantage of both these characteristics, has been proposed in [37] and analyzed through simulations in [38]. In this device, the substrate depletion region extends exactly to the surface resulting in minimal vertical electric field in the channel. Its threshold voltage is determined only by the flat band voltage (V_{FB}) and the built-in potential of the substrate-channel p-n junction (ϕ_{bi}) [37]. Thus, V_T is independent of oxide thickness. The Fermi FET can be identified as the boundary between the BCA and SCA device. It was also determined in [37]-[39] that this device yields the minimum long-channel subthreshold swing possible in bulk devices. Hence, in the search for the optimal bulk device for deep-submicron operation, the Fermi FET will be carefully considered. However, alternate gate materials [40] will be needed to obtain low V_T SCA/FF devices as a result of the flat-band limitation.

The accumulation FET has been reported to have a higher breakdown voltage [36] and reduced hot-carrier generation [41] compared to SCI FETs. The SCI FET has a higher $1/f$ noise because of interface states, whereas the BCA FET conducts partially or completely in the bulk and thus has a lower $1/f$ noise [42].

Since the channel region is doped the same type as the source/drain (S/D) region, the lateral electric field at the source/drain junction is smaller in accumulation FETs compared to inversion FETs [43]. This might lead to accumulation FETs being able to scale to lower channel lengths compared to inversion FETs. Due to a lower vertical field, QM V_T shift can be expected to be lower in accumulation FETs. Also, because of a lower source/drain-substrate area, accumulation FETs are expected to have a lower junction capacitance when compared to a similar inversion FET structure. The RD-SCI FET has a low dopant induced

V_T fluctuation because of its near-intrinsic channel [44]. Since the SCA FET too has a low channel doping, it is expected to show a similar dopant-induced V_T fluctuation behavior.

Only a few publications have made clear the distinction between BCA and SCA threshold voltage design [39],[45]-[47]. Although the SCA FET has a lower S and lower SCE compared to the BCA FET, few publications emphasize operation of the accumulation FET as an SCA FET [38]-[39],[47]. In addition, existing guidelines for design of accumulation FETs to operate as SCA devices are based on 1-D theory [38]-[39],[47]. and do not hold good for short channel devices. New design guidelines based on 2-D theory are needed for designing accumulation FETs so that they operate as SCA devices.

Previous comparisons between accumulation and inversion FETs have yielded conflicting results and have many shortcomings, primarily because of the limited design space considered [10]-[11],[38], [48] -[52]. In addition, many comparisons are for longer channel lengths and thicker oxides [10],[38],[49], where quantum mechanical effects and non-local carrier transport are not significant. In present day devices, both these effects are extremely important, and a re-evaluation of the comparison is needed in such scaled devices. In [11], accumulation FETs with ultra-shallow channel implants are found to be superior to inversion FETs. In [38], SCA FETs are found to be superior to SCI FETs. In [50] and [51], the comparisons are limited to that between metal-gate accumulation and poly-gate inversion FETs, and both find the poly-gate inversion FET to be superior to the accumulation FET. In [10], only hi-performance poly-gate accumulation and inversion FETs are compared, and inversion FETs are found to be scalable to lower supply voltages than accumulation FETs. In [52], a single tub-depth (y_i) and substrate doping (N_A) value are considered in evaluating the accumulation FET, and the accumulation FET is found to be less scalable than the inversion FET. This work attempts to compare accumulation and inversion MOSFETs for a variety of applications, gate materials and power supply voltages.

1.4 Organization

The rest of the dissertation is organized as follows. In Chapter II, a threshold voltage (V_T) rolloff model is proposed that works for both deep junction and shallow junction devices. A

source/drain junction depth dependent characteristic length is proposed that can be used to rapidly assess the impact of junction depth scaling on V_T rolloff. Then, long channel threshold voltage of accumulation FETs is derived. In Chapter III, subthreshold swing of accumulation and inversion MOSFETs is derived by using the volume inversion principle. Using the Swanson-Meindl methodology, a subthreshold current model for accumulation MOSFETs is developed.

In Chapter IV, a circuit simulation methodology is proposed that is applicable to all FET types, uses physically based I-V models, and is applicable over a wide range of supply voltages. The simulation method can be used for rapid optimization, and predictive circuit modeling.

In Chapter V, design guidelines based on 2-D models are presented on how to design an accumulation FET so that it has a low subthreshold swing and SCE. Surface channel accumulation operation is stressed not only to avoid the high SCE resulting from buried channel operation, but also high process sensitivity and high dopant fluctuation effect.

In Chapter VI, metal gate accumulation and inversion FETs are evaluated. It is found that in long channel regime, the Fermi-FET has the minimum subthreshold swing. In the short channel regime, accumulation FETs have higher SCE than inversion FETs if the gate work-function is tunable.

In Chapter VII, poly gate accumulation and inversion FETs are compared. It is found that accumulation FETs achieve a high V_T with a low channel doping; this results in a better channel mobility compared to inversion MOSFETs. Accumulation MOSFETs are also found to have a better subthreshold swing, a lower gate-channel tunneling current, and a lower band-to-band tunneling current at the drain-substrate junction.

In Chapter VIII, Monte Carlo simulations are used to accurately evaluate the performance advantage of mid-bandgap metal gate accumulation MOSFETs. Accumulation MOSFETs are found to perform better than inversion MOSFETs for LOP and HiP applications. In Chapter IX, scaling limits for accumulation and inversion MOSFETs are evaluated. Both 2-D simulations and SCE models are used to derive the minimum channel length. Chapter X presents conclusions and recommendations for future work.

CHAPTER II

THRESHOLD VOLTAGE MODELING

2.1 Threshold Voltage Rolloff of Uniformly Doped SCI MOSFETs

2.1.1 Introduction

MOSFET scaling is progressing into the nanometer range. ITRS 2002 [1] projects a channel length of 18 nm for the year 2010. The junction depth for this technology is projected to be 15 nm. Historically, devices have been of a deep junction type, where $d < y_j$, with d being the minimum channel depletion depth and y_j being the source/drain junction depth. But scaling theory [21] clearly shows that using deep junctions results in increased SCE. So ultra-shallow junction technology is being pursued as one of the key methods of controlling SCE in extremely scaled MOSFETs. In ultra shallow junction MOSFETs, the minimum channel depletion depth is comparable to or greater than the junction depth. Most MOSFET threshold voltage (V_T) models proposed to date only hold good for deep junction devices and do not model the impact of junction depth on SCE [53]-[58]. Charge sharing models were used to obtain the V_T rolloff in [53], and such models become less accurate at short channel lengths. Ratnakumar and Meindl [54] reduced the 2-D Poisson equation to an equivalent Laplace equation and solved it as a series solution. Liu [55] solved the quasi-2-D Poisson equation to obtain the channel potential. Suzuki [56] used a method similar to that in [57] and introduced a channel depletion depth parameter to model the effective depletion depth. Pang [58] used the quasi-2-D Poisson equation to model pocket MOSFETs. A few V_T rolloff models proposed to date include junction depth dependency but are not accurate for intermediate and shallow junction devices [47],[59]-[60]. Pimbley [59] used the variational approach to find V_T rolloff. Nguyen [60] considered cylindrical junctions and the equations are too complex to provide clear physical insight. Austin [47] used an approach similar to that in [54].

Existing circuit simulators mostly use BSIM4 models [61] to model bulk MOSFETs; the BSIM4 models use Liu's [55] V_T rolloff model which holds good only for deep junction devices. The lack of a physically based, junction depth dependent, V_T rolloff model in BSIM4 would necessitate more fitting parameters and more parameter binning for future technology. This in turn would increase the time required to extract device parameters and impede scalability.

A compact V_T model is proposed that holds good for both shallow and deep junction MOSFETs. The model provides a junction depth dependent scale length that can be used to rapidly assess the impact of source/drain scaling on minimum channel length (L_{min}). The model is physically based and is extremely useful for device optimization, technology prediction, and circuit simulation.

First, the solution methodology is introduced along with the boundary conditions used. The proposed model and some previously published models are then compared with 2-D numerical simulations. Finally, a junction depth dependent scale length is proposed, and L_{min} extracted from this scale length is compared with ITRS 2002 projections.

2.1.2 Model

Evanescent mode analysis is used to solve the 2-D Poisson equation. The device analyzed in this section has a uniform doping (N_A) in the channel and substrate. The analysis will be extended to devices with non-uniform doping in the depth direction, like retrograde doped MOSFETs and buried channel MOSFETs, in the next few sections. In the subthreshold regime, the potential distribution in the rectangular solution box (Fig. 2) is given by the 2-D Poisson equation:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = \frac{qN_A}{\epsilon_s} \tag{1}$$

The above equation is solved by the method of superposition [54] by assuming $\psi(x, y) = U(y) + \phi(x, y)$, where $U(y)$ is the solution to the 1-D Poisson equation and $\phi(x, y)$ is the solution to the 2-D Laplace equation. $\phi(x, y)$ can be thought of as the source/drain effect on the overall potential. The boundary conditions for $\psi(x, y)$, $U(y)$, and $\phi(x, y)$ are shown in Tables 3 and 4. $U(y)$ is solved by double integration. $\phi(x, y)$ is solved by separation of

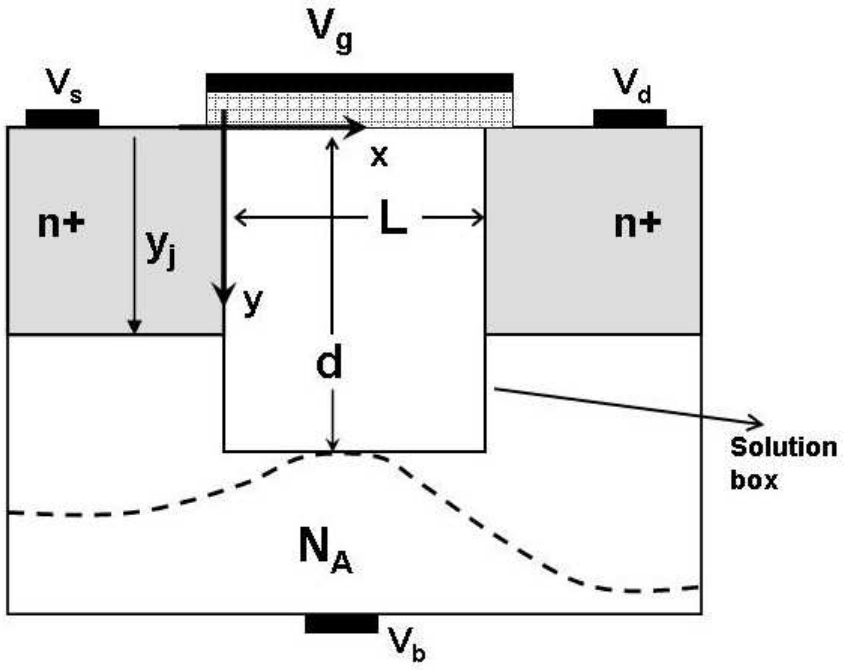


Figure 2: Device structure and solution box.

variables:

$$\phi(x,y) = \left(C^{'} exp(\lambda(x-L)) + D^{'} exp(-\lambda x)\right)(A.sin(\lambda y) + B.cos(\lambda y)) \tag{2}$$

The eigenvalue λ can be obtained from the top and bottom boundary conditions. The top boundary condition gives,

$$\frac{A}{B} = \frac{C_{ox}}{\lambda \epsilon_s} \tag{3}$$

From Table 4, it can be seen that there are two bottom boundary conditions for $\phi(x,y)$.

$$\frac{\partial \phi(x,d)}{\partial y} = 0 \tag{4}$$

$$\phi(x,d) = 0 \tag{5}$$

Only one of (4) or (5) is needed for evaluating λ. Ratnakumar [54], Pimbley [59] and Liu [55] assume the bottom boundary condition to be (4), whereas Frank [62] uses (5). This work clarifies which of the two boundary conditions needs to be used to accurately model V_T rolloff. The corresponding characteristic length equations obtained from the two bottom boundary conditions are

$$\frac{\partial \phi(x,d)}{\partial y} = 0 \Rightarrow tan(\lambda^{'} d) = \frac{C_{ox}}{\lambda^{'} \epsilon_s} \tag{6}$$

Table 3: Boundary conditions for $\psi(x,y)$; the depletion depth at the source/drain is denoted by d_s/d_d, and the built-in potential between the source/drain and substrate is denoted by ψ_{bi}.

Left	$0 \le y \le y_j$ $d_s \ge y \ge yj$	$\psi_{bi} - V_{BS}$ $\psi_{bi}\text{-}V_{BS}+\frac{1}{2}\frac{q}{\epsilon_{si}}N_A(y-y_j)^2$ $-(y-y_j)\sqrt{\frac{2qN_A(\psi_{bi}-V_{BS})}{\epsilon_s}}$
Right	$0 \le y \le y_j$ $d_d \ge y \ge yj$	$\psi_{bi} - V_{BS} + V_{DS}$ $\psi_{bi}\text{-}V_{BS} + V_{DS}+\frac{1}{2}\frac{q}{\epsilon_{si}}N_A(y-y_j)^2$ $-(y-y_j)\sqrt{\frac{2qN_A(\psi_{bi}-V_{BS})}{\epsilon_s}}$
Top	$0 \le x \le L$	$\frac{\partial\psi(x,0)}{\partial y}$ $= \frac{C_{ox}}{\epsilon_s}\left(V_G - V_{FB} - V_{BS} - \psi(x,0)\right)$
Bottom	$0 \le x \le L$	$\psi(x,d) = 0$ $\frac{\partial\psi(x,d)}{\partial y} = 0$

Table 4: Boundary conditions for $U(y)$ and $\phi(x,y)$.

		$U(y)$	$\phi(x,y)$
Left	$0 \le y \le y_j$ $d_s \ge y \ge yj$	- -	$\phi_{bi} - V_{BS}$ $\psi_{bi}\text{-}V_{BS}+\frac{1}{2}\frac{q}{\epsilon_{si}}N_A(y-y_j)^2$ $-(y-y_j)\sqrt{\frac{2qN_A(\psi_{bi}-V_{BS})}{\epsilon_s}}$
Right	$0 \le y \le y_j$ $d_d \ge y \ge yj$	- -	$\psi_{bi} - V_{BS} + V_{DS}$ $\psi_{bi}\text{-}V_{BS} + V_{DS}+\frac{1}{2}\frac{q}{\epsilon_{si}}N_A(y-y_j)^2$ $-(y-y_j)\sqrt{\frac{2qN_A(\psi_{bi}-V_{BS})}{\epsilon_s}}$
Top	$0 \le x \le L$	$\frac{\partial U(0)}{\partial y}$ $= \frac{C_{ox}}{\epsilon_s}\left(V_{GS} - V_{FB} - V_{BS} - U(0)\right)$	$\frac{\partial\phi(x,0)}{\partial y}$ $= \frac{C_{ox}}{\epsilon_s}$
Bottom	$0 \le x \le L$	$U(d) = 0$ $\frac{\partial U(d)}{\partial y} = 0$	$\phi(x,d) = 0$ $\frac{\partial\phi(x,d)}{\partial y} = 0$

$$\phi(x,d) = 0 \Rightarrow tan(\lambda^{''}d) = \frac{-\lambda^{''}\epsilon_s}{C_{ox}} \tag{7}$$

Thus, $\phi(x,y)$ can be written as

$$\phi(x,y) = \sum_{n=1}^{\infty} \left(C_n.exp(\lambda_n(x-L)) + D_n.exp(-\lambda_n x)\right)\left(sin(\lambda_n y) + \frac{\lambda_n \epsilon_s}{C_{ox}} cos(\lambda_n y)\right) \tag{8}$$

λ in (8) can be either $\lambda^{'}$ or $\lambda^{''}$. C_n and D_n are evaluated from the left and right boundary conditions. To enable an analytical solution for V_T rolloff, $\phi(x,y)$ is approximated by a single term solution. This approximation works very well since the terms involving higher order λ's decay rapidly[1]. V_T rolloff is determined as follows: short channel V_T is determined from the condition $V_T = V_G|_{\psi_{min}=2\phi_F}$, where ψ_{min} is the minimum potential, and $\phi_F = \frac{kT}{q} ln\left(\frac{N_A}{n_i}\right)$ is the bulk Fermi level. The short channel V_T is subtracted from the long channel V_T ($V_{T_{long}} = V_G|_{U(0)=2\phi_F}$) to yield the threshold voltage shift, ΔV_T.

Assuming $\lambda_1 L >> 1$ and $x_m = L/2$ (x_m is the x-coordinate of the location of the minimum potential) enables a compact V_T rolloff expression that can provide physical insights.

$$\Delta V_{T_{approx}} = \frac{\left[\chi_1(\psi_{bi}, y_j) + \chi_1(\psi_{bi} + V_{DS}, y_j)\right].exp(-\lambda_1 L/2)}{\kappa_1 - 2exp(-\lambda_1 L/2)} \tag{9}$$

The various terms in the above equation are as shown in Table 5. The assumption, $x_m = L/2$, does not work well when DIBL is high. When V_{DS} is greater than a few tenths of a volt, the exact expression for ΔV_T (shown in Table 5) would have to be used. The approximate expression for ΔV_T, (9), is similar to that in [47]. However, the χ_1 term is different because of the different boundary condition used in this work.

2.1.3 Results and Discussion

Fig. 3 shows the rolloff model using either $\lambda^{'}$ or $\lambda^{''}$, along with 2-D numerical simulations [63], on a V_T vs. junction depth plane. It can be seen that for the deep junction case ($d << y_j$), the ΔV_T model using $\lambda^{'}$ works better than that using $\lambda^{''}$; for ultra shallow junctions ($d >> y_j$), the reverse is true. For cases where $d \approx y_j$, neither the V_T rolloff model using $\lambda^{'}$ nor that using $\lambda^{''}$ gives a good match with simulations.

[1] A multiple term solution would have to be considered if SCE is very high. Such a scenario occurs usually when $\Delta V_T > 200mV$, at which point the device has too high a SCE and ceases to be of interest to the device engineer.

Table 5: Expression for ΔV_T.

All cases	$\Delta V_{T_{exact}} = 2\Gamma.\Delta V_{T_{approx}} \left\{1 + \sqrt{1 - \frac{\chi_1(\psi_{bi},y_j)\chi_1(\psi_{bi}+V_{DS},y_j)}{[\chi_1(\psi_{bi},y_j)+\chi_1(\psi_{bi}+V_{DS},y_j)]\Gamma.\Delta V_T}}\right\}$ $\chi_(V,y_j) = A_1(V) + B_1(V,y_j)$ $A_1(V) = V + V_{FB} - V_{T_{long}} - \frac{qN_A}{\epsilon_s\lambda_1^2}$ $\Gamma = \frac{exp(-\lambda_1 L/2)}{\kappa_1 + 2exp(-\lambda_1 L/2)}$ $\kappa_1 = \frac{1}{2}\left[1 + \frac{C_{ox}d}{\epsilon_s}\left(1 + \left(\frac{\lambda_1\epsilon_s}{C_{ox}}\right)^2\right)\right]$ $d = \sqrt{\frac{2\epsilon_s}{qN_A}2\phi_F} \qquad \phi_F = \frac{kT}{q}ln\left(\frac{N_A}{n_i}\right)$ $V_{T_{long}} = V_{FB} + 2\phi_F + \frac{\sqrt{2q\epsilon_s N_A.2\phi_F}}{C_{ox}}$
$d \leq y_j$ deep junction	$B_1(V,y_j) = \left[V - V_{BS} + \frac{qN_A}{\epsilon_s\lambda_1^2}\right]\left[\frac{\epsilon_s\lambda_1}{C_{ox}}sin(\lambda_1 d) - cos(\lambda_1 d)\right]$
$d > y_j$ shallow junction	$B_1(V,y_j) = \frac{qN_A}{\epsilon_s}\left(-\frac{1}{\lambda_1^2}\right)\left[cos(\lambda_1 y_j) - \frac{\epsilon_s}{C_{ox}}sin(\lambda_1 y_j)\right]$ $+ \left[\frac{qN_A(d_{s/d}(V)-y_j)}{\epsilon_s\lambda_1}\right]\left[sin(\lambda_1 y_j) + \frac{\epsilon_s\lambda_1}{C_{ox}}cos(\lambda_1 y_j)\right]$ $+ \left[\frac{qN_A(d_{s/d}-d)^2}{2\epsilon_s}\right]\left[cos(\lambda_1 d) - \frac{\epsilon_s\lambda_1}{C_{ox}}sin(\lambda_1 d)\right]$ $- \left[\frac{qN_A(d-d_{s/d}(V)}{\epsilon_s\lambda_1}\right]\left[\frac{\epsilon_s\lambda_1}{C_{ox}}cos(\lambda_1 d) + sin(\lambda_1 d)\right]$ $d_{s/d}(V) = y_j + \sqrt{\frac{2\epsilon_s}{qN_A}(V - V_{BS})}$

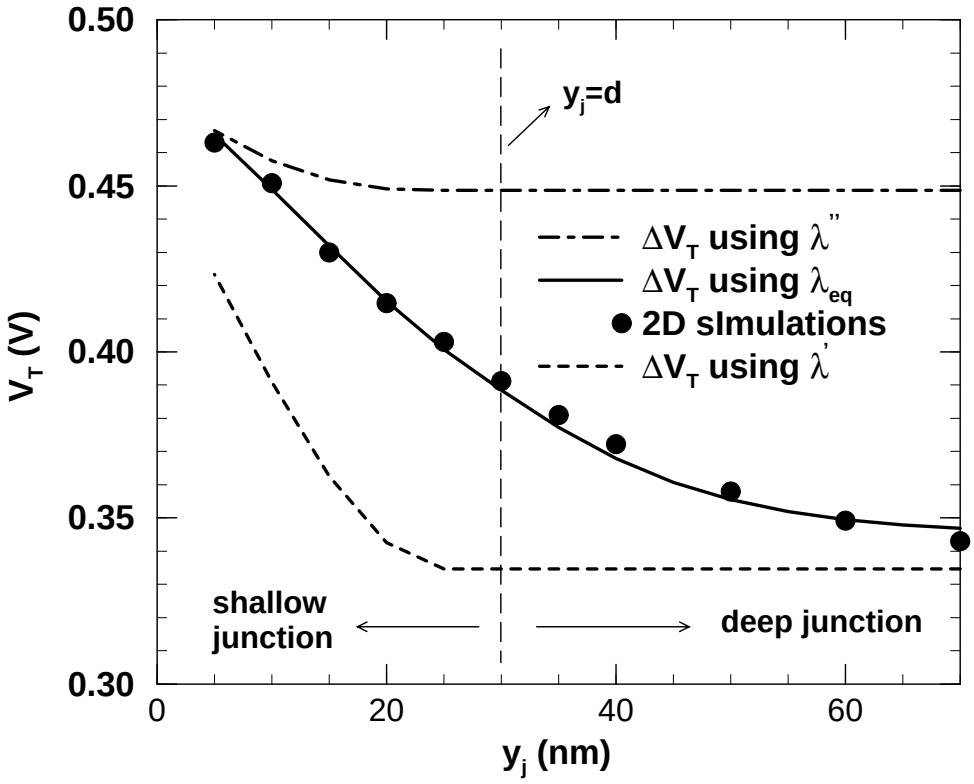


Figure 3: Threshold voltage versus junction depth. $L = 65\ nm$, N_A =2x10^{18} cm^{-3}, $t_{ox} = 1.5\ nm$, $V_{DS} = 1.0\ V$. The long channel depletion depth for this device is 30 nm.

The physical explanation for such a behavior is as follows. For the deep junction case, assuming (5) as the bottom boundary would mean that the source/drain potential is forced to go to zero at $y = d$. In reality, the source (drain) potential is fixed at $y = d$, and is equal to ψ_{bi} ($\psi_{bi}+V_{DS}$); here, ψ_{bi} is the built-in potential between the source/drain and substrate. Assuming (5) as the bottom boundary would underestimate the effect of source/drain on channel potential and thus would underestimate V_T rolloff. Near the source/drain, at $y = d$, the field in the y-direction is zero. Assuming (4) as the bottom boundary causes the source/drain potential to be fixed to $\psi_{bi}/(\psi_{bi}+V_{DS})$ at $y = d$, and thus is a better boundary condition to apply. For the shallow junction case, assuming (5) as the bottom boundary is more accurate since the source/drain potential would have almost decayed to zero at $y = d$. If (4) is assumed as the bottom boundary, the source/drain potential would be fixed to $\psi_{bi}/(\psi_{bi} + V_{DS})$ at $y = d$, and this would cause an overestimation of V_T rolloff.

From Fig. 3, it can be seen that when $d = y_j$, V_T is roughly the average of the values given by $\lambda^{'}$ and $\lambda^{''}$. For intermediate junction cases ($d \approx y_j$), a superposition of $\lambda^{'}$ and $\lambda^{''}$ would be needed. A function is proposed that does the necessary superposition:

$$\lambda_{eq} = (\lambda^{'} - \lambda^{''}).\frac{1}{1 + exp\left(\frac{d-y_j}{d/2}\right)} + \lambda^{''} \tag{10}$$

It can be seen that when $y_j >> d$, $\lambda_{eq} \approx \lambda^{'}$ and when $y_j << d$, $\lambda_{eq} \approx \lambda^{''}$. When $y_j = d$, λ_{eq} is the average of $\lambda^{'}$ and $\lambda^{''}$. When V_T obtained by using λ_{eq} is plotted against y_j, Fig. 3, it can be seen that the model now agrees well with simulation results over all y_j. When $d \gg t_{ox}$, (10) can be simplified by noting that $\lambda_n^{''} \approx (\pi + (n-1)\pi)/(d + 3t_{ox})$, and $\lambda_n^{'} \approx (\pi/2 + (n-1)\pi)/(d + 3t_{ox})$. It is to be noted that the proposed ΔV_T model only uses the primary eigen value (λ_{eq_1}).

In the proposed model, the minimum channel depletion depth value used is that of a long channel MOSFET. When the source/drain depletion depth is comparable to channel length, the average depletion depth in the channel is considerably greater than the long channel depletion depth. In such a situation, the proposed model will not work well. But in all well-scaled devices, the channel length is considerably greater than the source/drain depletion depth.

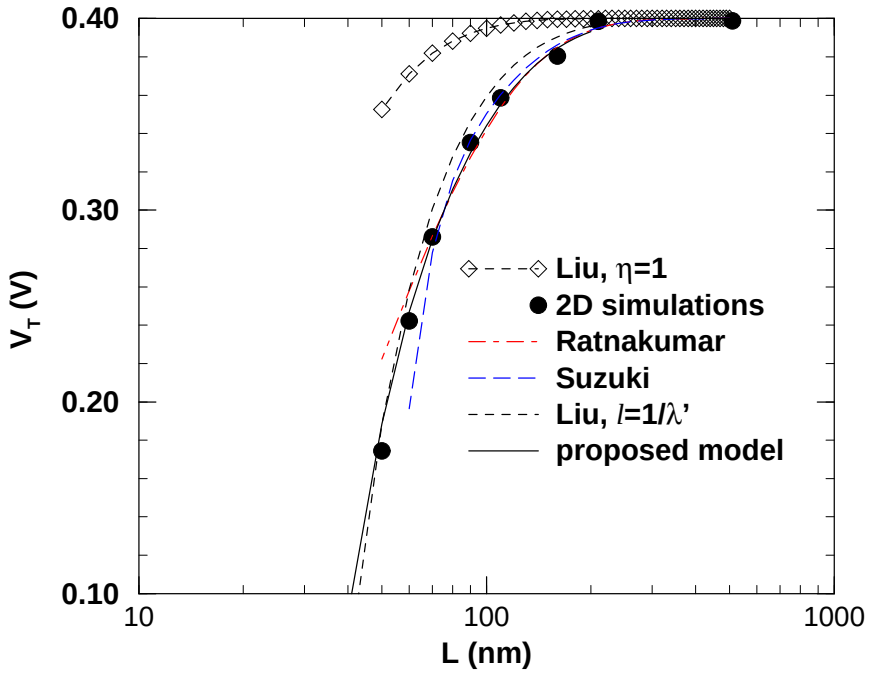


Figure 4: V_T versus L for a deep junction device; $y_j = 70\ nm$, N_A =1x10^{18} cm^{-3}, $t_{ox} = 1.5\ nm$, $V_{DS} = 0.05\ V$. The minimum channel depletion depth for this device is 35 nm.

Fig. 4 shows V_T vs. L for the deep junction case. For comparison, models of Ratnakumar, Liu and Suzuki are also shown. Ratnakumar's and Suzuki's models agree well with simulation results. Liu's model, with $\eta = 1$, severely underestimates rolloff. If $\lambda^{'}$ is used in Liu's model instead of $1/l$, then a good match is seen with the 2-D simulations. Also shown in the figure is the proposed model, which agrees well with simulations.

Fig. 5 shows V_T vs. L for the shallow junction case. It can be seen that the proposed model agrees well with simulations. Ratnakumar, Liu and Suzuki assume a deep junction in their models and their models cannot predict the effect of y_j on V_T. So a poor match is seen between their models and simulations.

Fig. 6 shows the proposed model versus simulations for various y_j, in the V_T vs. L plane. It can be seen that for all cases, the model agrees well with simulations. λ_{eq_1} can be used as a junction depth dependent scale length. Fig. 7 shows the universal V_T rolloff curve generated by plotting rolloff versus $L.\lambda_{eq_1}$. The rolloff data is generated for various N_A, y_j and L. It can be seen that all points lie on a common curve, indicating the fact

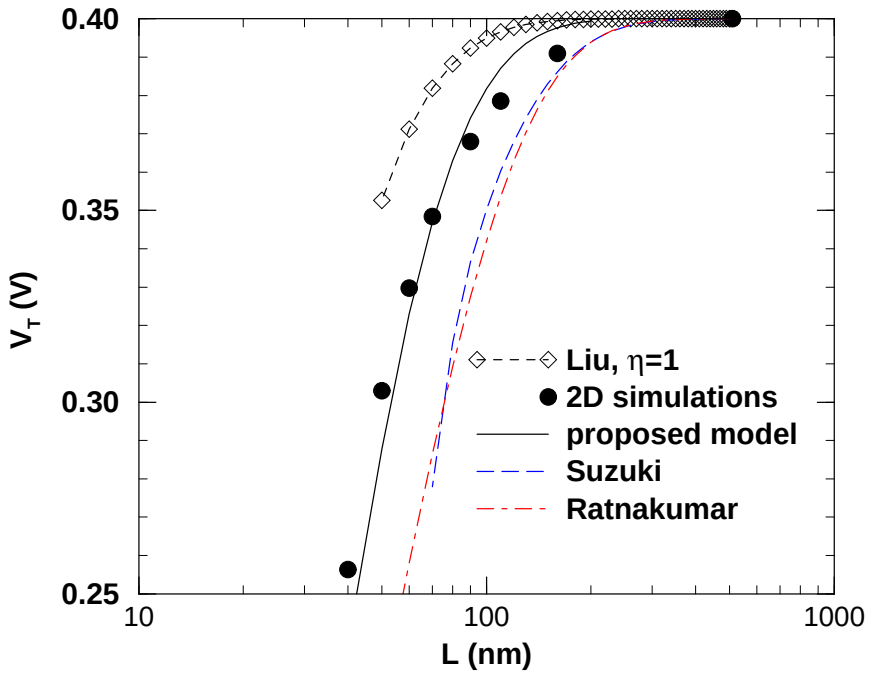


Figure 5: V_T versus L for a shallow junction device; $y_j = 20\ nm$, N_A =1x10^{18} cm^{-3}, $t_{ox} = 1.5\ nm$, $V_{DS} = 0.05\ V$. The minimum channel depletion depth for this device is 35 nm.

that λ_{eq_1} can be used as a junction depth dependent scale length. This scale length can be used to rapidly analyze the minimum channel length for a given t_{ox}, y_j and N_A. The relationship between L_{min} and λ_{eq_1} is calibrated using ITRS 2002 parameters for the year 2002. It is found that $L_{min} = 3.46/\lambda_{eq_1}$. Using this relation, L_{min} is extracted for various technology generations from 2003-2007 and compared with the ITRS projections, Fig. 8. A close correlation is seen between L_{min} predictions by the model, and ITRS projections.

2.2 Threshold Voltage Rolloff of Accumulation MOSFETs

2.2.1 Introduction

Most previous work on accumulation MOSFET modeling have been limited to 1-D analyses [64]-[65]. A threshold voltage (V_T) rolloff model for buried-channel MOSFETs is proposed in [66], but uses fit parameters, and thus cannot be used for predictive modeling; also, charge sharing principle is used to model drain-induced barrier lowering, and the method is not accurate for short channel lengths. Austin [47] proposed a source/drain junction depth dependent V_T rolloff model for bulk accumulation MOSFETs, but the model is not accurate

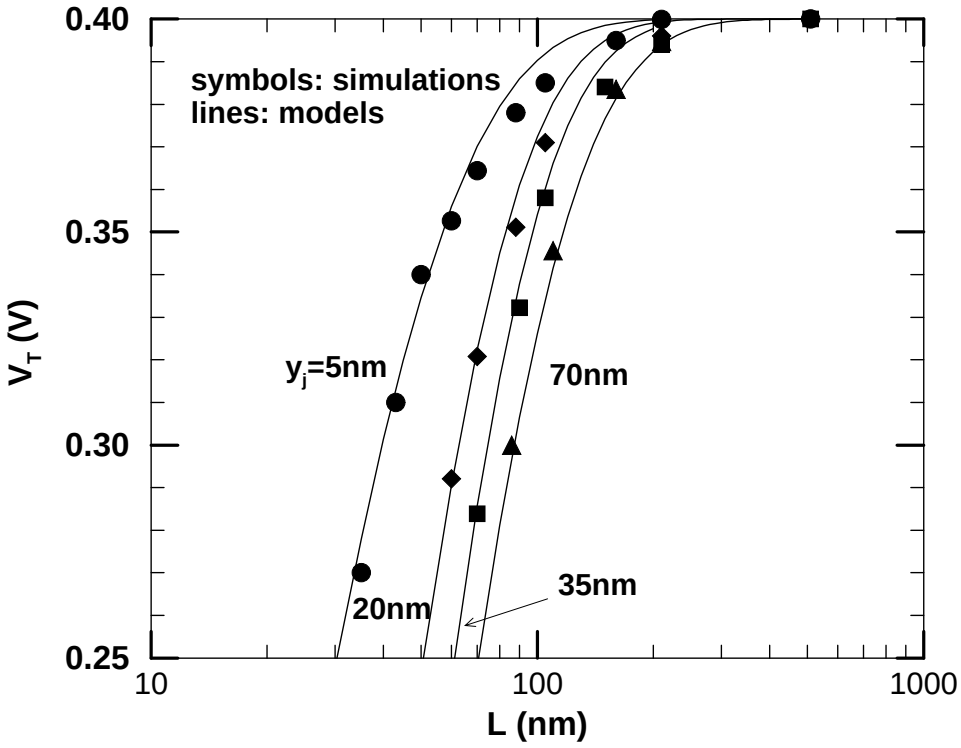


Figure 6: V_T vs. L for different junction depths; N_A =1x10^{18} cm^{-3}, t_{ox} = 1.5 nm, $V_{DS} = 1.0\ V$. The minimum channel depletion depth for this device is 35 nm.

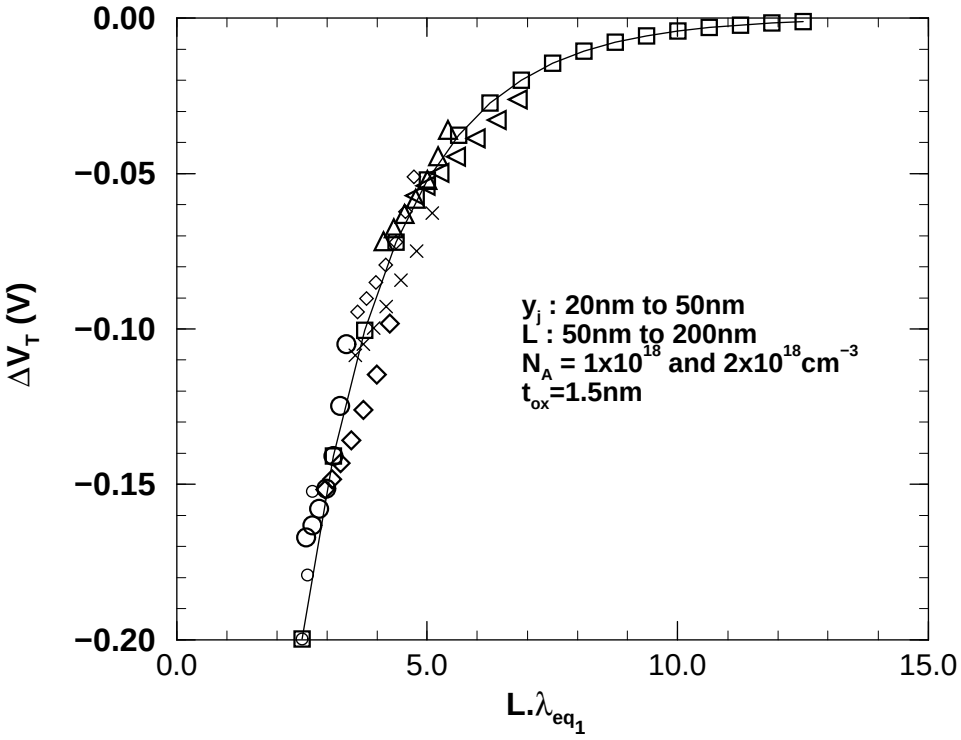


Figure 7: V_T vs. $L.\lambda_{eq_1}$ for various device parameters.

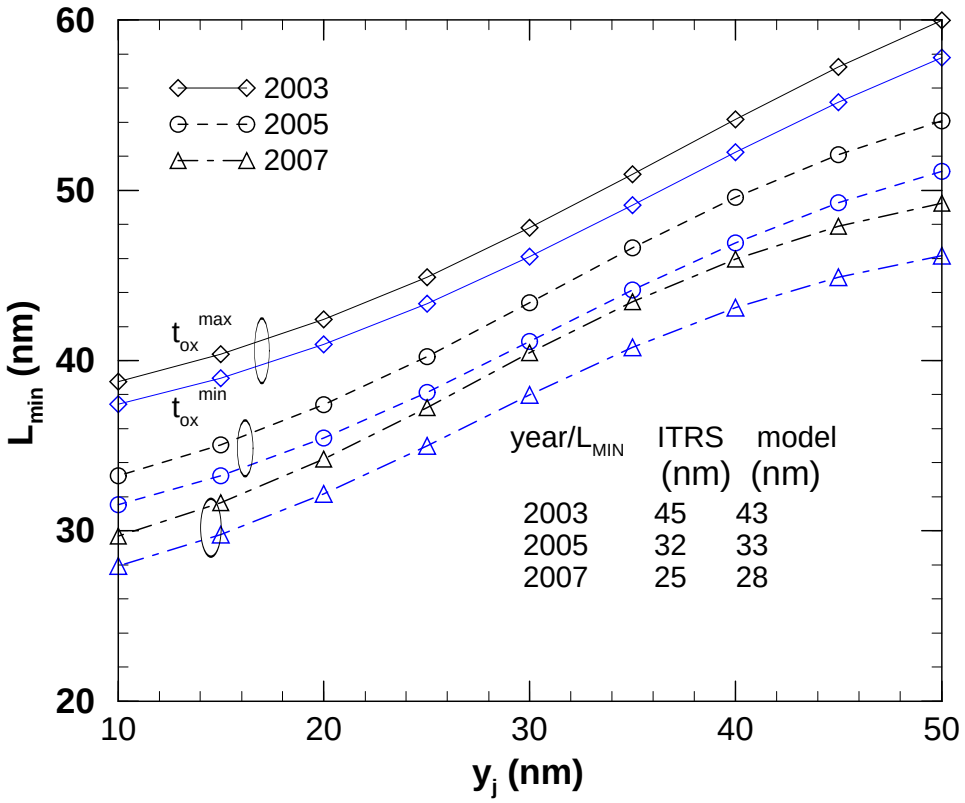


Figure 8: L_{min} projection using λ_{eq}. t_{ox}^{min} and t_{ox}^{max} values are taken from ITRS 2002 recommendations for the particular generation.

for intermediate and shallow junction devices. An accurate, source/drain junction depth dependent, V_T rolloff model for bulk accumulation MOSFETs is proposed, and verified against 2-D simulations.

2.2.2 Long Channel Threshold Voltage

The long-channel threshold voltage ($V_{T_{long}}$) of accumulation MOSFETs has been derived previously in various works [47], [64]. But existing $V_{T_{long}}$ models do not agree well with simulations. Thus, there is a need to look at the derivation of $V_{T_{long}}$ for accumulation MOSFETs. The device analyzed is shown in Fig. 9. In the subthreshold regime, the 1-D potential in the channel depth direction is given by the 1-D Poisson equation:

$$\begin{aligned} \frac{\partial^2 U}{\partial y^2} &= \frac{qN_D}{\epsilon_s}; \qquad 0 < x < L,\ 0 < y < y_i \\ &= \frac{qN_A}{\epsilon_s}; \qquad 0 < x < L,\ y_i < y < d \end{aligned} \tag{11}$$

L is the effective channel length, and y_i is the tub-depth. $U(y)$ is solved by double integration:

$$\begin{aligned} U(y) &= -\frac{qN_D}{2\epsilon_s}(y - y_i)^2 - \frac{qN_A}{\epsilon_s}(d - y_i)y + \frac{qN_A}{2\epsilon_s}(d^2 - y_i^2) \quad 0 \le y \le y_i \\ &= \frac{qN_A}{2\epsilon_s}(y - d)^2 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad y_i \le y \le d \end{aligned} \tag{12}$$

The expression for the minimum channel depletion depth (d) is shown in Table 6. The long channel threshold voltage ($V_{T_{long}}$) is obtained from $U(y)$ by finding the gate voltage that gives $U(y_{ch}) = 2\phi_F$, where $\phi_F = \frac{kT}{q} ln\left(\frac{N_A}{n_i}\right)$ is the bulk Fermi level. For SCA MOSFETs,

$$V_{T_{long}} = V_{FB} + 2\phi_F - \frac{q(N_A + N_D)y_i}{C_{ox}} + \frac{\sqrt{2q\epsilon_s N_A}}{C_{ox}}\left[2\phi_F + \frac{qN_A}{2\epsilon_s}\left(1 + \frac{N_D}{N_A}\right)y_i^2\right]^{1/2} \tag{13}$$

where V_{FB} is the flat-band voltage, and $C_{ox} = \epsilon_{ox}/t_{ox}$ is the oxide capacitance. For BCA MOSFETs,

$$\begin{aligned} V_{T_{long}} = V_{FB} + \frac{N_D}{N_D + N_A}2\phi_F + \left(\frac{1}{C_{ox}}\right)\left(\sqrt{\frac{2\epsilon_s N_A N_D}{N_D + N_A}2\phi_F} - qN_D y_i\right) \\ + \left(\frac{\epsilon_s}{y_i}\right)\left(\sqrt{\frac{2\epsilon_s N_A N_D}{N_D + N_A}2\phi_F} - \frac{qN_D y_i}{2}\right) \end{aligned} \tag{14}$$

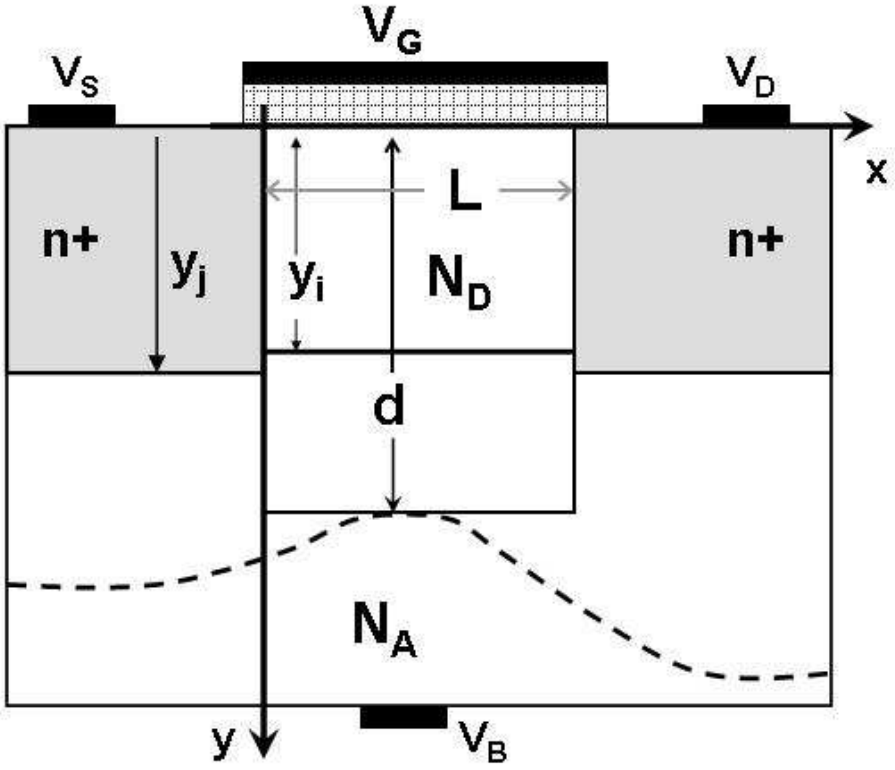


Figure 9: Accumulation n-FET structure and terminology.

For the Fermi MOSFET (which is the boundary between the BCA and SCA FET), $V_{T_{long}}$ is simply given by $V_{FB} + 2\phi_F$. For SCA MOSFETs (and inversion MOSFETs), the vertical channel location (y_{ch}) is zero. For BCA MOSFETs

$$y_{ch} = \left(1 + \frac{N_A}{N_D}\right) y_i - \frac{N_A}{N_D} d \tag{15}$$

In [47], $V_{T_{long}}$ is defined to be the gate voltage that gives $U(y_{ch}) = \phi_{bi}$; $\phi_{bi} = \frac{kT}{q} ln\left(\frac{N_A N_D}{ni^2}\right)$ is the built-in potential between the tub and the substrate. This definition results in a non-monotonic dependency of $V_{T_{long}}$ on channel doping, Fig. 10. It can be seen that the $V_{T_{long}}$ model based on the definition in [47] does not agree well with 1-D simulation results. On the other hand, the proposed $V_{T_{long}}$ definition ($V_T = V_G|_{U(y_{ch})=2\phi_F}$) agrees well with simulation results[2], Figs. 10 and 11. It can be seen that as y_i increases, the mismatch between the $V_{T_{long}}$ model and simulations increases. As y_i increases, the electrons spread more into the depth direction; an integration of the inversion charge in the depth direction would then

[2]In 2-D simulation results, V_T is defined as the gate voltage required to give a drain current of 0.1 $\mu A/\mu m$xW/L.

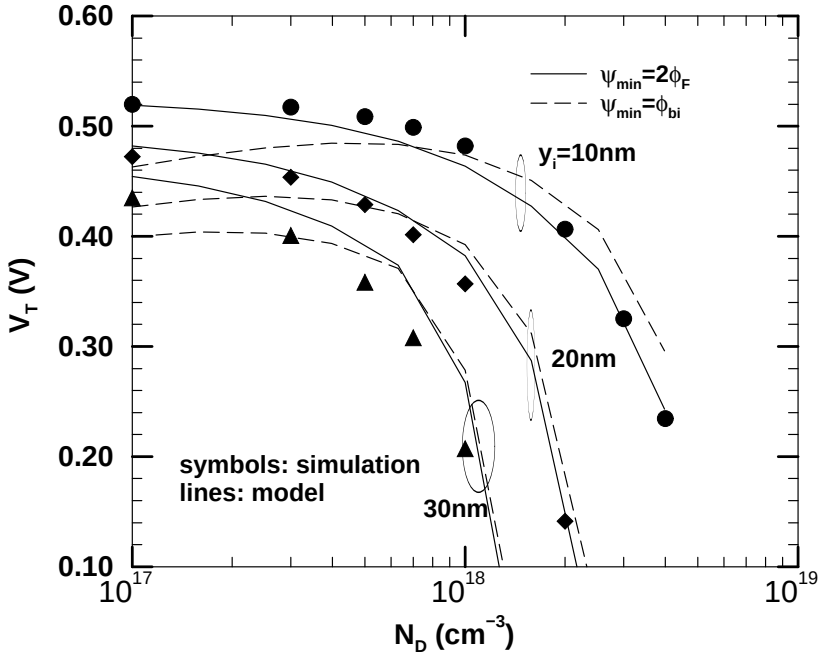


Figure 10: Verification of long channel threshold voltage model with 1-D simulations; N_A =7x10^{17} cm^{-3}, and $t_{ox} = 1.5$ nm.

give an accurate estimate of $V_{T_{long}}$. Appendix A shows how to extend the threshold voltage model to devices with deep tub-depths, and high SCE ($\Delta V_T > 200$ mV). It will be seen in later sections that for $y_i > 30$ nm, SCE is very high in accumulation FETs; thus the proposed compact model is sufficient for devices analyzed in this work.

2.2.3 Threshold Voltage Rolloff Model

The device analyzed is shown in Fig. 9. In the subthreshold region, the tub (N_D) and part of the substrate (N_A) are depleted of mobile carriers; the potential distribution in the depleted rectangular region is given by the 2-D Poisson equation:

$$\begin{aligned} \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} &= \frac{qN_D}{\epsilon_s}; \qquad 0 < x < L,\ 0 < y < y_i \\ &= \frac{qN_A}{\epsilon_s}; \qquad 0 < x < L,\ y_i < y < d \end{aligned} \tag{16}$$

L is the effective channel length, and y_i is the tub-depth. (16) is solved by the method of superposition [54]:

$$\psi(x, y) = U(y) + \phi(x, y) \tag{17}$$

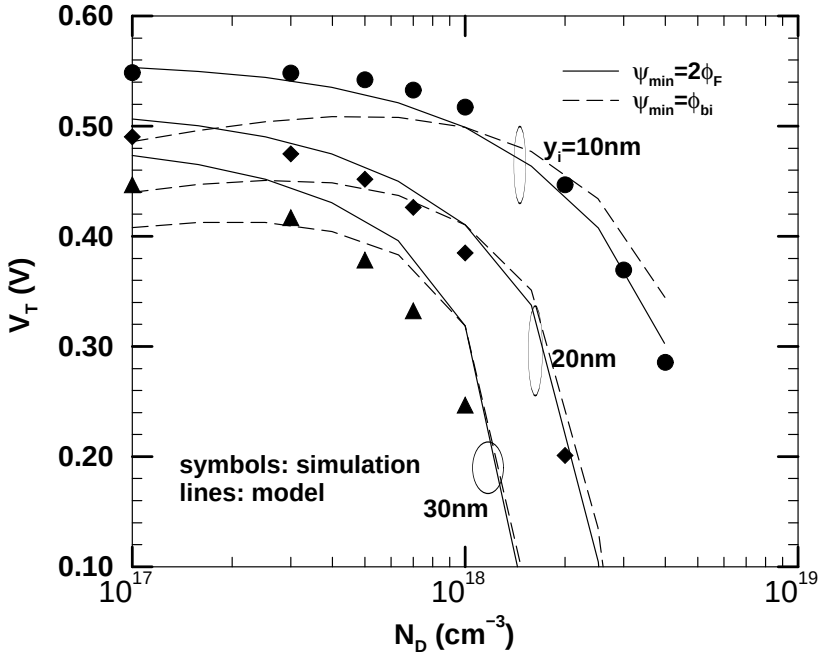


Figure 11: Verification of long channel threshold voltage model with 1-D simulations; N_A =1x10^{18} cm^{-3}, and $t_{ox} = 1.5$ nm.

where $U(y)$ is the solution to the 1-D Poisson equation and $\phi(x, y)$ is the solution to the 2-D Laplace equation. $\phi(x, y)$ can be thought of as the source/drain effect on the overall potential. $\phi(x, y)$ is solved by the method of separation of variables:

$$\phi(x,y) = \sum_{n=1}^{\infty} (C_n.exp(\lambda_n(x-L)) + D_n.exp(-\lambda_n x)) \left(sin(\lambda_n y) + \frac{\lambda_n \epsilon_s}{C_{ox}} cos(\lambda_n y) \right) \quad (18)$$

C_n and D_n are evaluated from the boundary conditions, and are as shown in Table 6. The two possible bottom boundary conditions for $\psi(x, y)$ result in two different eigenvalues, $\lambda^{'}$, and $\lambda^{''}$. $\lambda^{'}$ and $\lambda^{''}$ are superimposed to obtain λ, Table 6.

From the 2-D potential, an expression for the short-channel threshold voltage shift (ΔV_T) is obtained as follows: the minimum potential point (x_m, y_{ch}) is calculated from $\frac{\partial \psi(x_m, y_{ch})}{\partial x} = 0$; then the gate voltage that gives $\psi(x_m, y_{ch}) = 2\phi_F$ is the threshold voltage of the device. The difference between this threshold voltage and the long channel threshold voltage ($V_{T_{long}}$) is the short-channel threshold voltage shift (ΔV_T). To enable an analytical solution for V_T rolloff, $\phi(x, y)$ is approximated by a single-term solution. This approximation works very well since the terms involving higher order λ's decay rapidly. The expression for

ΔV_T is

$$\Delta V_T = 2\Gamma.\Theta \left\{ 1 + \sqrt{1 - \frac{\chi_1(\psi_{bi}, y_j)\chi_1(\psi_{bi} + V_{DS}, y_j)}{[\chi_1(\psi_{bi}, y_j) + \chi_1(\psi_{bi} + V_{DS}, y_j)]\Gamma.\Theta}} \right\} \tag{19}$$

where the various terms are shown in Tables 6 and 7. The tables also show how to extend the threshold voltage rolloff model to uniformly doped (UD) SCI MOSFETs, and retrograde doped (RD) SCI FETs. Fig. 12 compares the proposed threshold voltage rolloff model with 2-D simulations [67]. The model is seen to work well in predicting threshold voltage rolloff of both SCA and BCA MOSFETs.

2.3 Summary

A compact threshold voltage rolloff model has been proposed that accurately models the effect of junction depth on threshold voltage rolloff. The model is physically based, and is useful for device optimization, technology prediction, and circuit simulation. The model provides a scale length which can be used to rapidly analyze the impact of junction depth scaling on minimum channel length. The model is applicable to both inversion and accumulation MOSFETs, and thus allows easy comparison between the two MOSFET types.

Table 6: V_T rolloff expression for accumulation and inversion FETs.
[UD-SCI: $N_T = -N_A$, $N_W = N_A$, $y_i = 0$; RD-SCI: $N_T = -N_A^-$, $N_W = N_A$; SCA/BCA/FF: $N_T = N_D$, $N_W = N_A$]

All Cases:	$$\mathsf{D_n} = \frac{\mathsf{A_n}(\psi_{\mathsf{bi_S}}) + \mathsf{B_n}(\psi_{\mathsf{bi_S}}, \mathsf{y_j})}{\frac{\epsilon_\mathsf{s}\lambda_\mathsf{n}}{\mathsf{C_{ox}}}\mathcal{K}_\mathsf{n}}$$ $$\mathsf{C_n} = \frac{\mathsf{A_n}(\psi_{\mathsf{bi_D}} + \mathsf{V_{DS}}) + \mathsf{B_n}(\psi_{\mathsf{bi_D}} + \mathsf{V_{DS}}, \mathsf{y_j})}{\frac{\epsilon_\mathsf{s}\lambda_\mathsf{n}}{\mathsf{C_{ox}}}\mathcal{K}_\mathsf{n}}$$ $$\mathsf{tan}(\lambda^{'}\mathsf{d}) = \frac{\mathsf{C_{ox}}}{\lambda^{'}\epsilon_\mathsf{s}}; \qquad \mathsf{tan}(\lambda^{''}\mathsf{d}) = \frac{-\lambda^{''}\epsilon_\mathsf{s}}{\mathsf{C_{ox}}}$$ $$\lambda_\mathsf{n} = (\lambda_\mathsf{n}^{'} - \lambda_\mathsf{n}^{''}).\frac{1}{1 + \exp\left(\frac{\mathsf{d}-\mathsf{y_j}}{\mathsf{d}/2}\right)} + \lambda_\mathsf{n}^{''}$$ $$\Theta = \frac{-[\mathcal{X}_1(\Psi_{\mathsf{bi_S}}, \mathsf{y_j}) + \mathcal{X}_1(\Psi_{\mathsf{bi_D}} + \mathsf{V_{DS}}, \mathsf{y_j})]\exp(-\frac{\lambda_1\mathsf{L}}{2})\mathsf{F}(\mathsf{y_{ch}})}{\mathcal{K}_1 - 2\exp(-\frac{\lambda_1\mathsf{L}}{2})\mathsf{F}(\mathsf{y_{ch}})}$$ $$\Gamma = \frac{exp(-\lambda_1 L/2)F(y_{ch})}{\kappa_1 + 2exp(-\lambda_1 L/2)F(y_{ch})}$$ $$\mathcal{X}_1(V, y_j) = \mathsf{A_1(V)}\mid_{\mathsf{V_{GS}}=\mathsf{V_{T_{long}}}} + \mathsf{B_1(V, y_j)}$$ $$\mathsf{A_1(V)} = \mathsf{V} + \mathsf{V_{FB}} - \mathsf{V_{GS}} - \frac{\mathsf{qN_T}}{\epsilon_\mathsf{s}\lambda_1^2} + \left[\frac{\mathsf{q(N_W + N_T)}}{\epsilon_\mathsf{s}\lambda_1^2}\right]\left[\cos(\lambda_1\mathsf{y_i}) - \frac{\epsilon_\mathsf{s}\lambda_1}{\mathsf{C_{ox}}}\sin(\lambda_1\mathsf{y_i})\right]$$ $$\Psi_{\mathsf{bi_{S/D}}} = \frac{\mathsf{kT}}{\mathsf{q}}\ln\frac{\mathsf{N_W N_{D_{S/D}}}}{\mathsf{n_i^2}}$$ $$\mathsf{F(y_{ch})} = \cos(\lambda_1\mathsf{y_{ch}}) + \frac{\mathsf{C_{ox}}}{\lambda_1\epsilon_\mathsf{s}}\sin(\lambda_1\mathsf{y_{ch}});$$ $$\mathcal{K}_1 = \frac{1}{2}\left[1 + \frac{\mathsf{C_{ox}d}}{\epsilon_\mathsf{s}}\left(1 + \left(\frac{\lambda_1\epsilon_\mathsf{s}}{\mathsf{C_{ox}}}\right)^2\right)\right]$$ $$\mathsf{d} = -\frac{\epsilon_\mathsf{s}}{\mathsf{C_{ox}}} + \left[\frac{2\epsilon_\mathsf{s}}{\mathsf{qN_W}}(\mathsf{V_G} - \mathsf{V_{FB}} - \mathsf{V_B}) + \left(1 + \frac{\mathsf{N_T}}{\mathsf{N_W}}\right)\left(\mathsf{y_i^2} + \frac{2\epsilon_\mathsf{s}}{\mathsf{C_{ox}}}\mathsf{y_i}\right) + \left(\frac{\epsilon_\mathsf{s}}{\mathsf{C_{ox}}}\right)^2\right]^{1/2}$$

Table 7: V_T roll-off for accumulation and inversion devices.

$y_j > y_i;\quad d < y_j$	$B_1(V, y_j) = \left[V - V_{BS} + \frac{qN_W}{\epsilon_s \lambda_1^2}\right]\left[\frac{\epsilon_s \lambda_1}{C_{ox}} \sin(\lambda_1 d) - \cos(\lambda_1 d)\right]$
$y_j > y_i;\quad d > y_j$	$\begin{aligned} B_1(V, y_j) &= -\frac{qN_W}{\epsilon_s \lambda_1^2}\left[\cos(\lambda_1 y_j) - \frac{\epsilon_s \lambda_1}{C_{ox}} \sin(\lambda_1 y_j)\right] \\ &+ \left[\frac{qN_W(d_{s/d}(V) - y_j)}{\epsilon_s \lambda_1}\right]\left[\sin(\lambda_1 y_j) + \frac{\epsilon_s \lambda_1}{C_{ox}} \cos(\lambda_1 y_j)\right] \\ &- \left(\frac{qN_W(d_s - d)^2}{2\epsilon_s}\right)\left[\cos(\lambda_1 d) - \frac{\epsilon_s \lambda_1}{C_{ox}} \sin(\lambda_1 d)\right] \\ &+ \left[\frac{qN_W(d - d_{s/d}(V))}{\epsilon_s \lambda_1}\right]\left[\frac{\epsilon_s \lambda_1}{C_{ox}} \cos(\lambda_1 d) + \sin(\lambda_1 d)\right] \\ d_{s/d}(V) &= y_j + \sqrt{\frac{2\epsilon_s}{qN_W}(V - V_{BS})} \end{aligned}$
$y_j < y_i;\quad d > y_j$	$\begin{aligned} B_1(V, y_j) &= \frac{qN_T}{\epsilon_s \lambda_1^2}\left[\cos(\lambda_1 y_j) - \frac{\epsilon_s \lambda_1}{C_{ox}} \sin(\lambda_1 y_j)\right] \\ &+ \frac{qN_T}{\epsilon_s \lambda_1}\left[\frac{N_W(d_s - y_i)}{N_T} + y_j - y_i\right] \\ &\qquad \left[\sin(\lambda_1 y_j) + \frac{\epsilon_s \lambda_1}{C_{ox}} \cos(\lambda_1 y_j)\right] \\ &- \frac{q(N_W + N_T)}{\epsilon_s \lambda_1^2}\left[\cos(\lambda_1 y_i) - \frac{\epsilon_s \lambda_1}{C_{ox}} \sin(\lambda_1 y_i)\right] \\ &- \left(\frac{qN_W(d_s - d)^2}{2\epsilon_s}\right)\left[\cos(\lambda_1 d) - \frac{\epsilon_s \lambda_1}{C_{ox}} \sin(\lambda_1 d)\right] \\ &+ \left[\frac{qN_W(d - d_{s/d}(V))}{\epsilon_s \lambda_1}\right]\left[\frac{\epsilon_s \lambda_1}{C_{ox}} \cos(\lambda_1 d) + \sin(\lambda_1 d)\right] \\ d_{s/d}(V) &= y_j + \sqrt{\left(1 + \frac{N_T}{N_W}\right)(y_j - y_i)^2 + \frac{2\epsilon_s}{qN_W}(V - V_{BS})} \end{aligned}$

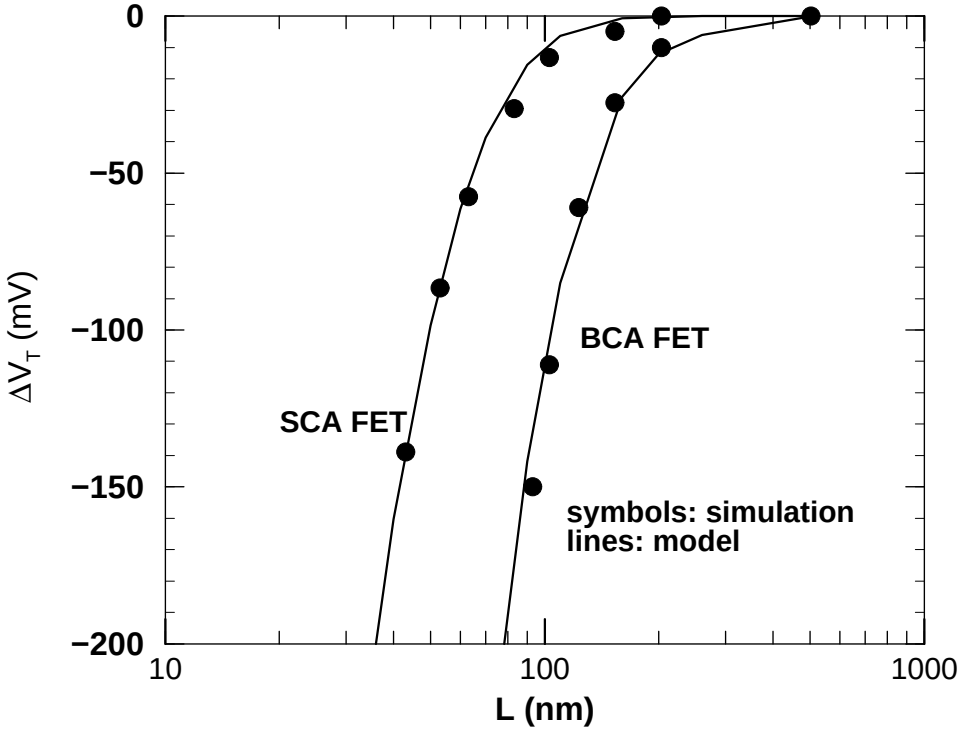


Figure 12: Threshold voltage rolloff for SCA and BCA MOSFETs: model vs. 2-D simulations; $L = 45\ nm$, $t_{ox} = 15$ Å, $y_j = 12\ nm$, and $V_{DS} = 1.0\ V$ for both SCA, and BCA MOSFETs. For the SCA FET, N_A =2x10^{18} cm^{-3}, N_D =1x10^{17} cm^{-3}, and $y_i = 10\ nm$. For the BCA MOSFET, N_A =1x10^{18} cm^{-3}, N_D =2x10^{18} cm^{-3}, and $y_i = 30\ nm$.

CHAPTER III

SUBTHRESHOLD CONDUCTION MODELING

3.1 Introduction

Static power in gigascale systems is becoming increasingly important, especially in portable systems. Subthreshold swing (S) is an indicator of how well the device can be turned off, and is important in predicting the subthreshold current. Existing models for S in the literature are either semi-empirical [61], [68] or when physical, are applicable only for relatively longer channel lengths [47]. Also, existing S models do not accurately model junction depth effects. An accurate subthreshold swing model is proposed that holds good for both accumulation and inversion MOSFETs. Existing off-current models for accumulation MOSFETs use fit parameters [69]-[70], and thus, are not suitable for predictive modeling. A subthreshold current model for accumulation MOSFETs is proposed based on the Swanson-Meindl approach for inversion MOSFETs.

3.2 Subthreshold Swing Model

Subthreshold slope is defined as

$$S^{-1} = \frac{\partial log(I_D)}{\partial V_g} \tag{20}$$

in the subthreshold region ($V_G < V_T$). The drain current I_D is dependent on the minimum potential point in the channel (assuming a charge sheet). Thus, $I_D \propto n$, where carrier concentration n is given by $n = e^{\beta(\psi_{min} - \phi_{Fn})}$. $\psi_{min}(y)$ is the minimum potential, and can be extracted from the 2-D potential derived earlier in this work; ϕ_{Fn} is the quasi-Fermi level of electrons (constant in the depth direction). Substituting n and I_D, in (20)

$$S^{-1} = \frac{\partial \psi_{min}}{\partial V_G} \tag{21}$$

Comparison of the above model for S with 2-D simulations [63] is shown in Fig. 13, for an RD-SCI device. It can be seen that the point-derivative model fails to capture the steep

subthreshold swing rollup of short channel length MOSFETs.

Fig. 14 plots electron concentration vs. distance from the $\mathsf{Si}-\mathsf{SiO_2}$ interface. Defining the inversion layer thickness (ILT) as the depth at which the electron concentration has decayed by two orders of magnitude from its peak value, it can be seen that the 100 nm device has an ILT of 2nm, while the 40 nm device has an ILT of 4nm; the 30 nm device has an ILT of around 7 nm. It can be seen that as the channel length is decreased, even the classical distribution of electrons occupies a reasonable depth below the semiconductor-oxide interface, and the assumption of inversion layer thickness to be zero is invalid. This is the same situation that Chen et al [71] faced in their work on double gate (DG) FET modeling - the DG-FETs have volume inversion and thus the electron concentration has to be integrated over the device depth to calculate S. A similar approach can be adopted for bulk devices.

When the inversion (or accumulation) layer thickness is assumed to be non-zero, the drain current in the subthreshold region is no more dependent on the carrier concentration at a single point; rather, it is now a function of the carrier concentration over the inversion layer thickness. Since the exact inversion layer thickness is not known, it is better to make the drain current a function of carrier concentration over the depletion depth:

$$I_D \propto \int_0^d n.dy \tag{22}$$

Substituting (22) in (20),

$$S^{-1} = \frac{\int_0^d n.\frac{d\psi_{min}}{dV_g}.dy}{\int_0^d n.dy}.\frac{\beta}{ln(10)} \tag{23}$$

(23) reduces to (21) as $d \to 0$. When this new model is used to calculate S and compared with 2-D simulations, Fig. 13, a good match can be seen between the model and simulations. Fig. 15 compares the subthreshold swing model for BCA/SCA MOSFETs with 2-D simulations; a favorable match is seen between the model and simulations.

3.3 Subthreshold Conduction Model

An expression for the subthreshold current of long-channel accumulation FETs is obtained by following the analysis of Swanson and Meindl [72] for bulk inversion MOSFETs. The

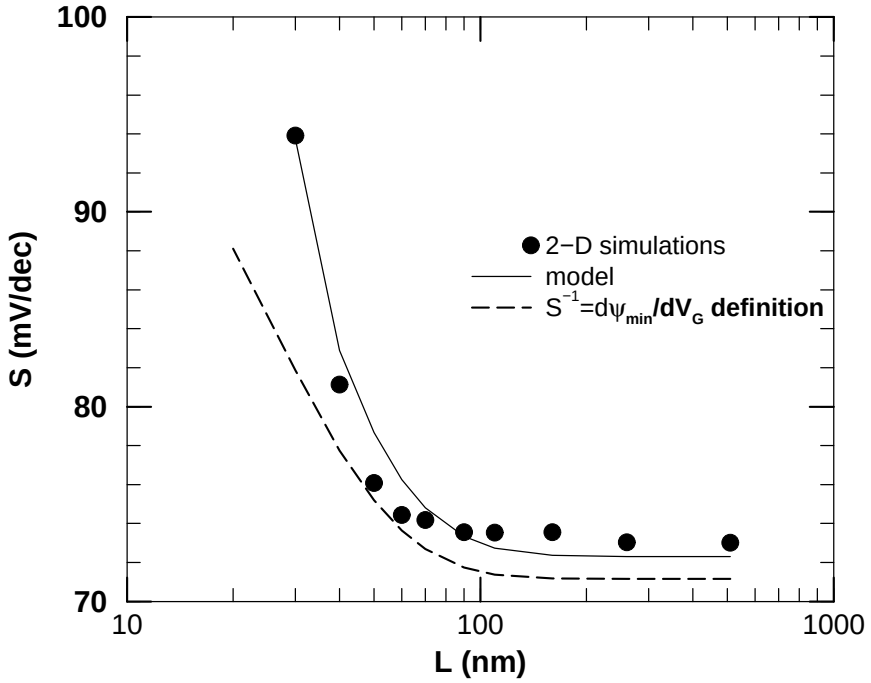


Figure 13: Comparison of subthreshold swing models with 2-D simulations for an RD-SCI MOSFET; t_{ox} =15 Å, $y_j = 20\ nm$, $V_{DS} = 0.05\ V$, N_A =2.5x10^{18} cm^{-3}, N_A^- =1x10^{17} cm^{-3}, and $y_i = 10\ nm$.

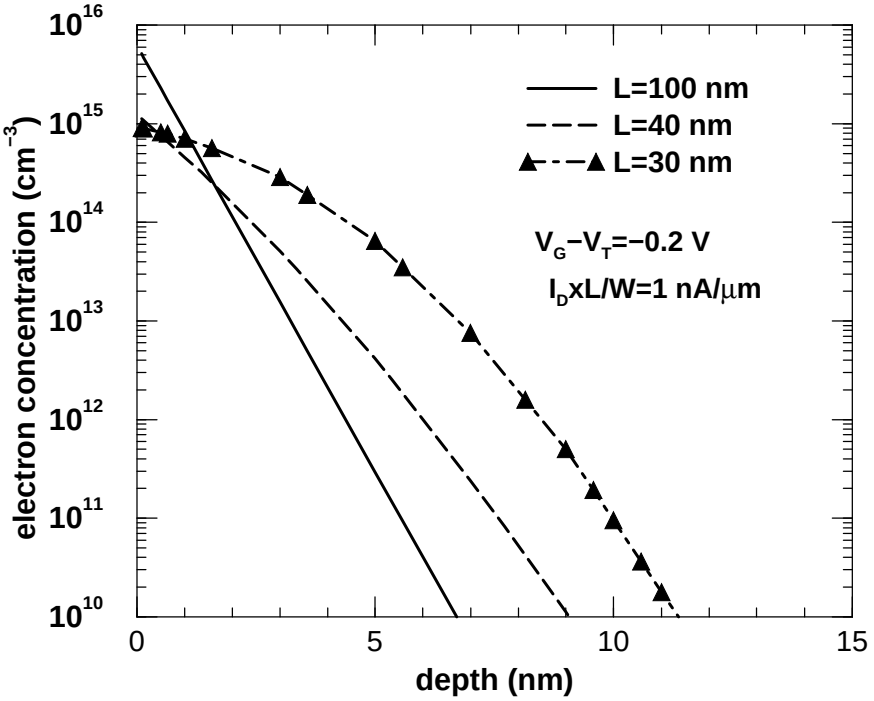


Figure 14: Electron concentration variation in the depth direction for different channel lengths (from 2-D simulations) for an RD-SCI MOSFET; t_{ox} =15 Å, $y_j = 12\ nm$, $V_{DS} = 1.0\ V$, N_A =2.5x10^{18} cm^{-3}, N_A^- =1x10^{17} cm^{-3}, and $y_i = 10nm$.

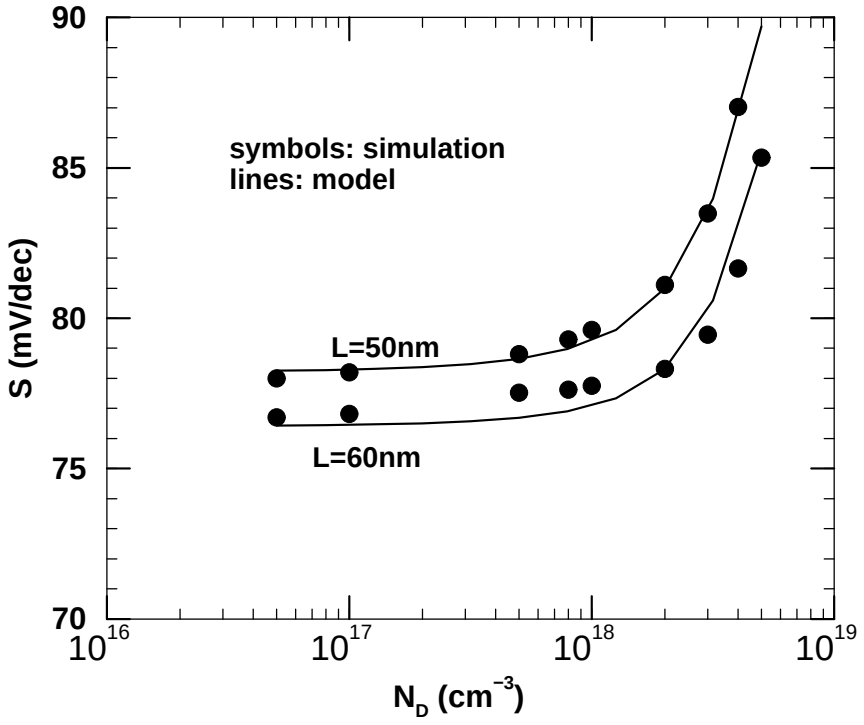


Figure 15: Comparison of subthreshold swing model for accumulation MOSFETs with 2-D simulations; t_{ox} =15 Å, $y_j = 12\ nm$, $V_{DS} = 1.0\ V$, N_A =2x10^{18} cm^{-3}, and $y_i = 10nm$.

short-channel subthreshold current is obtained by replacing S and V_T by their short-channel values:

$$I_{sub} = \frac{W}{L}\mu C_g \frac{\eta}{\beta^2} exp\left(\frac{V_G - V_T - \frac{n}{\beta}}{\frac{\eta}{\beta}}\right)\left[1 - exp\left(-\beta V_{DS}\right)\right] \tag{24}$$

Here, $\eta = \frac{S.\beta}{\ln(10)}$ is the subthreshold swing factor; $C_g = C_{ox}$ for SCA FETs, and $C_g = C_{ox}C_{Si}/(C_{ox} + C_{Si})$ for BCA FETs, where $C_{Si} = \epsilon_s/y_{ch}$. The subthreshold current model is compared with 2-D simulations for a wide range of channel doping, for both long, and short channel lengths, Fig. 16. A good match is seen between the model and simulations.

3.4 Summary

Static power is becoming increasingly important in gigascale systems. Accurate subthreshold swing modeling is critical for subthreshold current prediction, and optimization of V_T and V_{dd}. The canonical method of taking a point-derivative to calculate subthreshold swing fails at short channel lengths due to volume inversion. An approach originally formulated for subthreshold swing computation of double gate MOSFETs is applied to accumulation

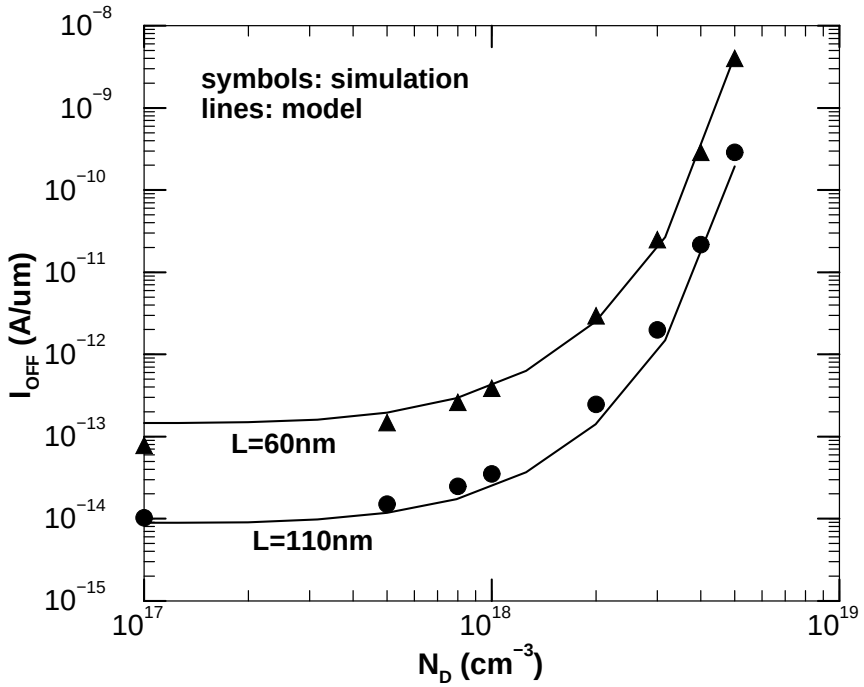


Figure 16: Comparison of off-current model with 2-D simulations for accumulation MOSFETs; t_{ox} =15 Å, $y_j = 12\ nm$, $V_{DS} = 1.0\ V$, N_A =2x10^{18} cm^{-3}, and $y_i = 10\ nm$.

and inversion MOSFETs. The resulting subthreshold swing model agrees well with 2-D simulations. An expression for subthreshold current of accumulation MOSFETs is formulated based on the Swanson-Meindl approach for inversion MOSFETs.

CHAPTER IV

PREDICTIVE CIRCUIT MODELING

4.1 Introduction

The trend toward device miniaturization makes it necessary to scan a wide range of supply and threshold voltages, and different types of FETs so that optimal performance is obtained. Such a scan is best done by physically based models so that predictive circuit modeling can be done to project, and optimize circuit performance well into the next decade. Circuit simulators like SPICE do not have an easy interface to input user-defined models; circuit simulators also have a higher computation time (compared to analytical techniques) which could result in large computation times for optimization in a multi-dimensional design space. Thus, a computationally efficient methodology that is also easy to implement is needed to use existing FET physical models to predict circuit performance.

The speed and accuracy of $V_{dd} - V_T$ optimization, and prediction of power-frequency performance depends on both the underlying FET models, and the way these models are used to extract power and timing information. In this work, the MOSFET I-V model proposed by Austin [73] is used. This I-V model is transregional in nature, i.e. all regions of operation are modeled by maintaining continuity across regional boundaries while including lateral and vertical high field effects. The model is entirely physical and avoids any parameter fitting. Therefore, the model is extremely useful for predicting the performance of technology generations over the next decade.

Predictive circuit modeling for near term generations can be done in a way similar to that in [74]. In fact, using physical FET models for threshold voltage and subthreshold swing derived in Chapters II and III, some of the critical BSIM parameters can be calculated physically. The expression for one such parameter - DVT0 (V_T rolloff coefficient), is

given:

$$DVT0 = \frac{4}{\zeta}\left(1 - \frac{\gamma\sqrt{2\phi_F} - \frac{qN_A}{\epsilon_{Si}\lambda_1^2}}{\phi_{bi} - 2\phi_F}\right) \tag{25}$$

where,

$$\zeta = \left[1 + \frac{C_{ox}d}{\epsilon_{Si}}\left(1 + \left(\frac{\lambda_1\epsilon_{Si}}{C_{ox}}\right)^2\right)\right](1 + exp(-\lambda_1 L)) \tag{26}$$

and γ is the body bias factor.

Predictive circuit modeling for long-term future generations cannot be done using the approach proposed in [74], since more than 90% of BSIM parameters are not predicted but kept the same as that for a previous generation. For long-term predictive circuit modeling, a chain of inverters could be considered since worst-case delay of a critical path can be simulated by replacing complex CMOS gates by their worst-case inverters [75]. A number of methods, mostly analytical [73], [76]-[78], have been proposed to model inverter delay. Either most of them are based on the alpha-power law proposed by Sakurai [79], which is unsuitable for predictive modeling since it makes use of fit parameters, or are valid only for a small range of supply voltages [73]. Also, for BCA FETs, which have five regions of operation [80], derivation of an analytical expression for gate delay would be subject to numerous assumptions, which would limit the validity of the delay equation to a small range of supply voltages.

A circuit simulation methodology is proposed that is valid over a wide range of supply voltages, makes use of physical models, and is applicable for a variety of FETs, so that device and technology optimization can be performed by scanning a wide design space. The method is fast as it avoids any numerical integration. The proposed method is called *tick*-based, wherein a "tick" is a point in time. Since BSIM decks are readily available for SCI devices, an SCI FET is considered to validate the tick-based methodology.

4.2 Tick Based Method

Consider a critical path modeled by a chain of inverters, Fig. 17. Assuming equal rise and fall times and a step input to the first stage, the delay of the critical path is [81]

$$T_{PD} = \frac{1}{f_{clk}} = \frac{n_{cp}T_{PDn}f_{ineff}}{b} \tag{27}$$

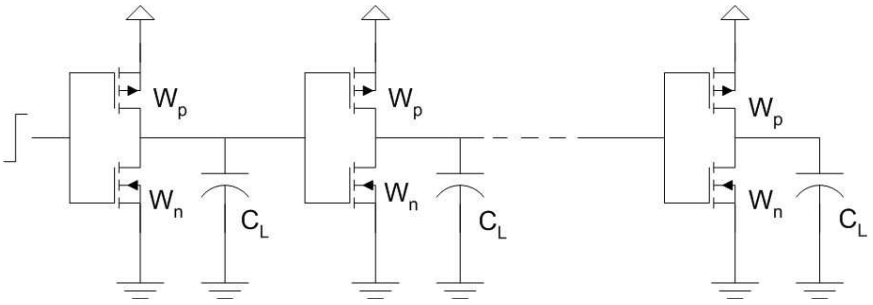


Figure 17: An inverter chain used in the tick-based method.

where n_{cp} is the number of gates in the critical path, f_{ineff} is the effective fan-in factor, b is the clock skew factor, and T_{PDn} is the delay of the second or subsequent stages of the inverter chain. Since short-circuit power is less than 10% of total power in the 50nm generation [81], the short circuit current can be neglected while estimating delay and power to a first order. Thus, T_{PDn} is just the discharge time through an n-MOSFET. At the output of any stage, it is known that

$$i = -C_L \frac{dV_{out}}{dt} \tag{28}$$

where C_L is the load capacitance. Integrating both sides,

$$\int_{V_{n-1}}^{V_n} dV_{out} = -\frac{1}{C_L} \int_{t_{n-1}}^{t_n} i.dt \tag{29}$$

The subscript n represents the current tick, whereas *n-1* represents the previous tick. The tick size, Δt $(= t_n - t_{n-1})$, is made small enough to avoid computational errors. Typically, this value is chosen to be around 1% of the expected delay value. At each tick, circuit parameters (V_T, V_{in}, V_{DSAT}, V_{out}) are updated. V_{T_n} is dependent on V_{DS_n} (because of DIBL), and thus needs to be updated at each time step.

An SCI device is used to exemplify the application of tick methodology. The SCI device has three regions of operation namely subthreshold, linear/non-saturated and saturated. The I-V equations [47] for the three regions of operation are shown in Table 8.

V_{GS_n} is the input voltage at the present tick. This value is determined based on the input waveform. If the input is a step, then its value is V_{dd} for $n \geq 0$. V_{DS} is the same as V_{out}. V_T is the sum of $V_{T_{long}}$ (not a function of V_{DS}) and ΔV_T (which is a function of

Table 8: I-V equations for the SCI FET [47].

Subthreshold	$\mathsf{I_d} = \frac{\mathsf{W}\mu_o\mathsf{C_{ox}}}{\mathsf{Lm}}\left[\frac{1}{2\theta}\left(\sqrt{1+\frac{4\eta\theta}{\beta}}-1\right)\right]^2\left\{1-\exp\left[-\frac{\beta\mathsf{m}}{\eta}\mathsf{V_{DS}}\right]\right\}$ $\exp\left\{\frac{\beta}{\eta}\left[\mathsf{V_{GS}}-\mathsf{V_T}-\frac{\eta}{\beta}\right]\right\}$
Linear/ Non-saturated	$\mathsf{I_d} = \frac{\mathsf{WC_{ox}}\mu_o}{\mathsf{L}\left(1+\frac{\mathsf{V_d}}{\mathsf{LE_c}}\right)(1+\theta(\mathsf{V_{GS}}-\mathsf{V_T}))}\left\{(\mathsf{V_{GS}}-\mathsf{V_T})\,\mathsf{V_{DS}}-\frac{1}{2}\mathsf{V_{DS}^2}\right\}$ $+\frac{\mathsf{WC_{ox}}\mu_o}{\mathsf{L}\left(1+\frac{\mathsf{V_d}}{\mathsf{LE_c}}\right)(1+\theta(\mathsf{V_{GS}}-\mathsf{V_T}))}\left\{\frac{4}{3}\phi_\mathsf{F}\frac{\mathsf{Q_{BO}}}{\mathsf{C_{OX}}}\left[\left(1+\frac{\mathsf{V_{DS}}}{2\lvert\phi_\mathsf{F}\rvert}\right)^{3/2}-\left(1+\frac{3}{2}\frac{\mathsf{V_{DS}}}{2\phi_\mathsf{F}}\right)\right]\right\}$
Saturated	$\mathsf{I_d} = \frac{\mathsf{WC_{ox}}\mu_o}{\mathsf{L}(1+\theta(\mathsf{V_{GS}}-\mathsf{V_T}))}\Bigg[\frac{1}{\left(1+\frac{\mathsf{V_{DSAT}}}{\mathsf{LE_c}}\right)}\{(\mathsf{V_{GS}}-\mathsf{V_T})\mathsf{V_{DSAT}}$ $-\frac{1}{2}\mathsf{V_{DSAT}^2}+\frac{4}{3}\phi_\mathsf{F}\frac{\mathsf{Q_{BO}}}{\mathsf{C_{ox}}}\left[\left(1+\frac{\mathsf{V_{DSAT}}}{2\lvert\phi_\mathsf{F}\rvert}\right)^{3/2}-\left(1+\frac{3}{2}\frac{\mathsf{V_{DSAT}}}{2\lvert\phi_\mathsf{F}\rvert}\right)\right]\Bigg\}$ $+\frac{1}{\mathsf{m}}\left[\frac{1}{2\theta}\left(\sqrt{1+\frac{4\eta\theta}{\beta}}-1\right)\right]^2\left\{1-\exp\left[-\frac{\beta\mathsf{m}}{\eta}\left(\mathsf{V_{DS}}-\mathsf{V_{DSAT}}\right)\right]\right\}\Bigg]$

V_{DS}). The value of ΔV_T is found by assuming $V_{DS} = V_{DS_{n-1}}$. If a small enough tick size is chosen, then the error resulting from this assumption is extremely small.

First consider the subthreshold region of operation. The term involving V_{DS} in the drain current equation can be neglected for the most part of the output transition. Denoting the time-independent term in the I-V equation as K,

$$I_d = K.exp\left\{\frac{\beta}{\eta}\left[V_{GS_n} - V_{T_n} - \frac{\eta}{\beta}\right]\right\} \tag{30}$$

Using (30) in (28) and integrating,

$$V_{out_n} - V_{out_{n-1}} = -\frac{1}{C_L}K.exp\left\{\frac{\beta}{\eta}\left[V_{GS_n} - V_{T_n} - \frac{\eta}{\beta}\right]\right\}\Delta t \tag{31}$$

$V_{out}|_{n=0} = V_{dd}$ for a falling output. Using this initial condition and the above equation, V_{out} can be evaluated at each n until V_{out} has reached 10% of the maximum value.

If V_{dd} is so low that the device always operates in the subthreshold region, then the delay can be found using analytical methods. It can be shown that the propagation delay in this case, assuming $V_{dd} >> \frac{3kT}{q}$, is

$$T_{PD} = \frac{C_L V_{DD}}{I_{ON}}\left(1 - \frac{\eta}{\beta V_{DD}}\right) \tag{32}$$

where I_{ON} is the on-current.

Next, consider the linear/non-saturated region of operation. In this region, the I-V equation can be rewritten as a sum of three components: $I_{linear} = I_1 - I_2 + I_3$, where,

$$\begin{aligned} I_1(V_{DS_n}) &= \kappa \left(V_{GS_n} - V_{T_n}\right) V_{DS_n} \\ I_2(V_{DS_n}) &= \kappa \left\{ \frac{1}{2} V_{DS_n}^2 \right\} \\ I_3(V_{DS_n}) &= \kappa \left\{ \frac{4}{3} \phi_F \frac{Q_{BO}}{C_{OX}} \left[\left(1 + \frac{V_{DS_n}}{2|\phi_F|} \right)^{3/2} - \left(1 + \frac{3}{2} \frac{V_{DS_n}}{2\phi_F} \right) \right] \right\} \end{aligned}$$

where,

$$\kappa = \frac{W C_{ox} \mu_o}{L \left(1 + \frac{V_d}{L E_c}\right) \left(1 + \theta (V_{GS_n} - V_{T_n})\right)} \tag{33}$$

Normally, I_3 contributes very little to the overall current. Thus, this part can be evaluated using the V_{DS} value of the previous tick ($=V_{DS_{n-1}}$). Fig. 18a shows V_{out} and the various components of I_{linear} plotted vs. time. To get a better idea of how large I_3 is w.r.t. I_{linear}, the same curves are plotted as a percentage of the total current. This is shown in Fig. 18b. I_3 is seen to be less than 10% of I_{total} almost all throughout the decay period . Thus, I_3 can be made time independent. The I-V equation can now be rewritten as

$$I_d = \kappa \left\{ (V_{GS_n} - V_{T_n}) V_{out} - \frac{1}{2} V_{out}^2 \right\} + I_3(V_{out_{n-1}}) \tag{34}$$

Using (34) in (28) and integrating,

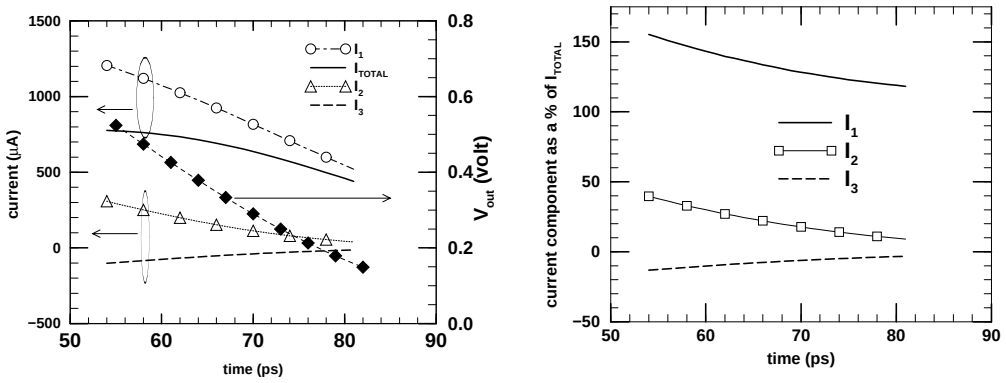
$$\int_{V_{out_{n-1}}}^{V_{out_n}} \frac{dV_{out}}{\left\{ (V_G - V_T) V_{out} - \frac{1}{2} V_{out}^2 + \frac{I_3(V_{out_{n-1}})}{\kappa} \right\}} = \frac{-\kappa}{C_L} \Delta t \tag{35}$$

$$\left[-2 \frac{tan^{-1} \left\{ \frac{-V_{out}+b}{\sqrt{2a+b^2}} \right\}}{\sqrt{2a+b^2}} \right]_{V_{out_{n-1}}}^{V_{out_n}} = \frac{-\kappa}{C_L} \Delta t \tag{36}$$

where $a = I_3(V_{out_{n-1}})/\kappa$, and $b = V_{GS_n} - V_{T_n}$. The above expression gives an analytical equation for V_{out_n}.

Next, consider the saturated region of operation. V_{DSAT_n} is a function of V_{GS_n}, and is a known quantity. Denoting the time independent parts in the I-V equation as K', and K'',

$$I_D = K' - K''.exp \left[-\frac{\beta m}{\eta} \left(V_{out} - V_{DSAT_n} \right) \right] \tag{37}$$



(a) (b)

Figure 18: (a) Various components of I_{linear}, and (b) components of I_{linear} shown as a percentage of the total current; $\left(\frac{W_p}{L}\right) = 2\left(\frac{W_n}{L}\right) = 40$, $FO = 1$, and $V_{dd} = 1.5\ V$.

Using (37) in (28) and integrating,

$$\left[\frac{ln\left(K' - K''.exp\left\{-\frac{\beta m}{\eta}(V_{out} - V_{DSAT_n})\right\}\right)}{\frac{\beta m}{\eta}K'} + \frac{(V_{out} - V_{DSAT_n})}{\frac{\beta m}{\eta}K'}\right]_{V_{out_{n-1}}}^{V_{out_n}} = -\frac{\Delta t}{C_L} \qquad (38)$$

(38) gives an analytical expression for V_{out_n}.

4.3 Tick Method Verification

TSMC 0.25 μm process was selected and BISM parameters were obtained from MOSIS [82]. The technology parameters for the TSMC process are shown in Table 9. To account for non-uniform substrate doping, one parameter in the transregional model was adjusted by around 10%; the match between the transregional I-V model and SPICE is shown in Fig. 19.

Table 9: Technology parameters for TSMC 0.25 μm process.

	N-channel	P-channel	units
V_T	0.54	-0.50	volt
t_{ox}	57	57	Å
y_j	100	100	nm
channel doping	2.35x10^{17}	4.16x10^{17}	cm^{-3}

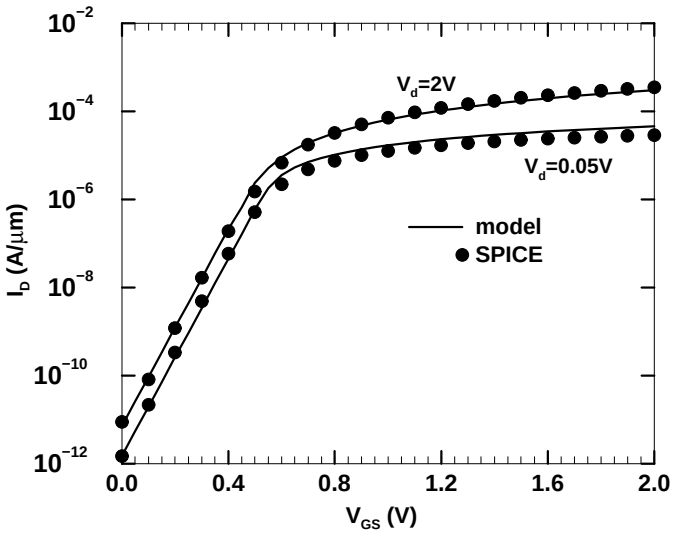


Figure 19: I-V curve comparison of transregional MOSFET model and HSPICE for TSMC 0.25 μm n-FETs.

Using the tick method, a chain of inverters is simulated. The input to the first stage is a step waveform. Waveforms at the first stage of the inverter chain are plotted in Fig. 20. A good match is seen between the tick method and HSPICE. From the simulation of the first stage of the inverter chain, delay information is extracted. The input to the second stage of the inverter chain is modeled as a ramp using the delay information from the first stage. Fig. 21 shows waveforms at the input and output of the second stage. Again, a good match is seen between the tick method and SPICE. Next, delay of the second stage of the inverter chain is plotted, for varying V_{dd}, Fig. 22. This is to confirm that assumptions made in the tick-based method are valid for a wide range of V_{dd}. A good match is seen between the tick methodology and SPICE, with the maximum error being less than 10% of the actual delay. The tick-based method runs more than 10X faster than a comparable HSPICE simulation. The speed-up is because of the fact that numerical integration is avoided and most calculations are kept analytical.

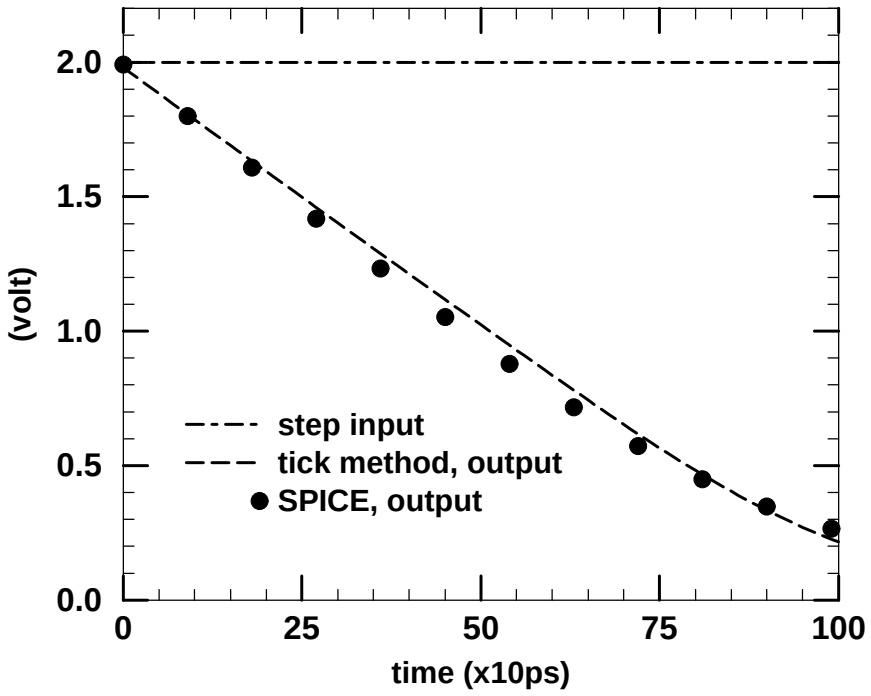


Figure 20: Transient waveforms at the first stage of inverter chain; $\left(\frac{W_p}{L}\right) = 2.5 \left(\frac{W_n}{L}\right) = 50$, $FO = 1$, $C_L = 1\ pF$, and $V_{dd} = 2.0\ V$.

4.4 Summary

The tick-based methodology is an extremely useful method to easily use physically based, transregional I-V models to *predict* delay and power-frequency performance of future technology. The method is valid over a wide range of supply voltages, and for a variety of FETs. Numerical integration is avoided, and thus, more than 10X speed improvement over HSPICE is obtained.

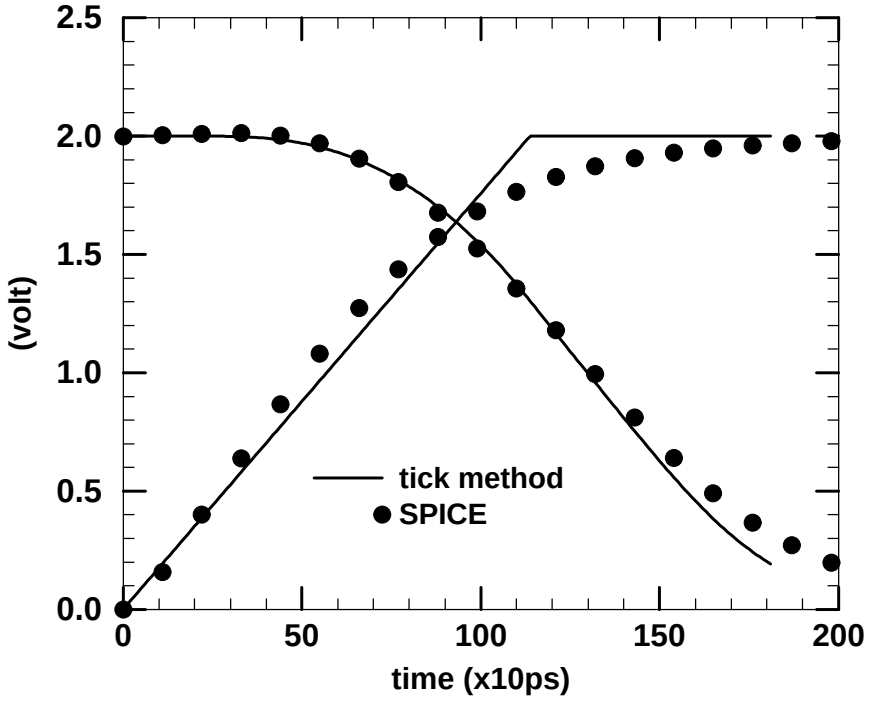


Figure 21: Input and output waveforms at the second stage of inverter chain; $\left(\frac{W_p}{L}\right) = 2.5\left(\frac{W_n}{L}\right) = 50$, $FO = 1$, $C_L = 1\ pF$, and $V_{dd} = 2.0\ V$.

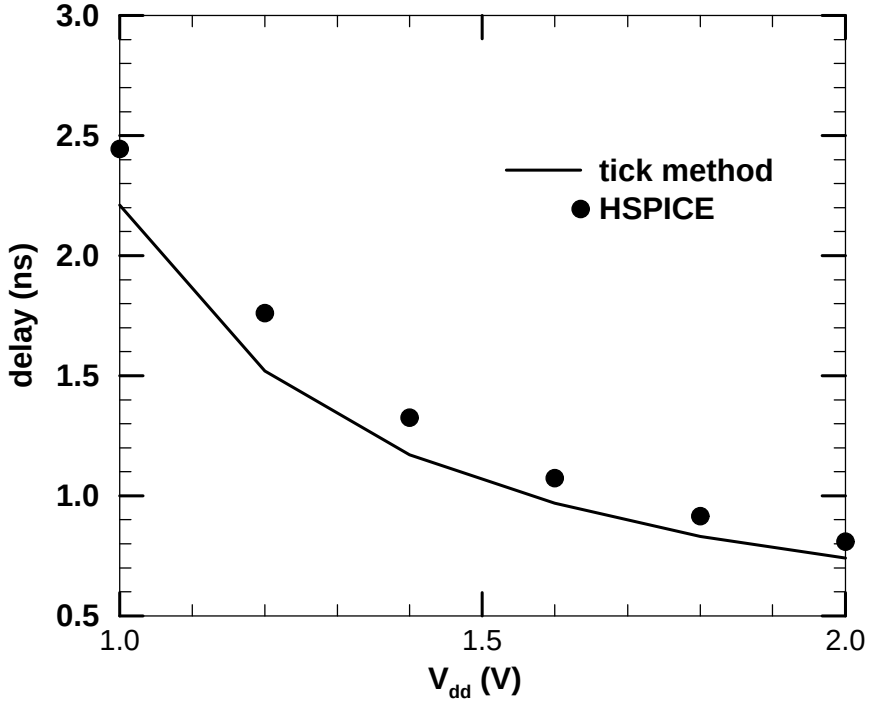


Figure 22: Delay of second stage of inverter chain: comparison between model and HSPICE; $\left(\frac{W_p}{L}\right) = 2.5\left(\frac{W_n}{L}\right) = 50$, $FO = 1$, and $C_L = 1\ pF$.

CHAPTER V

ACCUMULATION FET DESIGN

5.1 Introduction

The accumulation MOSFET may either operate as an SCA MOSFET or a BCA MOSFET. The main difference between BCA and SCA devices is the location where the channel first opens. BCA MOSFET conduction starts in a neutral buried channel, and with increasing applied gate bias, subsequently extends to an accumulated surface layer [47]. On the other hand, SCA MOSFET conduction is limited to a surface accumulation layer. Because of the steep subthreshold swing rollup and threshold voltage rolloff of BCA MOSFETs, it has been recognized previously that it is better to design accumulation MOSFETs such that they operate as SCA MOSFETs [38],[39]. Guidelines have been provided in [37]-[38], [39], on how to design an accumulation device such that it operates as an SCA MOSFET with minimum long channel subthreshold swing. These guidelines are based on a 1-D solution to the Poisson equation, and do not hold good for short channel devices. New guidelines are given based on a 2-D Poisson equation solution.

5.2 BCA vs. SCA FETs

Threshold voltage control of accumulation and inversion MOSFETs is shown in Fig. 23. For low channel doping, the accumulation MOSFET is an SCA MOSFET, whereas for high channel doping, it is a BCA MOSFET. The BCA/SCA MOSFET boundary is the Fermi MOSFET, the device with the minimum long-channel subthreshold slope, and zero oxide field at $V_G = V_T$ [37]. High channel doping that causes the accumulation MOSFET to be in the BCA regime also results in high process sensitivity (dV_T/dy_i), Fig. 24, and high dopant fluctuation effects [44]. Because of high SCE, high process sensitivity, and high dopant fluctuation effects of BCA MOSFETs, it is best to operate the accumulation MOSFET either as an SCA MOSFET or as a BCA MOSFET close to the BCA/SCA MOSFET

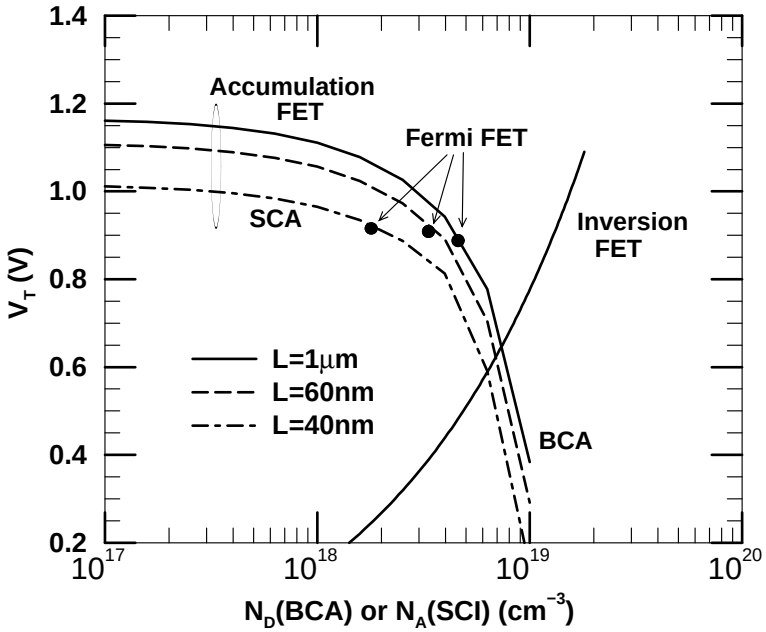


Figure 23: Threshold voltage (from model) vs. channel doping of poly-gate accumulation and inversion MOSFETs; t_{ox} =15 Å, $y_j = 12\ nm$, $V_{DS} = 1.0\ V$, N_A =2.5x10^{18} cm^{-3}, and $y_i = 10\ nm$.

boundary.

5.3 Accumulation FET design

Once the 2-D potential in the channel region of the accumulation MOSFET has been solved, the boundary between BCA and SCA MOSFETs can be identified by finding the channel doping that gives a zero vertical field at the $Si - SiO_2$ interface. Let N_{dbound} be the channel doping beyond which the SCA MOSFET becomes a BCA MOSFET. The channel doping that gives

$$\frac{\partial \psi}{\partial y}(x_m, 0)\Big|_{V_G = V_T} = 0 \tag{39}$$

is N_{dbound}. For $N_D > N_{dbound}$, the location where $\frac{\partial \psi}{\partial y}(x_m, y)\Big|_{V_G = V_T}$ is zero occurs for $y > 0$, resulting in buried channel operation. Since the LHS of (39) is a function of y_i, N_A, y_j, V_{DS}, L, t_{ox} and $N_{D_{S/D}}$ (the source/drain doping), N_{dbound} too is a function of all these parameters. In a manner similar to finding N_{dbound}, it is possible to find a y_{ibound} (the boundary channel depth) from (39). y_{ibound} is a function of N_D, N_A, y_j, V_{DS}, L, t_{ox} and $N_{D_{S/D}}$. This is as opposed to the long channel boundary depth found in [37]-[38], [39],

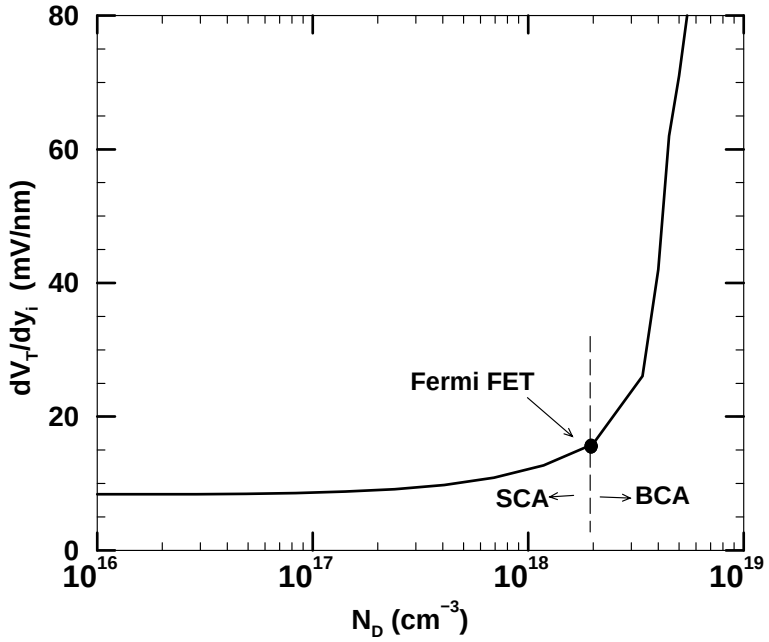


Figure 24: Effect of a $1nm$ tub-depth (y_i) variation on threshold voltage for BCA/SCA MOSFETs (from model); $L = 45\ nm$, t_{ox} =15 Å, $y_j = 12\ nm$, $V_{DS} = 1.0\ V$, N_A =2.5x10^{18} cm^{-3}, and $y_i = 10\ nm$.

which is only a function of N_D and N_A. Table 10 shows N_{dbound} as a function of channel length, for two different channel depths. The N_{dbound} derived from long-channel theory (for $L = 1\ \mu m$) is seen to be different from that derived from 2-D theory. It can be seen from Fig. 23 and Table 10 that an SCA MOSFET at long channel length becomes a BCA MOSFET at shorter channel lengths. Thus, an accumulation MOSFET would have to be designed such that the minimum channel length device (say, 3σ less than the nominal channel length) operates as an SCA FET.

Fig. 25 shows the design space for accumulation MOSFETs with a channel length of 30 nm, generated from 2-D simulations. Also shown are constant V_T contours. It can be seen that good SCE control makes necessary extremely shallow channel implants. The minimum channel/tub-depth is taken to be 5 nm due to process limitations. The V_T[1] range allowable by the design space is $0.27 < V_T < 0.40$ V. Fig. 26 shows the design space generated by the SCE models derived previously in this work. An additional constraint -

[1]ITRS defines V_T as the gate voltage that gives a drain current of 1 $\mu A/\mu m$. For $S = 100\ mV/dec$, and a channel length of 30 nm, $V_T|_{simulation} - V_T|_{ITRS} \approx 50\ mV$.

Table 10: N_{dbound} for various L; N_A =2.5x10^{18} cm^{-3}, and V_{DS} =1.0 V.

L	$y_i = 10\ nm$	$y_i = 20\ nm$
1 μm	5x10^{18} cm^{-3}	2x10^{18} cm^{-3}
60 nm	3.4x10^{18} cm^{-3}	1.07x10^{18} cm^{-3}
40 nm	1.5x10^{18} cm^{-3}	6x10^{17} cm^{-3}

that of V_T sensitivity to tub-depth variation - is included in the design space. It can be seen that at low tub-doping, the design space is bounded by the SCE limit, whereas at high tub-doping, it is bounded by the dV_T/dy_i limit.

The model-generated design space helps to greatly reduce the multi-dimensional search space for a given application. In the final optimization step, a 2-D simulator can be used to search the reduced design space. As an example, consider the design space shown in Fig. 26. To achieve an I_{OFF} of 0.1 $\mu A/\mu m$, a V_T of about 0.36 V is needed (assuming $S = 90\ mV/dec$). Reading off the tub-doping and tub-depth values from the $V_T = 0.35\ V$ contour in Fig. 26, a 2-D simulator (classical simulations with drift-diffusion transport) is used to scan this set of devices. The optimal device is found to be one with a tub-doping of 1x10^{18} cm^{-3} and a tub-depth of 10 nm, Fig. 27. Inclusion of velocity-overshoot in the simulations shown in Fig. 27 would shift the optimal device design point since a lower channel doping results in a higher velocity overshoot.

5.4 Summary

BCA FETs not only have a high subthreshold swing, but also a high process sensitivity, and high dopant fluctuation effects. So, it is best to design an accumulation FET such that it either operates as an SCA FET, or close to the BCA/SCA FET boundary. Based on 2-D theory, the BCA/SCA FET boundary has been derived, and is shown to be different from that derived using 1-D theory. Using the SCE models derived in this work, accumulation FETs can be designed to meet V_T rolloff, process sensitivity, and off-current requirements. An example design window is shown for $S < 110\ mV/dec$, and $dV_T/dy_i < 30\ mV/nm$.

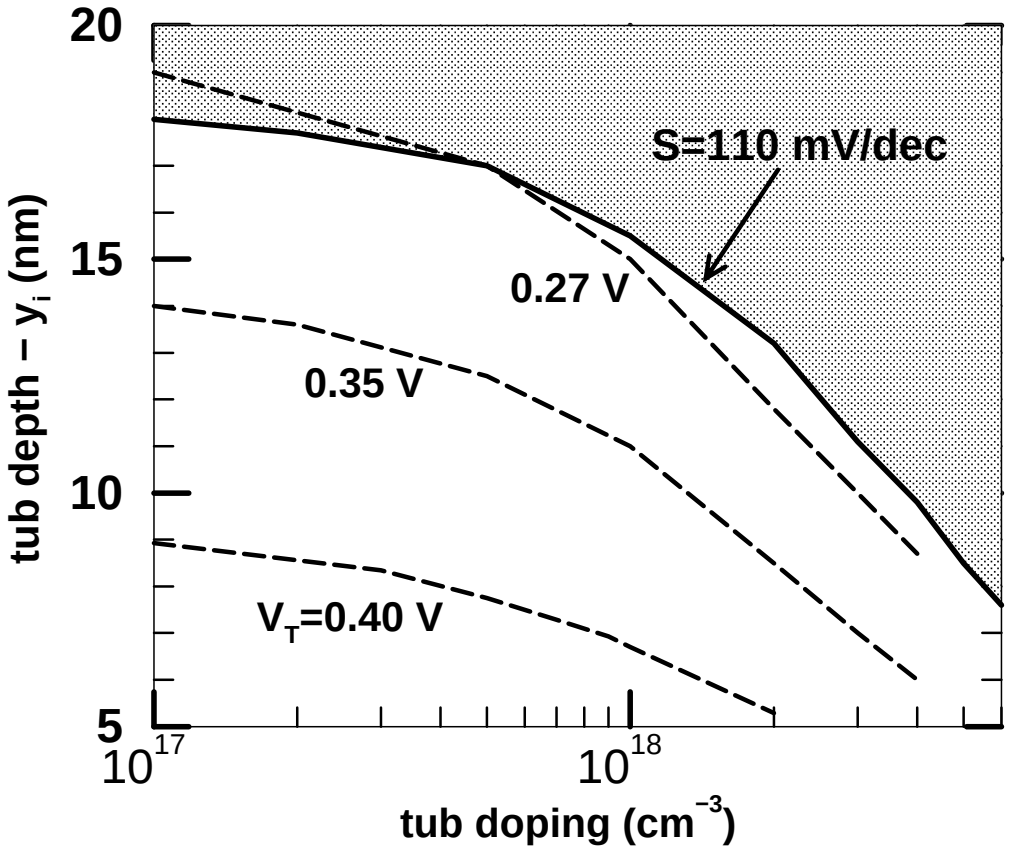


Figure 25: Design space for mid-bandgap metal gate, accumulation n-MOSFETs (from 2-D simulations). Substrate doping, N_A =4x10^{18} cm^{-3}, and the maximum subthreshold swing is taken to be 110 mV/dec. Also shown are constant V_T contours (dashed lines). $L = 30\ nm$, t_{ox} =10 Å, $y_j = 12\ nm$, and $V_{dd} = 1.0\ V$.

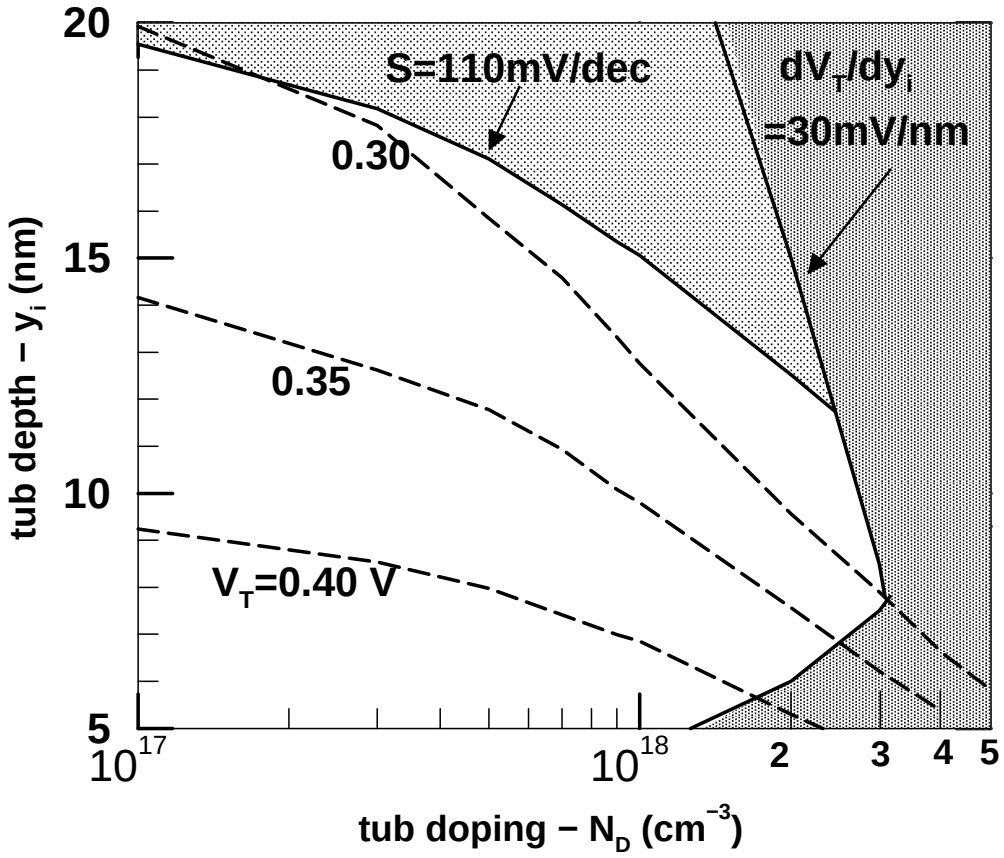


Figure 26: Design space for mid-bandgap metal gate, accumulation n-MOSFETs (from models). Substrate doping, N_A =4x10^{18} cm^{-3}; the maximum subthreshold swing is taken to be 110 mV/dec, and the maximum dV_T/dy_i is taken to be 30 mV/nm. Also shown are constant V_T contours (dashed lines). $L = 30\ nm$, t_{ox} =10 Å, $y_j = 12\ nm$, and $V_{dd} = 1.0\ V$.

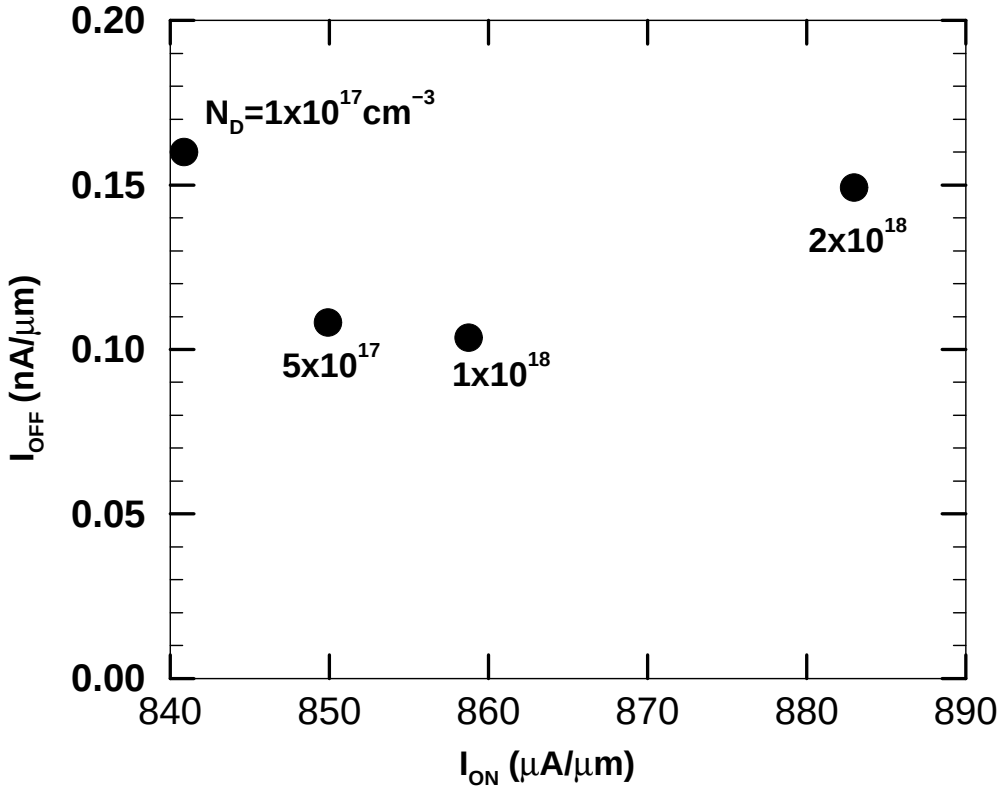


Figure 27: $I_{OFF} - I_{ON}$ from 2-D simulations (drift-diffusion transport); N_D and y_i are chosen so as to give a V_T of 0.35 V. $L = 30$ nm, t_{ox} =10 Å, $y_j = 12$ nm, $V_{dd} = 1.0$ V, N_A =4x10^{18} cm^{-3}, and gate workfunction is 4.63 eV.

CHAPTER VI

METAL GATE MOSFETS

Polysilicon gates have the disadvantage of gate depletion since it increases the effective oxide thickness [83]. Metal gate technology overcomes the gate-depletion problem of poly-gates, and so is being widely pursued as the next promising gate material. With the recent advent of tunable workfunction (WF) metal gates, it has become possible to fabricate metal gates with workfunctions from 4.2 eV to 5.2 eV [40]. CMOS technology can either use a single gate for both n-MOS and p-MOS; or it can use dual metal gate technology. When single metal gate technology is used, the workfunction of the metal is usually at the Si mid-bandgap. When dual metal gate technology is used, the workfunctions of the gates are optimized for the given application [84].

Short channel effect in accumulation MOSFETs is poorly understood in existing literature. Using the physical SCE models derived in previous chapters, accumulation and inversion MOSFET SCE are compared. This is followed by a comparison of accumulation and inversion MOSFETs for tunable metal gate CMOS technology. Then, mid-bandgap metal gate technology is considered.

6.1 SCE comparison of accumulation and inversion MOSFETs

Long-channel subthreshold swing of accumulation and inversion devices is plotted in Fig. 28. The left hand side of the figure shows accumulation FETs and the right side shows inversion FETs. The x-axis is the tub-doping (inverse of the tub-doping) of the inversion (accumulation) MOSFET. The minimum subthreshold swing occurs for an accumulation FET with a tub-doping around $1\text{x}10^{18}$ cm^{-3}, the device being a Fermi FET.

Short channel subthreshold swing is plotted in Fig. 29. For the accumulation FET, S increases rapidly beyond a certain tub-doping (this is the BCA/SCA FET boundary).

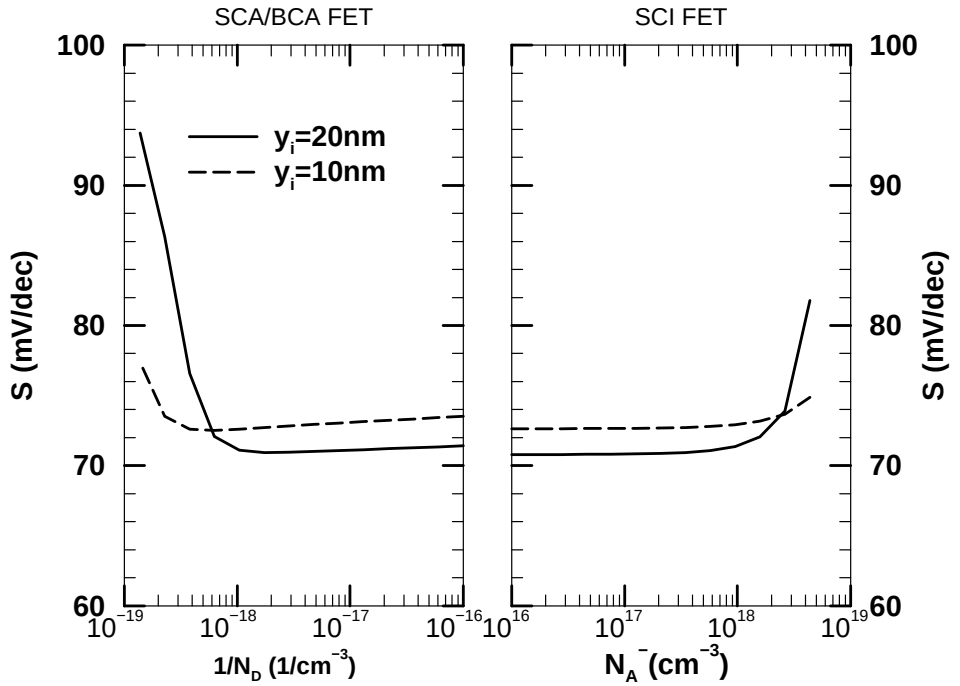


Figure 28: Long-channel subthreshold swing vs. tub-doping (from models); t_{ox} = 15 Å, and N_A =2.5x10^{18} cm^{-3}.

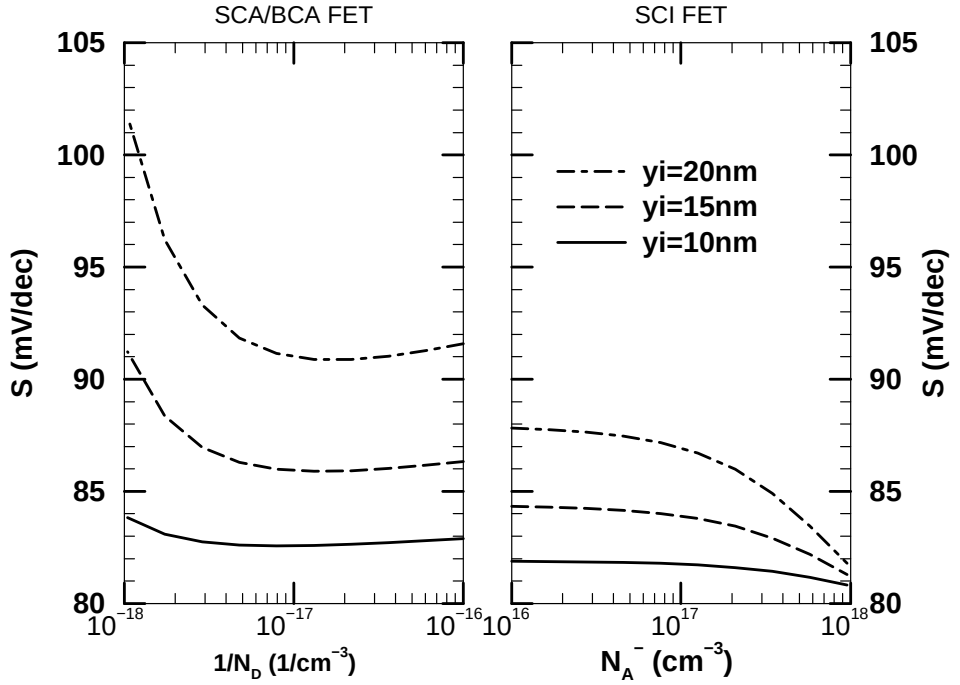


Figure 29: Short-channel subthreshold swing vs. tub-doping (from models); $L = 40\ nm$, t_{ox} =15 Å, $y_j = 20\ nm$, $V_{DS} = 1.0\ V$, and N_A =2.5x10^{18} cm^{-3}.

This is because the accumulation FET goes into the buried channel mode of operation when $N_D > N_{dbound}$. A different behavior can be seen in the case of inversion FETs - as tub-doping increases, S decreases; this is because SCE decreases with increasing tub-doping resulting in a decreased S rollup, and hence an overall decrease in S. Because of increased SCE, the Fermi-FET is no longer the device with the minimum subthreshold swing in the short-channel regime.

Fig. 30 shows V_T rolloff for various accumulation and inversion FETs. Because of the counter-doped region in the accumulation FET, the lateral electric field at the source/drain junction is expected to be lower in the accumulation FET than in the inversion FET, resulting in reduced SCE. Fig. 31 plots the lateral field at the surface of three example n-FETs. The RD-SCI device has a peak lateral field at the source that is 60% greater than that of the SCA FET. But accumulation FETs are seen to have a similar or worse V_T rolloff compared to inversion FETs. Fig. 30 shows that an SCA FET and a UD-SCI FET with the same channel doping are seen to have similar V_T rolloff curves; an increase in either the tub-doping or tub-depth of the SCA FET results in an increase in V_T rolloff.

The anomaly of reduced lateral field not leading to reduced SCE can be explained using Fig. 32, where depletion depth and channel location are plotted vs. channel doping. The accumulation FET is seen to have a greater depletion depth compared to the inversion FET. Any reduction in SCE caused by the lowered lateral field in the accumulation FET is offset by this increase in depletion depth. In the long channel regime, the increased depletion depth in the SCA FET compared to the SCI FET results in a reduced subthreshold swing. As the SCA FET channel doping is increased, d increases, resulting in decreased subthreshold swing. When $N_D > N_{dbound}$, the channel location becomes buried (Fig. 32) resulting in a steep subthreshold swing rollup. This is the reason why the device operating at the BCA/SCA FET boundary (i.e. the Fermi FET) has the minimum long-channel subthreshold swing. In the short channel regime, the increased depletion depth in the accumulation FET offsets any reduction in SCE resulting from the reduced lateral field, and results in increased SCE compared to the inversion FET.

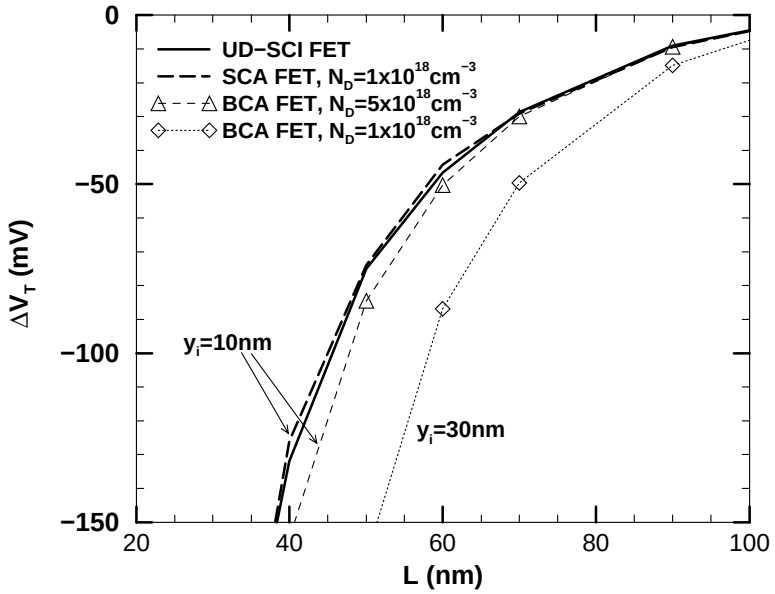


Figure 30: Threshold voltage rolloff comparison of accumulation and inversion MOSFETs (from 2-D simulations); t_{ox} =15 Å, $y_j = 10\ nm$, $V_{DS} = 1.0\ V$, and N_A =2.5x10^{18} cm^{-3}.

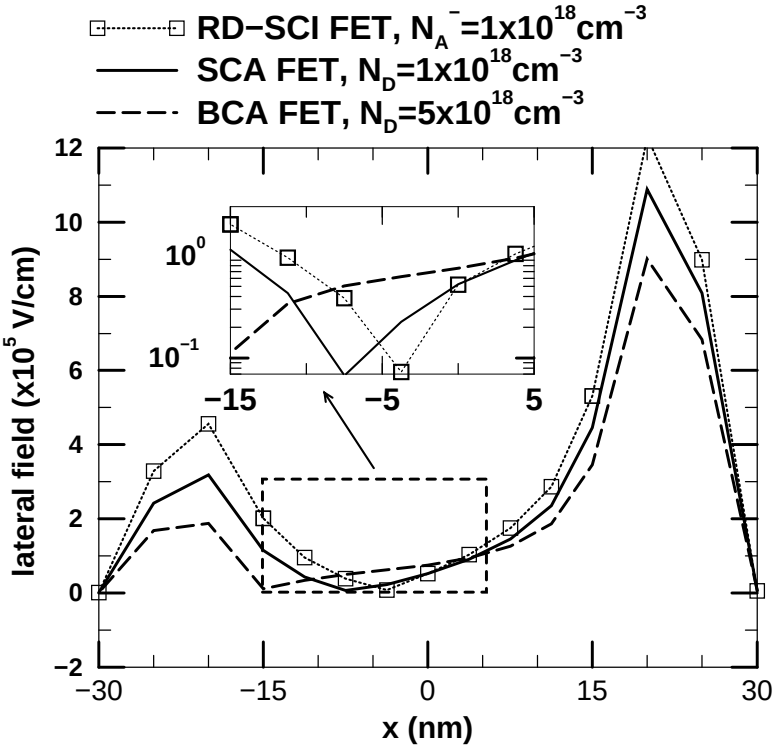


Figure 31: Lateral field (from 2-D simulations) at the $Si - SiO_2$ interface for accumulation and inversion FETs; $L = 45\ nm$, t_{ox} =15 Å, $y_j = 10\ nm$, $V_{DS} = 1.0\ V$, $V_{GS} = 0\ V$, N_A =2.5x10^{18} cm^{-3}, and $y_i = 10\ nm$.

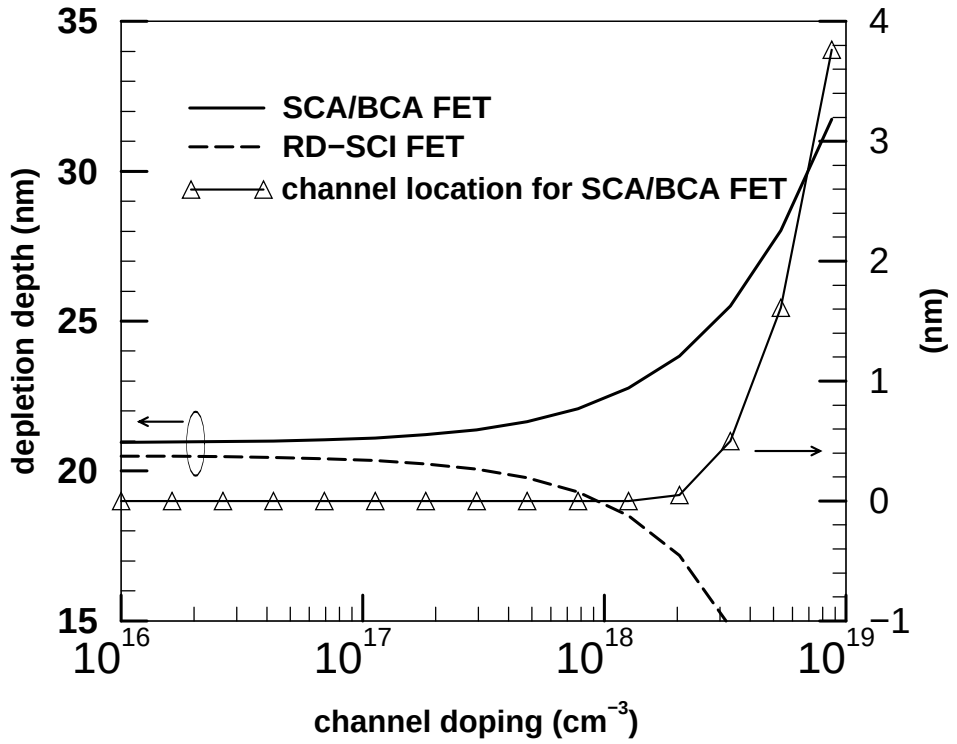


Figure 32: Long-channel depletion depth for accumulation and inversion MOSFETs (from models); t_{ox} =15 Å, $V_{GS} = 0.2\ V$, N_A =2.5x10^{18} cm^{-3}, and $y_i = 10\ nm$.

6.2 Tunable Workfunction, Metal Gate CMOS Technology

From the arguments in the previous section, it can be concluded that short-channel accumulation FETs offer no significant performance advantage over inversion FETs either in terms of subthreshold swing, or V_T rolloff, when the gate WF is tunable. Thus, for good SCE control, it is best to use an inversion device. The optimal inversion and accumulation FET are identified for a given I_{ON}, and I_{OFF}-I_{ON} from 2-D simulations using MEDICI [63] are plotted for these optimized structures, Fig. 33. The simulations are classical in nature, with drift-diffusion transport. It can be seen that over a wide range of I_{ON}, the SCA FET performs worse than the SCI FET. When body-bias is used (e.g. to control die-to-die parameter variations [85]), the situation remains the same. Thus, in the scenario where the gate workfunction is tunable, the inversion MOSFET is better than the accumulation MOSFET.

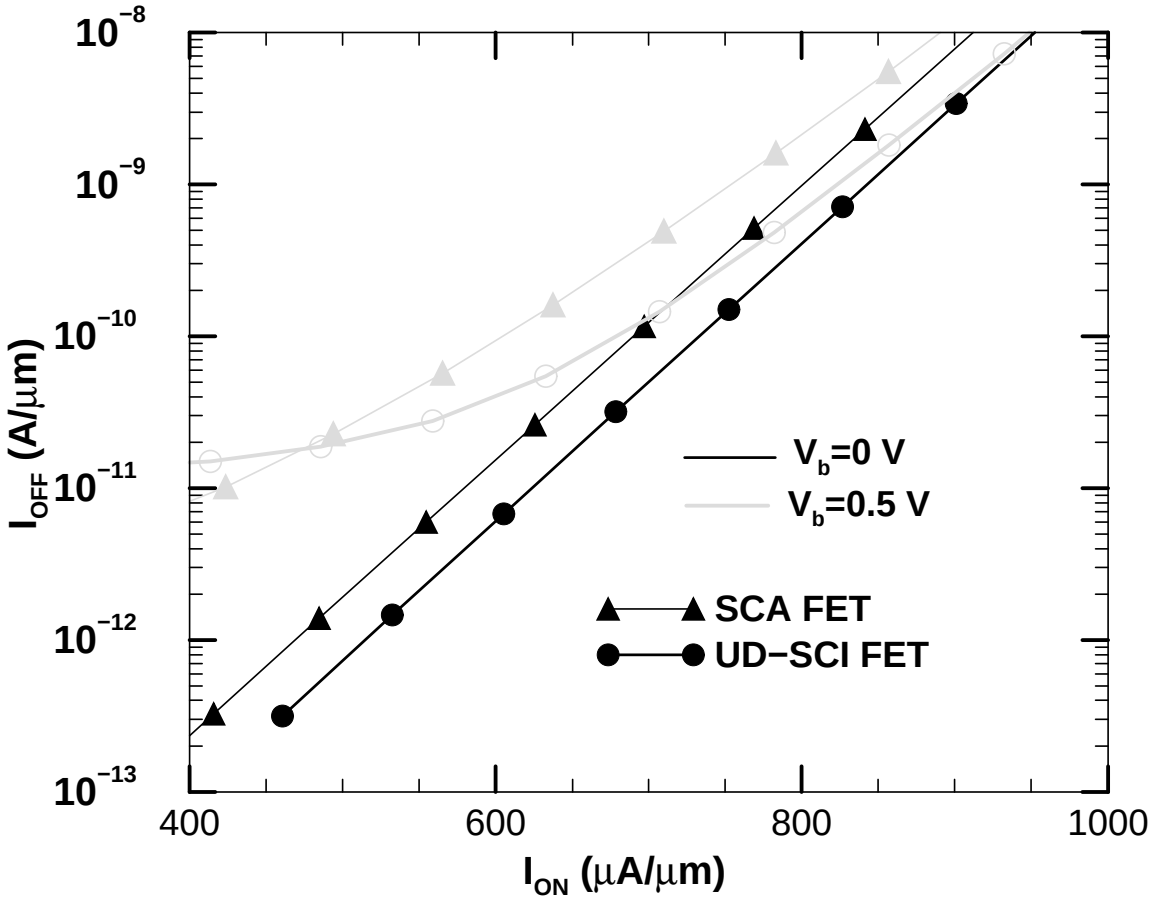


Figure 33: I_{OFF} -I_{ON} for tunable WF metal gate CMOS technology (from 2-D simulations); L = 55 nm, t_{ox} =15 Å, y_j = 10 nm, and V_{dd} = 1.0 V. For the SCA FET, N_D =1x10^{18} cm^{-3}, N_A =2.5x10^{18} cm^{-3}, and y_i = 10 nm. For the UD-SCI FET, N_A =1x10^{18} cm^{-3}.

6.3 Single metal gate CMOS

Many applications limit the gate material to a single metal for both the n-FET and the p-FET. This greatly reduces process complexity and increases circuit density [17]. In single metal gate technology, the gate workfunction will most likely be the Si mid-bandgap value so that equivalent drive currents are obtained from n-MOS and p-MOSFETs [86]. The best metal suitable for such applications is Tungsten (W), which has a workfunction of 4.63 eV. For such applications, the necessary V_T is obtained by adjusting the channel/substrate doping of the MOSFET.

When V_T of SCA and SCI FETs using a tungsten gate is plotted vs. channel doping, Fig. 34, it can be seen that SCA FETs achieve a V_T between 0.35 V and 0.5 V, whereas SCI FETs have a V_T greater than 0.45V. A lower value of V_T can be obtained with the SCI FET if a higher y_i and/or lower channel/substrate is used. But, this would result in increased SCE. Fig. 35 shows I_{OFF} vs. V_T from 2-D models derived in previous chapters. Accumulation and inversion devices are designed such that the minimum I_{OFF} is obtained for a given V_T. A cross-over is seen at $V_T = 0.46\ V$, where the off-current is about 20 $pA/\mu m$. At low V_T, the SCI FET has a lower channel/substrate doping compared to the accumulation FET, resulting in higher SCE, and hence a higher I_{OFF}.

The total channel doping in an accumulation FET is process dependent. If the process is able to achieve super steep Boron doping (for an n-FET), then the total doping in the channel is almost the same as the effective doping ($N_T = N_D$), Fig. 36. On the other hand, if the Boron doping in the n-FET is almost uniform, the total channel doping is the sum of the acceptor and donor impurity concentrations ($N_T = N_D + N_A$). The latter case results in lower channel mobility, and thus lower drive current, compared to the former case. In order to assess processing conditions on device performance, both cases of the accumulation FET will be considered while comparing it to the inversion FET.

Fig. 37 shows I_{OFF} vs. V_T for accumulation and inversion FETs, this time from 2-D simulations. Simulations are done using drift-diffusion transport. The I_{OFF} vs. V_T plane has a weak dependence on the mobility models used. Both cases of the accumulation FET ($N_T = N_D$, and $N_T = N_D + N_A$) result in the same curve since V_T is independent of

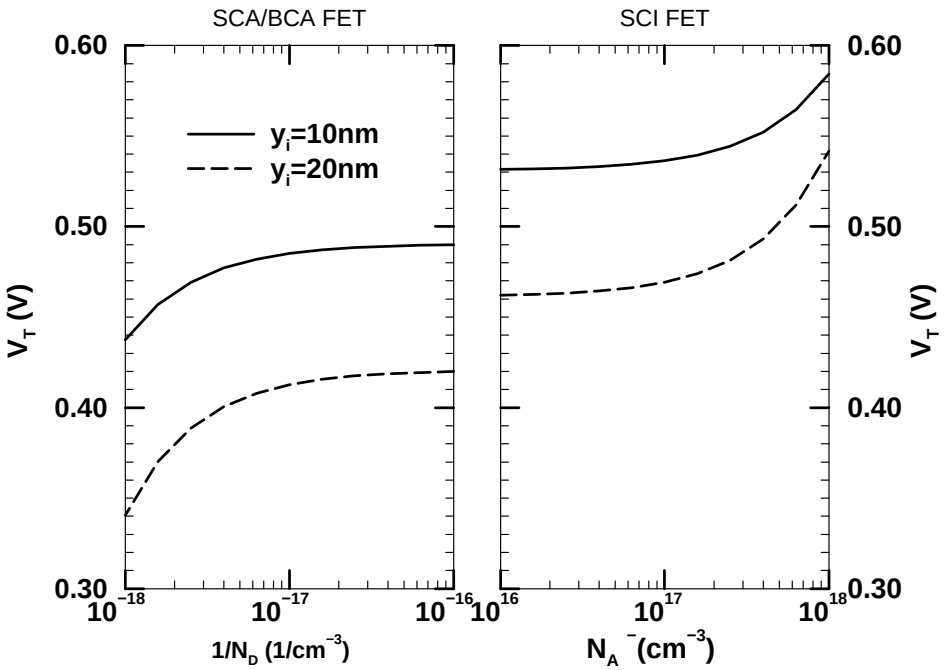


Figure 34: Threshold voltage (from models) vs. tub-doping for accumulation and inversion MOSFETs; $L = 45\ nm$, t_{ox} =15 Å, $y_j = 12\ nm$, $V_{DS} = 1.0\ V$, N_A =2.5x10^{18} cm^{-3}, and gate workfunction is 4.63 eV.

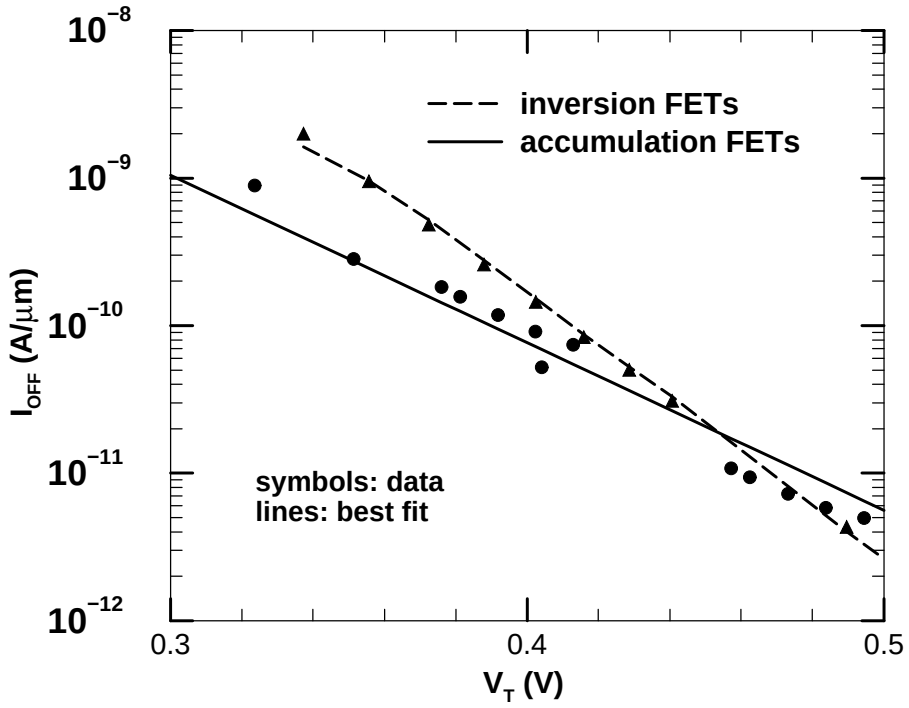


Figure 35: Off-current vs. threshold voltage for accumulation and inversion MOSFETs (from models); $L = 45\ nm$, t_{ox} =15 Å, $y_j = 12\ nm$, $V_{dd} = 1.0\ V$, and gate workfunction is 4.63 eV.

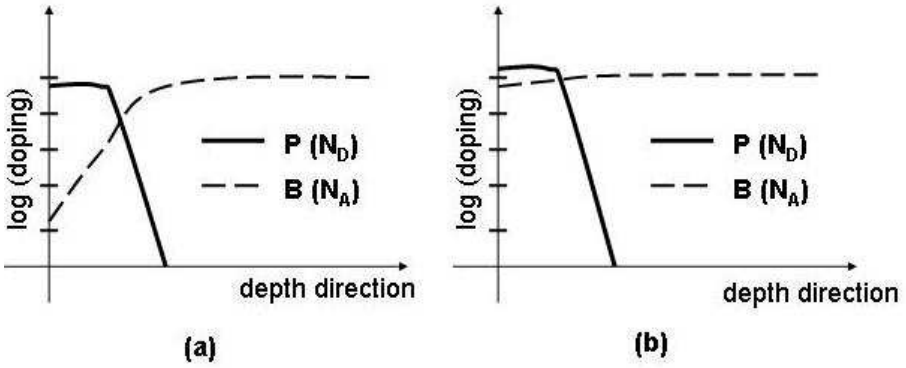


Figure 36: Effect of processing conditions on tub-doping; Phosphorus and Boron profiles are shown for two example n-FETs: (a) $N_T \approx N_D$, and (b) $N_T \approx N_D + N_A$.

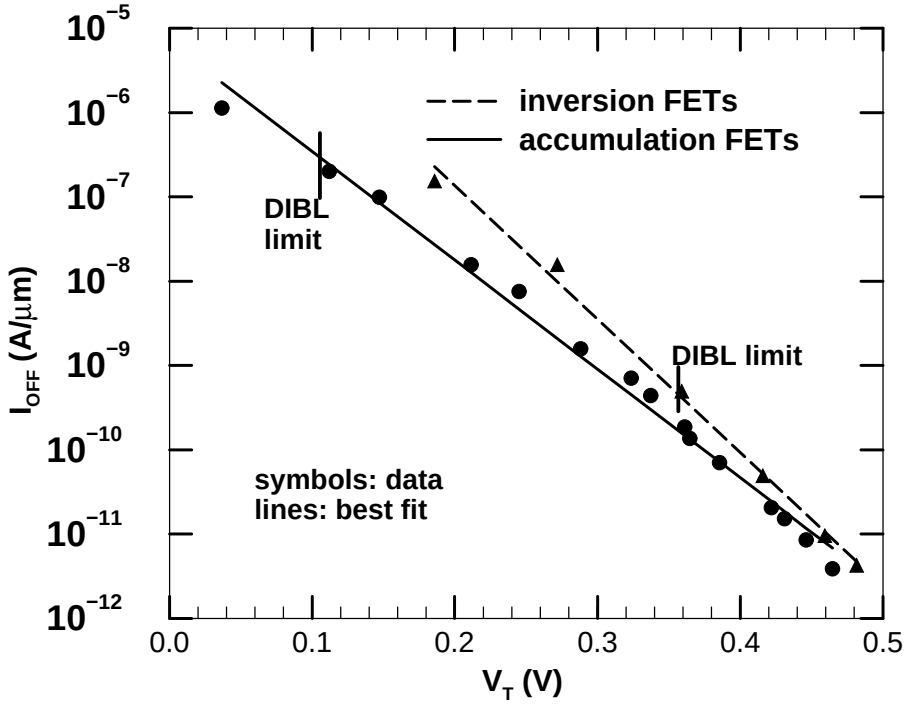


Figure 37: Off-current vs. threshold voltage for single metal gate CMOS technology (from 2-D simulations); symbols indicate data points, and lines indicate best fit. $L = 45\ nm$, t_{ox} =15 Å, $y_j = 12\ nm$, $V_{dd} = 1.0\ V$, and gate workfunction is 4.63 eV. The DIBL limit is taken to be 100 mV/V.

the channel mobility (μ). It can be seen that for $V_T < 0.4V$, the accumulation FET has a significantly lower I_{OFF} than the inversion FET. At lower V_T, the inversion FET has a significantly higher SCE compared to the accumulation FET. Fig. 37 shows the point where DIBL exceeds 100 mV/V for inversion and accumulation FETs. The inversion FET V_T is limited to $V_T > 0.35\ V$ because of DIBL. On the other hand, accumulation FETs can achieve a V_T as low as 0.1 V before reaching the DIBL limit. Additional comparisons between mid-band metal gate accumulation and inversion MOSFETs are given in Appendix B.

Fig. 38 shows $I_{OFF} - I_{ON}$ curves, from 2-D simulations using both drift-diffusion (DD) and hydro-dynamic (HD) transport. HD simulations include non-local carrier transport, and thus account for velocity overshoot. Both DD and HD simulations include QM V_T shift by means of the van Dort model [87]; the accuracy of using van Dort QM correction is discussed in Appendix C. In DD simulations, there is a crossover in inversion and accumulation FET curves. The reason for this is the same as that for the $I_{OFF} - V_T$ crossover. In HD simulations, there is no such crossover, and the inversion FET is always better than the accumulation FET.

The difference between on-current from DD and HD simulations (for a given off-current) is seen to be 100-150%. Non-local carrier transport causes an increase in on-current because of velocity overshoot. Using HD simulations to account for non-local transport has been known to result in spurious velocity overshoot [88]. Reported values of on-current increase due to velocity overshoot are between 10-15% [89]. The spurious velocity overshoot will be studied in detail and alternate simulation methodologies will be looked at to accurately predict the on-current.

To study the effect of velocity overshoot, an accumulation and an inversion device are selected having almost the same off-current (corresponding to $I_{OFF} \approx 0.5\ nA/\mu m$ from Fig. 38). Both have the same V_T, Table 11. It can be seen that channel doping of the accumulation FET is greater than that of the inversion FET. Hydro-dynamic simulations result in the inclusion of velocity overshoot (compared to drift-diffusion simulations). To understand the dependence of velocity overshoot on channel mobility, analytical expressions

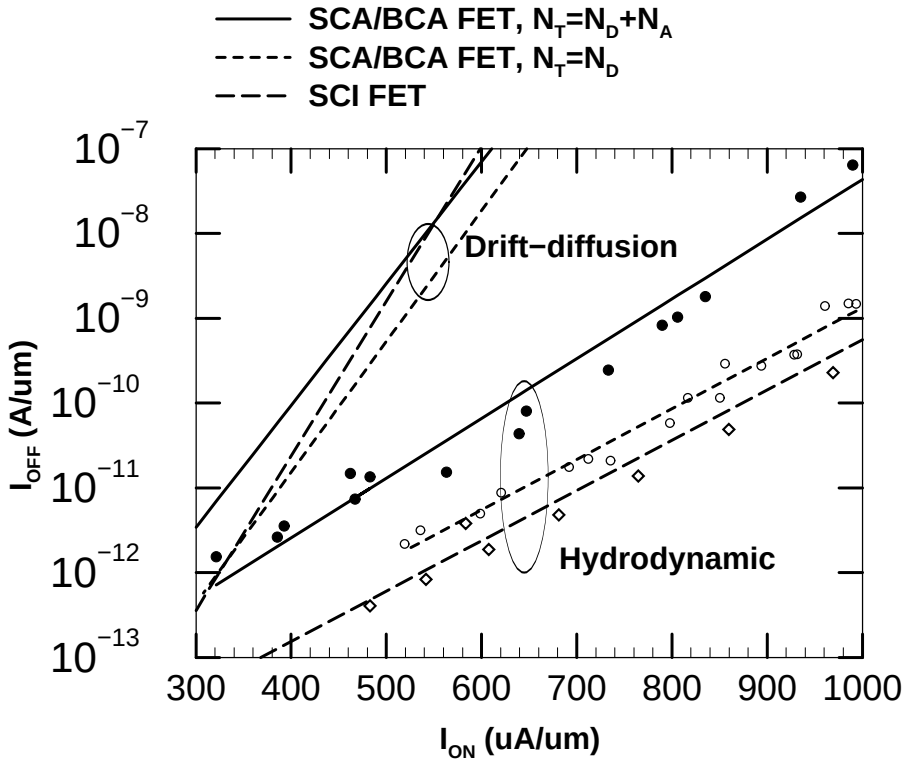


Figure 38: Off-current vs. on-current for single metal gate CMOS technology (from 2-D simulations); symbols indicate data points, and lines indicate best fit. $L = 45\ nm$, t_{ox} =15 Å, $y_j = 12\ nm$, $V_{dd} = 1.0\ V$, and gate workfunction is 4.63 eV.

from [47] are used. In [47], velocity overshoot is included by modifying the effective mobility as

$$\frac{\mu_{\mathsf{eff}}}{1+\frac{\mathsf{E(x)}}{\mathsf{E_C}}} \rightarrow \frac{\mu_{\mathsf{eff}}}{1+\frac{\mathsf{E(x)}}{\mathsf{E_C}}} + \frac{\lambda_{\mathsf{VO}}}{\mathsf{L}} \tag{40}$$

where μ_{eff} is the mobility including vertical high field effect, $E(x)$ is the lateral field, E_c is the critical field, and λ_{VO} is the velocity overshoot factor. λ_{VO} is given by

$$\lambda_{\mathsf{VO}} = 91\mathsf{ln}\left[0.16\mu_{\mathsf{eff}} - 14\right] \times 10^{-6} \tag{41}$$

From (40), and (41), it can be seen that the velocity overshoot effect increases with channel mobility. Since the inversion FET channel mobility is higher than that of the accumulation FET (Table 11), a higher velocity overshoot occurs in the inversion FET [90]. The amount of this velocity overshoot is overestimated by the un-calibrated HD simulations, resulting in a huge increase in on-current. Thus, using un-calibrated HD simulations, there

Table 11: Velocity Overshoot Effect for non-halo FETs; $L = 45\ nm$, t_{ox} =15 Å, $y_j = 12\ nm$, $V_{DS} = 1.0\ V$, and gate workfunction is 4.63 eV.

	SCI FET	BCA/SCA FET $N_T = N_A + N_D$	BCA/SCA FET $N_T = N_D$	units
N_D/N_A^-	4x10^{17}	5x10^{17}	5x10^{17}	cm^{-3}
N_A	4x10^{17}	8x10^{17}	8x10^{17}	cm^{-3}
y_i	-	14	14	nm
V_T - QM, DD	0.55	0.55	0.54	V
V_T - QM, HD	0.51	0.50	0.49	V
I_{ON} - QM, DD	438	437	480	$\mu A/\mu m$
I_{ON} - QM, HD	969	733	928	$\mu A/\mu m$
I_{OFF} - QM, HD	0.43	0.24	0.37	$nA/\mu m$

is no crossover in $I_{OFF} - I_{ON}$ curves of accumulation and inversion FETs, and inversion FETs perform better than accumulation FETs for a wide range of on-current.

Fig. 39 plots the lateral electric field and channel mobility of the two FET types (corresponding to devices in Table 11). Both accumulation and inversion FETs have almost the same lateral field at the source and in the channel. Channel mobility of the inversion FET is seen to be higher than that of accumulation FETs. Fig. 40 plots velocity in the channel. Velocity in the inversion FET is seen to be higher than that in accumulation FETs, and this is because of the increased channel mobility in the inversion FET. Thus, while drift-diffusion simulations predict better performance with well-processed accumulation MOSFETs, hydro-dynamic simulations predict the opposite. Because of the disparity between drift-diffusion and hydro-dynamic simulation results and because hydro-dynamic simulations predict an unrealistic (100-150%) increase in on-current compared to drift-diffusion simulations, alternative device simulation strategies would have to be used to accurately assess the scaling potential of accumulation and inversion MOSFETs. This will be done in chapter VIII, where Monte Carlo device simulations will be used to predict the drive-current.

The results presented until now are for MOSFETs without halo doping. The next plot

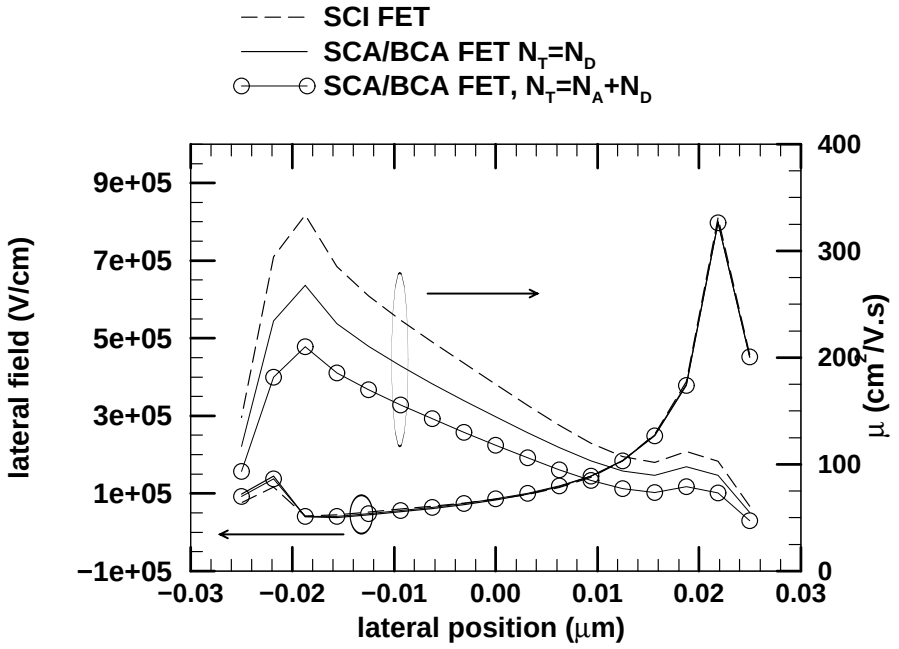


Figure 39: Lateral field and mobility vs. lateral position (from 2-D simulations); $L =$ 45 nm, t_{ox} =15 Å, $y_j = 12$ nm, $V_{DS} = 1.0$ $V, V_{GS} - V_T = 0.5$ V, and gate workfunction is 4.63 eV. Device doping is given in Table 11.

will be for halo-doped MOSFETs with a channel length of 30 nm. Fig. 41 shows the halo doping profile. The halo is generated by an analytic function, and is described in Appendix D. The $I_{OFF} - I_{ON}$ plot for halo devices, Fig. 42, shows that a BCA/SCA n-FET with a steep Boron doping can achieve better performance than the inversion FET for $I_{ON} > 700$ $\mu A/\mu m$. The increased velocity overshoot seen in non-halo SCI FETs is not present in the SCI FET with halo doping since halos reduce the effective carrier mobility. Table 12 shows doping, off-current and on-current of halo-doped devices. It can be seen that although channel and substrate doping of the accumulation FET is higher than that of the inversion FET, the on-current of the inversion FET is comparable to that of the accumulation FET. Both accumulation and inversion MOSFETs have $\Delta V_T(30\%\delta L)$ (i.e. reduction in V_T for a 30% reduction in channel length) less than 100 mV for the range of I_{ON} shown.

Fig. 43 plots $I_{OFF} - I_{ON}$ for non-halo FETs, with $V_{dd} = 0.5$ V. The simulations are classical in nature with drift-diffusion transport since both QM V_T shift and non-local

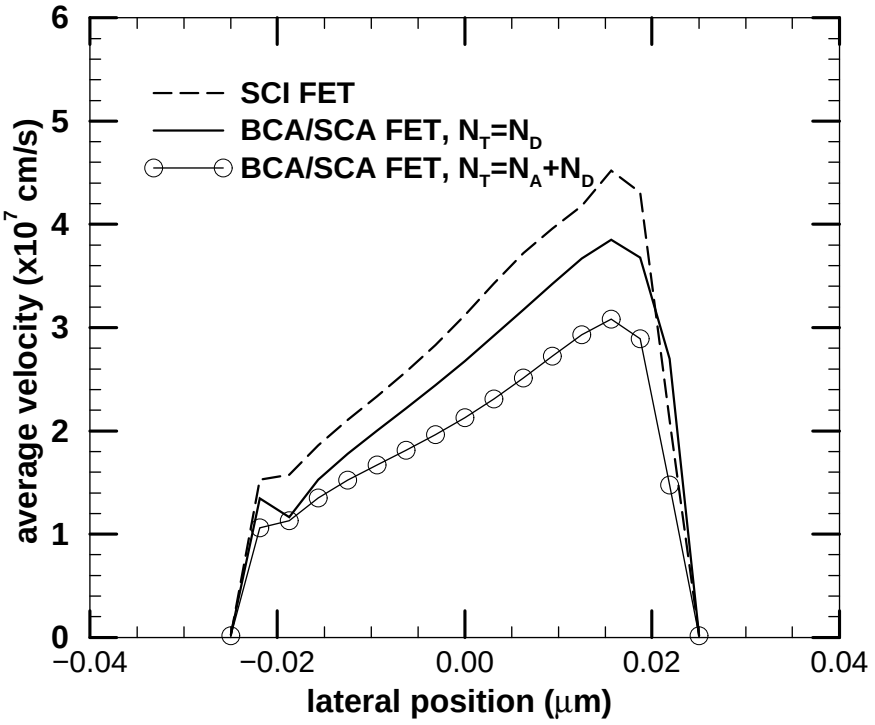


Figure 40: Average velocity vs. lateral position (from 2-D simulations); $L = 45\ nm$, t_{ox} =15 Å, $y_j = 12\ nm$, $V_{DS} = 1.0\ V, V_{GS} - V_T = 0.5\ V$, and gate workfunction is 4.63 eV. Device doping is given in Table 11.

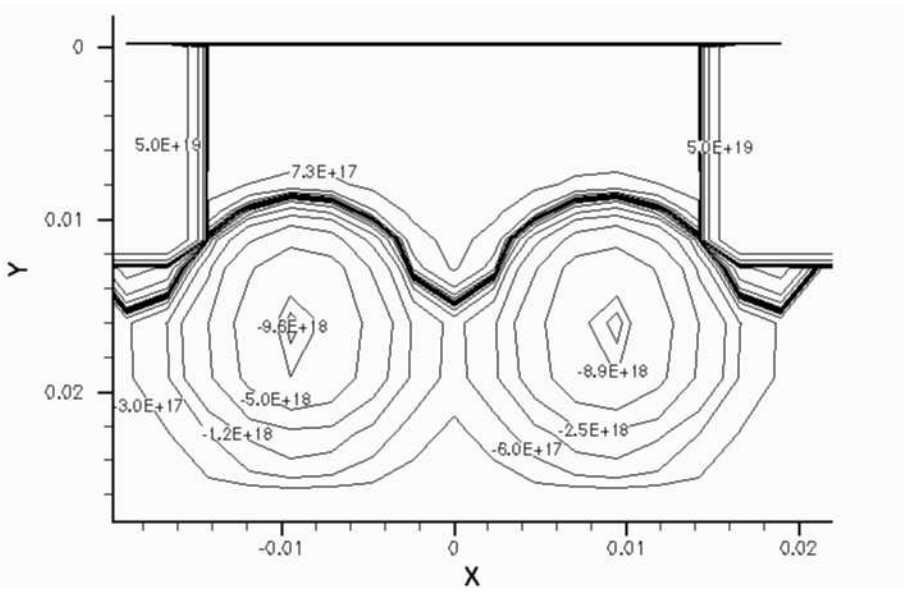


Figure 41: Halo doping profile for a 30 nm channel length accumulation FET.

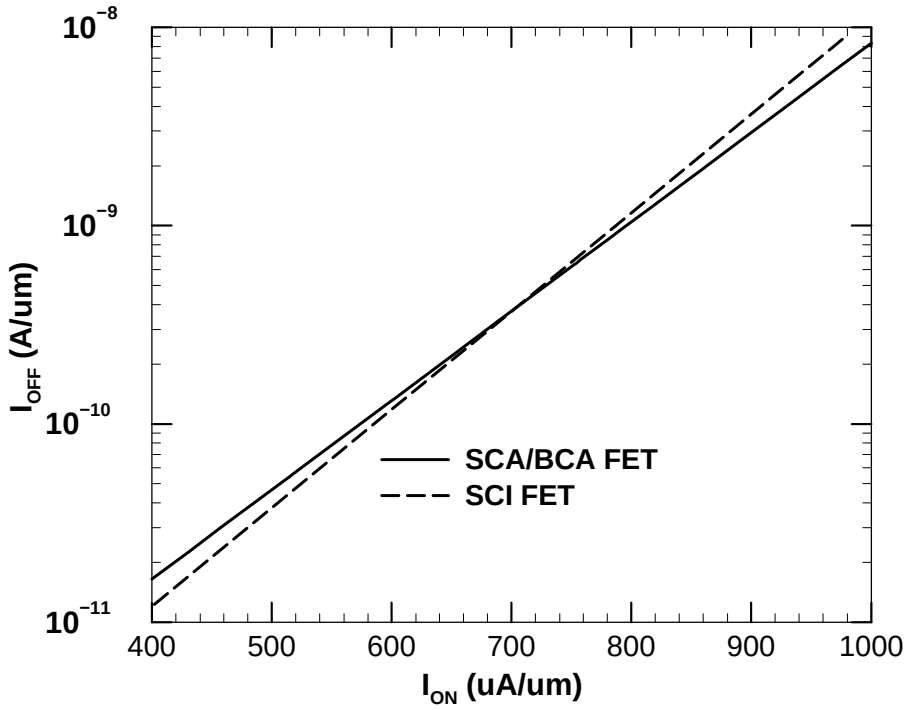


Figure 42: Off-current vs. on-current for $L = 30nm$ FETs (from 2-D simulations); $L =$ 30 nm, t_{ox} =15 Å, $y_j = 12$ nm, $V_{dd} = 0.9$ V, and gate workfunction is 4.63 eV. The MOSFETs are halo-doped with a peak halo doping of 1x10^{19} cm^{-3}.

Table 12: Velocity Overshoot Effect for halo-FETs; $L = 30nm$, t_{ox} =15 Å, $y_j = 12$ nm, $V_{dd} = 0.9$ V, and gate workfunction is 4.63 eV.

	SCI FET	BCA/SCA FET $N_T = N_A + N_D$	BCA/SCA FET $N_T = N_D$	units
N_D/N_A^-	4x10^{17}	5x10^{17}	5x10^{17}	cm^{-3}
N_A	4x10^{17}	6x10^{17}	8x10^{17}	cm^{-3}
y_i	-	14	14	nm
V_T - QM, HD	0.58	0.54	0.55	V
I_{ON} - QM, HD	594	621	656	$\mu A/\mu m$
I_{OFF} - QM, HD	0.10	0.40	0.21	$nA/\mu m$
$\Delta V_T(30\%\delta L)$	87	84	82	mV

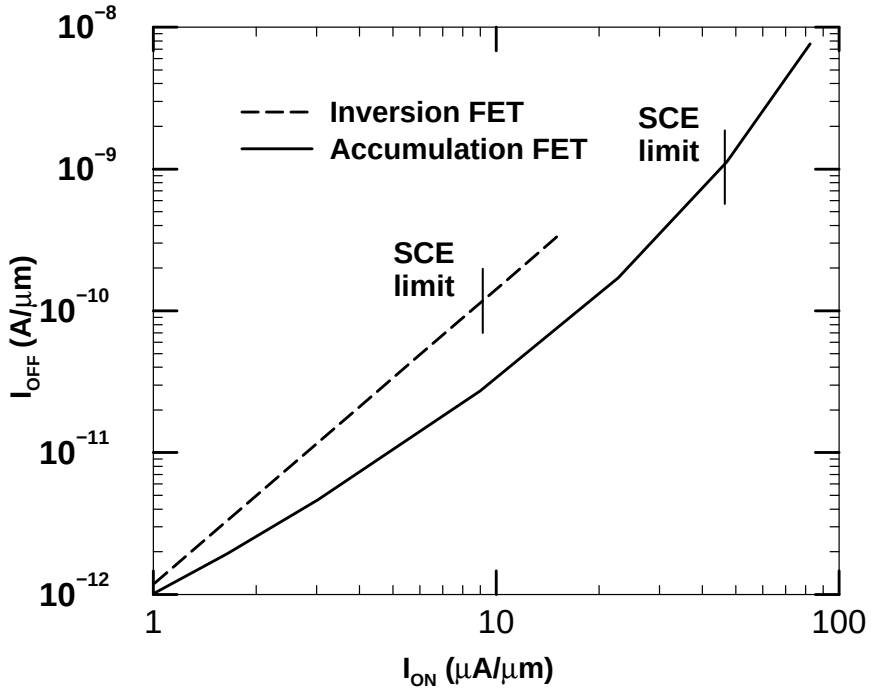


Figure 43: Off-current vs. on-current for low V_{dd} applications (from 2-D classical DD simulations); $L = 30\ nm$, t_{ox} =10 Å, $y_j = 12\ nm$, $V_{dd} = 0.5\ V$, and gate workfunction is 4.63 eV. The SCE limit is taken to be $S < 100\ mV/dec$.

transport can be neglected at low drive currents. The SCE limit is taken to be $S < 100\ mV/dec$. Accumulation FETs are seen to have a lower I_{OFF} and a lower SCE (for a given I_{ON}) compared to SCI FETs. Thus, single metal gate accumulation FETs are also useful for ultra-low power applications.

6.4 Summary

The accumulation FET has an increased depletion depth compared to the inversion FET. In the long-channel regime, this results in a lower subthreshold swing; in the short-channel regime, this results in worse SCE compared to the inversion FET. So, if the gate WF is tunable, short-channel inversion MOSFETs have better SCE control than accumulation FETs. For mid-band metal gate HiP and LOP applications, there is a disparity between drift-diffusion and hydro-dynamic simulations; Monte Carlo simulations would have to be used to accurately assess the scaling potential of accumulation and inversion MOSFETs for such applications. For HiP applications with single metal gate CMOS, well-designed

halos result in better performance using accumulation FETs than with inversion FETs. For ultra-low power, single metal gate CMOS applications, accumulation FETs perform much better than inversion FETs due to lower SCE.

CHAPTER VII

POLY GATE MOSFETS

7.1 Introduction

Poly-gate MOSFETs have been the workhorse of CMOS technology. Inversion n-FETs use n+ poly as the gate material to obtain enhancement operation. Accumulation n-FETs need to use p+ poly as the gate material in order to obtain enhancement mode operation. When poly gates are used, V_T of the transistors will have to be adjusted by changing the channel/substrate doping.

Fig. 44 plots V_T vs. channel doping, and it can be seen that the SCA FET has a V_T of around 1 V for a wide range of channel doping, whereas the SCI FET can achieve a high V_T only with very high channel and substrate doping. At such a high doping, the SCI FET has many shortcomings:

- very low channel mobility
- high band-to-band tunneling (BTBT) leakage at the drain-substrate junction
- high subthreshold slope (since d decreases as N_A increases)

In contrast, the SCA FET uses a lightly doped channel and has a low S, thus resulting in good on-current and off-current. Also, BTBT leakage in accumulation FETs can be kept to low levels since a p-n junction is absent in the lateral direction in the channel.

Fig. 44 shows that the UD-SCI FET needs a doping in excess of $1\text{x}10^{19}$ cm^{-3} to achieve a V_T of 0.8V. The curves in Fig. 44 were generated using the Boltzmann distribution, and did not include the effect of QM V_T shift. Even if the Fermi-Dirac distribution (which would be more accurate at high doping) is used and QM effects are included, a very high doping ($>6\text{x}10^{18}$ cm^{-3}) would still be needed to achieve a V_T of 0.8 V. Fig. 45 shows a plot of S vs. V_T for the inversion and accumulation FETs. Comparing this to Fig. 44, it can be seen that the SCA FET has a better S than the SCI FET for V_T >0.5V. In addition to

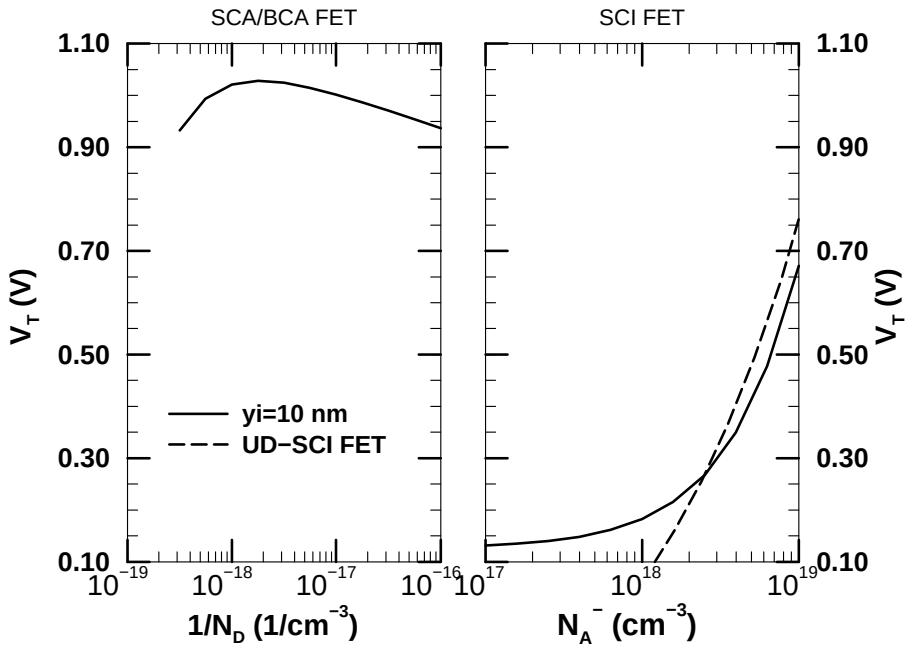


Figure 44: Threshold voltage vs. tub-doping for poly-gate accumulation and inversion MOSFETs (from models); $L = 70\ nm$, t_{ox} =15 Å, $y_j = 20\ nm$, $V_{dd} = 1.0\ V$, N_A =2.5x10^{18} cm^{-3}. The gate workfunction is 5.19 eV for accumulation FETs, and 4.19 eV for inversion FETs.

SCE models, models for BTBT leakage, and gate tunneling are needed to comprehensively compare poly gate accumulation and inversion MOSFETs.

7.2 Band to Band Tunneling Leakage

As the channel/substrate doping in a MOSFET increases, BTBT leakage at the drain-substrate and drain-channel region becomes appreciable. The BTBT leakage model is derived by making the triangular barrier approximation for a p-n junction [43]:

$$J_t = \frac{\sqrt{2m^*}q^3EV}{4\pi^2\hbar^2E_g^{1/2}} exp\left(-\frac{4\sqrt{2m^*}E_g^{1/2}}{3qE\hbar}\right) \tag{42}$$

where E is the electric field at the junction, E_g is the bandgap, V is the applied voltage, m^* is the effective mass, and $\hbar = h/(2\pi)$ (h is Planck's constant). The electric field at the drain-substrate junction can be estimated to a first order by 1-D electrostatics. The electric field expression is obtained from abrupt p-n junction theory:

$$E = A0.\frac{qN'_{D_D}x_n}{\epsilon_s} \tag{43}$$

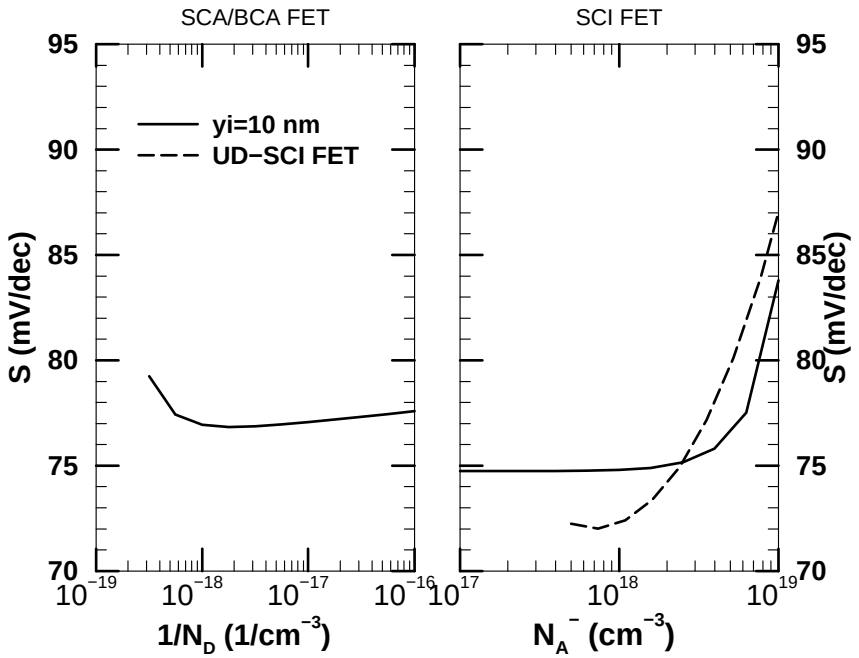


Figure 45: Subthreshold swing vs. tub-doping for poly-gate accumulation and inversion MOSFETs (from models); L = 70 nm, t_{ox} =15 Å, y_j = 20 nm, V_{dd} = 1.0 V, N_A =2.5x10^{18} cm^{-3}. The gate workfunction is 5.19 eV for accumulation FETs, and 4.19 eV for inversion FETs.

where $N'_{D_D} = A1.N_{D_D}$ is the modified drain doping, and x_n, the depletion depth on the drain side is given by

$$x_n = \sqrt{\frac{2\epsilon_s}{q}\left(\frac{N'_A}{(N'_A + N'_{D_D})N'_{D_D}}\right)\psi_{bi}} \tag{44}$$

$N'_A = A2.N_A$ is the modified channel doping. $A0$, $A1$, and $A2$ are model parameters to account for non-uniform doping at the drain-substrate junction. The model is compared to measurements from [91], Fig. 46. A good match is seen between the model and measurements.

7.3 Accumulation vs. Inversion FETs

In this chapter, BTBT leakage is assumed to occur only in the drain-channel region, since the drain-substrate region can be tailored to be p-i-n type to avoid BTBT leakage. In an accumulation FET, considerable BTBT leakage occurs only if the substrate doping is high and the tub-depth is much less than the junction depth. For poly-gate applications, the substrate doping of accumulation FETs is kept below 2.5x10^{18} cm^{-3}, resulting in negligible

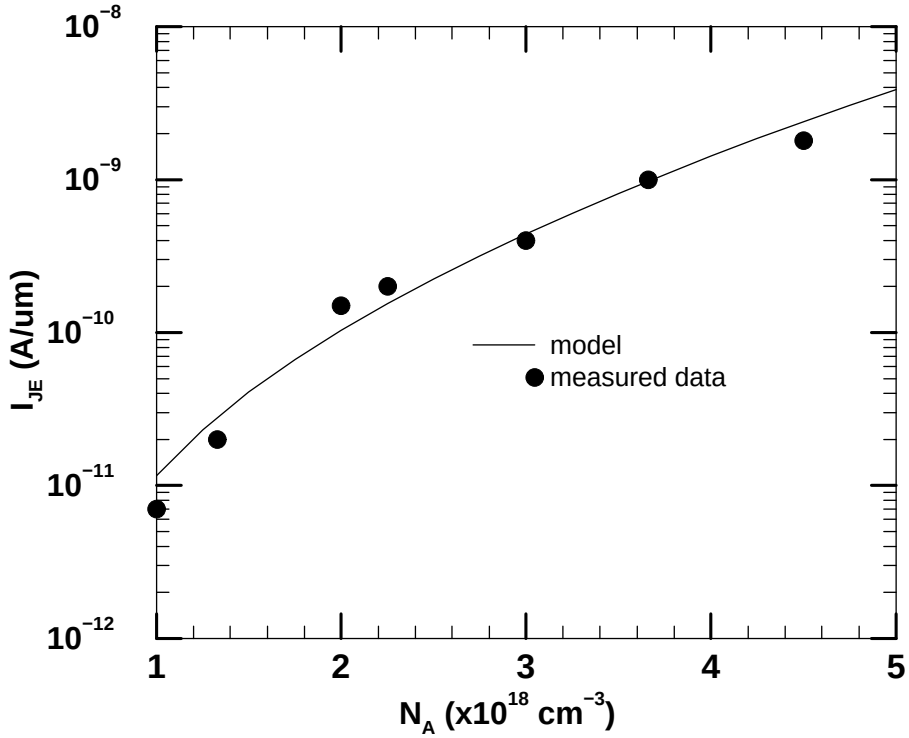


Figure 46: BTBT leakage at drain-channel region vs. channel doping: model verification with measured data from [91]; $V_{DS} = 1.3\ V$, and $T = 100\ ^oC$.

BTBT leakage. In an inversion FET, the channel and substrate doping need to be high to achieve high V_T. Fig. 47 shows subthreshold and BTBT leakage vs. doping, for inversion MOSFETs. It can be seen that BTBT leakage limits the minimum off-current of inversion MOSFETs to 60 pA/um. The ITRS [1] recommended off-current for low standby power (LSTP) systems is 1 pA/um or less. The SCI FET is unable to achieve I_{OFF} less than 1 pA/um because of the onset of BTBT leakage.

In an SCI FET, the electric field points into the oxide-semiconductor interface (looking from the top), and hence attracts carriers towards the interface. In a BCA FET, the electric field points out of the $\mathsf{Si} - \mathsf{SiO_2}$ interface; thus, in a BCA FET, electrons experience a force that repels them away from the $\mathsf{Si} - \mathsf{SiO_2}$ interface and deeper into the tub region. Hence, a BCA n-MOSFET has a higher barrier for electrons tunneling through the oxide, Fig. 48. Thus, the accumulation n-MOSFET can use a thinner oxide than the SCI n-MOSFET, and thereby achieve a better performance. The tunneling model used is similar to that in [92], and is explained in Appendix E.

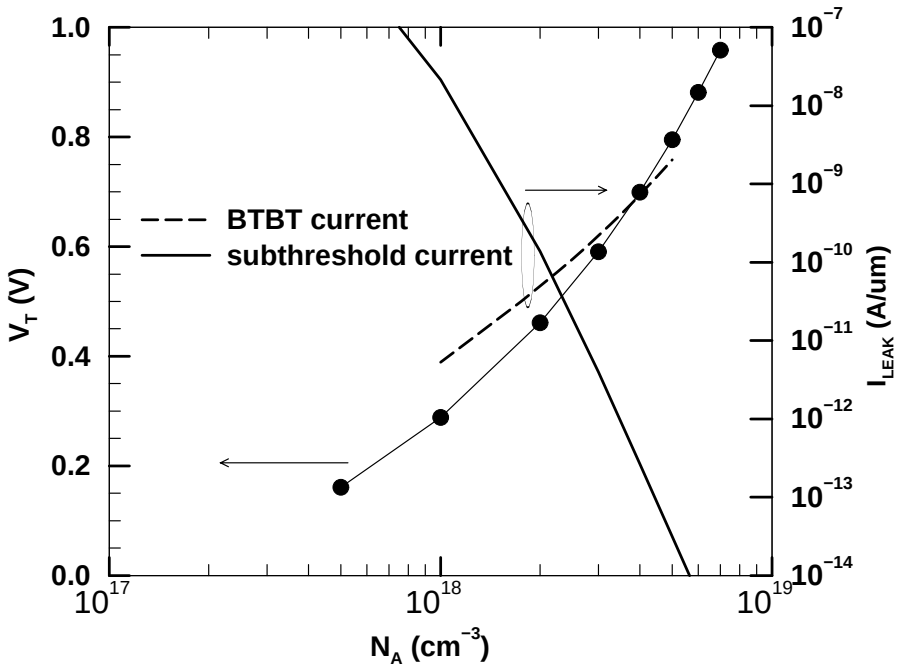


Figure 47: BTBT leakage at drain-channel region, and subthreshold leakage (from models) for poly-gate, uniformly-doped, inversion n-MOSFETs; $L = 45\ nm$, t_{ox} =15 Å, $V_{dd} = 1.0\ V$, and gate workfunction is 4.63 eV.

A chain of inverters is simulated using the tick-based circuit simulation methodology to find the power-frequency ($P - f_{clk}$) performance of circuits using either accumulation or inversion FETs. Fig. 49 shows $P - f_{clk}$ curves for inverter chains using either accumulation or inversion FETs. The devices use halo doping to obtain reduced SCE. It can be clearly seen that for high performance applications, the SCI FET chain performs better, whereas for LSTP applications, the SCA FET chain performs better. For LSTP applications, the SCA FET has more than an order of magnitude lower standby power than the SCI FET (for the same dynamic power and frequency).

7.4 Summary

For dual polysilicon gate CMOS, accumulation FETs have an advantage over inversion FETs for low standby power applications. To achieve a high V_T, the SCI FET needs a very high channel channel doping. This causes a low channel mobility (and hence a low I_{ON}), high BTBT leakage at the drain-channel junction, high σV_T and a high S. On the other

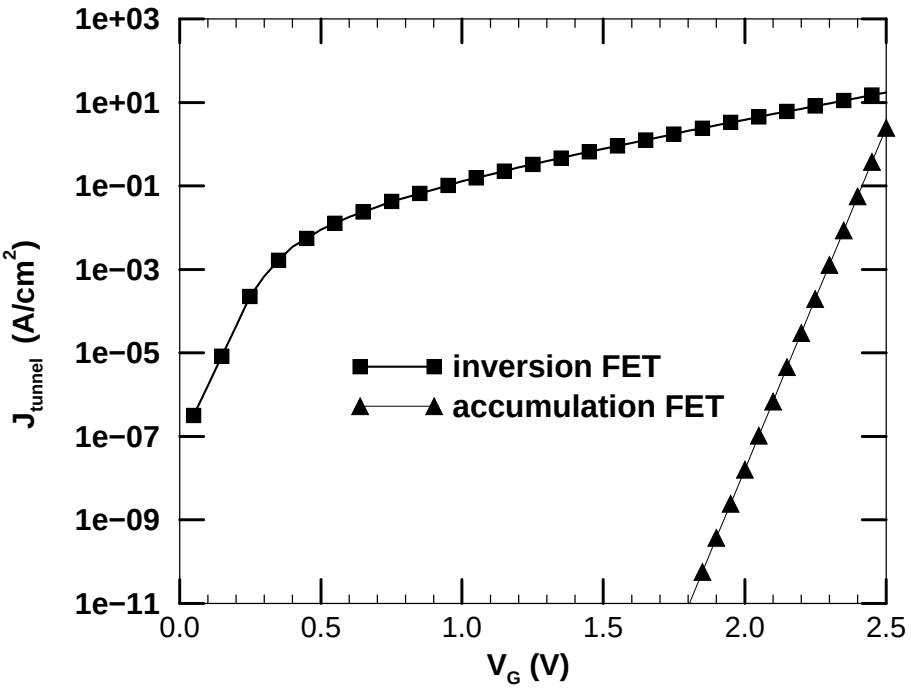


Figure 48: Gate tunneling current density (from models) comparison of poly-gate BCA/SCA and SCI n-FETs for 15 Å gate oxide.

hand, the SCA FET can achieve a high V_T by using a lightly doped channel and a substrate doping that is low enough to avoid BTBT leakage, but high enough to control SCE. Thus the SCA FET has a high channel mobility, negligible BTBT leakage and a low σV_T. Also, because of a higher barrier for gate-channel electron tunneling in an accumulation n-FET, it has a much lower gate oxide tunneling than an inversion n-FET. A chain of inverters using either inversion or accumulation FETs are simulated, and it is found that for LSTP applications, an SCA FET inverter chain has a 1-order of magnitude lower static power (for the same dynamic power and frequency) than an SCI FET inverter chain.

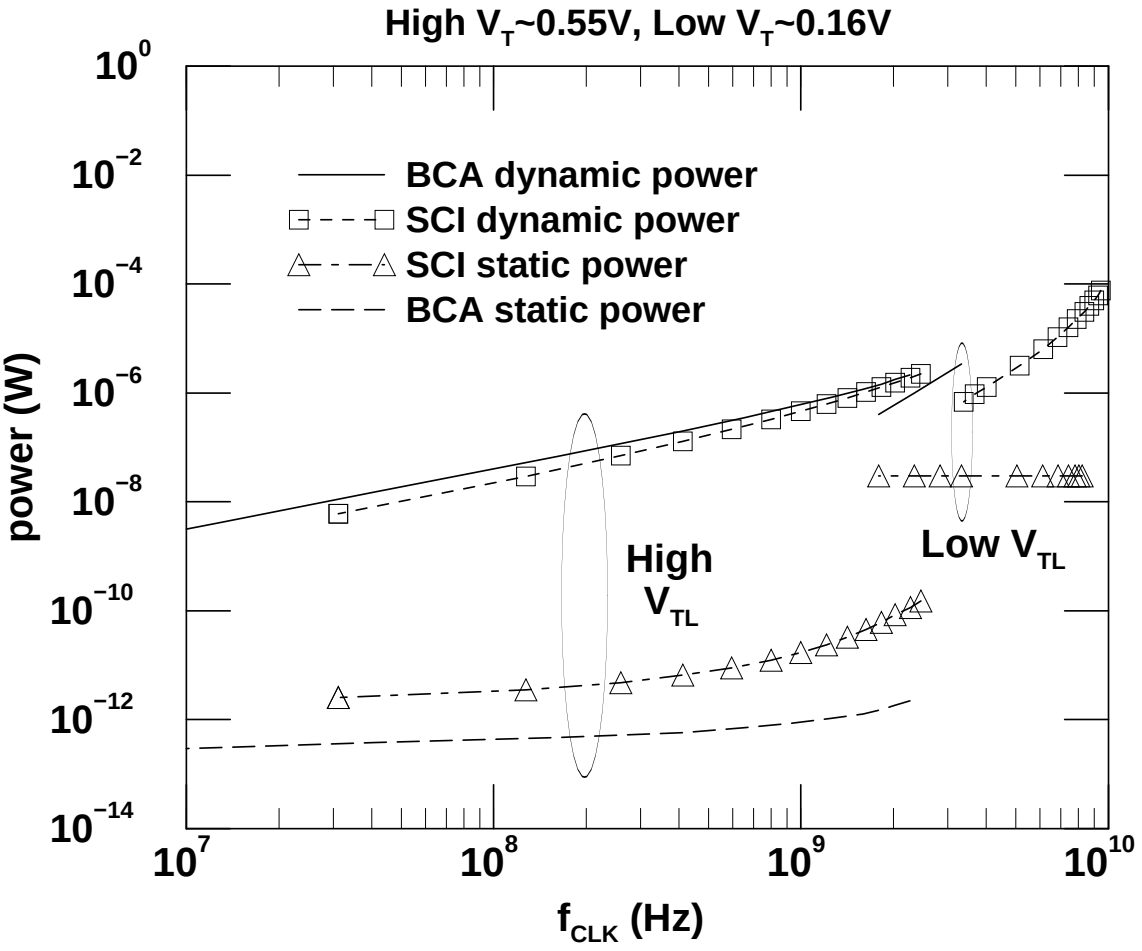


Figure 49: $P - f_{clk}$ for poly gate accumulation and inversion FET inverter chains with L_{eff}=50nm. High V_T case: t_{ox}=15Å for SCA FET, t_{ox}=23Å for SCI FET, y_j=32nm, y_{iBCA}=10nm, N_{ABCA}=2.4x10^{18} cm^{-3}, FO=2, C_w=0, n_{cp}=12 ; low V_T case: t_{ox}=15Å, y_j=y_{iBCA}=10nm, N_{ABCA}=8x10^{18} cm^{-3}, FO=2, C_w=10fF, n_{cp}=7. V_{dd} varies in both high and low V_T cases.

CHAPTER VIII

MONTE CARLO DEVICE SIMULATION

8.1 Introduction

When simulating semiconductor devices, a variety of transport models can be used: drift-diffusion, hydro-dynamic, solutions of the Boltzmann transport equation (BTE), and quantum transport. The accuracy of a device simulation depends on the transport model used. In the drift diffusion approach, the first two moments of the BTE are used to describe carrier transport (the two resulting equations are the continuity equation, and the current flow equation) [93]. The two equations are coupled to Poisson's equation, and solved self-consistently to find the carrier density and electrostatic potential. The drift-diffusion approach assumes that the current changes slowly on the scale of the momentum relaxation time, and does not treat nonlocal effects like velocity overshoot.

For Si MOSFET channel lengths less than a few hundred nanometres, non-local effects become important. These effects can be captured by the momentum-energy balance equations. In this approach, the energy gradient is included in the current equation, and mobility is a function of carrier energy (rather than the local electric field as in the drift-diffusion approach). In the hydro-dynamic/energy transport approach, the first three moments of the BTE are solved; the extra equation that results is the energy balance equation. There are many forms of the HD model because of different assumptions made in deriving them [94]. HD model parameters are usually extracted from a self-consistent MC simulation or from measurements.

For small devices (for Si MOSFET channel lengths less than a few tens of nanometres), the BTE should be solved. The BTE is a semi-classical approach to describing carrier transport - it describes the evolution of the carrier trajectory by using a combination of classical mechanics, and quantum mechanics (by way of probabilistic scattering rates) [95].

The BTE is given by

$$\frac{\partial f}{\partial t} + \vec{v}.\nabla_r f + \frac{\epsilon \vec{E}}{\hbar}.\nabla_p f = \left.\frac{\partial f}{\partial t}\right|_{coll} + s(\vec{r},\vec{p},t) \tag{45}$$

where,

$$\nabla_r f = \frac{\partial f_x}{\partial x}\hat{x} + \frac{\partial f_y}{\partial y}\hat{y} + \frac{\partial f_z}{\partial z}\hat{z} \tag{46}$$

$$\nabla_p f = \frac{\partial f_x}{\partial p_x}\hat{x} + \frac{\partial f_y}{\partial p_y}\hat{y} + \frac{\partial f_z}{\partial p_z}\hat{z} \tag{47}$$

$f(\vec{r},\vec{p},t)$ is the distribution function that gives the probability of finding carriers at time t, located at a position $\vec{r}$ with momentum $\vec{p}$, and $\vec{E}$ is the electric field. The BTE accounts for all possible mechanisms by which f might change including collisions and external fields. For classical particles, $\vec{v} = \vec{p}/m^*$, whereas for electrons in semiconductors, $\vec{v} = \nabla_p E(\vec{p})$. $\left.\frac{\partial f}{\partial t}\right|_{coll}$ describes collisions, whereas $s(\vec{r},\vec{p},t)$ describes carrier generation recombination processes.

Under low fields when the scattering is elastic and isotropic, the relaxation time approximation is widely used - this greatly simplifies the solution of the BTE. But there exist many situations where the collisions are neither elastic nor isotropic - e.g. electrons with a few tens of eV energy, and in these cases numerical techniques are necessary. Proposed numerical approaches to solve the BTE include Monte Carlo, cellular automata, scattering matrix, and spherical harmonic methods. Of these, the Monte Carlo method has achieved the most success because of its ease of programming. For Si MOSFET channel lengths less than a few nanometres, the quantum transport equation would have to be solved to accurately simulate device behavior. Some methods proposed for solving the quantum equation are Wigner transformation [96], and non-equilibrium Green's function method [97].

8.2 The Monte Carlo Approach

The motivation behind using MC methods is that it is an elegant way to solve the BTE. The BTE is an integro-differential equation, and very difficult to solve; it is much easier to simulate the trajectories of individual carriers as they move through a device under the influence of electric fields and random scattering forces. If a large number of such paths are simulated, then the average value of the extracted parameters will closely agree with

the exact solution of the BTE. MC method is frequently the standard against which the validity of simpler approaches is gauged.

There are two possible ways to use the Monte Carlo technique to simulate semiconductor devices [93]. In the ensemble Monte Carlo (EMC) approach, an ensemble of particles are followed within the device, in parallel. In the single particle Monte Carlo (SPMC) approach, carriers are followed one at a time - the carrier is injected from a contact and followed until it exits the device through the same or another contact. To simulate each operating point of a MOSFET using the single particle MC simulator SPARTA [98], takes about 12-16 hours. The same takes 24-30 hrs. using the ensemble MC simulator DEGAS [99]. For comparison, drift-diffusion (DD) simulations take less than a minute, and hydro-dynamic (HD) simulations take about 10 minutes. There are two possible ways to keep MC simulations to a minimum.

In the first method, a calibrated drift diffusion method can be used; the saturation velocity (v_{sat}) would have to be altered to account for velocity overshoot. The exact value of v_{sat} necessary in the DD simulation can be calibrated with the help of a MC simulation at high V_{DS} [100]. The drawback of such an approach is that the value of v_{sat} would be different for different channel doping (since velocity overshoot is dependent on channel mobility and channel length). Thus, v_{sat} would have to be individually extracted from MC simulations for different structures, which would not result in any savings in time compared to the MC simulation approach. The second approach is to use a calibrated hydro-dynamic method; surface roughness mobility parameters in the HD simulation are adjusted so that the on-current matches well with MC simulations. The calibration is explained in further detail in Appendix F. The reliability of this method is tested against MC simulation results, Table 13. It can be seen that HD simulations have an error within 6% of MC simulations, which is acceptable given the huge savings in computation time.

Using the single particle Monte Carlo (MC) simulator SPARTA [98], halo-doped accumulation and inversion MOSFETs are simulated. Before comparing the DD simulation output with that of MC simulation for high drain voltages, the DD mobility model parameters should be adjusted so that the output agrees well with MC simulations for low

Table 13: On-current comparison from various simulations; $L = 30\ nm$, t_{ox} =15 Å, $y_j = 12\ nm$, $V_{dd} = 1.0\ V$, N_A =1x10^{18} cm^{-3}, $N_D = N_A^-$ =1x10^{17} cm^{-3} and $y_i = 20\ nm$. The devices are halo-doped. The source/drain doping is 5x10^{19} cm^{-3}.

device	V_{dd} (V)	DD I_{ON} (uA/um)	HDWOP I_{ON} (uA/um)	HDWP I_{ON} (uA/um)	MC I_{ON} (uA/um)	% error in HDWP w.r.t. MC result
accumulation	0.9	295	647	351	354	0.0
accumulation	1.0	431	-	530	513	3.3
inversion	0.9	269	-	310	328	5.4

drain voltages [89]. In drift-diffusion simulations, the surface roughness mobility parameters need to be adjusted to match MC simulation output, at low drain voltages. The Darwish mobility model [101] is used to model mobility degradation due to surface roughness. The surface scattering ratio, s (where s=1.0 means that the scattering is purely specular, and s=0 means that the scattering is purely diffuse), is taken to be 0.85 in MC simulations [89].

Fig. 50 compares the channel velocity of four simulations - DD, MC, HD with parameters fit to MC simulation (HDWP) and HD without any parameter fit (HDWOP). It can be seen that HDWOP predicts a huge velocity overshoot in the channel compared to the MC simulation. The source side velocity is what determines the on-current in a MOSFET. Fig. 51 plots velocity (at different distances from the source) vs. on-current. It can be seen that the points lie on a straight line, confirming that source-side velocity (and not peak channel velocity) determines the on-current. From the MC simulations, it is found that for $L_{eff} = 30\ nm$ accumulation MOSFETs with halos, on-current increases by about 20% because of velocity overshoot, Table 13. HDWOP simulations predict more than a 100% increase in on-current (compared to DD simulations). The increase is about 21% with HDWP, for inversion MOSFETs.

8.3 Mid-bandgap Metal Gate MOSFETs Revisited

In section 6.3, mid-bandgap metal gate n-MOSFETs were compared. It was found that for halo doped MOSFETs with a channel length of 30 nm, accumulation MOSFETs had a

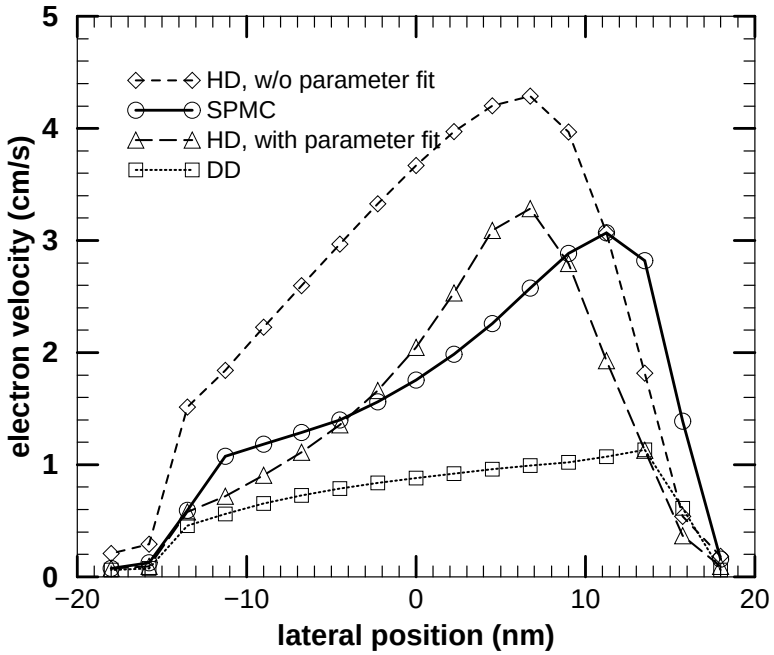


Figure 50: Absolute electron velocity vs. lateral position for a mid-bandgap metal gate accumulation n-MOSFET (from 2-D simulations), 1 nm below the $Si - SiO_2$ interface; $L = 30\ nm$, t_{ox} =15 Å, $y_j = 12\ nm$, $V_{GS} = V_{DS} = 1.0\ V$, N_A =1x10^{18} cm^{-3}, N_D =1x10^{17} cm^{-3}, and $y_i = 20\ nm$. QM V_T shift was included by means of a 70 mV shift in the gate workfunction. The accumulation MOSFET is halo-doped with a peak halo doping of 1x10^{19} cm^{-3}, and a source/drain doping of 5x10^{19} cm^{-3}. The surface-scattering ratio in the Monte Carlo simulations is taken to be 0.85. In the legend, SPMC refers to Single Particle MC simulation.

marginal advantage over inversion MOSFETs for hi-performance applications. The transport model used to arrive at this result was hydro-dynamic, with default mobility model parameters. However, it was seen in the previous section that HDWOP severely overestimates on-current increase due to velocity overshoot. Thus, the comparison would have to be repeated using a calibrated mobility model, using the HD transport model. Alternatively, MC simulations can be used to perform the comparison.

MOSFETs with a channel length of 30 nm, and with halos, are simulated using HDWP, and Monte Carlo simulations, Fig. 52. The simulations assume a gate workfunction of 4.70 eV, including a gate workfunction shift of 70 mV to account for quantum mechanical centroid shift. The maximum allowable subthreshold swing is assumed to be 110 mV/dec. It can be seen that once again HDWP simulations have less than a 5% error compared

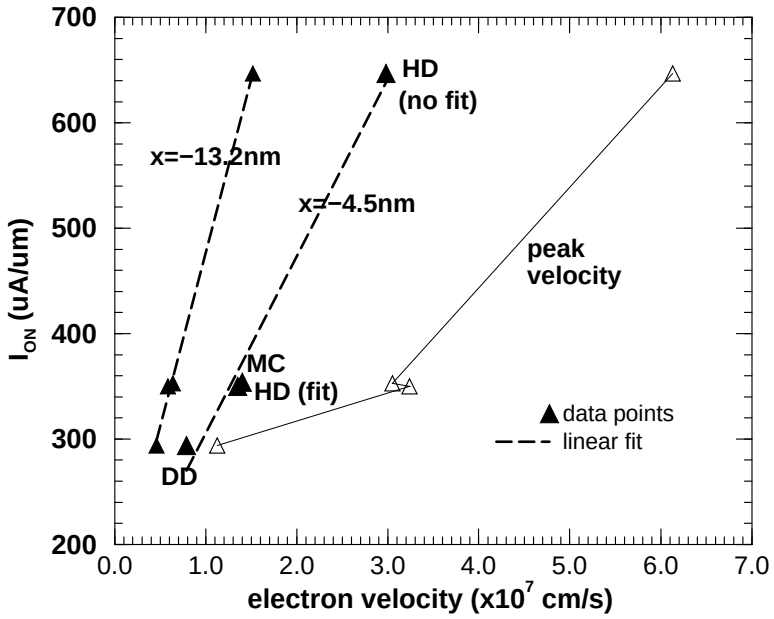


Figure 51: On-current vs. electron velocity for a mid-bandgap metal gate accumulation n-MOSFET (from 2-D simulations); the source is at $x = -15\ nm$. $L = 30\ nm$, t_{ox} =15 Å, $y_j = 12\ nm$, $V_{dd} = 1.0\ V$, N_A =1x10^{18} cm^{-3}, N_D =1x10^{17} cm^{-3}, and $y_i = 20\ nm$. QM V_T shift was included by means of a 70 mV shift in the gate workfunction. The accumulation MOSFET is halo-doped with a peak halo doping of 1x10^{19} cm^{-3}, and a source/drain doping of 5x10^{19} cm^{-3}. The surface-scattering ratio (s) in the MC simulation is taken to be 0.85.

to MC simulations. Accumulation MOSFETs are seen to have a 17% greater maximum on-current compared to inversion MOSFETs. A second set of simulations is done using the HDWP method with the van Dort model [87] to include quantum correction, Fig. 53. It can be seen that the on-current advantage of accumulation MOSFETs increases from 17% to 45% when quantum effects are included by means of the van Dort method. The increase in advantage is because accumulation MOSFETs have a lower vertical field resulting in a lower QM V_T shift compared to inversion MOSFETs.

8.4 Summary

- Drift-diffusion simulations do not include non-local effects, and thus cannot be applied to MOSFETs with channel lengths less than a few hundred nanometres. HD simulations need to be used to include non-local effects. However, there are a number of ways of implementing the HD models. In the literature, there is no clear consensus

as to which HD method is most effective to accurately simulate non-local effects.

- MC simulations give reference data in the absence of experimental data. HD mobility model parameters can be calibrated to give the same output as MC simulations. HD simulations with calibrated mobility model parameters give on-currents within 6% of the MC simulation output, with huge savings in computation time.

- For 30 nm MOSFETs with halos, with $V_{dd} = 1.0\ V$ and a 1 nm gate oxide thickness, accumulation MOSFETs have a 25% greater on-current compared to inversion MOSFETs (for the same off-current). Also, the maximum achievable on-current is 45% greater with accumulation MOSFETs (for $S < 110\ mV/dec$).

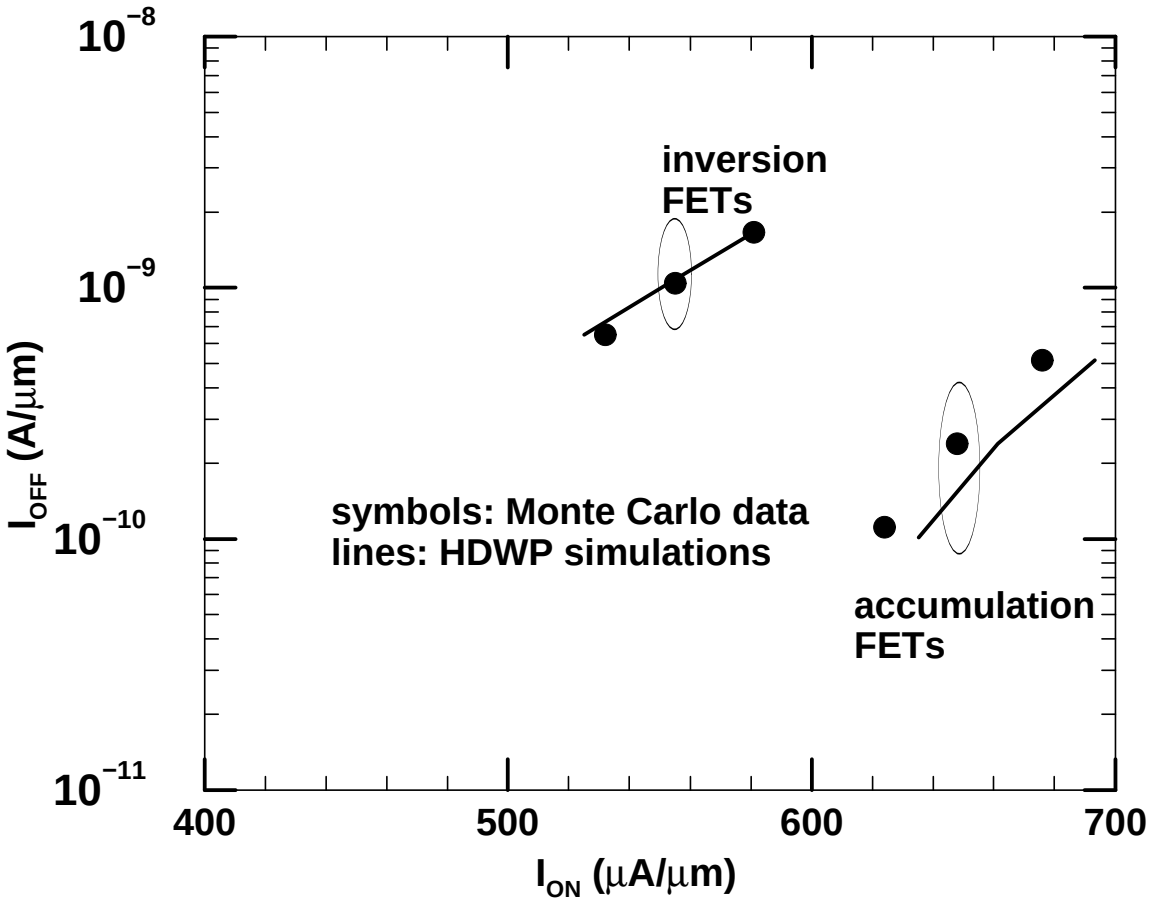


Figure 52: Off-current vs. on-current for mid-bandgap metal gate n-MOSFETs; $L = 30\ nm$, t_{ox} =10 Å, $y_j = 12\ nm$, and $V_{dd} = 1.0\ V$. The MOSFETs are halo-doped with a peak halo doping of 1x10^{19} cm^{-3}, and a source/drain doping of 5x10^{19} cm^{-3}. The maximum subthreshold swing is taken to be 110 mV/dec. Quantum mechanical V_T shift is included by means of a 70 mV shift in the gate workfunction.

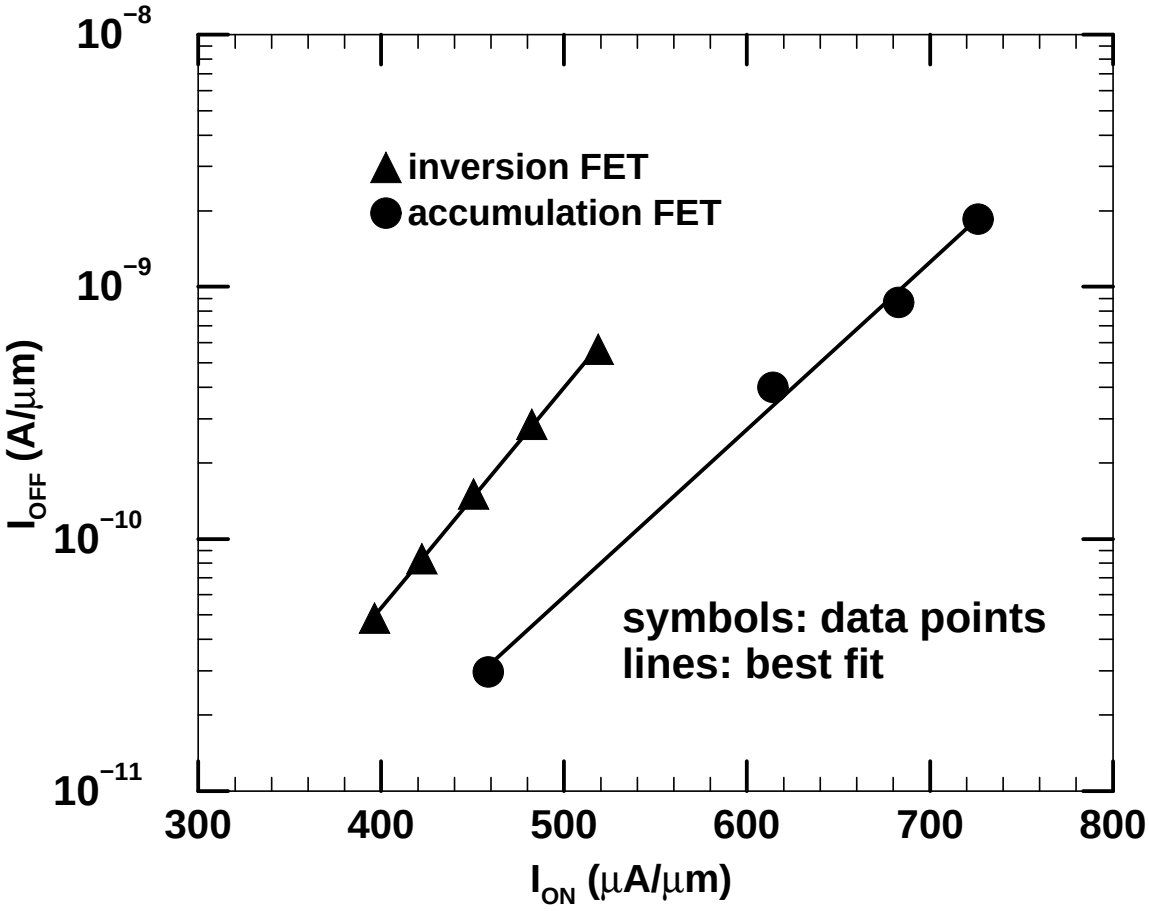


Figure 53: Off-current vs. on-current for mid-bandgap metal gate n-MOSFETs (from HDWP simulations); $L = 30\ nm$, t_{ox} =10 Å, $y_j = 12\ nm$, and $V_{dd} = 1.0\ V$. The MOSFETs are halo-doped with a peak halo doping of 1x10^{19} cm^{-3}, and a source/drain doping of 5x10^{19} cm^{-3}. The maximum subthreshold swing is taken to be 110 mV/dec. Quantum mechanical V_T shift is included by means of van Dort method.

CHAPTER IX

SCALING LIMITS

9.1 Introduction

MOSFET scaling is one of the main drivers of silicon technology - in addition to gains in increased functionality per unit area, and speed, there is also a reduction in the amount of dynamic power dissipated per cycle per device. But device scaling is limited by SCE and related phenomena, and process limitations. Process variations cause the MOSFET gate length to vary across the die, resulting in a 3σ value of about 10-30% of the gate length. This results in MOSFETs with various channel lengths across the wafer, and the MOSFET with the shortest length (say, L_G=$L_{nom} - 3\sigma$) will have the worst off-current. The nominal channel length MOSFET will have to be designed such that the worst case device (i.e. L_G=$L_{nom} - 3\sigma$) has a leakage within one to two orders of magnitude of the nominal device. This translates to a maximum V_T reduction of 100-200 mV for the worst case device compared to the nominal device ($I_{OFF} = I_0 10^{-V_T/S}$).

Process variations also cause variations in oxide thickness, and dopant fluctuation effect, but these effects have a far lesser impact on device scaling than channel length variation. In case of SOI MOSFETs, variation in silicon film thickness may also limit device scaling [102]. The MOSFET channel length limit depends on the particular application [103] and the gate technology [104]. The different application areas defined by ITRS (Table 1) have different off-current and V_T requirements. This leads to different limits on the maximum channel doping, and minimum oxide thickness for the various applications, leading to different minimum effective channel lengths (L_{min}). Previous efforts on deriving scaling limits have many short comings. Most previous works on determining the device scaling limit do not consider application dependent limits [105], [106]. Frank [103] derived scaling limits for bulk SCI MOSFETs for various applications, but considered tunable workfunction gates only. The minimum channel length in [103] is mainly based on BTBT leakage considerations

at the drain-channel region.

In this chapter, different bulk MOSFETs - accumulation, retrograde doped SCI (RD-SCI), and uniformly doped SCI (UD-SCI) - will be considered while evaluating device scaling limits. Different application areas as well as different gate technologies will be considered. The scaling limits will be derived both from 2-D simulations and SCE models. Such an effort clarifies the advantage of various device types, and gate technologies, for different applications.

9.2 Minimum Channel Length Determination

Random variations in gate length (and hence effective channel length) is the limiting factor in device scaling. The 3σ value of gate length is anywhere from 10-30% of the nominal value. The device with the worst-case on-current is one with a gate length of $L_{nom} + 3\sigma$, whereas the device with the worst-case off-current is one with a gate length of $L_{nom} - 3\sigma$. Both the reduction in on-current and increase in off-current will have to be considered when evaluating the nominal channel length for a given system. But for determining the minimum channel length, only the increase in off-current needs to be considered; the increased off-current will limit integration (i.e., the maximum number of devices on a die). The decreased on-current among longer-channel devices does not limit integration; it limits the maximum speed at which the system can operate, and this can be taken care of by tweaking V_{dd}, V_T, and width of the devices.

There are a number of ways to determine the minimum channel length. Assuming a maximum of one-order of magnitude off-current increase (when channel length reduces from L_{nom} to $L_{nom} - 3\sigma$), the maximum V_T shift is limited to 100 mV. This criterion is the one that is directly related to limits imposed on device design by integration requirements and process limits. Other criteria that can be used to quantify the minimum channel length are maximum allowable DIBL (usually less than 100 mV/V), maximum allowable subthreshold swing (usually less than 120 mV/dec), and maximum allowable V_T shift from the long-channel value. Note that these criteria are imposed on the nominal channel length device and are an indirect measure of the impact of integration requirements and process

limitations on device design.

Both 2-D simulations (with drift-diffusion transport), and SCE models are used to derive the minimum channel length. RD-SCI and accumulation devices are assumed to have a channel depth of 10 nm. For accumulation FETs, it is seen from Fig. 25 that for $y_i > 15\ nm$, the SCE limit is not satisfied for most channel doping. In case of RD-SCI FETs too, having a channel depth greater than about 15 nm will lead to a higher SCE. The RD-SCI FET is assumed to have a low-doped tub (with a doping of 1x10^{17} cm^{-3} or less). From Fig. 24 it can be seen that a high N_D results in high process sensitivity, and thus dV_T/dy_i limits the maximum allowable N_D. Accumulation MOSFETs are assumed to have a maximum allowable channel doping (N_D) of 5x10^{18} cm^{-3}. All devices are assumed to have a source/drain junction depth of 12 nm with a 5x10^{20} cm^{-3} doping. Poly-gate devices are assumed to have a poly-depletion of 0.2 nm [50]. The supply voltage is taken to be 1 V for all applications. All channel length values reported are effective values.

The following procedure is followed to obtain the minimum channel length.

- Based on BTBT leakage requirements at the drain-substrate region, the maximum allowable substrate doping is determined. BTBT leakage current is limited to 50% of the allowable subthreshold leakage current. In UD-SCI devices, BTBT leakage is assumed to occur over a 10 nm length, independent of channel length. In accumulation and RD-SCI devices, BTBT leakage is assumed to occur over a 2 nm length, independent of channel length, Fig. 54.

- The minimum allowable oxide thickness ($t_{ox}|_{min}$) would need to be calculated based on the allowable gate leakage current (I_{gate}). Assuming that hi-K dielectrics are possible in the future, the oxide thickness is assumed to be 10 Å for HiP and LOP applications. For LSTP applications using inversion devices, it is assumed to be 23 Å. For LSTP applications using accumulation MOSFETs, $t_{ox}|_{min}$ is calculated using the gate-channel tunneling current model (Appendix E), by making $I_{gate} < 0.1 I_{sub}$.

- The gate workfunction is fixed to 4.63 eV for mid-bandgap metal gate FETs, to

4.19 eV for poly-gate inversion n-FETs, and to 5.19 eV for poly-gate accumulation n-FETs. In case of tunable workfunction metal gates, the gate workfunction is assumed to be able to take on any value between 4 eV to 5.3 eV.

- The device with the highest substrate doping (N_A) results in the lowest L_{min}. So, in case of tunable workfunction gates, all devices are assumed to have the highest allowable N_A.

- A device with the highest allowable N_A may not meet the V_T criterion. In case the V_T has to be increased, either L can be increased, or t_{ox} can be increased. In case V_T has to be decreased, N_A can be decreased or if the device is an accumulation FET, N_D (tub-doping) can be increased.

- Using 2-D numerical simulations, the minimum channel length that meets the SCE requirement ($S < 110\ mV/dec$) is determined. The search can be done by a binary search.

- Using SCE models, the minimum channel length can be computed much faster than using 2-D simulations. From Fig. 7 it can be seen that V_T rolloff follows a universal curve. This curve shows that to achieve a V_T rolloff of less than 120 mV, $L\lambda_{eq_1}$ needs to be less than π. At this point, a 30% reduction in channel length results in about a 100 mV reduction in V_T. Thus, for $\Delta V_T < 120\ mV$, and $\Delta V_T(30\%\delta L) < 100mV$, $L_{min} = \pi/\lambda_{eq_1}$. V_T for the device with $L = L_{min}$ is simply $V_{T_{long}} - 0.12\ V$. λ_{eq_1} can be determined analytically from (10), and is a function of the minimum channel depletion depth, source/drain junction depth, and oxide thickness.

Using the BTBT leakage model described in section 7.2, the maximum allowable N_A is determined, Fig. 55. Tables 14-19 show the minimum channel lengths determined from both 2-D numerical simulations, and SCE models, for various gate technologies and applications. Table 20 shows L_{min} computed from both models and simulations. It can be seen that for tunable workfunction gates, inversion devices are always better than accumulation devices. For mid-bandgap metal gates, accumulation MOSFETs have a lower L_{min} than

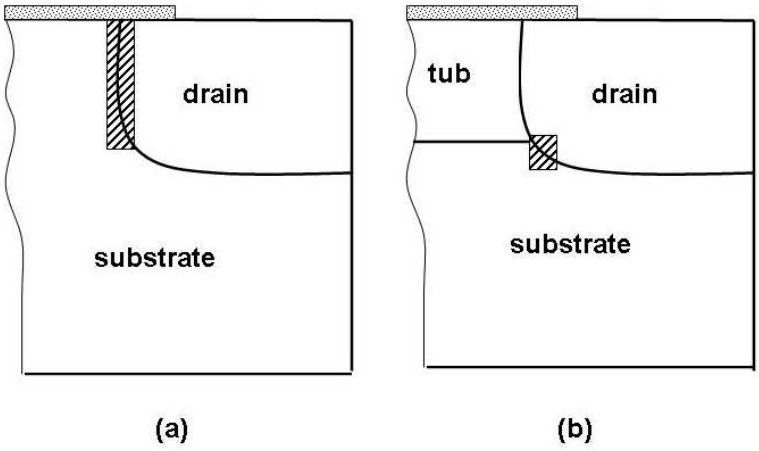


Figure 54: BTBT leakage - the hatched area shows the BTBT leakage region: (a) UD-SCI MOSFET showing BTBT leakage over a 10 nm length; (b) accumulation and RD-SCI MOSFETs showing BTBT leakage over a 2 nm length.

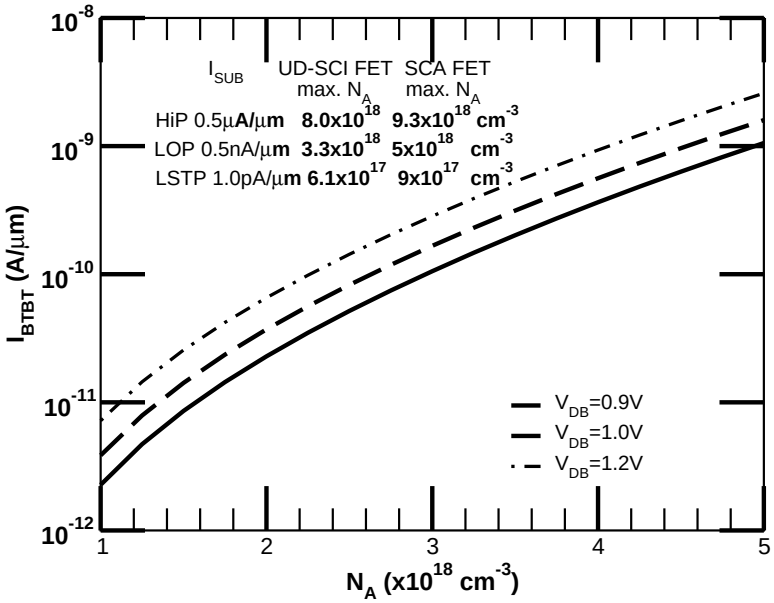


Figure 55: Maximum allowable substrate doping ($N_A|_{max}$) determined by BTBT leakage; $N_A|_{max}$ for RD-SCI FETs is the same as that for accumulation FETs.

Table 14: Minimum channel length for bulk accumulation MOSFETs for various gate technologies, and applications (from 2-D simulations). The maximum subthreshold swing is limited to 110 mV/dec. The channel depth (y_i) is assumed to be 10 nm. "Tun" refers to tunable workfunction metal gate.

application	gate WF (eV)	I_{OFF} ($nA/\mu m$)	V_T (V)	S (mV/dec)	N_A (cm^{-3})	N_D (cm^{-3})	t_{ox} (Å)	L_{min} (nm)
HiP	4.63	300	0.05	110	2.7x10^{18}	5x10^{18}	10	34
LOP	4.63	2.0	0.30	112	5x10^{18}	1x10^{17}	10	25
LSTP	4.63	0.002	0.48	84	9x10^{17}	1x10^{17}	23	62
HiP	Tun	300	0.05	110	9x10^{18}	1x10^{17}	10	22
LOP	Tun	1.0	0.30	110	5x10^{18}	1x10^{17}	10	25
LSTP	Tun	0.001	0.50	110	9x10^{17}	1x10^{17}	23	43
HiP	5.19	600	0.02	110	1x10^{17}	5x10^{18}	12	80
LOP	5.19	2.0	0.31	110	7x10^{17}	5x10^{18}	12	53
LSTP	5.19	0.0008	0.71	111	9x10^{17}	1x10^{17}	12	40

inversion MOSFETs for HiP and LOP applications. For poly-gate applications, accumulation MOSFETs show better performance for LSTP applications. These L_{min} advantages mirror the findings on accumulation MOSFET performance advantage in chapters VI-VIII.

9.3 Summary

The minimum effective channel length, L_{min}, has been derived for different gate technologies and applications. BTBT leakage at the drain-substrate region limits the maximum allowable N_A. Gate tunneling leakage limits the minimum allowable oxide thickness. Based on physical models, the maximum allowable N_A, and the minimum allowable t_{ox} have been quantified for various applications, and device types. Using SCE models, L_{min} is found to be π/λ_{eq_1}. L_{min} found using the models agrees well with that found using 2-D simulations. Accumulation MOSFETs show a lower L_{min} than inversion MOSFETs for three applications: mid-bandgap metal gate HiP, mid-bandgap metal gate LOP, and poly-gate LSTP applications.

Table 15: Minimum channel length for bulk RD-SCI MOSFETs for various gate technologies, and applications (from 2-D simulations). The maximum subthreshold swing is limited to 110 mV/dec. The channel depth (y_i) is assumed to be 10 nm. "Tun" refers to tunable workfunction metal gate.

application	gate WF (eV)	I_{OFF} ($nA/\mu m$)	V_T (V)	S (mV/dec)	N_A (cm^{-3})	N_D (cm^{-3})	t_{ox} (Å)	L_{min} (nm)
HiP	4.63	500	0.05	202	1x10^{16}	1x10^{16}	10	100
LOP	4.63	2.0	0.30	112	4.5x10^{18}	1x10^{17}	10	28
LSTP	4.63	0.002	0.51	84	9x10^{17}	1x10^{17}	23	60
HiP	Tun	300	0.05	110	9x10^{18}	1x10^{17}	10	21
LOP	Tun	1.0	0.30	110	5x10^{18}	1x10^{17}	10	24
LSTP	Tun	0.001	0.50	110	9x10^{17}	1x10^{17}	23	42
HiP	4.19	300	0.04	88	9x10^{18}	1x10^{17}	14	28
LOP	4.19	2.0	0.21	83	5x10^{18}	1x10^{17}	18	50
LSTP	4.19	0.001	0.50	110	9x10^{17}	1x10^{17}	67	82

Table 16: Minimum channel length for bulk UD-SCI MOSFETs for various gate technologies, and applications (from 2-D simulations). The maximum subthreshold swing is limited to 110 mV/dec. "Tun" refers to tunable workfunction metal gate.

application	gate WF (eV)	I_{OFF} ($nA/\mu m$)	V_T (V)	S (mV/dec)	N_A (cm^{-3})	t_{ox} (Å)	L_{min} (nm)
HiP	4.63	500	0.05	202	1x10^{16}	10	100
LOP	4.63	2.0	0.32	110	9x10^{17}	10	36
LSTP	4.63	0.002	0.48	85	6x10^{17}	23	60
HiP	Tun	500	0.05	110	8x10^{18}	12	16
LOP	Tun	1.0	0.30	110	3x10^{18}	12	23
LSTP	Tun	0.001	0.50	110	6x10^{17}	23	42
HiP	4.19	500	0.05	104	6x10^{18}	12	20
LOP	4.19	1.0	0.30	110	3x10^{18}	25	33
LSTP	4.19	0.001	0.52	108	6x10^{17}	70	86

Table 17: Minimum channel length for bulk accumulation MOSFETs for various gate technologies, and applications (from SCE models). The maximum V_T rolloff is limited to 120 mV. "Tun" refers to tunable workfunction metal gate.

application	gate WF (eV)	V_T (V)	N_A (cm^{-3})	N_D (cm^{-3})	yi (nm)	t_{ox} (Å)	L_{min} (nm)
HiP	4.63	0.05	2.7x10^{18}	5x10^{18}	10	10	36
LOP	4.63	0.30	5x10^{18}	1x10^{17}	10	10	26
LSTP	4.63	0.50	9x10^{17}	1x10^{17}	10	39	74
HiP	Tun	0.05	9x10^{18}	1x10^{17}	10	10	23
LOP	Tun	0.30	5x10^{18}	1x10^{17}	10	10	26
LSTP	Tun	0.50	9x10^{17}	1x10^{17}	10	23	62
HiP	5.19	0.07	1x10^{17}	5x10^{18}	10	12	130
LOP	5.19	0.31	4x10^{17}	5x10^{18}	10	12	74
LSTP	5.19	0.73	9x10^{17}	1x10^{17}	10	12	48

Table 18: Minimum channel length for bulk RD-SCI MOSFETs for various gate technologies, and applications (from SCE models). The maximum V_T rolloff is limited to 120 mV. "Tun" refers to tunable workfunction metal gate.

application	gate WF (eV)	V_T (V)	N_A (cm^{-3})	N_D (cm^{-3})	yi (nm)	t_{ox} (Å)	L_{min} (nm)
HiP	4.63	0.05	5x10^{16}	5x10^{16}	10	10	157
LOP	4.63	0.30	2.5x10^{18}	1x10^{17}	10	10	32
LSTP	4.63	0.50	9x10^{17}	1x10^{17}	10	33	67
HiP	Tun	0.05	9x10^{18}	1x10^{17}	10	10	23
LOP	Tun	0.30	5x10^{18}	1x10^{17}	10	10	26
LSTP	Tun	0.50	9x10^{17}	1x10^{17}	10	23	62
HiP	4.19	0.05	9x10^{18}	1x10^{17}	10	14	26
LOP	4.19	0.30	5x10^{18}	1x10^{17}	10	28	38
LSTP	4.19	0.50	9x10^{17}	1x10^{17}	10	70	105

Table 19: Minimum channel length for bulk UD-SCI MOSFETs for various gate technologies, and applications (from SCE models). The maximum V_T rolloff is limited to 120 mV. "Tun" refers to tunable workfunction metal gate.

application	gate WF (eV)	V_T (V)	N_A (cm^{-3})	t_{ox} (Å)	L_{min} (nm)
HiP	4.63	0.05	5x10^{16}	10	157
LOP	4.63	0.30	1.5x10^{18}	10	36
LSTP	4.63	0.50	6x10^{17}	33	76
HiP	Tun	0.05	8x10^{18}	10	20
LOP	Tun	0.30	3x10^{18}	10	27
LSTP	Tun	0.50	6x10^{17}	23	70
HiP	4.19	0.05	3.5x10^{18}	12	28
LOP	4.19	0.30	3x10^{18}	21	38
LSTP	4.19	0.50	6x10^{17}	70	119

Table 20: Minimum channel length for various FET types, gate technologies, and applications - comparison between 2-D simulations and models. In the 2-D simulations, the maximum subthreshold swing is limited to 110 mV/dec. In the models, the maximum V_T rolloff is limited to 120 mV. The channel depth (y_i) is assumed to be 10 nm for accumulation and RD-SCI MOSFETs. All channel length values are in nanometres.

Gate type		Ioff (nA/um)	Vt (V)	Accum. FET Sim.	Accum. FET Models	RD-SCI FET Sim.	RD-SCI FET Models	UD-SCI FET Sim.	UD-SCI FET Models
mid	HiP	500	0.05	34	36	>100	157	>100	157
band	LOP	1	0.3	25	26	28	32	36	36
gap	LSTP	0.001	0.5	62	74	60	67	60	76
tunable	HiP	500	0.05	22	23	21	23	16	20
metal	LOP	1	0.3	25	26	24	26	23	27
gate	LSTP	0.001	0.5	43	62	42	62	42	70
poly	HiP	500	0.05	80	130	28	26	20	28
gate	LOP	1	0.3	53	74	50	38	33	38
	LSTP	0.001	0.5	40	48	82	105	86	119

CHAPTER X

CONCLUSIONS

The primary objective of the current research has been to comprehensively compare accumulation and inversion bulk FETs, and find application areas where each is superior. To provide clear physical insight, physically based models for short channel effect (SCE) and related phenomena were derived. Both SCE models and 2-D numerical simulations are used to perform the comparison, and to project the minimum channel length (L_{min}).

- Threshold voltage models were derived that can accurately predict the impact of source/drain junction depth. A junction depth dependent characteristic length was derived that can be used to rapidly assess the impact of junction depth scaling on V_T.

- It was shown that existing theory does not correctly model long-channel V_T of accumulation MOSFETs. A simple modification was proposed that is accurate over a wide range of channel depths and doping.

- Subthreshold swing (S) rollup is modeled using the volume inversion approach. Using the models for V_T and S, a subthreshold conduction model is derived based on the Swanson-Meindl approach.

- A circuit simulation methodology is developed that can simulate both inversion and accumulation FET inverter chains, and is found to be accurate over a wide range of supply voltages. The simulation methodology can be used for rapid technology optimization, and performance prediction.

- Design guidelines for accumulation MOSFET design are proposed. Surface channel operation is stressed; buried channel operation not only results in increased SCE, but also high dopant fluctuation effects, and process sensitivity. Based on 2-D theory, the SCA/BCA boundary is derived. It is shown that an SCA FET at long channel lengths

can operate as a BCA FET at shorter channel lengths.

- A y_i (tub-depth) vs. N_D (tub-doping) design plane for accumulation MOSFETs is introduced. At lower N_D, the allowable design space is limited by V_T rolloff requirements, whereas at higher N_D, it is limited by process sensitivity requirements.

- The relative advantages of accumulation MOSFETs over inversion MOSFETs is found to be gate-technology dependent, and also application dependent. Three different gate technologies are considered while comparing accumulation and inversion MOSFETs: tunable metal gate, mid-bandgap metal gate, and dual poly gate. Three different application areas (as recommended by ITRS) are considered: hi performance (HiP), low operating power (LOP), and low standby power (LSTP). The off-current and V_T requirements for the various applications are shown in Table 1.

- The accumulation FET has an increased depletion depth compared to a similarly doped inversion FET. For long channel operation, this results in the Fermi-FET being the device with the minimum subthreshold swing. For short channel operation, this results in increased SCE in accumulation FETs. Thus, for tunable workfunction gates, inversion FETs have better SCE control than accumulation FETs.

- In case of mid-bandgap metal gate, accumulation FETs perform better than inversion devices for HiP and LOP applications. Accumulation devices have a lower SCE, and lower QM V_T shift compared to inversion devices. Full-band Monte Carlo simulations show that at $L_{eff} = 30\ nm$, the accumulation MOSFET can achieve a 30-40% better on-current than an inversion MOSFET with a comparable SCE.

- In case of dual poly applications, accumulation MOSFETs perform better than inversion MOSFETs for LSTP applications; because of a higher tunneling barrier, accumulation n-MOSFETs have orders of magnitude lower gate tunneling current compared to inversion n-MOSFETs. Thus, accumulation n-FETs can use a thinner oxide, and thereby achieve better performance than inversion n-MOSFETs. Also, BTBT leakage at the drain-substrate junction is much lower in case of accumulation MOSFETs.

This results in a higher allowable substrate doping, and thus better SCE control.

- L_{min} for accumulation and inversion MOSFETs was projected using both SCE models and 2-D simulations. For HiP applications with mid-bandgap metal gate, accumulation FETs provide more than four generations of further scaling compared to inversion FETs. The corresponding numbers for LOP with mid-bandgap gate, and LSTP with poly gate are one-half generation, and two generations respectively.

10.1 Future Work Recommendations

- This work did not include compact modeling of SCE in halo-doped MOSFETs. Halo-doped MOSFETs are widely used for better control of SCE, and thus need to be modeled accurately. Existing V_T models for halo-FETs are either parameter fit [61], or quasi-physical [58]; none of the existing models can predict the impact of source/drain junction depth on V_T.

- SOI structures were not considered in this work. The results presented in this work are equally applicable to partially depleted (PD) SOI structures. Fully depleted (FD) SOI accumulation FETs need to be compared with FD-SOI inversion MOSFETs for possible advantages.

- Accumulation MOSFETs have a higher g_m (transconductance), and thus perform better than inversion MOSFETs for many analog applications. This needs to be more thoroughly explored by actual circuit simulations.

- Accumulation MOSFETs have a lower $1/f$ noise compared to inversion MOSFETs. This aspect was not considered in this work since the emphasis was mostly on digital applications.

APPENDIX A

THRESHOLD VOLTAGE MODEL FOR DEVICES WITH A DEEP TUB OR HIGH SCE

In Chapter 2, V_T rolloff is derived by assuming a single term solution for $\phi(x, y)$. The single term solution approximation for deriving a compact model for V_T rolloff works well only when $\Delta V_T < 200\ mV$. To accurately model high SCE devices, two additional factors need to be considered: a multiple term solution would have to be used for evaluating the 2-D potential (since the higher order terms become comparable to the first term in the expansion for $\phi(x, y)$), and subthreshold current is no more determined by just the carrier concentration at the minimum potential point.

While deriving the long channel V_T of accumulation MOSFETs in section 2.2.2, it was assumed that V_T is determined by the minimum potential point. In reality, as the tub-depth increases in an accumulation MOSFET, the device tends to become buried channel type, and the electrons spread somewhat uniformly in a 5-10 nm range around the minimum potential point, Fig. 56. Thus, an integration of electron concentration in the depth direction would be needed to accurately determine V_T of accumulation devices with a deep tub. It can be seen from Fig. 14 that as channel length is reduced with all other parameters remaining constant, the inversion layer thickness increases even with the classical distribution (i.e. without including the quantum centroid shift). Thus, for devices with very high SCE, V_T would have to be determined by integrating the electron concentration in the depth direction. The drain current using the integration approach is

$$I \propto \int_0^d n_i . e^{(\psi(x_m, y) - \phi_{Fn})/kT}\ dy \tag{48}$$

x_m is the x-coordinate of the minimum potential point, and ϕ_{Fn} is the electron quasi-Fermi level. It has been established from 2-D MEDICI simulations that the difference between the electron quasi-Fermi level and the bulk Fermi level is about 15 mV for a moderately doped

substrate ($N_A > 1\text{x}10^{17}$ cm^{-3}), for both high and low drain voltages. So ϕ_{Fn} in (48) is taken to be ϕ_F, the bulk Fermi level; as will be seen later, this approximation introduces little error in the calculation of V_T rolloff and DIBL. Let the right hand side of (48) be denoted by Q_i, the inversion areal charge density (IACD). The threshold voltage is determined as the gate voltage at which Q_i reaches a certain density, and thus the model is labeled constant inversion areal charge density (CIACD) model. Let Q_{T_i} be the threshold IACD. From studying a handful of long-channel devices, Q_{T_i} is determined to be $1.8\text{x}10^{11}$ cm^{-2}.

Fig. 57 shows threshold voltage vs. tub-doping for long channel MOSFETs. Only a single term solution is used to calculate $\psi(x_m, y)$ in all the graphs shown in this appendix. The CIACD model is seen to give a good match with 2-D simulations for a wide range of tub-depth and tub-doping. From Fig. 11 it can be seen that the compact model starts diverging from simulations for high y_i, whereas the CIACD model is accurate even for high y_i. Fig. 58 shows the CIACD model vs. 2-D simulations for a short channel device. Also shown is the point where ΔV_T becomes greater than 200 mV, for $N_A = 4\text{x}10^{18}$ cm^{-3}. The CIACD model is able to predict V_T rolloff accurately for devices with ΔV_T as high as 400 mV. Fig. 59 shows threshold voltage vs. tub-depth for a short channel device, this time on a $V_T - y_i$ plane. Once again, the CIACD model works well for high SCE devices. The compact model is clearly limited to devices with $\Delta V_T < 200\ mV$. The disadvantage of using the CIACD model is that it takes more time (5X-10X) than the compact model to evaluate ΔV_T. But the time taken by the CIACD model is still far less than the time taken by a 2-D simulator to evaluate ΔV_T.

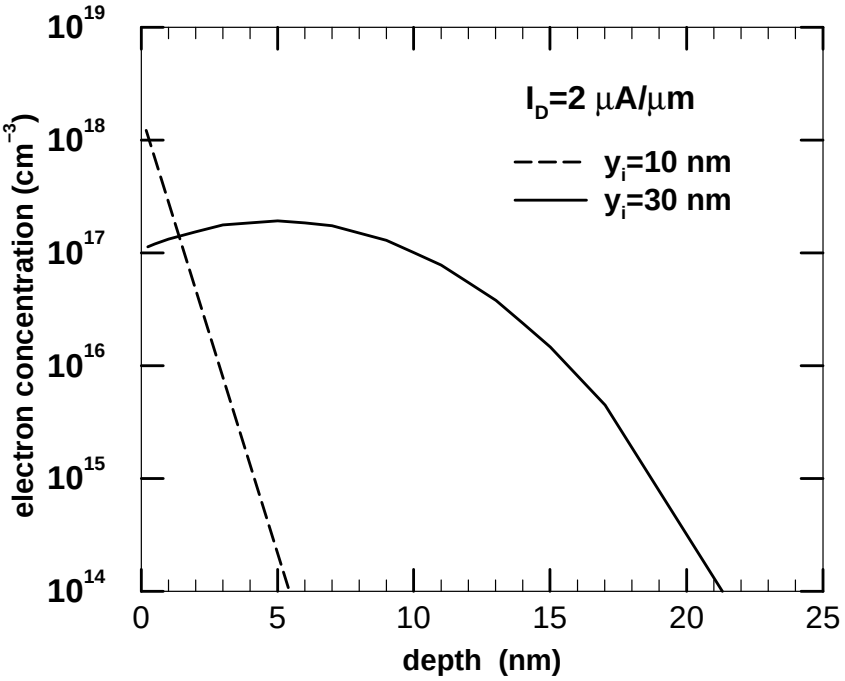


Figure 56: Carrier concentration in the middle of the channel in long channel accumulation MOSFETs (from 2-D drift-diffusion simulations); $L = 200\ nm$, $V_{DS} = 0.05\ V$, N_A =1x10^{18} cm^{-3}, N_D =1x10^{18} cm^{-3}, and $t_{ox} = 1\ nm$. The gate workfunction of the devices are adjusted so that both devices have the same drain current of 2 $\mu A/\mu m$ for $V_{GS} = 0.5\ V$.

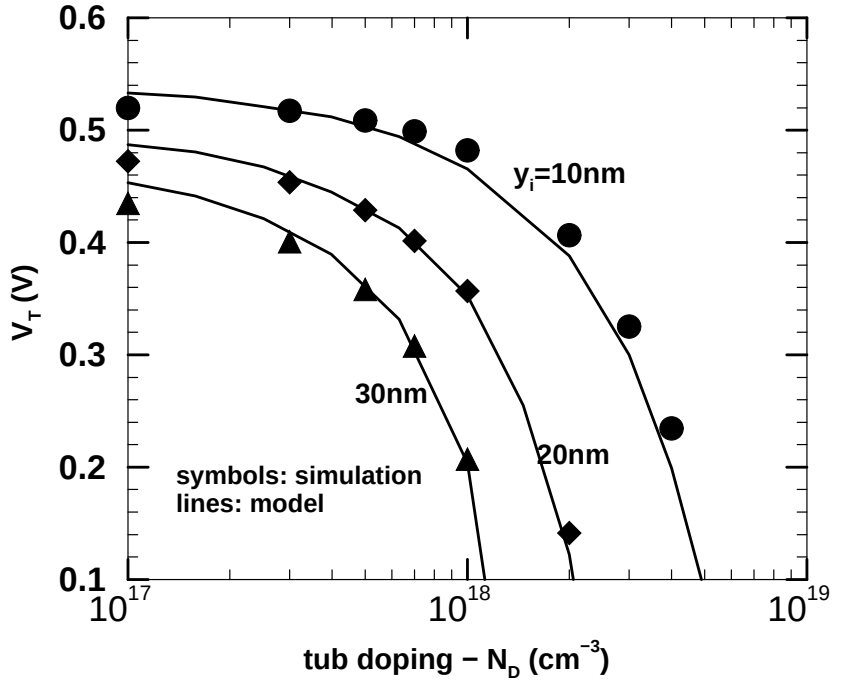


Figure 57: Long channel V_T of accumulation MOSFETs: CIACD model vs. simulations; N_A =1x10^{18} cm^{-3}, and $t_{ox} = 1.5\ nm$.

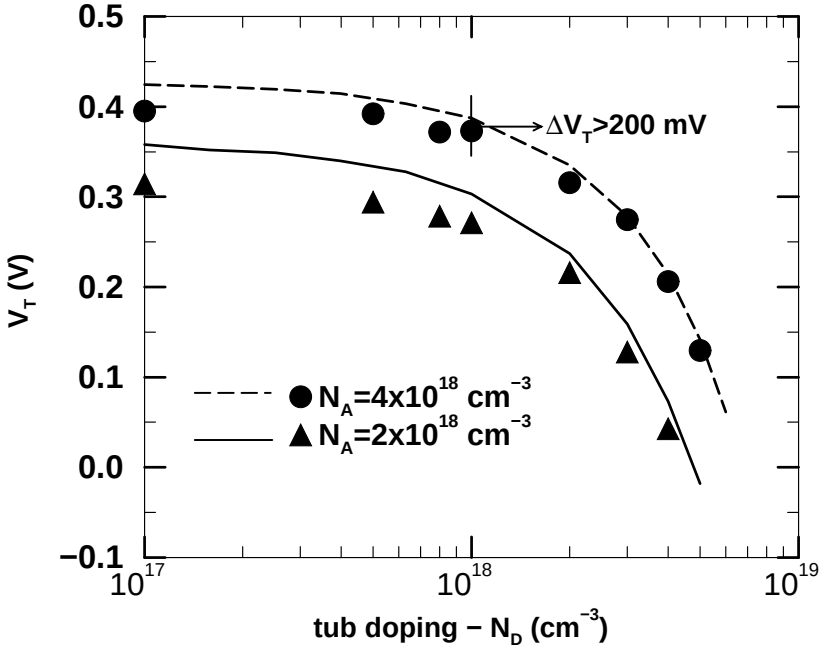


Figure 58: Short channel V_T of accumulation MOSFETs: CIACD model vs. simulations; $L = 30\ nm$, t_{ox} =10 Å, $y_j = 12\ nm$, $V_{dd} = 1.0\ V$, y_i =10 nm, and gate workfunction is 4.63 eV.

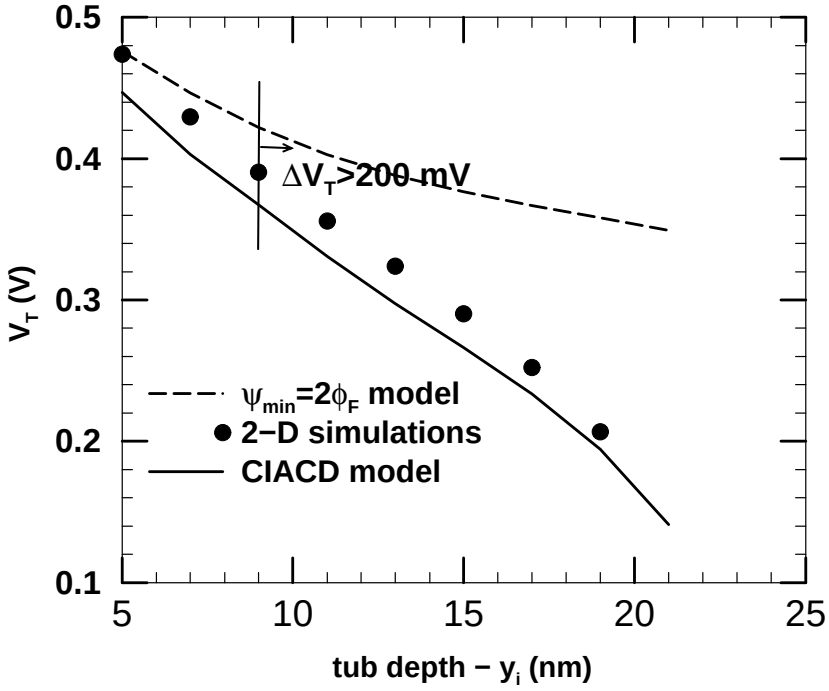


Figure 59: Short channel V_T of accumulation MOSFETs: CIACD model vs. simulations; $L = 30\ nm$, t_{ox} =10 Å, $y_j = 12\ nm$, $V_{dd} = 1.0\ V$, N_A =4x10^{18} cm^{-3}, N_D =1x10^{18} cm^{-3}, and gate workfunction is 4.63 eV.

APPENDIX B

MID-BAND METAL GATE MOSFETS

Additional comparisons between mid-band metal gate accumulation and inversion MOSFETs are provided in this appendix. Fig. 60 shows off-current vs. threshold voltage for $L = 30\ nm$ n-MOSFETs. Both RD-SCI and UD-SCI MOSFETs are considered while evaluating inversion FETs. For $V_T < 0.3\ V$, accumulation MOSFETs show a better performance than inversion MOSFETs in terms of a better off-current for a given threshold voltage. DIBL limits the minimum achievable V_T of inversion n-MOSFETs to 0.3 V and that of accumulation n-MOSFETs to 0.2 V. Fig. 61 shows off-current vs. threshold voltage for $L = 30\ nm$ p-MOSFETs. Accumulation MOSFETs are seen to perform better than inversion MOSFETs for $V_T < 0.45\ V$. DIBL limits the minimum achievable V_T of inversion p-MOSFETs to 0.4 V, and that of accumulation p-MOSFETs to 0.3 V.

Fig. 62 shows off-current vs. on-current for accumulation and inversion n-MOSFETs. Poly-gate inversion MOSFETs are assumed to have a poly-doping of $5\text{x}10^{19}\ \text{cm}^{-3}$. Mid-band metal gate accumulation MOSFETs are seen to have a better on-current (for a given off-current) than either mid-band metal-gate or poly-gate inversion MOSFETs.

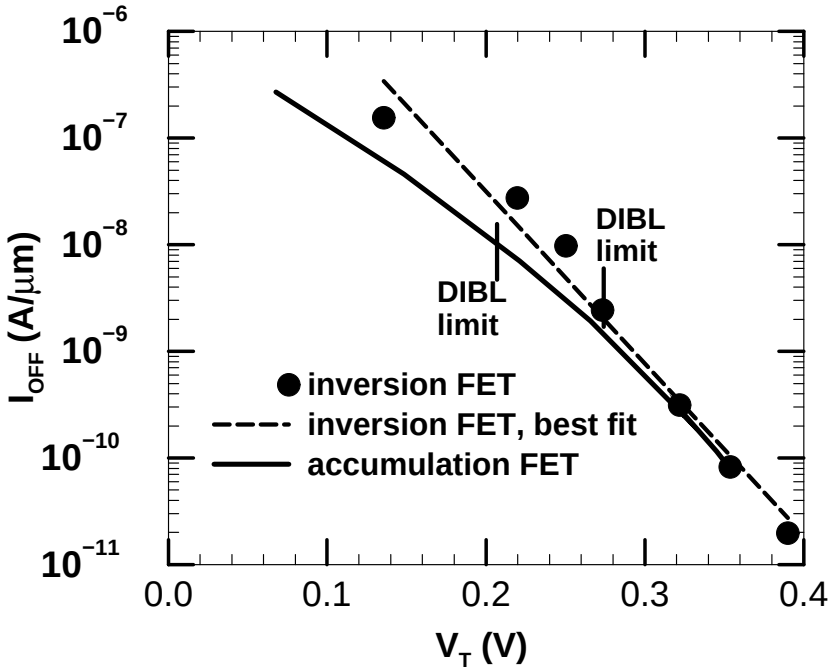


Figure 60: Off-current vs. threshold voltage for n-MOSFETs (from 2-D classical DD simulations). $L = 30\ nm$, t_{ox} =10 Å, $y_j = 12\ nm$, $V_{dd} = 1.0\ V$, and gate workfunction is 4.63 eV. For accumulation MOSFETs, N_A =4x10^{18} cm^{-3}, and $y_i = 10\ nm$. The DIBL limit is taken to be 100 mV/V.

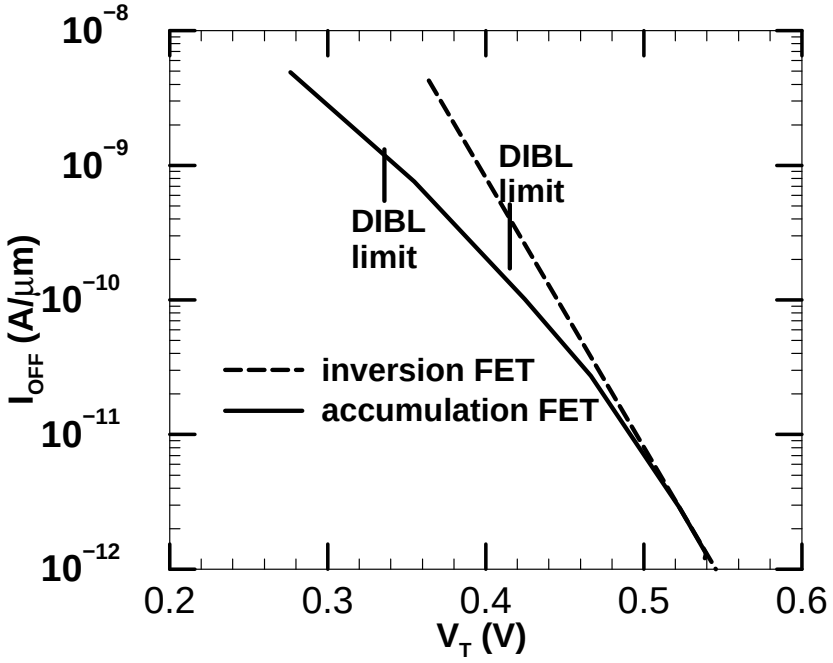


Figure 61: Off-current vs. threshold voltage for p-MOSFETs (from 2-D classical DD simulations). $L = 30\ nm$, t_{ox} =10 Å, $y_j = 12\ nm$, $V_{dd} = 1.0\ V$, and gate workfunction is 4.63 eV. For accumulation MOSFETs, N_A =4x10^{18} cm^{-3}, and $y_i = 10\ nm$. The DIBL limit is taken to be 150 mV/V.

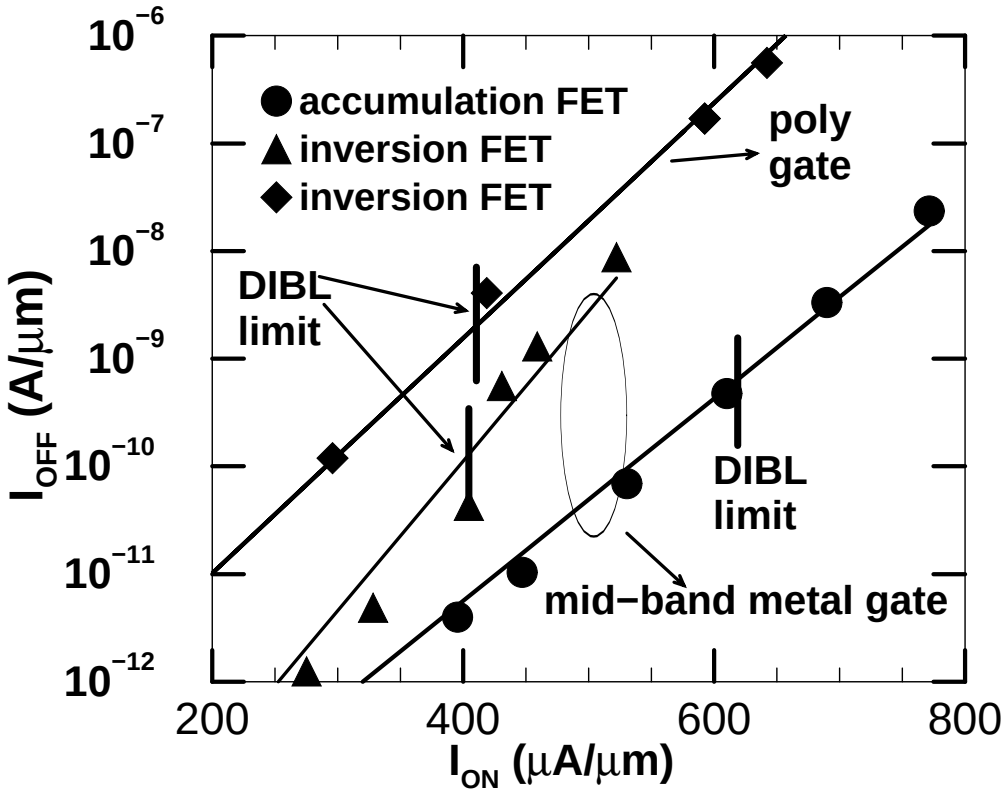


Figure 62: Off-current vs. on-current for n-MOSFETs (from 2-D DD simulations including QM V_T shift); $L = 30\ nm$, t_{ox} =10 Å, $y_j = 12\ nm$, and $V_{dd} = 1.0\ V$; the gate workfunction is 4.63 eV for mid-band metal gate MOSFETs, and 4.19 eV for poly gate inversion MOSFETs. For accumulation MOSFETs, N_A =4x10^{18} cm^{-3}, and $y_i = 10\ nm$. Symbols indicate data points and lines indicate best fit. The DIBL limit is taken to be 100 mV/V.

APPENDIX C

QUANTUM CORRECTION FOR V_T IN SIMULATIONS

Device simulators provide the user with a variety of options to include quantum mechanical (QM) V_T shift in simulations. These include 1-D Schrödinger-Poisson solution coupled to 2-D Poisson solution, the density gradient method, and the van Dort method [87]. The 1-D Schrödinger solution is the most accurate among the three. But simulations using it tend to be slow, and often lead to convergence problems. The van Dort method is less physical, but leads to fast convergence. The density gradient method is numerically robust but needs a lot of calibration effort for the mobility and recombination-generation models.

In this work, van Dort correction is used to include QM effects. Fig. 63 shows a comparison of on-current output from van-Dort and 1-D Schrödinger solutions. It can be seen that the two agree over a wide range of on-current. This confirms that van Dort method is an accurate way of including quantum mechanical V_T shift in 2-D device simulations.

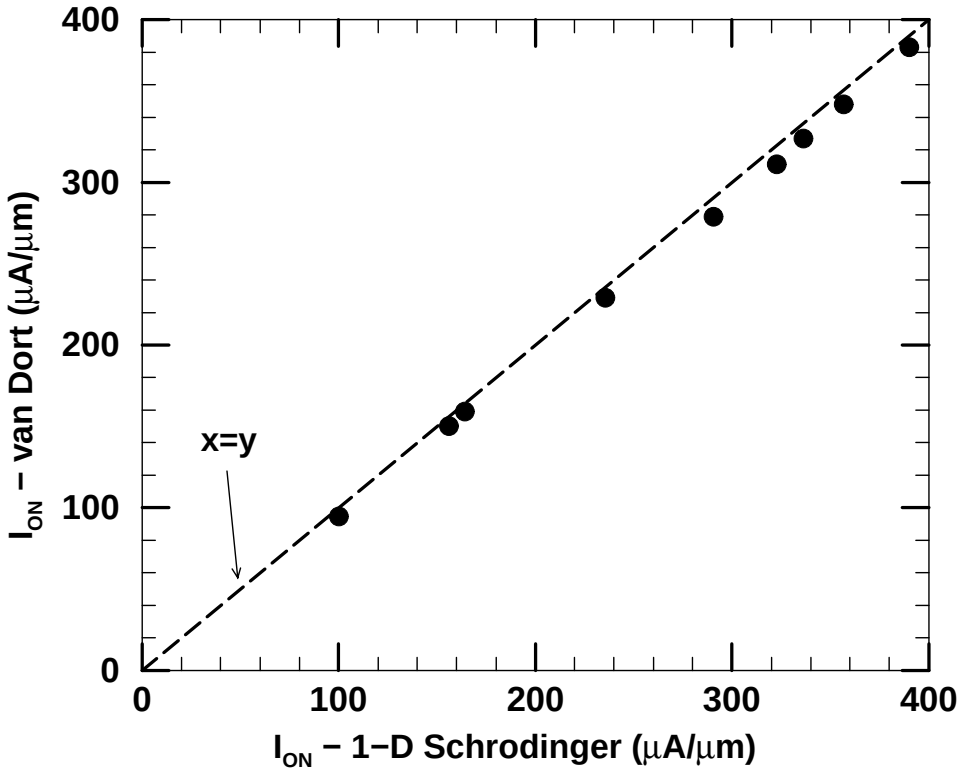


Figure 63: Comparison of on-current from 1-D Schrödinger and van Dort methods. The devices are inversion type, and have various channel and substrate doping. Symbols indicate data points, and the dashed line is the "$x = y$" line. $L = 45\ nm$, $t_{ox} = 15$ Å, $y_j = 12\ nm$, and $V_{dd} = 1.0\ \ V$.

APPENDIX D

HALO PROFILE DESCRIPTION

Halo-doped devices used in this work use analytical formulae to generate the halo profiles. Source and drain halos are symmetrical, and are generated by a product of two Gaussian functions [107]. Suppose (x_p, y_p) is the co-ordinate of the peak halo doping, Fig. 64. Let N_p be the peak halo doping. Then the drain-halo is generated by

$$f(x,y) = N_p.exp\left(\frac{x-x_p}{\sigma_x}\right).exp\left(\frac{y-y_p}{\sigma_y}\right) \tag{49}$$

The origin of the co-ordinate system is assumed to be in the middle of the channel, at the $\mathsf{Si}-\mathsf{SiO_2}$ interface. For the 30 nm channel length MOSFETs considered in chapter VIII, the halo parameters are as follows: $x_p = L_{eff}/2 + 5$, $y_p = 22$, $N_p = 1\text{x}10^{19}$ cm^{-3}, $\sigma_x = 4$, and $\sigma_y = 3.4$. x, y, and σ values are in nanometres.

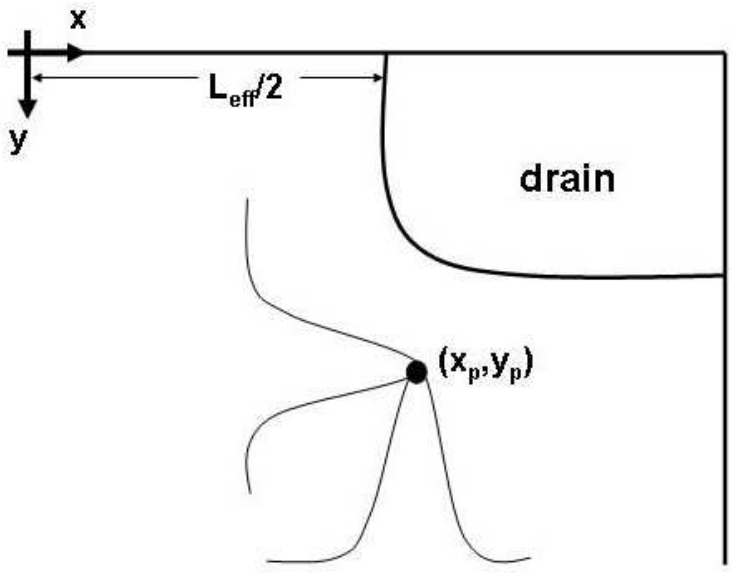


Figure 64: Halo-doping at the drain-end of a MOSFET. Two 1-D Gaussian functions are superposed to obtain the halo doping, with the peak doping being N_p, at (x_p, y_p).

APPENDIX E

GATE TUNNELING MODEL

The gate tunneling model from [92] is used in this work. The model is physically based and uses no fitting parameters. The model can be easily modified to include accumulation devices. The gate tunneling current density is calculated by multiplying the electron density and the tunneling probability at the $\mathsf{Si}-\mathsf{SiO_2}$ interface [92]. Since the potential profiles of accumulation and inversion devices are different, the tunneling probabilities too are different. The tunneling probability is given as

$$D(E_x) = exp\left(-\gamma\sqrt{E_B - E_x}\right) \tag{50}$$

where γ is $4\pi t_{ox}\sqrt{2m_{ox}}/h$. E_B is the barrier height, and E_x is the electron energy. t_{ox} is the physical oxide thickness, and $m_{ox} = 0.32m_0$ is the effective electron mass in the oxide. The barrier height is approximated by a rectangular barrier.

$$E_B = q\left(3.1 - \frac{V_{ox}}{2}\right) \tag{51}$$

V_{ox} is the voltage drop across the oxide layer, and it is positive when the field points into the oxide (looking from the top of the $\mathsf{Si}-\mathsf{SiO_2}$ interface). When the field points out of the oxide, electrons no longer experience an attractive force towards the $\mathsf{Si}-\mathsf{SiO_2}$ interface, and V_{ox} is then negative.

The gate tunneling density is given as

$$J_{tunnel} = \frac{4\pi m^* q}{h^3}(kT)^2\left(1 + \frac{\gamma kT}{q\sqrt{E_B}}\right) exp\left(\frac{q\phi_S - q\phi_F - E_G/2}{kT}\right) exp\left(-\gamma\sqrt{E_B}\right) \tag{52}$$

where,

$m^* = 0.19m_0$ is the electron transverse mass

ϕ_S is the surface potential

ϕ_F is the Fermi level in the bulk

E_G is the Si bandgap energy

k is Boltzmann's constant

T is the temperature in °K

In case of buried channel MOSFETs, the barrier for gate-channel electron tunneling is higher compared to inversion devices, Figs. 65 and 66; this results in orders or magnitude lower gate tunneling current in BCA FETs.

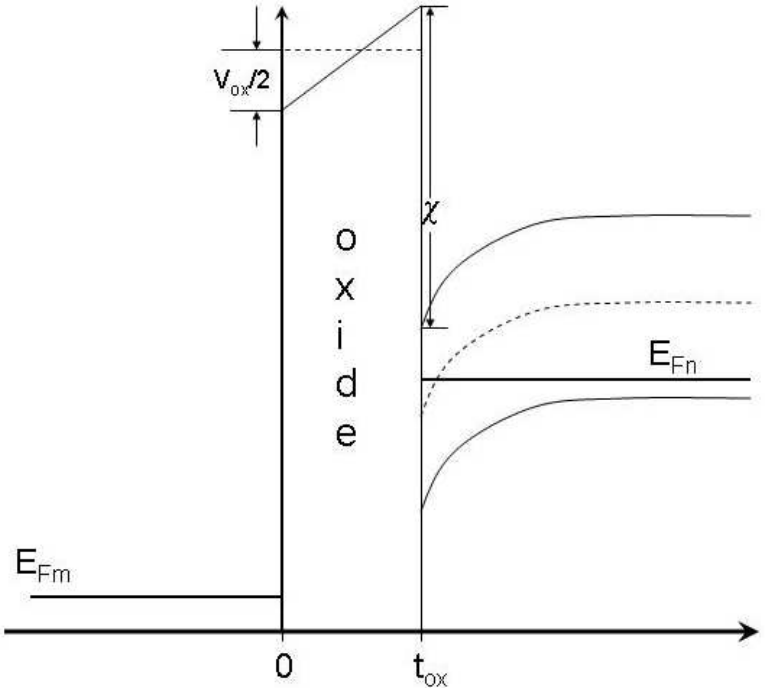


Figure 65: Band-diagram for an inversion device in moderate-inversion regime.

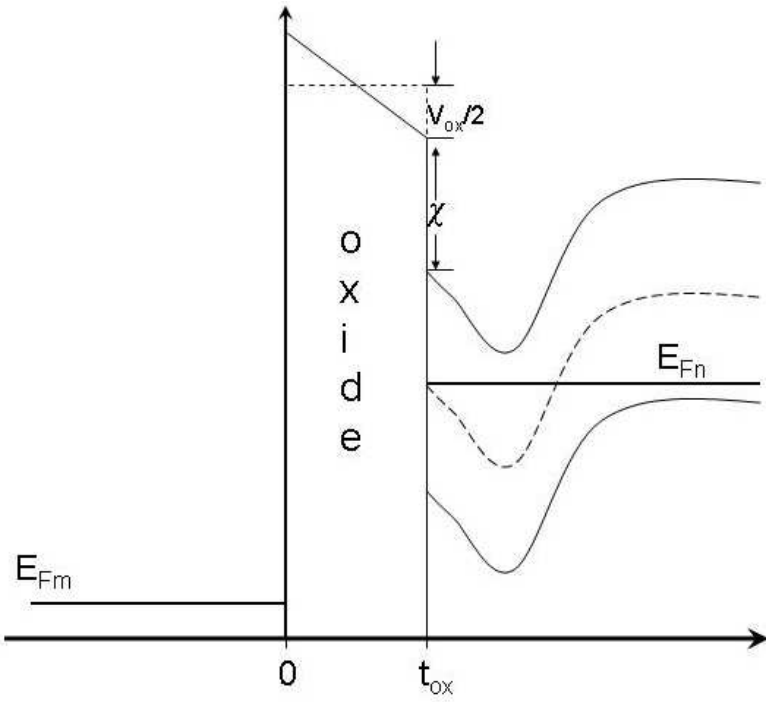


Figure 66: Band-diagram for an accumulation device in moderate-conduction regime.

APPENDIX F

HYDRO-DYNAMIC MODELS

The simulator DESSIS is used for all hydro-dynamic simulations reported in this work. In DESSIS [67], the set of equations solved for electrons in the hydro-dynamic approach is

$$\vec{J_n} = \mu_n \left(n\nabla E_C + k_B T_n \nabla_n + f_n^{td} k_B n \nabla T_n - 1.5 n k_B T_n \nabla ln m_E \right) \tag{53}$$

$$\frac{\partial W_n}{\partial t} + \nabla . \vec{S_n} = \vec{J_n} . \nabla E_C + \left. \frac{dW_n}{dt} \right|_{coll} \tag{54}$$

$$\vec{S_n} = -\frac{5 r_n}{2} \left(\frac{k_B T_n}{q} \vec{J_n} + f_n^{hf} \hat{\kappa_n} \nabla T_n \right) \tag{55}$$

$$S_L = -\kappa_L \nabla T_L \tag{56}$$

where $\hat{\kappa_n} = \frac{k_B^2}{q} n \mu_n T_n$. T_n is the electron temperature, E_C is the conduction band energy, n is the electron concentration, and m_E is the effective electron mass. r_n, f_n^{td}, and f_n^{hf} are the energy flux, thermal diffusion, and heat flux parameters respectively. They are fixed at the default values of 0.6, 0 and 1 respectively, which correspond to Stratton's [108] energy balance formulation. A similar set of equations can be formulated for holes.

To fit HD simulation output to Monte Carlo (MC) data, the parameters mentioned above can be varied; or the electron (or hole) relaxation time can be varied; or the mobility model parameters related to surface roughness mobility can be varied. Varying the HD model parameters does not give a satisfactory match with MC data. Varying the electron relaxation time is not physical [94]. So, the best option is to vary surface roughness mobility parameters to achieve a fit with MC data.

Surface roughness mobility is modeled by using the Darwish mobility model [101]. The surface mobility contribution due to acoustic phonon scattering has the form [67]

$$\mu_{ac} = \frac{B}{F_\perp} + \frac{C \left(N_i / N_O \right)^\lambda}{F_\perp^{1/3} \left(T / T_0 \right)^k} \tag{57}$$

The contribution due to surface roughness scattering is given by

$$\mu_{sr} = \left(\frac{\left(F_\perp / F_{ref} \right)^{A^*}}{\delta} + \frac{F_\perp^3}{\eta} \right)^{-1} \tag{58}$$

The contributions from μ_{ac} and μ_{sr} are combined with the contribution from bulk mobility (μ_b) according to Mathiessen's rule:

$$\frac{1}{\mu} = \frac{1}{\mu_b} + \frac{1}{\mu_{ac}} + \frac{1}{\mu_{sr}} \tag{59}$$

The exponent A^* is given by

$$A^* = A + \frac{\alpha_\perp (n+p)}{((N_i + N_1)/N_{ref})^\nu} \tag{60}$$

In the above equations, $F_\perp$ is the vertical field. The normal electric field can be damped by adjusting the the exponential decay factor:

$$F_\perp \rightarrow F_\perp exp\left(-\frac{l}{l_{crit}}\right) \tag{61}$$

The various parameters are calibrated and their values are shown in Table 21 for a 1 nm gate oxide thickness.

Table 21: Darwish model parameters for electrons in Silicon.

parameter	accumulation FET	inversion FET	units
B	1.97x10^7	$(2 + (N_A/10^{17} - 4)$x0.1)x10^7	cm/s
C	3.01x10^6	2.41x10^6	$cm^{5/3}/(V^{2/3}s)$
N_0	1	1	cm^{-3}
λ	-0.168	-0.168	-
k	1	1	-
δ	3.5x10^{18}	3.510^{18}	$cm^2/(Vs)$
A	2.58	2.58	-
$\alpha_\perp$	0	0	cm^{-3}
N_1	1	1	cm^{-3}
ν	0.0767	0.0767	-
η	5.82x10^{30}	5.82x10^{30}	$V^2/(cm.s)$
l_{crit}	10	10	nm

REFERENCES

[1] Semiconductor Industry Association, "The International Technology Roadmap for Semiconductors," 2002.

[2] G. G. Shahidi, D. A. Antoniadis, and H. I. Smith, "Electron velocity overshoot at room and liquid nitrogen temperatures in silicon inversion layers," *IEEE Electron Dev. Lett.*, vol. 9, no. 2, pp. 94–96, Feb. 1988.

[3] M. Aoki, T. Ishii, T. Yoshimura, Y. Kiyota, S. Iijima, T. Yamanaka, T. Kure, K. Ohyu, T. Nishida, S. Okazaki, K. Seki, and K. Shimohigashi, "Design and performance of 0.1-um CMOS devices using low-impurity-channel transistors (LICT's)," *IEEE Electron Dev. Lett.*, vol. 13, no. 1, pp. 50 –52, Jan 1992.

[4] K. Yamaguchi, Y. Shiraki, Y. Katayama, and Y. Muriyama, "A new short-channel MOSFET with an atomic-layer-doping impurity profile," *Japanese J. of Appl. Phys.*, vol. 22, pp. 267–270, 1983.

[5] H. Tian, R. B. Hulfachor, J. J. Ellis-Monaghan, K. W. Kim, M. A. Littlejohn, J. R. Hauser, and N. A. Masnari, "An evaluation of super-steep-retrograde channel doping for deep-submicron MOSFET applications," *IEEE Transactions on*, vol. 41, no. 10, pp. 1880 –1882, Oct 1994.

[6] T. Ohguro, H. Naruse, H. Sugaya, E. Morifuji, S. Nakamura, T. Yoshitomi, T. Morimoto, H. Kimijima, S. MoMose, H. Katsumata, and Y. Iwai, "An 0.18-um CMOS for mixed digital and analog applications with zero-volt-Vth epitaxial-channel MOSFETs," *IEEE Trans. on Electron Dev.*, vol. 46, no. 7, pp. 1378 –1383, Jul 1999.

[7] H. K. J. Ihantola and J. L. Moll, "Design theory of a surface field-effect transistor," *Solid St. Electronics*, vol. 7, no. 4, pp. 423–430, Apr. 1964.

[8] S. R. Hofstein, "An analysis of deep depletion thin film MOS transistors," *IEEE Trans. on Electron Dev.*, vol. ED-13, pp. 846–855, Dec. 1966.

[9] F. P. Heiman, "Thin-film silicon-on-sapphire deep depletion MOS transistors," *IEEE Trans. on Electron Dev.*, vol. ED-13, pp. 855–862, Dec. 1966.

[10] A. H. Montree, V. M. H. Meijssen, and P. H. Woerlee, "Comparison of buried and surface channel PMOS devices for low voltage 0.5 um CMOS," in *Symp. on VLSI Tech.*, 1993, pp. 11 –14.

[11] T. Yoshitomi, M. Saito, H. Oguma, Y. Akasaka, M. Ono, H. N. nd Y. Ushiku, H. Iwai, and H. Hara, "Ultra-shallow buried-channel p-MOSFET with extremely high transconductance," in *Symp. on VLSI Tech.*, 1993, pp. 99–100.

[12] H. Abiko, A. Ono, R. Ueno, S. Masuoka, S. Shishiguchi, K. Nakajima, and I. Sakai, "A channel engineering combined with channel epitaxy optimization and TED supression for 0.15μm n-n gate CMOS technology," in *Symp. on VLSI Tech.*, 1995, pp. 23–24.

[13] H. Inokawa, Y. Okazaki, K. Nishimura, S. Date, T. Ishihara, T. Mizusawa, M. Miyake, T. Kobayashi, and T. Tsuchiya, "Highly robust 0.25-μm single-poly-gate CMOS with inter-well deep trenches," in *Symp. on VLSI Tech.*, 1996, pp. 218–219.

[14] H. Matsuhashi, T. Ochiai, M. Kasai, T. Nakamura, and S. Nishikawa, "High-performance double-layer epitaxial-channel pMOSFET compatible with single gate CMOSFET," in *Symp. on VLSI Tech.*, 1996, pp. 36–37.

[15] M. Nandakumar, A. Chatterjee, G. Stacey, and I.-C. Chen, "A 0.25μm gate length CMOS technology for 1V low power applications - device design and power/performance considerations," in *Symp. on VLSI Tech.*, 1996, pp. 68–69.

[16] T. Yoshitomi, T. Ohguro, M. Saito, E. Morifuji, H. Momose, and H. Iwai, "High performance 0.5 micron single gate Co salicide CMOS," in *Digest of Papers Symp. on VLSI Technology*, June 1996, pp. 34–35.

[17] S. Inaba, R. Katsumata, H. Akatsu, R. Rengarajan, P. Ronsheim, C. S. Murthy, K. Sunouchi, and G. B. Bronner, "Threshold voltage roll-up/roll-off characteristic control in sub-0.2-um single workfunction gate CMOS for high-performance DRAM applications," *IEEE Trans. on Electron Dev.*, vol. 49, no. 2, pp. 308 –313, Feb 2002.

[18] B. Agrawal, V. K. De, and J. D. Meindl, "Device parameter optimization for reduced short-channel effects in retrograde doping MOSFETs," *IEEE Trans. on Electron Dev.*, vol. 43, no. 2, pp. 365–368, Feb. 1996.

[19] T. Skotnicki and P. Bouillon, "Electrical performances of retrograde versus conventional profile MOSFETs," in *Symp. on VLSI Tech.*, 1996, pp. 152 –153.

[20] I. De and C. M. Osburn, "Impact of super-steep-retrograde channel doping profiles on the performance of scaled devices," *IEEE Trans. on Electron Dev.*, vol. 46, no. 8, pp. 1711 –1717, Aug 1999.

[21] G. Baccarani, M. R. Wodeman, and R. H. Dennard, "Generalized scaling theory and its application to a 1/4 micrometer MOSFET design," *IEEE Trans. on Electron Dev.*, vol. ED-31, no. 4, pp. 452–462, Apr. 1984.

[22] S. Yamakawa, K. Sugihara, T. Furukawa, Y. Nishioka, T. Nakahata, Y. Abe, S. Maruno, and Y. Tokuda, "Drivability improvement on deep-submicron MOSFETs by elevation of source/drain regions," *IEEE Electron Device Letters*, vol. 20, no. 7, pp. 366 –368, Jul 1999.

[23] J. Tanaka, T. Toyabe, S. Ihara, S. Kimura, H. Noda, and K. Itoh, "Simulation of sub-0.1-um MOSFETs with completely suppressed short-channel effect," *IEEE Electron Device Letters*, vol. 14, no. 8, pp. 396 –399, Aug 1993.

[24] H. Noda, F. Murai, and S. Kimura, "Short channel characteristics of Si MOSFET with extremely shallow source and drain regions formed by inversion layers," *IEEE Trans. on Electron Dev.*, vol. 41, no. 10, pp. 1831 –1836, Oct 1994.

[25] H. Sangyeon, S.-I. Chang, L. Jongho, and S. Hyungcheol, "50 nm MOSFET with electrically induced source/drain (S/D) extensions," *IEEE Trans. on Electron Dev.*, vol. 48, no. 9, pp. 2058 –2064, Sep 2001.

[26] S. Kimura, H. Noda, D. Hisamoto, and E. Takeda, "1 A 0.1 um-gate elevated source and drain MOSFET fabricated by phase-shifted lithography," in *IEEE Int. Electron Dev. Meeting*, 1991, pp. 950 –952.

[27] T. Horiuchi, T. Homma, Y. Murao, and K. Okumura, "An asymmetric sidewall process for high performance LDD MOSFET's," *IEEE Trans. on Electron Dev.*, vol. 41, no. 2, pp. 186 –190, Feb 1994.

[28] S. Odanaka and A. Hiroki, "Potential design and transport property of 0.1-um MOSFET with asymmetric channel profile," *IEEE Trans. on Electron Dev.*, vol. 44, no. 4, pp. 595 –600, Apr 1997.

[29] M. Jurczak, T. Skotnicki, R. Gwoziecki, M. Paoli, B. Tormen, P. Ribot, D. Dutartre, S. Monfray, and J. Galvier, "Dielectric pockets-a new concept of the junctions for deca-nanometric CMOS devices," *IEEE Trans. on Electron Dev.*, vol. 48, no. 8, pp. 1770 –1775, Aug 2001.

[30] P. Ranade, H. Takeuchi, W.-C. Lee, V. Subramanian, and T.-J. King, "Application of Silicon-Germanium in the fabrication of ultra-shallow extension junctions for sub-100 nm PMOSFETs," *IEEE Trans. on Electron Dev.*, vol. 49, no. 8, pp. 1436 –1443, Aug 2002.

[31] H. Wakabayashi, M. Ueki, M. Narihiro, T. Fukai, N. Ikezawa, T. Matsuda, K. Yoshida, K. Takeuchi, Y. Ochiai, T. Mogami, and T. Kunio, "Sub-50-nm physical gate length CMOS technology and beyond using steep halo," *IEEE Trans. on Electron Dev.*, vol. 49, no. 1, pp. 89 –95, Jan 2002.

[32] R. Gwoziecki and T. Skotnicki, "Smart pockets-total suppression of roll-off and roll-up," in *Symp. on VLSI Tech.*, 1999, pp. 91 –92.

[33] T. Tanaka, Y. Momiyama, K. Goto, Y. Sambonsugi, M. Deura, and T. Sugii, "Realization of 0.1 um buried-channel PMOSFETs by device restructuring using tilted well implantation technology," in *Symp. on VLSI Tech.*, 1999, pp. 109 –110.

[34] H. V. Deshpande, C. Baohong, and J. C. S. Woo, "Channel engineering for analog device design in deep submicron CMOS technology for system on chip applications," *IEEE Trans. on Electron Dev.*, vol. 49, no. 9, pp. 1558 –1565, Sep 2002.

[35] J. Xu and M.-C. Cheng, "Design optimization of high-performance low-temperature 0.18 um MOSFETs with low-impurity-density channels at supply voltage below 1 V," *IEEE Trans. on Electron Dev.*, vol. 47, no. 4, pp. 813 –821, Apr 2000.

[36] K. Nishiuchi, H. Oka, T. Nakamura, H. Ishikawa, and M. Shinoda, "A normally-off type buried-channel MOSFET for VLSI circuits," in *IEEE Int. Electron Dev. Meeting*, 1978, pp. 26–29.

[37] A. W. Vinal, "Fermi threshold field effect transistor," United States Patent Office, (Patent Number 4,990,974), Feb. 1991.

[38] T. Enda and N. Shigyo, "Alleviation of subthreshold swing and short-channel effect in buried-channel MOSFETs: the counter-doped surface-channel MOSFET structure," *Electronics and Communications in Japan, Part 2*, vol. 79, no. 11, pp. 43–49, Nov. 1996.

[39] C. Bulucea and D. Kerr, "Threshold voltage control in buried-channel MOSFETs," *Solid St. Electronics*, vol. 41, pp. 1345–1354, 1997.

[40] P. Ranade, Y.-K. Choi, D. Ha, A. Agarwal, M. Ameen, and T.-J. King, "Tunable work function Molybdenum gate technology for FDSOI-CMOS," in *IEEE Int. Electron Dev. Meeting*, 2002, pp. 363 –366.

[41] T. Takeda, H. Kume, T. Toyabe, and S. Asai, "Submicrometer MOSFET structure for minimizing hot carrier generation," *IEEE Trans. on Electron Dev.*, vol. ED-29, pp. 611–618, Apr. 1982.

[42] M. Aoki, Y. Sakai, and T. Masuhara, "Low 1/f noise design of hi-CMOS devices," *IEEE Trans. on Electron Dev.*, vol. ED-29, pp. 296–299, Feb. 1982.

[43] S. M. Sze, *Physics of Semiconductor Devices.* Wiley Interscience, 1981.

[44] A. Asenov and S. Saini, "Suppression of random dopant-induced threshold voltage fluctuations in sub-0.1-um MOSFET's with epitaxial and δ-doped channels," *IEEE Trans. on Electron Dev.*, vol. 46, no. 8, pp. 1718 –1724, Aug.: 1999.

[45] F. M. Klassen, J. J. Bastiaens, W. Hes, and M. Sprokel, "Scaling of compensated MOSFETs towards submicron dimensions," in *IEEE Int. Electron Dev. Meeting*, 1984, pp. 613–616.

[46] F. M. Klaassen and W. Hes, "Compensated MOSFET devices," *Solid St. Electronics*, vol. 4, no. 28, pp. 359–373, Apr. 1985.

[47] B. L. Austin, "Performance analysis and scaling opportunities of bulk CMOS inversion and accumulation devices," Ph.D. dissertation, Georgia Tech, Atlanta, GA, May 2001.

[48] H. Oka, K. Nishiuchi, T. Nakamura, and H. Ishikawa, "Two-dimensional numerical analysis of normally-off type buried-channel MOSFETs," in *IEEE Int. Electron Dev. Meeting*, 1979, pp. 30–33.

[49] G. J. Hu and R. H. Bruce, "Design tradeoffs between surface and buried-channel FET's," *IEEE Trans. on Electron Dev.*, vol. ED-32, no. 3, pp. 584–588, Mar. 1985.

[50] E. Josse and T. Skotnicki, "Polysilicon gate with depletion-or-metallic gate with buried channel: what evil worse ?" in *IEEE Int. Electron Dev. Meeting*, 1999, pp. 661 –664.

[51] Y. Abe, T. Oishi, K. Shiozawa, Y. Tokuda, and S. Satoh, " Simulation study on comparison between metal gate and polysilicon gate for sub-quarter-micron MOSFETs," *IEEE Electron Device Letters*, vol. 20, no. 12, pp. 632 –634, Dec 1999.

[52] C. Fiegna, H. Iwai, T. Wada, M. Saito, E. Sangiorgi, and B. Ricco, "Scaling the MOS transistor below 0.1 um: methodology, device structures, and technology requirements," *IEEE Trans. on Electron Dev.*, vol. 41, no. 6, pp. 941 –951, Jun 1994.

[53] T. A. Fjeldly and M. Shur, "Threshold voltage modeling and the subthreshold regime of operation of short-channel MOSFETs," *IEEE Trans. on Electron Dev.*, vol. 40, no. 1, pp. 137–145, Jan. 1993.

[54] K. N. Ratnakumar and J. D. Meindl, "Short-channel MOST threshold voltage model," *IEEE J. Solid-State Circuits*, vol. SC-17, no. 5, pp. 937–947, Oct. 1982.

[55] Z.-H. Liu, C. Hu, J.-H. Huang, T.-Y. Chan, M.-C. Jeng, P. K. Ko, and Y. C. Cheng, "Threshold voltage model for deep-submicrometer MOSFET's," *IEEE Trans. on Electron Dev.*, vol. 40, no. 1, pp. 86–95, Jan. 1993.

[56] K. Suzuki, "Short channel MOSFET model using a universal channel depletion width parameter," *IEEE Trans. on Electron Dev.*, vol. 47, no. 6, pp. 1202–1208, June 2000.

[57] T. Toyabe and S. Asai, "Analytical models of threshold voltage and breakdown voltage of short-channel MOSFETs derived from two-dimentional analysis," *IEEE J. Solid-State Circuits*, vol. SC-14, no. 2, pp. 375–383, Apr. 1979.

[58] Y.-S. Pang and J. R. Brews, "Analytical subthreshold surface potential model for pocket n-MOSFETs," *IEEE Trans. on Electron Dev.*, vol. 49, no. 12, pp. 2209–2216, Dec. 2002.

[59] J. M. Pimbley and J. D. Meindl, "MOSFET scaling limits determined by subthreshold conduction," *IEEE Trans. on Electron Dev.*, vol. 36, no. 9, pp. 1711–1721, Sept. 1989.

[60] T. N. Nguyen, "Small-geometry MOS transistors: physics and modeling of surface and buried channel MOSFETs," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1984.

[61] "BSIM4.2.1 MOSFET Model - Users Manual," University of California, Berkeley, 2001.

[62] D. J. Frank, Y. Taur, and H.-S. P. Wong, "Generalised scale length for 2D effects in MOSFETs," *IEEE Electron Dev. Lett.*, vol. 19, no. 10, pp. 385–387, Oct. 1998.

[63] *MEDICI: Two Dimensional Device Simulation Program, Version 2002.4*, SYNOPSYS, Feb. 2003.

[64] M. J. Van der Tol and S. G. Chamberlain, "Potential and electron distribution model for the buried-Channel MOSFET," *IEEE Trans. on Electron Dev.*, vol. ED-36, no. 4, pp. 670–689, Apr. 1989.

[65] Y. Yin and J. A. Cooper Jr., "Simple equations for the electrostatic potential in buried-channel MOS devices," *IEEE Trans. on Electron Dev.*, vol. 39, no. 7, pp. 1770–1772, Jul. 1992.

[66] Y.-T. Lee, D.-S. Woo, J. D. Lee, and B.-G. Park, "Threshold voltage reduction model for buried channel pMOSFETs using quasi-2-D poisson equation," *IEEE Trans. on Electron Dev.*, vol. 47, no. 12, pp. 2326–2333, Dec. 2000.

[67] *ISE DESSIS: Release 8.0*, Integrated Systems Engineering, 2002.

[68] S.-W. Kang, K.-S. Min, and K. Lee, "Parametric expression of subthreshold slope using threshold voltage parameters for MOSFET statistical modeling," *IEEE Trans. on Electron Dev.*, vol. 43, no. 9, pp. 1382 –1386, Sep 1996.

[69] T. Skotnicki, G. Merckel, and T. Pedron, "Analytical study of punchthrough buried-channel P-MOSFETs," *IEEE Trans. on Electron Dev.*, vol. 36, no. 4, pp. 690–705, Apr. 1989.

[70] M. J. Van der Tol and S. G. Chamberlain, "Drain-induced barrier lowering in buried-channel MOSFETs," *IEEE Trans. on Electron Dev.*, vol. 40, no. 4, pp. 741–749, Apr. 1993.

[71] Q. Chen, B. Agrawal, and J. D. Meindl, "A comprehensive analytical subthreshold swing (S) model for double-gate MOSFETs," *IEEE Trans. on Electron Dev.*, vol. 49, no. 6, pp. 1086 –1090, Jun 2002.

[72] R. M. Swanson and J. D. Meindl, "Ion-implanted complementary MOS transistors in low-voltage circuits," *IEEE J. Solid-State Circuits*, vol. SC-7, no. 2, pp. 146–153, Apr. 1972.

[73] B. L. Austin, K. A. Bowman, X. Tang, and J. D. Meindl, "A low power transregional MOSFET model for complete power-delay analysis of CMOS gigascale integration (GSI)," in *IEEE International ASIC Conference*, 1998, pp. 125–129.

[74] M. Orshansky, J. An, C. Jiang, B. Liu, C. Riccobene, and C. Hu, "Efficient generation of pre-silicon MOS Model parameters for early circuit design," *IEEE J. Solid-State Circuits*, vol. 36, no. 1, pp. 156–159, Jan. 2001.

[75] A. Nabavi-Lishi and N. C. Rumin, "Inverter models of CMOS gates for supply current and delay evaluation," *IEEE Trans. on CAD of Integrated Ckts and Systems*, vol. 13, no. 10, pp. 1271–1279, Oct. 1994.

[76] L. Bisdounis, S. Nikolaidis, and O. Koufopavlou, "Analytical transient response and propagation delay evaluation of the CMOS inverter for short channel devices," *IEEE J. Solid-State Circuits*, vol. 33, no. 2, pp. 302–306, Feb. 1998.

[77] S. H. K. Embabi and R. Damodaran, "Delay models for CMOS, BiCMOS and BiNMOS Circuits and their applications to timing simulation," *IEEE Trans. on CAD of Integrated Ckts and Systems*, vol. 13, no. 9, pp. 1132–1142, Sept 1994.

[78] A. A. Hamoui and N. C. Rumin, "An analytical model for current, delay, and power analysis of submicron CMOS logic circuits," *IEEE Trans. on Ckts and Systems-II: Analog and Digital Signal Processing*, vol. 47, no. 10, pp. 999–1007, Oct. 2000.

[79] T. Sakurai and A. R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE J. Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, Apr. 1990.

[80] B. L. Austin, X. Tang, J. D. Meindl, M. Dennen, and W. Richards, "Threshold voltage roll-off model for low power bulk accumulation MOSFETs," in *IEEE International ASIC Conference*, 1998, pp. 175–179.

[81] A. J. Bhavnagarwala, B. L. Austin, K. A. Bowman, and J. D. Meindl, "A minimum total power methodology for projecting limits on CMOS GSI," *IEEE Trans. on VLSI Systems*, vol. 8, no. 3, pp. 235–251, June 2000.

[82] MOSIS, "Mosis parametric test results," *http://www.mosis.org/cgi-bin/params/tsmc-025/t02d-params.txt*.

[83] N. D. Arora, R. Rios, and C.-L. Huang, "Modeling the polysilicon depletion effect and its impact on submicrometer CMOS circuit performance," *IEEE Trans. on Electron Dev.*, vol. 42, no. 5, pp. 935–943, May 1995.

[84] B. Cheng, B. Maiti, S. Samavedam, J. Grant, B. Taylor, P. Tobin, and J. Mogab, "Metal gates for advanced sub-80-nm SOI CMOS technology," in *Symp. on VLSI Tech.*, 2001, pp. 91 –92.

[85] J. W. Tschanz, S. Narendra, R. Nair, and V. De, "Effectiveness of adaptive supply voltage and body bias for reducing impact of parameter variations in low power and high performance microprocessors," *IEEE J. Solid-State Circuits*, vol. 38, no. 5, pp. 826–829, May 2003.

[86] S. Matsuda, H. Yamakawa, A. Azuma, and Y. Toyoshima, "Performance improvement of metal gate CMOS technologies," in *Symp. on VLSI Tech.*, 2001, pp. 63 –64.

[87] M. J. van Dort, P. H. Woerlee, and A. J. Walker, "A simple model for quantization effects in heavily-doped silicon MOSFETs at inversion conditions," *Solid St. Electronics*, vol. 37, pp. 411–414, 1994.

[88] T. Grasser, H. Kosina, and S. Selberherr, "Investigation of spurious velocity overshoot using Monte Carlo data," *Applied Physics Letters*, vol. 79, no. 12, pp. 1900-1902, 2001.

[89] F. Bufler, Y. Asahi, H. Yoshimura, C. Zechner, A. Schenk, and W. Fichtner, "Monte Carlo simulation and measurement of nanoscale n-MOSFETs," *IEEE Trans. on Electron Dev.*, vol. 50, no. 2, pp. 418–424, Feb. 2003.

[90] J. B. Roldan, F. Gamiz, J. A. Lopez-Villanueva, and J. E. Carceller, "Modeling effects of electron-velocity overshoot in a MOSFET," *IEEE Trans. on Electron Dev.*, vol. 44, no. 5, pp. 841–846, May 1997.

[91] R. Chau, R. Arghavani, D. Barlage, G. Dewey, B. Doyle, M. Doczy, J. Kavalieros, D. Lionberger, A. Murthy, B. Roberds, and R. Schenker, "30nm physical gate length CMOS transistors with 1.0 ps n-MOS and 1.7 ps p-MOS gate delays," http://www.intel.com/research/silicon/micron.htm.

[92] K. A. Bowman, L. Wang, X. Tang, and J. D. Meindl, "A circuit level perspective of the optimum gate oxide thickness," *IEEE Trans. on Electron Dev.*, vol. 48, no. 8, pp. 1800–1810, Aug 2001.

[93] M. S. Lundstrom, *Fundamentals of Carrier Transport.* Reading, Massachussets: Addison-Wesley, 2000.

[94] M. Ieong and T.-W. Tang, "Influence of hydrodynamic models on the prediction of submicrometer device characteristics," *IEEE Trans. on Electron Dev.*, vol. 44, no. 12, pp. 2242–2251, Dec. 1997.

[95] K. Banoo, "Direct solution of the Boltzmann transport equation in nanoscale Si devices," Ph.D. dissertation, Purdue University, Dec. 2000.

[96] P. Bordone, M. Pascoli, R. Brunetti, A. Bertoni, C. Jacoboni, and A. Abramo, "Quantum transport of electrons in open nanostructures with the Wigner function formalism," *Phys. Rev.*, vol. 59, pp. 3060–69, 1999.

[97] S. Datta, *Electronic Transport in Mesoscopic Systems.* New York: Cambridge University Press, 1995.

[98] *ISE SPARTA: Release 8.0*, Integrated Systems Engineering, 2002.

[99] *ISE DEGAS: Release 8.0*, Integrated Systems Engineering, 2002.

[100] J. Bude, "MOSFET modeling into the ballistic regime," in *Simulation of Semiconductor Processes and Devices (SISPAD)*, 2000, pp. 23–26.

[101] M. Darwish, J. Lentz, M. Pinto, P. Zeitzoff, T. Krutsick, and H.-H. Vuong, "An improved electron and hole mobility model for general purpose device simulation," *IEEE Trans. on Electron Dev.*, vol. 44, no. 9, pp. 1529–1538, Sept. 1997.

[102] Q. Chen, "Scaling limits and opportunities for double gate MOSFETs," Ph.D. dissertation, Georgia Tech, Atlanta, GA, May 2003.

[103] D. Frank, R. Dennard, E. Nowak, P. Solomon, Y. Taur, and H.-S. Wong, "Device scaling limits of Si MOSFETs and their application dependencies," *Proc. IEEE*, vol. 89, no. 3, pp. 259–288, March 2001.

[104] R. Murali, B. L. Austin, L. Wang, and J. D. Meindl, "Short channel modeling of bulk accumulation MOSFETs," Submitted to *IEEE Trans. on Electron Dev.*

[105] C. Fiegna, H. Iwai, T. Wada, T. Saito, and E. Sangiorgi, "A new scaling methodology for the 0.1-0.025μm MOSFET," in *Symposium on VLSI Technology*, 1993, pp. 33–34.

[106] B. Agrawal, V. K. De, J. M. Pimbley, and J. D. Meindl, "Short channel models and scaling limits of SOI and bulk MOSFETs," *IEEE J. Solid-State Circuits*, vol. 29, no. 2, pp. 122–125, Feb. 1994.

[107] D. Antoniadis, I. Djomehri, K. Jackson, and S. Miller, "Well-tempered bulk-Si NMOSFET device home page," http://www-mtl.mit.edu/Well/.

[108] R. Stratton, "Diffusion of hot and cold electrons in semiconductor barriers," *Phys. Rev.*, vol. 126, no. 6, pp. 2002–14, 1962.

VITA

Raghunath Murali was born in Bangalore, India, on Feb. 9, 1978. He received a Bachelor of Engineering degree from Karnataka Regional Engineering College, Surathkal (now National Institute of Technology, Karnataka), in 1999, a M.S. in Electrical Engineering from Georgia Institute of Technology, Atlanta in 2000, and a Ph.D. in Electrical Engineering, also from Georgia Tech, in 2004.