A Charge-Controlled Model for MOS Transistors

Thesis by

Mary Ann Maher

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1989

(Submitted May 30, 1989)

©1989

Mary Ann Maher

All Rights Reserved

Acknowledgments

I express my deepest gratitude to Carver Mead for being my advisor. He has provided inspiration, encouragement and emotional support as well as a wonderful environment in which to learn to do research. It has been a privilege and a pleasure to observe and learn from his philosophy of science and his insights into device physics. Among other things he taught me the value of "real data", and to solve the simple problem first. I had a lot of fun; I hope you did too, Carver.

Thanks to the members of our research group, for their friendship, support, and technical interaction: Tobi Delbruck, Steve DeWeerth, Michael Emerling, David Feinstein, Dave Gillespie, John Lazarro, John LeMoncheck, Tzu Mu Lin, Jin Luo, Misha Mahowald, John Platt, Sylvie Rychbusch, Mass Sivilotti, John Tanner, and John Wawryznek. A special thanks to my office mate, Massimo Sivilotti and to Misha Mahowald for many fruitful discussions. I also thank Chuck Seitz and Ralph Cavin for their encouragement over the years. I express my appreciation to Al Barr, Chuck Seitz, Tom McGill, Carver Mead and John Linvill for serving on my committee.

I gratefully acknowledge the assistance of the many people who contributed to my thesis through technical support. The tools I used to design my test structures were written by Mass Sivilotti, Glen Gribble and John Tanner. The figures were drawn using a graphical

editor written by Glen Gribble and data plotting software written by Dave Gillespie. I truly appreciate the willingness of Dave and Glen to help with any software problems I had. Jim Campbell helped me to order equipment. Cal Jackson provided support of TEX. Dick Lyon and Lyn Dupre helped me to improve my writing style. Thanks to Carolyn Collins, Helen Derevan, and Diane Dysland for their administrative support and kindness.

This work was supported by the System Development Foundation and the Semiconductor Research Corporation. Experimental devices were kindly donated by Amr Mohsen and Gerry Parker of Intel Corporation. Computer support was provided by an equipment grant from Hewlett Packard. I have been supported by a faculty development fellowship from Hewlett Packard.

Thanks to my parents and my friends, especially Zoya Popovic, for their encouragement and understanding. Special thanks to John LeMoncheck who made the final weeks a real joy. I am deeply indebted to my dear friends Al Barr, Rajiv Gupta and Tim Kay, who were always there and helped me in countless ways.

Abstract

As MOS (metal-oxide-semiconductor) devices scale to submicron lengths, short-channel effects begin to dominate device behavior, and designers of VLSI (very-large-scale-integrated) circuits see an improved transistor model as a necessary tool. A new physically based, charge-controlled model for the DC current, the intrinsic terminal charges, and the transcapacitances in the MOS transistor under quasistatic conditions has been developed. The model expresses the current in the MOS transistor in terms of the mobile charge per unit area in the channel, and uses a complete set of natural units for velocity, voltage, length, charge, and current. The current-flow equation for the transistor includes both a drift term and a diffusion term, so that the formulation applies equally over the subthreshold, saturation, and "ohmic" regions of transistor operation and includes the effect of velocity saturation. The solution of this dimensionless current-flow equation using these units is a simple, continuous expression, and is suitable for the computer simulation of integrated circuits. The expression allows the regimes of transistor operation and the behavior of long-channel devices versus short-channel devices to be discerned easily. In particular, the expressions for source and drain terminal charges combine the mobile charge in the channel at the source and drain ends in simple polynomials. Analysis of the model shows a fundamental relation between the transistor transcapacitances and transconductances, and

permits the development of efficient simulation models of them.

Our physically based transistor model uses parameters derived from the fabrication process by direct measurement and from the dimensions of the device. The zero-order model agrees closely with measurements on the scaling of current with channel length down to submicron channel lengths. For more detailed analog simulations, the model contains several first-order effects calculated as perturbations on the simple model. Comparisons among calculated and measured curves of conductance, capacitance, and drain currents demonstrate the accuracy of the model both above and below threshold for a number of experimental devices of different channel lengths. Results from the model concur with measurements on short-channel transistors down to 0.6-micron channel length. Several analog circuit simulators now contain the model.

Contents

Acknowledgments iii

Abstract v

List of Figures x

List of Symbols xiii

1 Introduction 1

1.1 Overview 3

1.2 Chapter 2 4

1.3 Chapter 3 5

1.4 Chapter 4 6

1.5 Chapter 5 6

2 The General MOS Model 8

2.1 Electrostatics 9

2.2 Channel Current 14

2.3 Natural Units 16

2.4 Source Boundary Condition 18

2.5 Saturation 20

2.6 Drain Boundary Condition 21

2.7 Model Evaluation 23

2.8 Experimental Results 23

2.9 Summary 27

3 First-Order Corrections to the Basic Model **28**

3.1 Mobility Variation 30

3.2 Channel-Length Modulation 31

3.3 Resistance 36

3.4 The Upgraded Model 36

3.5 Experimental Results 37

3.6 Conclusions 43

4 Model Behavior **44**

4.1 The Subthreshold Model 45

4.1.1 Current 45

4.1.2 Drain Conductance 48

4.1.3 Body Effect 51

4.2 The Model Above Threshold 52

4.2.1 Long-Channel Limit in Saturation 52

4.2.2 Short-Channel Transistor in Saturation 53

4.2.3 Drain Conductance 54

4.3 Conclusions 56

5 The Charge Model 57

5.1 Calculation of Mobile Charge 58

5.1.1 Calculation of the Total Mobile Charge 61

5.2 The Terminal Charges 62

5.3 Definition of Transcapacitance 64

5.4 Calculation of Transcapacitance and Transconductance 64

5.5 Comparision of Theory and Experiment 66

5.5.1 Simulation Results 66

5.5.2 Comparison with Ward's Data 66

5.5.3 Experimental Data 66

5.5.4 Comparison of the Model with Our Data 66

5.6 Conclusions 69

6 Conclusions and Future Work 70

6.1 Summary 70

6.2 Circuit Simulation 72

6.3 Future Work 73

A Transcapacitance and Transconductance Calculations 75

A.1 Transcapacitance 75

A.2 Terminal Derivatives 76

A.3 Transconductance 77

References 78

List of Figures

2.1 Cross-section of an MOS transistor. 9

2.2 Energy diagram of a cross-section through an n-channel MOS transistor . . 10

2.3 Contributions of the surface-potential and of the mobile charge to the gate voltage 13

2.4 Saturation current above threshold versus $1/l$ 20

2.5 Determination of substrate doping. 24

2.6 Drain current versus v_{ds} for fixed v_{gs} for a 50-micron channel length transistor. 25

2.7 Drain current versus v_{ds} for fixed v_{gs} for a 2.8-micron channel length transistor. 25

2.8 Drain current versus v_{ds} for fixed v_{gs} for a 0.7-micron channel length transistor. 26

2.9 Drain current versus v_{ds} for fixed v_{gs} for a 0.35-micron channel length transistor. 26

3.1 Conductance and slope of conductance for a 50-micron channel length transistor. 32

3.2 Drain current versus v_{ds} for $v_{gs} = 0.5$ volts for a 0.7-micron channel length transistor. 32

3.3 Drain Current versus v_{ds} for $v_{gs} = 3$ volts for a 0.7-micron channel length transistor 34

3.4 Conductance and slope of conductance for a 0.7-micron channel length device 35

3.5 Drain current versus v_{ds} for $v_{gs} = 0 - 3$ volts for a 50-micron channel length transistor. 38

3.6 Drain current versus v_{ds} for $v_{gs} = 0 - 3$ volts for a 2.8-micron channel length transistor. 38

3.7 Drain current versus v_{ds} for $v_{gs} = 0 - 3$ volts for a 2-micron channel length transistor. 39

3.8 Drain current versus v_{ds} for $v_{gs} = 0 - 3$ volts for a 0.7-micron channel length transistor. 39

3.9 Drain current versus v_{ds} for $v_{gs} = 0 - 3$ volts for a 0.6-micron channel length transistor. 40

3.10 Drain current versus v_{ds} for $v_{gs} = 0 - 3$ volts for a 0.35-micron channel length transistor. 40

3.11 Drain current versus v_{ds} for $v_{gs} = 0.5$ volts for a 50-micron channel length transistor. 41

3.12 Drain current versus v_{ds} for $v_{gs} = 0.5$ volts for a 2.8-micron channel length transistor. 41

3.13 Drain current versus v_{ds} for $v_{gs} = 0.5$ volts for a 0.6-micron channel length transistor. 42

3.14 Conductance versus v_{gs} for a 2.8-micron channel length transistor. 42

3.15 Conductance versus v_{gs} for a 0.6-micron channel length transistor. 43

4.1 Saturation current times channel length versus v_{ds} for transistors in subthreshold 49

4.2 V_0 versus channel length 50

4.3 V_0 versus channel length for various V_g. 55

5.1 Mobile charge distribution for the subthreshold region 59

5.2 Mobile charge distribution for an above-threshold, long-channel, transistor. 59

5.3 Mobile charge distribution for a velocity-saturated transistor. 60

5.4 Mobile charge distribution for a 3-micron channel length transistor above threshold 60

5.5 $\mathcal{Q}_s/\mathcal{Q}_m$ and $\mathcal{Q}_d/\mathcal{Q}_m$ versus v_{ds} 63

5.6 Model calculation of transcapacitance versus gate voltage. 67

5.7 C_{dg} and C_{gd} versus gate voltage 68

5.8 C_{gd} versus gate voltage for $v_{ds} = 5$ volts and $v_{sb} = 0$ volts. 68

List of Symbols

C channel capacitance

C_{b} depletion capacitance

C_{d} channel capacitance at the drain end of the channel

C_{ij} transcapacitance between terminals i and j

C_{ox} oxide capacitance

C_{s} channel capacitance at the source end of the channel

D diffusion constant

E energy

E_{c} energy at the conduction band edge

E_{f} quasi-fermi level

E_{T} thermal energy, natural unit for energy

$\mathcal{E}_{\mathrm{ch}}$ electric field along the channel

$\mathcal{E}_{\mathrm{d}}$ electric field along the channel at the drain end

$\mathcal{E}_{\mathrm{f}}$ mobility degradation parameter

$\mathcal{E}_{\mathrm{g}}$ vertical electric field

$\mathcal{E}_{\mathrm{gd}}$ vertical electric field at the drain end

$\mathcal{E}_{\mathrm{gs}}$ vertical electric field at the source end

g conductance

g_{d} drain conductance

G_{j} transconductance $\frac{\partial I}{\partial V_j}$

i DC drain–source current

I normalized DC drain–source current

I_{b} normalized bulk current

I_{d} normalized drain current

I_{g} normalized gate current

I_{s} normalized source current

I_{tot} normalized total current

I_0 natural unit for current

k Boltzmann's constant

l transistor channel length

L normalized transistor channel length

L_{eff} effective channel length

l_0 natural unit for length

N_A density of acceptors in the substrate

N_c density of states in the channel at the drain end

N_d density of states in the drain

n_0 effective density of states in the channel

P_c Probability of a state in the channel at the drain end being occupied

P_d Probability of a state in the drain being occupied

P_cd Probability of transition of a carrier from channel to drain

P_dc Probability of transition of a carrier from drain to channel

q charge on the electron

q_b depletion charge

q_d Mobile charge at the drain end of the channel

Q_d normalized mobile charge at the source end of the channel

q_m mobile charge per unit area in the channel

Q_m normalized mobile charge per unit area in the channel

q_s Mobile charge at the source end of the channel

Q_s normalized mobile charge at the source end of the channel

Q_{T} thermal charge, natural unit for charge

q_{tot} total charge per unit area in the channel

q_0 effective charge per unit area in the channel

Q_0 normalized effective charge per unit area in the channel

T absolute temperature

t_{ox} oxide thickness

v voltage

V normalized voltage

V_{b} normalized bulk voltage

v_{b} bulk voltage

v_{d} drain voltage

V_{d} normalized drain voltage

V_{g} normalized gate voltage

v_{g} gate voltage

V_{s} normalized source voltage

v_{s} source voltage

v_{ds} drain–source voltage

V_{ds} normalized drain–source voltage

v_{fb} flat band voltage

v_{gs} gate–source voltage

V_{gs} normalized gate–source voltage

V_T thermal voltage, natural unit for voltage

V_{th} normalized threshold voltage

v_0 saturated velocity of mobile carriers in silicon

V_0 Early voltage

w transistor width

x distance into the silicon perpendicular to the surface

X normalized distance into the silicon perpendicular to the surface

x_0 depletion layer width along the channel

z distance along the channel

Z normalized distance along the channel

γ back gate coefficient

ΔL_s width of depletion layer at the source end of the channel

ΔL_d width of depletion layer at the drain end of the channel

ϵ_{ox} relative permittivity of the oxide

ϵ_s relative permittivity of the silicon

ϵ_0 permittivity of free space

κ derivative of the mobile charge with respect to surface potential

μ mobility in the channel

ξ quasi-fermi potential

Ξ normalized quasi-fermi potential

Ξ_d normalized quasi-fermi potential at the drain end of the channel referred to the fermi potential

ρ substrate doping density

ρ_{eff} effective doping density

ρ_m effective mobile charge density

ψ surface potential

Ψ normalized surface potential

ψ_f fermi potential

ψ_{ms} built-in barrier potential

Ψ_0 normalized operating point for the surface potential

Ψ_s normalized surface potential at the source end of the channel

Ψ_d normalized surface potential at the drain end of the channel

ψ_s surface potential at the source end of the channel

ψ_d surface potential at the drain end of the channel

$\mathcal{Q}_g$ normalized gate terminal charge

$\mathcal{Q}_s$ normalized source terminal charge

$\mathcal{Q}_d$ normalized drain terminal charge

$\mathcal{Q}_b$ normalized bulk terminal charge

Chapter 1

Introduction

As MOS (metal-oxide-semiconductor) devices scale to submicron lengths, short-channel effects begin to dominate device behavior, and designers of VLSI (very large-scale integrated) circuits see an improved transistor model as an essential tool. Accurate models for MOS transcapacitances and charges are necessary for understanding and simulating the detailed operation of traditional dynamic circuit forms and charge-dependent MOS circuits. Signal margins in dynamic RAMS, error voltages on the floating nodes of analog switches, switched capacitor filters, and the timing and voltage excursion of bootstrapped drivers all need more care in simulation than do signals in strictly digital integrated circuits. Many charge-conserving numerical-integration routines for use in circuit simulation require the model to supply equations for both the transistor charges and the transistor transcapacitances [Yang 82]. If the device model does not provide an analytical method for evaluating the device charge, the simulation algorithm calculates the device charge by integrating the capacitance equations. The path-dependent result of this integration often leads to nonconservation of charge. In addition, since the accuracy of the transistor model directly

determines the accuracy of a transient simulation, simulators demand complete and detailed charge models.

The evolution of a wider range of circuit techniques for CMOS (complementary metal-oxide-semiconductor) technology compels designers to understand and model accurately the DC as well as the AC transistor characteristics. New circuits working in subthreshold, hybrid analog/digital chips, BICMOS and speed-independent systems present modeling challenges. Large-scale analog systems emulating the behavior of the sensory systems of animals operate over many orders of magnitude in signal level [Mead 89]. The circuits in these systems require transistors operating above and below threshold, in saturation and in the "ohmic" (low drain–source voltage v_{ds}) region. Furthermore, these circuits cross the boundaries between these regions during normal operation. Models that describe the different regimes of operation with separate equations lead to discontinuous derivatives in the transition from subthreshold and in the transition into saturation. The device models used in circuit simulation must provide accurate, continuous expressions for charge, transcapacitance and current that are valid over all regimes of transistor operation.

Device models that are developed for circuit simulation must trade off speed for accuracy. This tradeoff is becoming more and more crucial as VLSI circuit designers attempt to simulate the hundreds of thousands of transistors found on a chip. As transistor channel lengths shrink, modeling becomes increasingly difficult as two- and three- dimensional effects become important and add complexity to the modeling process. Finally, if circuit designers are to truly take advantage of the physics of the devices in designing circuits, they must have access to models that are physically based and that will predict circuit properties as the device geometries and the processes vary.

1.1 Overview

In this thesis, we describe a new model for the DC current, transcapacitances, and intrinsic terminal charges of the MOS transistor under quasistatic operation that applies equally in the subthreshold, superthreshold, saturation, and ohmic regions of transistor operation. We use a charge-controlled approach related to the work of Linvill [Linvill 58], and we write the current in the MOS transistor in terms of the mobile charge per unit area in the channel. The model equations are simple and efficient enough to be suitable for computer simulation of integrated circuits, but capture the essential physical behavior needed to describe accurately the transistor operation for analog circuits.

General analyses of the transistor have been published, and there have been many types of models proposed [Pao 66,Brews 78,El-Mansy 77a] [Sheu 84,Baccarani 78,Vogel 85]. However, many of these models are too complicated for circuit simulation. Empirical models, although based on nonphysical assumptions, may still be able to predict I-V curves because they often contain a large number of adjustable parameters. They may not perform well outside the range of the parameter fit, however, and may contain nonphysical, length- and width-dependent parameters. Also, it is difficult to reason about them on the basis of physical knowledge. Table look-up models, while efficient for a well characterized process, are limited in scope as they cannot model a large range of device sizes or process parameters.

Our physically based model uses a small set of parameters that are derived from the fabrication process by direct measurement and from the dimensions of the devices. Extraction of parameters by global curve-fitting and optimization programs can lead to parameters that have misleading and nonphysical values. We develop a model that is physically meaningful

and intuitive so that circuit designers can reason about it and can rely on basic physical knowledge to verify model performance. By bridging the gap between the device physics and the circuit designer, the model provides a synthesis tool as well as an analysis tool. The charge-controlled approach allows us to make the model extendable and to develop models at different levels of abstraction. In this dissertation, we also derive a simplified model from our simulation model for circuit reasoning.

1.2 Chapter 2

We give a review of our charge-controlled DC transistor model in Chapter 2. We first develop a zero-order model assuming constant channel length and mobility for a given device. We describe the model in terms of saturation, where charge at the drain moves at saturated velocity. The more involved description of the transistor when it is not in saturation is treated as an extension of the saturation case. The current-flow equation includes both a drift term and a diffusion term, so the formulation applies equally over all regions of transistor operation. This treatment takes into account the finite velocity attained by carriers as they are accelerated in very high electric fields (velocity saturation). We define a complete set of natural units for velocity, voltage, length, charge, and current that relates the physical current-flow mechanisms and provides the key to developing the model equations. The simple, continuous solution of the dimensionless current-flow equation using these units allows us to discern easily the regimes of transistor operation and the behavior of long-channel versus short-channel devices.

Because the model is charge-controlled and continuous, it contains no explicit threshold voltage. However, we show that the common notion of threshold occurs at the point of

unit surface charge, where drift and diffusion make equal contributions to the current. The model includes the threshold shift due to the source–substrate bias (body effect) directly, with no new terms or parameters added to the model.

We present results on the zero-order model, comparing theory and experiment on transistors down to 0.35-micron channel length. The theoretical curves use the same set of parameters in all cases. We investigate the scaling of the device characteristics as the transistor dimensions are made smaller and we demonstrate the model's accuracy by presenting plots of saturation current versus channel length.

1.3 Chapter 3

In Chapter 3, for the more detailed analog simulations, we examine several first-order effects as perturbations on the zero-order model. Some of the effects, such as channel-length modulation, increase the current; others, such as mobility variation and internal resistances, decrease the current. All the effects, and the first-order corrections to the model itself, are of roughly the same magnitude. We find a particular regime of operation in which one of the effects is dominant, and we evaluate the effect there. Accurately modeling the channel-length modulation by drain voltage is crucial, as devices scale to submicron lengths. We include this effect by calculating an effective channel length as a function of drain and source bias. Observing a subthreshold, short-channel transistor in saturation shows this process nearly independently of other physical effects. We then use transistors above threshold to study the effect of mobile charge on the depletion-layer length. We also include the variation of mobility with gate electric field in the model by looking at the conductance of long-channel transistors. We discuss the consequences of parasitic resistance; an effect

that shows up at the higher current levels found in short-channel transistors. In addition, we examine higher-order approximations for the depletion capacitance and the gate electric field as a function of distance along the channel.

Comparisons among calculated and measured curves of conductance, the slope of the conductance, and the drain currents demonstrate the accuracy of the model both above and below threshold for a number of experimental devices of different channel lengths. We present detailed results on 50-micron, 3-micron, 2-micron, 0.7-micron, 0.6-micron, and 0.35-micron channel length devices supplied by Intel Corporation.

1.4 Chapter 4

In Chapter 4, we examine the behavior of the model in different regions of transistor operation. We compare the charge-controlled model expressions with those of typical voltage-based models in both the subthreshold and above-threshold cases. We also derive simplified forms of the transistor model useful to circuit designers in hand calculations. We deduce an expression for the drain conductance of the transistor in saturation in subthreshold, based on the Early voltage model used in bipolar transistors [Early 52]. We extend this model above threshold and include the effects of velocity saturation. Because the simplified models approximate the more complicated simulation model, designers can use their physical intuition about these models to verify and reason about the simulation results.

1.5 Chapter 5

In Chapter 5, we extend the DC model to describe a new model for the intrinsic charges and capacitances in the MOS transistor under quasistatic conditions. The nonlinear equations

describing the charge–voltage relations cannot be described by a reciprocal, two-terminal capacitor [Ward 81,Yang 82]. Instead, we define a total intrinsic charge $\mathcal{Q}$ associated with each terminal and we calculate a set of transcapacitances $C_{ij} = \frac{\partial \mathcal{Q}_i}{\partial V_j}$ at each terminal from the intrinsic terminal charges. We use the method of Ward [Ward 81] to divide the channel charge into source and drain components.

Our new charge and capacitance expressions are ratios of simple polynomial expressions of quantities needed for calculating the DC model. The calculated charges are continuous, and the results are valid over all regimes of transistor operation. We also develop a fundamental relation between the transistor's transcapacitances and transconductances, and we propose an efficient method for calculating these quantities when they are used in a circuit simulator.

We derive expressions for the mobile charge in the channel as a function of distance along the channel, and we calculate the total mobile charge. The expression for the mobile charge in the channel correctly predicts the behavior of long- and short-channel transistors, and holds above and below threshold. Simulation results displaying the 16 MOS transcapacitances versus gate voltage show the nonreciprocal nature of the transcapacitances. Plots of the total mobile charge splitting into source and drain components reflect the proper behavior under various boundary conditions. We describe how the charges and transcapacitances scale with channel length. We then verify our results by comparison with measured experimental data. Comparison of our complete model with experimental data in our laboratory taken from chips fabricated through MOSIS, the ARPA silicon foundry [Cohen 81], and with published transcapacitance data from Ward [Ward 81], demonstrate the model's applicability.

Chapter 2

The General MOS Model

In this chapter, we develop our zero-order model for the DC current in the MOS transistor. We first derive a solution assuming that the channel length and the mobility are constant for a given device. We begin by obtaining the channel current for a transistor in saturation. We examine the electrostatics of the MOS transistor to relate the gate voltage to the mobile charge and the surface potential. We treat the case of a uniformly doped substrate with a threshold-adjustment implant at the surface. We define a fixed relation between the mobile charge and the surface potential, which we call the effective channel capacitance. This definition leads to a complete set of natural dimensionless units for velocity, voltage, length, charge, and current. We obtain the channel current i by integrating the current-flow equations from one end of the channel to the other, using the mobile charge per unit area at the source q_s as a boundary condition. The current-flow equation for the transistor includes both a drift term and a diffusion term, so the formulation applies equally over all regions of transistor operation. This treatment takes into account mobile carrier velocity saturation. The more detailed treatment of the transistor when it is not in saturation is an extension

of the saturation case where q_d, the mobile-charge density at the drain end of the channel, is set not by velocity saturation, but rather by a boundary condition involving the drain voltage. Finally, we relate the mobile charge in the channel to the terminal voltages.

We compare the model with experimental data that were taken on a set of transistors with several channel lengths, down to 0.35 microns. These devices were fabricated with an advanced process (by 1988 standards), and have a gate-oxide thickness of 125 angstroms.

2.1 Electrostatics

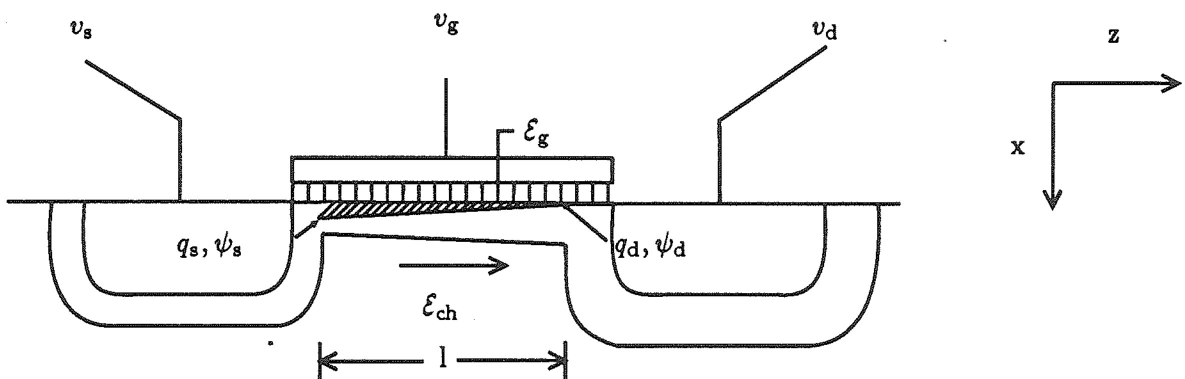

Figure 2.1: Cross-section of an MOS transistor.

Figure 2.1 is a visualization of a cross-section of an MOS transistor; it indicates the overall coordinate system. The energy diagram of a cross-section through the channel region at the barrier maximum of an n-channel MOS transistor is displayed in Figure 2.2. We assume

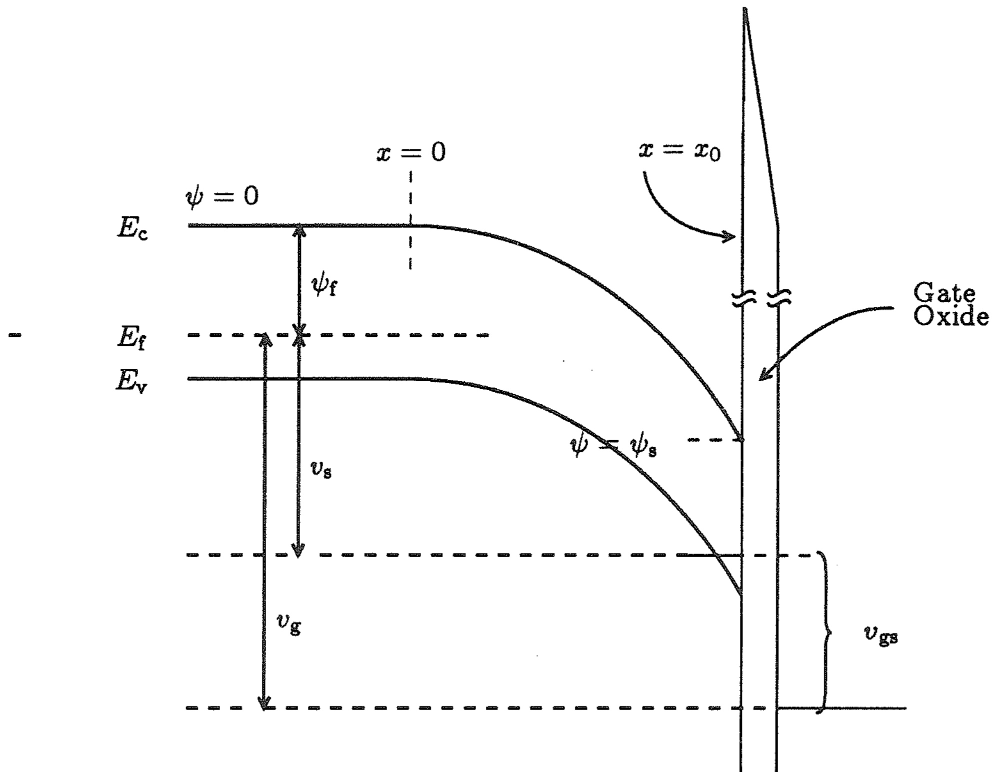

Figure 2.2: Energy diagram of a cross-section through an n-channel MOS transistor. The cross-section runs normal to the surface through the potential maximum. The zero of surface potential is the conduction band edge deep in the bulk, where the electric field is zero.

that the substrate is uniformly doped with N_A acceptors per unit volume; hence, under a depletion approximation, the depletion layer contains a constant charge density per unit volume $\rho = qN_A$. In our coordinate system, x is measured perpendicular to the surface, with $x = 0$ at the substrate edge of the depletion layer, and $x = x_0$ at the surface of the silicon. The mobile charge is quantum-mechanically distributed over some depth [El-Mansy 77a], but we will assume that it is all located at the surface, so that conditions in the bulk are not affected by the presence of mobile charge at the surface. Applying Gauss's law and integrating the electric field in the depletion layer, $\frac{\rho x}{\epsilon_s}$, from the bulk to surface yields a relation between the surface potential ψ and the depletion-layer thickness x_0, where ϵ_s is the permittivity of the silicon:

$$\psi = -\frac{1}{\epsilon_s}\frac{\rho x_0^2}{2} \tag{2.1}$$

We will use the conduction band edge deep in the substrate as the reference for the surface potentials. Using this convention, the surface potential is zero at flat-band, when the potential is flat all the way to the interface. Using the depletion-layer thickness, the total charge per unit area q_b uncovered in the depletion layer is

$$q_b = -\rho x_0 = \pm\sqrt{-2\rho\epsilon_s\psi} \tag{2.2}$$

This relation allows us to define a depletion-layer capacitance C_b:

$$C_b = -\frac{\partial q_b}{\partial \psi} = \sqrt{-\frac{\rho\epsilon_s}{2\psi}} = \frac{\epsilon_s}{x_0} \tag{2.3}$$

The applied gate voltage appears partially across the gate oxide and partially across the silicon substrate. The voltage across the gate oxide is the oxide electric field times the oxide thickness t_{ox}. The electric field is proportional to the total charge to the left of the oxide q_{tot}: the depletion-layer charge q_b, the mobile charge per unit area q_m, and the surface fixed

charge q_{ss} (which includes the interface charge q_{it} and any threshold-adjustment charge). Consequently, the gate voltage v_g is

$$\begin{aligned} v_g &= \psi - \frac{t_{ox}}{\epsilon_{ox}}(q_{tot}) + \psi_{ms} \\ &= \psi - \frac{1}{C_{ox}}(q_b + q_m + q_{ss}) + \psi_{ms} \end{aligned} \quad (2.4)$$

where $C_{ox} = \epsilon_{ox}/t_{ox}$ is the oxide capacitance per unit area, and ψ_{ms} is the the built-in potential that exists between the gate and silicon surface with zero applied gate voltage [Grove 67]. We take into account the fixed charge q_{ss} and the built-in potential by defining the term $v_{fb} = \psi_{ms} - q_{ss}/C_{ox}$. When $v_g = v_{fb}$, the surface potential is equal to zero; hence, v_{fb} is called the ***flat-band voltage*** [Grove 67].

$$v_g = \psi + v_{fb} - \frac{1}{C_{ox}}(q_b + q_m) \quad (2.5)$$

The mobile-charge density q_m is a smooth function of the position z along the channel. The surface potential ψ depends on q_m; hence, both ψ and q_b will be functions of z as well. Solving Equation 2.5 for q_m, we obtain

$$-q_m = C_{ox}(v_g - \psi - v_{fb}) + q_b \quad (2.6)$$

In subthreshold, the gate voltage follows the surface potential, because the mobile charge can be neglected. Above threshold, however, the surface potential changes little, and most of the new electric-field lines from the gate terminate on mobile charges. These relationships are shown in Figure 2.3.

We can extract useful information from Equation 2.5 for small changes in voltage around some operating point. Differentiating this equation with respect to ψ and applying Equation 2.3, we obtain

$$\frac{\partial v_g}{\partial \psi} = 1 + \frac{C_b}{C_{ox}} - \frac{1}{C_{ox}}\frac{\partial q_m}{\partial \psi} \quad (2.7)$$

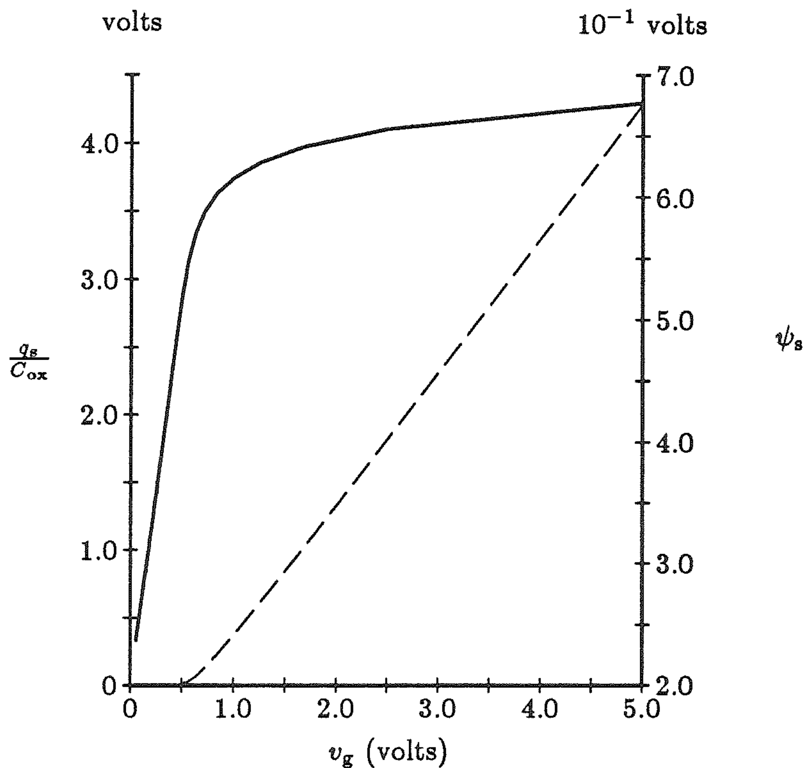

Figure 2.3: Contributions of the terms for the surface potential ψ_s (solid line) and for the mobile charge q_s/C_{ox} (dashed line) to the gate voltage, as expressed in Equation 2.5 and evaluated at the source end of the channel. In subthreshold, the gate voltage follows the surface potential, because the mobile charge is negligible. Above threshold, the major contribution to the change in surface potential is the mobile charge, as the surface potential becomes pinned.

For any given operating point, the gate is an equipotential; hence, v_g does not depend on the coordinate z along the channel, whereas the surface potential ψ changes considerably. The left-hand side of Equation 2.7 vanishes, and we can relate ψ to q_m by defining an *effective channel capacitance* C per unit area:

$$C = \frac{\partial q_m}{\partial \psi} = C_b + C_{ox} \tag{2.8}$$

This quantity represents the additional mobile charge that must be injected into the channel per unit increase in surface potential. As the mobile charge flows through the channel, there is a fixed relation between the mobile charge and the surface potential, given by Equation 2.6. Thus, the presence of mobile charge creates an electric field $\mathcal{E}_{ch}$ along the channel.

2.2 Channel Current

Given the boundary conditions on mobile-charge density and surface potential at the source, we can evaluate the channel current i in saturation. Under some circumstances, the current in the channel can be dominated by diffusion; in other regimes of operation, the same transistor may have charge carriers with drift velocities near saturation over the entire length of the channel. We will approximate the current flow by a drift and a diffusion term, and will include the effects of velocity saturation in the drift term:

$$\frac{i}{w} = q_m v_{drift} - D q_m' \tag{2.9}$$

where w is the width of the channel, D is the diffusion constant, v_{drift} is the carrier's drift velocity, and q_m' is the derivative of q_m with respect to z. Several relations have been proposed for the drift velocity as a function of electric field [Sodini 84,Baum 70,Garverick 87] We adopt a simple relation for the drift velocity that has the proper behavior at both high

and low fields [Hoeneisen 72]:

$$v_{\text{drift}} = v_0 \left(\frac{\mu \mathcal{E}_{\text{ch}}}{v_0 + \mu \mathcal{E}_{\text{ch}}} \right) \tag{2.10}$$

where v_0 is the carrier saturated velocity [Jacoboni 77]. Substituting this relation into Equation 2.9 gives

$$\frac{i}{w} = q_{\text{m}} v_0 \left(\frac{\mu \mathcal{E}_{\text{ch}}}{v_0 + \mu \mathcal{E}_{\text{ch}}} \right) - D q'_{\text{m}} \tag{2.11}$$

We now introduce an approximation with which we can, for any particular operating point, evaluate Equation 2.9 in closed form. In general, the effective channel capacitance C defined in Equation 2.8 between the mobile charge and the bulk and gate (fixed potentials) is a weak function of z, as is the mobility μ. The width of the depletion layer in the channel increases with distance toward the drain for nonzero drain voltage. This effect corresponds to a decrease in the channel capacitance. As devices scale to smaller channel lengths and correspondingly thinner oxides, C becomes less dependent on the mobile-charge density q_{m}. We will first derive the zero-order result, taking C as constant and equal to the value at the potential maximum C_{s}. In the simple form of our model, we also assume that the gate electric field $\mathcal{E}_{\text{g}}$ (and hence μ) is constant along the channel and is equal to the value at the source $\mathcal{E}_{\text{gs}}$. However, we still allow the depletion-layer charge q_{b} used in the calculation of the electrostatics to varys with distance. These restrictions on C and μ are relaxed in Chapter 3.

In the following analysis, we compute all currents for a channel of unit width. Because we know that

$$\frac{\partial q_{\text{m}}}{\partial z} = \frac{\partial q_{\text{m}}}{\partial \psi} \frac{\partial \psi}{\partial z} \approx C_{\text{s}} \mathcal{E}_{\text{ch}} \qquad \text{and} \qquad \mathcal{E}_{\text{ch}} = -\frac{\partial \psi}{\partial z}$$

we can write Equation 2.11 as

$$\frac{i}{w} = q_{\rm m} v_0 \left(\frac{-\mu C_{\rm s} q'_{\rm m}}{v_0 - \mu C_{\rm s} q'_{\rm m}} \right) - D q'_{\rm m} \tag{2.12}$$

2.3 Natural Units

Using the value of the channel capacitance at the source C_s, we can define a set of natural units for describing the transistor equations:

velocity	v_0
energy	$\mathcal{E}_{\rm T} = kT$
voltage	$V_{\rm T} = \frac{kT}{q}$
charge	$Q_{\rm T} = \frac{kT}{q} C_{\rm s}$
current	$I_0 = v_0 Q_{\rm T}$
length	$l_0 = \frac{D}{v_0} = \frac{\mu kT}{q v_0}$

The thermal charge $Q_T = C_s kT/q$ is the mobile charge per unit area at the source required to change the surface potential by exactly kT/q. By expressing all variables in terms of these natural units, we simplify all equations into a dimensionless form that allows us to discern easily the natural relationships and regimes of operation of the transistor.

We use $Q_{\rm m}$ for the dimensionless form of the mobile charge density $q_{\rm m}$, and I for the dimensionless form of the current i. We write Equation 2.11 in dimensionless form as

$$-I = \frac{Q_{\rm m} Q'_{\rm m}}{Q'_{\rm m} + 1} + Q'_{\rm m} \tag{2.13}$$

$$-I(-Q'_{\rm m} + 1) = Q_{\rm m} Q'_{\rm m} + Q'_{\rm m}(Q'_{\rm m} + 1) \tag{2.14}$$

where $Q'_{\rm m}$ indicates the derivative with respect to Z, the dimensionless distance along the channel.

The first term on the right-hand side of Equation 2.13 is the drift term, and the second is the diffusion term. Using natural units helps to show which terms dominate in a given region of transistor operation. The shortest channel length for which a device can be made to operate is still much longer than is the mean-free path of a carrier, so velocity saturation is not achieved for the diffusion process. We therefore assume that the ${Q'_{\rm m}}^2$ term is negligible compared with either the $Q_{\rm m}Q'_{\rm m}$ term (when $Q_{\rm m}$ is large) or the $Q'_{\rm m}$ term (when $Q_{\rm m}$ is small). This approximation is excellent as long as l_0 is much less than the device dimensions. For a typical n-channel process, l_0 is about 0.007 micron. We thus can write Equation 2.14 as

$$-I(-Q'_{\rm m}+1) \cong Q_{\rm m}Q'_{\rm m} + Q'_{\rm m} = Q'_{\rm m}(Q_{\rm m}+1)$$

or as

$$-I = Q_{\rm m}Q'_{\rm m} + Q'_{\rm m}(1-I) = \frac{\partial}{\partial Z}\left(\frac{Q_{\rm m}^2}{2} + Q_{\rm m}(1-I)\right) \tag{2.15}$$

We now integrate both sides of Equation 2.15 along the channel from source ($Z=0$) to drain ($Z=L$), where L is the normalized channel length. We note that I is not a function of Z and apply the boundary conditions $Q_{\rm m}=Q_{\rm s}$ at the source and $Q_{\rm m}=Q_{\rm d}$ at the drain:

$$IL = \frac{Q_{\rm s}^2 - Q_{\rm d}^2}{2} + (Q_{\rm s}-Q_{\rm d})(1-I) \tag{2.16}$$

Solving Equation 2.16 explicitly for I gives

$$I = \frac{Q_{\rm s}-Q_{\rm d}}{Q_{\rm s}-Q_{\rm d}+L}\left(\frac{Q_{\rm s}+Q_{\rm d}}{2}+1\right) \tag{2.17}$$

We can gain important insights into the operation of MOS devices from these equations. For sufficiently large L, the current is small compared with unity, and $1-I \approx 1$.

This approximation corresponds to the usual treatment, ignoring velocity-saturation effects. Tracing through the derivation, we see that the quadratic term in Equation 2.16 is a result of the drift term in Equation 2.9, and the linear term is a result of the diffusion term in Equation 2.9 and velocity saturation. The two terms make approximately equal contributions to the saturation current for $q_s = Q_T$ (in natural units $Q_s = 1$). The gate voltage at which Q_s is equal to 1 (called V_{th} in natural units), corresponds to the common notion of *threshold* and, at threshold, there is no discontinuity. We conclude that *below threshold, current flows by diffusion; above threshold, current flows by drift.* For larger Q_s, the surface potential is dominated by mobile charge; for smaller Q_s, the surface potential is determined by the charge in the depletion layer.

2.4 Source Boundary Condition

We use a charge sheet approximation and assume that the mobile charge is all located at the surface of the channel [Brews 78]. The more general case has been worked out by [Pao 66]. We formulate the boundary condition on q_{m} by the integral with respect to the energy E of the carrier density in the source region (a Fermi distribution), times the two-dimensional density of states $n(E)$ (number per unit area, per unit energy) in the channel:

$$q_{\mathrm{m}} = q \int_{E_c}^{\infty} \left(n(E) \frac{1}{e^{(E-E_{\mathrm{f}})/(kT)} + 1} \right) dE \qquad = q \int_{q\psi}^{\infty} \left(n(E) \frac{1}{e^{(E-q\xi)/(kT)} + 1} \right) dE$$

where E_{c} is the energy at the conduction band edge E_{f} is the quasi-fermi level and ξ is a quasi–Fermi potential or *imref* [Shockley 60]. For any realizable bias conditions, even for submicron devices, the Fermi level is always many kT below the surface potential. The usual approximation, in which the Fermi function is replaced by a Boltzmann approximation, is

thus valid. The resulting expression can be written as

$$q_{\mathrm{m}} = q_0 e^{(-q\psi+q\xi)/(kT)} \tag{2.18}$$

where $q_0 = qn_0$, is an effective charge density and n_0, the effective density of states in the channel, is given by

$$n_0 = \int_{q\psi}^{\infty} n(E) e^{-(E-q\xi)/(kT)}\, dE$$

For a real channel, the form of $n(E)$ is not known; thus, we cannot compute n_0 with confidence. The spacing of the energy levels in the channel is a function of the electric field at the surface; hence, $n(E)$ is proportional to $1/\sqrt{\psi}$, a slow dependence that we neglect compared with the exponential due to the Boltzmann function. As long as we have the value of n_0 in the right general order of magnitude, however, the effect of an error in the pre-exponential factor is easily masked by small changes in surface potential. Solving Equation 2.18 for ψ, we obtain

$$-\psi = -\xi + \frac{kT}{q} \ln\left(\frac{q_{\mathrm{m}}}{q_0}\right) \tag{2.19}$$

At the source, we know that the quasi–Fermi potential is equal to the fermi potential ψ_{f} plus the source voltage v_{s}. Rewriting Equations 2.18 and 2.19 at the source yields:

$$Q_{\mathrm{s}} = Q_0 e^{(-\Psi_{\mathrm{s}}+\Psi_{\mathrm{f}}+V_{\mathrm{s}})} \tag{2.20}$$

$$-\Psi_{\mathrm{s}} = -\Psi_{\mathrm{f}} - V_{\mathrm{s}} + \ln\left(\frac{Q_{\mathrm{s}}}{Q_0}\right) \tag{2.21}$$

where, in natural units, Q_0 is the effective charge, Ψ_{s} is the surface potential at the source end of the channel, Ψ_{f} is the fermi potential, and V_{s} is the source voltage. We thus can compute the surface potential for any mobile-charge density from Equation 2.21. Given the surface potential, the gate voltage can be determined from Equation 2.5, as described.

2.5 Saturation

For channel lengths where the short-channel effects are important but where the limiting behavior has not been reached, we can proceed as follows. For a transistor in saturation, the boundary condition on the charge at the drain is (in natural units) $Q_{\mathrm{d}} \approx I$, and Equation 2.16 becomes a simple quadratic form in Q_s:

$$2I_{\mathrm{sat}}L + 1 = (Q_s + 1 - I_{\mathrm{sat}})^2 \tag{2.22}$$

where I_{sat} is the saturation current.

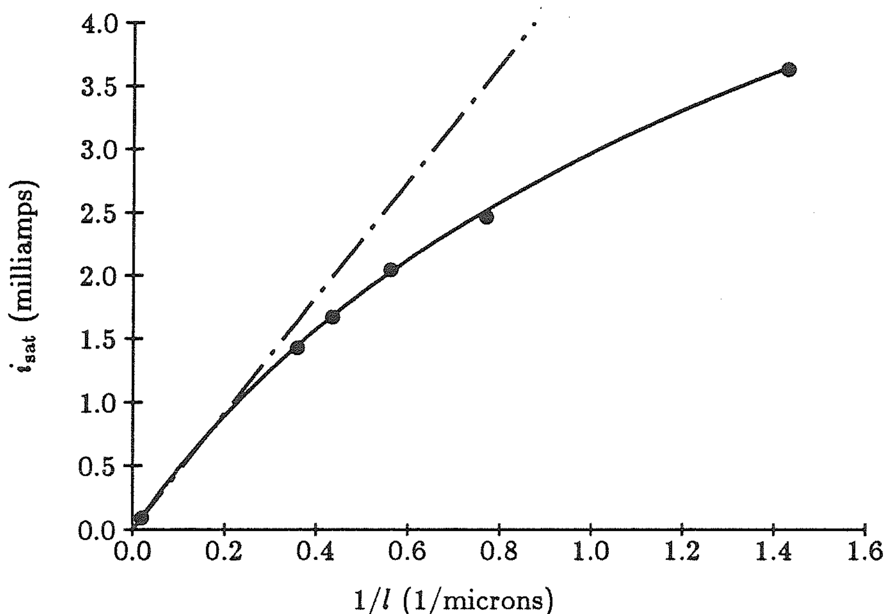

Figure 2.4: Saturation current above threshold versus $1/l$. The solid line represents Equation 2.23; dots indicate data from experimental devices. The dashed line shows the long-channel approximation, in which saturation current is inversely proportional to channel length. In a transistor made with current technology (2-micron feature size), the curve deviates substantially from the long-channel behavior.

The saturation currents of a number of transistors of different lengths are plotted as a

function of $1/l$ in Figure 2.4. The dashed line is the long-channel dependence, neglecting velocity saturation. Clearly, devices with channel lengths of the order of 1 micron show large deviations from the long-channel limit. We can solve Equation 2.22 explicitly for I_{sat}:

$$\begin{aligned} I_{\text{sat}} &= Q_s + (L+1)\left(1 - \sqrt{1 + 2Q_s \frac{L}{(L+1)^2}}\right) \\ &\approx Q_s + (L+1)\left(1 - \sqrt{1 + \frac{2Q_s}{L}}\right) \end{aligned} \tag{2.23}$$

The solid line in Figure 2.4 represents Equation 2.23, evaluated with the parameters mentioned in the introduction to this chapter.

2.6 Drain Boundary Condition

For sufficiently low drain voltages, the mobile-charge density q_{d} at the drain is no longer moving at saturated velocity. We must determine the drain boundary condition for Equation 2.17 to evaluate Q_{d} as a function of V_{d}. Writing Equation 2.19 for the source and drain in natural units and subtracting the equations yields a relation between the surface potentials at source and drain (Ψ_s and Ψ_{d}), the mobile-carrier densities at source and drain (Q_s and Q_{d}), and the imrefs at source and drain ($\Xi = \Psi_{\text{f}} + V_s$ and $\Xi = \Psi_{\text{f}} + \Xi_{\text{d}}$):

$$\Psi_{\text{d}} - \Psi_s + V_s - \Xi_{\text{d}} = -\ln\left(\frac{Q_{\text{d}}}{Q_s}\right) \tag{2.24}$$

We will derive the drain boundary condition by the following argument. Let the density of states in the drain be N_{d} and that at the drain end of the channel be N_{c}. The probability P_{cd} of a carrier in the channel making a transition to a state in the drain is just the probability P_{c} of the state in the channel being occupied multiplied by the probability $1 - P_{\text{d}}$ that the corresponding state in the drain is unoccupied. A similar argument produces the

probability that a carrier in the drain makes a transition back into the channel.

$$P_{\mathrm{cd}} = N_{\mathrm{c}} P_{\mathrm{c}} N_{\mathrm{d}} (1 - P_{\mathrm{d}})$$

and

$$P_{\mathrm{dc}} = N_{\mathrm{d}} P_{\mathrm{d}} N_{\mathrm{c}} (1 - P_{\mathrm{c}})$$

with N_{c} and N_{d} equal to the density of states in the channel and drain. The net drain current is proportional to the difference of these two probabilities:

$$\frac{I}{K} = P_{\mathrm{net}} = P_{\mathrm{cd}} - P_{\mathrm{dc}} = N_{\mathrm{c}} N_{\mathrm{d}} (P_{\mathrm{c}} - P_{\mathrm{d}}) \tag{2.25}$$

The actual value of the current is known only to within the constant K.

Writing P_{c} and P_{d} in terms of the imrefs as given in Equation 2.19, Equation 2.25 becomes (using natural units)

$$I = K Q_{\mathrm{d}} \left(1 - e^{-\Xi_{\mathrm{d}} + V_{\mathrm{d}}}\right) \tag{2.26}$$

We notice that we can evaluate the constant K by considering operation at large drain voltages (V_{d} much greater than Ξ_{d}). This condition corresponds to saturation. Using the saturation boundary condition that $Q_{\mathrm{d}} \approx I$ in Equation 2.26 yields $K = 1$. We therefore can write Equation 2.26 as

$$\Xi_{\mathrm{d}} = -\ln\left(1 - \frac{I}{Q_{\mathrm{d}}}\right) + V_{\mathrm{d}} \tag{2.27}$$

Substituting Equation 2.27 into Equation 2.24, we arrive at the final form of the relation among carrier density, current, and drain voltage:

$$V_{\mathrm{d}} - V_{\mathrm{s}} = \Psi_{\mathrm{d}} - \Psi_{\mathrm{s}} - \ln\frac{Q_{\mathrm{s}}}{Q_{\mathrm{d}}} + \ln\left(1 - \frac{I}{Q_{\mathrm{d}}}\right) \tag{2.28}$$

The $\Psi_{\mathrm{d}} - \Psi_{\mathrm{s}}$ term is just the difference in the imrefs at the two ends of the channel. The $\ln(1 - I/Q_{\mathrm{d}})$ term is due to the "drain drop"; that is, to the difference between Ξ_{d} and V_{d}.

Because Equation 2.5 is valid for any surface potential, we can use Equations 2.1, 2.2, and 2.5 to solve for $\psi_{\rm d}$, yielding

$$\psi_{\rm d} = \left(\sqrt{\frac{\epsilon_{\rm s}\rho}{2C_{\rm ox}^2} + v_{\rm g} - \frac{q_{\rm d}}{C_{\rm ox}}} - \sqrt{\frac{\epsilon_{\rm s}\rho}{2C_{\rm ox}^2}}\right)^2 \tag{2.29}$$

2.7 Model Evaluation

To generate model curves for comparison with experimental data, we employ the following algorithm. We use voltages $V_{\rm g}$ and $V_{\rm s}$ with Equations 2.21 and 2.5 to determine the source charge $Q_{\rm s}$. Then, we use the drain voltage $V_{\rm d}$ with Equations 2.17 and 2.29 to determine the drain charge and the drain current. Alternatively, we could pick several values of the drain charge varying between $Q_{\rm d} = Q_{\rm s}$ ($I_{\rm ds} = 0, V_{\rm ds} = 0$) and $Q_{\rm d} = I_{\rm sat}$ ($I = I_{\rm sat}$, large $V_{\rm ds}$), to sweep out the drain characteristic.

2.8 Experimental Results

We compared the model predictions against data taken from a number of experimental devices with oxide thickness $\approx$ 125 angstroms provided by Intel Corporation. Detailed comparisons were made for devices from the same wafer, all of width 50 microns and of length ranging from 50 microns to 0.35 micron. Mobility was taken from the channel conductance of the 50-micron channel length device at very low drain–source voltage. Oxide thickness was obtained from the capacitance of a large MOS dot. We found the substrate doping by plotting the threshold voltage versus the square root of the source–substrate reverse bias, as shown in Figure 2.5. The fixed charge at the surface $q_{\rm ss}$ was computed directly from the threshold voltage once the substrate doping was known. This charge includes any threshold-adjustment implant dose. The saturated velocity of electrons in

silicon was taken from the literature [Jacoboni 77].

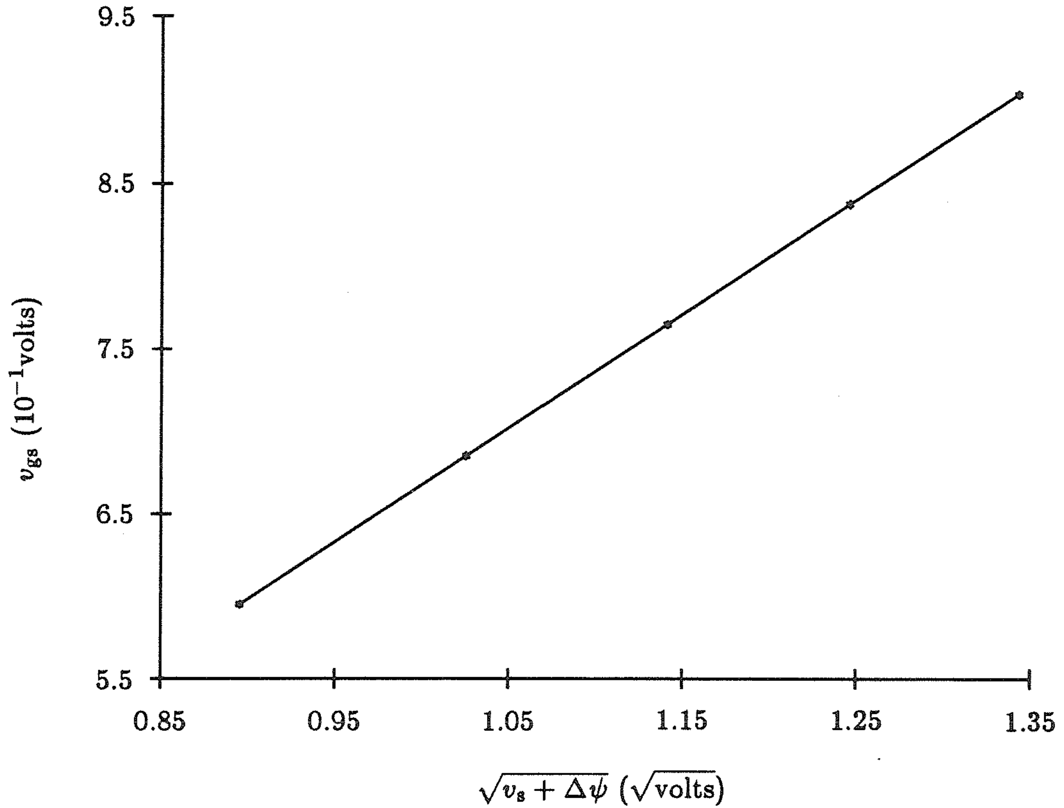

Figure 2.5: Determination of substrate doping. Gate–source voltage is plotted versus $\sqrt{v_s + \Delta\psi}$, for fixed current level. The slope of the line is proportional to the doping density. Using a value of 125 angstroms for the oxide thickness, we determined the density N_A to be $10^{17}/\text{cm}^3$.

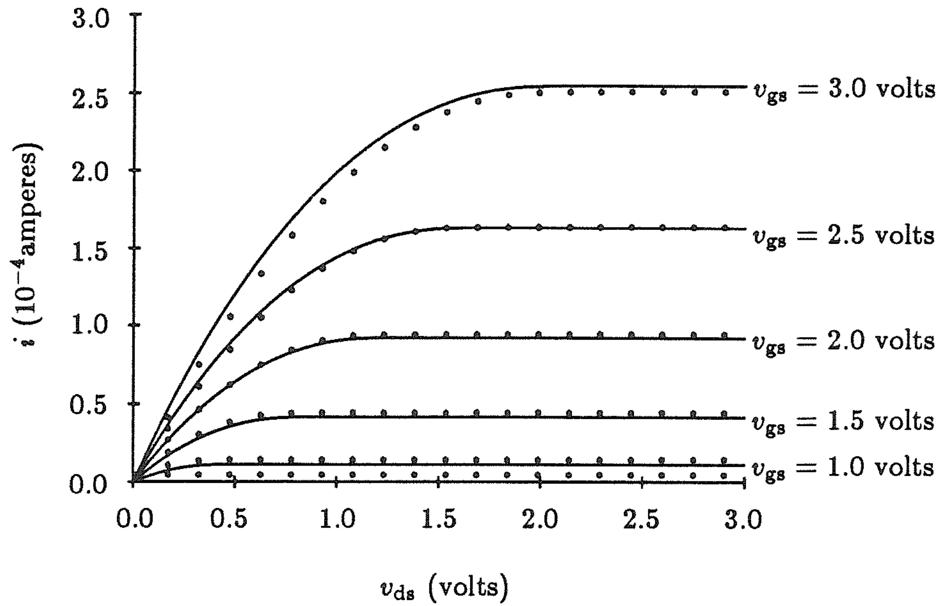

Figure 2.6: Drain current versus v_{ds} for fixed v_{gs} for a 50-micron channel length transistor. Solid lines are the values predicted by the model. Dots indicate experimental data.

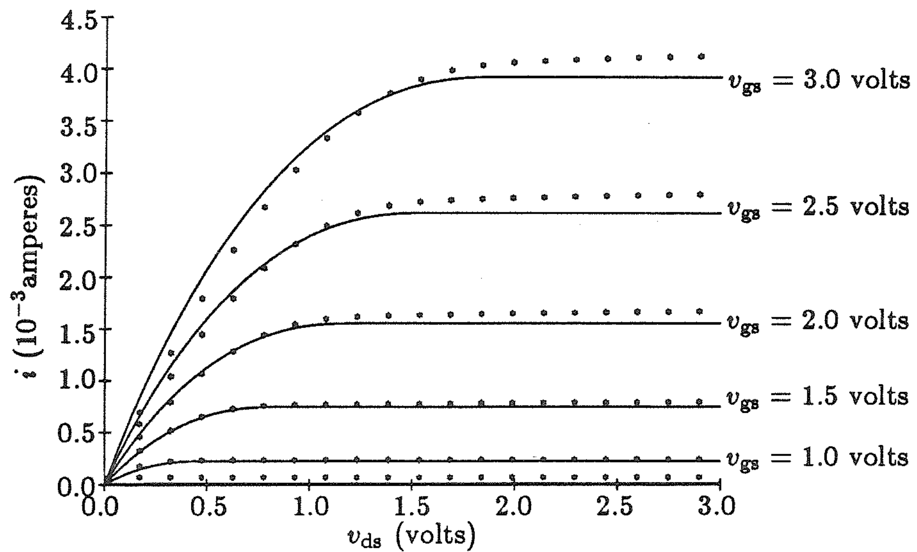

Figure 2.7: Drain current versus v_{ds} for fixed v_{gs} for a 2.8-micron channel length transistor. Solid lines are the values predicted by the model. Dots indicate experimental data.

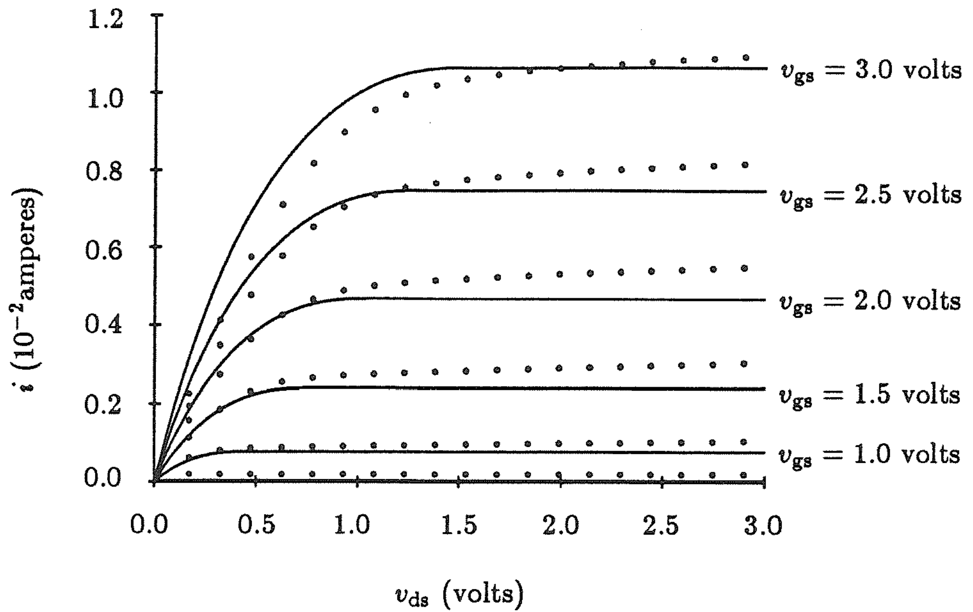

Figure 2.8: Drain current versus v_{ds} for fixed v_{gs} for a 0.7-micron channel length transistor. Solid lines are the values predicted by the model. Dots indicate experimental data.

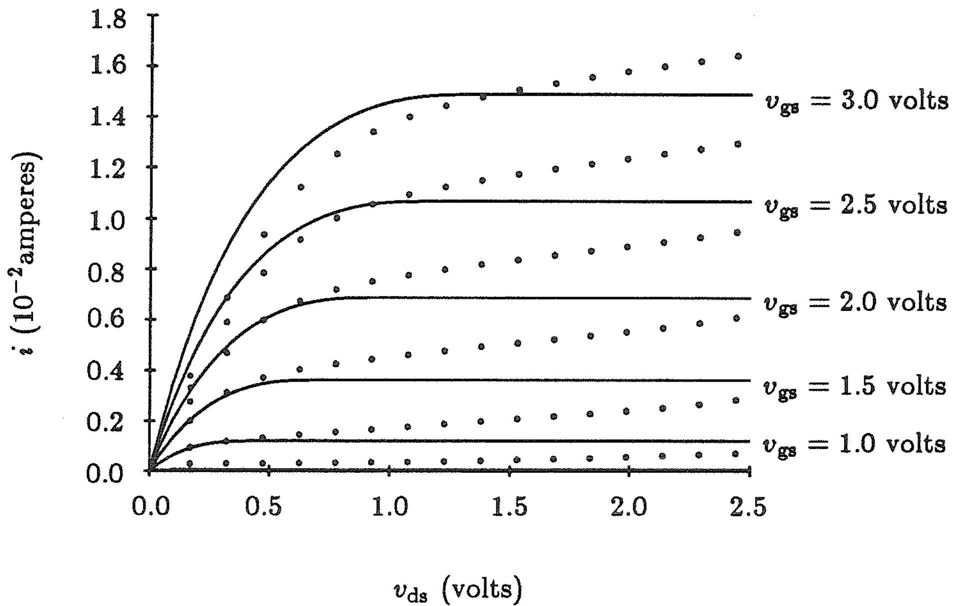

Figure 2.9: Drain current versus v_{ds} for fixed v_{gs} for a 0.35-micron channel length transistor. Solid lines are the values predicted by the model. Dots indicate experimental data.

Experimental drain families and conductance curves were generated using the Keithley 619 programmable electrometer and Keithley 230 voltage sources under computer control using the Hewlett Packard HP9836 computer. Experiments to determine capacitance used

the Boonton capacitance meter. Software for control and graphics was written in Pascal under the UCSD operating system. All parameters were either derived from the process by direct measurement, or were physical dimensions of the layout of a particular device.

The results of comparing the zero-order model for 50-micron, 2.8-micron, 0.7-micron, and 0.35-micron channel length devices are shown in Figures 2.6, 2.7, 2.8, and 2.9. Drain curves are shown for fixed v_{gs} ranging from 0 to 3 volts. The theoretical curves use the same set of parameters in all cases.

2.9 Summary

We have formulated a new charge-controlled model for the DC current in the MOS transistor. The model we developed is valid above or below threshold, in saturation or in the ohmic range, and is continuous over all these regions of operation. We have defined a set of natural units for the physical quantities involved in transistor operation. These units have allowed us to express the equations describing transistor operation in dimensionless form, and thus to simplify the analysis. We have seen that the functional form of the equations describing above-threshold operation are far from the long-channel limit, even with 1-micron channel lengths. Subthreshold devices, however, continue to conform to the simple model, even at very small dimensions, as we shall see in Chapter 4. These conclusions have important consequences for the design of digital, as well as of analog, systems.

Chapter 3

First-Order Corrections to the Basic Model

There are several first-order effects that need to be added to the model for detailed comparison with submicron devices. The consequences of these first-order effects are visible in Figure 2.9. We have not yet considered the slope of the drain curves in saturation. This dependence of saturation current on drain voltage is due to the change in channel length with drain bias in Equation 2.17. The drain-voltage modulation of channel length is important in today's devices, and becomes crucial as devices scale to submicron lengths. In addition, the conductance of an actual device near the origin is less than that predicted by the simple model, and the discrepancy is larger for larger gate voltages. This behavior is in part due to the dependence of mobility on vertical electric field perpendicular to the direction of current flow, leading to mobility degradation with increasing v_{gs}. Finally, the experimental saturation currents are less dependent on gate voltage than they should be, even accounting for mobility degradation. This discrepancy is caused by the parasitic

resistances of the source and drain. Although resistance is not a device property in the purest sense, it is a necessary and unavoidable byproduct of any real fabrication process. In addition to these phenomena, we will add two other first-order effects to the model. We account for the varying width of the depletion layer in the channel, which corresponds to a decrease in the capacitance $C = C_{\mathrm{b}} + C_{\mathrm{ox}}$, and for the variation of the vertical electric field $\mathcal{E}_{\mathrm{g}}$ along the channel.

The several effects mentioned, along with the first-order corrections to the model itself, are of roughly the same magnitude. Some, such as channel-length modulation, increase the current. Others, such as mobility variation and internal resistances, decrease the current. Our *modis operandi* is to find a particular regime of operation in which one of the effects is dominant, and to evaluate the effect there.

The mobility-degradation effect can be seen quite clearly in a 50-micron channel length transistor in which velocity saturation and series resistance are negligible. If the mobility were constant, the conductance plot shown in Figure 3.1 would be a straight line intersecting at threshold, and the derivative would be constant above threshold. The slope of the derivative curve directly measures the mobility variation with the gate electric field. The conductance of a short-channel transistor shows the effects of the parasitic resistance, which are masked in the smaller currents of a longer channel transistor. Channel-length modulation is best observed in a short-channel device in subthreshold, where there are no mobile-charge carriers to reduce the effect. Current flows by pure diffusion, without velocity saturation to confuse the issue. Electric-field lines from the drain can terminate on mobile electrons as well as on the negative fixed charges in the depletion layer. The density of mobile charge increases at higher gate voltages. We therefore expect a smaller change

in channel length at high gate voltages than in subthreshold. To account for this effect, we include the mobile charge in the solution to Poisson's equation for the depletion-layer length.

These effects have been well-studied by other researchers and two and three dimensional analyses have been done, but the results are not computationally efficient to simulate circuits. Because we want a model that is simple enough for circuit simulation, we adopt simple and approximate descriptions of these effects but we stay as close to the physics of the device as possible. The effects mentioned were incorporated into a unified model that was used to generate theoretical curves for measured properties of transistors over a wide range of lengths. We found excellent agreement between theory and experiment. The experimental data we show in this chapter were taken from the devices described in Chapter 2. We present plots of conductance and drain curves both above and below threshold.

3.1 Mobility Variation

The vertical electric field acting perpendicular to the channel leads to mobility degradation with increasing v_{gs} [Garverick 87]. Intuitively, the field from the gate attracts electrons in the channel toward the oxide interface. Conditions at this interface are not as ideal as they are in the silicon crystal, and an electron is more likely to be scattered if it spends more time there. This additional scattering decreases the electron's mean free time, and hence reduces its mobility. This effect can be seen quite clearly in a 50-micron channel length transistor, in which velocity saturation and series resistance are negligible. A plot of the low-field, channel conductance versus gate voltage is shown in Figure 3.1. Also shown in Figure 3.1 is the derivative of the conductance curve $\frac{\partial g_d}{\partial v_{gs}}$. If the mobility were constant, the conductance

plot would be a straight line with x intercept at threshold, and the derivative would be constant above threshold. The slope of the derivative curve is a direct measurement of the mobility variation with the gate electric field, $\mathcal{E}_g$. By adding another term to the scattering model used to derive the velocity saturation, we obtain a form for μ:

$$\mu_{\text{eff}} = \frac{\mu}{1 + \frac{\mathcal{E}_g}{\mathcal{E}_f}}$$

We add mobility variation to the model by replacing μ in Equation 2.10 with μ_{eff}. We calculate the vertical electric field $\mathcal{E}_g$ from the relation

$$\mathcal{E}_g \approx \frac{q_{\text{tot}}}{\epsilon_s}$$

where q_{tot} is the total charge as used in Equation 2.4. For the curve shown in Figure 3.1, the gate electric field was assumed to be constant, and to be equal to the value at the source. Variation of the gate electric field will be discussed in Section 3.4. The parameters μ and $\mathcal{E}_f$ can be evaluated directly from the data in Figure 3.1. We found the value of the mobility to be 490 $\text{cm}^2/\text{v} - \text{s}$. The results of this refinement to the model and of the uncorrected model are plotted, with the original experimental data, in the figure.

3.2 Channel-Length Modulation

We investigate drain-voltage modulation of channel length by considering a short-channel device in subthreshold, where there are no mobile-charge carriers to reduce the effect. Current flows by pure diffusion, so there is no velocity saturation in the channel proper. We know the surface potential at the source edge of the channel, and know that it is constant to the very edge of the drain depletion region. Hence, a measurement of the slope of saturation current amounts to a direct measurement of the change in channel length with drain

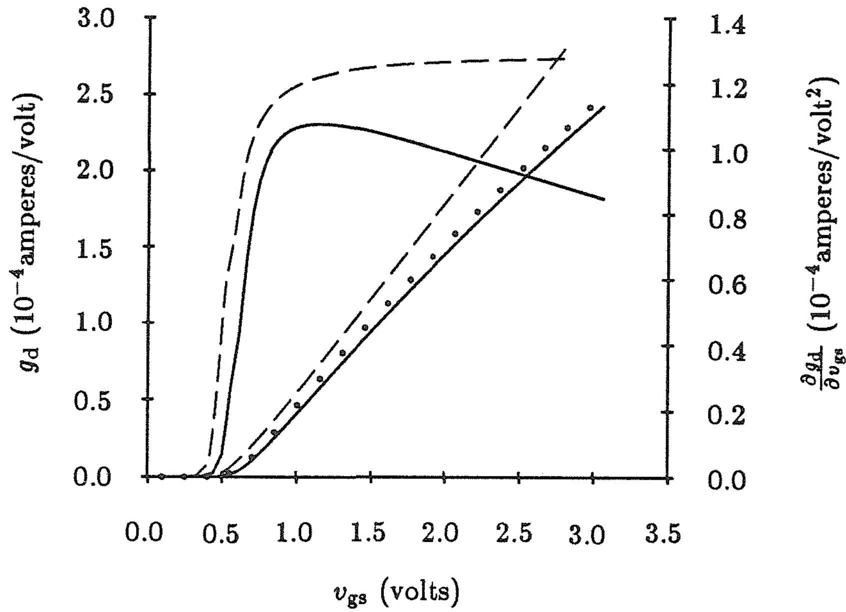

Figure 3.1: Conductance and slope of conductance for a 50-micron channel length transistor. Dots indicate experimental data; solid lines indicate corrected model curves; dotted lines indicate the model without mobility variation.

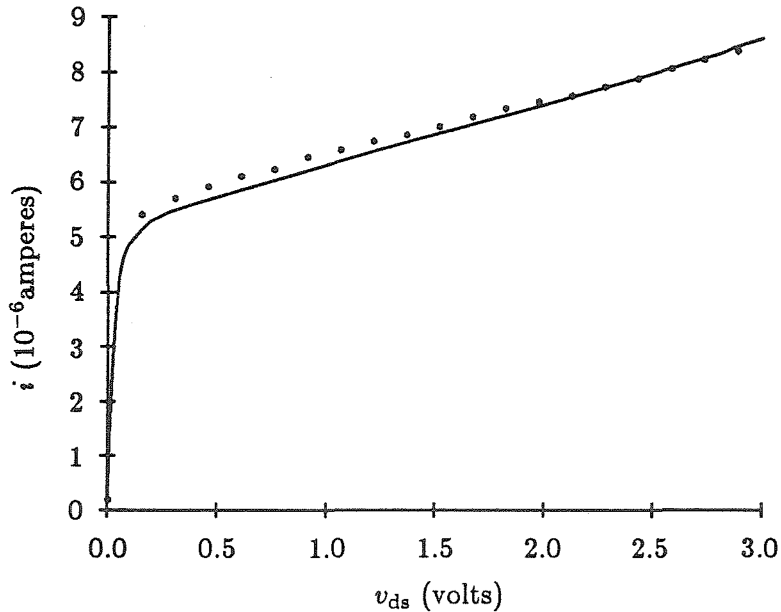

Figure 3.2: Drain current versus v_{ds} for v_{gs} = 0.5 volts for a 0.7-micron channel length transistor. Dots indicate experimental data; the solid line indicates the corrected model curve.

voltage:

$$\frac{\partial I_{\text{sat}}}{\partial V_{\text{d}}} = \frac{1}{L}\frac{\partial L}{\partial V_{\text{d}}} I_{\text{sat}} \tag{3.1}$$

where I_{sat} is the saturation current in the absence of the channel-length–modulation effect. An example of the direct manifestation of the channel-length–modulation effect can be seen in the subthreshold current of a 0.7 micron channel length device in Figure 3.2.

Many models have been proposed for channel length modulation [El-Mansy 77b], [Frohman-Bentchkowsky 69], [Reddi 65], [Popa 72]. To incorporate channel-length modulation into our model, we calculate the effective channel length L_{eff} by subtracting the lengths of the depletion layers at source and drain, ΔL_{s} and ΔL_{d}, from the physical channel length:

$$L_{\text{eff}} = L - \Delta L_{\text{s}} - \Delta L_{\text{d}}$$

The actual boundary around the depletion layer near the drain is complicated, involving not only the fixed charges in the depletion layer, but also the mobile charge in the channel in three dimensions [Baum 70,Frohman-Bentchkowsky 69]. Assuming a planar junction at the depletion layer, and using the surface potential at the drain end of the channel, the horizontal electric field at the drain end of the channel $\mathcal{E}_{\text{d}}$, and the drain voltage as the boundary conditions in the solution of Gauss's law yields

$$\Delta L_{\text{d}} \approx \frac{-\mathcal{E}_{\text{d}} + \sqrt{\mathcal{E}_{\text{d}}^2 + 8\frac{\rho}{\epsilon_s}(v_{\text{d}} + \psi_{\text{f}} - \psi_{\text{d}})}}{2\frac{\rho}{\epsilon_s}} \tag{3.2}$$

A similar equation holds at the source. A simple cylindrical approximation to the two-dimensional solution around the drain "corner" in subthreshold gives a value for ΔL_{d} that is $\sqrt{2}$ times as large as that predicted for a planar junction. We expect a value between 1 and $\sqrt{2}$ to occur in real devices. The factor derived from Figure 3.2 is 1.4. This value is a parameter of the process and is not a function of the dimensions of a particular device.

To add the channel-length–modulation effect to our model, we replaced the value of L in Equation 2.17 by L_{eff}. The result of this correction is the model curve shown in Figure 3.2.

Electric-field lines from the drain can terminate on mobile electrons as well as on the negative fixed charges in the depletion layer [Frohman-Bentchkowsky 69,Popa 72]. The density of mobile charge increases at higher gate voltages. We therefore expect a smaller change in channel length at high gate voltages than in subthreshold. A graphic illustration of the effect of mobile charge on the channel-length modulation can be seen in Figure 3.3, which is a drain curve for higher gate voltage (3.0 volts, well above threshold). The theoretical curve shown is that predicted by the channel-length–modulation effect, ignoring the contribution of mobile charge.

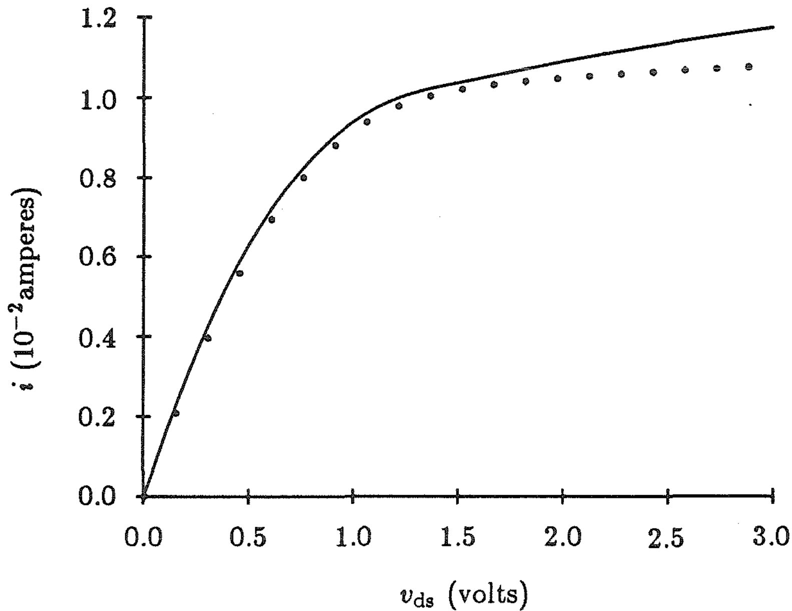

Figure 3.3: Drain current versus v_{ds} for v_{gs} = 3 volts for a 0.7-micron channel length transistor. Dots indicate experimental data; the solid line is the model curve without the mobile-charge effect.

We define an effective volume charge density ρ_m at the source and at the drain by normalizing the mobile charge per unit area by the width of the depletion layer normal to

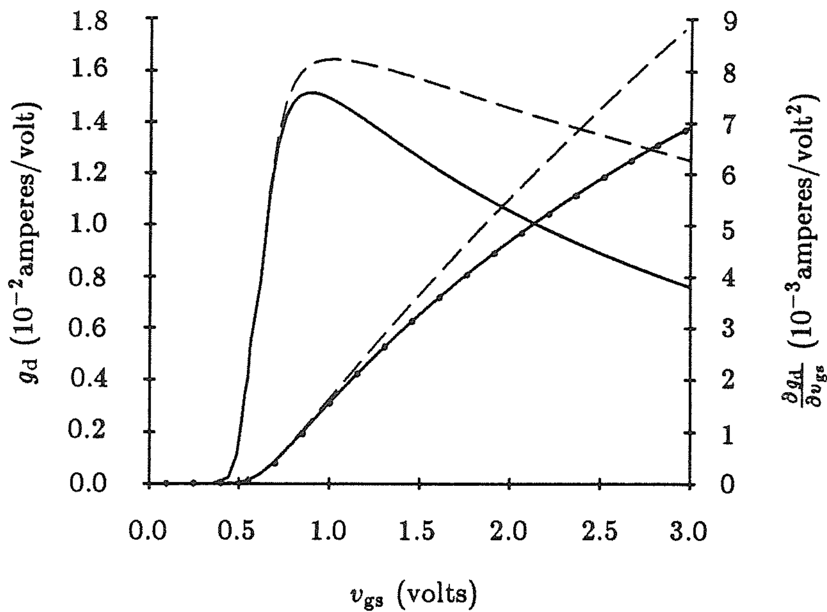

Figure 3.4: Conductance and slope of conductance for a 0.7-micron channel length device. Dots indicate experimental data; solid lines indicate corrected model curves; dashed lines indicate the model without resistance.

the surface, as calculated from Equation 2.1. So ρ now becomes

$$\rho_{\text{eff}} = \rho_{\text{m}} + \rho$$

We substitute the value of ρ_{eff} in place of ρ in the calculation of the depletion-layer lengths ΔL_s and ΔL_{d}.

Since the mobile charge increases the total charge in the electrostatic equations, it decreases the extent to which the saturation current changes with drain voltage. We call this decrease in the channel-length–modulation effect with mobile charge density the *late effect.* It is clear that the late effect makes the device a better current source and hence, in some sense, more ideal. However, the effect decreases the current-driving capability of the device. The corrected model curve is shown as part of the drain characteristics in Figure 3.8.

3.3 Resistance

Although they are not strictly part of a device model, source and drain resistances must be included as part of the model so that we can compare it with any real measurements. Direct measurements of the sheet resistance of the diffusion layer gave 68 ohms per square. In the test devices, the distance from the metal contact cuts to the edge of the gate was measured to be 5 microns. Since the devices were 50 microns wide, the source and drain resistances were 7 ohms. For a short, wide device 50 microns by 0.7 microns, we measured drain currents up to 10 milliamps. The voltage drop across the 7 ohm source resistance was 70 millivolts, corresponding to an error current of about 15 percent of the spacing between adjacent drain curves. To compute accurate device characteristics, we use a simple model for the parasitic resistance and add lumped source and drain resistors to the model. We simulate this equivalent circuit by iteration, alternately updating V_s and V_d and evaluating the model. Figure 3.4 shows the experimental data and the model results with and without resistance. The model uses the same value for the resistivity for all the devices taken from the same wafer.

3.4 The Upgraded Model

In addition to the phenomena described above, we added two other first-order effects to the model. The width of the depletion layer in the channel increases with distance towards the drain. This result corresponds to a decrease in the channel capacitance C in Equation 2.8. To add this effect to the model, we interpolate the capacitance linearly between the values at source and drain:

$$C = C_s + \beta z$$

where $\beta = \frac{(C_{\rm d}-C_{\rm s})}{l}$, and we know the value for $C_{\rm d}$ from Equation 2.7. Using this technique is much more accurate than is using a first-order Taylor series. We integrate the equation for current (Equation 2.17) to get a new model equation.

The vertical electric field $\mathcal{E}_{\rm g}$ also varies along the channel, reaching a minimum at the drain, affecting the mobility variation in Equation 3.1 at high drain voltages. We use

$$\mathcal{E}_{\rm g} = \mathcal{E}_{\rm gs} + \alpha z$$

where $\alpha = \frac{(\mathcal{E}_{\rm gd}-\mathcal{E}_{\rm gs})}{l}$, and $\mathcal{E}_{\rm gd}$ is calculated from Equation 3.1 evaluated at the drain. We substitute this value for the vertical electric field, and we derive a new current equation.

3.5 Experimental Results

We incorporated the effects mentioned into a unified model, which we used to generate theoretical curves for measured properties of transistors of a wide range of lengths. The family of drain measurements for 50-micron, 3-micron, 2-micron, 0.7-micron, 0.6-micron and 0.35-micron channel length devices is shown in Figures 3.5, 3.6, 3.7, 3.8, 3.9, and 3.10 respectively. The drain characteristics are shown for gate voltages of 0 to 3 Volts in 0.5 Volt steps. For the 50-micron, 2.8-micron, and 0.6-micron channel length devices, we show a subthreshold drain curve at $v_{\rm g} = 0.5$ Volts in Figures 3.11, 3.12, and 3.13. The subthreshold drain curve for the 0.7-micron channel length device was shown in Figure 3.2. We present conductance plots for the 2.8-micron and 0.6-micron channel length devices in Figures 3.14 and 3.15. The conductance plots for the 50-micron and 0.7-micron devices were shown in Figures 3.1 and 3.4. The experimental devices were described in Section 2.8. The experimental technique used to acquire the data also was described in that section.

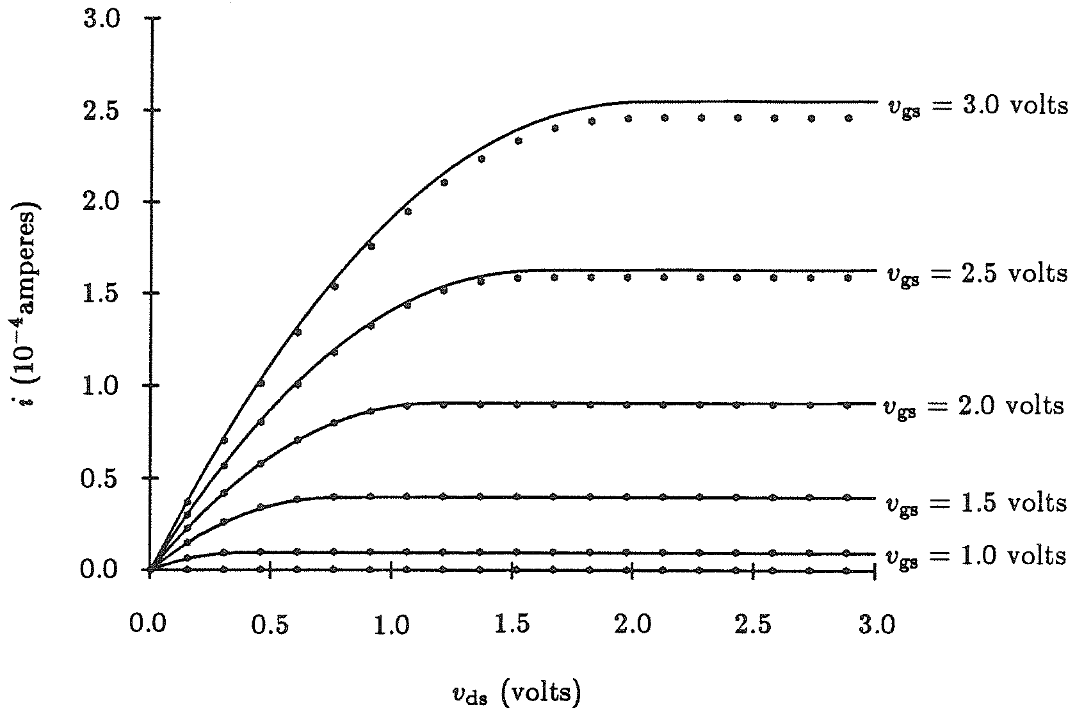

Figure 3.5: Drain current versus v_{ds} for $v_{gs} = 0 - 3$ volts for a 50-micron channel length transistor. Dots indicate experimental data; solid lines indicate model curves.

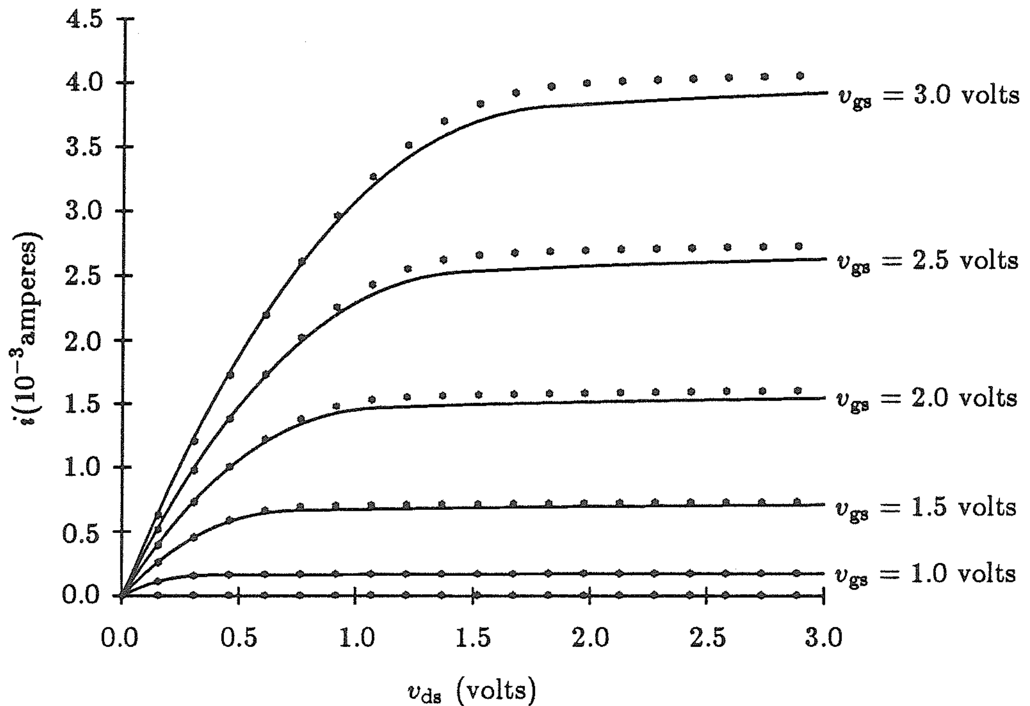

Figure 3.6: Drain current versus v_{ds} for $v_{gs} = 0 - 3$ volts for a 2.8-micron channel length transistor. Dots indicate experimental data; solid lines indicate model curves.

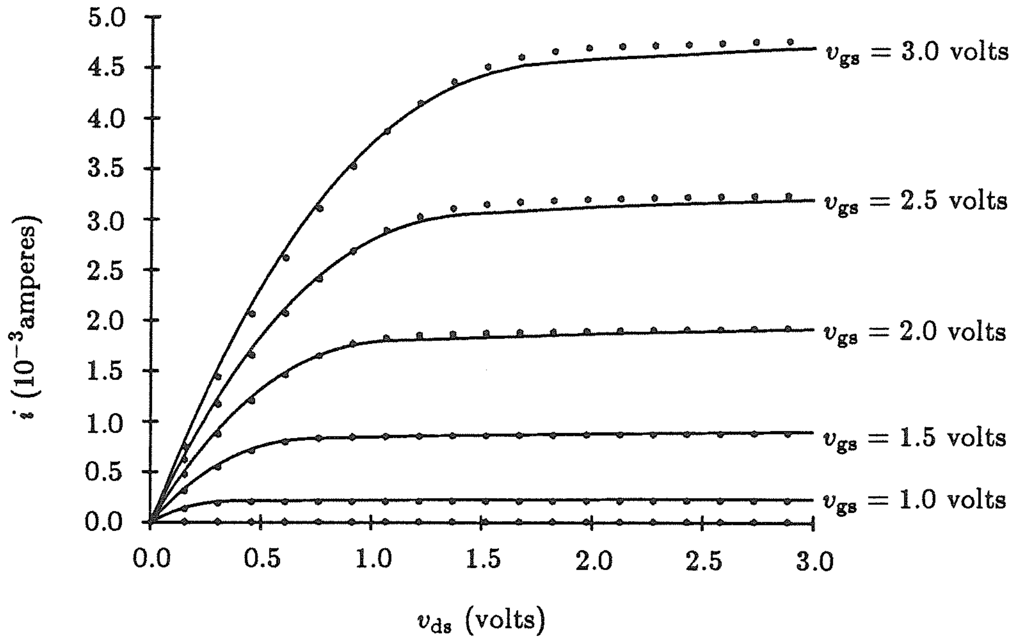

Figure 3.7: Drain current versus v_{ds} for $v_{gs} = 0 - 3$ volts for a 2-micron channel length transistor. Dots indicate experimental data; solid lines indicate model curves.

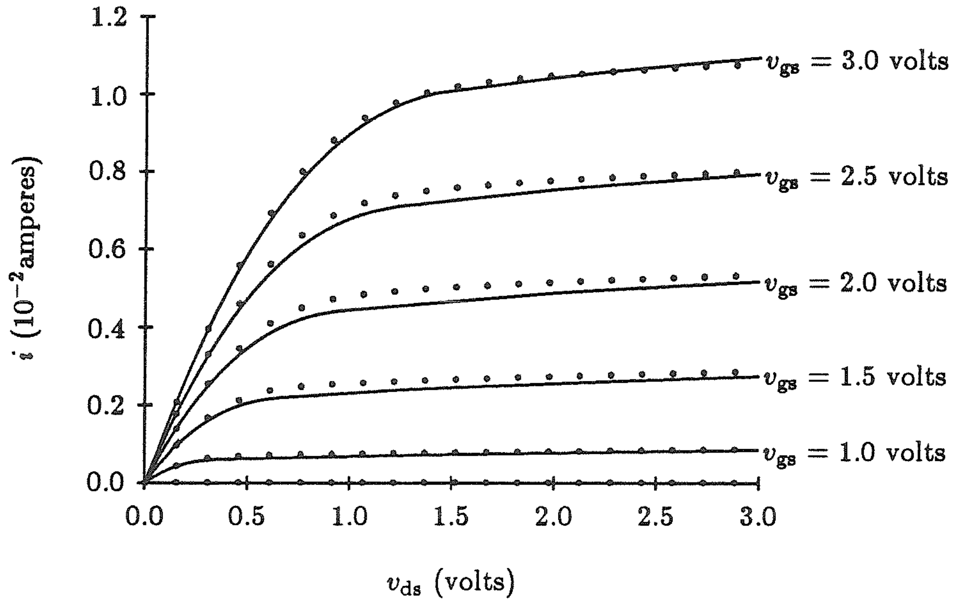

Figure 3.8: Drain current versus v_{ds} for $v_{gs} = 0 - 3$ volts for a 0.7-micron channel length transistor. Dots indicate experimental data; solid lines indicate model curves.

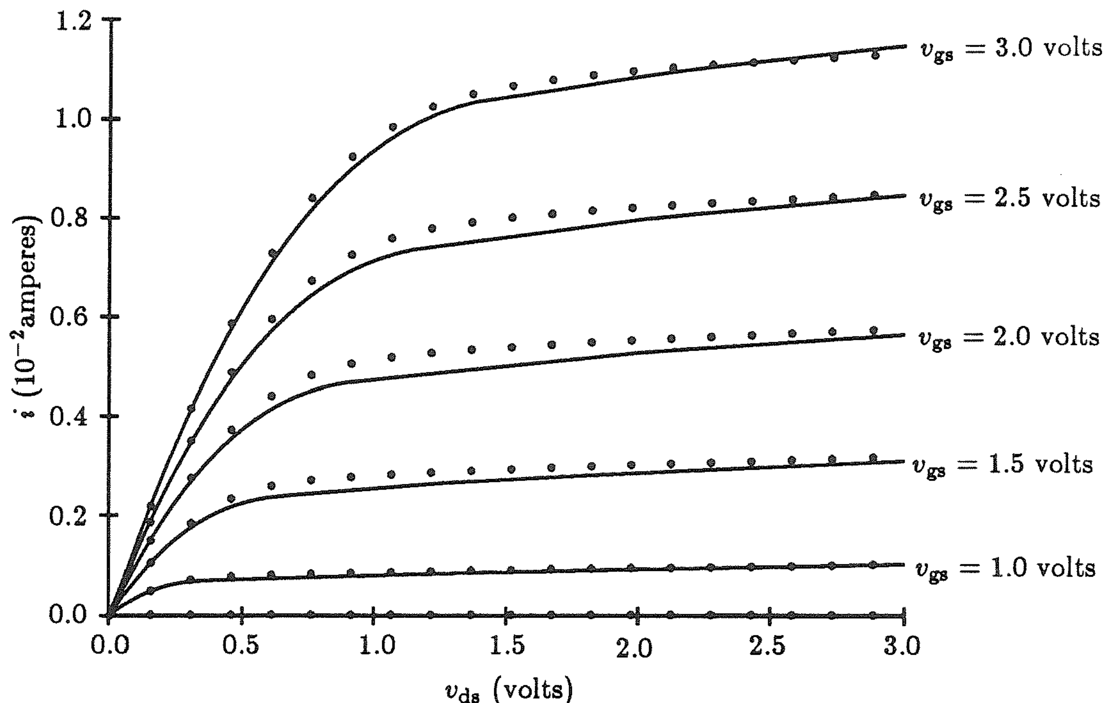

Figure 3.9: Drain current versus v_{ds} for $v_{gs} = 0 - 3$ volts for a 0.6-micron channel length transistor. Dots indicate experimental data; solid lines indicate model curves.

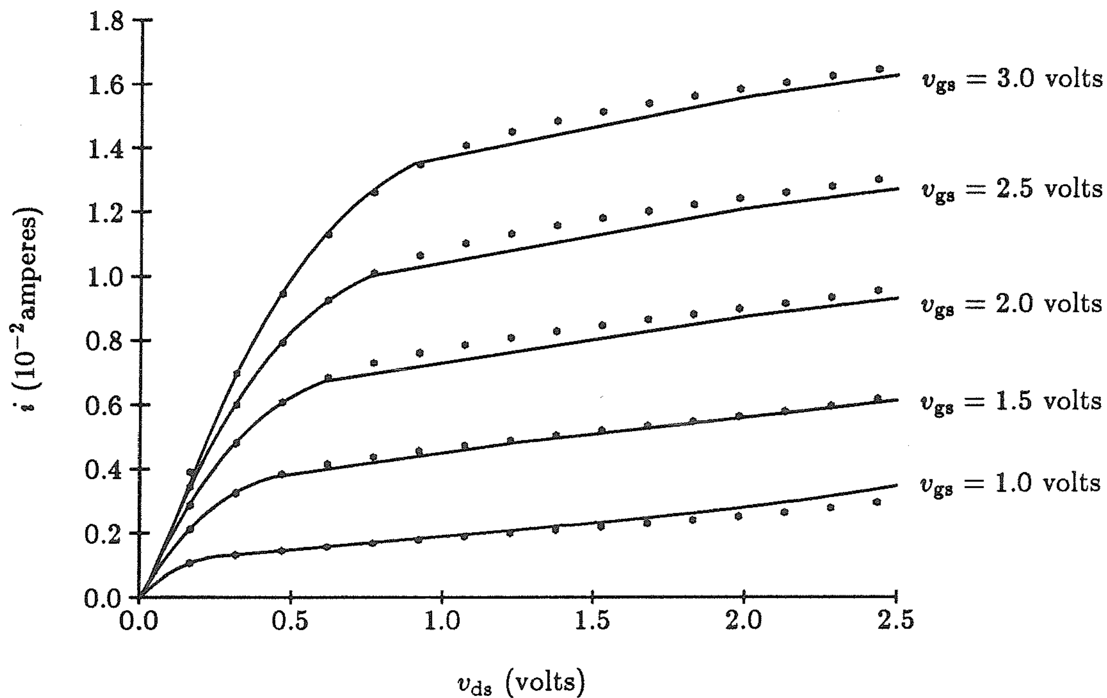

Figure 3.10: Drain current versus v_{ds} for $v_{gs} = 0 - 3$ volts for a 0.35-micron channel length transistor. Dots indicate experimental data; solid lines indicate model curves.

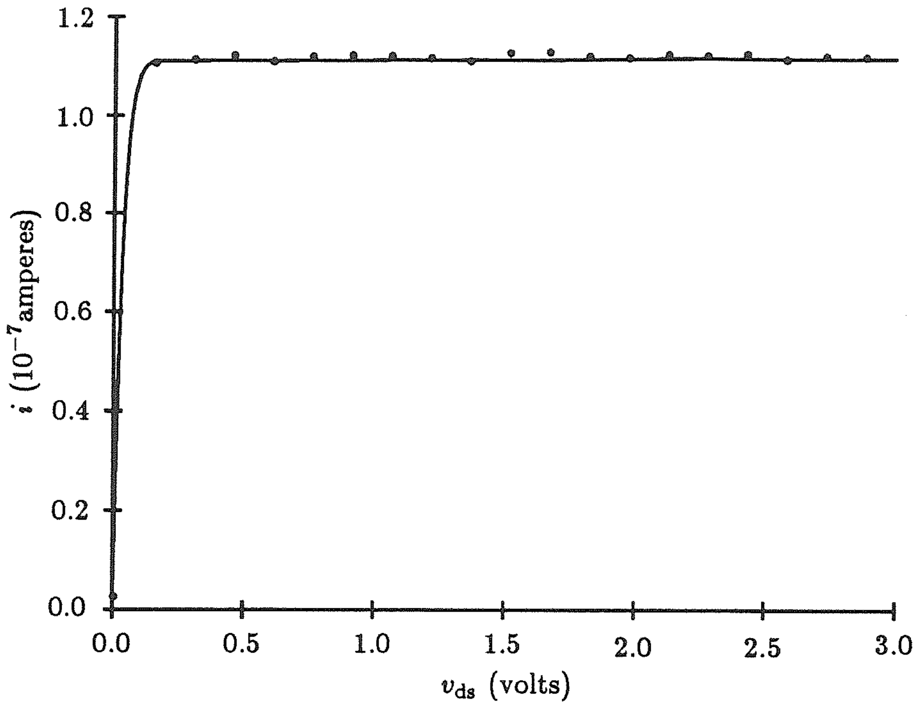

Figure 3.11: Drain current versus v_{ds} for v_{gs} = 0.5 volts for a 50-micron channel length transistor. Dots indicate experimental data; the solid line indicates the model curve.

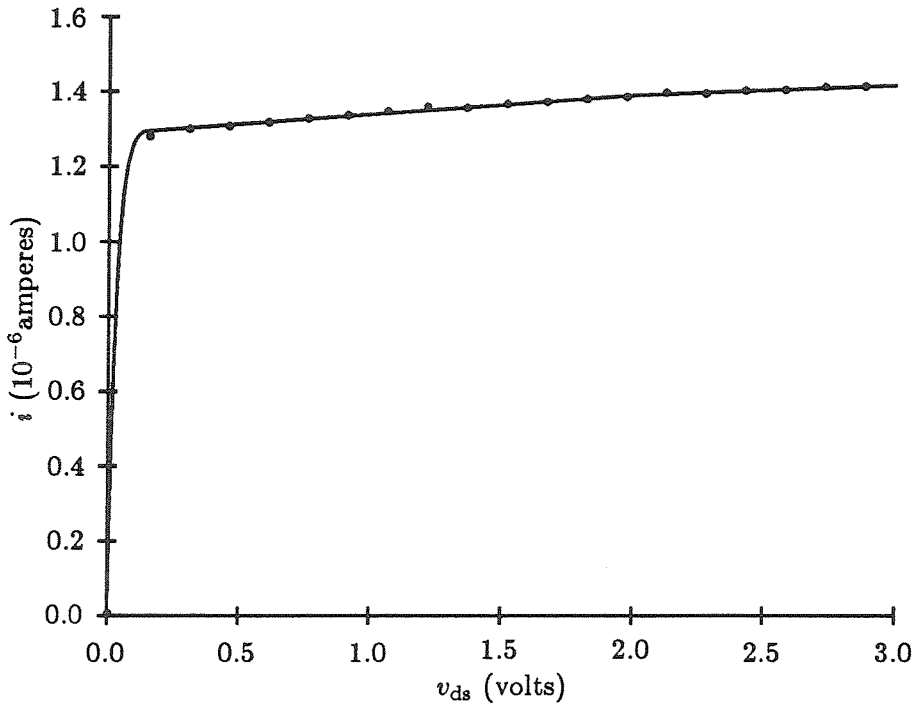

Figure 3.12: Drain current versus v_{ds} for v_{gs} = 0.5 volts for a 2.8-micron channel length transistor. Dots indicate experimental data; the solid line indicates the model curve.

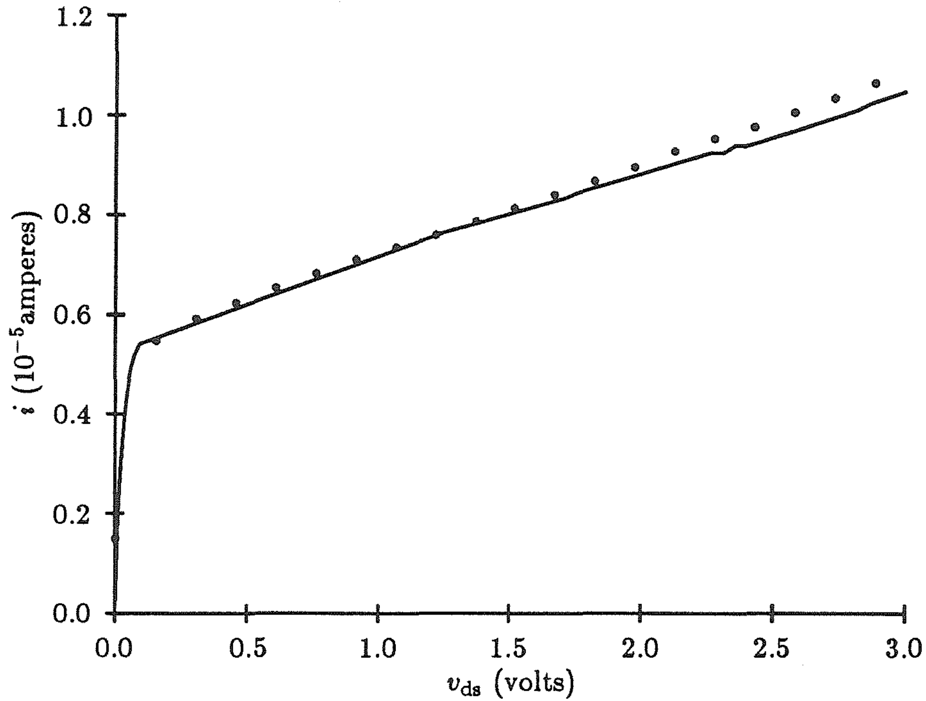

Figure 3.13: Drain current versus v_{ds} for v_{gs} = 0.5 volts for a 0.6-micron channel length transistor. Dots indicate experimental data; the solid line indicates the model curve.

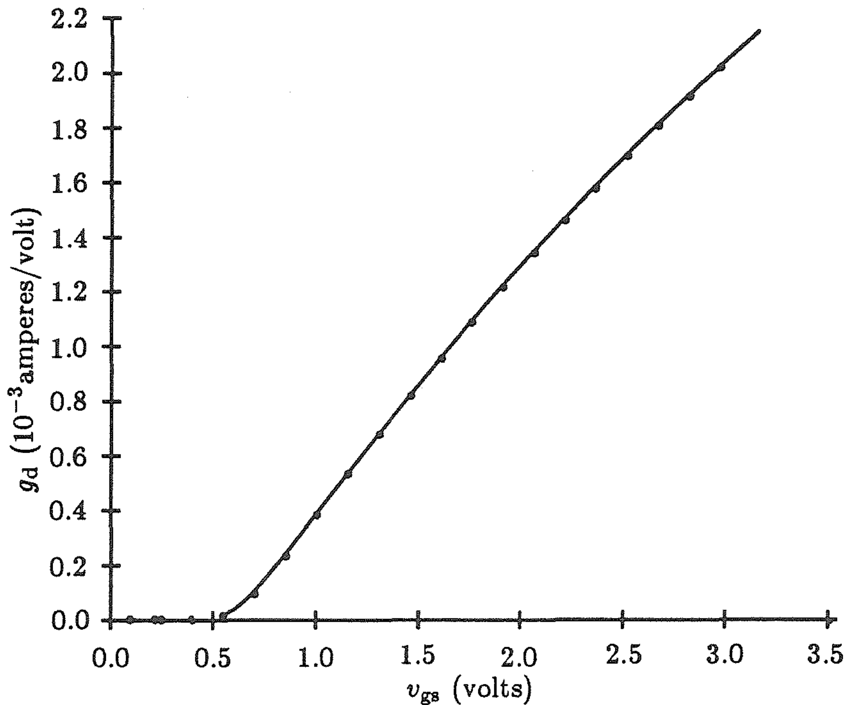

Figure 3.14: Conductance versus v_{gs} for a 2.8-micron channel length transistor. Dots indicate experimental data; the solid line indicates the model curve.

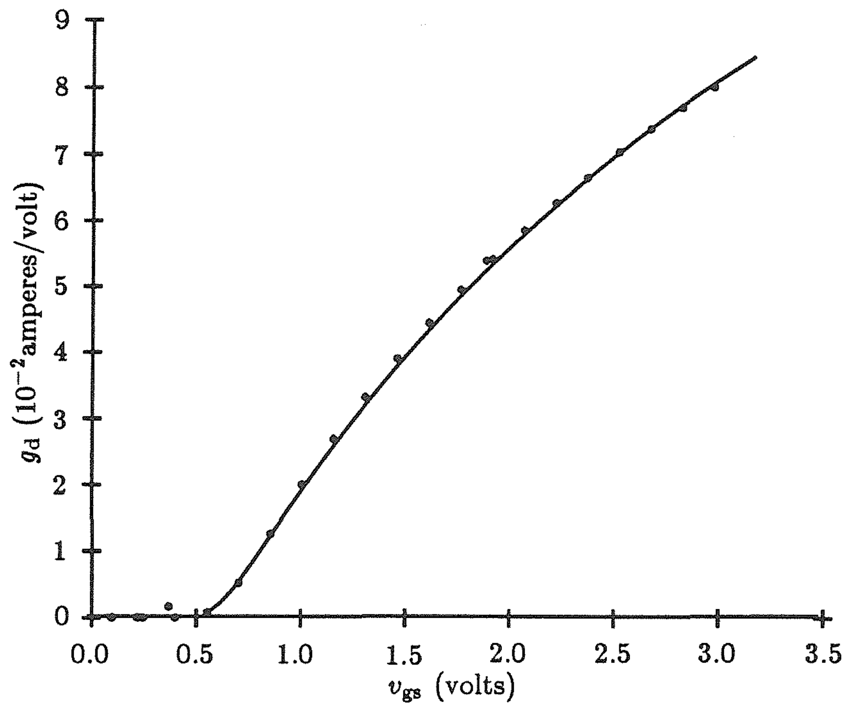

Figure 3.15: Conductance versus v_{gs} for a 0.6-micron channel length transistor. Dots indicate experimental data; the solid line indicates the model curve.

3.6 Conclusions

We can see that the model generates curves that are in excellent agreement with experimental results over a wide range of device sizes, without resorting to ad hoc parameters. All parameters either are derived from the process by direct measurement, or are physical dimensions of the layout of a particular device. These results demonstrate that a simple first-principles model with physically meaningful, measurable parameters is capable of quantitatively predicting the behavior of MOS devices.

Chapter 4

Model Behavior

In this chapter, we look in detail at the behavior of the model in the subthreshold and above-threshold regions. We compare the model to other models. We also derive simplified models useful for analog-circuit designers, similar to models developed by Vittoz [Vittoz 85]. We generalize the simple model of collector conductance, called the *Early effect*, to the drain conductance in MOS transistors. This simple approximation, used in the modeling of bipolar transistors, assumes that, for certain bias conditions, the output conductance is proportional to the collector current and is inversely proportional to a voltage V0 (the Early voltage [Early 52]), which is a constant with respect to the emitter, collector, and base voltages. As part of our simplified transistor model, we define an Early voltage for the saturated drain conductance of MOS transistors, expressing V0 as a function of gate voltage and channel length that includes the effect of velocity saturation. We compare the theoretical expressions with data from transistors with channel lengths from 0.7 to 50 microns for several different values of gate voltage. The match between theory and experiment is quite good.

4.1 The Subthreshold Model

In a subthreshold MOS transistor, the current flow i from source to drain is due to the diffusion of charge carriers. We look at the resulting model using just this current mode, and compare the results with the limiting case of the full model. We use simplified boundary conditions on the energy barrier at the drain end of the channel. We revisit the basic electrostatics of the device, because, in subthreshold, the mobile charge Q_{m} per unit area in the channel is much smaller than is the depletion charge in the substrate. This treatment allows us to understand the interaction of the surface potential of the channel with the substrate potential. We define a constant κ that expresses the effectiveness of the gate in determining the surface potential, and we derive an expression for it, and for the gate voltage required to produce a given drain current as the source voltage V_{s} is changed. We also analyze how the drain current in saturation increases with increasing drain voltage.

4.1.1 Current

Rewriting Equation 2.9 to include only the diffusion term gives

$$i = -wqD\frac{\partial q_{\mathrm{m}}}{\partial z} \tag{4.1}$$

Because no carriers are lost on their way from source to drain, q_{m} decreases linearly with distance z along the channel. Using this fact, we can rewrite Equation 4.1 in natural units to yield the subthreshold current:

$$I \approx \frac{Q_{\mathrm{s}} - Q_{\mathrm{d}}}{L} \tag{4.2}$$

In subthreshold, with Q_{s} and $Q_{\mathrm{d}} << L$, the $Q_{\mathrm{s}} - Q_{\mathrm{d}} + L$ in the denominator of Equation

2.17 describing the full model is approximately equal to L, and the current becomes

$$IL \approx \frac{Q_s^2 - Q_d^2}{2} + Q_s - Q_d \tag{4.3}$$

For operation in the subthreshold region, both I and Q are much less than 1; $Q_s^2/2$ and $Q_d^2/2$ are less than either Q_s or Q_d, and Equation 4.3 reduces to Equation 4.2.

The boundary conditions, Q_s and Q_d, are generated by the Boltzmann distribution from Equation 2.19. Because the bulk and gate potentials are constant along the channel, the fact that the mobile charge can be neglected compared with the gate and bulk charges in subthreshold implies that the surface potential is constant along the channel. Because there is no electric field along the channel, current cannot flow by drift; hence, diffusion must be the dominant current-flow mechanism. This self-consistent set of conditions defines the subthreshold regime of operation. If we assume that the surface potential does not vary along the channel, and use the approximation that $\Xi_d \approx V_d$, then

$$Q_s = Q_1 e^{-\Psi_s} e^{V_s} \tag{4.4}$$

$$Q_d = Q_1 e^{-\Psi_s} e^{V_d} \tag{4.5}$$

where $Q_1 = Q_0 e^{\Psi_f}$. The boundary conditions Q_s and Q_d actually apply at the boundaries where the channel meets the source and drain depletion layers, as calculated in the full model. Applying the boundary conditions from Equations 4.4 and 4.5 to Equation 4.2, we obtain

$$I = Q_1 e^{-\Psi_s} \left(e^{V_s} - e^{V_d} \right) \tag{4.6}$$

To express the current solely as a function of the terminal voltages, we will adopt an approximation in which the surface potential is affected by the substrate ("bulk") potential; we define a constant κ such that, in the region of operation, the surface potential ψ is related

to the gate voltage v_g by

$$\kappa = \frac{\partial \psi}{\partial v_g} \tag{4.7}$$

Modifying Equation 2.5 to neglect the effect of mobile charge on the electrostatics of the device gives

$$v_g = \psi - \frac{t_{ox}}{\epsilon_{ox}} q_b = \psi + \frac{1}{C_{ox}} \sqrt{-2\rho\epsilon_s\psi} \tag{4.8}$$

$$v_g = \psi + v_{fb} - \frac{1}{C_{ox}} q_b \tag{4.9}$$

For small changes in voltage around some operating point, we can derive the dependence of surface potential on gate voltage, and hence can derive an explicit expression for κ. Differentiating Equation 4.8 with respect to ψ, we obtain

$$\frac{1}{\kappa} = \frac{\partial v_g}{\partial \psi} = 1 + \frac{1}{C_{ox}} \sqrt{\frac{\rho\epsilon_s}{2\psi}} = 1 + \frac{C_b}{C_{ox}} \tag{4.10}$$

We notice that, as the doping approaches zero, κ approaches unity. We also observe that κ is inversely related to the doping. For a highly doped substrate, more ions per unit depletion width are available to be uncovered, and the depletion width is shorter. As a result, the capacitance C_b is greater and the current increases with gate voltage less rapidly than it does for a lightly doped substrate.

If we assume that, for small excursions around the particular operating point Ψ_0, $\Psi_s = \Psi_0 + \kappa V_g$, and if we absorb the pre-exponential factors into a constant I_1, Equation 4.6 becomes

$$I = I_1 \, e^{-\kappa V_g} \left(e^{V_s} - e^{V_d} \right) \tag{4.11}$$

Using $V_{ds} = V_d - V_s$, we obtain a working form for the current:

$$I = I_1 \, e^{-\kappa V_g} \, e^{V_s} \left(1 - e^{V_{ds}} \right) \tag{4.12}$$

The current is exponential in the gate-source voltage and saturates after a few $\frac{kT}{q}$, as shown in Figures 3.2 and 3.8. This model is equivalent to the one given in [Vittoz 85].

4.1.2 Drain Conductance

In the preceding discussion, we treated the channel length L as a constant. As described in Chapter 3, L is determined by the distance between the depletion regions surrounding the source and drain. The widths of these regions are functions of the source-to-substrate and drain-to-substrate biases. The width of the drain depletion layer increases with drain voltage, thereby decreasing the channel length. This decrease in channel length increases the gradient of carrier density, and therefore increases the channel current. Because the dependence of I on L is explicit in Equation 4.2, we can solve directly for the drain conductance g_{d} of the transistor in saturation. Because the current in Equation 4.2 is inversely proportional to L,

$$\frac{\partial I_{\mathrm{sat}}}{\partial V_{\mathrm{d}}} = g_{\mathrm{d}} = \frac{\partial I_{\mathrm{sat}}}{\partial L}\frac{\partial L}{\partial V_{\mathrm{d}}} = -\frac{I_{\mathrm{sat}}}{L}\frac{\partial L}{\partial V_{\mathrm{d}}} \tag{4.13}$$

which is the result of Equation 3.1. Thus, the drain conductance g_{d} (which manifests itself as a nonzero slope on the drain characteristic curves for large V_{d}) is proportional to $I \approx I_{\mathrm{sat}}$ and is inversely proportional to L.

For hand calculations, we can approximate the derivative in Equation 4.13 by

$$\frac{1}{L}\frac{\partial L}{\partial V_{\mathrm{d}}} \approx -\frac{1}{V_0} \tag{4.14}$$

where V_0 is taken to be a constant for a given process and for a given transistor length. Because the conductance is proportional to the drain current, in this approximation the extrapolated drain curves all intersect the voltage axis at a single point, which is $-V_0$. As mentioned in the chapter introduction, this parameter is known as the *Early voltage*

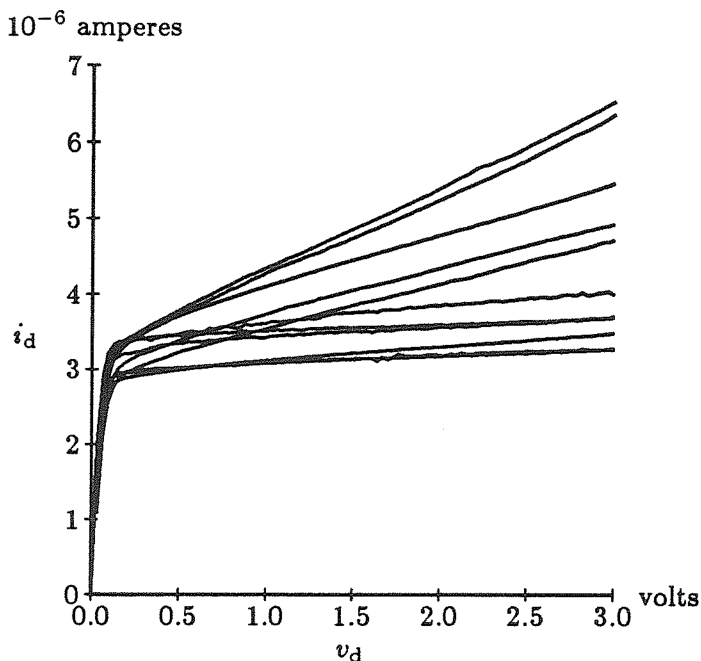

Figure 4.1: Saturation current times channel length in microns is plotted versus v_{ds} for transistors in subthreshold ($v_g = 0.5$ volts). Transistor lengths varied from 0.7 to 50 microns. All devices were measured on the same die, with gate-oxide thickness of 125 angstroms.

[Early 52] in the literature on bipolar transistors, and also was used by Vittoz [Vittoz 85].

Taking the drain conductance into account, our equation for the current becomes

$$I = I_1 \, e^{-\kappa V_g} \, e^{V_s} \left(1 - e^{V_{ds}}\right) + g_d V_{ds} \tag{4.15}$$

where we have used the approximation $g_d \Delta V_d \approx g_d \Delta V_{ds}$ [Vittoz 85]. In the V_0 approximation,

$$I = I_1 \, e^{-\kappa V_g} \, e^{V_s} \left(1 - e^{V_{ds}} + \frac{V_{ds}}{V_0}\right) \tag{4.16}$$

For transistors in saturation, with a given voltage on their gates, Equation 4.16 reduces to the following simple dependence on V_0, and therefore on L:

$$I = \text{constant} \left(1 + \frac{V_{ds}}{V_0}\right) \tag{4.17}$$

This functional form can represent the behavior of real transistors surprisingly well, as we can see in Figure 4.1. The product of the saturation drain current and the channel length

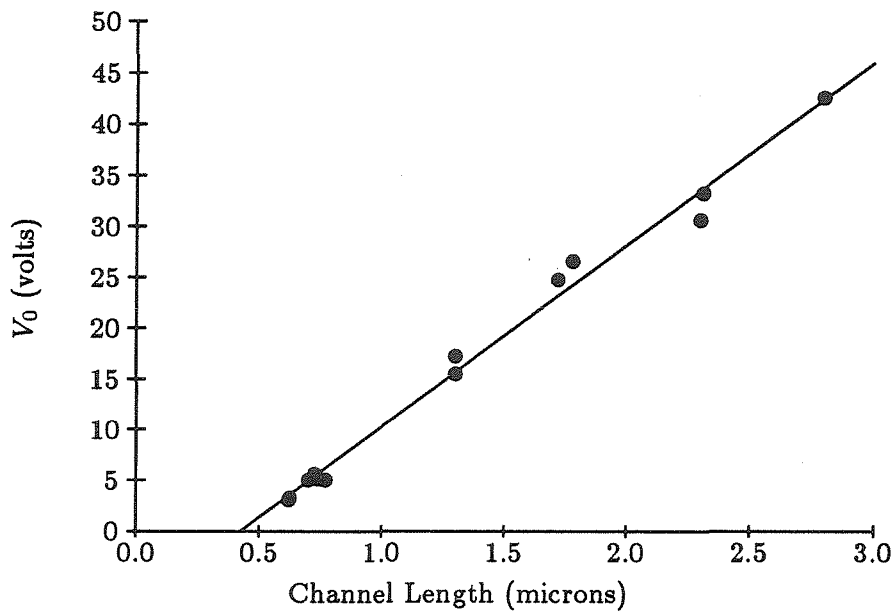

Figure 4.2: V_0 versus channel length for the devices of Figure 4.1. We determined V_0 by extrapolating the least-squares fit to the drain conductance to the v_d axis. Longer devices have larger V_0 and hence higher drain resistance. This result can be seen in Figure 4.1, where the slopes of the current versus voltage plots are higher for shorter devices.

for each transistor is a nearly straight line. Furthermore, within the voltage offsets among the transistors, all these lines extrapolate to a single value for $V_{ds} = 0$. From Equation 4.14, the value of V_0 should be directly proportional to the channel length L. This dependence can be seen in Figure 4.2.

4.1.3 Body Effect

We can see from the expressions for drain current given in the previous section that the gate voltage has less effect on the current flow than does the source voltage. As we increase the source voltage, we must increase the gate voltage even more to keep the current constant. We must deplete more charge in the substrate to lower the barrier energy (to make the channel potential more positive). That additional charge increases the electric field in the gate oxide, and thus increases the voltage across the gate oxide. This increase in the gate–source voltage required to maintain the same current as we increase the source voltage is commonly called the *body effect* (also known as the bulk effect or the back-gate effect).

For any given current, there must be a certain charge in the channel. Whether the mobile charge is large or small compared with the bulk charge, it will be constant if the current is held constant. Hence, for computing the change in gate voltage required to keep the current constant as the source voltage is increased, the charge in the channel can, to first order, be ignored. The effect is thus captured in Equation 4.9, for operation either above or below threshold. The gate–source voltage required would be constant if the square-root term in Equation 4.9 were zero. Hence, that term represents the body effect directly. The effect usually is expressed in terms of the increase in threshold voltage of the transistor with source voltage, but, as we have seen, it applies equally well to any level of current flowing in the transistor. From Equation 4.9, the increase in gate voltage is proportional to the

square root of the surface potential ψ, which, for a given current, is a constant $\Delta\psi$ above the source voltage:

$$v_g = v_s + \Delta\psi + v_{fb} + \frac{1}{C_{ox}}\sqrt{2\rho\epsilon_s(v_s + \Delta\psi)} \quad (4.18)$$

$$= v_s + \Delta\psi + v_{fb} + \gamma\sqrt{v_s + \Delta\psi} \quad (4.19)$$

where $\gamma = \sqrt{2\rho\epsilon_s}/C_{ox}$ often is called the back-gate coefficient, and γ is a function of the doping density, as a result of the quantity of charge that must be depleted in the substrate, or "body," to lower the energy barrier. The most instructive way to evaluate the effect in a real transistor is to plot v_{gs} as a function of $\sqrt{v_s + \Delta\psi}$, for a fixed current level measured on a long-channel device. The result is a straight line, as shown in Figure 2.5. The slope of the line is γ. The threshold shift due to the source substrate bias is modeled directly, because the source potential in Equation 2.19 does not need to be zero. No new terms need to be added to Equation 2.5, and no new parameters need to be added to the model.

4.2 The Model Above Threshold

Now we look at the above-threshold behavior of the model, and we generalize the results for saturated drain conductance discussed in the previous sections.

4.2.1 Long-Channel Limit in Saturation

For L large compared with Q_s and Q_d, the $Q_s - Q_d + L$ in the denominator of Equation 2.17 describing the full model is approximately equal to L, and the form of the current is the same as Equation 4.3. This approximation corresponds to the usual treatment, which ignores velocity- saturation effects. For a long-channel device in saturation, Q_d is much less

than Q_s, and Equation 4.3 becomes

$$I_{\text{sat}} L \approx \frac{Q_s^2}{2} + Q_s \tag{4.20}$$

Well above threshold, Q_s is much greater than 1; and the electric-field lines from additional charges on the gate virtually all end on mobile charges in the channel. The mobile-charge density above threshold is thus $q_s \approx (v_{gs} - v_{\text{th}})C_{\text{ox}}$, where v_{th} is the threshold voltage of the transistor. This relationship is shown in Figure 2.3. In natural units,

$$Q_s \approx (V_{gs} - V_{\text{th}})\frac{C_{\text{ox}}}{C} = \kappa(V_{gs} - V_{\text{th}}) \tag{4.21}$$

where κ is defined as in Equation 4.7. For large gate voltages, the Q_s^2 term dominates in Equation 2.16 and we have the familiar long-channel behavior, which corresponds to the standard treatment of the device characteristics:

$$I_{\text{sat}} L \propto \frac{Q_s^2}{2} \approx \frac{\kappa^2}{2}(V_{gs} - V_{\text{th}})^2 \tag{4.22}$$

In these expressions, the threshold voltage V_{th} depends on V_s through the body effect, as we discussed previously.

4.2.2 Short-Channel Transistor in Saturation

The effects of velocity saturation can be seen in Equation 2.17. For extremely short-channel devices operated in saturation, well above threshold, the $Q_s - Q_d$ in the denominator of Equation 2.17 is much greater than L, so we can neglect L. Then the first fraction of Equation 2.17 reduces to 1. Also, both Q_s and I are much greater than 1. We use the velocity-saturation boundary condition, that the charge density q_d at the drain is moving at saturated velocity v_0. Therefore, in this short-channel limit, Equation 2.17 becomes

$$I \approx \frac{Q_s + I}{2} \Rightarrow I \propto V_{gs} - V_{\text{th}} \tag{4.23}$$

So, for a highly velocity-saturated device, there is a linear dependence of current on gate voltage, rather than a quadratic dependence. We see this behavior readily by comparing the drain characteristics of the transistor with 50-micron channel length shown in Figure 3.5 with the transistor with 0.7-micron channel length shown in Figure 3.8. We can develop an intuition for this result by the following line of reasoning. For a very short-channel device, the entire population of electrons that surmount the barrier is moving at saturated velocity. In natural units, we can write this condition as $Q_s = I$, in agreement with the limit given in Equation 4.23.

It is easy to push a small transistor into the short-channel limit. For a 100-angstrom oxide thickness, a gate voltage of 3 volts produces a source charge density Q_s of about 120 in natural units. The length unit l_0 is about 0.007 micron. Thus, the natural length of a device with 0.8-micron channel length is about equal to the dimensionless charge at that gate voltage. Shorter devices will rapidly approach the limiting behavior of Equation 4.23.

4.2.3 Drain Conductance

As in our discussion of subthreshold operation, we can once again develop a simple model for the Early effect in MOS transistors. From Equation 2.17, it is obvious that the Early effect is more complicated in the general case than it was in subthreshold. Aside from the additional complexity, our previous approach to analyzing the drain conductance is still applicable. A more general form of Equation 4.14 is

$$V_0 = \frac{I}{\frac{\partial I_{\mathrm{sat}}}{\partial V_{\mathrm{d}}}} = \frac{I_{\mathrm{sat}}}{\frac{\partial I_{\mathrm{sat}}}{\partial L}\frac{\partial L}{\partial V_{\mathrm{d}}}} \tag{4.24}$$

Differentiating Equation 2.22 with respect to L, we obtain

$$\frac{\partial I_{\mathrm{sat}}}{\partial L}(L + Q_s + 1 - I_{\mathrm{sat}}) = -I_{\mathrm{sat}}$$

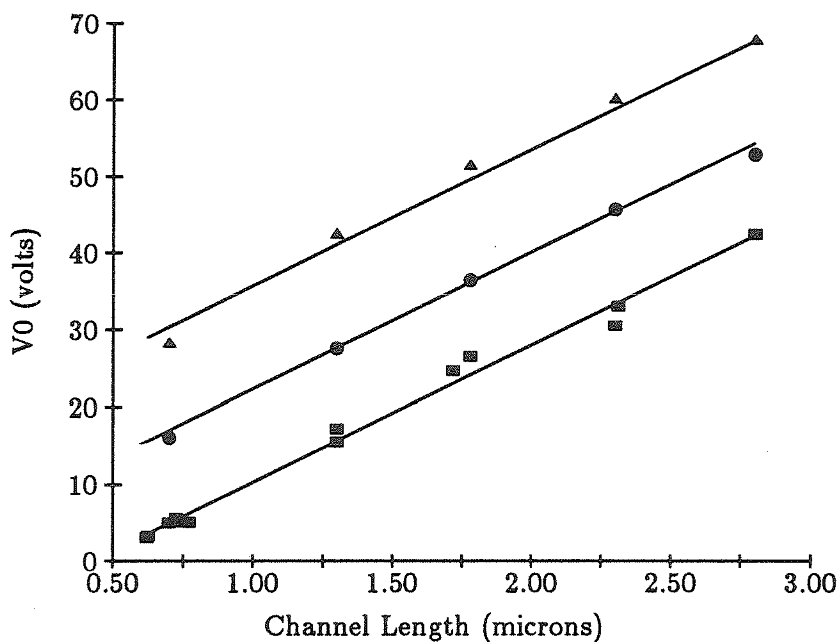

Figure 4.3: V_0 versus channel length. Curves are plotted for three different gate voltages. Solid lines are the values predicted by Equation 4.26. Dots indicate experimental data. The figure shows that, as in the subthreshold case, V_0 is proportional to L and the y-intercept value is proportional to Q_s.

or

$$\frac{I_{\mathrm{sat}}}{\frac{\partial I_{\mathrm{sat}}}{\partial L}} = -(L + Q_{\mathrm{s}} + 1 - I_{\mathrm{sat}}) \tag{4.25}$$

Substituting Equation 4.25 into Equation 4.24, we obtain the final form for the Early effect at any gate voltage:

$$V_0 = \frac{L + Q_{\mathrm{s}} + 1 - I_{\mathrm{sat}}}{-\frac{\partial L}{\partial V_{\mathrm{d}}}} \tag{4.26}$$

Measured values of V_0 versus length are plotted in Figure 4.3 for several different values of V_{g}, each corresponding to a value for Q_{s}. The effect of above-threshold operation is to shift the entire curve to higher values, maintaining approximately the same slope. This behavior is suggested by Equation 4.26. It is a direct result of the saturation of carrier velocity at high electric fields, which makes the current less dependent on channel length. It is significant for devices that are several microns in length, and is a dominant characteristic of submicron devices.

4.3 Conclusions

In this chapter, we have reviewed the behavior of transistors operating both above and below threshold, paying particular attention to the dependence of the transistors' characteristics on channel length. We have shown the correspondence between our model and other models currently being used. We have developed simple forms of the model that are useful for hand calculations, and have shown how the model takes into account the body effect. We have generalized the concept of the Early voltage to subthreshold and above-threshold MOS transistors, including to velocity-saturated devices.

Chapter 5

The Charge Model

We shall now extend our DC model to calculate the transistor's intrinsic terminal charges and transcapacitance. Because of the nonlinear nature of the underlying device physics, the transcapacitances are *nonreciprocal*: The charge induced in the second terminal by a change in voltage on the first terminal is not equal to the charge induced in the first terminal by a similar change in voltage on the second. The MOS transistor charge–voltage relations cannot be described by a set of two-terminal capacitors in an equivalent circuit model as proposed by Meyer [Meyer 71]. In order to be used in a charge conserving circuit simulator the model must define charges and transcapacitances at each terminal of the transistor [Ward 81,Yang 82].

We define the total charge at each terminal $\mathcal{Q}$, using the formulation proposed by Ward [Ward 81]; we split equating the current into each terminal of an MOS transistor in quasistatic operation to a DC transport component, I_{dc}, and a capacitive charging term, $\frac{d\mathcal{Q}}{dt}$. We use Ward's method [Ward 81] to divide the channel charge into source and drain components. with our model, these charges become ratios of simple polynomial expressions

of quantities needed for calculating the DC model. The continuous charge equations behave properly under different biasing conditions. The capacitive charging term may then be written in terms of the transcapacitances defined from these charges.

To compute the transcapacitances using our model, we must find the derivatives of the total charges with respect to the terminal voltages. We use the chain rule to factor the transcapacitances into simple polynomial expressions of the same variables used for the charge calculation and of the derivatives $\frac{\partial Q_{\mathrm{d}}}{\partial V_{\mathrm{j}}}$ and $\frac{\partial Q_{\mathrm{s}}}{\partial V_{\mathrm{j}}}$ for terminal j.

Because current is also written as a function of Q_{s} and Q_{d}, the calculation of transconductance $\frac{\partial I}{\partial V_{\mathrm{j}}}$ also can be decomposed by the chain rule and can use the same derivatives as those we calculated for the transcapacitance. Thus, we develop a relationship between transcapacitance and transconductance. The transconductances are important in solving the nonlinear equations used in circuit simulation.

We compare the predictions of the model with data taken in our own laboratory and with data taken from other researchers.

5.1 Calculation of Mobile Charge

Equation 2.15 was used to calculate the current of the MOS transistor as a function of the boundary conditions on mobile charge at source and drain. By integrating this relation from the source to an arbitrary point in the channel, we find the mobile charge per unit area in the channel Q_{m} as a function of distance Z along the channel:

$$IZ = \frac{Q_{\mathrm{s}}^2 - Q_{\mathrm{m}}^2}{2} + (Q_{\mathrm{s}} - Q_{\mathrm{z}})(1 - I)$$

$$Q_{\mathrm{m}} = -(1 - I) + \sqrt{(Q_{\mathrm{s}} + 1 - I)^2 + -2IZ} \tag{5.1}$$

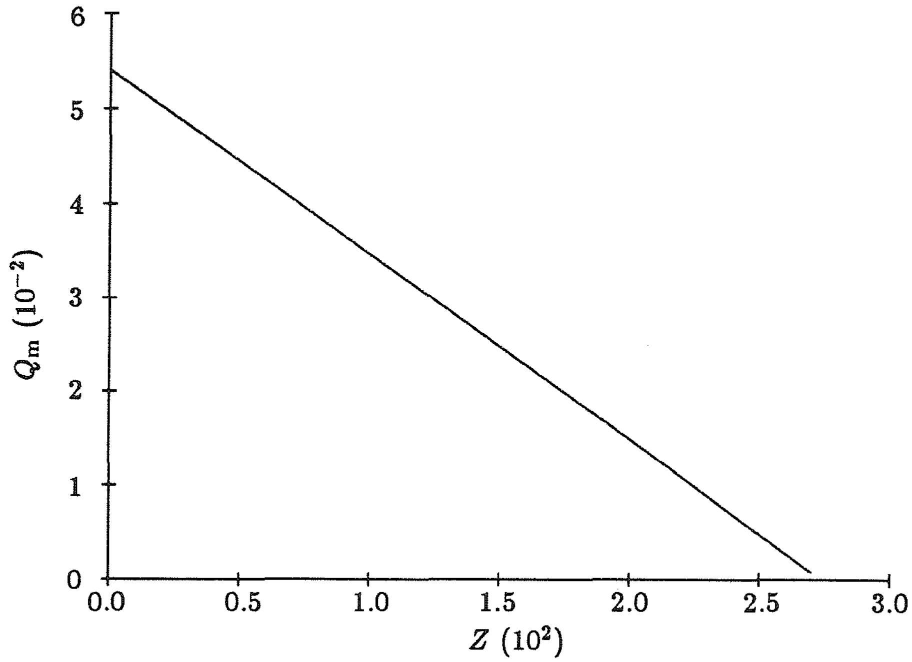

Figure 5.1: Mobile charge versus distance along the channel for the subthreshold region.

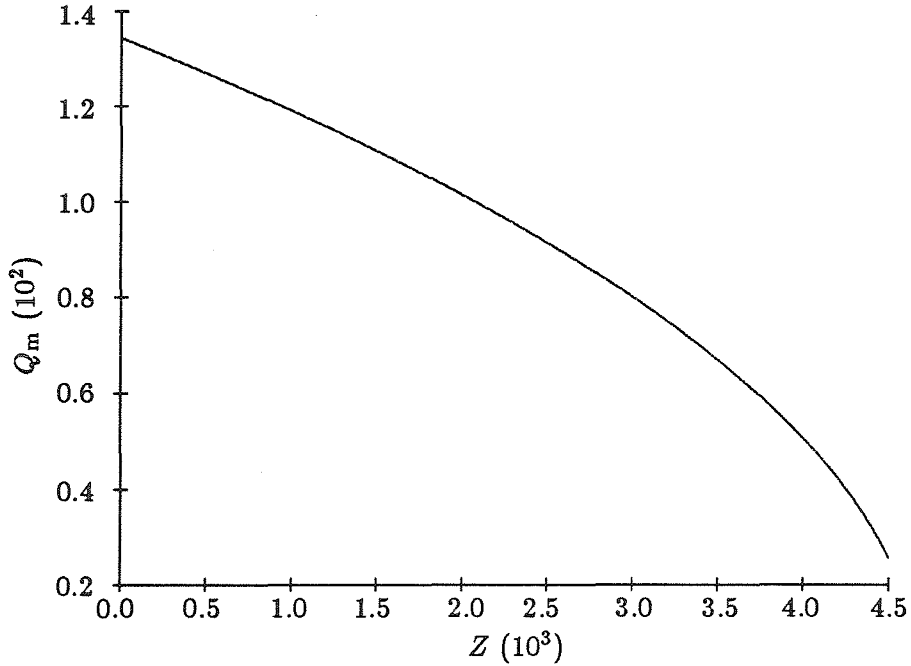

Figure 5.2: Mobile charge versus distance along the channel for an above-threshold, long-channel transistor ($l = 50$ microns).

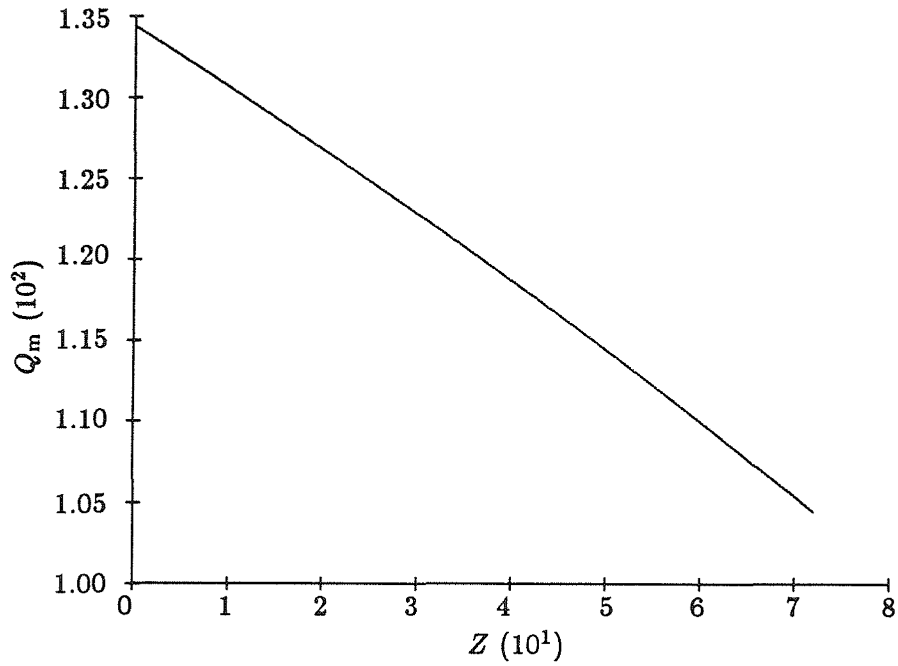

Figure 5.3: Mobile charge versus distance along the channel for a velocity-saturated transistor ($l = 0.8$ micron).

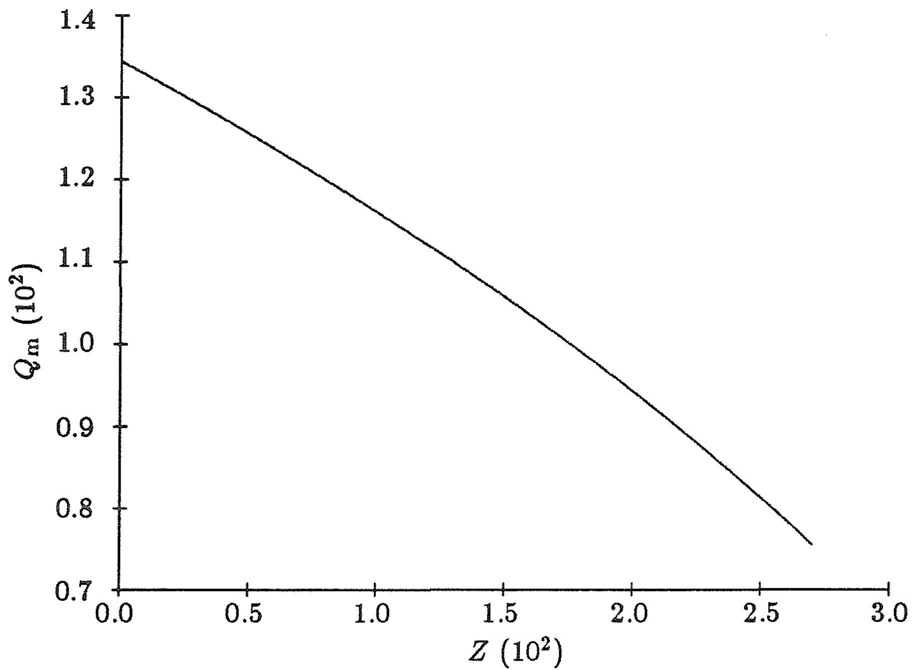

Figure 5.4: Mobile charge distribution for a 3-micron channel length transistor above threshold.

The term $(1 - I)$ is a result of velocity saturation, and for a long-channel device I will be small compared to 1, and $(1 - I) \approx 1$. As we discussed in Chapter 4, in the subthreshold region of transistor operation, Q_{m} is a linear function of distance along the channel, as shown in Figure 5.1. In subthreshold, the current flow is by diffusion and the current is proportional to the charge gradient. The charge distribution must be linear to have constant current. We derive this result by expanding Equation 5.1 and noting that, in natural units, below threshold $Q_{\text{s}} < 1$ and $Q_{\text{s}}^2 << Q_{\text{s}}$.

Above threshold, current flows by drift, and the mobile charge varies as the square root of the distance along the channel for a long-channel device, as shown in Figure 5.2. As channel lengths become shorter, however, the effects of velocity saturation can be seen. The charge distribution becomes linear even above threshold. Figure 5.3 is a plot of the charge distribution for a transistor with short-channel length (0.8 micron). So, for a short-channel device, the distribution becomes linear over all regions of operation. Above threshold, however, the charge varies over only a small range. At the feature sizes used today, we see velocity-saturation effects (see Figure 5.4). Equation 5.1 accurately predicts each of these behaviors.

5.1.1 Calculation of the Total Mobile Charge

We calculate the total mobile charge $\mathcal{Q}_{\text{m}}$ in the channel by integrating Q_{m} from the source to the drain. Let $x = (Q_{\text{s}} + 1 - I)$ and $y = (Q_{\text{d}} + 1 - I)$; then,

$$\mathcal{Q}_{\text{m}} = -(1 - I)L + \frac{2L}{3}\frac{x^2 + y^2 + xy}{x + y} \tag{5.2}$$

A simple case occurs when the source voltage is equal to the drain voltage. The boundary conditions for the mobile charge are the same, and the charge distribution is flat. Substi-

tuting the relation $Q_d = Q_s$ into Equation 5.2 gives the result that the total mobile charge is just the source charge Q_s multiplied by the transistor length L, as we expect.

5.2 The Terminal Charges

We split the current into each terminal of an MOS transistor in quasistatic operation into two parts: a DC transport component, I_{dc}, and a capacitive charging term, $\frac{d\mathcal{Q}}{dt}$, which defines a total charge $\mathcal{Q}$ associated with each terminal:

$$I_{tot} = I_{dc} + \frac{d\mathcal{Q}}{dt} \tag{5.3}$$

We use our DC model to calculate the DC current into the source and drain terminals, and we assume that I_{dc} is equal to zero at the gate and bulk terminals. The gate charge $\mathcal{Q}_g$ is equal to the sum of the total mobile charge and the total depletion charge:

$$\mathcal{Q}_g = \mathcal{Q}_m + \mathcal{Q}_b$$

The charge entering the source and drain terminals becomes the mobile charge in the channel; hence,

$$\mathcal{Q}_s + \mathcal{Q}_d = \mathcal{Q}_m$$

It is not true, however, that $\mathcal{Q}_s$ represents charge that physically entered the source terminal. A plausible method for dividing the channel charge into source and drain components [Ward 81] is given by:

$$\mathcal{Q}_d = \int_0^L \frac{Z}{L} Q_m dZ \tag{5.4}$$

$$\mathcal{Q}_s = \int_0^L \frac{1 - Z}{L} Q_m dZ \tag{5.5}$$

These definitions are based on the fact that the right-hand sides of Equations 5.4 and 5.5, when differentiated with respect to time, give the same current for the charging term in Equation 5.3 as would result from the continuity equation [Ward 81].

Substituting for Q_{m} and integrating provides the results:

$$\mathcal{Q}_{\mathrm{s}} = -\frac{L(1-I)}{2} + \frac{2L}{15}\frac{2x^3+3y^3+4x^2y+6xy^2}{(x+y)^2}$$

$$\mathcal{Q}_{\mathrm{d}} = -\frac{L(1-I)}{2} + \frac{2L}{15}\frac{2y^3+3x^3+4y^2x+6yx^2}{(x+y)^2}$$

where x and y are defined as in Equation 5.2. These simple polynomial expressions are valid over all regimes of transistor operation. The expressions are symmetric, and, with identical conditions at the source and drain, the total mobile charge splits equally between $\mathcal{Q}_{\mathrm{d}}$ and $\mathcal{Q}_{\mathrm{s}}$. In saturation, 60 percent of the charge goes to the source, and 40 percent goes to the drain, for a long-channel, above-threshold device, as shown in Figure 5.5.

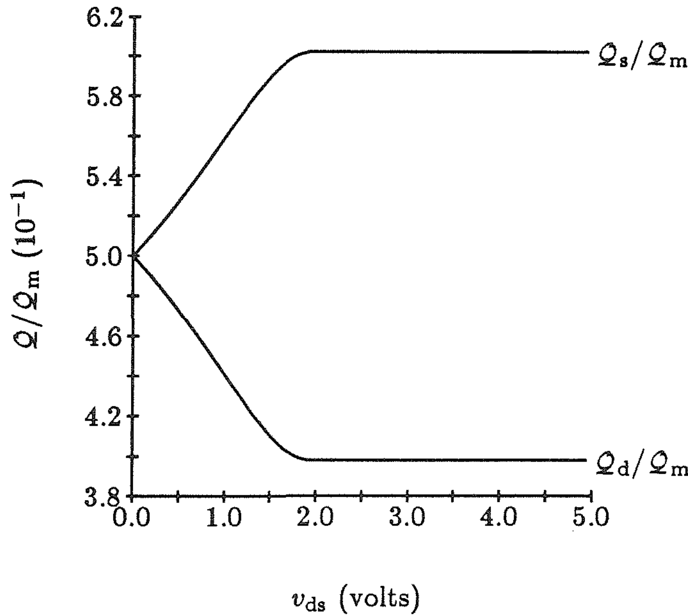

Figure 5.5: $\mathcal{Q}_{\mathrm{s}}/\mathcal{Q}_{\mathrm{m}}$ and $\mathcal{Q}_{\mathrm{d}}/\mathcal{Q}_{\mathrm{m}}$ versus v_{ds} for $v_{\mathrm{g}} = 3$ volts, $v_{\mathrm{s}} = 0$ volts, and $v_{\mathrm{b}} = 0$ volts.

The total bulk charge $\mathcal{Q}_{\mathrm{b}}$ is the sum of the charges associated with the depletion layer.

To calculate this charge, we first find an expression for the depletion-layer charge as a function of the mobile charge and of the terminal voltages from the electrostatic equations:

$$q_{\rm b} = \sqrt{2\epsilon_{\rm s}\rho}\left(\sqrt{\frac{\epsilon_{\rm s}\rho}{2C_{\rm ox}^2} + v_{\rm g} - \frac{q_{\rm m}}{C_{\rm ox}}} - \sqrt{\frac{\epsilon_{\rm s}\rho}{2C_{\rm ox}^2}}\right) \tag{5.6}$$

We then rewrite Equation 5.6 to express the depletion-layer charge $q_{\rm b}$ as a function of the distance z along the channel by substituting for $q_{\rm m}$. We integrate this expression along the channel from source to drain to find the total charge.

5.3 Definition of Transcapacitance

Because the nonlinear equations describing the charge–voltage relations in the MOS transistor cannot be written as functions of the difference in terminal voltages, these relations cannot be modeled by a set of simple, two-terminal reciprocal capacitors [Ward 81]. Instead we define a set of transcapacitances $C_{ij} = \frac{\partial Q_i}{\partial V_j}$ at each terminal. The transcapacitance relates the change in the charge associated with terminal i to a change in the voltage at terminal j. The capacitive-charging term can then be written in terms of the transcapacitance:

$$\frac{dQ_{\rm g}}{dt} = C_{\rm gg}\frac{dV_{\rm g}}{dt} + C_{\rm gs}\frac{dV_{\rm s}}{dt} + C_{\rm gd}\frac{dV_{\rm d}}{dt} + C_{\rm gb}\frac{dV_{\rm b}}{dt}$$

Note that C_{ij} is not necessarily equal to C_{ji}.

5.4 Calculation of Transcapacitance and Transconductance

To compute the transcapacitances, we must find the derivatives of the total charges with respect to the terminal voltages. We can write the total charges as functions of $Q_{\rm s}$ and $Q_{\rm d}$ by eliminating I, using Equation 2.17, or by ignoring I, in the long-channel case. The

transcapacitance calculation can then be performed by the chain rule:

$$\frac{\partial Q_i}{\partial V_j} = \frac{\partial Q_i}{\partial Q_s}\frac{\partial Q_s}{\partial V_j} + \frac{\partial Q_i}{\partial Q_d}\frac{\partial Q_d}{\partial V_j}$$

The derivatives $\frac{\partial Q_i}{\partial Q_s}$ and $\frac{\partial Q_i}{\partial Q_d}$ are simple polynomial expressions of quantities already calculated in the DC solution. The derivatives $\frac{\partial Q_s}{\partial V_j}$ and $\frac{\partial Q_d}{\partial V_j}$ are shared among all capacitance calculations for efficient calculation. We calculate these derivatives by differentiating Equations 2.2, 2.5, 2.17, and 2.21 with the appropriate boundary conditions. The result is a set of equations, each of which is linear in the derivatives and thus can be solved in closed form. An inspection of the results shows that, within these calculations, there are many common subexpressions. Further simplifications result from the fact that, by the conservation of charge, the sum of all the charges and the sum of all the currents must be equal to zero:

$$I_g + I_b + I_s + I_d = 0$$

$$Q_g + Q_s + Q_d + Q_b = 0$$

The calculations of the transcapacitances are given in Appendix A. Because these equations are analytic, the resulting transcapacitance equations are continuous.

For the model to be of use for circuit simulation, the transconductances (the derivatives of current with respect to the terminal voltages), $\frac{\partial I}{\partial V_j}$ must be calculated. The transconductances are used by most analog circuit simulators for linearizing the nonlinear transistor model about the operating point [Nagel 75]. The calculation of transconductance can also be decomposed in the same way as the calculation of transcapacitance was. The derivatives of current with respect to Q_s and Q_d are simple polynomial expressions of quantities that are needed for the DC current calculation, and the derivatives $\frac{\partial Q_s}{\partial V_j}$ and $\frac{\partial Q_d}{\partial V_j}$ are needed for the capacitance calculation. These results show a deep similarity between capacitance and

conductance. The conductance and capacitance calculations become particularly efficient. The conductance calculations are given in Appendix A. All the derivatives are analytic.

5.5 Comparision of Theory and Experiment

5.5.1 Simulation Results

Figure 5.6 shows a plot of the 16 MOS transcapacitances for a long-channel MOS transistor. The model accurately predicts the nonreciprocal nature of the MOS transcapacitances, as shown in the Figure.

5.5.2 Comparison with Ward's Data

Figure 5.7 shows a comparison between our model predictions and published experimental data from Ward [Ward 81]. We plot normalized transcapacitance versus gate voltage.

5.5.3 Experimental Data

Using the Keithley 595 quasistatic capacitance meter, we measured capacitance curves under computer control using coaxial probes. We took capacitance curves from large, wide-channel MOS (50 microns by 50 microns) transistors from a number of different runs fabricated by MOSIS. Our goal was to verify the performance of the model over different processes.

5.5.4 Comparison of the Model with Our Data

Figure 5.8 shows the comparison between curves generated by the model and data taken in our laboratory. Curves of C_{gd} are plotted versus gate voltage.

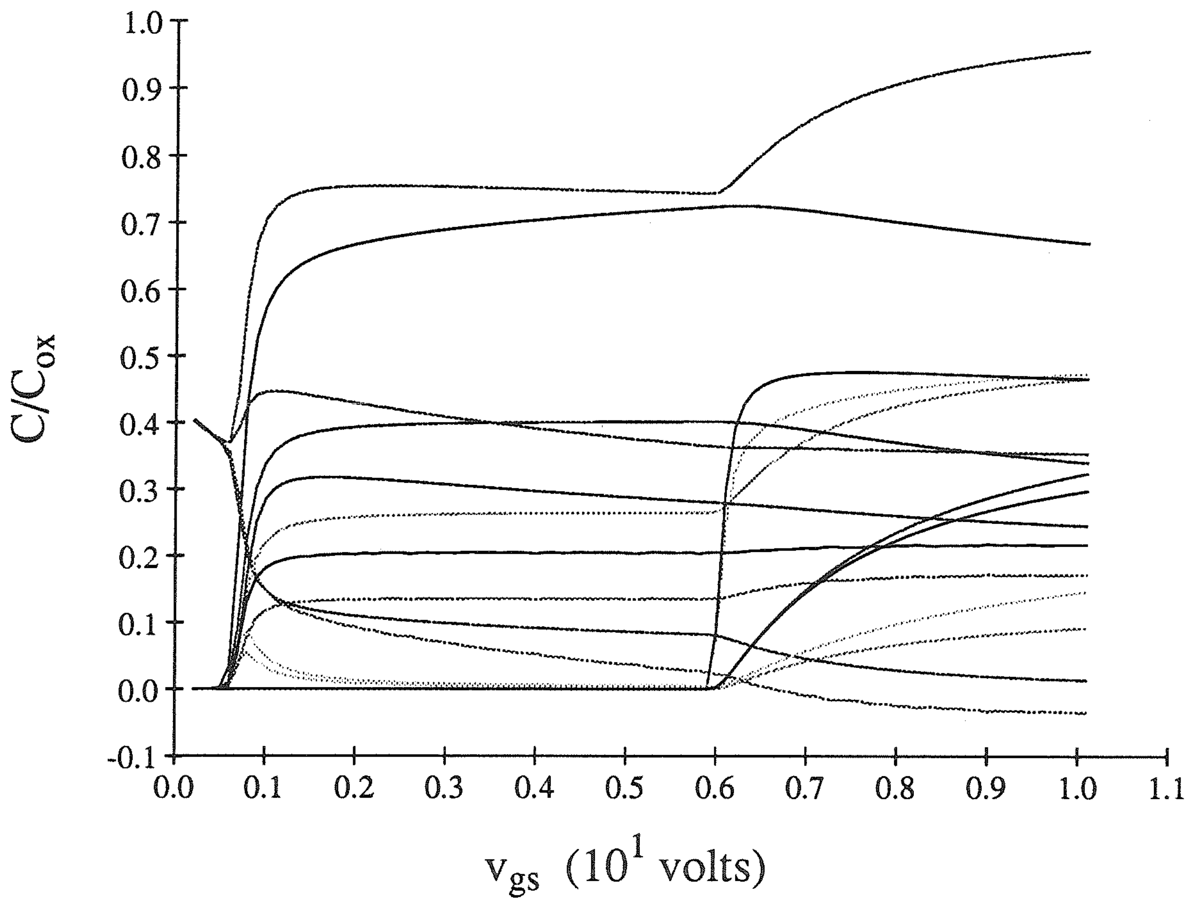

Figure 5.6: Model calculations of the 16 transcapacitances normalized by the oxide capacitance versus gate voltage.

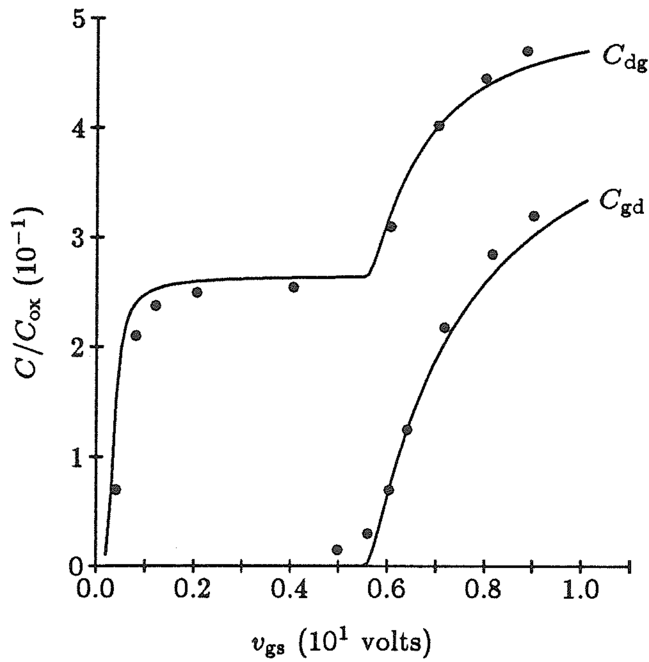

Figure 5.7: C_{dg} and C_{gd} versus gate voltage for $v_{ds} = 5$ volts and $v_{sb} = 0$ volts.

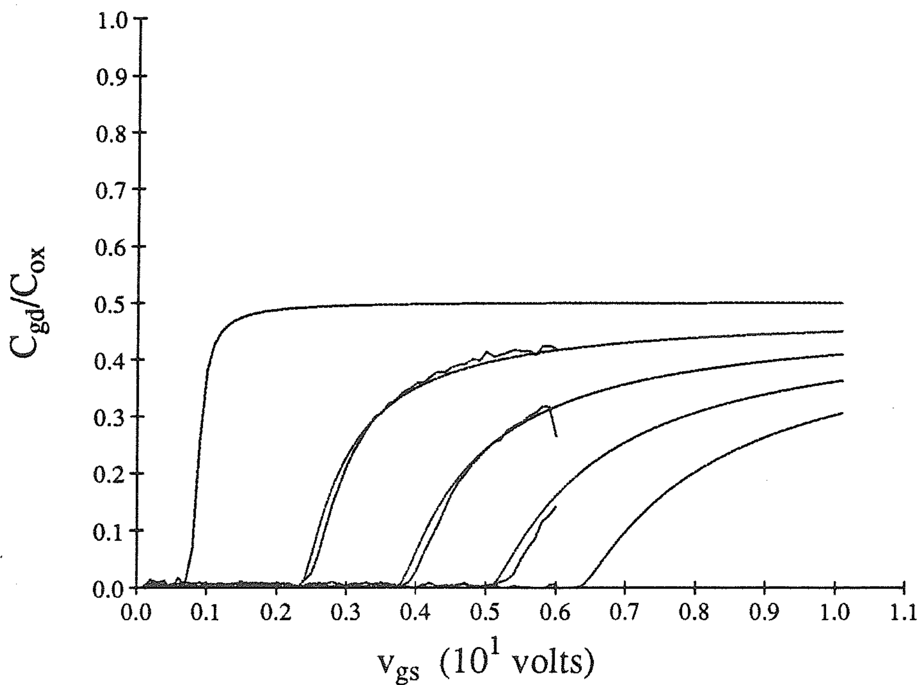

Figure 5.8: C_{gd} versus gate voltage for $v_{ds} = 5$ volts and $v_{sb} = 0$ volts.

5.6 Conclusions

We have developed a model for the mobile charge in the channel, the intrinsic charges, and the transcapacitances in the MOS transistor, and have verified it by comparing its predictions with experimental data. The source and drain charges are continuous and symmetric. Capacitances show nonreciprocity and can be calculated efficiently because they share many subexpressions and are simple polynomial combinations of quantities calculated for the DC model.

Chapter 6

Conclusions and Future Work

6.1 Summary

In this thesis, we have developed a model for the MOS transistor that includes the following new results:

- a new expression for the DC current in the transistor
- new expressions for the intrinsic terminal charges and transcapacitances
- a set of natural units for describing transistor operation
- a relation between the transcapacitance and conductance
- a new model for drain conductance based on the early voltage model that includes velocity saturation

The properties of the new model are:

- it is continuous and applicable over all regions of transistor operation

- it is charge controlled
- it is physically based
- its parameters are either derived from the process by direct measurement or are physical dimensions of the layout of a particular device.
- it is simple and computationally efficient enough for circuit simulation. The charges and transcapacitances are calculated as polynomial expressions of quantities previously calculated from the DC model.
- simplified models for circuit reasoning may be developed from it

The model generates curves that are in excellent agreement with experiment over a wide range of device sizes and predicts the scaling of the DC current with channel length. We show results on the DC model with plots of drain current and transconductance taken on transistors from 50 micron to 0.7 micron channel length. We have compared our new model with published transcapacitance data from Ward and with transcapacitance data taken in our own laboratory. Our results clearly display the non-reciprocal nature of the MOS transcapacitances. The model's performance has been verified on different processes as we have included results on data taken from chips fabricated through MOSIS, data taken on devices from Intel and published experimental data from Ward.

These results demonstrate that a simple first-principles model with physically meaningful, measurable parameters is quite capable of quantitatively predicting the behavior of MOS devices.

6.2 Circuit Simulation

In order to test our model's suitability for circuit simulation we added it to the MSIM circuit simulator. The simulator was written as a masters project and contains physically based bipolar and I^2L models. The circuit simulator uses Newton Raphson iteration to linearize the circuit equations and a novel iteration technique at the lowest level to solve the linear matrix equations. The simulator was written in Pascal and runs on the HP9836 workstations with graphical output.

Next, we added the MSIM simulator and our zero order model to the analog simulation environment of the AnaLog design tool written by graduate students John Lazzaro and David Gillespie. AnaLog has an interactive, graphical, workbench-like user interface and is written in Pascal. The AnaLog environment already contained an analog functional simulator, a digital simulator and a schematic editor. We felt that since the AnaLog simulator is used by many Caltech students and our research group, it would be a good way to test the model on a user community.

We have been using the model in the MSIM simulator developed at Caltech to simulate circuit properties and compare them with data from fabricated circuits. We are currently comparing data taken from a fast ring oscillator with the model, matching speed with power-supply scaling. We found that velocity saturation was an extremely important effect in limiting the power supply voltage at today's device sizes.

We plan to add the model to other simulators, particularly the widely used SPICE family of simulation programs The model is being added to SPICE2G5 and we will recode the model in the "C" programming language and add it to SPICE3.

6.3 Future Work

We have demonstrated the model's applicability to modeling long-channel capacitance behavior. As pointed out in [Ward 81] there is some question as to whether the assumptions made in deriving the decomposition of the channel charge into source and drain components is valid as transistors scale to submicron lengths. Our model includes the effects of velocity saturation, a major effect as the transistor channel lengths shrink. In order to test the validity of our model, we have designed a set of short-channel capacitance test structures and we are developing the capability to measure them in our laboratory.

We believe that our quasi-static model can be extended to describe fully dynamic operation using a lumped approximation to the channel.

In order to discover the charge properties of the MOS transistor, a test chip was designed, verified, and sent for fabrication to the MOSIS fabrication facility and is currently being testing. The test chip consists of a transistor under test, a p-channel device and an n-channel device in series with the test transistor and voltage followers attached to the source and drain. Using the Tektronix digital oscilloscope, we measured the time course of the voltage on the source due to a step in gate voltage.

This test chip will have many interesting applications; one being the calculation of the error voltage on the undriven node of a CMOS switch caused by the turn-off transient and dumping of the channel charge onto this node. We will also be able to study the time course of the diffusion of carriers from the source to the channel. For a long channel device the time course of this diffusion produces a time varying capacitance that could be used for for analog integrated circuit applications.

We showed how the model could be simplified for circuit reasoning. We can use better

approximations for the mobile charge boundary condition and still use the basic charge controlled modelling strategy.

Appendix A

Transcapacitance and Transconductance Calculations

A.1 Transcapacitance

The transcapacitance between terminals i and j is calculated by employing the chain rule:

$$C_{ij} = \frac{\partial \mathcal{Q}_i}{\partial V_j} = \frac{\partial \mathcal{Q}_i}{\partial Q_s}\frac{\partial Q_s}{\partial V_j} + \frac{\partial \mathcal{Q}_i}{\partial Q_d}\frac{\partial Q_d}{\partial V_j}$$

We further decompose the calculation of the gate charge by using the relation $\mathcal{Q}_g = \mathcal{Q}_m + \mathcal{Q}_b$ and

$$\frac{\partial \mathcal{Q}_g}{\partial V_j} = \frac{\partial \mathcal{Q}_m}{\partial V_j} + \frac{\partial \mathcal{Q}_b}{\partial V_j}$$

To simplify the derivation we will use the substitutions $x = (Q_s + 1)$ and $y = (Q_d + 1)$ and we will assume $1 - I \approx 1$, (the long channel case).

From the equations for the total charges we calculate the partial derivatives $\frac{\partial \mathcal{Q}_i}{\partial Q_s}$ and $\frac{\partial \mathcal{Q}_i}{\partial Q_d}$

$$
\begin{aligned}
\frac{\partial \mathcal{Q}_{\mathrm{d}}}{\partial Q_{\mathrm{s}}} &= -\frac{4L}{15}\frac{2x^3+3y^3+4x^2y+6xy^2}{(x+y)^3}+\frac{2L}{15}\frac{6x^2+8xy+6y^2}{(x+y)^2} \\
\frac{\partial \mathcal{Q}_{\mathrm{d}}}{\partial Q_{\mathrm{d}}} &= -\frac{4L}{15}\frac{2x^3+3y^3+4x^2y+6xy^2}{(x+y)^3}+\frac{2L}{15}+\frac{9x^2+12xy+4y^2}{(x+y)^2} \\
\frac{\partial \mathcal{Q}_{\mathrm{s}}}{\partial Q_{\mathrm{s}}} &= -\frac{4L}{15}\frac{2y^3+3x^3+4y^2x+6yx^2}{(x+y)^3}+\frac{2L}{15}\frac{6y^2+8xy+6x^2}{(x+y)^2} \\
\frac{\partial \mathcal{Q}_{\mathrm{s}}}{\partial Q_{\mathrm{d}}} &= -\frac{4L}{15}\frac{2y^3+3x^3+4y^2x+6yx^2}{(x+y)^3}+\frac{2L}{15}+\frac{9y^2+12xy+4x^2}{(x+y)^2} \\
\frac{\partial \mathcal{Q}_{\mathrm{m}}}{\partial Q_{\mathrm{s}}} &= \frac{-\mathcal{Q}_{\mathrm{m}}+L+\frac{2L}{3}(2x+y)}{x+y} \\
\frac{\partial \mathcal{Q}_{\mathrm{m}}}{\partial Q_{\mathrm{d}}} &= \frac{-\mathcal{Q}_{\mathrm{m}}+L+\frac{2L}{3}(2y+x)}{x+y}
\end{aligned}
$$

A.2 Terminal Derivatives

We calculate the derivatives of the mobile charge with respect to the terminal voltages $\frac{\partial Q_s}{\partial V_j}$ and $\frac{\partial Q_{\mathrm{d}}}{\partial V_j}$ by taking the derivatives with respect to voltage of Equations 2.2, 2.5, 2.17, and 2.21, and then solving the resulting set of linear equations.

Let:

1. $C = 2\rho\epsilon_{\mathrm{s}}$
2. $z = Q_{\mathrm{s}} + Q_{\mathrm{d}} + L$
3. $w = \frac{1}{-C_{\mathrm{ox}}-0.5\sqrt{\frac{C}{V_{\mathrm{d}}}}}$
4. $u = \frac{A}{Q_t}\frac{1}{Q_d-I}$
5. $v = \frac{A}{I}\frac{1}{Q_d-I}$

then the expressions for the partial derivatives become:

$$
\frac{\partial Q_{\mathrm{s}}}{\partial V_{\mathrm{s}}} = \frac{-2Q_s C_{\mathrm{ox}} - Q_s\sqrt{C/V_s}}{2V_{\mathrm{T}}C_{\mathrm{ox}} + V_{\mathrm{T}}\sqrt{C/V_s} + 2Q_s}\frac{V_{\mathrm{T}}}{Q_{\mathrm{T}}}
$$

$$
\begin{aligned}
\frac{\partial Q_{\mathrm{s}}}{\partial V_{\mathrm{g}}} &= \frac{2Q_s C_{\mathrm{ox}}}{2V_{\mathrm{T}}C_{\mathrm{ox}} + V_{\mathrm{T}}\sqrt{C/V_s} + 2Q_s}\frac{V_{\mathrm{T}}}{Q_{\mathrm{T}}} \\
\frac{\partial Q_{\mathrm{d}}}{\partial V_{\mathrm{g}}} &= \frac{-C_{\mathrm{ox}}w + u\frac{x}{z}\frac{\partial Q_s}{\partial V_g}}{-w + u(1+\frac{y}{z})}\frac{V_{\mathrm{T}}}{Q_{\mathrm{T}}} \\
\frac{\partial Q_{\mathrm{d}}}{\partial V_{\mathrm{s}}} &= = \frac{u\frac{x}{z}\frac{\partial Q_s}{\partial V_g}}{-w + u(1+\frac{y}{z})}\frac{V_{\mathrm{T}}}{Q_{\mathrm{T}}} \\
\frac{\partial Q_{\mathrm{d}}}{\partial V_{\mathrm{d}}} &= \frac{-1}{-w + u(1+\frac{y}{z})}\frac{V_{\mathrm{T}}}{Q_{\mathrm{T}}}
\end{aligned}
$$

A.3 Transconductance

The transconductances are decomposed in the same way as the transcapacitances were:

$$
G_j = \frac{\partial I}{\partial V_j} = \frac{\partial I}{\partial Q_{\mathrm{s}}}\frac{\partial Q_{\mathrm{s}}}{\partial V_j} + \frac{\partial I}{\partial Q_{\mathrm{d}}}\frac{\partial Q_{\mathrm{d}}}{\partial V_j}
$$

$$
\frac{\partial I}{\partial Q_{\mathrm{s}}} = \frac{2x}{L}
$$

$$
\frac{\partial I}{\partial Q_{\mathrm{d}}} = \frac{2y}{L}
$$

For the calculation of the conductance, we use the partial derivatives defined above in section A.2.

References

[Baccarani 78] G. Baccarani, M. Rudan, and G. Spadini. Analytical i.g.f.e.t. model including drift and diffusion currents. *IEEE Journal of Solid State Circuits*, 2:62–68, 1978.

[Baum 70] G. Baum and H. Beneking. Drift velocity saturation in MOS transistors. *Correspondence IEEE Transactions on Electron Devices*, ED-17, 1970.

[Brews 78] J. R. Brews. A charge-sheet model of the MOSFET. *Solid State Electronics*, 21, 1978.

[Cohen 81] D. Cohen and G. Lewicki. MOSIS–the ARPA silicon broker. In *Proceedings of the Second Caltech Conference on VLSI*, pages 29–44, California Institute of Technology, Pasadena, CA, 1981.

[Early 52] J. M. Early. Effects of space charge layer widening in junction transistors. *Proceedings of the IRE*, 40:1401–1406, November 1952.

[El-Mansy 77a] Y. A. El-Mansy and A. R. Boothroyd. A new approach to the theory and modeling of insulated-gate field-effect transistors. *IEEE Transactions on Electron Devices*, ED-24:241–253, 1977.

[El-Mansy 77b] Y. A. El-Mansy and A. R. Boothroyd. A simple two-dimensional model for IGFET operation in the saturation region. *IEEE Transactions on Electron Devices*, ED-24:254–262, 1977.

[Frohman-Bentchkowsky 69] D. Frohman-Bentchkowsky and A. S. Grove. Conductance of MOS transistors in saturation. *IEEE Transactions on Electron Devices*, ED-16, 1969.

[Garverick 87] S. L. Garverick and C. G. Sodini. A simple model for scaled MOS transistors that includes field-dependent mobility. *IEEE Journal of Solid State Circuits*, SC-22(1):111–114, February 1987.

[Grove 67] A. S. Grove. *Physics and Technology of Semiconductor Devices*. Wiley, New York, 1967.

[Hoeneisen 72] B. Hoeneisen and C. A. Mead. Current-voltage characteristics of small size MOS transistors. *IEEE Transactions on Electron Devices*, ED-19(3):382–383, March 1972.

[Jacoboni 77] C. C. Jacoboni, C. C. Canali, G. Ottaviani, and A. A. Quaranta. A review of some charge transport properties of silicon. *Solid State Electronics*, 20:77–89, 1977.

[Linvill 58] J. G. Linvill. Lumped models of transistors and diodes. *Proceedings of the IRE*, 46(6):1141–1152, June 1958.

[Mead 89] C. A. Mead. *Analog VLSI and Neural Systems.* Addison Wesley, 1989.

[Meyer 71] J. E. Meyer. MOS models and circuit simulation. *RCA Review*, 32:42–63, March 1971.

[Nagel 75] L. W. Nagel. *Spice2: A Computer Program to Simulate Semiconductor Circuits.* Technical Report ERL-M520, University of California, Berkeley, 1975.

[Pao 66] H. C. Pao and C. T. Sah. Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors. *Solid State Electronics*, 10:927–937, 1966.

[Popa 72] A. Popa. An injection level dependent theory of the MOS transistor in saturation. *IEEE Transactions on Electron Devices*, ED-19:774–781, 1972.

[Reddi 65] V. Reddi and C. Sah. Source to drain resistance beyond pinch-off in metal-oxide -semiconductor transistors (MOST). *IEEE Transactions on Electron Devices*, ED-12:139–141, March 1965.

[Sheu 84] B. J. Sheu, D. L. Scharfetter, and H. C. Poon. *Compact Short-Channel IGFET Model(CSIM).* Technical Report ERL-M84/9, University of California, Berkeley, 1984.

[Shockley 60] W. Shockley. *Electrons and Holes in Semiconductors.* D. Van Nostrand, 1960.

[Sodini 84] G. Sodini, P. K. Ko, and J. L. Moll. The effect of high fields on MOS device and circuit performance. *IEEE Transactions on Electron Devices*, ED-31:1386–1393, October 1984.

[Vittoz 85] E. Vittoz. *Design of MOS circuits for telecommunications*, chapter Micropower Techniques. Prentice Hall, 1985.

[Vogel Ron F. Vogel. Analytical MOSFET model with easily extracted parameters. *IEEE Transactions on Computer-Aided Design*, CAD-4(2):127–134, April 1985.

[Ward 81] D. E. Ward. *Charge-Based Modeling of Capacitance in MOS Transistors.* Technical Report G201-11, Integrated Circuits Lab -Stanford University, Stanford, CA, June 1981.

[Yang 82] P. Yang and P. K. Chatterjee. Spice modeling for small geometry MOSFET circuits. *IEEE Transactions on Computer-Aided Design*, CAD-1:169–182, 1982.