A New Inverse-Modeling-Based Technique for Sub-100-nm MOSFET Characterization

by

Zachary K. Lee

B.Sc. in Physics, University of British Columbia, 1989
M.A.Sc. in Electrical Engineering, University of British Columbia, 1992

Submitted to the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

November 3, 1998
[February, 1999]

Author ....
Department of Electrical Engineering and Computer Science
November 3, 1998

Certified by ....
Dimitri A. Antoniadis
Professor of Electrical Engineering
Thesis Supervisor

Accepted by ....
Arthur C. Smith
Professor of Electrical Engineering
Graduate Officer

A New Inverse-Modeling-Based Technique for Sub-100-nm MOSFET Characterization

by

Zachary K. Lee

Submitted to the Department of Electrical Engineering and Computer Science
on November 3, 1998 in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Performance and density of VLSI circuits have been rapidly improving over the years as transistors are miniaturized. As MOSFET transistor gates are scaled to the sub-100 nm regime, however, subtle details of the two-dimensional (2D) and three-dimensional (3D) redistribution of dopants, due to thermal diffusion during the fabrication process, strongly determine the short-channel effects, which ultimately limit device operation and performance.

In order to suppress short-channel effects, extensive use of non-uniform doping profiles are found in modern devices. Among these are the super-steep retrograde (SSR) channel profile, characterized by a low dopant concentration near the surface and a high dopant concentration at some depth from the surface, and the halo doping, characterized by a laterally non-uniform doping profile across the device channel. In order to engineer a device having good short-channel characteristics and performance, through dopant engineering, the 2D dopant distribution must be known accurately. A 2D doping profile characterization technique is therefore very important. Not only should it enable device engineering, but it can also be used as a tool for process monitoring and characterization. One dimensional (1D) profiling techniques such as the C-V method and SIMS have been widely used. Direct 2D techniques, however, have met with less success.

The first goal of this thesis is to offer a solution to this problem by developing an inverse-modeling-based 2D doping profile characterization technique using subthreshold I-V characteristics that (1) does not require special test structures, (2) is able to extract 2D doping profiles of devices with very short channel-lengths, (3) is immune to parasitic capacitances and noise, (4) has low sensitivity to gate area variations, (5) has low dependence on mobility, (6) is non-destructive, and (7) uses easily obtainable data. A second goal of the thesis is to demonstrate that knowledge of the 2D distribution of dopants can be used to calibrate transport (mobility) models, leading to highly accurate predictive capabilities useful for device design. An inverse-modeling-based transport model calibration methodology is then discussed. Lastly, parasitic effects such as source/drain series resistance, that are important for device design, and may have an impact on the accuracy in simulating high-current I-V characteristics, are discussed and analyzed.

Thesis Supervisor: Dimitri A. Antoniadis
Title: Professor of Electrical Engineering

Acknowledgments

I would like to express my appreciation for my thesis advisor, Prof. Dimitri Antoniadis, for his enormous support and encouragement. Without his vision, technical guidance, patience, and willingness to let me pursue this work, this thesis would not have been possible. And to my former advisor, Prof. Don Heiman, from whom I have learned many values as a researcher, as well as life in general. I am grateful to Dr. Nadim Khalil for his many valuable discussions and help throughout the course of this work. I would like to thank Prof. Terry Orlando and Former Dean Frank Perkins for their advice during the time when I needed it most. I would like to thank Prof. Clifton Fonstad and Prof. Jacob White, amidst their busy schedules, for acting as readers of this thesis. The funding of this work by DARPA and SRC is also acknowledged.

Many thanks to the MTL staff and individuals in our group: Mike McIlrath, Nadir Rahman, Cristina Gordy, Ilia Sokolinski, Hasan Nayfeh, Ihsan Djomehri, Dr. Jong Ho Lee, Mark Armstrong, Tony Lochtefeld, Andy Ritenour, Andy Wei, Keith Jackson, Melanie Sherony, Jim Fiorenza, Vicky Diadiuk, Patricia Burkhart, Bernard Alamariu, Paul Tierney, Barry Farnsworth, and Kurt Broderick.

Various individuals have made my stay in Boston more enjoyable: Gervais Favrot, Margaret O'Meara, Patrick Wu, King-Chung Yu, Ernest Yeh, Dorcas and Wing-Wah Sung, and brothers and sisters of the Daniel Fellowship. To my parents, sisters and their families, thank you for your patience and support throughout these years. To Kim, thank you for your understanding and willingness to listen to my endless complaints. And finally, to God, for making all things possible!

> "Even though I walk through the valley of the shadow of death, I will fear no evil, for you are with me..." - Psalms 23:4

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Problem with Modern Sub-100-nm MOSFETs

Performance and density of VLSI circuits have been rapidly improving over the years as transistors are miniaturized. As MOSFET transistor gates are scaled to the sub-100 nm regime, however, subtle details of the two-dimensional (2D) and three-dimensional (3D) redistribution of dopants[1-5], due to thermal diffusion during the fabrication process (e.g., gate oxidation and dopant drive-in/activation ... etc.), strongly determine the short-channel effects, which ultimately limit device operation and performance. Moreover, phenomena caused by complicated 2D (and 3D) interactions, such as the enhanced diffusion of dopants near the source/drain junction regions [6-16], give rise to the reverse short-channel effect (RSCE) [17-21] that must be suppressed.

In order to suppress short-channel effects, extensive use of non-uniform doping profiles are found in modern devices. The super-steep retrograde (SSR) channel profile [Fig. 1.1.1(a)], which has a low dopant concentration near the surface and a high dopant concentration at some depth from the surface, is useful for providing minimal mobility degradation due to impurity scattering near the surface while maintaining good short-channel effect control [22-26]. The halo or pocket doping, characterized by a laterally non-uniform doping profile which has a peak concentration near the source/drain extension to channel junction [Fig. 1.1.1(b)], is also useful in reducing short-channel effects [22-24, 27-30]. Due to the high non-uniformity (peakiness) of these profiles, however, fabrication requires that dopants diffuse very slowly. Heavy dopants such as indium and antimony are usually used for this purpose. However, the diffusion characteristics of these dopants are very little known, making it difficult to obtain reliable and accurate predictions from process simulations. Even for dopant species whose diffusion characteristics are well-known, transient diffusion [6-21] effects are appreciable, presenting yet another obstacle in accurate process simulation.

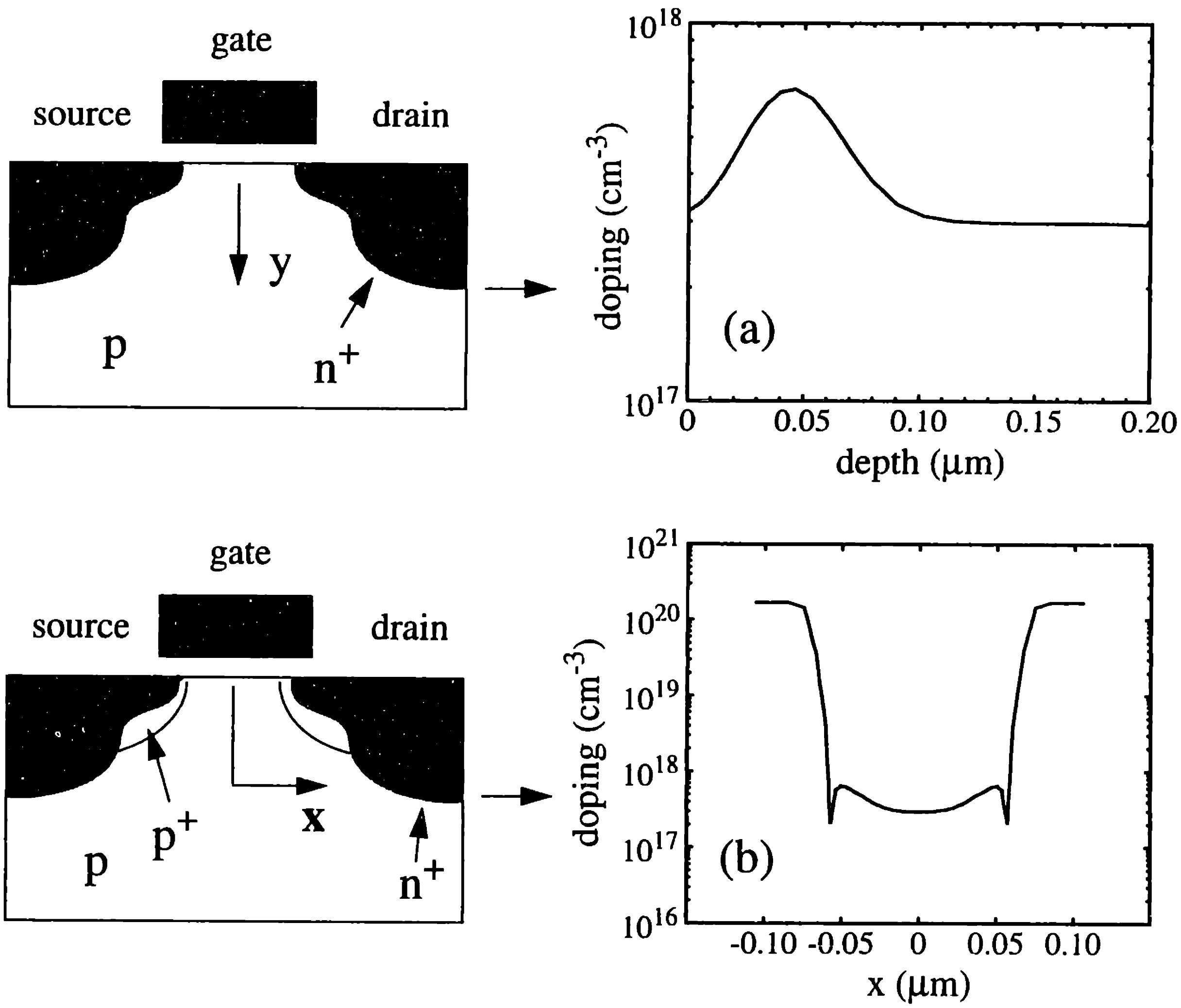

Figure 1.1.1: (a) Typical super-steep retrograde (SSR) channel doping profile having a vertically non-uniform doping concentration; and (b) typical halo doping having a laterally non-uniform doping concentration.

1.2 Goals of the Thesis

In order to engineer a device having robust and predictable short-channel characteristics and performance, through dopant engineering, the 2D dopant distribution must be known accurately. A 2D doping profile characterization technique is therefore very important. Not only should it enable device engineering, but it can also be used as a tool for process monitoring and characterization.

One dimensional (1D) profiling techniques such as the C-V [31-33] method and SIMS [34] have been widely used. Direct 2D techniques [35], however, have met with less success. An inverse modeling approach by finding a doping profile such that its simulated C-V characteristics match the corresponding experimentally determined characteristics has been used [36-40]. Due to the small capacitances involved, however, special transistor test structures having very large widths are needed.

The first goal of this thesis is to offer a solution to this problem by developing an inverse-modeling-based 2D doping profile characterization technique [41-42] using subthreshold I-V characteristics that (1) does not require special test structures, (2) is able to extract 2D doping profiles of devices with very short channel-lengths, (3) is immune to parasitic capacitances and noise, (4) has low sensitivity to gate area variations, (5) has low dependence on mobility, (6) is non-destructive, and (7) uses easily obtainable data. A second goal of the thesis is to demonstrate that knowledge of the 2D distribution of dopants can be used to calibrate transport (mobility) models, leading to highly accurate predictive capabilities useful for device design. An inverse-modeling-based transport model calibration methodology is then discussed. Lastly, parasitic effects such as source/drain series resistance, that are important for device design, and may have an impact on the accuracy in simulating high-current I-V characteristics, are discussed and analyzed in detail.

1.3 Organization of the Thesis

This thesis is organized into seven chapters and three appendices. Detailed discussions of the 2D doping profile characterization technique, the applications and results are provided in the seven chapters. Miscellaneous information such as the tool required for the application of the technique is provided in the appendices.

Chapter 2 reviews some of the basics of modern short-channel MOSFETs. Particular emphasis is placed on the subthreshold region, including discussions of short-channel effects such as drain-induced-barrier-lowering, threshold voltage roll-off, and punchthrough.

Chapter 3 presents the theoretical framework for the inverse-modeling-based 2D doping profiling technique using subthreshold I-V characteristics. More advanced discussions of short-channel effects than those covered in Chapter 2 are presented here.

Chapter 4 presents details of the 2D doping profile characterization technique as well as results for devices having various doping profiles. Sensitivity issues are also examined.

Chapter 5 presents an important application of the technique -- transport model calibration. Using 2D doping profiles obtained from subthreshold I-V inverse modeling, it is shown that highly accurate calibrations of transport models are possible, leading to excellent predictive capabilities of device characteristics.

Chapter 6 presents a detailed analysis of parasitic source/drain series resistance. The dependence of this resistance on source/drain configuration is systematically analyzed using numerical simulations. This is useful for device design as well as for the accurate simulation of high-current I-V characteristics.

Chapter 7 presents a general conclusion of this work. A suggestion for further research is also provided.

Appendix A presents a review of the Levenberg-Marquardt optimization algorithm used in the inverse modeling part of this work. Miscellaneous information that is not included elsewhere is provided in Appendix B. The “shift and ratio” technique for effective channel-length extraction that is used many times in this thesis is reviewed in Appendix C.

Chapter 2

Background: Operation of MOSFETs

2.1 Introduction

In this chapter the basic concepts of MOSFET operation are reviewed, with emphasis placed on the subthreshold region. Long-channel devices are discussed first, followed by an extension to short-channel devices. Topics including drain-induced-barrier-lowering (DIBL), threshold voltage roll-off, and punchthrough are also discussed.

2.2 General Concepts

The MOSFET (Metal Oxide Semiconductor Field Effect Transistor) operates by using a vertical electric field (controlled by a "gate") to modulate a lateral conductance between a "source" and a "drain", so that the current that flows between them is controlled by the "gate". A schematic diagram of a MOSFET is shown in Fig. 2.2.1.

The basic equations applicable to MOSFET devices include Poisson's equation and the current continuity equations. For steady-state operation, and in the absence of generation/recom-

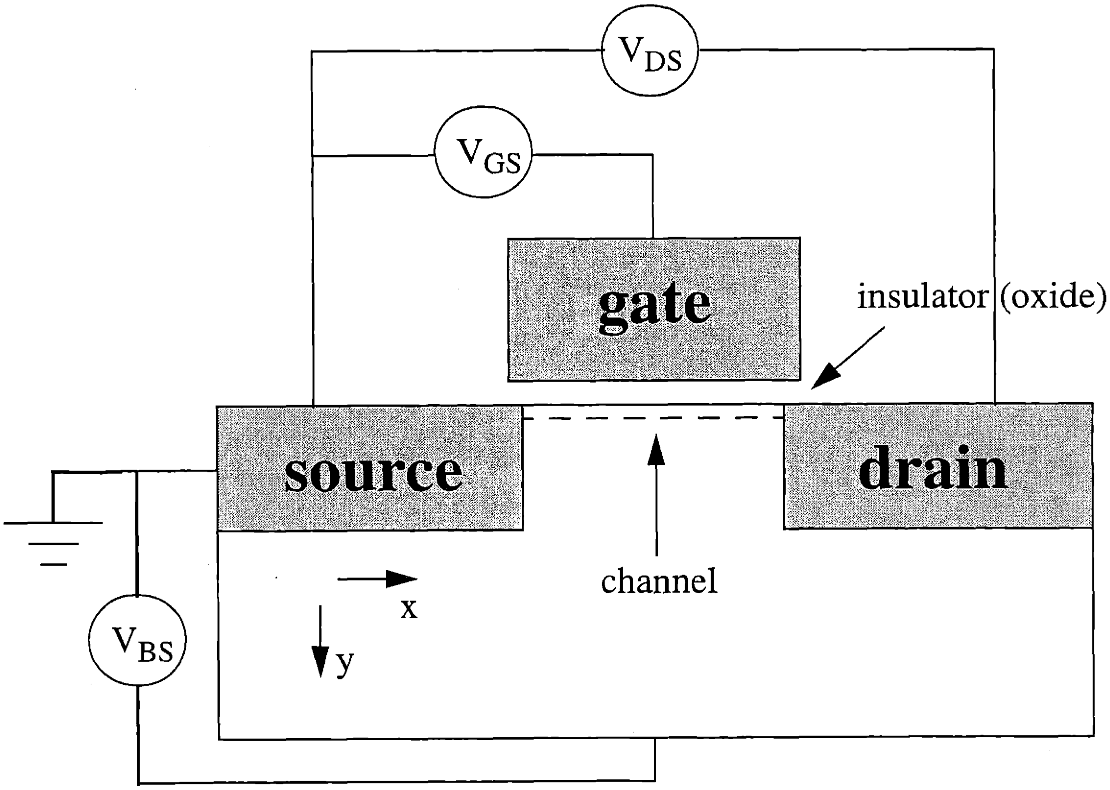

Figure 2.2.1: Schematic diagram of a typical MOSFET, showing the source, drain, gate, and the applied biases V_{GS}, V_{DS}, and V_{BS}.

bination (i.e., conditions relevant to the discussion in this thesis), these equations can be written as:

$$\varepsilon\nabla^2\psi + q(p - n + N_D^+ - N_A^-) + \rho_s = F_\psi(\psi, n, p), \tag{2.2.1}$$

$$\frac{1}{q}\nabla \bullet \overline{J_n} = F_n(\psi, n, p), \tag{2.2.2}$$

$$\frac{1}{q}\nabla \bullet \overline{J_p} = F_p(\psi, n, p) \tag{2.2.3}$$

where ε, ψ, q, p, n, N_D^+, N_A^-, ρ_s, $\overline{J}_n$, and $\overline{J}_p$ are, respectively, the dielectric constant, electrostatic potential, elementary charge, hole concentration, electron concentration, ionized donor concentration, ionized acceptor concentration, interface charge, electron current density, and hole current density. These quantities are related by the following equations:

$$\overline{J_n} = q\mu_n n\bar{E} + qD_n \nabla n, \tag{2.2.4}$$

$$\overline{J_p} = q\mu_p p\bar{E} - qD_p \nabla p, \tag{2.2.5}$$

$$\bar{E} = -\nabla\psi, \tag{2.2.6}$$

where μ_n, μ_p, $\bar{E}$, D_n, and D_p are, respectively, the electron mobility, hole mobility, electric field, electron diffusivity, and hole diffusivity. Equations (2.2.4) and (2.2.5) describe the drift and diffusion transport while eqn. (2.2.6) relates the electrostatic potential to the electric field. The primary function of any device simulator is to solve eqns. (2.2.1) to (2.2.3) self-consistently, i.e.,

$$\begin{aligned} F_\psi(\psi, n, p) &= 0 \\ F_n(\psi, n, p) &= 0, \\ F_p(\psi, n, p) &= 0 \end{aligned} \tag{2.2.7}$$

Since MOSFETs are unipolar (i.e., either one of electrons or holes is important to the current), usually only one of eqns. (2.2.2) and (2.2.3) will need to be solved in practice. For device simulation, it is convenient to use the mid-gap potential ψ, referenced to the Fermi potential $\phi_F = -E_F/q$ at equilibrium (where E_F is the Fermi energy level), so that if Maxwell-Boltzmann statistics is used, the electron and hole concentrations are given by:

$$n = n_i \exp\left[\frac{q}{kT}(\psi - \phi_n)\right], \tag{2.2.8}$$

$$p = n_i \exp\left[-\frac{q}{kT}(\psi - \phi_p)\right], \tag{2.2.9}$$

where n_i is the intrinsic carrier concentration given by

$$n_i = \sqrt{N_C N_V} \exp\left(-\frac{E_G}{2kT}\right). \tag{2.2.10}$$

Here, N_C, N_V, E_G, k, T, ϕ_n, and ϕ_p are, respectively, the conduction band density of states, valence band density of states, band-gap, Boltzmann's constant, absolute temperature, electron quasi-Fermi potential, and hole quasi-Fermi potential. If Fermi-Dirac statistics is used, more complicated expressions are used in place of eqns. (2.2.8) - (2.2.9) (see [43] for details). Using the mid-gap potential is convenient because in thermal equilibrium, when no biases are present, ϕ_n and ϕ_p can be set to zero, and are equal to ϕ_F (i.e., the reference is chosen such that the Fermi level is at zero energy). When bias is present at an ohmic contact, as boundary conditions, ϕ_n and ϕ_p at the contact are simply set to the applied bias, and ψ is set to a value consistent with the doping concentration, ϕ_n, and ϕ_p, as given by eqns. (2.2.8) and (2.2.9). Therefore, if bias is applied at an ohmic contact, ψ at the contact will be raised by an amount equal to the applied bias, from its equilibrium value. The same also applies to Schottky contacts, except that at a Schottky contact, ψ is determined by its workfunction Φ_{work}, so that

$$\psi = \chi + \frac{E_G}{2q} + \frac{kT}{2q}\ln\left(\frac{N_C}{N_V}\right) - \Phi_{work} + V_{applied}, \qquad (2.2.11)$$

where χ and $V_{applied}$ are, respectively, the semiconductor electron affinity and the applied bias. For more discussions on the implementation in device simulators, the reader is referred to, for example, [43].

2.3 Operation of Long-Channel MOSFETs

While most of the important device physics is captured by the equations in section 2.2, it is intuitive to explore analytical models in order to gain some insight of how devices operate. In this section, the operation of long-channel MOSFETs is reviewed, followed by an extension to short-channel MOSFETs in section 2.4.

Since this thesis is primarily concerned with n-channel devices, the discussion is based on n-channel devices, although it is also applicable to p-channel devices. For an n-channel device, the channel (substrate) region is doped p-type, while the source, drain, and gate are doped heavily with n-type dopants. The heavy n-type doping of the gate allows it to be conveniently modeled as a metal.

Figure 2.3.1 shows a band diagram of a MOSFET in the depth direction across the center of the channel. As positive gate bias is applied, band bending occurs near the surface of the silicon, relative to the Fermi level, inducing electrons. When the gate bias is increased to a certain point, as shown in the figure, the surface electron concentration equals the concentration of the ionized dopants, resulting in the onset of strong inversion. In the normal operation of MOSFETs, the drain current is controlled by the conductance of the inversion layer, which in turn is controlled by the gate potential.

The analysis of long-channel MOSFETs can be simplified using the assumption that $\partial^2\psi/\partial x^2 \ll \partial^2\psi/\partial y^2$ (the gradual channel approximation) [44-46], so that Poisson's equation becomes:

$$\frac{\partial^2\psi}{\partial y^2} = \frac{q}{\varepsilon_{Si}}(N_A^- - N_D^+ + n - p), \tag{2.3.1}$$

where ε_{Si} is the silicon dielectric constant. For analytical modeling, it is convenient to reference the potential to the neutral (p-type) substrate. Defining a potential ϕ with this new reference, and using eqns. (2.2.8) and (2.2.9), eqn. (2.3.1) becomes:

$$\frac{\partial^2\phi}{\partial y^2} = \frac{qN_A}{\varepsilon_{Si}}\left[1 - \exp\left(-\frac{2\phi_F}{\phi_T}\right) + \exp\left(\frac{\phi - 2\phi_F}{\phi_T}\right) - \exp\left(-\frac{\phi}{\phi_T}\right)\right], \tag{2.3.2}$$

where $\phi_T = kT/q$ is the thermal potential. Here, a p-type substrate with a uniform doping concentration is assumed. Integrating eqn. (2.3.2), and using the boundary conditions: (1) $\phi(y=0) = \phi_s$ (ϕ_s is called the "surface potential"), (2) $\phi|_{\text{depletion edge}} = 0$, and (3) $d\phi/dy|_{\text{depletion edge}} = 0$, an expression for the total charge density Q_C is obtained [44]:

$$Q_C = \mp\gamma C_{ox}\sqrt{\exp\left(-\frac{2\phi_F}{\phi_T}\right)\left(\phi_T\exp\left(\frac{\phi_s}{\phi_T}\right) - \phi_s - \phi_T\right) + \phi_T\exp\left(-\frac{\phi_s}{\phi_T}\right) + \phi_s - \phi_T}, \tag{2.3.3}$$

where $C_{ox} = \varepsilon_{ox}/T_{ox}$ is the oxide capacitance (ε_{ox} is the oxide dielectric constant, and T_{ox} is the oxide thickness), and

$$\gamma = \frac{\sqrt{2q\varepsilon_{Si}N_A}}{C_{ox}} \tag{2.3.4}$$

and is called the "body factor". For p-type silicon channel, the negative sign applies to $\phi_s > 0$ while the positive sign applies to $\phi_s < 0$. When $\phi_s < 0$ (accumulation), Q_C is the accumulation charge. When $\phi_s > 0$, Q_C is the total charge density equal to the sum of the inversion (mobile) charge Q_I and the bulk depletion charge Q_B (i.e., $Q_C = Q_I + Q_B$).

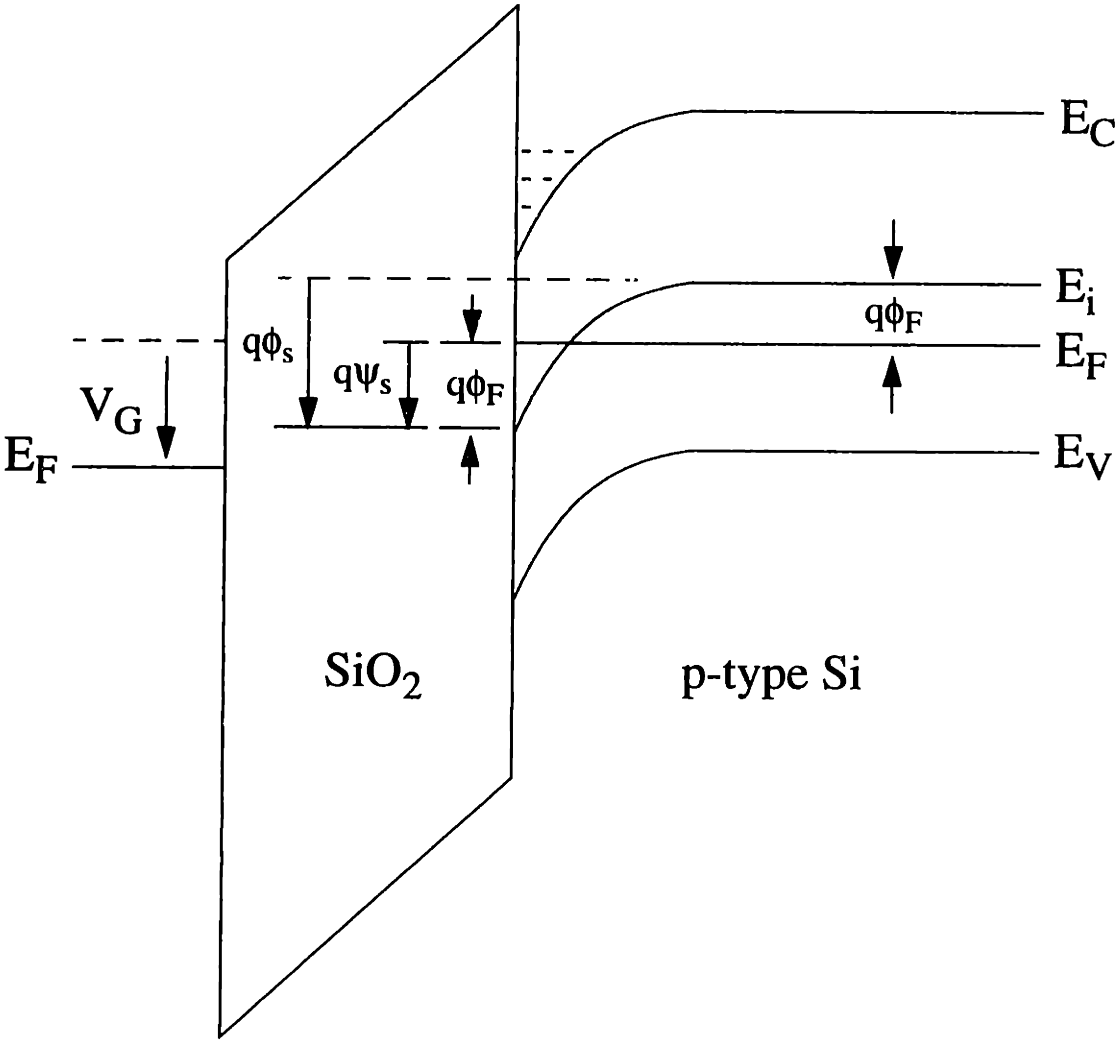

Figure 2.3.1: Energy band diagram of a MOSFET in the depth direction across the center of the channel. Here, E_C, E_i, E_F, E_V, and ϕ_F denote, respectively, the conduction band, mid-gap energy, Fermi level, valence band, and Fermi potential. As positive gate bias is applied, band bending occurs near the surface of the silicon, relative to the Fermi level, and electrons are induced (as shown). When the gate bias is increased to a certain point, so that $\phi_s = 2\phi_F$, as shown in the figure, the surface electron concentration equals the concentration of the ionized dopants, resulting in the onset of strong inversion. The surface potential with reference to the Fermi level ψ_s, and the surface potential with reference to the silicon bulk ϕ_s, are also shown.

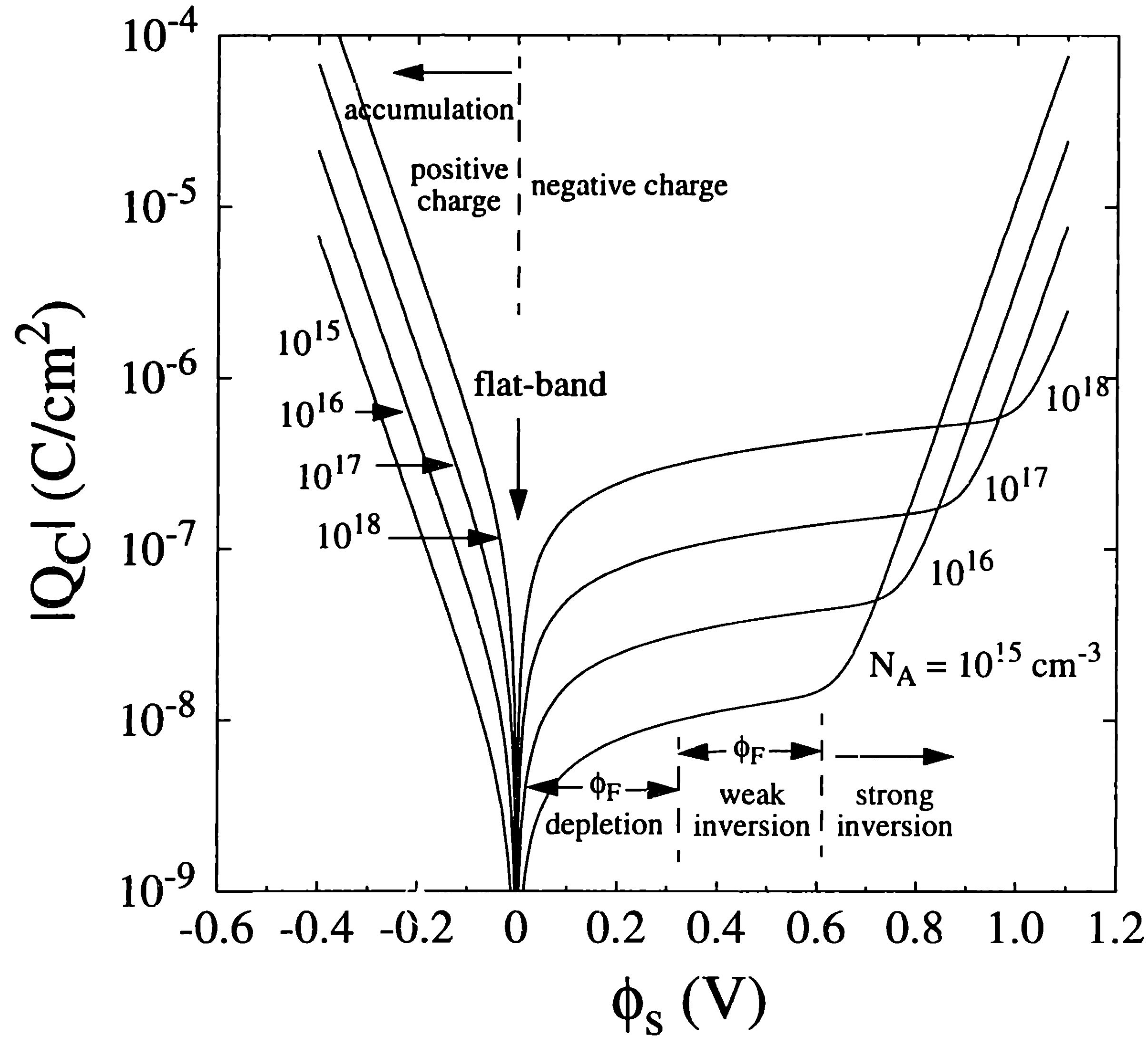

Figure 2.3.2: Total charge density Q_C as a function of surface potential ϕ_s for various substrate doping concentrations N_A. For $N_A = 10^{15}$ cm^{-3}, the different regions are identified, including accumulation, depletion, weak inversion, and strong inversion. At higher N_A, the onset of strong inversion is moved to higher ϕ_s, indicating that a higher gate bias is needed to strongly invert the silicon surface.

A plot of Q_C as a function of ϕ_s, with various N_A, is given in Fig. 2.3.2. Note that eqn. (2.3.3) contains $\gamma C_{ox} = \sqrt{2q\varepsilon_{Si}N_A}$, and is independent of T_{ox}. As can be seen, when ϕ_s is negative, positive charges are induced in the silicon (accumulation), and when ϕ_s is positive, negatives charges are induced. For positive ϕ_s, three distinct regions can be identified: depletion, weak inversion (subthreshold), and strong inversion. In depletion, $\phi_s < \phi_F$, and the electron concentration near the silicon surface is less than the intrinsic concentration (see Fig. 2.3.1). Only extremely low currents can flow between the source and the drain of the MOSFET. The total charge Q_C is dominated by the bulk depletion charge Q_B (Fig. 2.3.2). In weak inversion (subthreshold), $\phi_F < \phi_s < 2\phi_F$, and the electron concentration near the silicon surface is greater than the intrinsic concentration. In a MOSFET, if the drain is positively biased, the induced electrons near the surface can flow from the source to the drain, resulting in a "subthreshold current". The total charge Q_C is still dominated by Q_B. In strong inversion, $\phi_s > 2\phi_F$, and the induced electron concentration near the silicon surface exceeds Q_B. Because a small increase in ϕ_s will result in a large increase in Q_C, the rate of increase in ϕ_s, as gate bias is increased, is expected to decrease rapidly. In other words, ϕ_s is expected to be somewhat "pinned" near $2\phi_F$ when strong inversion is reached. The pinning of ϕ_s also results in pinning of the depletion depth and depletion charge, so that for any additional charge put on the gate, the same amount but with opposite sign will be put on the inversion layer (silicon surface) instead of the silicon bulk. Of the three regions considered, Q_C depends most strongly on Q_B in the depletion and weak inversion regions. However, since only currents in the weak inversion region are large enough to be detectable in practice, device characteristics in the weak inversion region are most useful for extracting dopant concentrations in the substrate (i.e., useful for inverse modeling).

An expression relating Q_B and ϕ_s can be obtained:

$$Q_B = -\gamma C_{ox}\sqrt{\phi_s}. \qquad (2.3.5)$$

Since $Q_I = Q_C - Q_B$, the areal inversion charge density can be calculated from ϕ_s. Using Gauss's law, the gate bias V_{GB} (referenced to the substrate) is related to ϕ_s by:

$$V_{GB} = V_{FB} + \phi_s - \frac{Q_c}{C_{ox}}, \qquad (2.3.6)$$

where V_{FB} is the "flat-band" voltage, and is an offset caused by workfunction difference between the gate and substrate, as well as any interface charge that may exist. Substituting eqn. (2.3.3) into (2.3.6), an expression relating V_{GB} and ϕ_s can be obtained. Then ϕ_s can be solved for as a function of V_{GB}, and consequently expressions relating Q_C, Q_I, and Q_B to V_{GB} can be obtained. It is cumbersome to do this, however. If appropriate approximations are used, much simplified but accurate expressions can be obtained.

(a) Electron Concentration in Weak Inversion (Subthreshold)

In weak inversion, $|Q_I| << |Q_B|$, and the following expressions apply [44, 47]:

$$\phi_s = \left[-\frac{\gamma}{2} + \left(\frac{\gamma^2}{4} + V_{GB} - V_{FB} \right)^{\frac{1}{2}} \right]^2, \qquad (2.3.7)$$

$$Q_I = -\frac{\gamma C_{ox} \phi_T}{2\sqrt{\phi_s}} \exp\left(\frac{\phi_s - 2\phi_F}{\phi_T} \right) = -\frac{\sqrt{2q\varepsilon_{Si} N_A}\,\phi_T}{2\sqrt{\phi_s}} \exp\left(\frac{\phi_s - 2\phi_F}{\phi_T} \right). \qquad (2.3.8)$$

In this region, ϕ_s increases approximately linearly with V_{GB}, with an inverse slope n given by:

$$n = 1 + \frac{\gamma}{2\sqrt{1.5\phi_F}}, \qquad (2.3.9)$$

so that

$$Q_I \propto \exp\left(\frac{V_{GB}}{n\phi_T} \right). \qquad (2.3.10)$$

(b) Electron Concentration in Strong Inversion

In strong inversion, $|Q_I| >> |Q_B|$, $\phi_s \sim 2\phi_F + m\phi_T$ (the term $m\phi_T$ is added to ensure that the device is in strong inversion, and m is a small constant typically between 3 and 6 [44, 47]), and the following expressions apply:

$$Q_I = -\sqrt{\frac{2q\varepsilon_{Si}\phi_T}{N_A}}n_i\exp\left(\frac{\phi_s}{2\phi_T}\right) = -C_{ox}(V_{GB}-V_t), \tag{2.3.11}$$

where V_t is called the "threshold voltage", and is given by:

$$V_t = V_{FB} + 2\phi_F + m\phi_T + \gamma\sqrt{2\phi_F + m\phi_T}. \tag{2.3.12}$$

Equations (2.3.7) - (2.3.12) are useful for deriving expressions for the drain current I_D.

(c) Current in Weak Inversion (Subthreshold)

In weak inversion, Q_I is very small, and ϕ_s is flat across the channel, independent of drain bias [44-49]. Consequently both the lateral electric field ($-d\phi_s/dx$) and Q_I are small, making the drift current ($-\mu_n W Q_I d\phi_s/dx$; where W is the device width) vanishingly small. The drain current I_D in weak inversion is therefore dominated by the diffusion component, i.e.,

$$I_D = D_n W\frac{dQ_I}{dx} = \mu_n W\phi_T\frac{dQ_I}{dx}, \tag{2.3.13}$$

where the Einstein relation $D_n = \mu_n\phi_T$ is used. Integrating both sides of eqn. (2.3.13) and rearranging terms,

$$I_D = \mu_n\frac{W}{L}\phi_T[Q_I(L)-Q_I(0)], \tag{2.3.14}$$

where L is the channel-length. Using eqn. (2.3.8), and noting that if positive biases are applied at the source and drain contacts, the quasi-Fermi potential near those contacts will be increased by an amount equal to the applied bias, reducing the electron concentration nearby in the channel, so that $Q_I(0)$ and $Q_I(L)$ are given by:

$$Q_I(0) = -\frac{\sqrt{2q\varepsilon_{Si}N_A}\phi_T}{2\sqrt{\phi_s}}\exp\left(\frac{\phi_s-2\phi_F}{\phi_T}\right)\exp\left(-\frac{V_{SB}}{\phi_T}\right), \tag{2.3.15}$$

$$Q_I(L) = -\frac{\sqrt{2q\varepsilon_{Si}N_A}\phi_T}{2\sqrt{\phi_s}}\exp\left(\frac{\phi_s-2\phi_F}{\phi_T}\right)\exp\left(-\frac{V_{DB}}{\phi_T}\right). \tag{2.3.16}$$

Therefore the subthreshold current is given by:

$$I_D = \mu_n \frac{W}{L} \frac{\sqrt{2q\varepsilon_{Si}N_A}\phi_T^2}{2\sqrt{\phi_s}} \exp\left(\frac{\phi_s - 2\phi_F}{\phi_T}\right)\left[\exp\left(-\frac{V_{SB}}{\phi_T}\right) - \exp\left(-\frac{V_{DB}}{\phi_T}\right)\right]. \qquad (2.3.17)$$

A more compact expression is given by [44]:

$$I_D = \frac{W}{L} I_x \exp\left(\frac{V_{GS} - V_x}{n\phi_T}\right)\left[1 - \exp\left(-\frac{V_{DS}}{\phi_T}\right)\right], \qquad (2.3.18)$$

where

$$I_x = \mu_n C_{ox} \phi_T^2 \frac{\gamma}{2\sqrt{1.5\phi_F - V_{BS}}} \exp\left(-\frac{\phi_F}{2\phi_T}\right), \qquad (2.3.19)$$

$$V_x = V_{FB} + 1.5\phi_F + \gamma\sqrt{1.5\phi_F - V_{BS}}, \qquad (2.3.20)$$

$$n = 1 + \frac{\gamma}{2\sqrt{1.5\phi_F - V_{BS}}}. \qquad (2.3.21)$$

At V_{DS} greater than a few ϕ_T, the drain bias has negligible effect on I_D. This is characteristic of diffusion, which, physically, means that at V_{DS} greater than a few ϕ_T, the electron concentration near the drain becomes so low that its value does not affect the concentration gradient. Note that expressions for V_x [eqn. (2.3.20)] and V_t [eqn. (2.3.12)] are very much alike, and each represents an offset in V_{GS} that corresponding approximations in each region applies.

Due to the exponential dependence of I_D on V_x (which contains terms relating to N_A and T_{ox}), a small change in N_A and T_{ox} will result in a substantial change in I_D. For example, consider the term $\exp(-V_x/n\phi_T)$, which is proportional to I_D, and assuming T_{ox} = 5 nm, a ratio of 1/30 is obtained if N_A is changed from 10^{17} cm^{-3} to 2x10^{17} cm^{-3}. In other words, I_D is reduced by approximately 30 times if N_A is doubled. This makes the subthreshold current very sensitive to doping concentrations, useful for inverse modeling (discussed in Chapters 3 and 4). The mobility μ_n, however, only affects I_D linearly, and therefore any uncertainty related to it will have very little effect on the results in inverse modeling. Note that the I_D sensitivity on N_A can be increased substantially if ϕ_T is decreased. This implies that using low temperature subthreshold I-V data can dramatically improve the sensitivity of the inverse modeling technique. However, since N_A affects the exponential term through γ, which also contains T_{ox}, it is clear that the inverse modeling technique is also sensitive to T_{ox}. Consequently, an accurate T_{ox} value needs to be obtained in order to obtain accurate results from inverse modeling.

(d) Current in Strong Inversion

Following a similar formulation as above, the drift and diffusion components in strong inversion can be obtained [44]:

$$\int_0^L I_D^{\text{drift}} dx$$

$$= \mu_n W C_{ox}\left[(V_{GB} - V_{FB})(\phi_s(L) - \phi_s(0)) - \frac{1}{2}(\phi_s^2(L) - \phi_s^2(0)) - \frac{2}{3}\gamma(\phi_s^{3/2}(L) - \phi_s^{3/2}(0))\right], \tag{2.3.22}$$

$$\int_0^L I_D^{\text{diffusion}} dx = \mu_n W C_{ox}\phi_T[(\phi_s(L) - \phi_s(0)) + \gamma(\phi_s^{1/2}(L) - \phi_s^{1/2}(0))]. \tag{2.3.23}$$

In this regime, however, the diffusion component is unimportant. Assuming strong inversion throughout the channel and using

$$\phi_s(0) = 2\phi_F + m\phi_T + V_{SB}, \tag{2.3.24}$$

$$\phi_s(L) = 2\phi_F + m\phi_T + V_{DB}, \tag{2.3.25}$$

in eqn. (2.3.22), the drain current I_D can be obtained:

$$I_D = \mu_n \frac{W}{L} C_{ox}\left\{(V_{GS} - V_{FB} - 2\phi_F - m\phi_T)V_{DS} - \frac{1}{2}V_{DS}^2 - \frac{2}{3}\gamma(2\phi_F + m\phi_T - V_{BS} + V_{DS})^{3/2} + \frac{2}{3}\gamma(2\phi_F + m\phi_T - V_{BS})^{3/2}\right\}. \tag{2.3.26}$$

This expression is valid up to V_{DS}', beyond which I_D remains constant because the drain side is no longer strongly inverted, and the channel is "pinched off". V_{DS}' is given by:

$$V_{DS}' = V_{GS} - V_{FB} - 2\phi_F - m\phi_T + \frac{\gamma^2}{2} - \gamma\sqrt{V_{GS} - V_{FB} - V_{BS} + \frac{\gamma^2}{4}}. \tag{2.3.27}$$

At $V_{BS} = 0$, and $V_{DS} << 2\phi_F$, eqn. (2.3.26) reduces to

$$I_D = \mu_n \frac{W}{L} C_{ox}\left[(V_{GS} - V_t)V_{DS} - \frac{1}{2}V_{DS}^2\right], \tag{2.3.28}$$

where V_t is given by eqn. (2.3.12).

2.4 Subthreshold Behaviors of Long-Channel and Short-Channel MOSFETs

Figures 2.4.1 - 2.4.3 show the simulated results of the surface potential distribution (referenced to the substrate), electron concentration, and conduction band E_C with quasi-Fermi level E_{Fn} of a long-channel NMOSFET device (L = 1 μm, $N_A = 3\text{x}10^{17}$ cm^{-3}, and T_{ox} = 5 nm). From Fig. 2.4.1, it can be seen that in the weak inversion (subthreshold) region, the surface potential distribution is flat (i.e., constant) across most of the channel. Since the lateral electric field is very low in the flat region, diffusion, as already discussed, is the primary mechanism. The source/channel transition is similar to a n^+/n^- junction. There is a potential energy barrier that electrons must surmount in order to enter the flat region of the channel (i.e., electrons in the source are "repelled" by the electric field of the channel). Since current is made up of drift and diffusion components [eqn. (2.2.4) - (2.2.5)], the diffusion component at the source/channel transition must be greater than the opposing drift component in order for current to flow. At the channel/drain transition, however, the electric field is oriented in such a way that electrons are "attracted" to the drain. Note that the surface potential distribution is flat, regardless of V_{DS}.

The exponential dependence of surface electron concentration on ϕ_S, which in turn depends approximately linearly on V_{GS}, as suggested by eqn. (2.3.8), is shown in Fig. 2.4.2. Here, the insensitivity of the subthreshold current on V_{DS} for V_{DS} greater than a few ϕ_T is evident. The higher V_{DS} moves the electron quasi-Fermi level away from the conduction band near the drain (Fig. 2.4.3), reducing the surface electron concentration in the channel nearby. However, further away from the drain and into the channel, the electron concentration and its gradient are hardly changed even if V_{DS} is increased, as shown on a linear scale in Figure 2.4.4, which also shows the surface potential distribution. Examination of Fig. 2.4.4 reveals a number of important facts. First, as explained earlier, there is a potential barrier that electrons from the source must surmount in order to enter the channel. The current in the region before the barrier represents a balance of drift and diffusion components, with the diffusion component dominating so that there is a net current flow. The sharp rise in electron concentration towards the source is an indication that there are both drift and diffusion components in that region. After the barrier, the surface potential is flat and the electron concentration gradient is constant, indicating that the current in this region is almost entirely diffusive (similar to the base of a short-base diode in which there is no recombination so that the concentration gradient is constant, and the current is entirely diffusive). Near the end of the channel, the electron concentration gradient changes, indicating that a drift component begins to set in. Second, while eqn. (2.3.14) is still appropriate, eqn. (2.3.17) is not accurate because it assumes that the surface potential distribution is flat across the entire channel-length L. In fact, as Fig. 2.4.4 shows, "diffusion-only transport" occurs approximately between $x = -0.35\ \mu m$ and $x = 0.4\ \mu m$. The length L used in eqn. (2.3.14) should be replaced by approximately $L - 2\lambda$, as shown in Fig. 2.4.4. Here, λ arises from the 2D solution of Poisson's equation and represents a characteristic length with which the potential of the source/drain decays to that of the channel. As shown in Fig. 2.4.4, λ is approximately $0.1\ \mu m$. Third, the beginning of the flat region of the surface potential distribution (i.e., the first point of the peak of the barrier) coincides with the beginning of the "diffusion-only" region. With these points taken into account, it becomes obvious that, on a first order, the current in weak inversion (subthreshold) of a long-channel device is proportional to the electron concentration at the first point of the peak of the barrier (denoted by a solid circle in Fig. 2.4.4).

Because of the characteristic length λ, which represents the lateral extension of the source/drain potential into the channel, the surface potential distribution, length of diffusion region, position of the barrier peak, barrier height, electron concentration, and consequently the current, are all expected to be dramatically changed if the source-to-drain distance is reduced to the order of 2λ, as in modern-day short-channel devices.

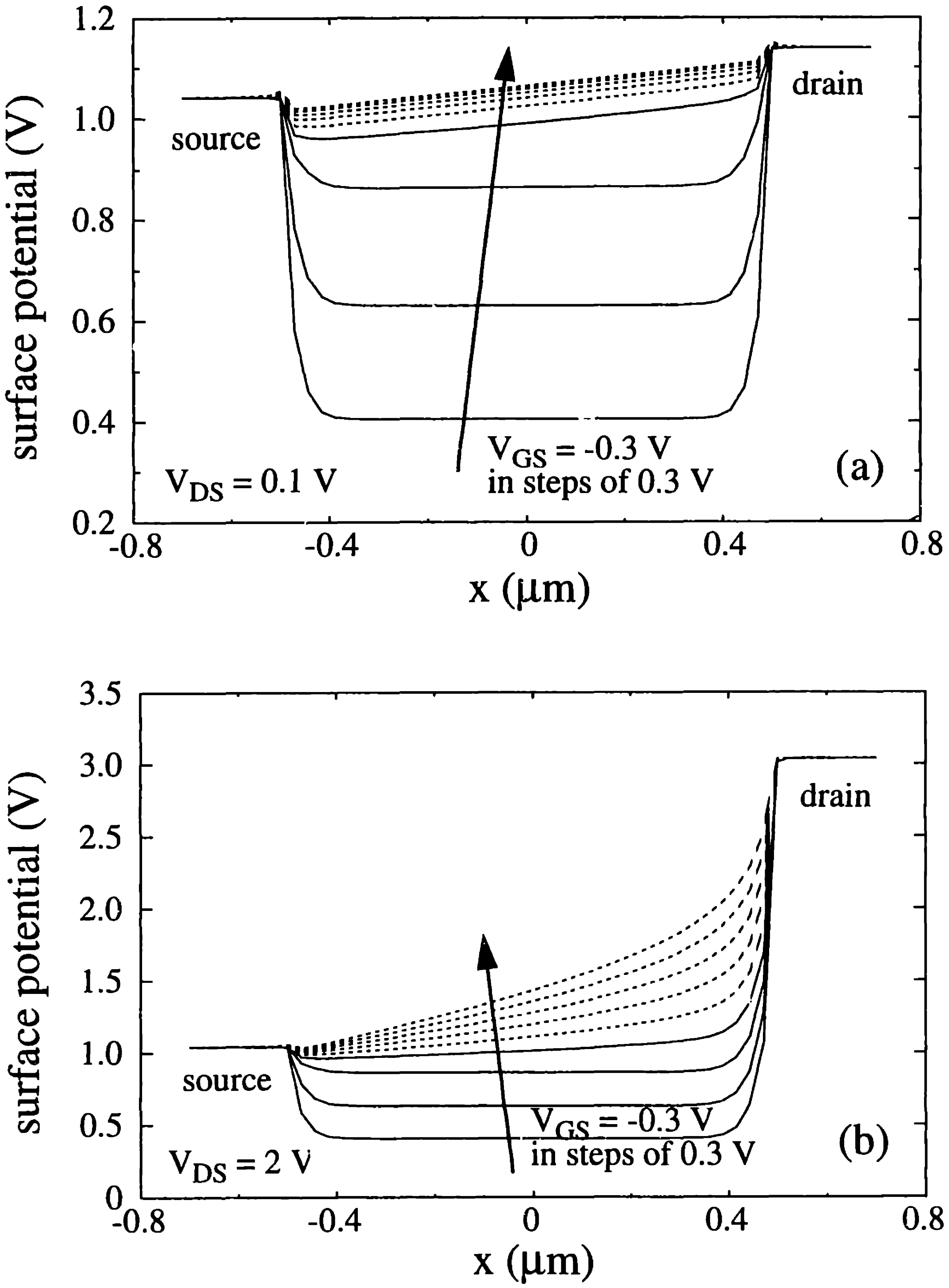

Figure 2.4.1: Simulated surface potential distribution (referenced to the substrate) of a 1 μm channel-length NMOSFET having $N_A = 3\times10^{17}\ \text{cm}^{-3}$ and T_{ox} = 5 nm at (a) V_{DS} = 0.1 V, and (b) V_{DS} = 2 V, and at V_{GS} in steps of 0.3 V. Approximately, the weak inversion (subthreshold) region is denoted by solid lines, while the strong inversion region is denoted by dashed lines.

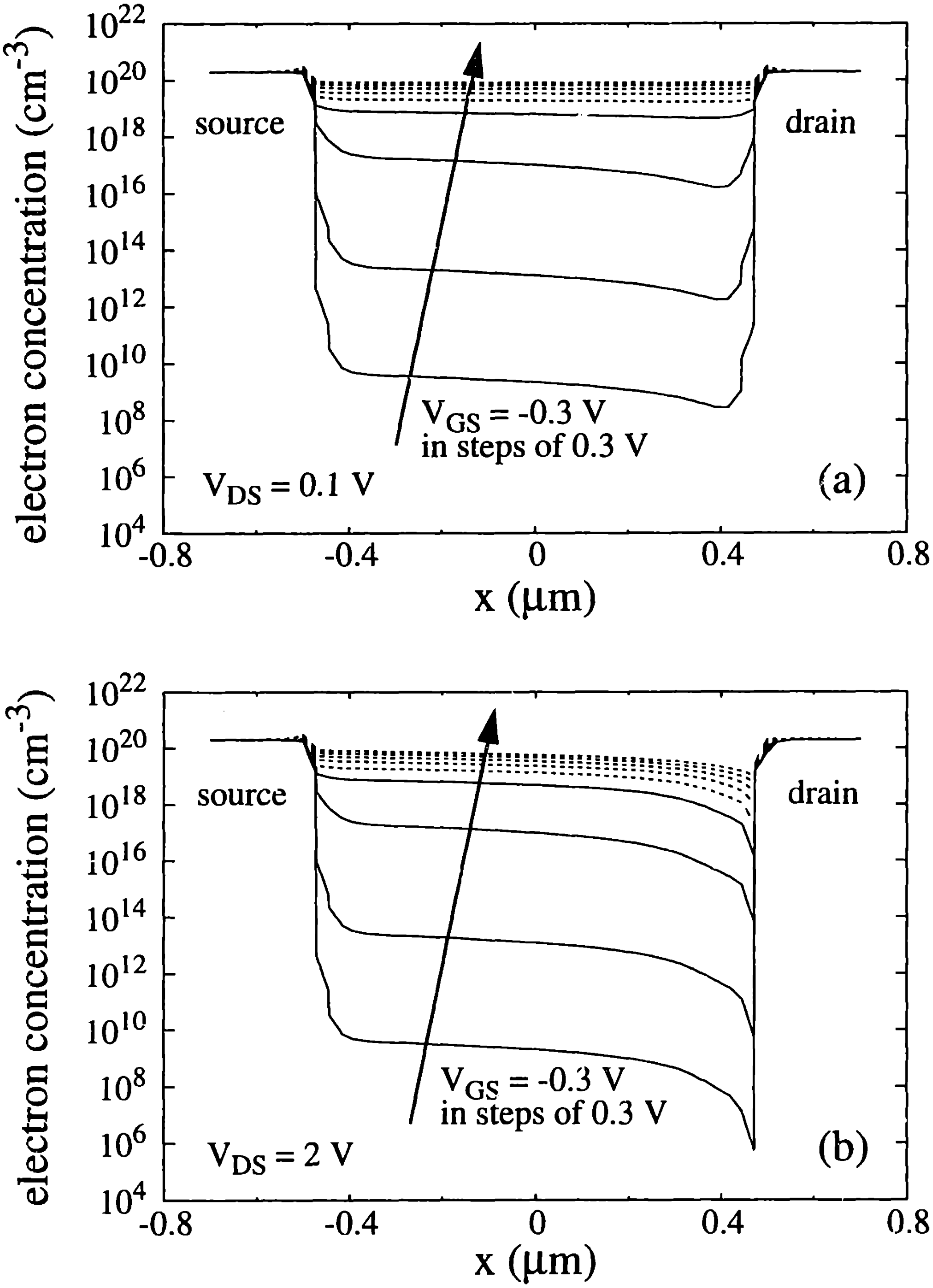

Figure 2.4.2: Simulated surface electron concentration of the device of Fig. 2.4.1 at (a) V_{DS} = 0.1 V, and (b) V_{DS} = 2 V, and at V_{GS} in steps of 0.3 V. Approximately, the weak inversion (subthreshold) region is denoted by solid lines, while the strong inversion region is denoted by dashed lines.

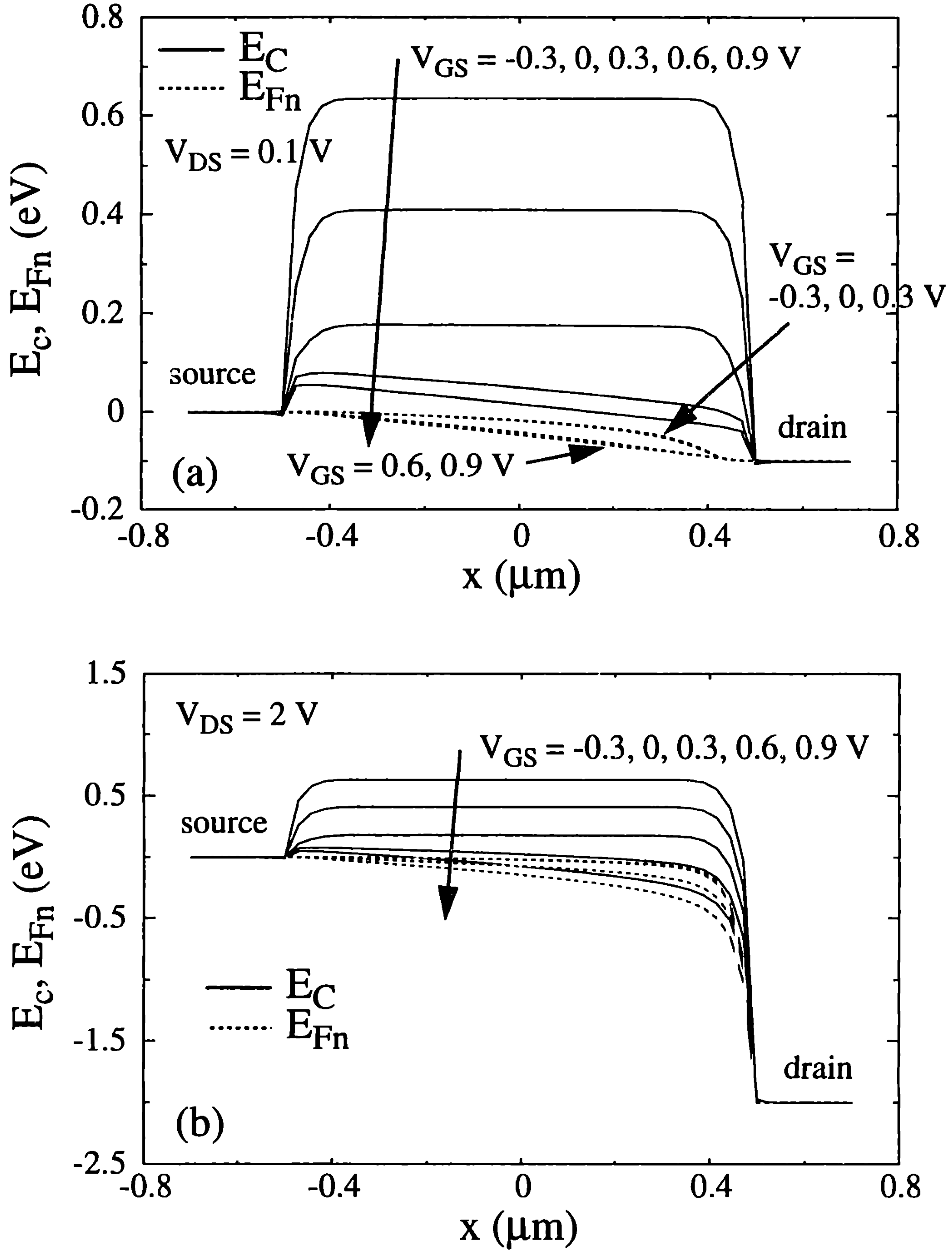

Figure 2.4.3: Simulated electron quasi-Fermi level E_{Fn} (dashed lines), with conduction band E_c (solid lines) of the device of Fig. 2.4.1 at (a) V_{DS} = 0.1 V, and (b) V_{DS} = 2 V, and at V_{GS} in steps of 0.3 V.

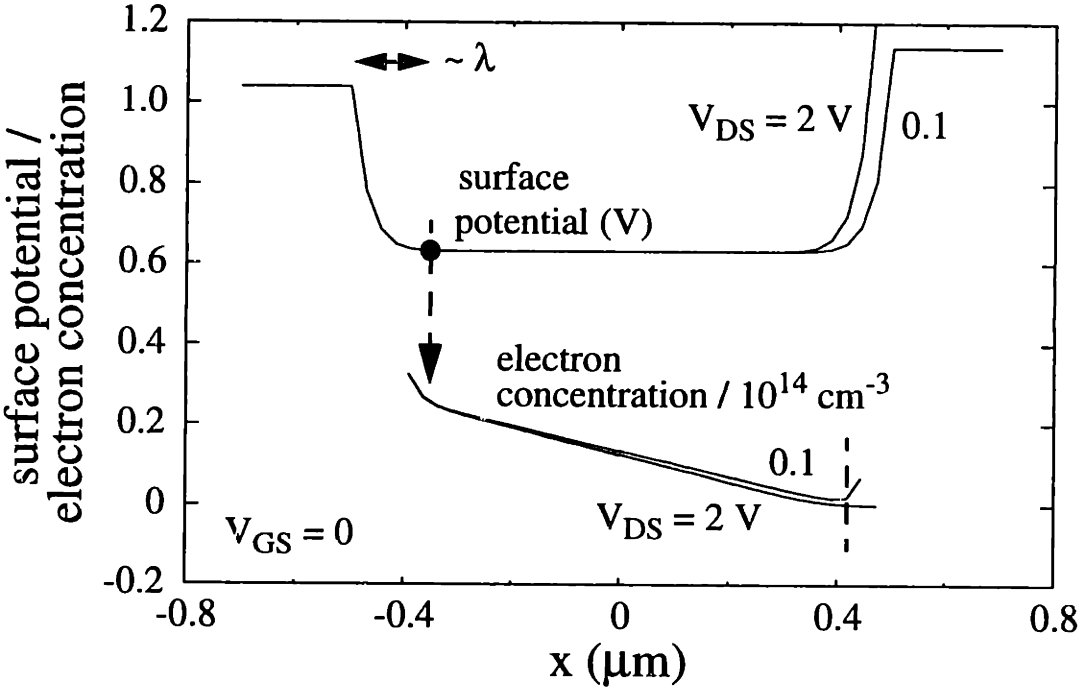

Figure 2.4.4: Simulated surface potential distribution and surface electron concentration of a 1 μm channel-length device at $V_{GS} = 0$, with $V_{DS} = 0.1$ V and 2 V.

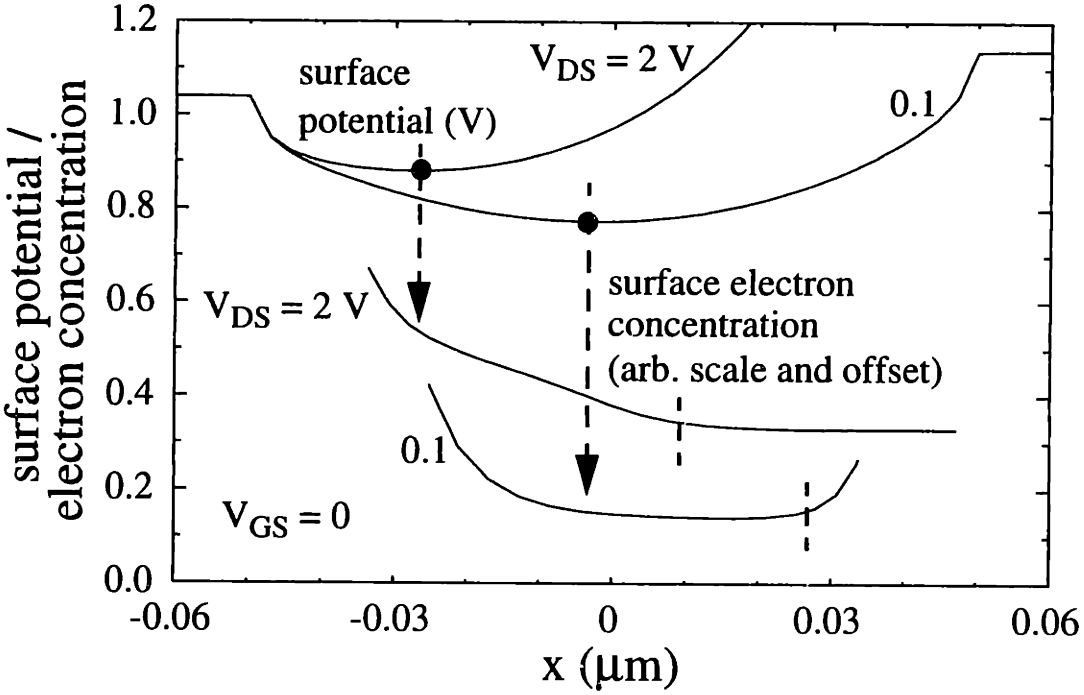

Figure 2.4.5: Simulated surface potential distribution and surface electron concentration of a 100 nm channel-length device at $V_{GS} = 0$, with $V_{DS} = 0.1$ V and 2 V.

Figure 2.4.5 shows the simulated surface potential distribution and surface electron concentration of a 100 nm channel-length device. Except for the channel-length, all device parameters including N_A and T_{ox} are identical to the 1 μm channel-length device discussed earlier. Plots of surface potential distribution, surface electron concentration, and conduction band with quasi-Fermi level are given in Figs. 2.4.6 - 2.4.8. As can be seen in Fig. 2.4.5, no extended flat region in the surface potential distribution as in long-channel devices exists here because the lateral extension of the source potential (i.e., λ) merges with that of the drain. Also, the surface potential in the channel region is generally increased, especially when V_{DS} is increased, in contrast to long-channel devices in which the surface potential distribution does not depend on V_{DS} (see Fig. 2.4.4 for comparison). Furthermore, the distance from the source to the first occurance of the region in surface potential distribution in which the slope is zero reduces with V_{DS} (i.e., the first point of the peak of the barrier moves towards the source as V_{DS} is increased), again in contrast to long-channel devices in which the barrier does not move with V_{DS} (Fig. 2.4.5 - 2.4.6).

From Fig. 2.4.5, the same three regions as in long-channel devices can still be identified. Near the source/channel transition, both drift and diffusion components exist, with the diffusion component dominating so that there is a net current flow. At around the peak of the barrier, the slope of the surface potential distribution is approximately zero, so that diffusion dominates the transport in this region, as evident by the region of constant slope in electron concentration at both $V_{DS} = 0.1$ V and $V_{DS} = 2$ V. At the end of the "diffusion-dominated" region, a drift component sets in and the electron concentration gradient changes. As in the 1 μm channel-length device, the length of the diffusion region is shorter than the source-to-drain distance, and this discrepancy becomes large as the device becomes short. The beginning of the "diffusion-dominated" region also starts at the first occurance of the peak of the potential barrier (denoted by solid circles), similar to long-channel devices. Consequently, as in long-channel devices, the subthreshold current in short-channel MOSFETs is also proportional to the electron concentration near the peak of the potential barrier (i.e., the point at which the slope of the surface potential is zero).

Although the transport mechanisms are very similar, there are important differences. Because the surface potential in the channel region is increased, in particular the potential at the barrier (i.e., reduced barrier height), as V_{DS} is increased, the electron concentration at the barrier will also be increased. Consequently the current also increases. Due to the exponential dependence of electron concentration on the surface potential (or barrier height), the current in a short-channel device can become strongly dependent on V_{DS}. This phenomenon, referred to as drain-induced-barrier-lowering (DIBL) [50-59], is a major obstacle limiting the operation of modern-day short-channel devices, but it can be used as a vehicle in obtaining 2D doping profiles in subthreshold I-V inverse modeling (discussed in Chapters 3 and 4). As seen in Fig. 2.4.5, at $V_{DS} = 2$ V, the potential minimum is increased by about 0.1 V compared to when $V_{DS} = 0.1$ V. In fact, this difference will be even larger at higher V_{GS}. Besides the current sensitivity on V_{DS}, another consequence of reducing the source-to-drain distance is the threshold voltage roll-off (V_t roll-off). Decreasing the channel-length to within 2λ in short-channel devices raises the surface potential minimum even at $V_{DS} = 0$, so that a smaller V_{GS} is needed to bring the device into inversion. In other words, the threshold voltage is reduced. As seen in Fig. 2.4.5, the surface potential minimum in the 100 nm channel-length device is increased by about 0.2 V over the 1 μm channel-length device of Fig. 2.4.4. Examination of Fig. 2.4.7 indicates that the surface electron concentration in the 100 nm channel-length device is about two orders of magnitude higher even at $V_{DS} = 0.1$ V.

From a simple analytical model [58-59] of a MOSFET having a uniform channel doping concentration and constant depletion depth, the surface potential distribution $\Phi_s(x)$ is given by:

$$\Phi_s(x) = \Phi_{sL} + (V_{bi} + V_{DS} - \Phi_{sL})\frac{\sinh\left(\frac{x}{l}\right)}{\sinh\left(\frac{L}{l}\right)} + (V_{bi} - \Phi_{sL})\frac{\sinh\left(\frac{L-x}{l}\right)}{\sinh\left(\frac{L}{l}\right)}, \qquad (2.4.1)$$

where Φ_{sL} is the surface potential if the device were very long, V_{bi} is the built-in potential of the source/drain junctions, L is the channel-length, and l is given by:

$$l = \sqrt{\frac{\varepsilon_{Si} T_{ox} W_D}{\varepsilon_{ox} \eta}}, \qquad (2.4.2)$$

where W_D is the depletion depth, and η is a fitting parameter. Note that l represents a length scale that describes when short-channel effects become important as L is reduced. It can be seen that when W_D is small, l is also small, so that $\Phi_s(x)$ approaches the constant Φ_{sL}, and the device behavior approaches that of a long-channel device (e.g., small DIBL and V_t roll-off). On the other hand, when W_D is large, l also becomes large, and $\Phi_s(x)$ deviates greatly from Φ_{sL}. The device therefore exhibits large amounts of short-channel effects. According to this model, the V_t roll-off for $l << L$ (i.e., device not extremely short) is given as:

$$\Delta V_t = [3(V_{bi} - \Phi_{sL}) + V_{DS}]\exp\left(-\frac{L}{l}\right) + 2\sqrt{(V_{bi} - \Phi_{sL})(V_{bi} - \Phi_{sL} + V_{DS})}\exp\left(-\frac{L}{2l}\right). \tag{2.4.3}$$

Since ΔV_t represents a shift of the I_D vs. V_{GS} curves, eqn. (2.3.18) can be extended to include short-channel devices by incorporating this information, so that I_D becomes proportional to:

$$I_D \propto \mu_n \exp\left(-\frac{V_x}{n\phi_T}\right)\exp\left(\frac{\Delta V_t}{n\phi_T}\right). \tag{2.4.4}$$

The first exponential factor corresponds to the dependence on doping concentration in a long-channel device, while the second exponential factor corresponds to the perturbation due to short-channel effects. As discussed earlier, even for a long-channel device, I_D is very sensitive to the doping concentration. This is because V_x, which contains a term proportional to $(N_A T_{ox}^2)^{1/2}$, is inside the exponential. For a short-channel device, it can be seen that ΔV_t, which is already inside the exponential of eqn. (2.4.4), in turn depends on l (which is related to N_A and T_{ox}) through another exponential [eqn. (2.4.3)]. This double exponential dependence makes I_D extremely sensitive to the doping concentration. On the other hand, I_D is only linearly proportional to μ_n. As explained in section 2.3, the high sensitivity on doping concentration but only linear dependence on μ_n enables inverse modeling to be unaffected by uncertainties in μ_n (see discussion in Chapter 4).

Another common short-channel phenomenon is punchthrough [60]. In short-channel devices having low channel doping concentrations and deep source/drain junctions, punchthrough can occur, signified by a rapid change in subthreshold slope as a function of V_{DS}. This condition is a result of the 2D solution of Poisson's equation. A schematic diagram explaining this phenomenon is shown in Fig. 2.4.9. At low V_{GS} and high V_{DS} (subthreshold region), the potential near the surface of the silicon is low (i.e., controlled by the gate) while the drain potential forces the potential deeper from the surface to be high (i.e., controlled by the drain), and at even larger depth, the quasi-neutral region of the substrate forces the potential to be low again (controlled by the substrate contact). As a result, the potential variation in the depth direction has a peak (i.e., lowest potential energy) that is away from the surface of the silicon. The diffusion current that flows from the source to the drain therefore travels deep from the surface and follows this path of lowest potential energy. Simulations showing the 2D potential distribution and current paths in a device where punchthrough occurs (channel-length = 120 nm, $N_A = 10^{17}$ cm^{-3}, T_{ox} = 5 nm, and source/drain junction depth = 50 nm), at V_{DS} = 2 V and V_{GS} = -0.6 V, are given in Figs. 2.4.10 and 2.4.11 respectively. The simulated subthreshold I-V characteristics are shown in the inset of Fig. 2.4.11. In Fig. 2.4.10, there is a broad potential peak in the y-direction that is centered at approximately y = 0.08 μm. The current flow is therefore centered at this depth, as shown in Fig. 2.4.11. As V_{GS} becomes higher, the gate progressively gains more control of the channel by pulling the potential peak towards the surface, and eventually punchthrough stops. The signature of the subthreshold I-V characteristics is a result of the gate modulation of the current paths. Note that at high channel doping levels the potential in the depth direction is tightly coupled to the gate and the substrate so that it is more difficult for the drain potential to penetrate the channel region. As a result, the potential peak remains at the surface under normal biases, and punchthrough does not occur.

2.5 Conclusion

In this chapter the basics of MOSFET operation are reviewed, with emphasis on the subthreshold behavior. short-channel effects, including drain-induced-barrier-lowering (DIBL), threshold voltage roll-off, and punchthrough are covered.

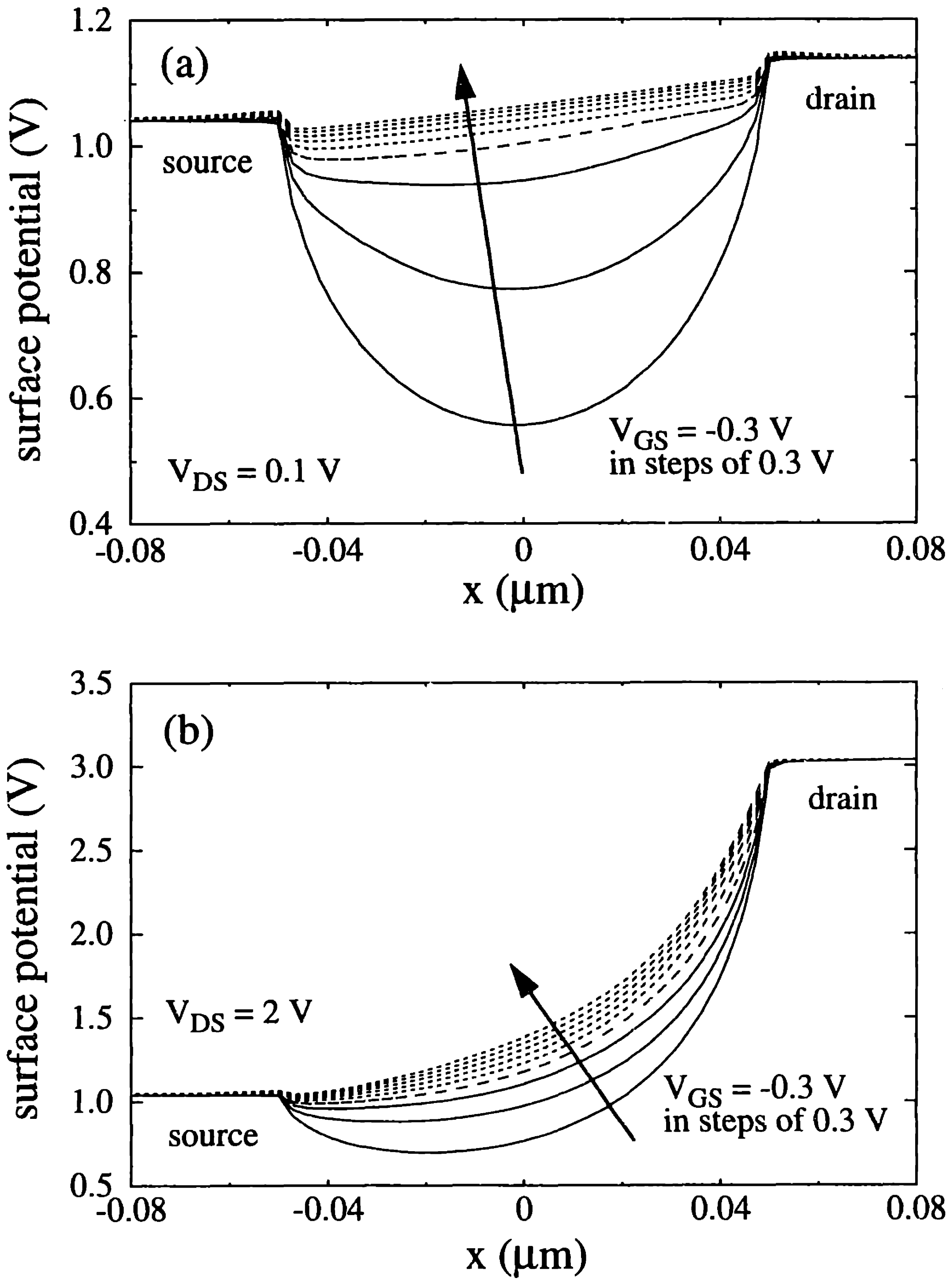

Figure 2.4.6: Simulated surface potential distribution (referenced to the substrate) of a 100 nm channel-length NMOSFET having $N_A = 3\times10^{17}$ cm^{-3} and T_{ox} = 5 nm at (a) V_{DS} = 0.1 V, and (b) V_{DS} = 2 V, and at V_{GS} in steps of 0.3 V. Approximately, the weak inversion (subthreshold) region is denoted by solid lines, while the strong inversion region is denoted by dashed lines.

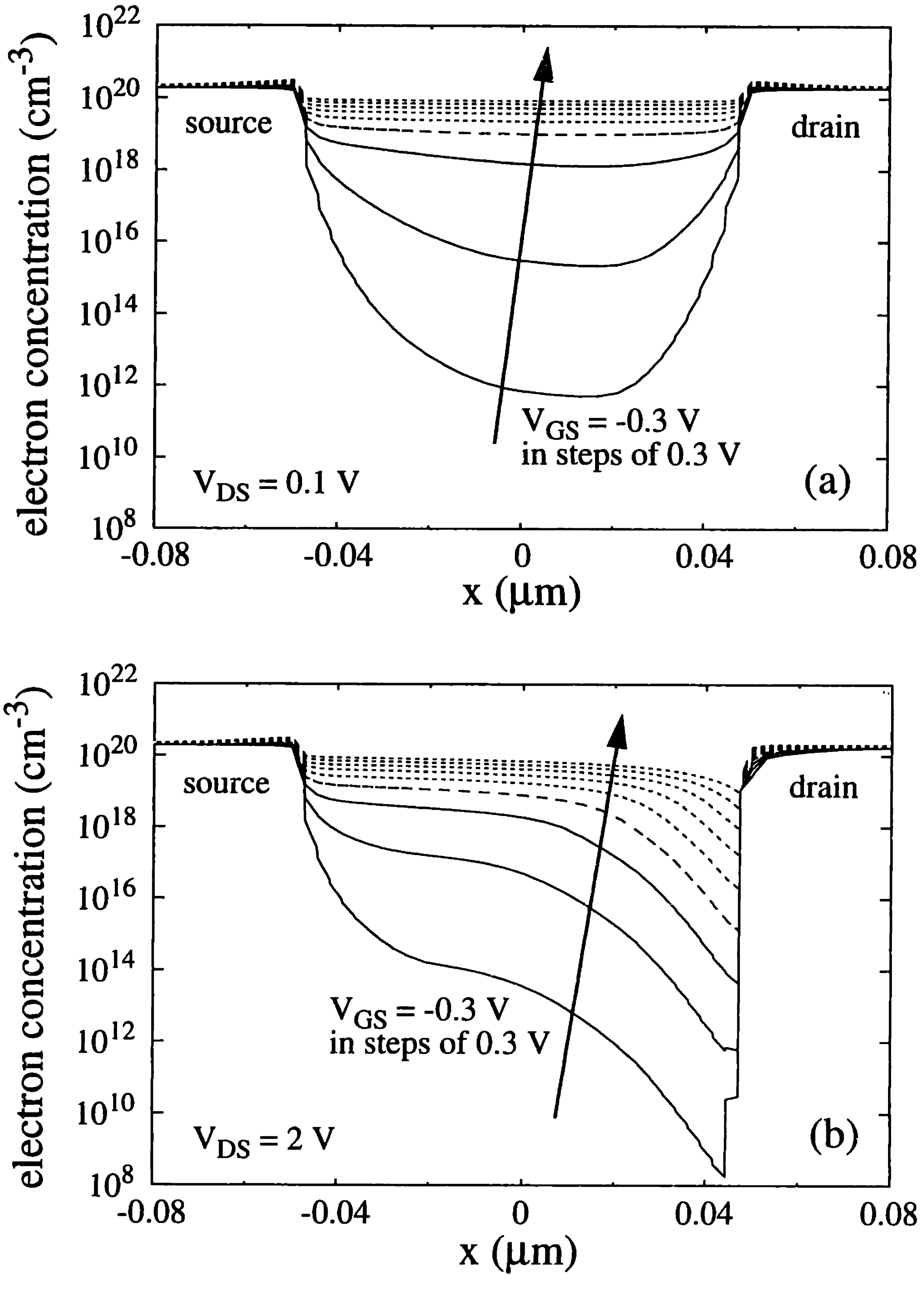

Figure 2.4.7: Simulated surface electron concentration of the device of Fig. 2.4.6 at (a) V_{DS} = 0.1 V, and (b) V_{DS} = 2 V, and at V_{GS} in steps of 0.3 V. Approximately, the weak inversion (subthreshold) region is denoted by solid lines, while the strong inversion region is denoted by dashed lines.

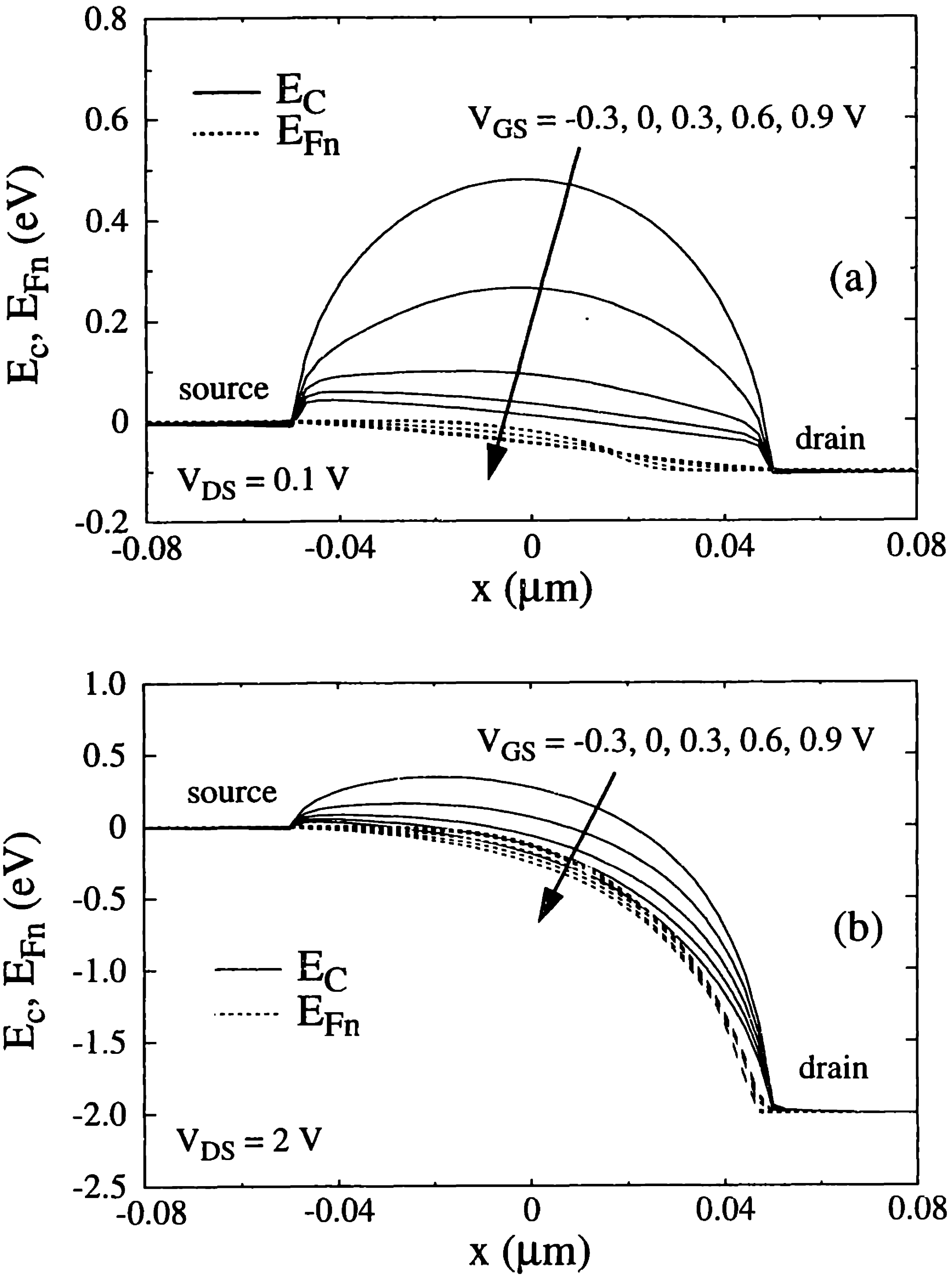

Figure 2.4.8: Simulated electron quasi-Fermi level E_{Fn} (dashed lines), with conduction band E_c (solid lines) of the device of Fig. 2.4.6 at (a) V_{DS} = 0.1 V, and (b) V_{DS} = 2 V, and at V_{GS} in steps of 0.3 V.

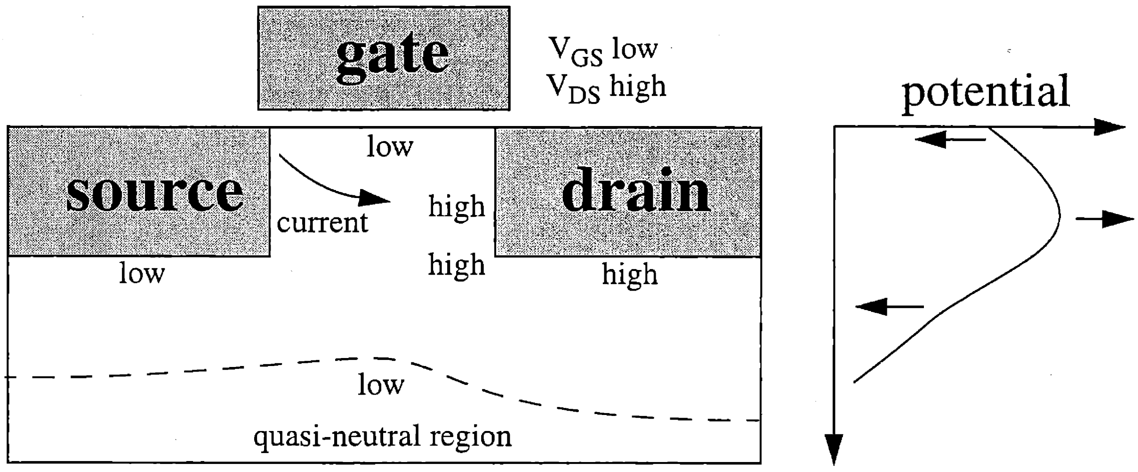

Figure 2.4.9: Schematic diagram showing how and why punchthrough occurs. When punchthrough happens, the potential has a peak occuring at a larger depth away from the surface, so that current travels at a depth away from the surface.

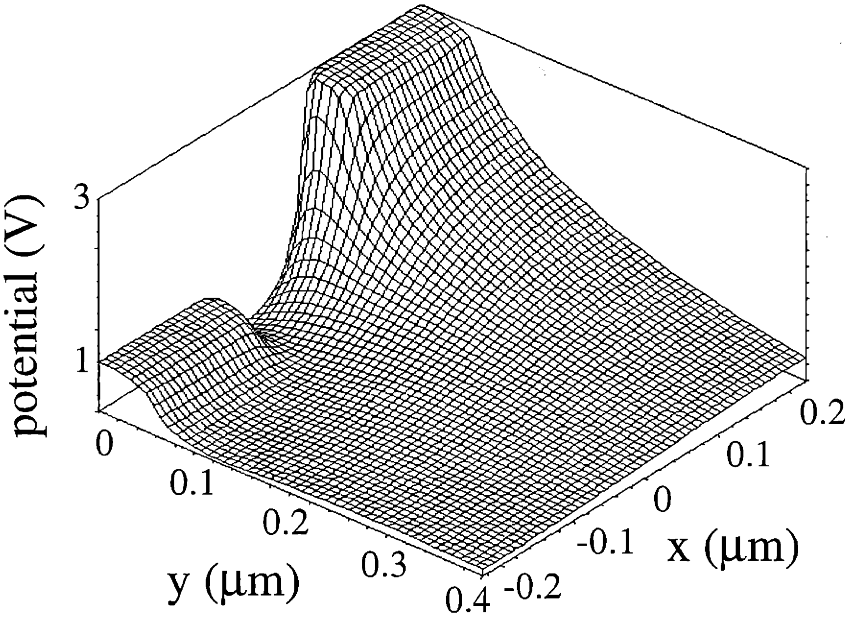

Figure 2.4.10: Simulated potential distribution in a device experiencing punchthrough, at V_{DS} = 2 V and V_{GS} = -0.6 V. In the channel region, the potential at the surface is lower than that deeper from the surface, with a broad peak (in the y-direction) centered at approximately y = 0.08 μm.

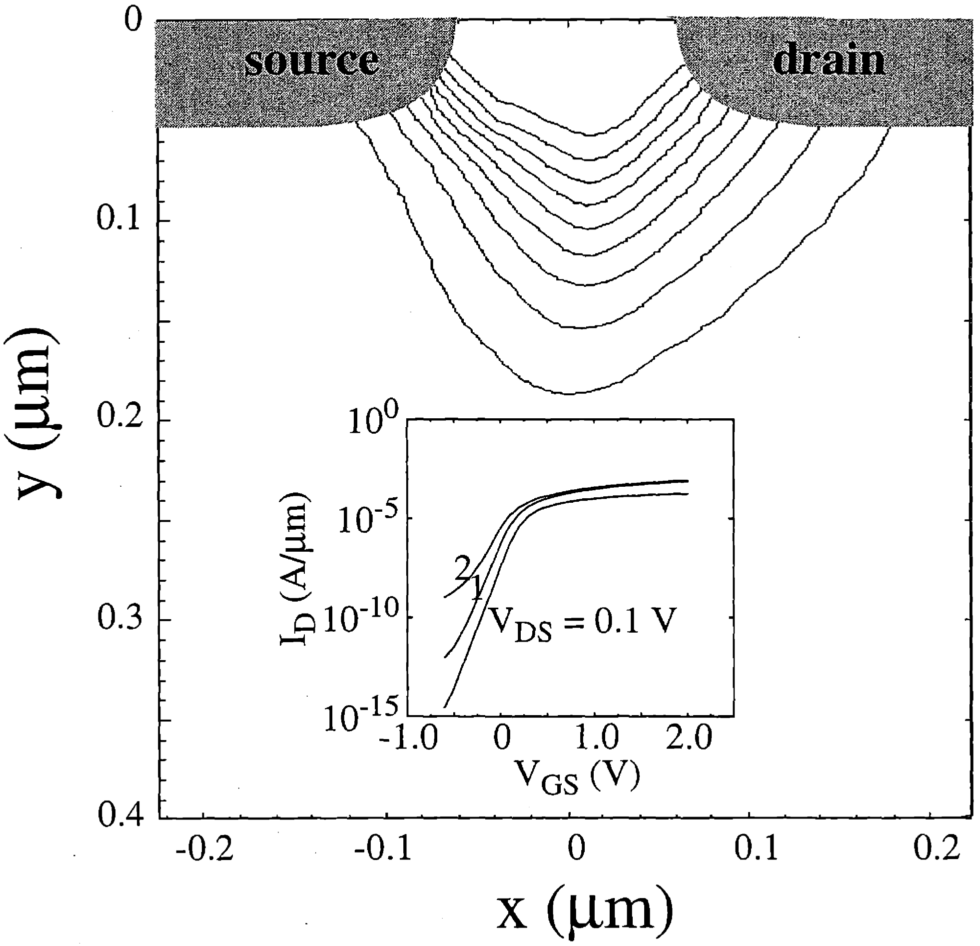

Figure 2.4.11: Simulated current paths at V_{DS} = 2 V and V_{GS} = -0.6 V. In the channel region, the potential distribution has a broad peak (in the y-direction) centered at approximately y = 0.08 μm. Consequently the current paths are centered at this depth. The punchthrough current, most evident at V_{DS} = 2V, is shown in the inset.

Chapter 3

How Does Subthreshold I-V Inverse Modeling Work?

3.1 Introduction

Subthreshold I-V inverse modeling is based on obtaining a 2D doping profile such that the simulated subthreshold I-V characteristics, over a broad range of bias conditions (i.e., V_{GS}, V_{DS}, and V_{BS}), match the corresponding experimental data. The basic tool required is a 2D device simulator embedded in a non-linear least-square optimization loop which minimizes the RMS error between the simulated and experimental data. The 2D device simulator used in this work is MEDICI [43], while the optimizer is a custom program implementing the Levenberg-Marquardt [61-62] non-linear least square algorithm incorporating a Broyden [63] scheme for Jacobian approximation. A flow chart showing the inverse modeling process is given in Fig. 3.1.1.

In the inverse modeling process, an initial guess of the doping profile is first provided by the user. The device simulator then generates a set of subthreshold I-V curves, which are fed into the optimizer. The optimizer compares the difference between the simulated and experimental data (i.e., the RMS error) and calculates the doping profile needed to minimize this difference. This process is iterated until the convergence criteria are met. Details of the optimization process are provided in Appendix A.

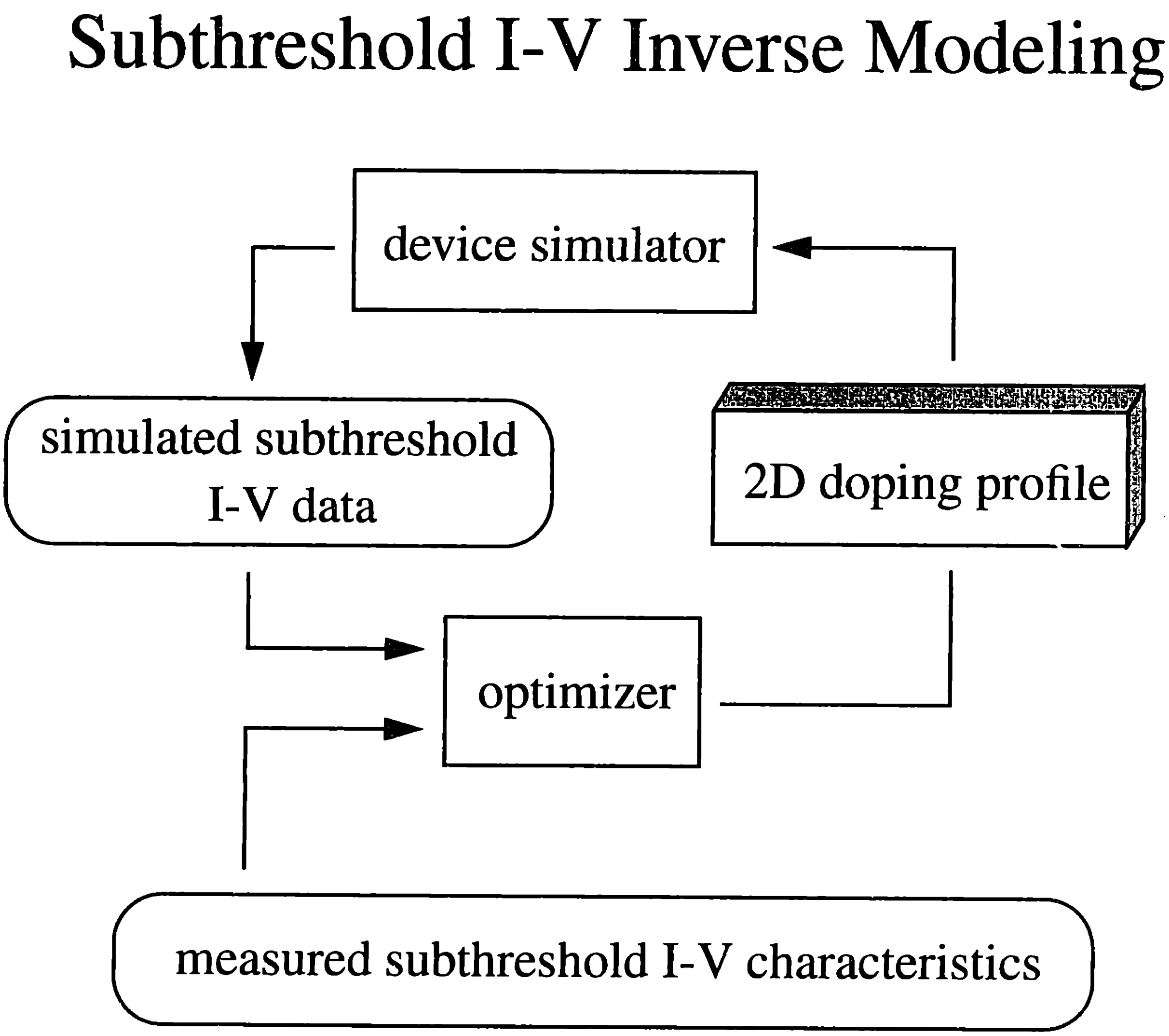

Figure 3.1.1: Flow chart showing the process of subthreshold I-V inverse modeling.

3.2 Parameterization of Doping Profile

The 2D doping profiles are represented using a sum of 2D Gaussian functions (representing the source/drain/halo), and a 1D B-spline or Gaussian function (representing the depth-wise variation of the channel profile not included in the 2D Gaussian functions) [64]. Other basis functions such as the complementary error function and exponential function may also be used, although it was found that the Gaussian functions often resulted in the best results (i.e., lowest RMS errors). During optimization, the parameters representing the 2D profile are varied until a best-fit is achieved. Since the subthreshold I-V characteristics are primarily a manifestation of the device electrostatics, which are determined by dopings in the channel and near source/drain extensions, the deep source/drain junctions commonly used to form the source/drain contact areas are not included in the simulations. The only structural information needed to be known *apriori* includes the gate width, gate dielectric (oxide) thickness, and dielectric constant. However, as will be discussed in section 4.4 of Chapter 4, the gate width need not be specified very accurately for accurate results. Moreover, the channel-length, which is determined by the 2D doping profile, is automatically obtained in the optimization. The gate electrode length, which has little effect on the subthreshold I-V characteristics, need not be known.

3.3 Equivalent Oxide Thickness

For the present work, the gate material is assumed to be metallic. Any polysilicon gate depletion as well as quantum mechanical effects in the inversion layer are absorbed into an equivalent oxide thickness $T_{ox,\ eqv}$, obtained by extrapolating regular C-V data to the weak-inversion region, as shown schematically in Fig. 3.3.1. The use of $T_{ox,\ eqv}$ allows for a first order approximation of polysilicon depletion and quantum mechanical effects. This is a reasonable approximation especially if the effects are not very pronounced (e.g., oxide not too thin). More advanced treatments should be exercised if more accurate simulations are required, particularly for very thin oxides (see, for example, [65-68]).

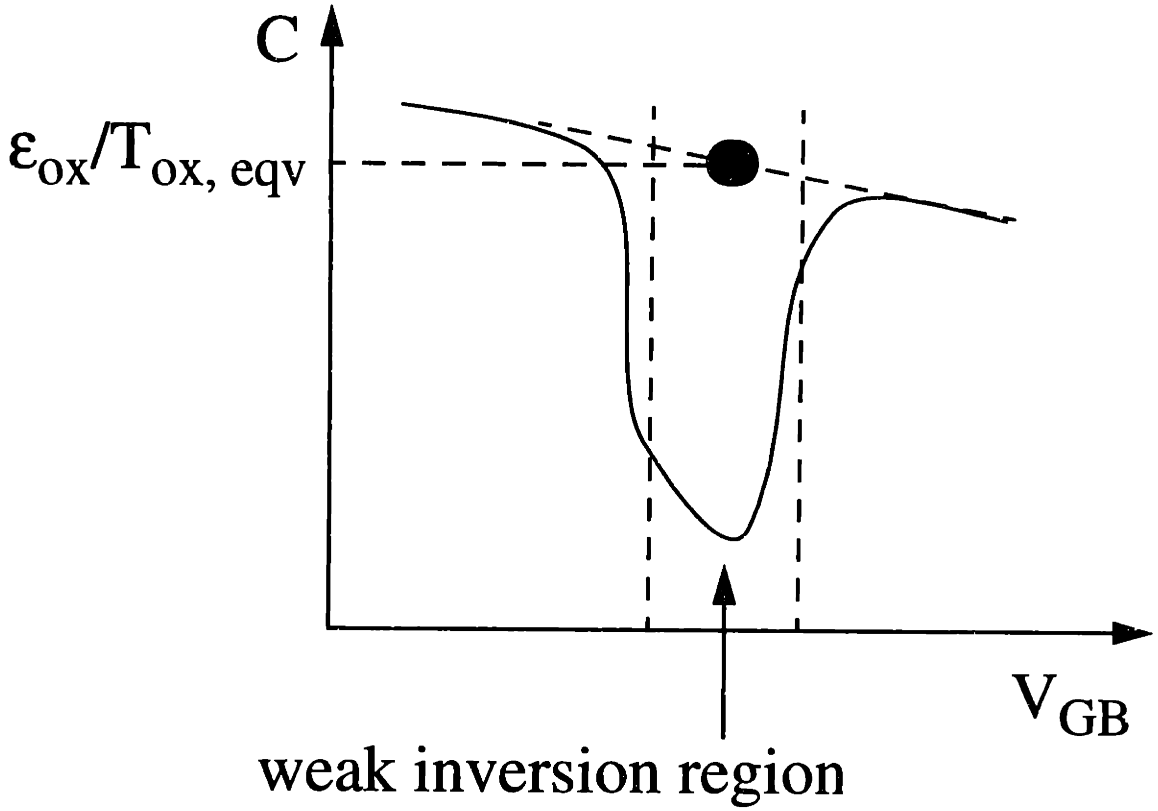

Figure 3.3.1: Schematic diagram illustrating the concept of equivalent oxide thickness $T_{ox,eqv}$, used in approximating polysilicon gate depletion and quantum mechanical effects in thin oxides. $T_{ox,eqv}$ is obtained by extrapolating C (capacitance normalized to the gate area) to the weak inversion region where the simulations are carried out. ε_{ox} denotes the oxide dielectric constant.

3.4 Subthreshold I-V Inverse Modeling

For a MOSFET in weak-inversion, the drain current I_D is very sensitive to the 2D electrostatic potential distribution in the depletion region of the device channel, which in turn depends on the applied potential at the source, drain, gate, and substrate in ways that are strongly determined by the 2D dopant distribution. In this section, the theoretical aspects of 1D doping profile extraction by subthreshold I-V inverse modeling is provided, followed by an extension to 2D.

3.4.1 Inverse Modeling in One Dimension

It is a well known fact that 1D channel doping profiles of long-channel MOSFETs can be obtained from the change in V_{GS} required to maintain a constant I_D as V_{BS} is changed [69-71]. In a long-channel device, where short-channel effects can be ignored, charge balance between the gate and the silicon requires that

$$Q_G = Q_I + Q_B, \tag{3.4.1.1}$$

where Q_G, Q_I, and Q_B are, respectively, the gate charge, inversion charge, and bulk depletion charge. If the inversion charge is maintained constant (so that I_D remains constant), then

$$\frac{dQ_G}{dV_{GS}} = \frac{dQ_B}{dV_{GS}} \text{ and } \frac{dQ_G}{dV_{BS}} = \frac{dQ_B}{dV_{BS}}. \tag{3.4.1.2}$$

Using depletion approximation, the following results are obtained [70-71]:

$$Q_G = \frac{\varepsilon_{ox}}{T_{ox}}(V_{GS} - V_{FB} - \Phi_s), \tag{3.4.1.3}$$

$$Q_B = qNW_D, \tag{3.4.1.4}$$

$$\Phi_s - V_{BS} = \frac{qN}{2\varepsilon_{Si}}W_D^2. \tag{3.4.1.5}$$

Here, V_{FB}, Φ_s, N, W_D, ε_{ox}, ε_{Si}, and T_{ox} are, respectively, the flat-band voltage, surface potential, doping concentration, depletion depth, oxide dielectric constant, silicon dielectric constant, and oxide thickness. Using eqns. (3.4.1.2) - (3.4.1.5), and upon some differentiations and algebraic manipulations, the followings can be obtained [70-71]:

$$W_D = \frac{\varepsilon_{Si}}{\varepsilon_{ox}}T_{ox}\left|\frac{dV_{BS}}{dV_{GS}}\right|, \tag{3.4.1.6}$$

$$N(W_D) = \frac{1}{q\varepsilon_{Si}}\left(\frac{\varepsilon_{ox}}{T_{ox}}\right)^2\left|\frac{d^2V_{BS}}{dV_{GS}^2}\right|^{-1}. \tag{3.4.1.7}$$

In the derivation above, it is assumed that $d\Phi_s/dV_{GS} = d\Phi_s/dV_{BS} = 0$, which, physically, means that changing V_{GS} to compensate for changes in V_{BS} to maintain constant I_D does not affect the band bending near the surface. In effect, changing V_{BS} changes the amount of bulk charge, which causes the inversion charge to change in order to maintain charge balance with the gate, and in order for the same amount of inversion charge to be maintained, V_{GS} also has to be changed. The incremental amount of change in V_{GS} per incremental amount of change in V_{BS} in order to maintain constant I_D is therefore related to the incremental amount of bulk charge that is added or subtracted at the depletion depth. Consequently this method can be considered as an I-V analogue of the C-V profiling method [32], which also samples the incremental charge at the depletion depth.

There are limitations of the method described above, however. For instance, the depletion approximation falls apart near the surface of the silicon. As shown in [70], the minimum depth that can be probed is approximately $3\lambda_D$, where λ_D is the extrinsic Debye length. More accurate profiles can be obtained using numerical simulation, through inverse modeling, which is the focus of this thesis.

In subthreshold I-V inverse modeling, a set of numerically simulated subthreshold I_D vs. V_{GS} vs. V_{BS} curves are fitted to the corresponding experimental data by optimizing for the 1D doping profile. Besides the fact that no approximations are needed, inverse modeling is more accurate because a set of curves is used to obtain a profile, as opposed to only a set of points, namely, a set of V_{GS} vs. V_{BS} points for a constant I_D. Inverse modeling also automatically optimizes for the subthreshold slope, which is a function of V_{BS} and doping concentration. Note that this is an improvement over fitting only the threshold voltage vs. V_{BS} data [69-71]. Moreover, the inclusion of low current data (low V_{GS} for a n-channel device) allows the doping concentration very near the silicon surface to be sampled, since the depletion depth can be quite small at low V_{GS}, especially when positive V_{BS} is applied.

There is an important difference between subthreshold I-V inverse modeling and the analytical I-V technique discussed above (which is similar to the C-V profiling technique [32] and the 2D C-V technique [36-40]). Note that eqns. (3.4.1.6) and (3.4.1.7) contain only the derivatives of V_{BS} and V_{GS}; the actual values of V_{GS} and V_{BS} are unimportant. In constructing a profile using the analytical I-V technique, a range of V_{BS} is used, and the depletion depth is swept from a small range to a large range, so that the profile is constructed incrementally from near the surface to the maximum depletion depth. Unless zero depletion depth is reached (which, in practice, is impossible for the I-V technique), the doping concentration at less than the minimum depletion depth cannot be determined. In subthreshold I-V inverse modeling, however, both V_{BS} and V_{GS} are also matched to the experimental data (e.g., including threshold voltages). As a result, in addition to building up the profile incrementally from near the depletion edge, the integrated profile from the surface to the depletion depth is also automatically matched. Therefore, in situations where the minimum depletion depth cannot reach the surface (such as due to current leakage), the profile within the minimum depletion depth, when integrated from the surface, is still correctly obtained. An example of I_D vs. V_{GS} vs V_{BS} characteristics of a device having the channel doping profile given in the inset is shown in Fig. 3.4.1.1.

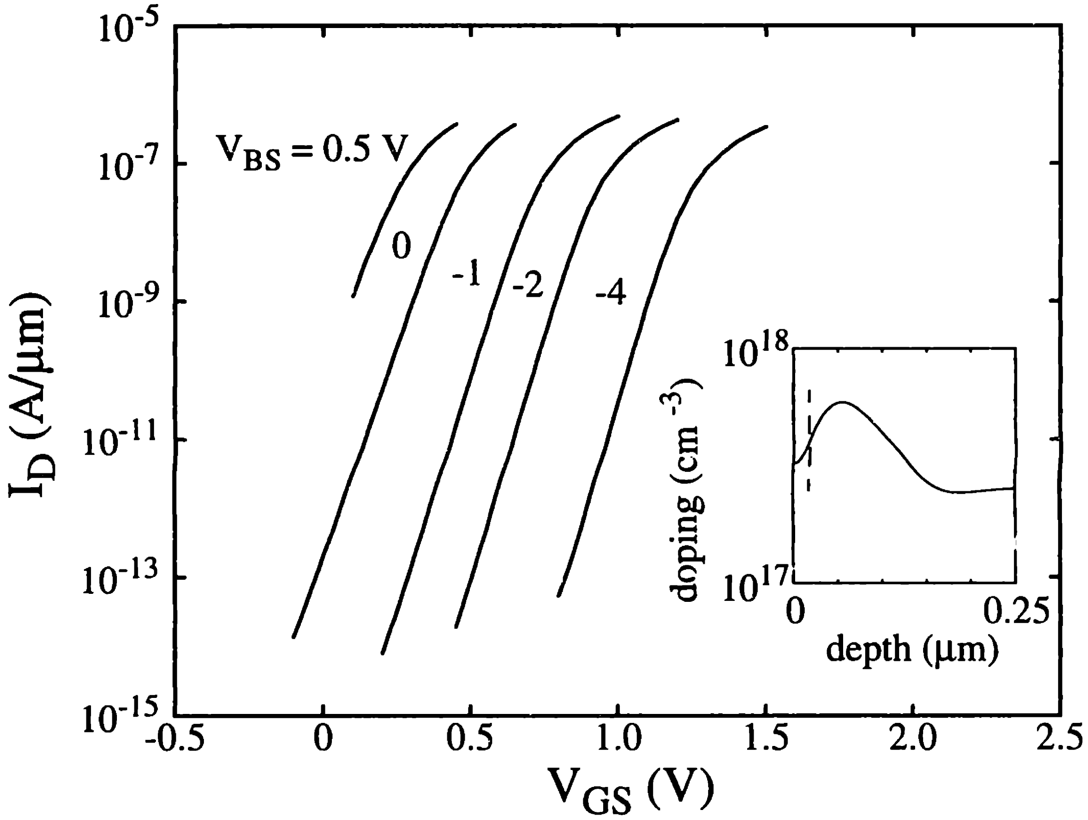

Figure 3.4.1.1: I_D vs. V_{GS} vs. V_{BS} characteristics of a 5 μm channel-length device. The channel doping profile is shown in the inset. The dashed line represents the minimum depletion depth in the data.

3.4.2 Inverse Modeling in Two Dimensions

The subthreshold I-V inverse modeling technique presented above can be extended to two dimensions. Figure 3.4.2.1 shows a schematic diagram of the depletion region and surface potential distribution of a n-channel MOSFET at a fixed V_{GS}, but at different V_{DS}. As V_{DS} is increased, the depletion region moves away from the drain, so that additional negative charges (i.e., ionized dopants) are exposed near a region close to the depletion edge, while additional positive charges are exposed within the drain depletion region. As the region of the depletion edge is incrementally increased in depth, the surface potential in the same region is also incrementally increased. This is because the surface potential, which is a solution of 2D Poisson's equation, is determined by the depleted charge from the surface to the depletion edge as well as the potential at the drain. If the region near the minimum of the surface potential is perturbed (increased), as in a short-channel device, an incremental increase in inversion charge, and therefore I_D, will result (see Chapter 2). If I_D is to be maintained constant, V_{GS} has to be reduced. Consequently the amount of change in V_{GS} as V_{DS} is changed, in order to maintain a constant I_D, is related to the doping concentration near the perturbed depletion edge, analogous to changing V_{GS} to maintain a constant I_D while changing V_{BS} as in the 1D case. In 2D, however, V_{BS} data also has to be included in order to probe the doping concentration in the depth direction. At this point, if only the changes in bias are considered (not the magnitude), this technique will be the analogue of the 2D C-V technique [36-40]. Beyond this point, however, is where the differences begin.

In 2D subthreshold I-V inverse modeling, a set of numerically simulated subthreshold I_D vs. V_{GS} vs. V_{DS} vs. V_{BS} curves are fitted to the corresponding experimental data by optimizing for the 2D doping profile [41-42]. Similar to 1D subthreshold inverse modeling, the values of I_D, V_{GS}, V_{DS}, and V_{BS} are matched, instead of only their changes. Therefore, the surface potential distribution, which is related to the integrated doping concentration within the depletion region, is also matched. Consequently, in constructing a 2D doping profile, in addition to building up the profile incrementally from near the depletion edge, the profile within the depletion region is also constructed so that the total depleted charge at all bias points and current is matched to the experimental data. Although it is impossible in practice to totally eliminate the depletion region in a

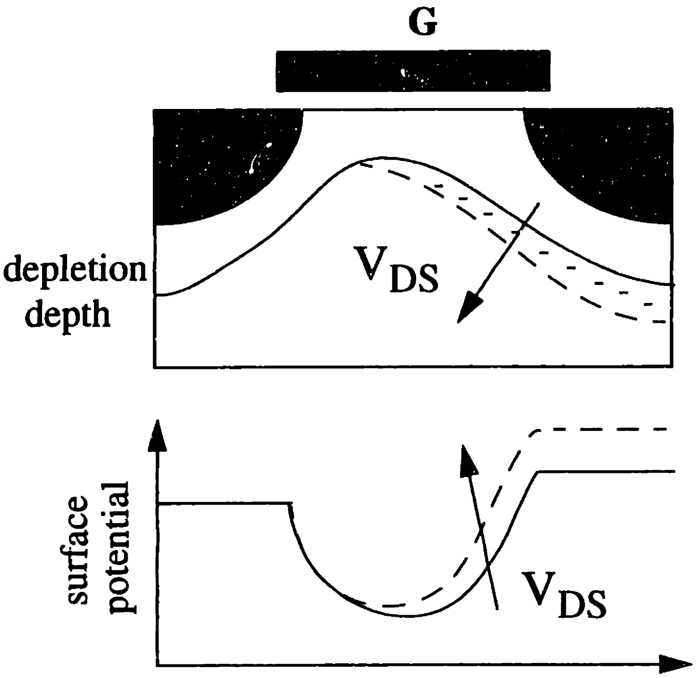

Figure 3.4.2.1: Schematic diagram showing the depletion region and surface potential of a n-channel MOSFET at fixed V_{GS}, but at different V_{DS}. As V_{DS} is increased, additional positive charges are exposed in the drain depletion region (not shown), and additional negative charges are exposed in the channel depletion region. The surface potential near the region where the charge is increased is perturbed.

short-channel device by applying reasonable biases so that details of the depthwise variation of the doping profile very near the surface can be resolved, the doping profile, when integrated from the surface to the minimum achievable depletion depth, can still be obtained correctly.

The construction of the doping profile can be illustrated by examining the depletion region and surface potential under different V_{GS} and V_{DS}, as shown in Figs. 3.4.2.3 and 3.4.2.4. In this 120 nm n-channel MOSFET example, the channel doping concentration is 3×10^{17} cm^{-3}, with an oxide thickness of 5 nm and a lateral doping profile at the silicon surface as shown in Fig. 3.4.2.2. As shown in Fig. 3.4.2.3, for a fixed V_{GS} (-0.25 V), and at low V_{DS} (0.01 V), the depletion region has a "peak" (i.e., smallest depletion depth) near the center of the channel due to symmetry. The surface potential distribution, also due to symmetry, has a minimum near the center of the channel. At the same V_{GS}, but a higher V_{DS} (2 V), the depletion region is changed, and the amount of V_{GS} change needed to maintain the same I_D provides information of the incremental charge that is depleted as well as the depletion depth, analogous to the 2D C-V technique [36-40]. It can be seen that in this range of biases, the position of the surface potential minimum coincides with the "peak" of the depletion region. This is because near the center of the channel (i.e., away from the strong influence of the source/drain), the depleted charge integrated from the surface to the depletion depth is minimum where the depletion depth is minimum. It can be seen that in Fig. 3.4.2.3, similar to the "peak" of the depletion region, the position of the surface potential minimum is shifted towards the source as V_{DS} is increased, allowing the doping concentration near the center of the channel to be probed laterally.

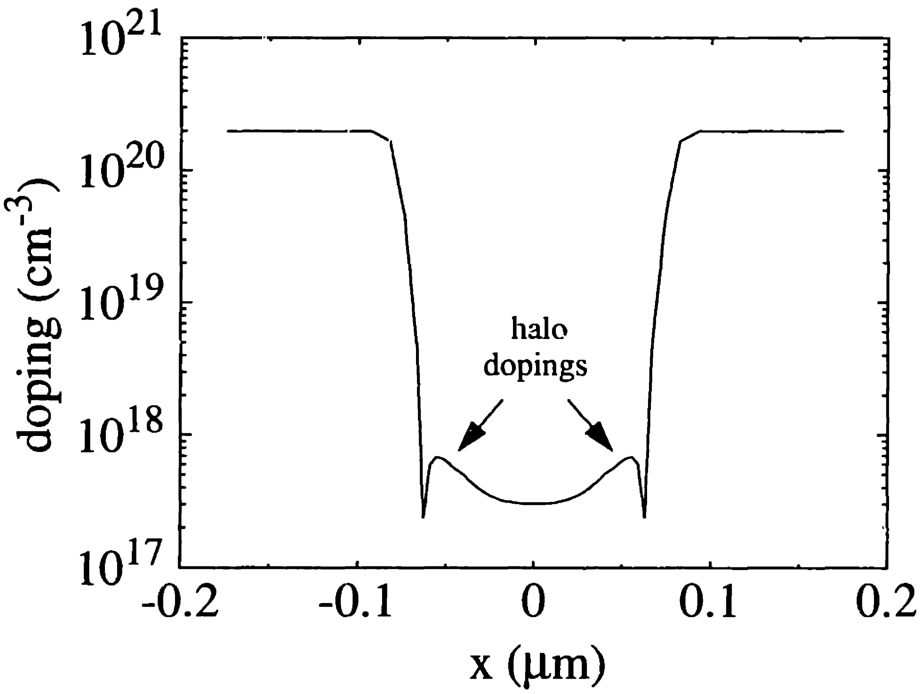

Figure 3.4.2.2: Lateral doping profile at the silicon surface of the n-channel MOSFET example (see text).

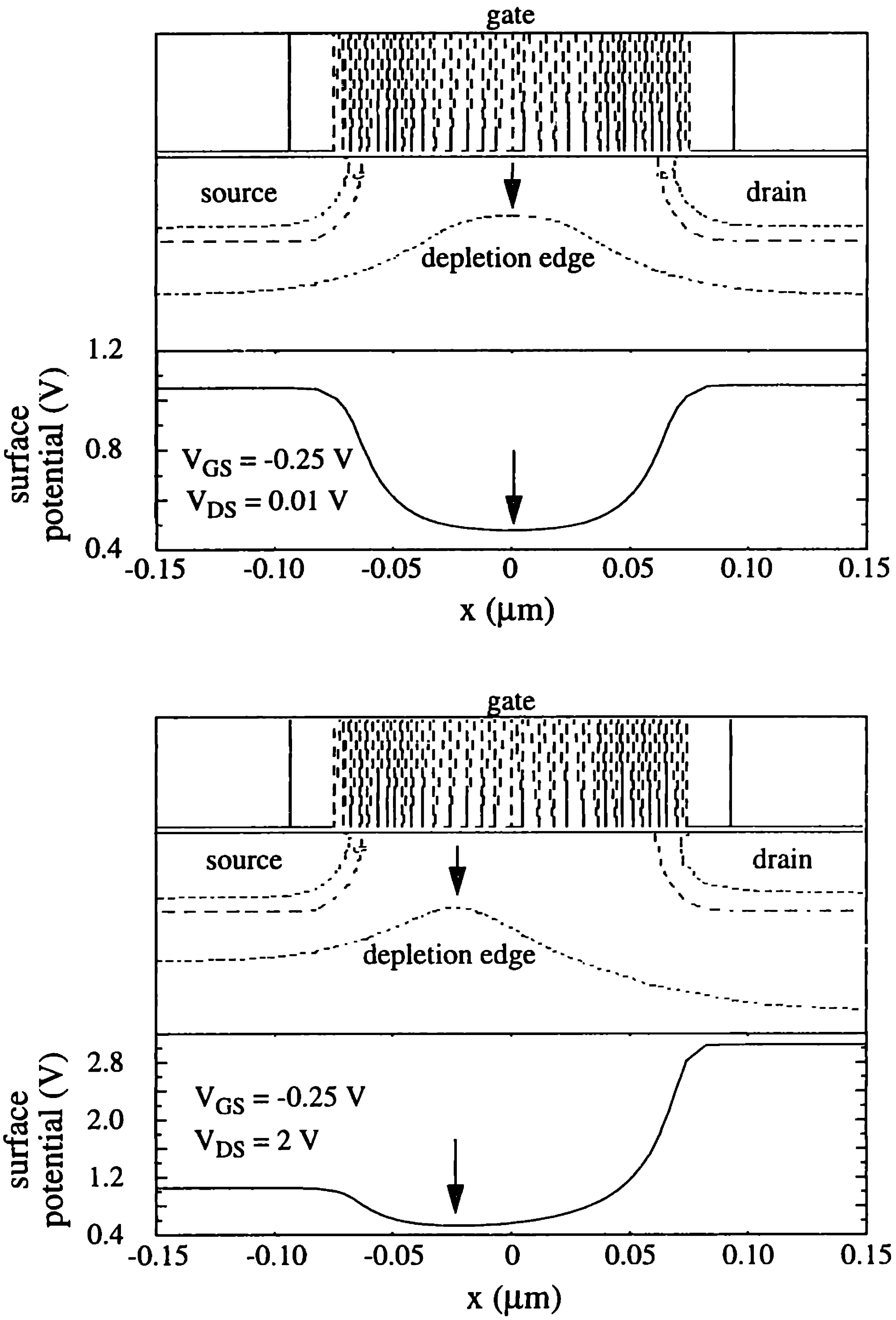

Figure 3.4.2.3: Depletion region and surface potential distribution of the MOSFET example discussed in text, at V_{GS} = -0.25 V, with V_{DS} = 0.01 V and 2 V. The arrows indicate the positions of the depletion region "peak" and surface potential minimum.

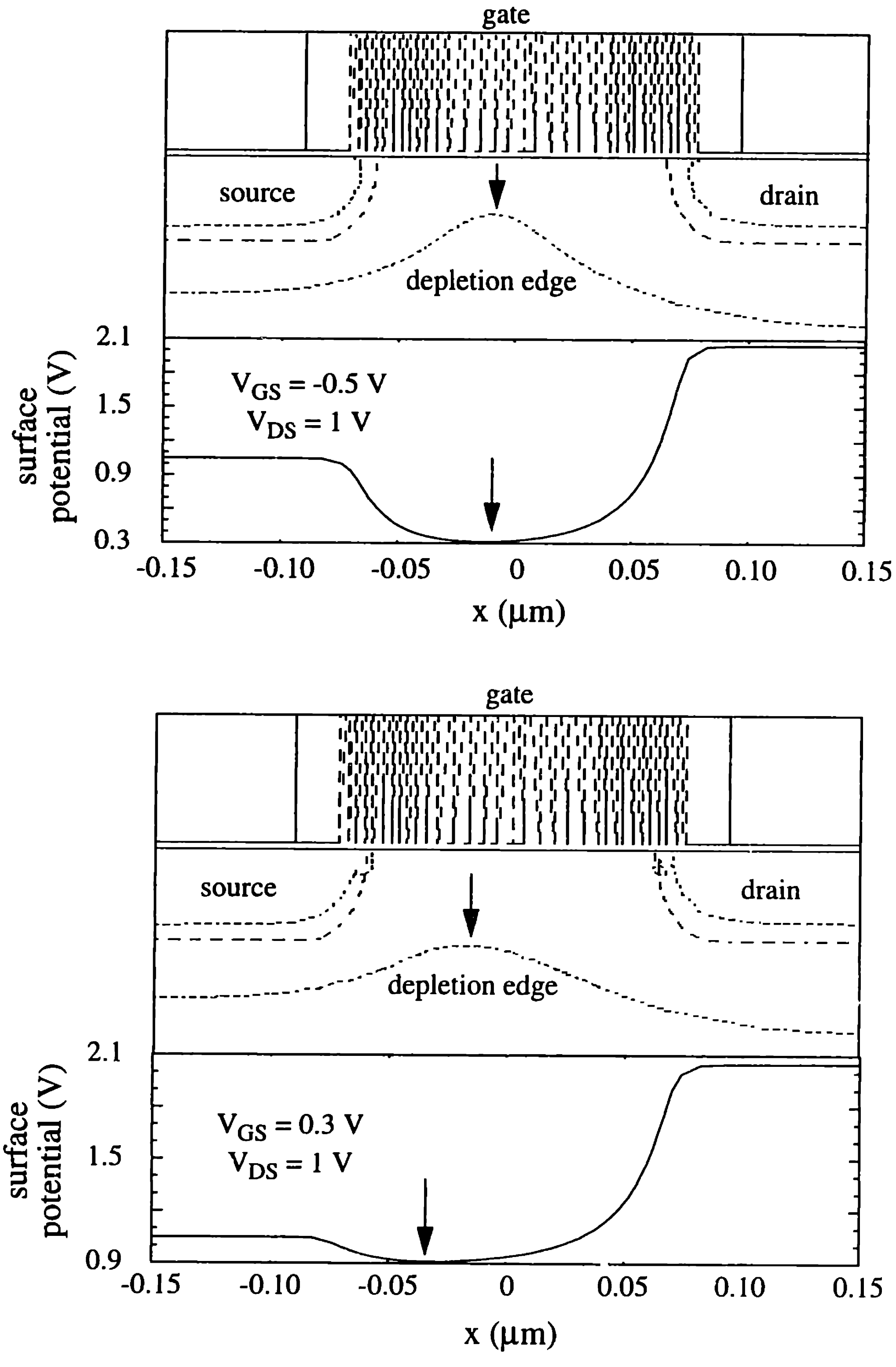

Figure 3.4.2.4: Depletion region and surface potential distribution of the MOSFET example discussed in text, at V_{DS} = 1 V, with V_{GS} = -0.5 V and 0.3 V. The arrows indicate the positions of the depletion region "peak" and surface potential minimum.

Figure 3.4.2.4 shows the depletion region and surface potential distribution of the same device as above, but with $V_{DS} = 1$ V, and at $V_{GS} = -0.5$ V and 0.3 V. At $V_{GS} = -0.5$ V, the depletion region has a "peak" slightly shifted towards the source, but still near the center of the channel. The same also applies to the surface potential minimum, which approximately coincides with the "peak" of the depletion region. As V_{GS} is increased, the depletion region "peak" moves towards the source. More dopants are depleted near the depletion edge, and the incremental charge that gets depleted is related to the incremental change in I_D, providing information of the doping concentration near the depletion edge. This can be thought of as a "lateral body effect". For instance, as the depletion edge sweeps across the halo, the higher doping concentration relative to the center of the channel makes the halo more difficult to deplete, resulting in a higher "body factor" and higher incremental subthreshold swing, affecting the subthreshold slope as well as the general shape of the I-V curve.

In addition to the incremental changes, however, the magnitude of the various biases and current also need to be matched to the experimental data. Consequently the surface potential distribution (and lateral doping profile) also has to be consistent with that suggested by the experimental data. As seen in Fig. 3.4.2.4, as the surface potential minimum moves towards the source, the potential of the source begins to strongly affect the surface potential distribution, so that the surface potential minimum moves towards the source faster than the "peak" of the depletion edge. Because I_D is determined by the minimum of the surface potential distribution (i.e., maximum potential barrier), which, in turn, is determined by the lateral doping profile and various biases through the solution of 2D Poisson's equation, the process of reconstructing the surface potential distribution (through fitting the set of experimental data), and therefore the lateral doping profile within the depletion region, can be likened to solving for the coupled set of equations:

$$\begin{aligned} \Phi_{m111} &= F(\bar{P}; V_{GS1}, V_{DS1}, V_{BS1}) \\ \Phi_{m211} &= F(\bar{P}; V_{GS2}, V_{DS1}, V_{BS1}) \\ \Phi_{m121} &= F(\bar{P}; V_{GS1}, V_{DS2}, V_{BS1}) \\ \Phi_{m112} &= F(\bar{P}; V_{GS1}, V_{DS1}, V_{BS2}) \\ &\cdots\cdots\cdots \end{aligned} \tag{3.4.2.1}$$

where Φ_{m11} ... etc. represent the set of surface potential minima (which are measured through measuring I_D, and are known variables), V_{GS1}, V_{DS1}, and V_{BS1} ... etc. represent the set of applied biases (also known variables), and F represents the surface potential minimum which arises from the 2D solution of Poisson's equation having a known functional form dictated by the parameterization of the 2D doping profile having a finite set of parameters $\overline{P}$ (which are to be solved for). It has to be noted that the solution of Poisson's equation is a very smooth function in space (since the potential is a double integral of the charge in space). Consequently very fine details of the doping profile within the depletion region are susceptible to be averaged out. This, however, may not be a serious issue because in the processing of MOSFETs, the diffusion of dopants also results in smooth doping profiles.

Therefore, in subthreshold I-V inverse modeling, it is the incremental changes (that allow the doping concentration near the depletion edges to be determined) coupled with the absolute values of the current and biases (that allow the doping concentration within the depletion region to be determined) that allows the complete 2D doping profile to be extracted.

From the discussion above, it becomes clear that the DIBL dependence on V_{BS} is very sensitive to the source/drain doping configuration. For example, in a n-channel device, the depletion region shrinks when positive V_{BS} is applied, and expands for negative V_{BS}. When the depletion depth is small (i.e., positive V_{BS}), the quasi-neutral region (i.e., undepleted region) partially restricts the drain electric field from affecting the channel because the depletion edge is an equipotential surface, where the normal component of the electric field vanishes (note that this is different from the plate of a capacitor where surface charge can exist). Consequently the DIBL effect is small. On the other hand, when the depletion depth is large (i.e., negative V_{BS}), the depletion edge is far away from the surface and the source/drain junctions, so that the shielding effect is diminished, resulting in large DIBL. A simple analytical model [59] of a MOSFET having a uniform channel doping concentration and constant depletion depth shows that the surface potential distribution $\Phi_s(x)$ is given by:

$$\Phi_s(x) = \Phi_{sL} + (V_{bi} + V_{DS} - \Phi_{sL})\frac{\sinh\left(\frac{x}{l}\right)}{\sinh\left(\frac{L}{l}\right)} + (V_{bi} - \Phi_{sL})\frac{\sinh\left(\frac{L-x}{l}\right)}{\sinh\left(\frac{L}{l}\right)}, \qquad (3.4.2.2)$$

where Φ_{sL} is the surface potential if the device were very long, V_{bi} is the built-in potential of the source/drain junctions, L is the channel-length, and l is a characteristic length proportional to $(T_{ox}W_D)^{1/2}$. It can be seen that when the depletion depth W_D is small (e.g., positive V_{BS}), l is also small, so that $\Phi_s(x)$ approaches the constant Φ_{sL}, and the device behavior approaches that of a long-channel device (e.g., small DIBL). On the other hand, when W_D is large (e.g., large negative V_{BS}), l also becomes large, and $\Phi_s(x)$ deviates greatly from Φ_{sL}. The device therefore exhibits large amounts of short-channel effects (e.g., large DIBL). Figure 3.4.2.5 shows the simulated I-V characteristics of the example device discussed is this section.

Device breakdown and gate-induced-drain-leakage (GIDL) [72], however, are not direct electrostatic effects associated with doping, and are difficult to model precisely; data strongly affected by these effects should not be used in the optimization.

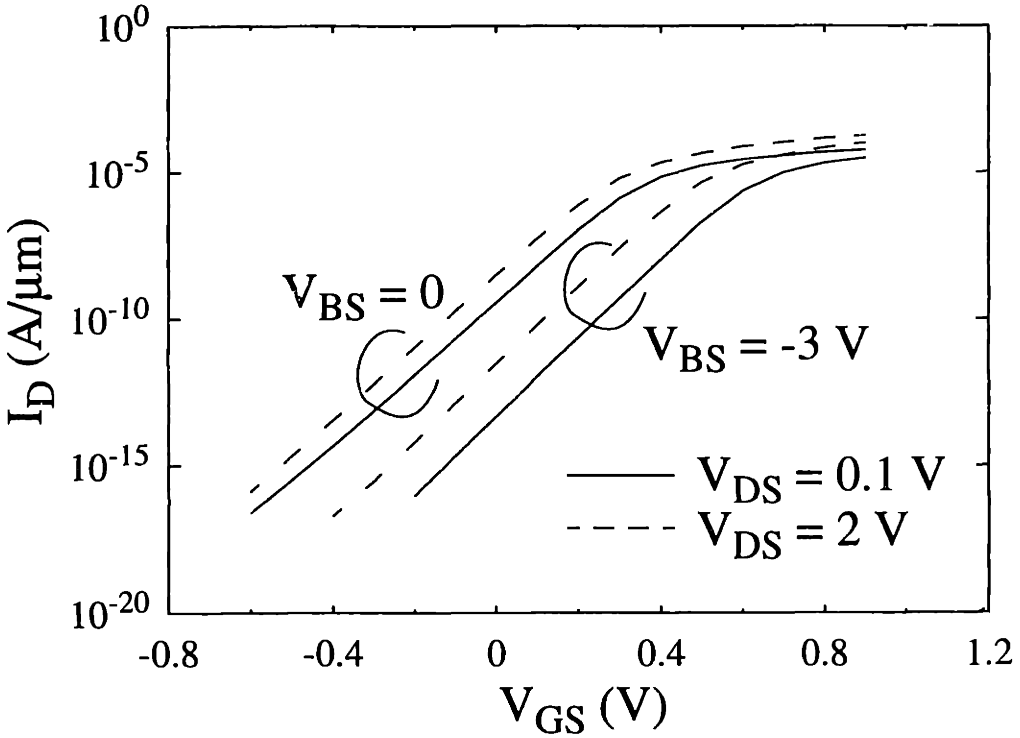

Figure 3.4.2.5: Effect of V_{BS} on DIBL of the example device discussed in this section.

The range of doping concentration that can be extracted is set by the limits determined by the maximum depletion depth, limited by device breakdown. Doping profiles within the minimum depletion depth can also be extracted; however, very fine details may be averaged out due to the double integral characteristics of Poisson's equation. For this reason, within the minimum depletion depth, the technique may have reduced sensitivity to distinguish between doping concentration at the surface and doping concentration away from the surface but within the minimum depletion depth. Therefore, proper biases have to be used so that the minimum depletion depth will be as small as possible (e.g., including positive V_{BS} and low V_{GS} data for n-channel devices). At very high doping levels (e.g., 2×10^{19} cm^{-3} and higher), the extraction method also loses its sensitivity because the potential variation diminishes there (see discussion of results in Chapter 4). Since short-channel effects are exploited, the technique is applicable only to short-channel devices which exhibit DIBL effects. The best sensitivity in fact lies in the channel region, a complement of the C-V technique [36-40] which has good sensitivity in the source/drain region because it allows the gate to source/drain overlap capacitance to be matched. For this reason, the two techniques should be viewed as complementary rather than exclusionary. It should be noted that the 2D Gaussian/B-spline representation presented here is only one of many possibilities; other functional representations, provided that the same quality of fit can be achieved, may also be used as well. However, the 2D Gaussian/B-spline representation is found to be sufficient for the applications attempted in the course of this work. The application and sensitivity analysis of the technique are discussed in Chapter 4.

3.5 Conclusion

In this chapter the theoretical aspects of subthreshold I-V inverse modeling are addressed. In subthreshold I-V inverse modeling, 2D doping profiles are obtained by matching a large set of simulated subthreshold I-V characteristics to the corresponding experimental data by optimizing for the doping profiles. Physically, the technique combines the matching of incremental changes in current and various biases (that allow the doping concentration near the depletion edge to be determined) and the absolute values of the current and biases (that allow the doping concentration within the depletion region to be determined). The process of inverse modeling, parameterization of doping profiles, and concept of equivalent oxide thickness are also discussed.

Chapter 4

Methodology and Results

4.1 Introduction

The theoretical explanations why subthreshold I-V characteristics can provide useful information pertaining to the 2D distribution of dopants have been presented in Chapter 3. In this chapter a 2D doping profile extraction methodology using subthreshold I-V characteristics [41-42] is illustrated using a number of examples. The chapter will conclude with a discussion of the sensitivity issues as well as other special features of the technique.

The devices being examined here were fabricated at MIT, and have different channel-lengths and channel dopings, but similar source/drain extensions with “halo” implants [23]. They are n-channel devices having either a boron and indium doped super-steep retrograde (SSR) or a BF_2 doped step-like channel doping profile (STEP), with indium “halo” and arsenic source/drain extensions. The gate oxide thickness is 4.5 nm. Other structural information, although not important for the extraction here, includes a spacer length of approximately 180 nm, a polysilicon gate thickness of 300 nm, and a deep junction depth of 110 nm. A schematic diagram of the device structure, drawn roughly to the actual aspect ratio, is given in Fig. 4.1.1.

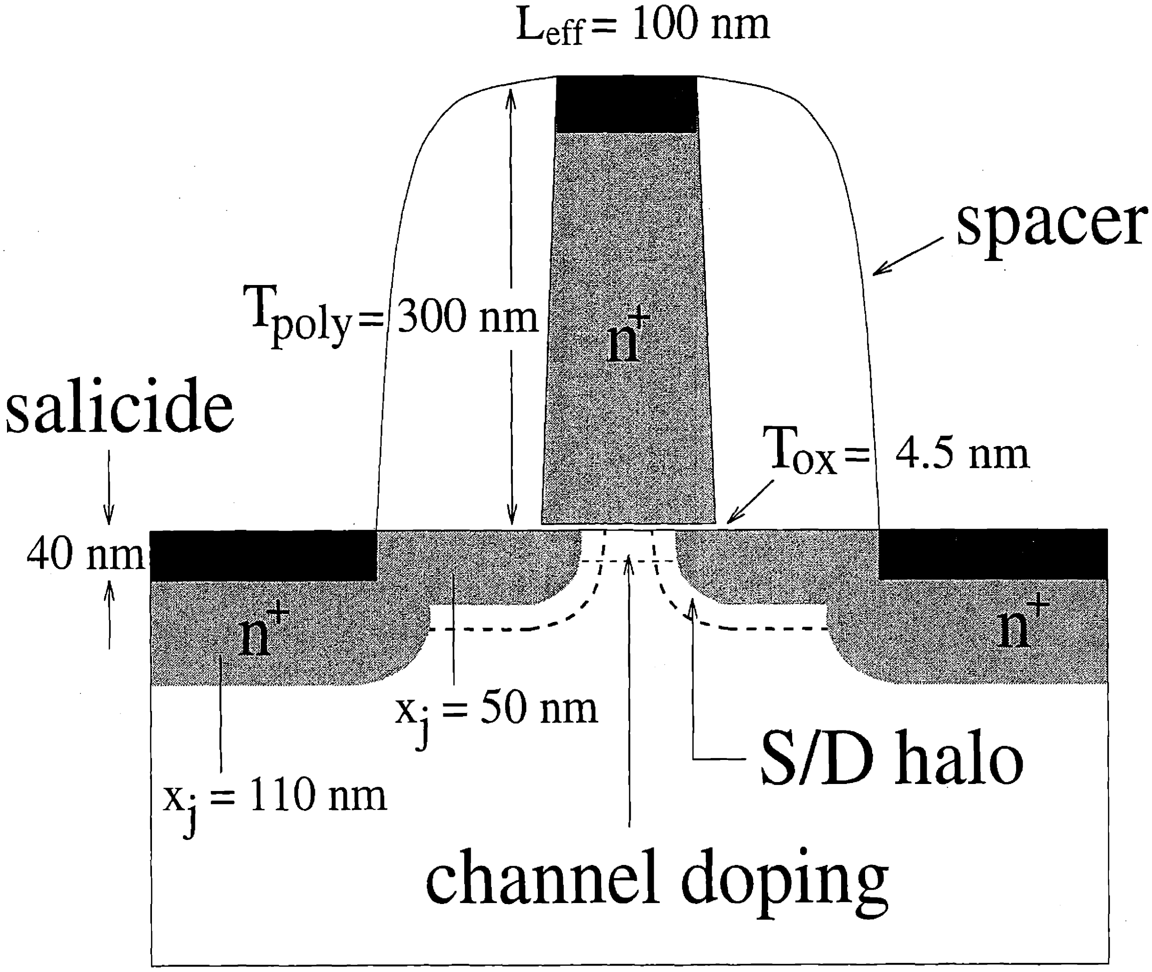

Figure 4.1.1: Schematic diagram of a 100 nm channel-length device, showing the halos, source/drain extensions, and deep source/drain junctions. The channel doping profile can be either a super-steep retrograde (SSR) or a step (STEP).

4.2 Extraction Procedure

(a) SSR device

The procedure starts with the extraction of the 1D doping profile (i.e., doping in the depth direction) in a long-channel device co-fabricated with the short-channel devices of interest. This can be accomplished by the standard C-V method (e.g., [31]), advanced C-V methods that also model quantum mechanical and polysilicon gate depletion effects [32, 65], SIMS [34], or, as illustrated here, by optimizing the 1D doping profile so that the simulated subthreshold I_D vs. V_{GS} vs. V_{BS} data (from the device simulator MEDICI) of a long-channel device match the corresponding experimental data [41-42]. By applying V_{BS}, the depletion depth is changed (for an NMOS device, it is increased for negative V_{BS}). The change in depletion depth, and thus depletion charge, which is related to the 1D doping profile, causes a corresponding change in the inversion charge in order to maintain charge balance with the gate. Consequently the current, which directly measures the amount of inversion charge, is changed as a result of applying V_{BS}. Important 1D doping information can therefore be extracted. Note that fitting I_D vs. V_{GS} vs. V_{BS} data in the weak inversion region, as opposed to fitting the threshold voltage V_t vs. V_{BS} data [69-71], allows the doping profile very near the surface to be extracted. This is because the range of I_D vs. V_{GS} vs. V_{BS} data can be chosen so that the depletion depth is small (e.g., negative V_{GS} and slightly positive V_{BS} for an NMOS device). On the other hand, the minimum depletion depth achievable by applying $V_{GS} = V_t$, as required by fitting V_t vs. V_{BS} data, is usually much larger, even for slightly positive V_{BS}. Moreover, fitting the complete subthreshold I-V curve also results in more accurate profiles because the subthreshold slope is in general a function of V_{BS}.

Figure 4.2.1 shows a subset of the subthreshold I-V data of a 5 μm channel-length device having an extracted SSR channel profile as shown in the inset. Here, the profile is represented by a B-spline function [64], with the corresponding coefficients as fitting parameters. The knot locations are chosen so as to achieve the best match to the data (see, for example [32]). In general, a good fit should have very small errors in the V_t (a few mV) and subthreshold slope (a few mV/dec average) between the simulated and experimental data, and occur uniformly at all biases (see Fig. 4.2.1). The maximum depth of the extracted profile is determined by the maximum depletion depth, which, in this case, is approximately 250 nm.

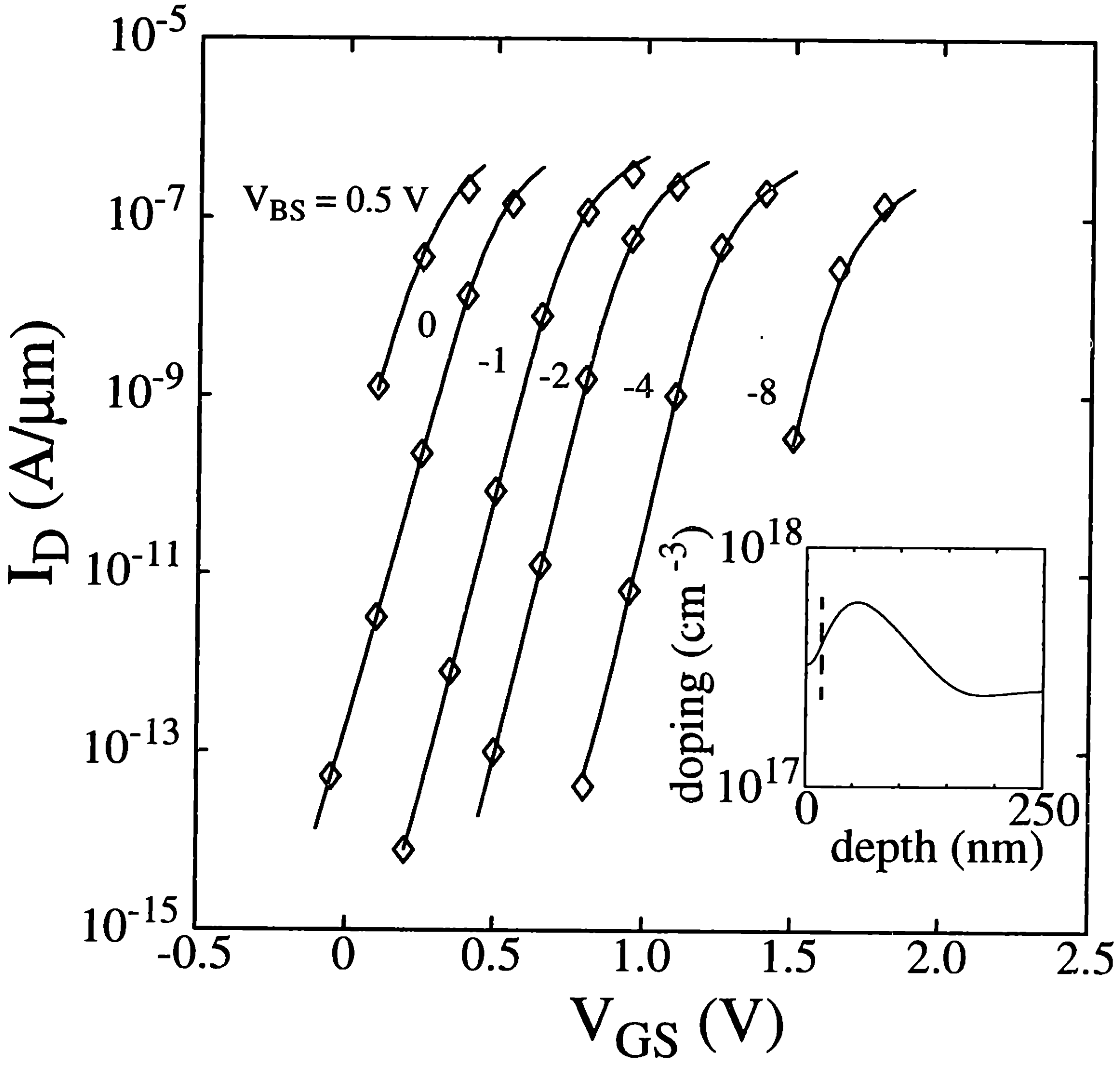

Figure 4.2.1: I_D vs. V_{GS} vs. V_{BS} data of a 5 μm channel-length SSR device. Simulated (optimized) data are denoted by symbols. The device width is 50 μm. The extracted 1D channel profile is shown in the inset. The dashed line in the inset represents the minimum depletion depth achievable in the data set, limited by leakage currents.

The extracted 1D profile is subsequently used as a starting point for the approximate profile in the depth direction of the short-channel device. The complete 2D profile of the short-channel device, as parameterized using a sum of B-spline functions and two 2D Gaussian functions, which represent the source/drain/halo regions, is extracted by fitting simulated I_D vs. V_{GS} vs. V_{DS} vs. V_{BS} data to the corresponding experimental data in the subthreshold region. Here, the center and width of each Gaussian function, in both the depth and lateral directions, are varied in the optimization. The amplitude of one of the Gaussian functions, which represents the peak "halo" doping, is also allowed to be optimized. The amplitude of the second Gaussian function, which represents the peak source/drain doping, however, is kept fixed at the doping level determined by 1D process simulations and/or SIMS, since the technique is unable to resolve very high (i.e., degenerate) doping levels (see the discussion in section 4.4). The coefficients corresponding to the B-spline function, which represent the channel doping contribution, are also varied in the optimization. Figure 4.2.2 shows a comparison between the experimental I-V data and the simulated data of the optimized 2D profile. The bias conditions are chosen to accentuate the short-channel effects so that the doping information is revealed. It should be noted that the data shown here represent only a small sample of the actual data used in the optimization. Typically, about thirty I-V curves, which cover all bias conditions, are used in each optimization. A number of interesting phenomena can be observed in Fig. 4.2.2. At V_{BS} = 0.4 V, the amount of DIBL, as signified by the shift of the I-V curve from V_{DS} = 0.21 V to V_{DS} = 2.01 V is very small as compared to the curves at the same V_{DS}'s but with V_{BS} = -3.5 V. This is due to the DIBL vs. V_{BS} dependence as discussed in Chapter 3. The shift of the I-V curves due to the body-effect as a result of applying V_{BS} is also seen. Moreover, the subthreshold slope dependence on V_{DS} and V_{BS} can also be observed. Figure 4.2.3 shows the extracted lateral profile at the oxide-silicon interface of the device corresponding to that of Fig. 4.2.2., along with that of another device having a slightly longer channel-length. The mid-channel doping profiles in the depth direction are given in Fig. 4.2.4. Note that the lateral doping profiles are very similar (e.g., the halos have similar peaks and curvatures). The difference in the vertical doping profiles are indeed very small. Reasons for the difference include: (1) the doping profiles are indeed different due to various 2D diffusion effects [1-21]; (2) the set of experimental V_{BS} data used in the optimization is limited (this problem can be eliminated or reduced by using a larger set of V_{BS} data, with smaller spacings, so that the sensitivity can be improved); and (3) the knot locations for the B-spline (Appendix B) are not optimal (this can be improved by trying different locations).

It should be noted that when the B-spline coefficients that represent the 1D (depthwise) contribution of the doping profile are also optimized, a large set of data to cover the V_{BS} dependence is needed in the optimization. In practice, the 1D (depthwise) profile obtained from a long-channel device is kept fixed while optimizing for the 2D source/drain/halo profile, and when a good match of the data is found, the 1D B-spline coefficients are relaxed and allowed to be varied to improve the match.

A variation of the procedure described above is also possible. For instance, the 1D channel profile that is extracted from the long-channel device can be fixed in the entire optimization (i.e., the contribution from the depthwise variation of the doping concentration in the short-channel device is assumed to be the same as that in the long-channel device). This is especially important for devices in which the channel doping concentration is low (such as the "STEP" devices to be shown next), so that the shift of the I-V curves as V_{BS} is applied becomes small. Fixing the channel doping concentration while optimizing for the 2D doping profile helps improve the sensitivity as well as the stability in the optimization process.

(b) STEP device

For devices having low background channel doping concentration, the coefficients corresponding to the 1D channel doping profile are kept fixed in the optimization for the short-channel device. This helps improve the sensitivity as well as the stability of the optimization because the shift of the I-V curves as V_{BS} is applied can be small due to the reduced body effect. Figure 4.2.5 shows the subthreshold I_D vs. V_{GS} vs. V_{BS} data of a 5 μm channel-length STEP device having a step-like channel doping profile. The extracted 1D profile, as parameterized using a sum of two 1D Gaussian functions with the amplitude, center, and width of each Gaussian function a fitting parameter, is shown in the inset. Having obtained this 1D profile, a sum of two 2D Gaussian functions are then added and optimized to match the simulated I_D vs. V_{GS} vs. V_{DS} vs. V_{BS} data to the corresponding experimental data from the short-channel device. Figure 4.2.6 shows a sample of the data of a 120 nm channel-length device that are used in the optimization. Since the channel doping is relatively low for these devices, punch-through can be readily observed at higher V_{DS}. The signature of the punch-through current provides a self-consistent piece of 2D doping information, as evident in Fig. 4.2.6. The corresponding lateral profile at the oxide-silicon interface, along with those of other devices having different channel-lengths, is shown in Fig. 4.2.7. Other than the gate length, these devices are identical (i.e., they have the same source/drain/halo implant).

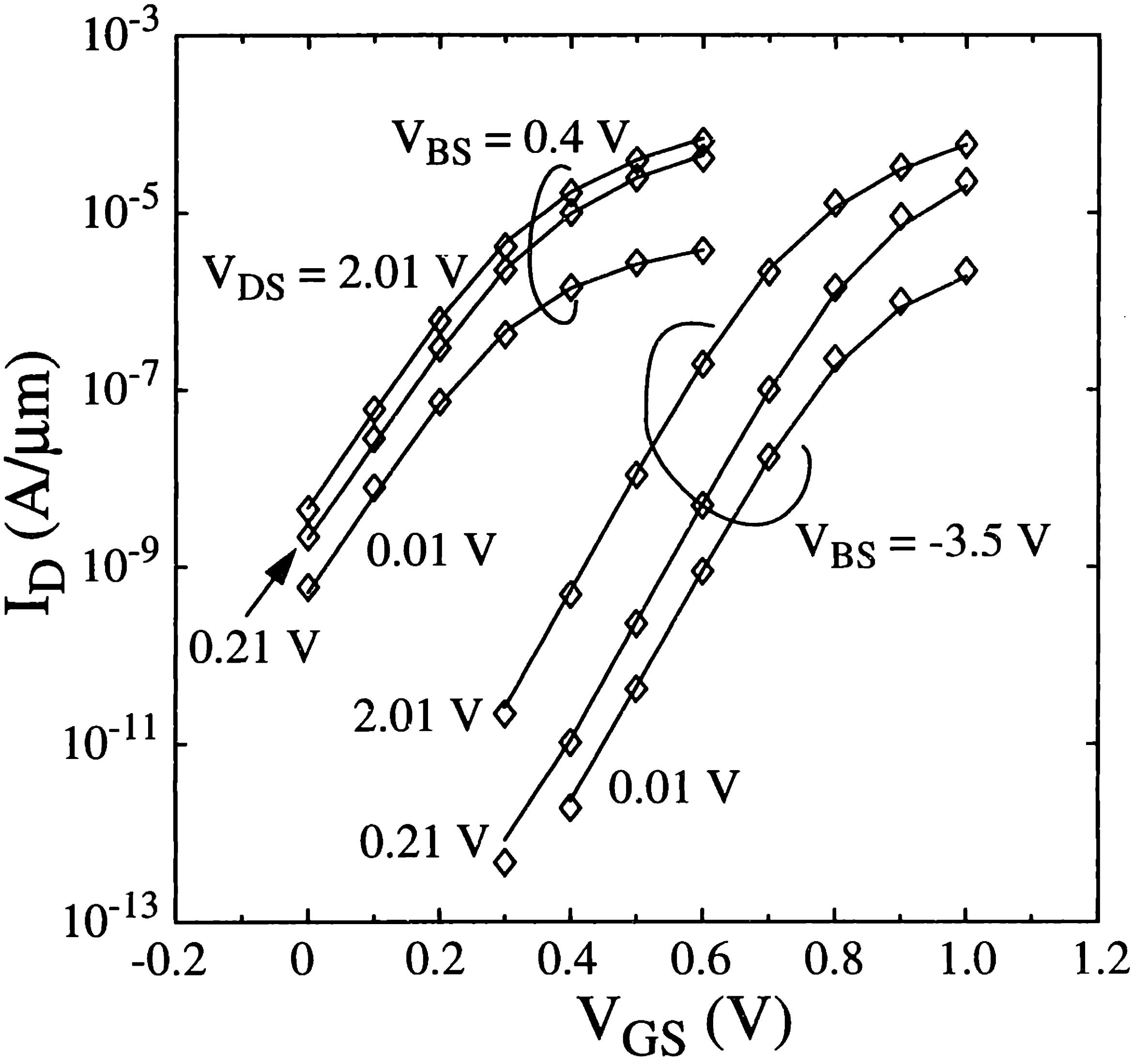

Figure 4.2.2: Comparison between experimental and simulated (optimized) I_D vs. V_{GS} vs. V_{DS} vs. V_{BS} data of a 140 nm channel-length SSR device (SSR3 R2). Simulated data are denoted by symbols. The device width is 10 μm.

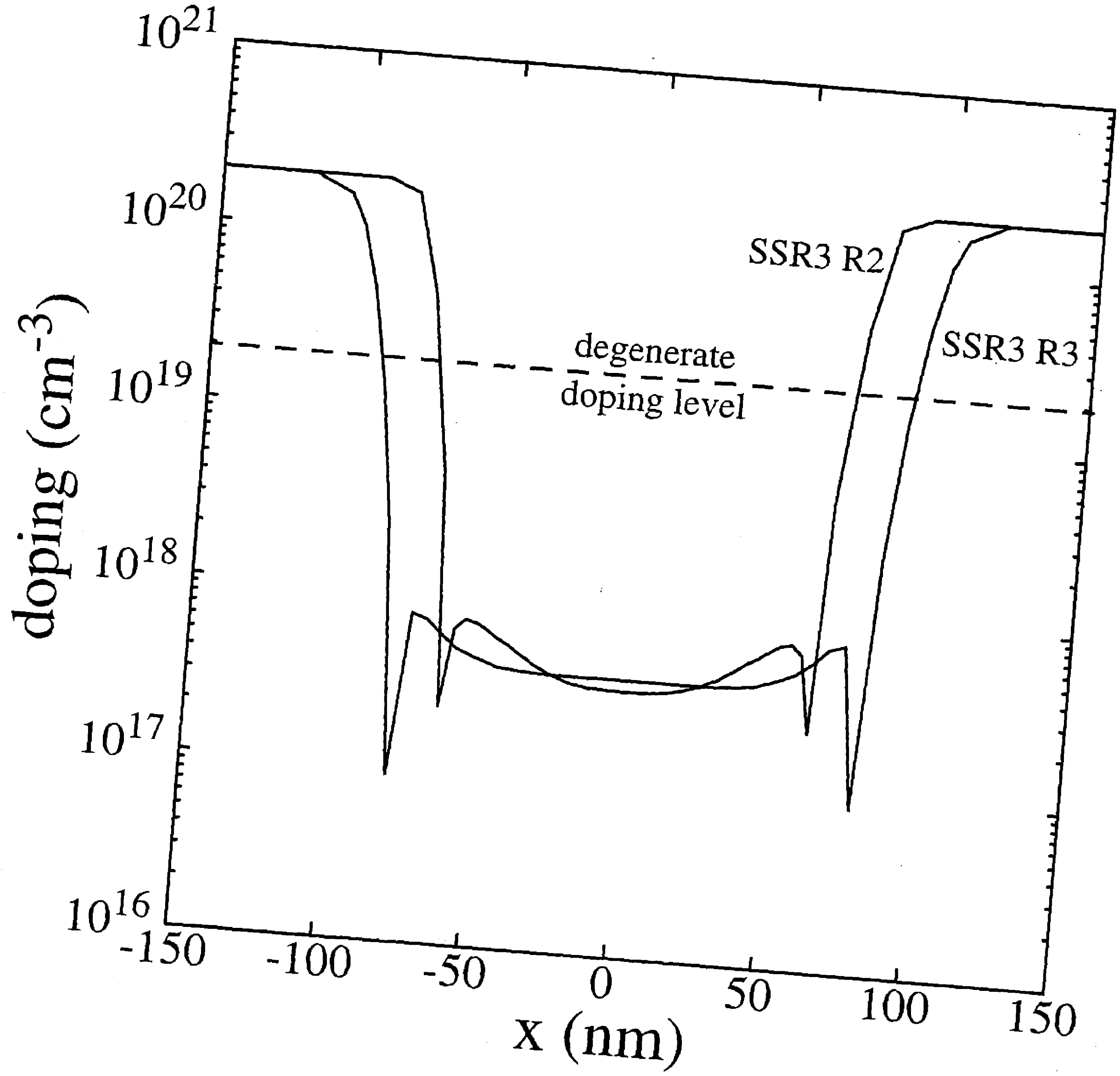

Figure 4.2.3: Extracted lateral profile at the oxide-silicon interface of the device of Fig. 4.2.2 (SSR3 R2), along with that of another device having a slightly longer channel-length (SSR3 R3). The degenerate doping level represents the doping level above which the sensitivity of the technique is reduced (see section 4.4).

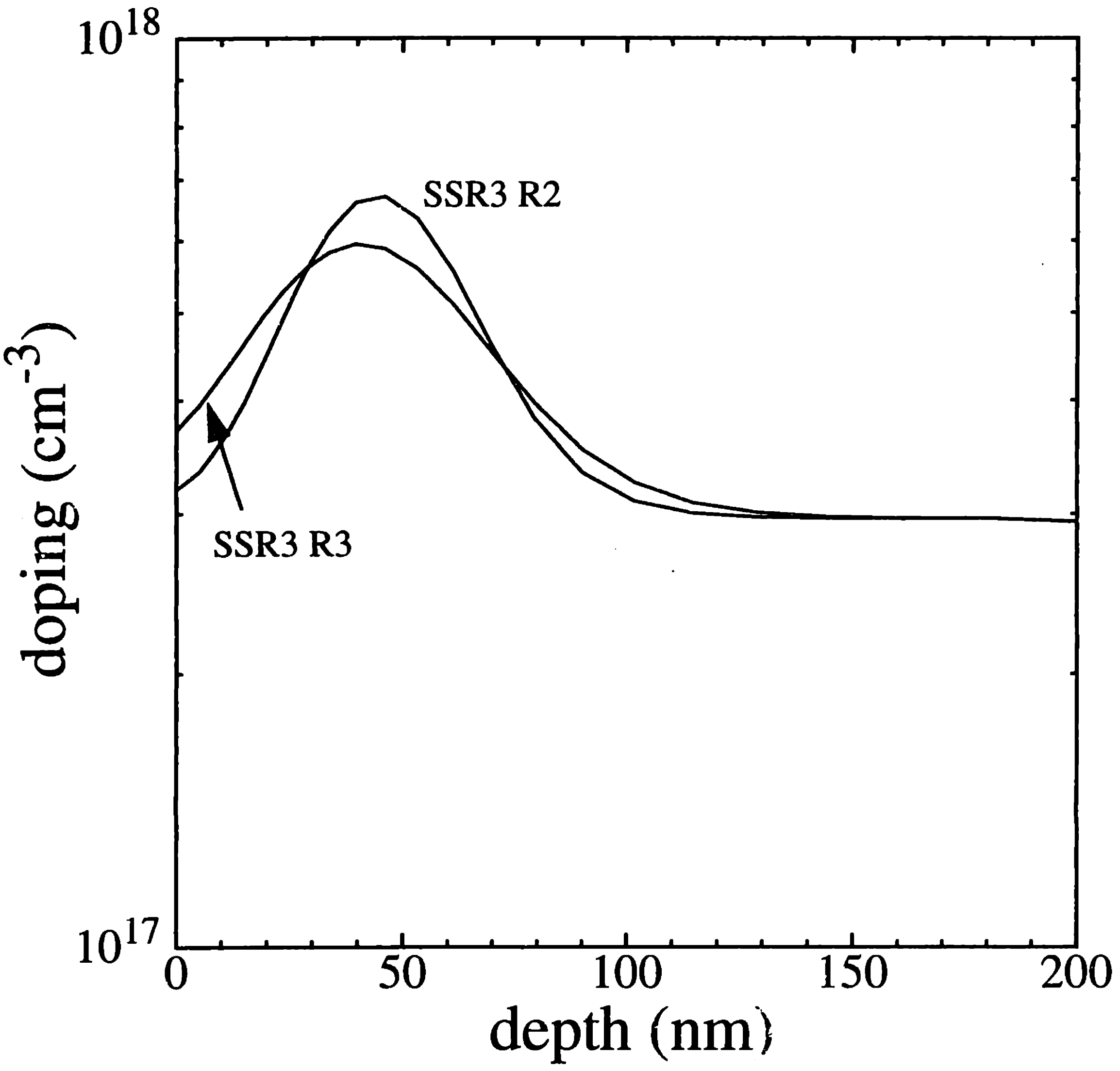

Figure 4.2.4: Extracted mid-channel doping profile in the vertical (depth) direction of the device of Fig. 4.2.2 (SSR3 R2), along with that of another device having a slightly longer channel-length (SSR3 R3).

It can be seen that the source/drain/halo regions of the longer devices are very similar, particularly the shape and peak doping level of the halo, which occurs just below 10^{18} cm^{-3}. This is an expected result since the devices were processed simultaneously. For the shorter devices, however, the slope as well as the peak doping level of the halo increases as the gate length decreases. This is believed to be due to transient enhanced diffusion (TED) [6-16], which causes the indium dopants in the halos to pile up near the source/drain junctions where the silicon point defect gradient is greatest. Figure 4.2.8 shows an expanded view of the halo profiles of Fig. 4.2.7 near the metallurgical junctions. The x-axis (Δx) has been adjusted with its origin defined at the metallurgical junctions, while the halo doping profiles are extrapolated to the metallurgical junctions. Although the differences are small, the trend can be clearly seen. While there is hardly any difference between the longer devices (devices STEP50 R3 and STEP50 R4), the pile-up effect is most evident in the shorter ones, with device STEP50 R1 showing the greatest pile-up.

4.3 Effective Channel-Length

As part of the 2D doping profile, this technique automatically extracts the channel-length, which is an important parameter for gauging technologies. Table 4.3.1 shows a comparison of the extracted channel-lengths using this technique with the experimental effective channel-lengths L_{eff} extracted using the C_{gds} [73] and "shift and ratio" [74] (see Appendix C) methods. The C_{gds} method is a combination of capacitance and the Terada and Muta method [75] of effective channel-length extraction. Using the gate-to-channel capacitance in inversion and subtracting the fringing capacitance in accumulation, the gate electrode length can be extracted. This gate length is subsequently used to obtain the L_{eff} by measuring the channel resistance (Terada and Muta method, see [75] for details).

For the profiles extracted from inverse modeling, the channel-length is arbitrarily "defined" by the positions in the source/drain at which the doping concentration is sufficiently high (i.e., $2x10^{19}$ cm^{-3}). Note that the L_{eff} obtained by the "shift and ratio" method tends to be slightly longer than that obtained by inverse modeling when this definition of "channel-length" for the extracted profiles is used. Indeed, when the "shift and ratio" method is applied on simulated data, the L_{eff}'s thus obtained are typically longer than those defined by the $2x10^{19}$ cm^{-3} points. This discrepancy is especially pronounced for the devices having low channel doping concentrations, such as the STEP devices. For example, using simulated I-V characteristics on the doping profile obtained from the STEP10 R2 device, the L_{eff} extracted by the "shift and ratio" method is 207 nm, while the channel-length at $2x10^{19}$ cm^{-3} is only 150 nm. Possible reason for this discrepancy is that the halo doping raises the local threshold voltage so that the inversion charge concentration is not uniform across the channel, contrary to the basic assumption in the "shift and ratio" method that the channel conductance varies as $1/L_{eff}$. Another cause of the differences is that the experimental data used in inverse modeling were taken after repeated measurements (which could include some stressing), as well as cooling and warming cycles, which could have caused degradation.

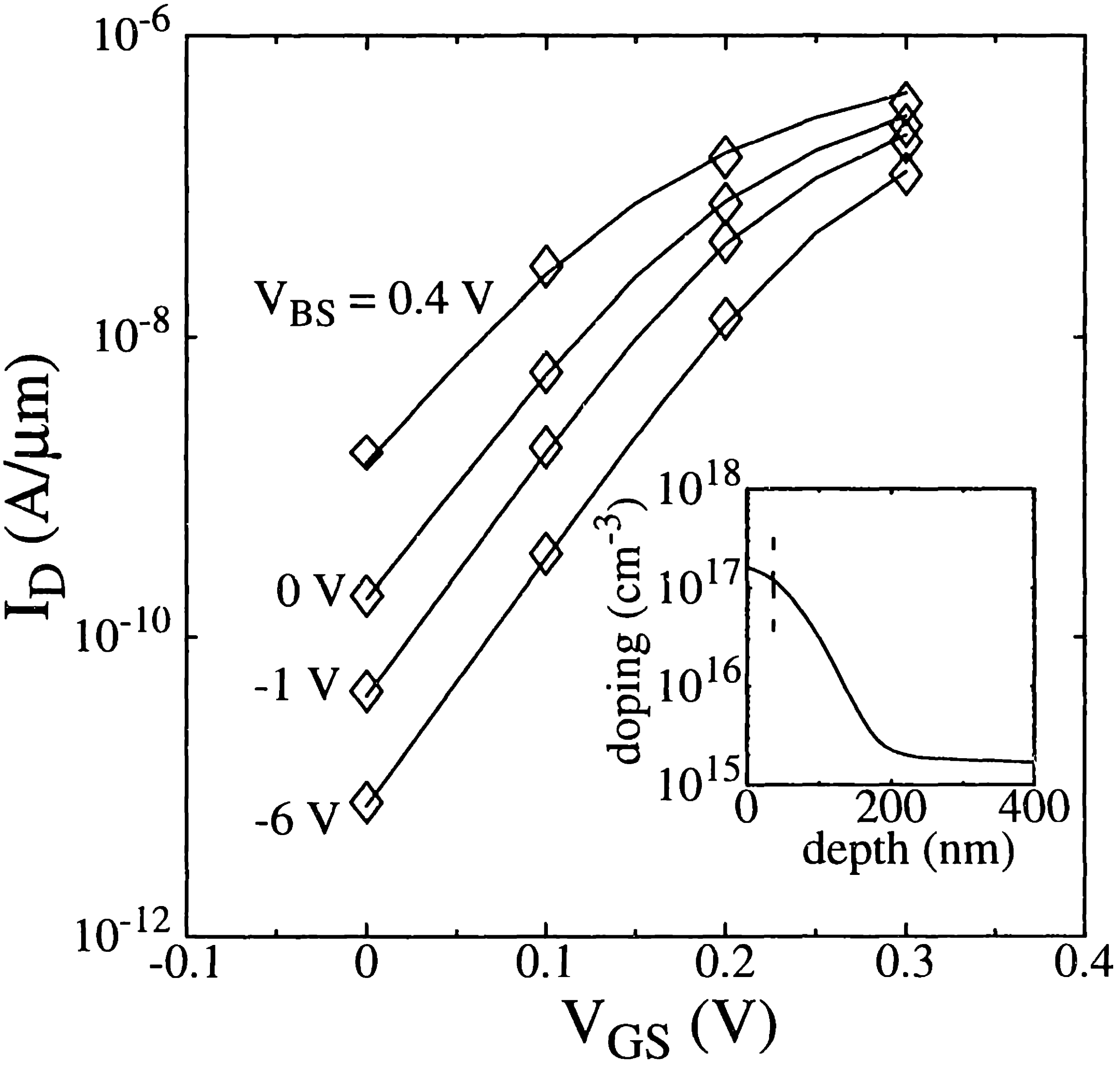

Figure 4.2.5: I_D vs. V_{GS} vs. V_{BS} data of a 5 μm channel-length STEP device. Simulated (optimized) data are denoted by symbols. The device width is 50 μm. The extracted 1D channel profile is shown in the inset. The minimum depletion depth is represented by the dashed line in the inset.

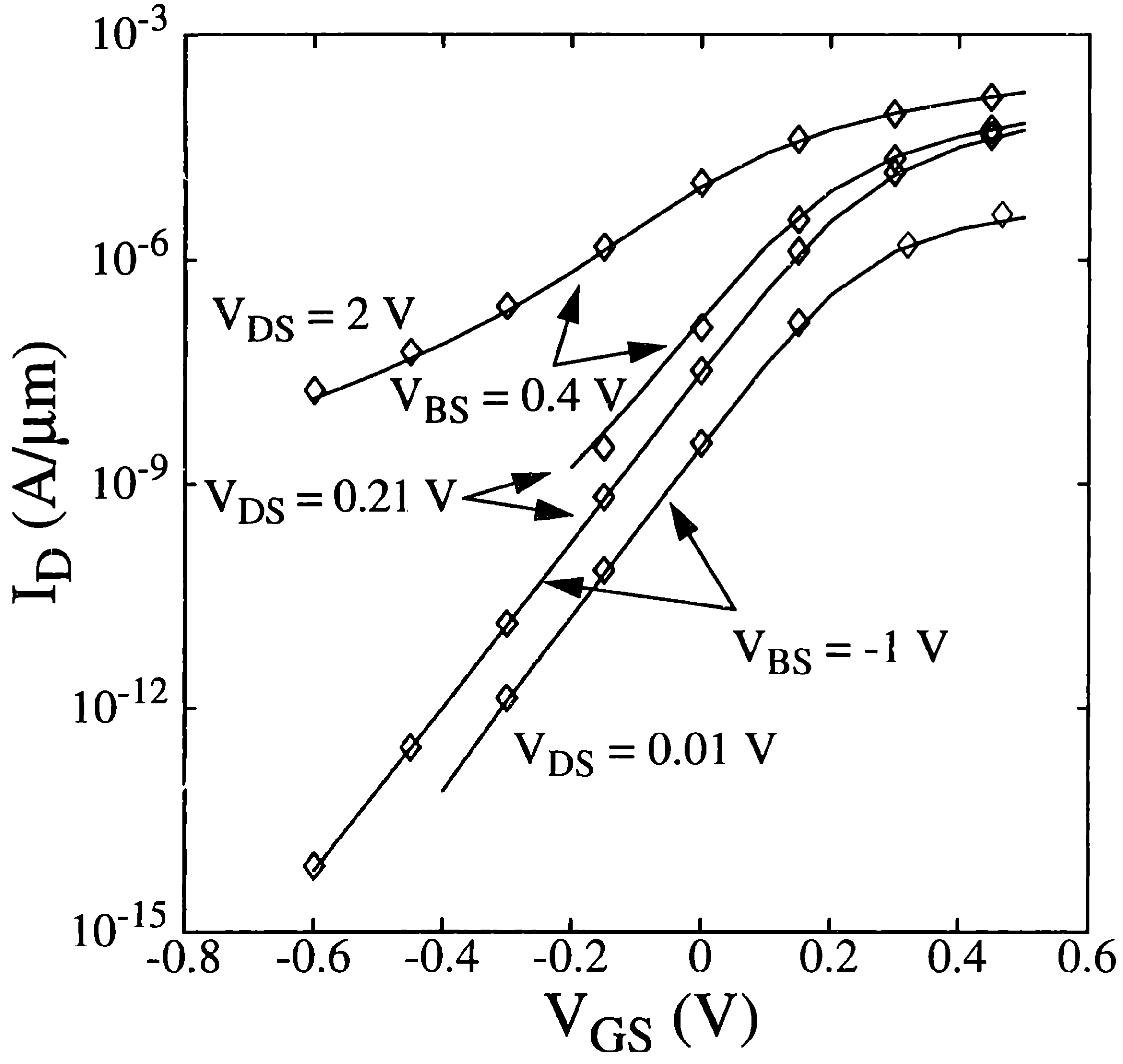

Figure 4.2.6: Comparison between experimental and simulated (optimized) I_D vs. V_{GS} vs. V_{DS} vs. V_{BS} data of a 120 nm channel-length STEP device (STEP50 R1). Simulated data are denoted by symbols. The device width is 50 µm. Note that the punch-through information is fully reproduced in the simulated (optimized) data.

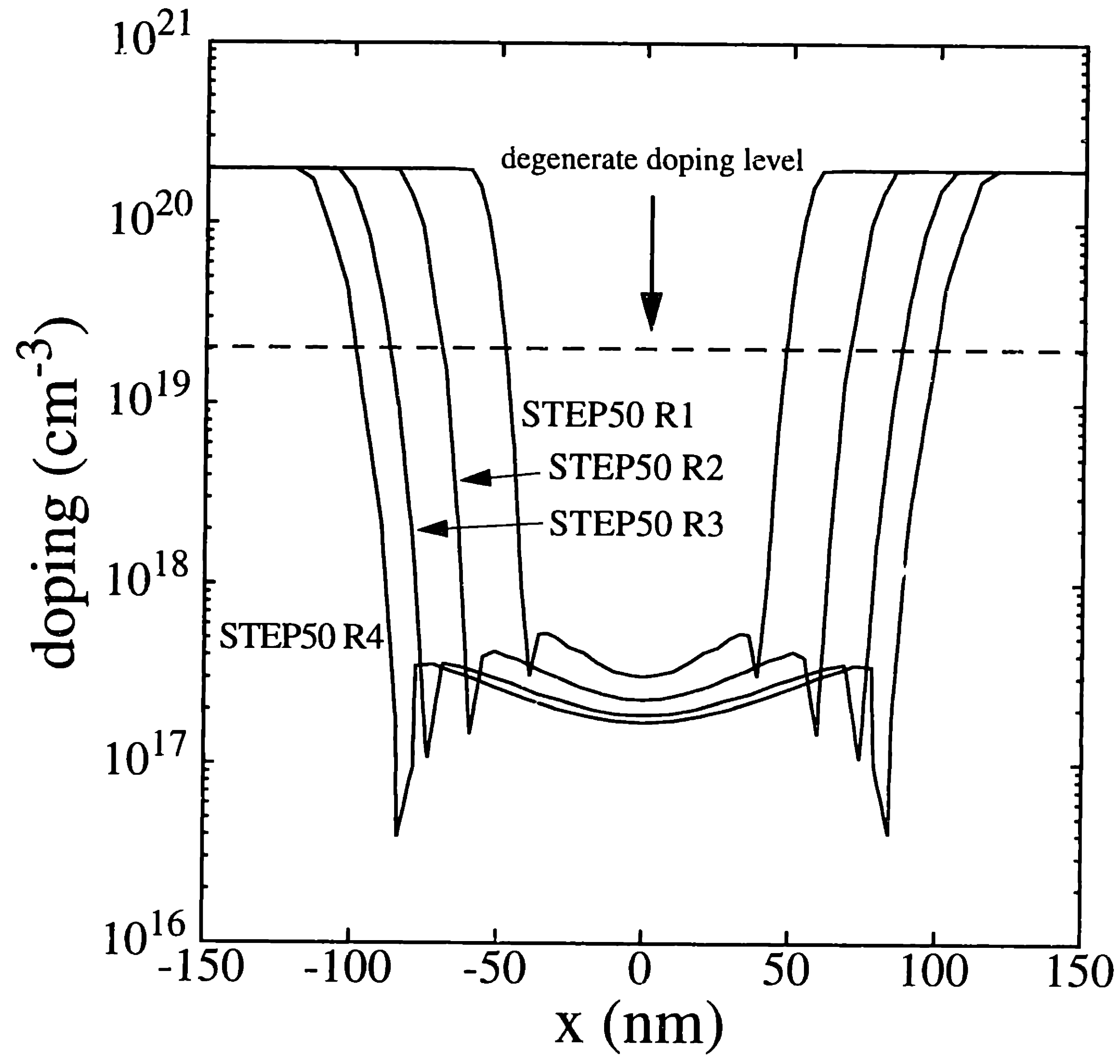

Figure 4.2.7: Comparison of lateral doping profiles at the oxide-silicon interface of various STEP devices. Other than the gate length, the devices are identical (i.e., they have the same source/drain/halo doping implants). The width of the devices is 50 μm. The degenerate doping level represents the doping level above which the sensitivity of the technique is reduced (see section 4.4).

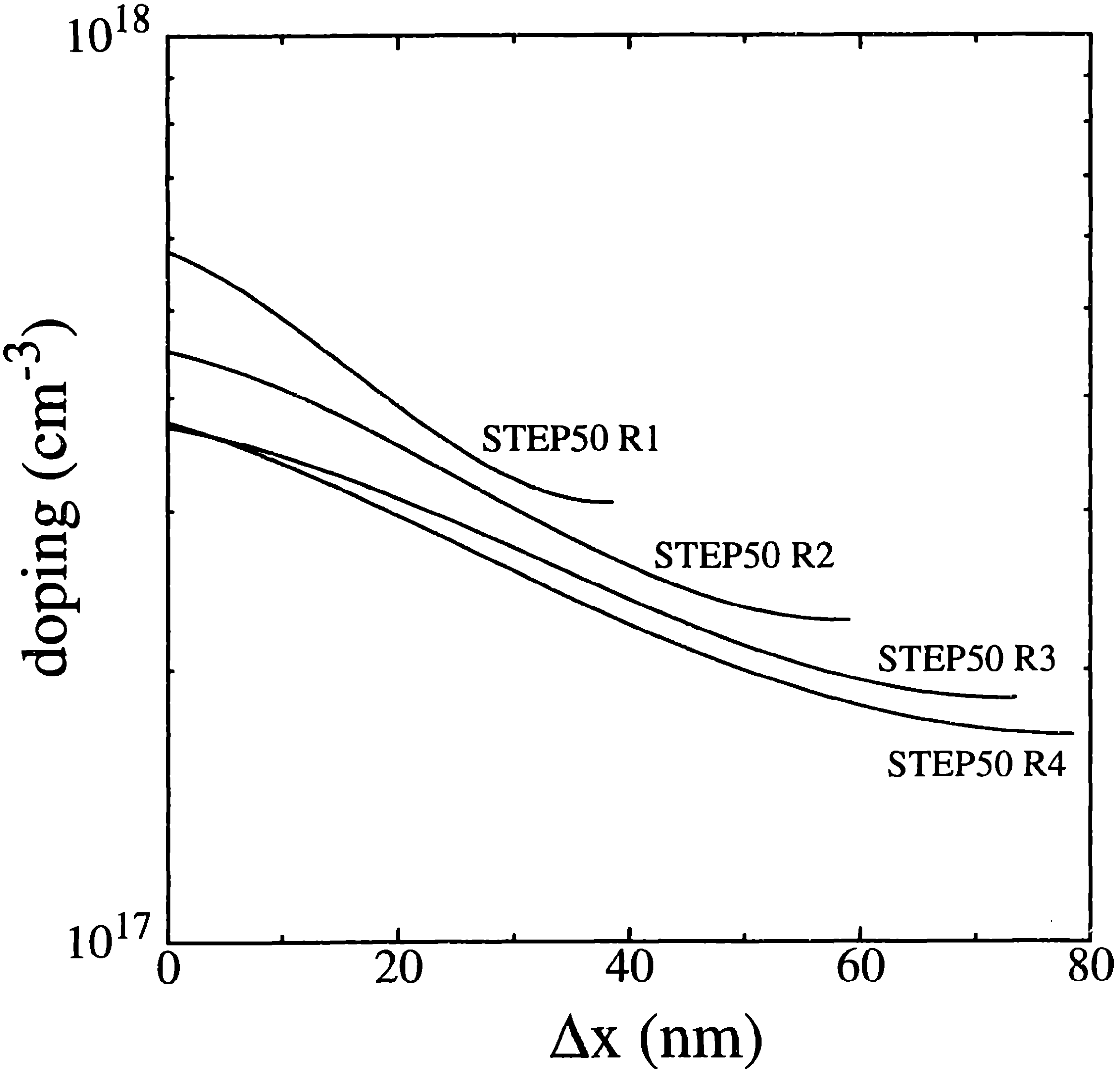

Figure 4.2.8: Expanded view of the halo profiles of Fig. 4.2.7 near the metallurgical junctions. The x-axis (Δx) has been adjusted so that the metallurgical junctions lie at the origin. The doping profiles of the halos are extrapolated to the metallurgical junctions. Although the differences are small, the trend that the indium piles up near the source/drain junctions is evident as the devices become short.

Table 4.3.1: Comparison of the channel-length of various devices extracted by inverse modeling, C_{gds} [73], and "shift and ratio" [74] methods. The channel-length obtained by inverse modeling is arbitrarily "defined" by the points in the source/drain at which the doping level is 2×10^{19} cm^{-3}. All devices have a width of 10 μm, except for the STEP50 devices, which have a width of 50 μm.

	Channel-Length (nm)	Effective Channel-Length (nm)	
Device	Inverse Modeling	C_{gds} method	"Shift and Ratio" method
SSR1 R2	90	--	100
SSR1 R4	148	--	160
SSR2 R2	95	117	120
SSR2 R3	149	166	180
SSR3 R2	145	142	160
SSR3 R3	181	185	180
STEP10 R1	125	122	130
STEP10 R2	150	162	210
STEP10 R3	203	203	230
STEP10 R4	234	232	260
STEP50 R1	94	--	130
STEP50 R2	138	--	160
STEP50 R3	172	--	200
STEP50 R4	196	--	230

4.4 Discussion

One of the major advantages of the inverse modeling technique is the low sensitivity to mobility and mobility models. Another advantage is its insensitivity to gate electrode dimensions, since, unlike capacitance-based extraction methods [32, 36-40], it is the channel-length, as opposed to the gate electrode area, that determines device electrical characteristics. From a simple model of a short-channel device having a uniform channel doping concentration, the subthreshold current is approximately proportional to [see eqns. (2.3.18), (2.3.20), and (2.4.3)]:

$$I_D \propto \mu_n W \exp\left(-\frac{V_x}{n\phi_T}\right)\exp\left(\frac{\Delta V_t}{n\phi_T}\right), \tag{4.4.1}$$

where

$$V_x = V_{FB} + 1.5\phi_F + \gamma\sqrt{1.5\phi_F - V_{BS}} \tag{4.4.2}$$

and

$$\Delta V_t = [3(V_{bi} - \Phi_{sL}) + V_{DS}]\exp\left(-\frac{L}{l}\right) + 2\sqrt{(V_{bi} - \Phi_{sL})(V_{bi} - \Phi_{sL} + V_{DS})}\exp\left(-\frac{L}{2l}\right). \tag{4.4.3}$$

The first exponential factor in eqn. (4.4.1) corresponds to the dependence on doping concentration in a long-channel device, while the second exponential factor corresponds to the perturbation due to short-channel effects. As discussed in Chapter 2, due to the exponential dependence, a very small change in the doping concentration will result in a very large change in I_D even for a long-channel device. This is because the body factor γ, which is proportional to $(N_A T_{ox}^2)^{1/2}$, is inside the exponential. For a short-channel device, it can be seen that ΔV_t, which is already inside the exponential of eqn. (4.4.1), in turn depends on l (which contains a factor T_{ox}^2 / N_A) through another exponential. This double exponential dependence makes I_D extremely sensitive to the doping concentration. On the other hand, I_D is only linearly proportional to $\mu_n W$. Consequently any uncertainties in μ_n and W will have an insignificant effect on the extracted profiles. Moreover, since to a large extent the inverse modeling technique makes use of the incremental changes in bias, unless the mobility changes very rapidly over a very narrow range of biases (which is unphysical), the doping profile thus constructed at the depletion edge will not be affected by μ_n and W because only the information pertaining to the incremental changes is needed here (see

Chapter 3).

However, since T_{ox} is always associated with N_A, the sensitivity to T_{ox} is very high. This can be verified easily by simulations. Figures 4.4.1 and 4.4.2 show, respectively, the effects on the extracted profiles due to the use of different mobility models and assumed oxide thicknesses. As expected, different mobility models have very small effects on the extracted profiles, while a ten percent change in the assumed oxide thickness results in approximately the same percentage change in the extracted channel-length, although the shape and magnitude of the profile remain essentially the same in this case. In order to obtain accurate profiles and channel-lengths, the oxide thickness should be modeled properly, with polysilicon gate depletion and quantum mechanical effects taken into account [65-68].

It should be noted that the range of the doping profile that can be probed by the technique is set by the limits determined by the maximum depletion depth, since only ionized dopants in the depletion region have an effect on the potential distribution in the device. Note that the doping concentration inside the minimum depletion depth can also be extracted; however, fine details may be averaged out due to the characteristics of Poisson's equation (see Chapter 3), so that the resolution is reduced. Therefore, the profile thus extracted that is within the minimum depletion depth should more appropriately be viewed as a spatial average.

At very high doping levels (e.g., 2×10^{19} cm^{-3} and higher), the extraction method loses its sensitivity. This is because the potential variation diminishes at high concentrations. This is confirmed by simulations as shown in Fig. 4.4.3. Here, profiles are extracted using the same set of data but different peak source/drain doping levels, which are intentionally fixed. The same quality of fit to the data is achieved in both cases. Note that the doping profiles are almost identical at lower (non-degenerate) doping levels, despite the fact that the peak dopings are different. Since the extraction method is unable to resolve degenerate doping levels, the peak doping values can be fixed arbitrarily in the optimization without affecting the extracted non-degenerate doping profiles. It should also be noted that the best sensitivity of this technique lies in the channel region, whereas the C-V technique [36-40] works best to provide doping information within the source/drain regions. For this reason, the two techniques should be viewed as complementary rather than exclusionary.

It should be noted that the 2D Gaussian/B-spline representation presented here is only one of many possibilities; other functional representations, provided that the same quality of fit can be achieved, may also be used as well. However, the 2D Gaussian/B-spline representation is found to be sufficient for the devices examined in this thesis.

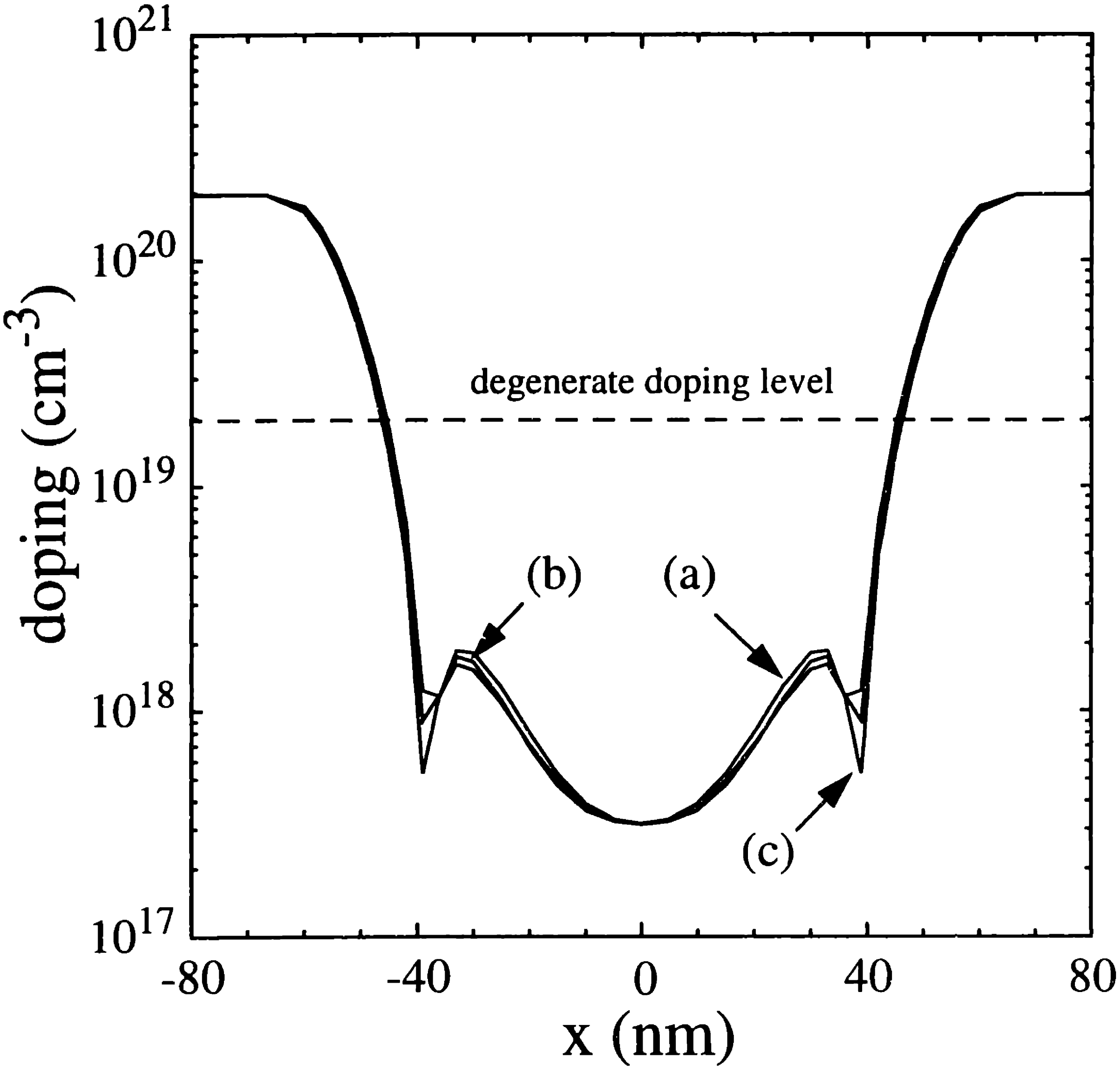

Figure 4.4.1: Comparison of extracted lateral profiles at the oxide-silicon interface obtained using (a) concentration dependent mobility model; (b) concentration dependent mobility model with approximately half the mobility as in (a); and (c) concentration dependent mobility model with vertical electric field degradation and velocity saturation taken into account.

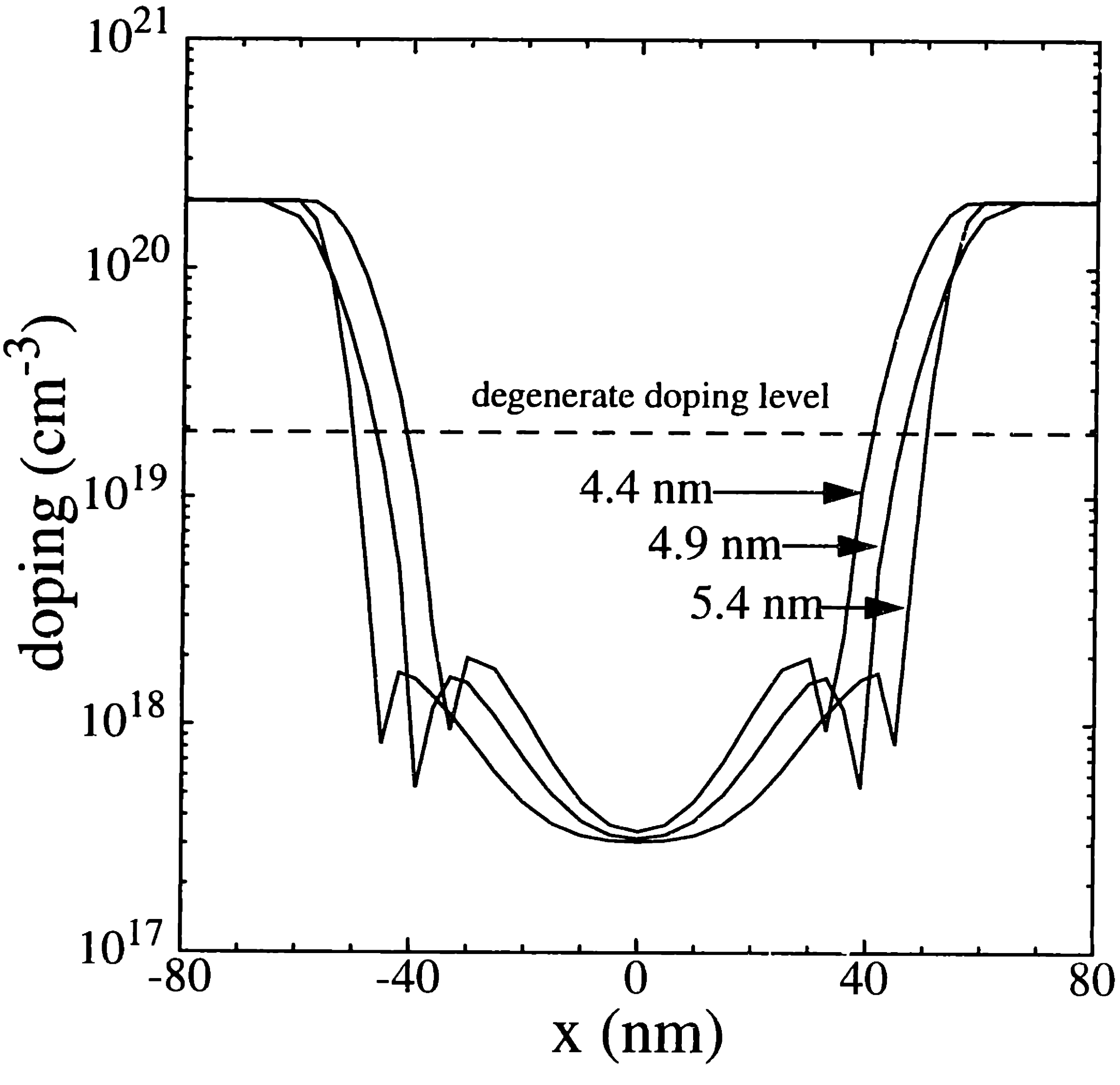

Figure 4.4.2: Comparison of extracted lateral profiles at the oxide-silicon interface obtained using various assumed oxide thicknesses. The expected oxide thickness of the device is 4.9 nm.

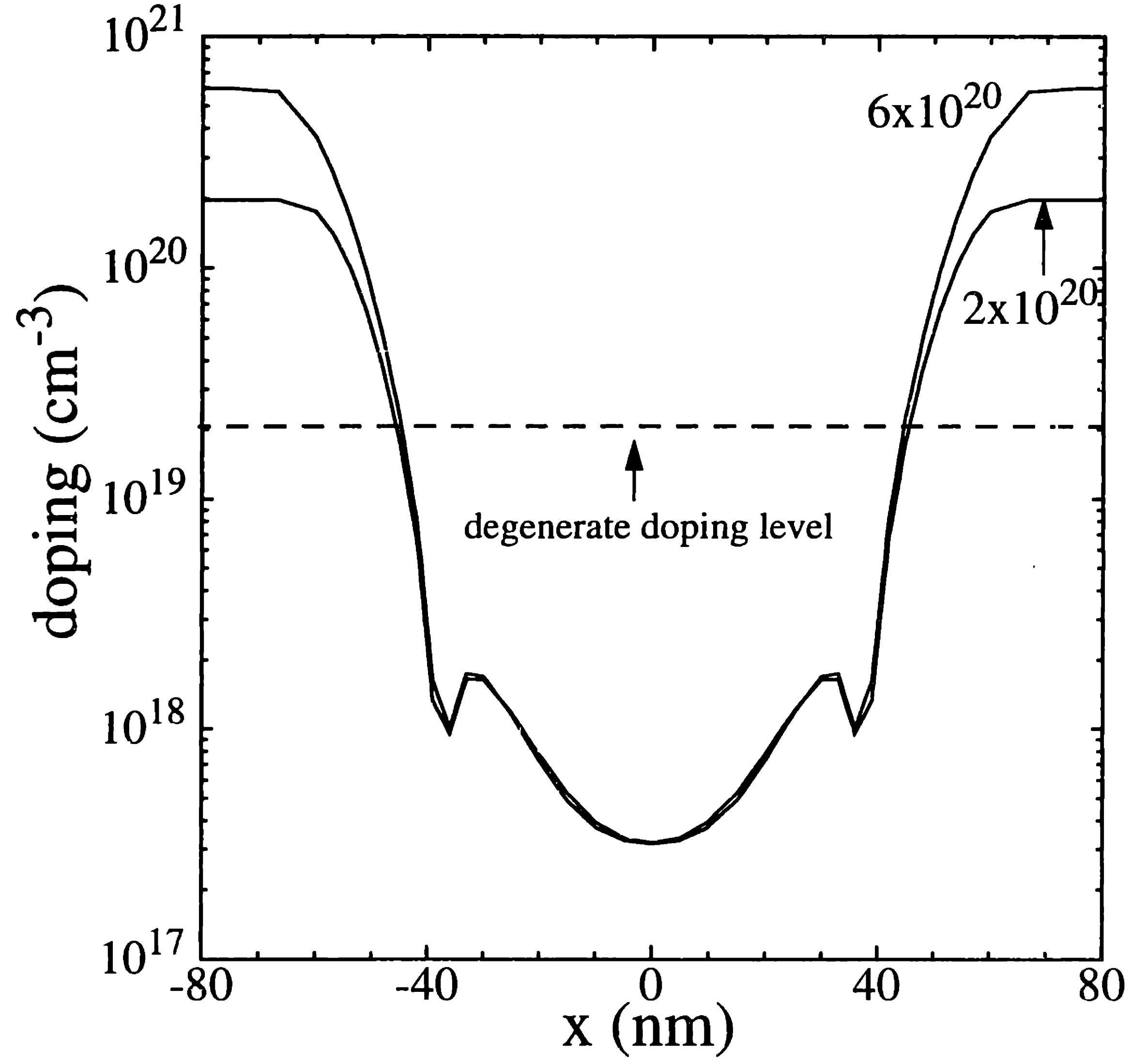

Figure 4.4.3: Lateral profiles at the oxide-silicon interface extracted using different assumed peak source/drain doping concentrations. Note that the doping profiles below approximately $2x10^{19}$ cm^{-3} are almost identical. Sensitivity of the technique decreases above the degenerate doping level. The quality of fit to the data is almost identical in both cases.

4.5 Conclusion

This chapter covers the subthreshold I-V inverse modeling methodology for the extraction of 2D doping profiles. Examples using short-channel (~ 100 nm) SSR and STEP devices having halo dopings are used. The limitations and sensitivity issues of the technique are discussed using simulations. The main advantages include: (1) it is capable of extracting 2D doping profile (including channel-length) of deep submicron devices because of its immunity to parasitic resistance, capacitance, noise, and fringing electric fields; (2) it does not require any special test structures since only subthreshold I-V data are used; (3) it is non-destructive; (4) it has very little dependence on mobility and mobility models; (5) it is easy to use since data collection and preparation are straightforward; and (6) it can be extended to the accurate calibration of carrier transport models using I-V characteristics at high current levels, because errors associated with uncertainties in doping profiles are removed. The calibration of transport models is discussed in Chapter 5.

Chapter 5

Transport Model Calibration

5.1 Introduction

The accurate simulation of I-V characteristics requires accurate transport (mobility) models [76-91]. Critical considerations for these models include: (1) they should be able to model the essential physical phenomena encountered in the device, such as vertical electric field degradation of mobility and velocity saturation/overshoot; (2) in the case of empirical or semi-empirical models, the model parameters have to be accurately calibrated to match I-V characteristics in all regions of operation, and for a wide range of devices; and (3) although not strictly required, it is preferred that the models be as simple as possible, without using an excessive number of parameters, because having too many parameters obscures the physical meaning, and makes calibration difficult. In this work, the applicability of the different transport models available in the device simulator MEDICI has been evaluated using experimental data of the SSR and STEP devices discussed in Chapter 4. It was found that the Lombardi mobility model [89] provides very accurate results for modeling the vertical electric field effects, while the Caughey-Thomas [90] expression was sufficient for modeling high lateral field (i.e., velocity saturation) effects, even for sub-100-nm channel-length devices. Moreover, for the experimental devices studied, it was found that by proper calibration, drift-diffusion models are found to produce almost identical results as more sophisticated hydrodynamic models [43, 91] that also take into account carrier heating effects in rapidly varying electric fields.

One of the most common difficulties in calibrating transport models, however, is the uncertainty of the doping profile to be used. Accurate knowledge of the doping profile from inverse modeling (see Chapter 4) removes this uncertainty and reduces the ambiguities in the transport model evaluation. As it will become clear later on, the use of accurate doping profiles allows for highly accurate calibration, leading to excellent predictive capability of I-V characteristics. In this chapter, the Lombardi and Caughey-Thomas models are reviewed, and a transport model calibration methodology using inverse modeling is described. The calibrated mobility, as function of electric field and doping concentration, is presented.

5.2 Lombardi Mobility Model

The Lombardi mobility model [89] is a semi-empirical model for electron (and hole) mobility in silicon inversion layers. It combines mobility expressions for both surface (silicon-oxide interface) and bulk silicon, and is composed of terms describing surface acoustic phonon scattering, surface roughness scattering, and concentration-dependent bulk silicon mobility. Unlike many other models that are applicable only at the silicon-oxide interface, and require using vertical grid spacings that are larger than the inversion layer thickness near the interface [43], the Lombardi model is a "local" model in that for every grid point, the mobility is a function only of the vertical electric field, doping concentration, and temperature, independent of which grid point it is or how far it is from the interface. Consequently it is able to seamlessly model both the subthreshold and above-threshold regions. On the other hand, this is not possible for mobility models that are applied only at the interface because the large grid spacing near the interface that is required by these models makes simulation of the subthreshold region highly inaccurate (simulation of the subthreshold region requires a fine 2D grid). Although, strictly speaking, the Lombardi model is intended only for modeling mobility in silicon inversion layers, it is also applicable in modeling source/drain series resistance (Chapter 6) because in the absence of large vertical electric field, the model reduces to a concentration dependent bulk mobility model, which is what should be used when modeling source/drain series resistance. For modeling lateral electric field (velocity saturation) effects, however, a separate mobility model such as the Caughey-Thomas model is needed.

The Lombardi model is based on the Mathiessen rule:

$$\frac{1}{\mu_s} = \frac{1}{\mu_{ac}} + \frac{1}{\mu_{sr}} + \frac{1}{\mu_{bulk}}, \tag{5.2.1}$$

where μ_s is the mobility accounting for surface and bulk effects, μ_{ac} is the mobility due to surface acoustic phonon scattering, μ_{sr} is the mobility due to surface roughness scattering, and μ_{bulk} is the concentration dependent bulk silicon mobility. The acoustic phonon mobility, obtained from a 2D deformation potential theory [86-88], is assumed to have the function form

$$\mu_{ac} = \frac{B}{E_\perp} + \frac{CN_A^D}{T\sqrt[3]{E_\perp}}, \tag{5.2.2}$$

where B, C, and D are fitting parameters, and $E_\perp$, T, and N_A are, respectively, the vertical electric field, temperature, and doping concentration. The surface roughness mobility is assumed to have the form

$$\mu_{sr} = \frac{\delta}{E_\perp^2}, \tag{5.2.3}$$

where δ is a fitting parameter. The concentration dependent bulk mobility is taken from [84] and has the same form

$$\mu_{bulk} = \mu_o + \frac{\mu_{max}(T) - \mu_o}{1 + (N_A/C_r)^\alpha} - \frac{\mu_1}{1 + (C_s/N_A)^\beta}, \tag{5.2.4}$$

where μ_o, μ_1, C_r, C_s, α, and β are all fitting parameters, and $\mu_{max}(T)$ is given by

$$\mu_{max}(T) = \mu_{maxo}\left(\frac{T}{300}\right)^{-\gamma}, \tag{5.2.5}$$

with μ_{maxo} and γ being fitting parameters.

The values of the fitting parameters are presented in section 5.4, where a calibration methodology is discussed. The Lombardi model is appropriate when V_{DS} is low. At high V_{DS}, velocity saturation effects should be included. The Caughey-Thomas model is found to be sufficient in providing reasonably accurate results.

5.3 Caughey-Thomas Expression

The Caughey-Thomas [90] expression is given by:

$$\mu = \frac{\mu_s}{\left[1 + \left(\frac{\mu_s E_{//}}{v_{sat}}\right)^{\lambda}\right]^{1/\lambda}} \tag{5.3.1}$$

where μ is the final mobility that takes both vertical and lateral electric field degradations into account, μ_s is the mobility obtained using the Lombardi mobility model, $E_{//}$ is the lateral electric field, v_{sat} is the saturation velocity, and λ is a fitting parameter. At 300 K, $v_{sat} = 10^7$ cm/s for electrons [43]. The value of λ is obtained by matching simulated data to experimental data at high V_{DS}, as explained in the next section.

5.4 Transport (Mobility) Model Calibration Methodology using Inverse Modeling

Besides obtaining 2D doping profiles using subthreshold I-V characteristics, inverse modeling can also be used to calibrate transport models using above-threshold I-V characteristics. The general procedure is to first obtain a 2D doping profile from the subthreshold I-V data, the doping profile is then fixed, and appropriate transport model parameters are optimized so as to match the simulated above-threshold I-V data to the corresponding experimental data.

As discussed in previous sections, the Lombardi mobility model coupled with the Caughey-Thomas expression was found to provide very good results with the SSR and STEP devices. Consequently the calibration methodology that follows is based on the use of these models. Since the Lombardi model only describes effects associated with the vertical electric field, while the Caughey-Thomas expression the lateral field, the two models can be calibrated separately. First, only the Lombardi model is calibrated using low V_{DS} data (low lateral field). Model parameters are then fixed, followed by the inclusion and calibration of the Caughey-Thomas model using all V_{DS} data. A schematic diagram of the entire procedure is shown in Fig. 5.4.1. Details are provided in the following sections.

5.4.1 Calibration of Vertical Electric Field (Lombardi) Model

When calibrating only vertical field models, low V_{DS} data should be used. Using low V_{DS} data, however, the parasitic series resistance of the source/drain (hereby denoted by R_{ext}) that is external to the simulation structure becomes important because the voltage dropped in this resistance can be a sizable fraction of the total V_{DS}. Therefore, an appropriate value of R_{ext} should be used. However, R_{ext} is in general not exactly known. As a result, an approximate value of R_{ext} is used as an initial guess (see Chapter 6 for a discussion of R_{ext}), and it is allowed to be varied around this value (i.e., as a fitting parameter).

A schematic diagram describing this step of the calibration procedure is shown in Fig. 5.4.1.1, which is an expanded view of the optimization loop shown in Fig. 5.4.1. The main goal here is to keep the parameters as close as possible to their original values (i.e., values given in [89]). For instance, if the end result has a low sensitivity on a particular parameter, it is best to keep that parameter fixed. Initially, only one parameter is allowed to be varied (namely, R_{ext}). If changing R_{ext} (to within certain reasonable bounds) is sufficient in meeting the convergence criterion of the optimization, no more parameters are to be varied, and the calibration for this step is complete; otherwise, an additional parameter is varied and is optimized simultaneously with the previous one(s). This procedure is repeated until convergence criterion is met.

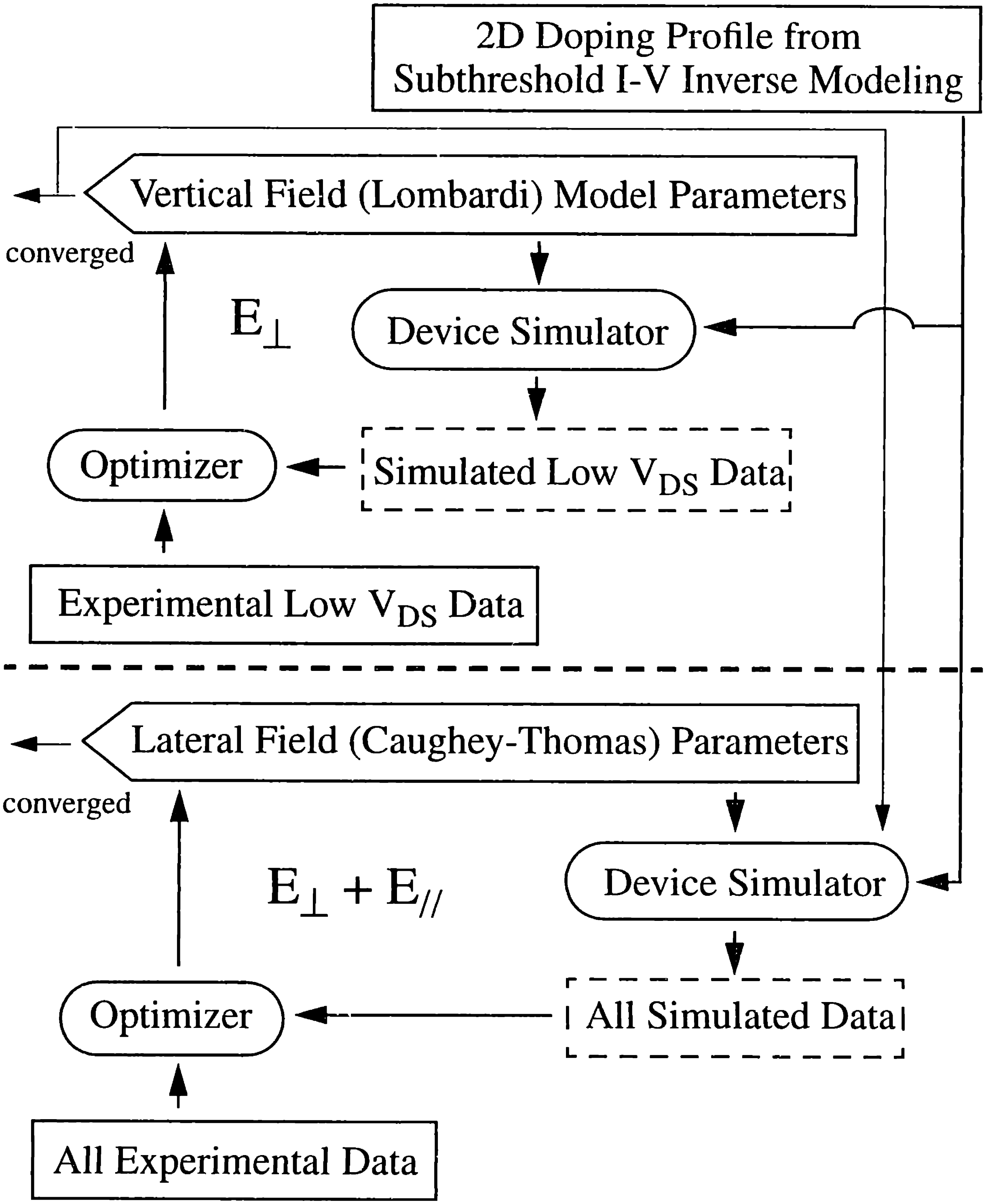

Figure 5.4.1: Schematic diagram showing the transport model calibration methodology.

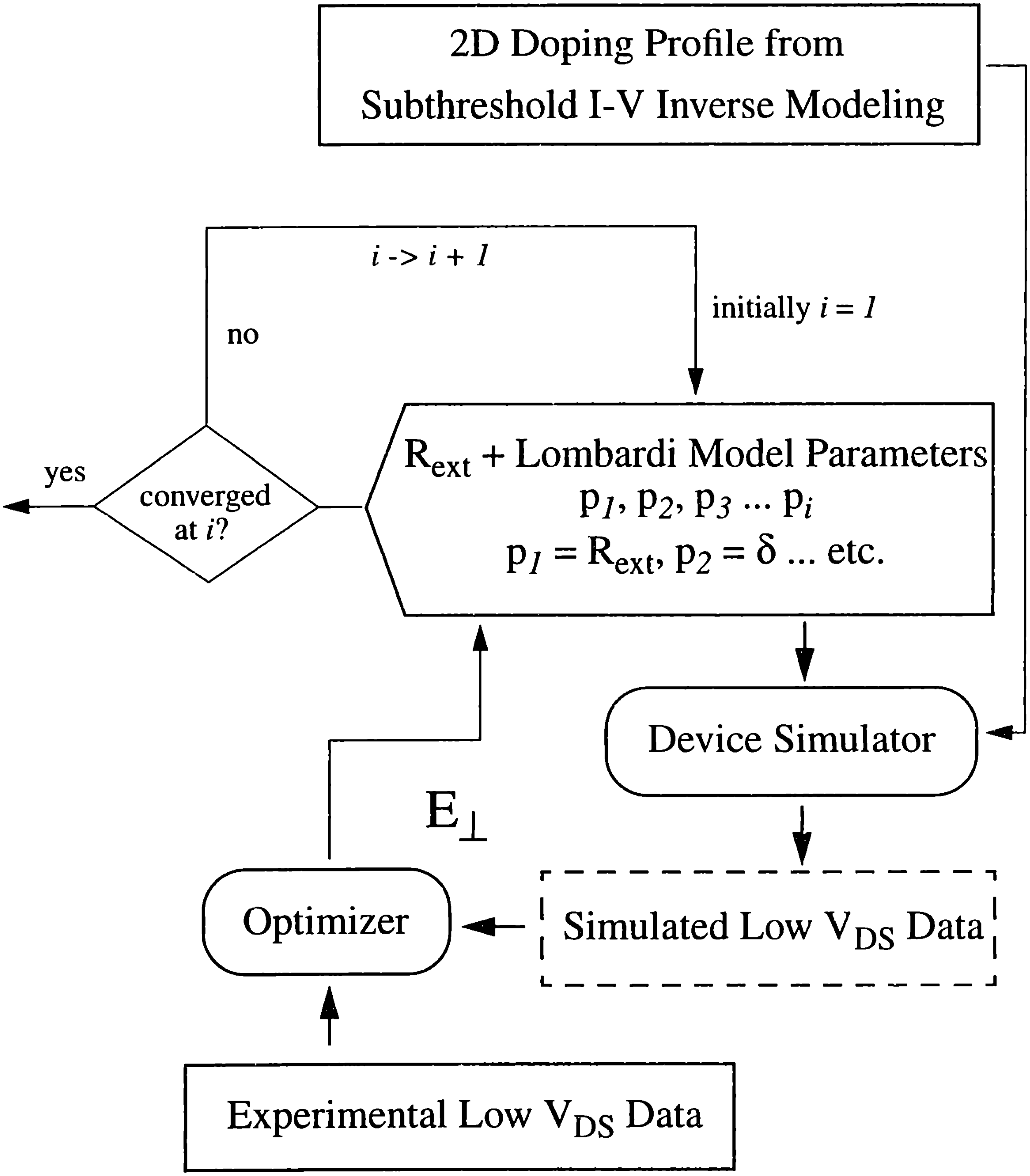

Figure 5.4.1.1: Expanded view of the optimization loop for vertical field calibration in Fig. 5.4.1. Initially, only one parameter is allowed to be varied. If convergence criterion for optimization is met, no more parameters are optimized; otherwise, an additional parameter is allowd to be varied and optimized. The procedure is repeated until convergence criterion is met.

Except for R_{ext}, which is the first parameter to be optimized, the model parameters to be varied in the optimization should begin with the one that has the largest effect on the end result. For the Lombardi model, it is δ because at high $E_{\perp}$, the surface roughness scattering is the dominant mechanism limiting μ_s, except at very high doping concentrations, when the bulk mobility becomes dominant. The next parameter to be varied should be B, followed by C.

Quite unexpectedly, however, it was found that no changes for the Lombardi model parameters from those given in [89] were required to match all available experimental SSR and STEP device data (at low V_{DS}). The only parameter that needed be adjusted was R_{ext}, which, for most of the 10 μm width devices, was found to be 100 - 200 Ω-μm, a very reasonable value when parasitic source/drain series resistance is taken into account (see Chapter 6). This is believed to be a consequence of the accuracy of the doping profiles used, as well as the accuracy of the model parameters given in [89]. An example showing the match between simulated and experimental I-V data at low V_{DS} for a SSR and a STEP device is shown in Fig. 5.4.1.2. The model parameters for the Lombardi model are summarized in Table 5.4.1.1.

Table 5.4.1.1: Model parameters for Lombardi mobility model (electrons).

μ_{ac}	B	4.75×10^{7}	Note: all parameters here are chosen such that electric fields have dimensions of V/cm, and doping concentrations have dimensions of cm^{-3}.
	C	1.74×10^{5}	
	D	0.125	
μ_{sr}	δ	5.82×10^{14}	
μ_{bulk}	μ_o	52.2	
	μ_1	43.4	
	μ_{maxo}	1417	
	C_r	9.68×10^{16}	
	C_s	3.43×10^{20}	
	α	0.68	
	β	2.00	
	γ	2.50	

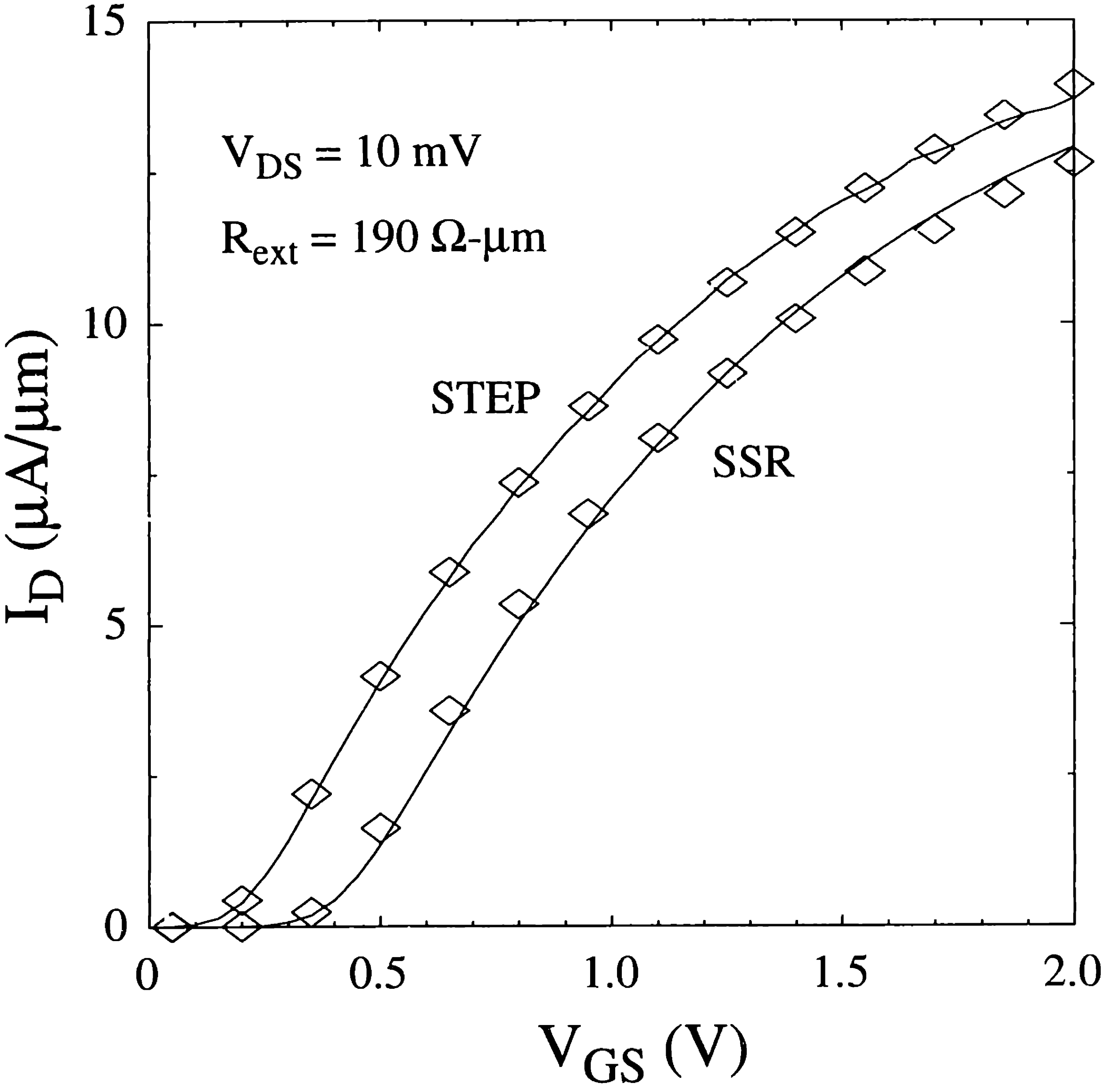

Figure 5.4.1.2: Comparison between experimental and simulated (symbols) I-V data at low V_{DS} of a SSR and a STEP device. Both devices have approximately a 100 nm channel-length and a 10 μm width. Using a R_{ext} = 190 Ω-μm, very good match between experimental and simulated data was obtained in both devices.

5.4.2 Calibration of Lateral Electric Field (Caughey-Thomas) Model

The vertical field (Lombardi) model is calibrated at low lateral field so that velocity saturation effects do not obscure the calibration. Once the vertical field model has been calibrated, the lateral field (Caughey-Thomas) model is added to the simulation and optimization, and the data set that is required to be matched is extended to contain both low and high V_{DS} values. Since the vertical field model has already been calibrated, the associated model parameters are fixed; only the parameters related to the lateral field model are optimized.

For the Caughey-Thomas model, the only two fitting parameters are λ and v_{sat}. A procedure similar to the calibration of the Lombardi model was used (section 5.4.1). As it turned out, the only parameter that needed be changed was λ (i.e., there was no need to alter v_{sat} from its original value of 10^7 cm/s). It was found that a λ value of 1.25 +/- 0.03 would fit all of the device data (devices having various SSR and STEP channel doping profiles), and under all bias conditions.

Figure 5.4.2.1 shows a comparison between the experimental and simulated I-V characteristics of a SSR device having a 100 nm channel-length and a 10 μm width using an optimized $\lambda = 1.25$ and a $R_{ext} = 190\ \Omega.\mu m$. Note that R_{ext} that must be added is somewhat arbitrary, since it depends on the peak doping of the source/drain extensions (which the subthreshold I-V inverse modeling technique cannot extract), the length of the extensions used in the simulation, and other parasitic resistances. Nevertheless, it is a very reasonable value (see Chapter 6 for a discussion of R_{ext}). It can be seen that once an accurate doping profile was used, only one parameter (namely, λ) in the transport models was needed to be modified in order to match all of the device data.

In order to demonstrate the accuracy, applicability, and universality of the calibrated transport models, simulations using the calibrated models have been performed on devices having various doping profiles. Figure 5.4.2.2 shows a comparison between the simulated and measured I-V characteristics of a 120 nm channel-length STEP device. Here, the doping profile of the device obtained from subthreshold I-V inverse modeling was used in the simulations. Despite the differences in channel-length, doping profile, and threshold voltage between the SSR and STEP devices, the universality of the calibrated transport models demonstrates their predictive capability and correctness, made possible by the use of accurate doping profiles obtained from inverse modeling.

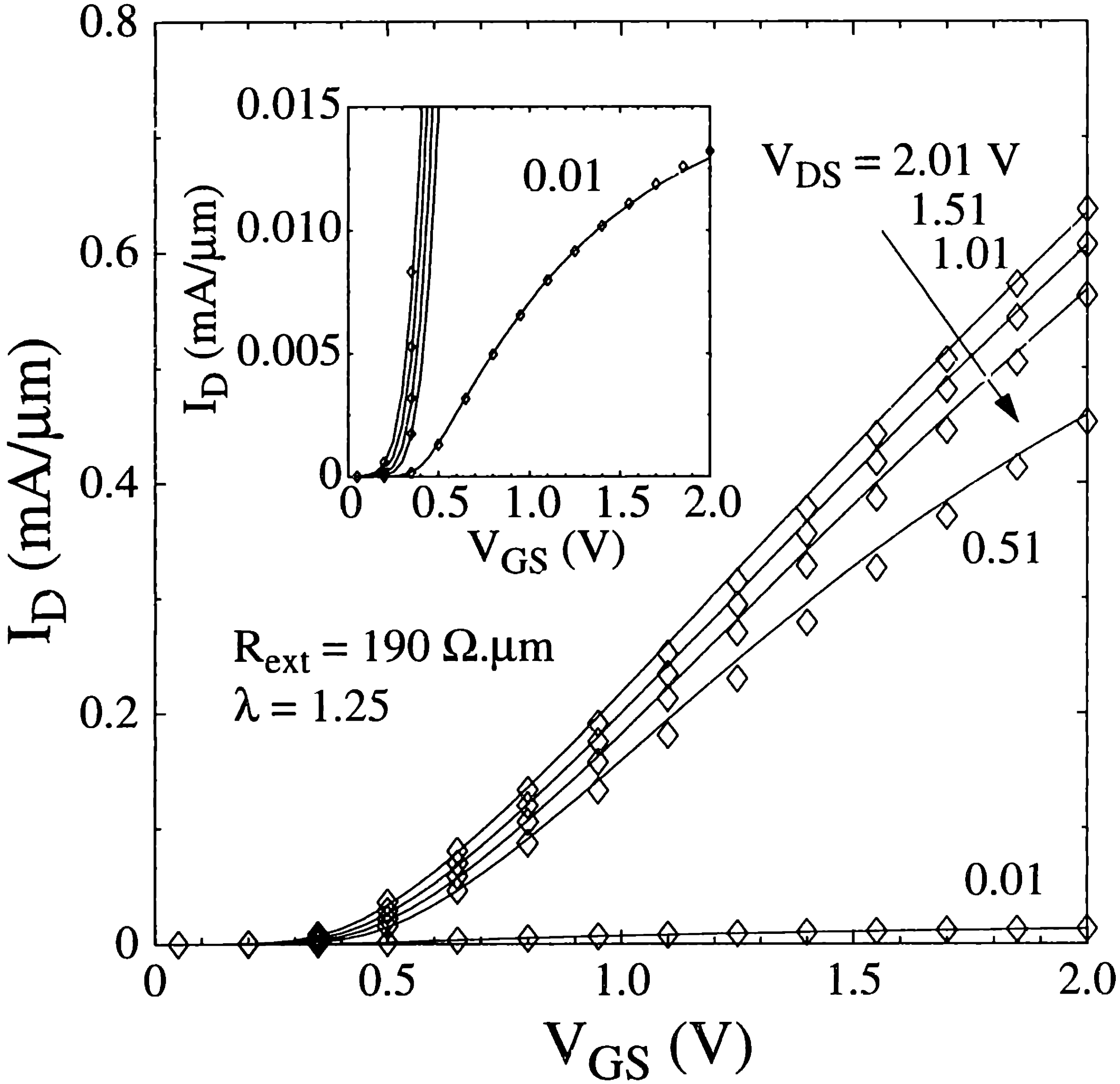

Figure 5.4.2.1: Comparison between experimental and simulated (symbols) I-V characteristics of a SSR device having a 100 nm channel-length and a 10 μm width using an optimized $\lambda = 1.25$ and a $R_{ext} = 190\ \Omega.\mu m$.

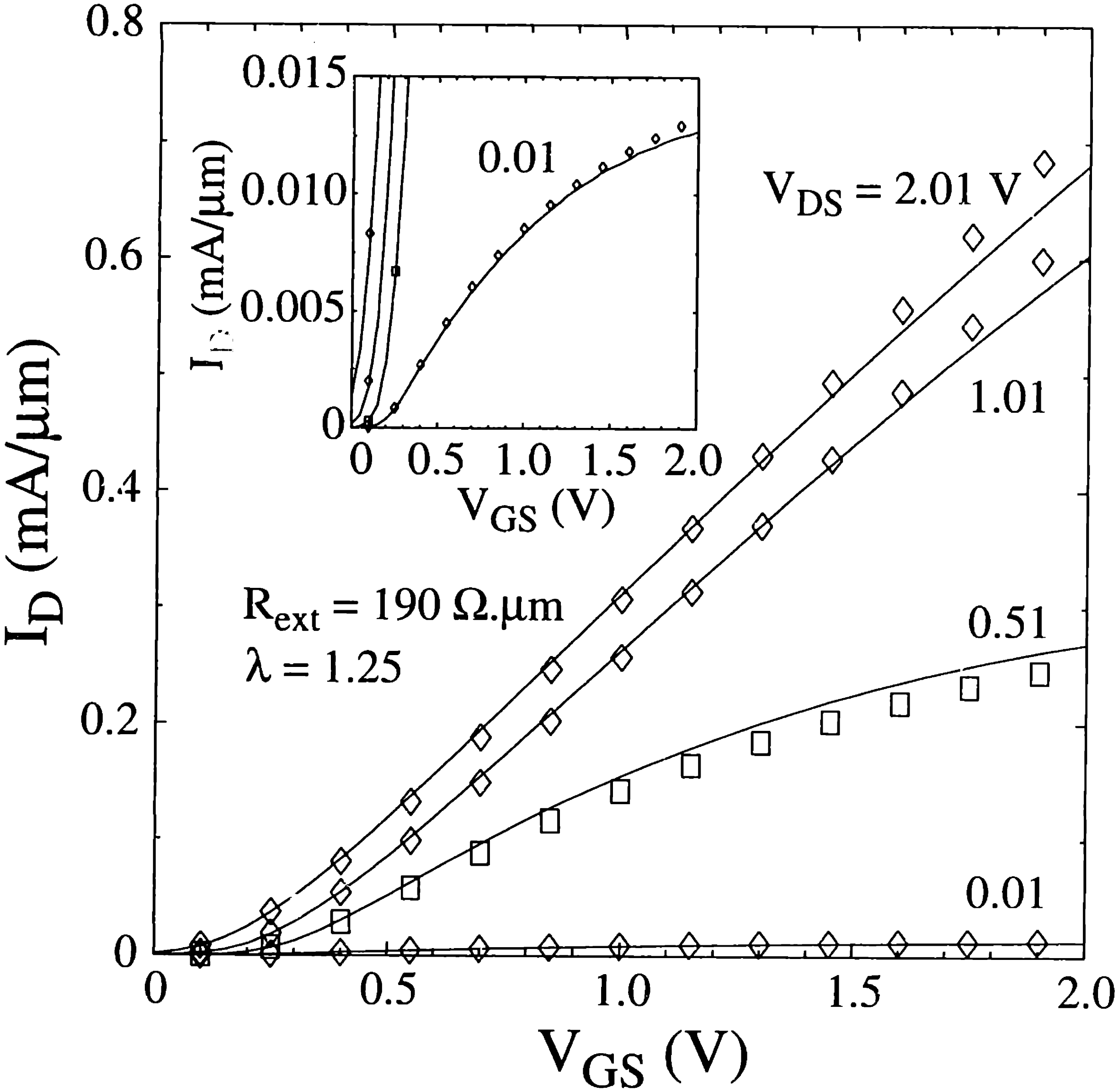

Figure 5.4.2.2: Comparison between experimental and simulated (symbols) I-V characteristics of a STEP device having a 120 nm channel-length and a 10 μm width using the same mobility models and parameters as those in Fig. 5.4.2.1. Here, the doping profile of the device extracted from inverse modeling was used in the simulations.

5.5 Calibrated Transport Models

The transport models discussed in this chapter have mobilities μ that are functions of vertical electric field $E_\perp$, lateral electric field $E_{//}$, doping concentration N_A, and, though only briefly discussed, temperature T. In this section, plots showing the various dependences of μ (calibrated) at 300 K are presented.

Figure 5.5.1 shows the dependence of μ on N_A, with $E_\perp$ as a parameter. Here, $E_{//} \sim 0$. The dependence of μ on $E_\perp$, with N_A as a parameter, is shown in Fig. 5.5.2. Again, $E_{//} \sim 0$. The various components of the Lombardi model at $E_\perp = 5\text{x}10^4$ V/cm and 10^6 V/cm are presented in Figs. 5.5.3(a) and (b) respectively. It can be seen that at low $E_\perp$, the mobilities associated with the acoustic phonon μ_{ac} and surface roughness μ_{sr} scattering are very large, so that the bulk mobility μ_{bulk} is the limiting factor for μ. At high N_A, μ is again limited by μ_{bulk}. Consequently, although the model is intended for the inversion layer, it is also applicable for the source/drain because μ reduces to μ_{bulk} at the low $E_\perp$ and high N_A of the source/drain (see Chapter 6).

Figure 5.5.4 shows the combined effects of the Lombardi and Caughey-Thomas models as a function of N_A, at various $E_\perp$ and $E_{//}$. It can be seen that at high $E_{//}$, μ is quickly degraded due to velocity saturation.

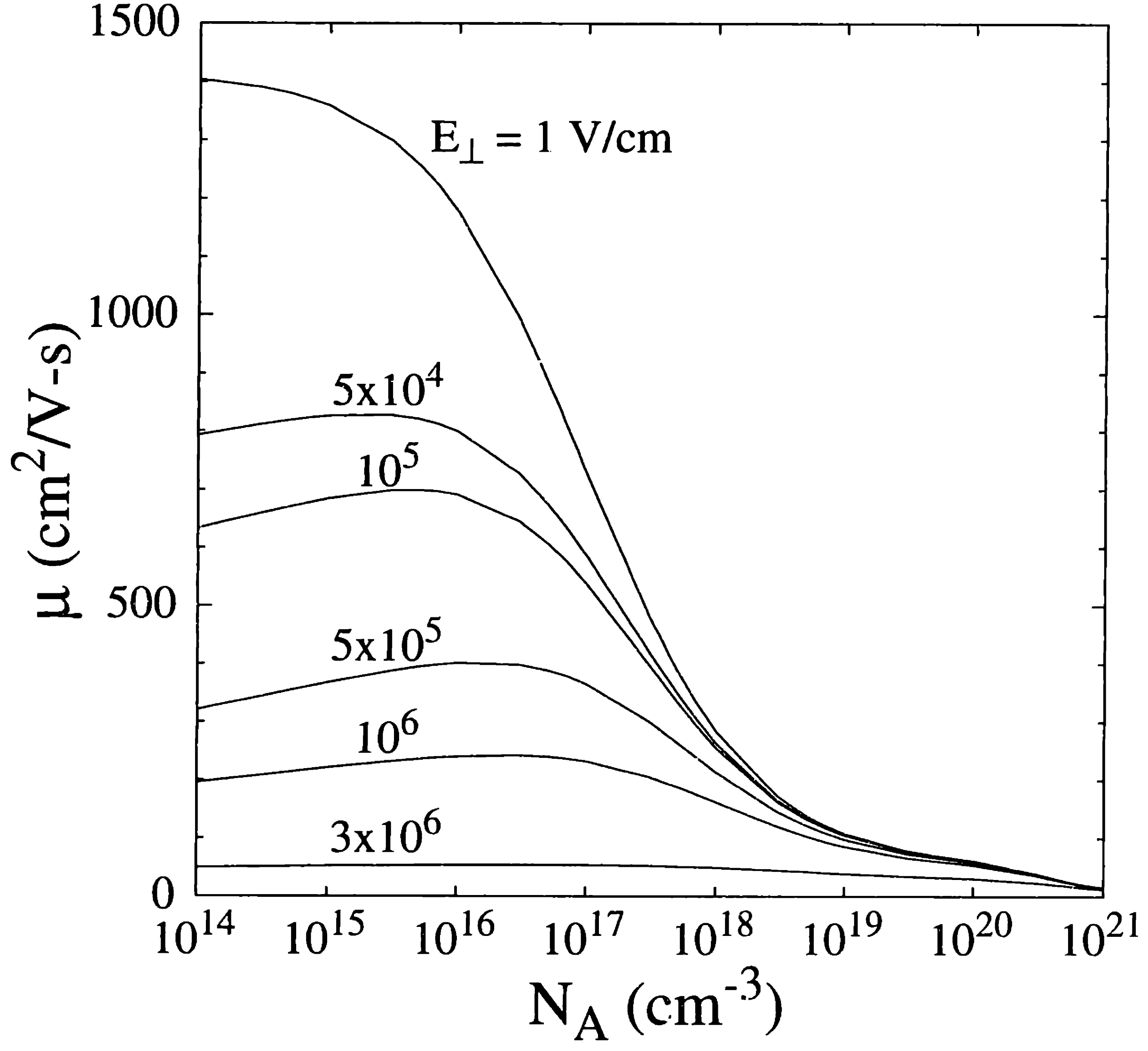

Figure 5.5.1: Dependence of μ on N_A, with $E_\perp$ as a parameter. $E_{//} \sim 0$.

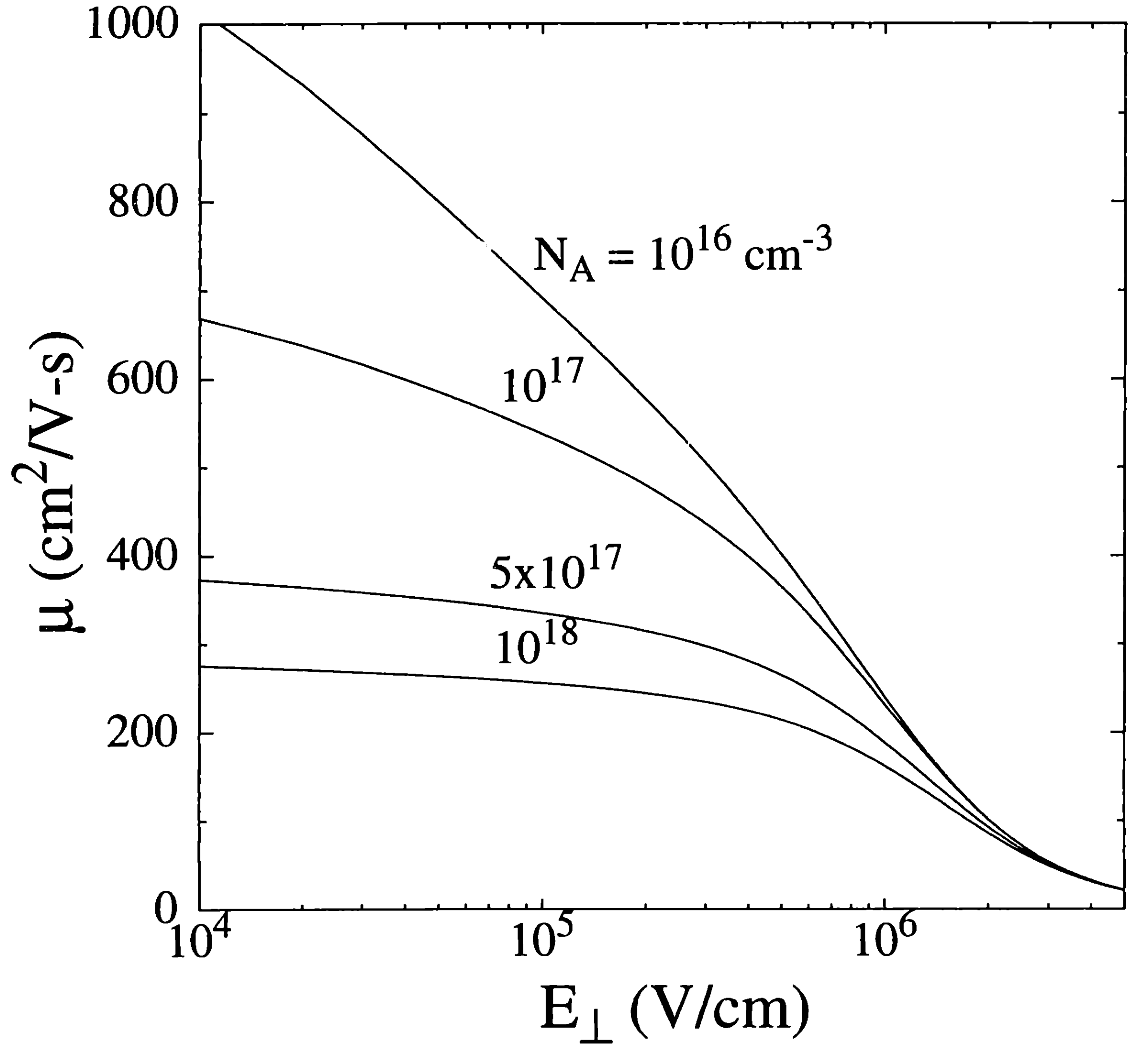

Figure 5.5.2: Dependence of μ on $E_\perp$, with N_A as a parameter. $E_{//} \sim 0$.

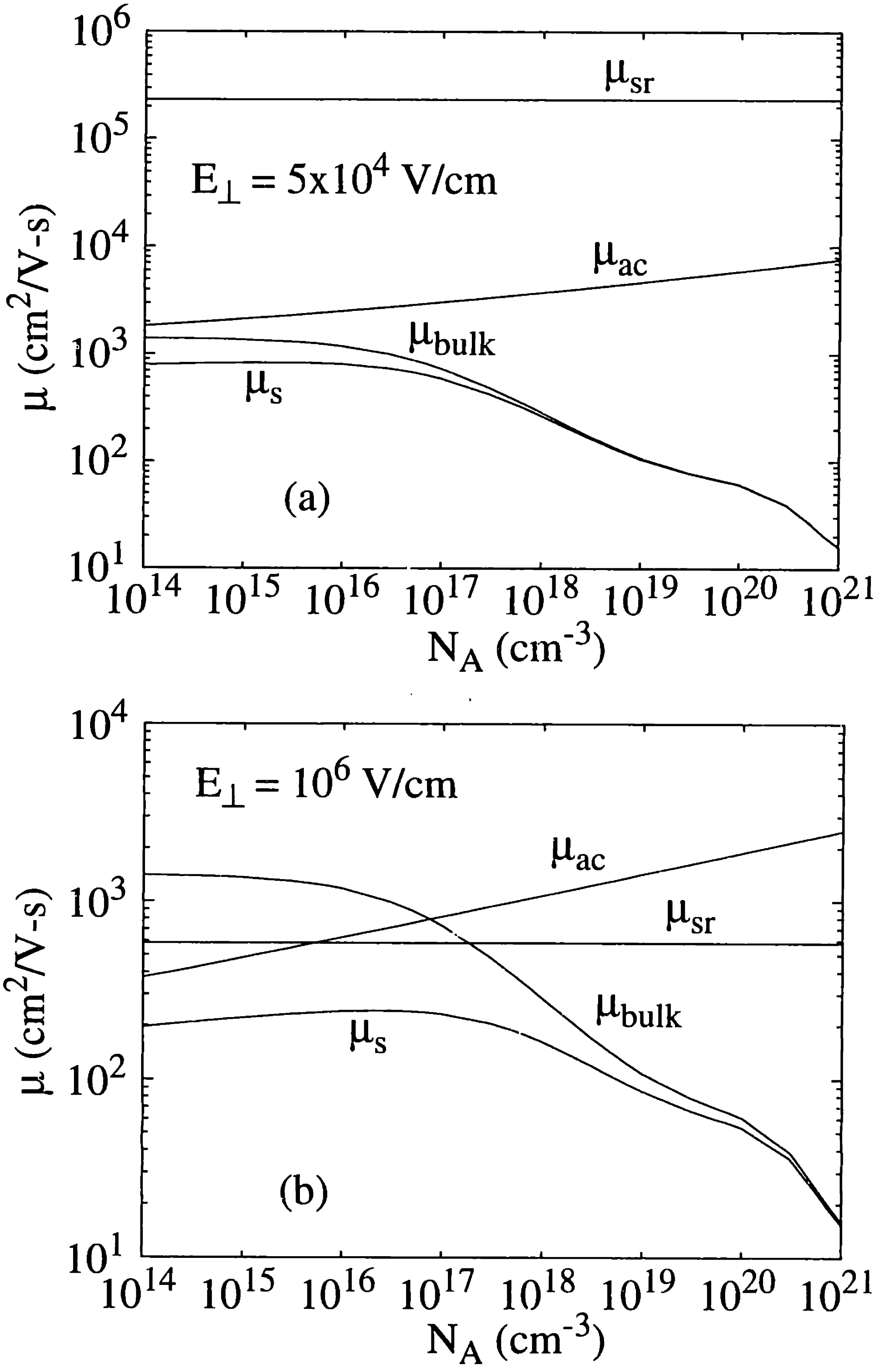

Figure 5.5.3: Dependence of μ on N_A showing the μ_{ac}, μ_{sr}, and μ_{bulk} components for (a) $E_\perp = 5\text{x}10^4$ V/cm; and (b) $E_\perp = 10^6$ V/cm. $E_{//} \sim 0$.

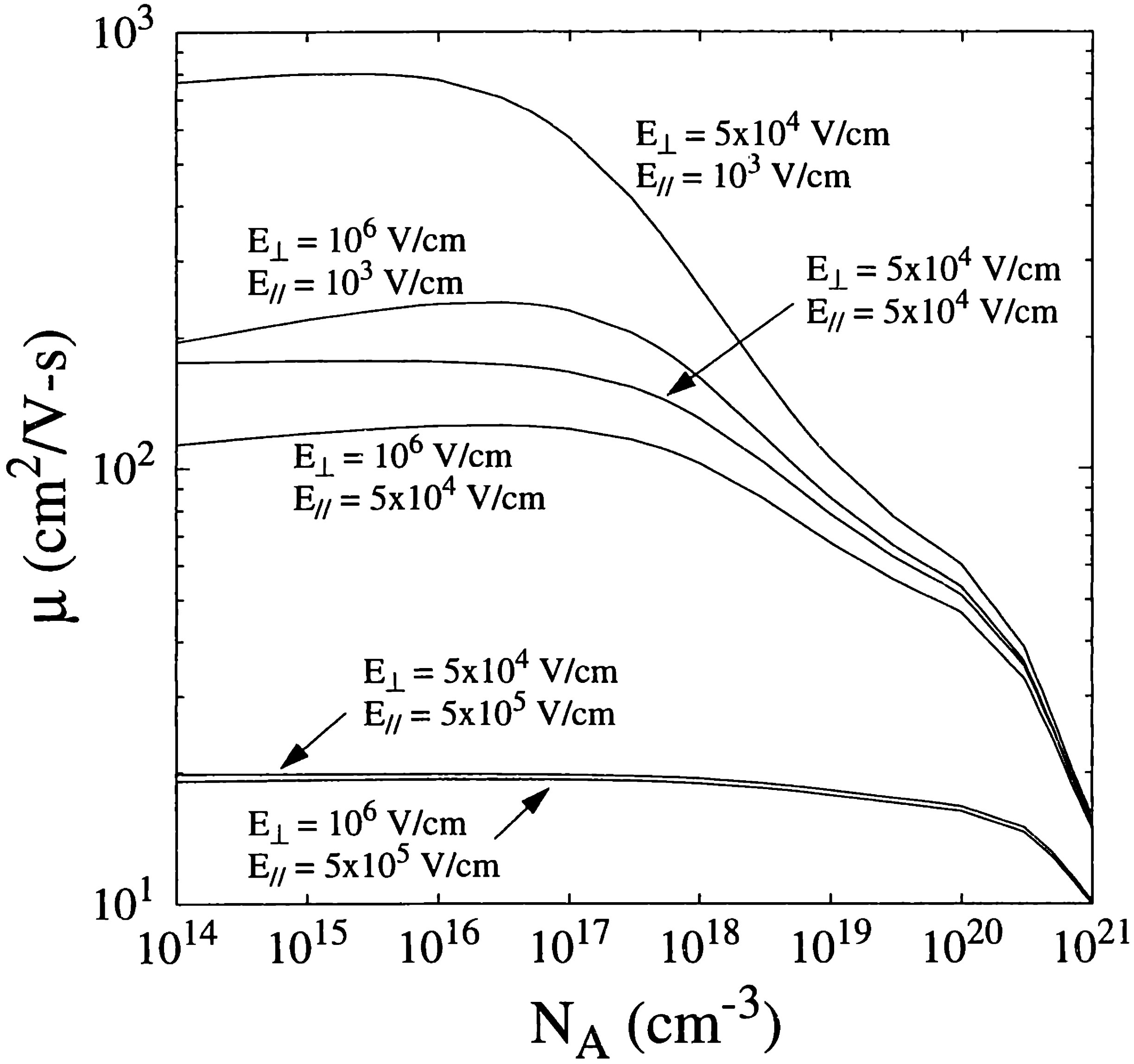

Figure 5.5.4: Dependence of μ on N_A, with $E_\perp$ and $E_{//}$ as parameters.

5.6 Hydrodynamic Simulations

Previous sections show that for the devices studied in this work, if appropriately calibrated transport models are used, simple drift-diffusion simulations (see Chapter 2) are able to produce I-V characteristics that accurately match the corresponding experimental data, and time-consuming hydrodynamic simulations are not needed. As a result, the purpose of this section is to demonstrate that simulation results produced by the hydrodynamic models in MEDICI can be accurately reproduced using simple drift-diffusion theory.

In hydrodynamic simulation, in addition to the drift-diffusion equations discussed in Chapter 2, energy balance equations are solved so that carrier heating effects are also included (for the equations implemented in MEDICI, the reader is referred to the MEDICI user's manual [43]). In MEDICI, an effective field approach is implemented. Based on the local carrier temperature, an effective field E_{eff} is calculated using the Caughey-Thomas expression, i.e.,

$$\mu(E_{eff})E_{eff}^{2} = \frac{3(u-u_o)}{2\tau}, \tag{5.6.1}$$

where $\mu(E_{eff})$ is the mobility calculated using eqn. (5.3.1), with $E_{//} = E_{eff}$, u is the carrier (e.g., electron) thermal potential (= kT/q), u_0 is the lattice thermal potential (= kT_0/q), and τ is the energy relaxation time. Upon solving for E_{eff} in eqn. (5.6.1), the E_{eff} thus calculated is inserted back into eqn. (5.3.1) to calculate the local mobility. Consequently, if this model is used, an appropriate value of an additional parameter, i.e., τ, needs to be used. For the results that follow, it is found that when the hydrodynamic model is invoked, very good agreement between simulation and experimental data can be achieved if appropriate values of λ [from eqn. (5.3.1)] and τ are used. Nevertheless, very good agreement can also be achieved using drift-diffusion simulations.

Figures 5.6.1 and 5.6.2 show, respectively, comparisions of experimental I-V characteristics with simulated I-V characteristics at high V_{DS} (0.21, 1.01, and 2.01 V) and low V_{DS} (0.01 V) of the same device as in Fig. 5.4.2.2. As can be seen, when appropriate parameters are used, very good match with the experimental data can be achieved. It is found that if the value of λ calibrated for drift-diffusion simulations is to be retained (i.e., $\lambda = 1.25$), τ should be set to $\tau = 0.1$ ps in order to match the simulated results to the experimental data. On the other hand, if $\tau = 0.2$ ps is used (default value in MEDICI [43]), the value of λ has to be changed to $\lambda = 1$ in order to match the data. As expected, at low lateral electric field, i.e., low V_{DS}, hardly any difference between the models can be observed (Fig. 5.6.2).

Results above show that even for devices having channel-lengths of around 100 nm, simple drift-diffusion simulations are capable to produce highly accurate results, even though hydrodynamic simulations are equally capable. With the added complication of uncertainties in τ to be used in hydrodynamic simulations, as well as the substantial increase in CPU time, drift-diffusion simulations are found to be more practical for this range of channel-lengths.

5.7 Conclusion

In this chapter, the extension of inverse modeling to the calibration of transport (mobility) models is presented. Using 2D doping profiles obtained from subthreshold I-V inverse modeling, transport models can be calibrated using above-threshold I-V characteristics. It was found that the Lombardi and Caughey-Thomas models produced very good results for the SSR and STEP devices, and hydrodynamic models are not necessary even for channel-lengths down to approximately 100 nm. A transport model calibration methodology is discussed, and the calibration results presented.

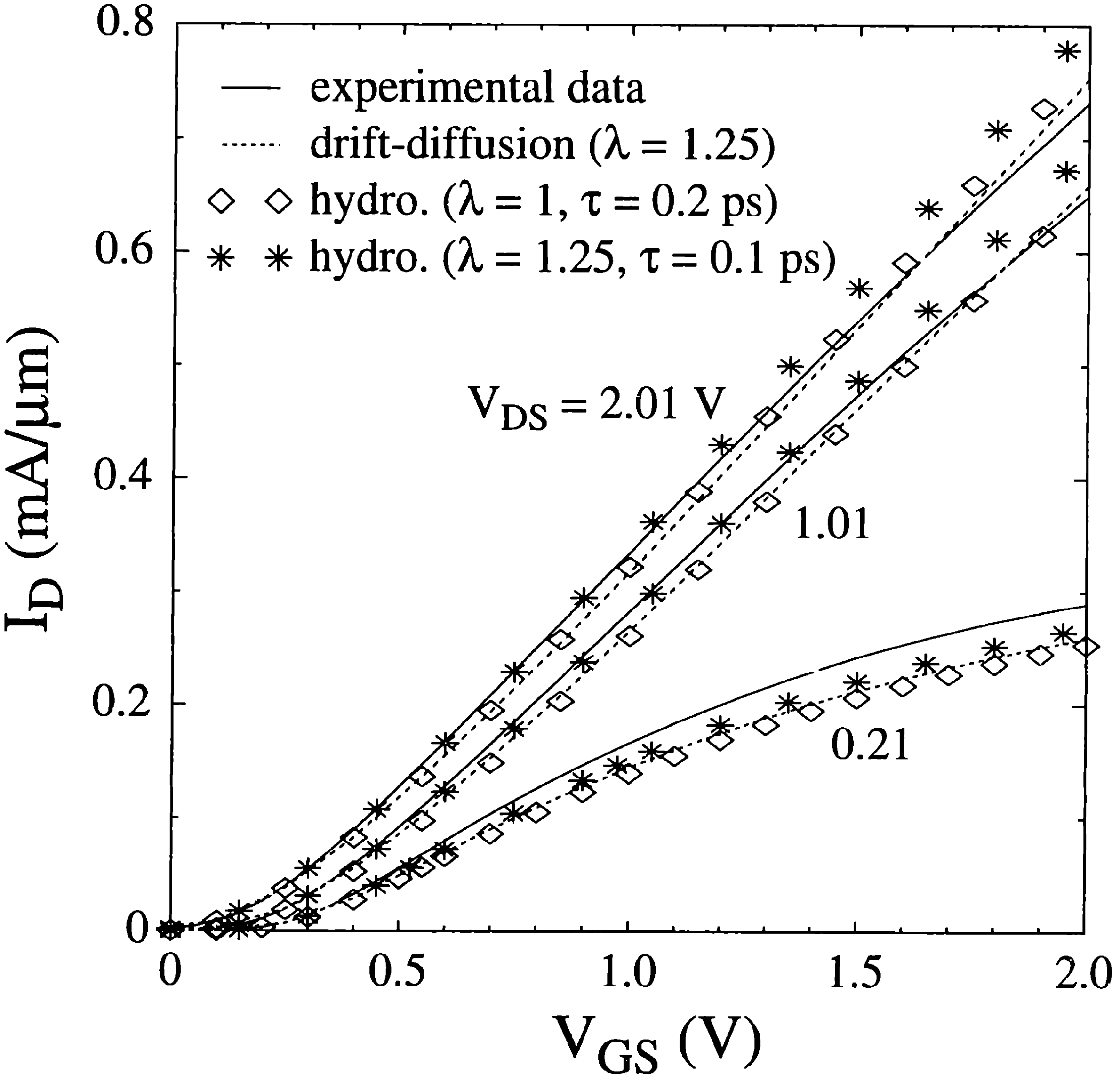

Figure 5.6.1: Comparison of experimental I-V characteristics with simulated I-V characteristics using drift-diffusion and hydrodynamic models at $V_{DS} = 0.21$ V, 1.01 V, and 2.01 V.

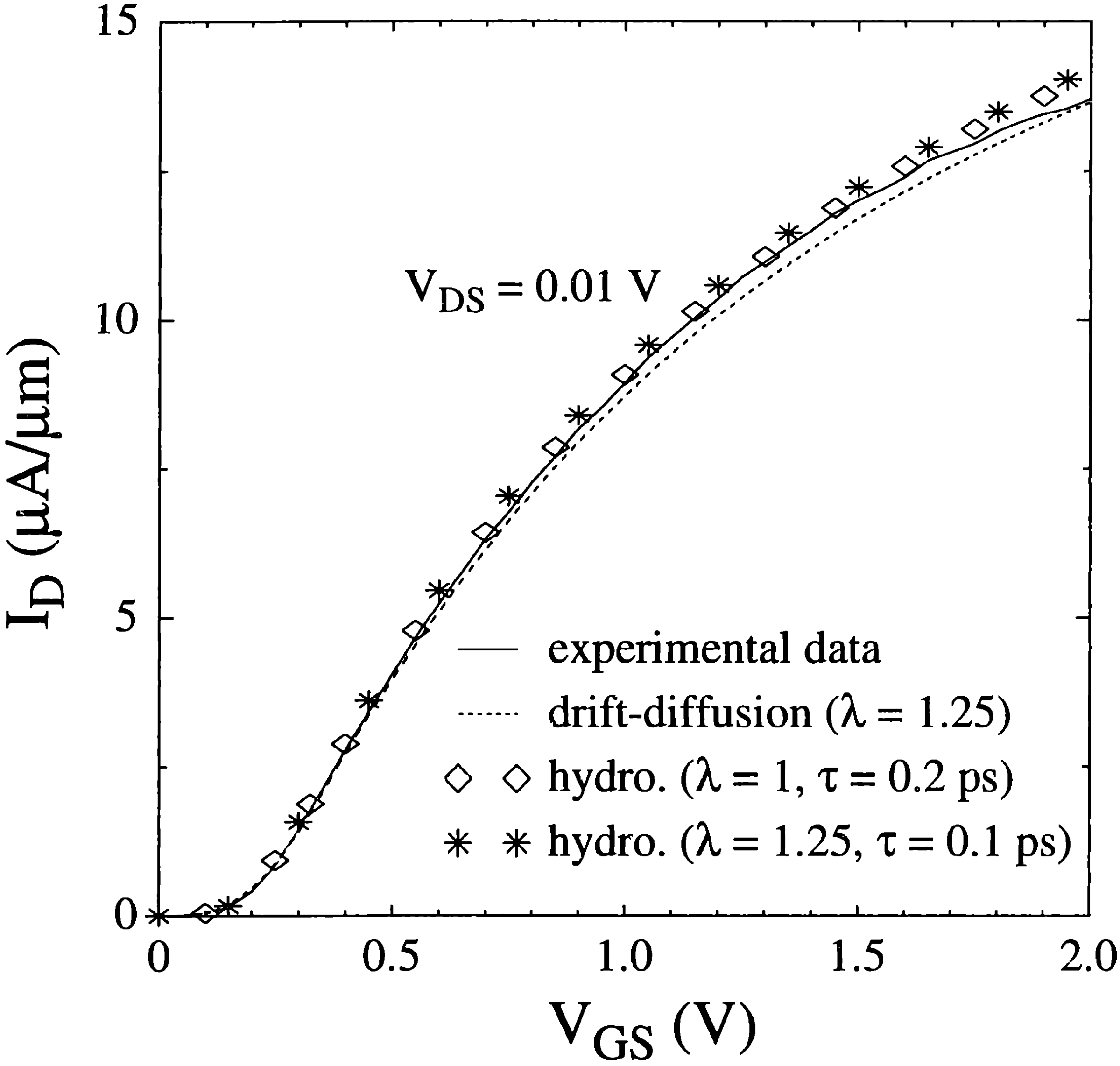

Figure 5.6.2: Comparison of experimental I-V characteristics with simulated I-V characteristics using drift-diffusion and hydrodynamic models at V_{DS} = 0.01 V.

Chapter 6

Parasitic Series Resistance

6.1 Introduction

It has been shown in previous chapters that using 2D doping profiles obtained from inverse modeling, it is possible to calibrate mobility models, leading to highly accurate predictive capabilities of device characteristics. In order to yield accurate results at low V_{DS}, however, it is important that an appropriate value of R_{ext} be used. This is especially important as devices are miniaturized [92-104], when the channel resistance becomes low while R_{ext} does not decrease at the same rate (or it may even increase). An understanding of the various contributions to R_{ext} is therefore an inseparable part of the characterization process. For this reason, this chapter is focused on a detailed analysis of the various contributions to this resistance.

The contributions to R_{ext} are the source/drain series resistance R_{SD}, the silicide/metal contact resistance, the parasitic resistance associated with the contact pads and probes, and, to a much lesser extent the cables and other parts of the measurement system. It is found that the total parasitic resistance associated with the contact pads, probes, and silicide/metal contact account for between 2 - 6 Ω of resistance, depending on the condition of the pads and probes, and quite independent of the device width. Consequently, R_{SD} is the most important contribution to R_{ext}, except for devices having very large widths.

6.2 Contributions to Source/Drain Series Resistance R_{SD}

R_{SD} is composed of the source resistance R_S and the drain resistance R_D (i.e., $R_{SD} = R_S + R_D$). Throughout the remaining discussion of this chapter, R_S and R_D are assumed to be equal, although in general they are not always the same and are in fact dependent upon biases applied at the gate and drain. This is a very reasonable assumption since the I-V characteristics are most sensitive to resistances at low V_{DS} (i.e., the depletion within the drain is negligible compared to that within the source).

There are four major components that contribute to R_{SD}: (1) spreading resistance R_{sp}; (2) accumulation resistance R_{ac}; (3) sheet resistance R_{sh}; and (4) contact resistance R_{co}. These resistances are shown schematically in Fig. 6.2.1. Not shown is the resistance contributed by the silicide/metal contact because this resistance can be made very small compared to the four components above.

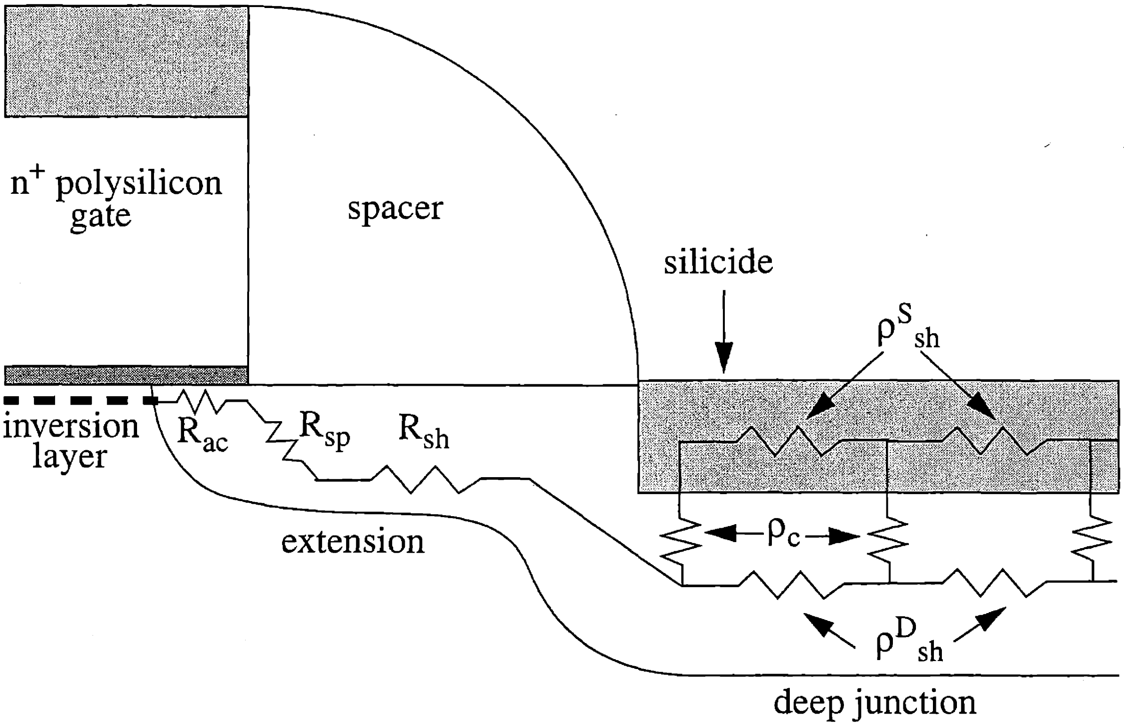

Figure 6.2.1: Schematic diagram of a source/drain junction showing various series resistance components. The distributed components, ρ^S_{sh},ρ^D_{sh}, and ρ_c are, respectively, the silicide sheet resistivity, diffusion sheet resistivity, and silicide/diffusion specific contact resistivity.

R_{sp} is the resistance due to the crowding of the current flow lines near the constriction between the source/drain junctions and the inversion layer [97-100]. The range that this current spreading takes place is on the same order as the junction depth. R_{sp} is therefore independent on the length of the source/drain, as long as it is longer than this range. Analytical expressions of R_{sp} (in Ω) have been derived by several groups, and can be written as:

$$R_{sp} = \frac{2\rho_{eff}}{\pi W} \ln\left(A\frac{X_j}{X_i}\right), \tag{6.2.1}$$

where W is the device width, ρ_{eff} the effective bulk resistivity, X_j the junction depth, X_i the inversion layer thickness, and, depending on the assumptions made as well as details of the derivations, the value of A is given as 0.58 [99], 0.75 [97], and 0.90 [100]. The logarithmic dependence makes the exact value of A unimportant. In general, since ρ_{eff} is both source/drain doping and bias dependent (e.g., through accumulation in the source/drain), so is R_{sp} [100].

R_{ac} is the component of the resistance in the region of the source/drain under the gate that becomes accumulated. It is given by [100]:

$$R_{ac} = \int_{region} \frac{dx}{\mu_{ac} W |Q_{ac}|}, \tag{6.2.2}$$

where μ_{ac} is the mobility of the accumulation layer, and $|Q_{ac}|$ is the accumulated charge, given by:

$$|Q_{ac}| = C_{ox}\left|V_{GS} - V_t^*\right|, \tag{6.2.3}$$

where V^*_t is the source/drain to gate flat-band voltage [e.g., approximately zero for n^+ (p^+) polysilicon gate and n^+ (p^+) source/drain].

R_{sh} (in Ω) is the sheet resistance given as:

$$R_{sh} = \frac{\rho_{sh} l}{W}, \tag{6.2.4}$$

where ρ_{sh} is the sheet resistivity (in Ω/sq), and l is the length associated with the region. The two main regions that contribute to sheet resistance are the source/drain extension and the deep junction.

R_{co} (in Ω) is the resistance that arises due to the finite specific contact resistivity ρ_c (in Ω-cm^2) between the silicide and the diffused region (diffusion). If ρ_c is low, current from the diffusion can readily enter the silicide, so the majority of the current flow takes a short path in the diffusion before entering the silicide. On the other hand, if ρ_c is high, it becomes more difficult (resistive) for the current to enter the silicide, and it will travel a longer distance in the diffusion before eventually entering the silicide. It is therefore the competition between ρ_c and the sheet resistivity of the diffusion $\rho^D{}_{sh}$ that determines this distance at which the majority of the current will have entered the silicide. A transmission line model [101-104] predicts this distance l_c (referred to as the contact transfer length) to be:

$$l_c = \sqrt{\frac{\rho_c}{\rho_{sh}^D + \rho_{sh}^S}} \approx \sqrt{\frac{\rho_c}{\rho_{sh}^D}}, \qquad (6.2.5)$$

since $\rho^S{}_{sh}$, the sheet resistivity of the silicide, is much lower than that of the diffusion. R_{co} is expressed as:

$$R_{co} = \frac{\sqrt{\rho_c \rho_{sh}^D}}{W} \coth\left(\frac{l}{l_c}\right). \qquad (6.2.6)$$

R_{co} is therefore highly dependent on the contact length l if $l < l_c$, and is constant if $l >> l_c$. Physically, for $l >> l_c$ most of the current will have already entered the silicide at l_c, and the remaining part of the contact will have little effect on R_{co}, and $R_{co} = (\rho_c \rho^D{}_{sh})^{1/2}/W$. Conversely, if $l << l_c$, R_{co} will be dominated by the resistance at the silicide/diffusion interface, and will be inversely proportional to the contact area (i.e., $R_{co} = \rho_c/Wl$).

The transmission line model discussed above assumes that the silicide is infinitely thin. Extension to the model to take into account finite silicide thickness has been reported [103]. The mathematical expression is complicated, but the physical concepts are simple. There are two consequences to having a thick silicide. First, the area corresponding to the sidewall of the silicide becomes significant (see Fig. 6.2.1), and some current flow lines from the diffusion may terminate on this sidewall (i.e., this sidewall can receive current). Second, as the silicide thickness approaches the diffusion thickness, $\rho^D{}_{sh}$ of the diffusion becomes large, resulting in a larger effective l_c that becomes a function of the silicide and diffusion thicknesses.

6.3 Effects of Source/Drain Junction Configuration

The different components described above all contribute to the total source/drain series resistance R_{SD}. In order to examine the effects on R_{SD} due to different source/drain junction configurations, the junction is parameterized, and the dependence of R_{SD} on these parameters is studied by numerical simulation, as discussed below.

6.3.1 Simulation Structure

The junction, including the extension and the deep junction, is approximated by the structure shown in Fig. 6.3.1.1, which is parameterized into eight parameters. All parameters have the dimension of length (nm), except for D_{EXT} and D_{DEEP}, which refer to the peak doping concentrations of the extension and deep junction, respectively. In order to approximate realistic junctions as much as possible, the junctions here have doping profiles with characteristic lengths in both the x- and y-directions that closely resemble those obtained by inverse modeling of experimental devices (see Chapter 4). Doping profiles corresponding to the cross-sections A-A' and B-B' of Fig. 6.3.1.1 are shown in Fig. 6.3.1.2. The parameters Y_{EXT} and Y_{DEEP} refer to the junction depth on a hypothetical substrate background concentration of 10^{15} cm^{-3}.

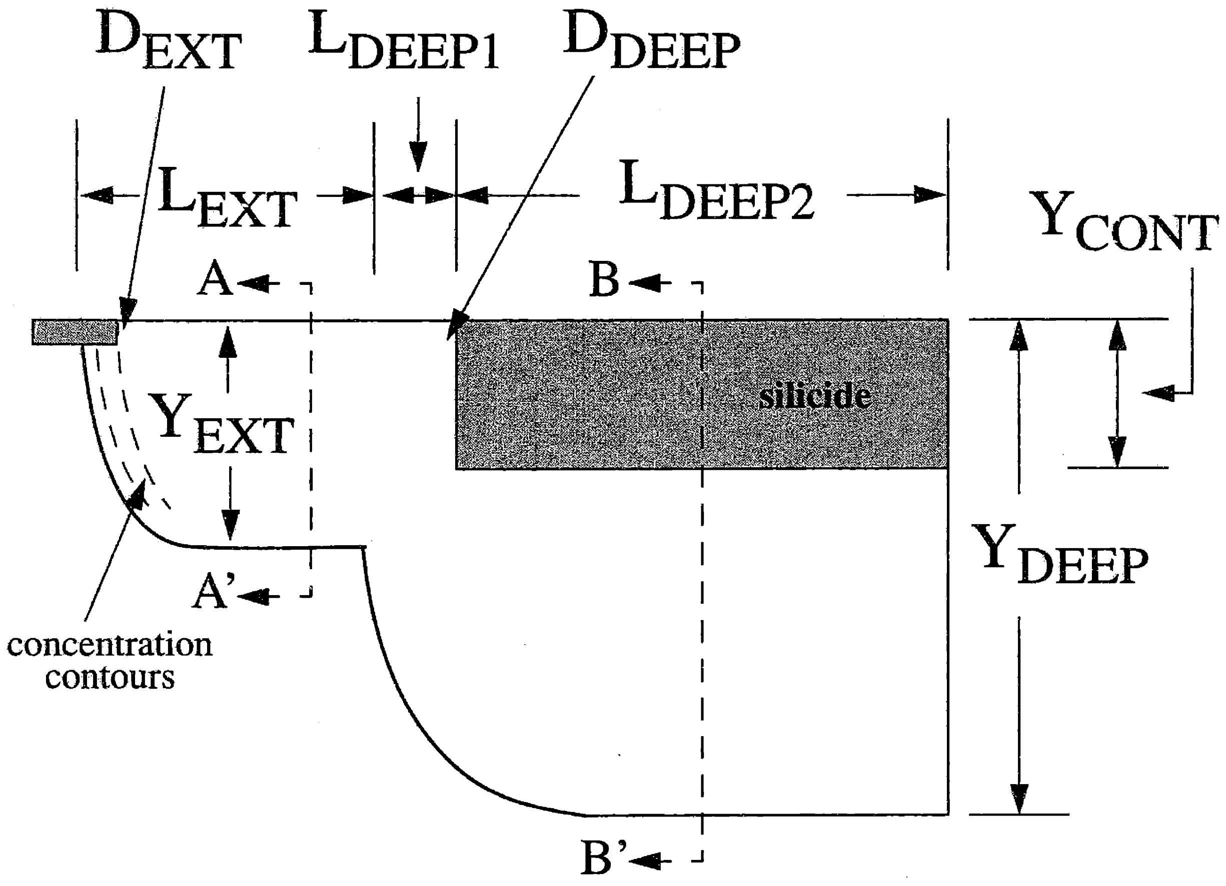

Figure 6.3.1.1: Schematic diagram showing the parameterization of a junction. All parameters have the dimension of length, except for D_{EXT} and D_{DEEP}, which refer to the peak concentrations of the extension and deep junction, respectively. Doping profiles along the cross sections A-A' and B-B' are shown in Figs. 6.3.1.2. The shaded area on the left indicates the ohmic contact that mimics the inversion layer.

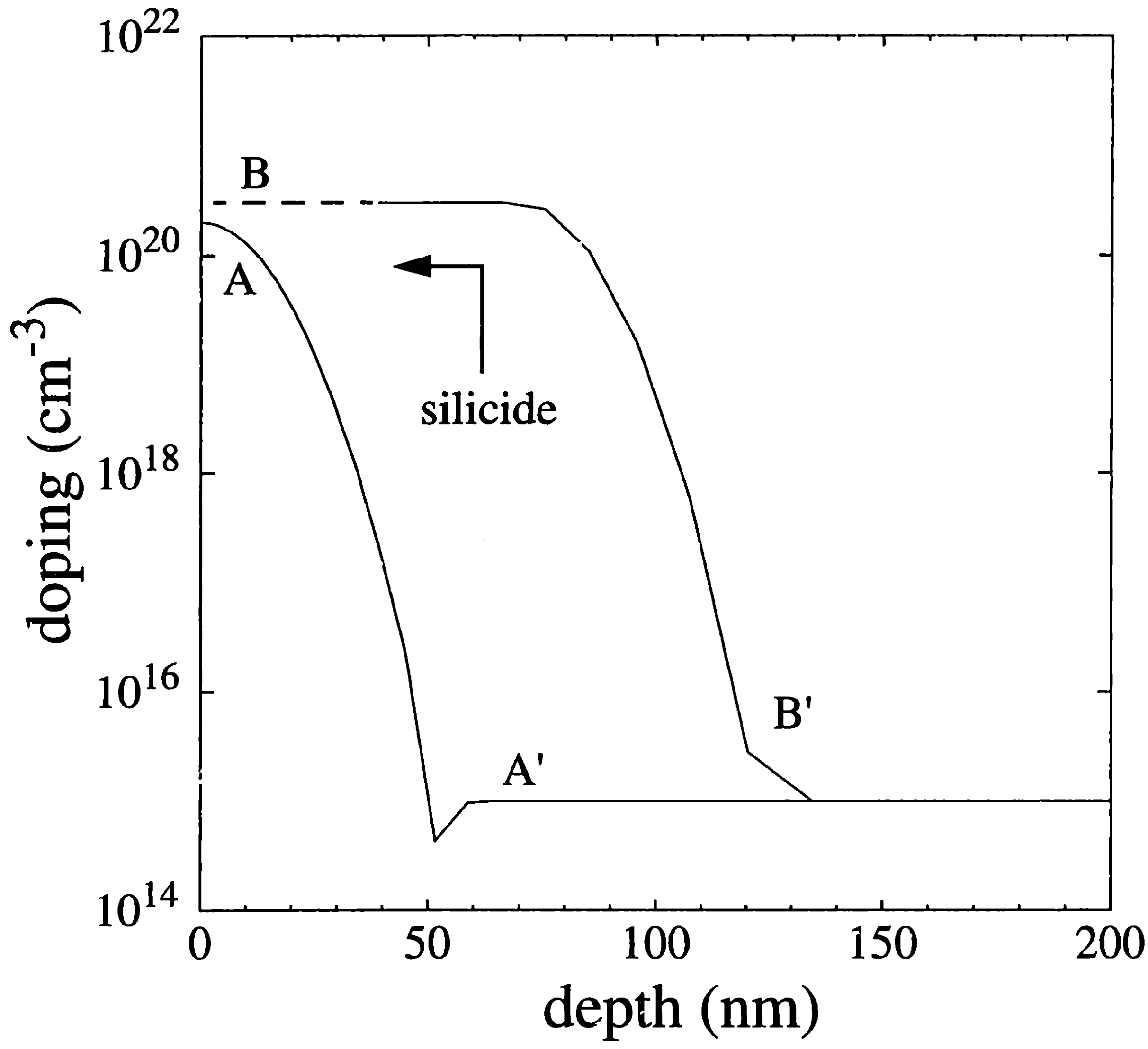

Figure 6.3.1.2: Doping profiles along the cross-sections A-A' and B-B' of Fig. 6.3.1.1.

The center values of the parameters used in the simulation are given in Table 6.3.1.1. Each of these parameters is systematically varied, while keeping the rest fixed, and the resultant structure is simulated by MEDICI. As mentioned earlier, R_S is assumed to be equal to R_D. The total resistance R_{SD} is therefore twice that obtained from the simulation structure. The bias-dependent components are ignored here.

The channel inversion layer is modeled as an ohmic contact having zero contact resistance with a depth of 5 nm. The exact value here is unimportant due to the *ln* dependence [see eqn. (6.2.1)], as easily confirmed by simulations. Since the junctions are graded (non-abrupt), in order to obtain consistent results, the ohmic contact position is chosen such that it is always contacting the peak concentration position of the junction, as shown schematically in Fig. 6.3.1.1. The silicide is modeled as an ohmic contact having various specific contact resistivities. Typical reported values are in the range of 10^{-7} to 10^{-6} Ω-cm^2 [95, 105-106]. Since the sheet resistance of the silicide is typically an order of magnitude lower than that of the diffusion [107], it is ignored. Here, the same calibrated mobility models discussed in Chapter 5 are used. The mobility across the cross-sections A-A' and B-B' of Fig. 6.3.1.1 are shown in Fig. 6.3.1.3. These values are consistent with the widely accepted mobility at the corresponding high doping levels [84]. Note that the mobility models used here are in fact intended (and calibrated) for the channel inversion layer, as discussed in Chapter 5. Near the metallurgical junctions, the electric field in the depletion region creates artifacts in the mobility (see Fig. 6.3.1.3). For instance, the vertical electric field in the depletion region is treated as the vertical field that causes surface acoustic phonon and surface roughness scatterings which do not exist here (see Chapter 5 for details). Nevertheless, due to the exponential nature of the doping profile, the carrier concentrations near the metallurgical junctions are very small, and the artifacts in mobility are insignificant in affecting the results.

Since the simulation structure does not have a gate to induce accumulation near the surface of the junctions, the results cannot account for R_{ac} effects. However, at high doping concentrations the effects of accumulation are very low. Moreover, as discussed in section 6.2, the spreading resistance R_{sp} depends on the carrier concentration in the junction at the point where the current spreads, which means that the position of the ohmic contact that mimics the inversion layer may affect R_{sp}, so that the R_{SD} thus obtained may slightly differ from that of an actual device. Nevertheless, since the focus of this work is to study the dependence of R_{SD} on the junction configuration, these effects are ignored for the moment, and a discussion of the corrections needed is provided in section 6.3.3. As the results show, effects related to the contact resistance dominate the total R_{SD}, making R_{ac} and R_{sp} relatively unimportant.

Table 6.3.1.1: Center values of the parameters of the source/drain structure used in the simulations.

Parameter	Value	Unit	Remark
L_{EXT}	130	nm	extension length
Y_{EXT}	50	nm	extension junction depth
D_{EXT}	2×10^{20}	cm^{-3}	extension peak concentration
L_{DEEP1}	70	nm	silicide/deep junction offset
L_{DEEP2}	1700	nm	silicide contact length
Y_{DEEP}	125	nm	deep junction depth
D_{DEEP}	3×10^{20}	cm^{-3}	deep junction peak concentration
Y_{CONT}	40	nm	silicide thickness

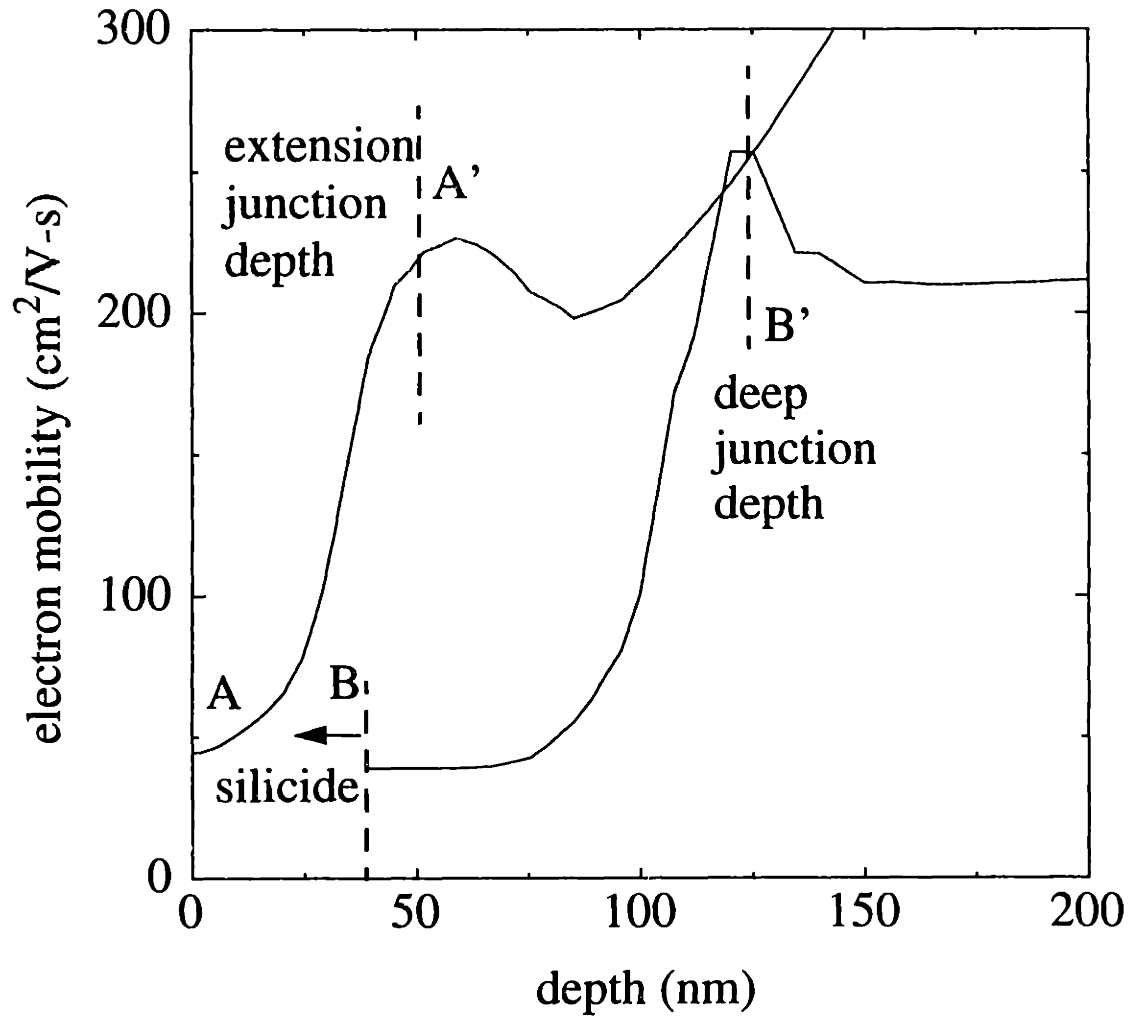

Figure 6.3.1.3: Electron mobility across the cross-sections A-A' and B-B' of Fig. 6.3.1.1.

6.3.2 Simulation Results

Figures 6.3.2.1 - 6.3.2.4, 6.3.2.7, 6.3.2.9, 6.3.2.10, and 6.3.2.12 show the simulated dependence (sensitivity) of R_{SD} on each parameter. Since a very wide range of junction configuration is covered, these figures can be used as an initial estimate of R_{SD} for any modern junction configuration, a valuable tool for design space exploration.

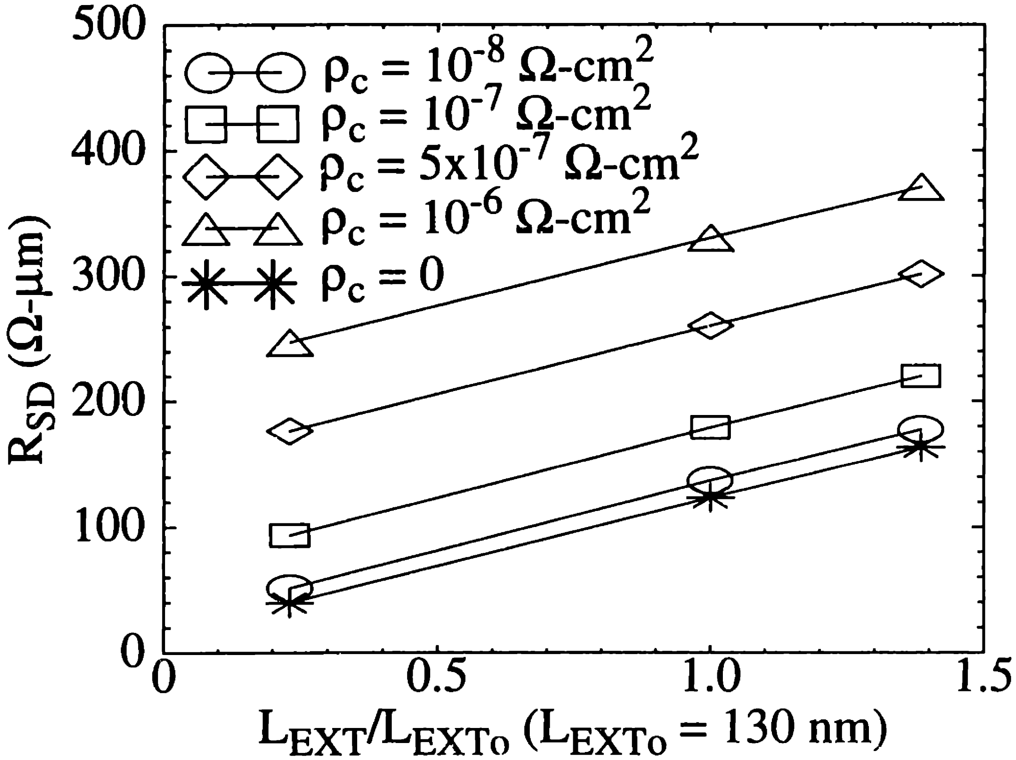

Figure 6.3.2.1: R_{SD} dependence on normalized extension length L_{EXT}/L_{EXTo}.

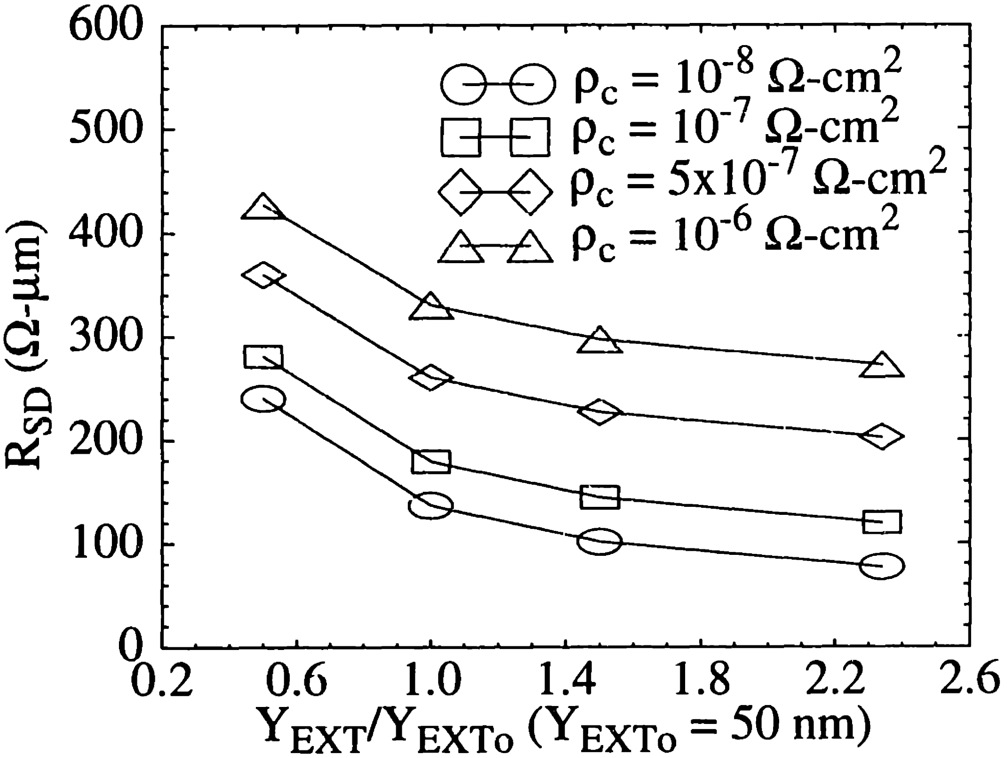

Figure 6.3.2.2: R_{SD} dependence on normalized extension junction depth Y_{EXT}/Y_{EXTo}.

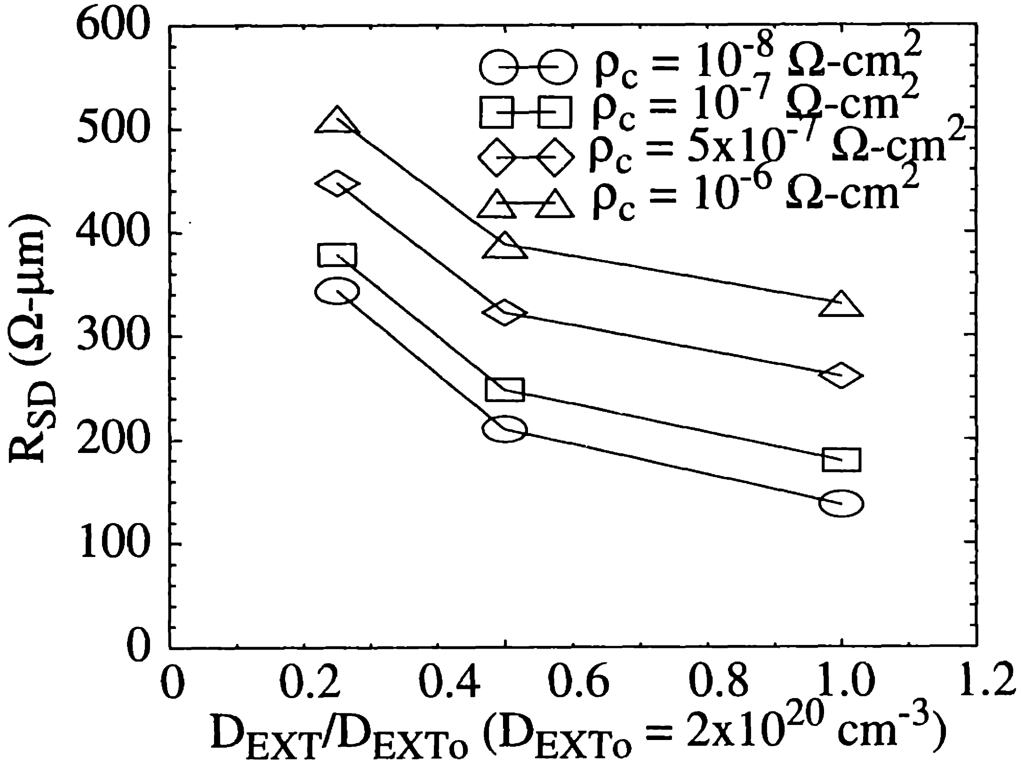

Figure 6.3.2.3: R_{SD} dependence on normalized extension peak doping concentration D_{EXT}/D_{EXTo}.

Several observations can be made of the figures. Fig. 6.3.2.1 shows the R_{SD} dependence on L_{EXT}. The linearity of the dependence indicates that the resistance is mainly due to the sheet resistance of the extension. Also, the fact that the curve corresponding to $\rho_c = 0$ does not extrapolate to zero indicates that there is a component of resistance that is due neither to the sheet resistance nor to the contact resistance. This resistance is attributed to the spreading resistance R_{sp}. Fig. 6.3.2.2 shows the R_{SD} dependence on Y_{EXT}. The approximate $1/Y_{EXT}$ dependence qualitatively suggests that the total resistance of the extension region is dominated by the sheet resistance, similar to the previous case. This is because $R_{sh} \propto \rho_{ext}/x_j$, where ρ_{ext} is the bulk resistivity of the source/drain extension, and x_j is the junction depth ($Y_{EXT} = x_j$). The R_{SD} dependence on D_{EXT} also exhibits a similar behavior, since $\rho_{ext} \propto 1/N_D$, where N_D is the doping concentration, as shown in Fig. 6.3.2.3.

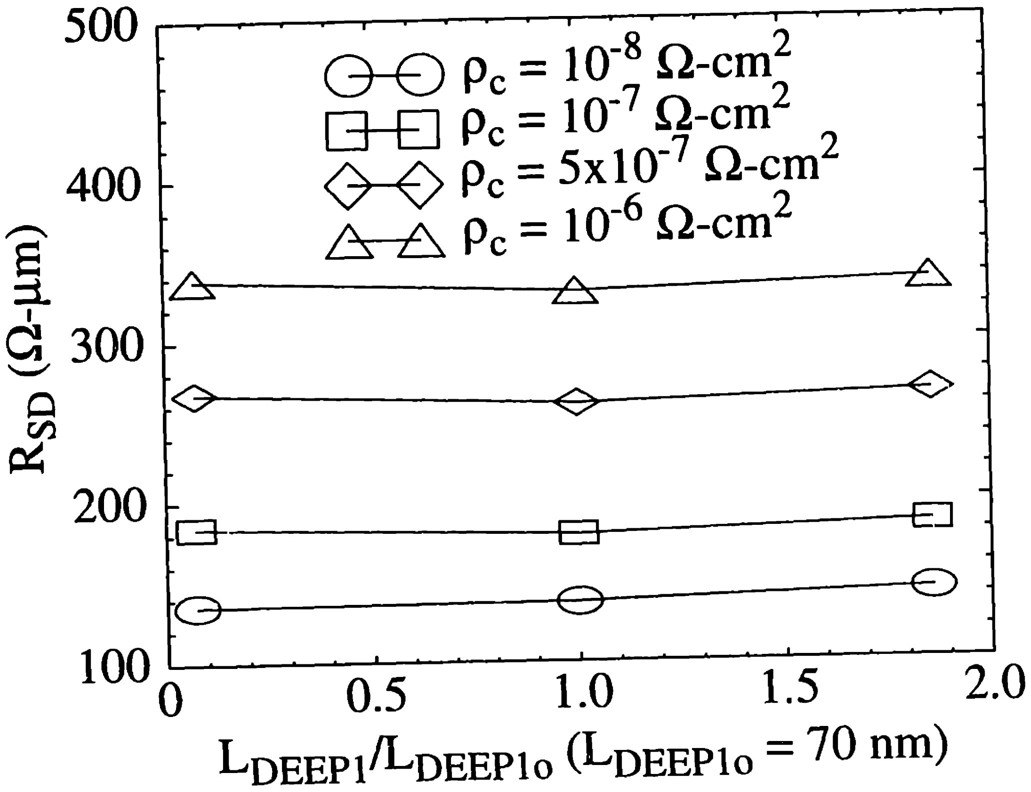

Figure 6.3.2.4: R_{SD} dependence on normalized silicide/deep junction offset L_{DEEP1}/L_{DEEP1o}.

The R_{SD} dependence on L_{DEEP1} is shown in Fig. 6.3.2.4. Obviously, there is little dependence. This may seem to be a surprising result, especially considering that there is a relatively strong dependence on L_{EXT}. There is in fact nothing surprising once we consider the fact that current within the region defined by L_{DEEP1} actually has a large component that flows in the vertical direction, shown schematically in Fig. 6.3.2.5. Within a reasonable range of L_{DEEP1}, an increase in path length traveled by the current is offset by a reduced current crowding. Consequently the R_{SD} dependence on L_{DEEP1} is small. Simulated current flow lines for L_{DEEP1} = 5 nm and L_{DEEP1} = 130 nm are shown in Figs. 6.3.2.6(a) and (b), respectively. Note that most of the current flow lines terminate underneath the silicide, as opposed to the sidewall, because the area of the sidewall is negligible compared to the area underneath (note also that the vertical and horizontal scales are different).

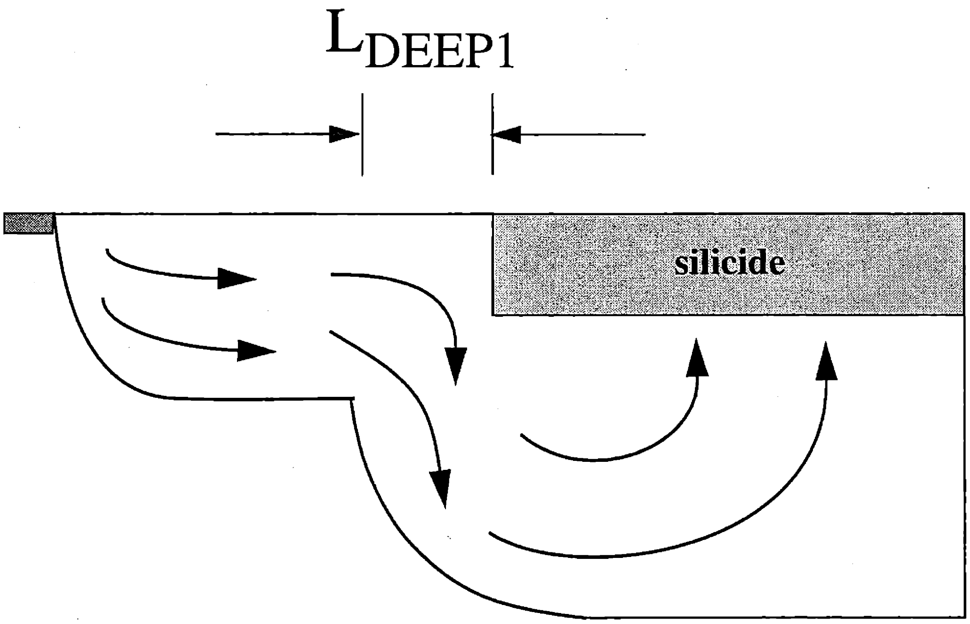

Figure 6.3.2.5: Schematic diagram showing the current flow lines in a junction. Within the region defined by L_{DEEP1}, the current has a large component in the vertical direction.

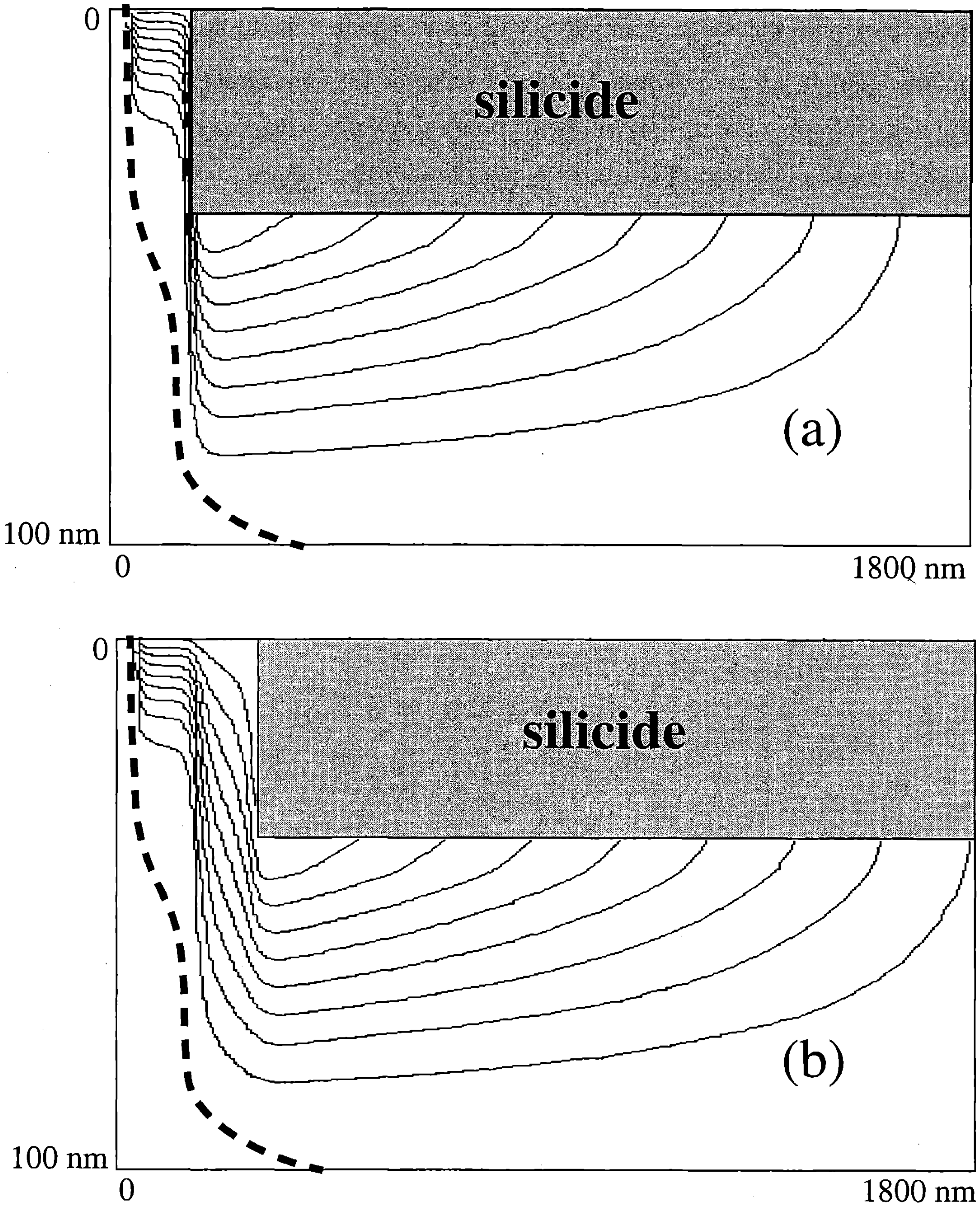

Figure 6.3.2.6: Simulated current flow lines for (a) L_{DEEP1} = 5 nm, and (b) L_{DEEP1} = 130 nm. The dashed lines denote the approximate positions of the junctions. Note that the vertical and horizontal scales are different.

The R_{SD} dependence on L_{DEEP2} is shown in Fig. 6.3.2.7. It can be seen that the specific contact resistivity ρ_c plays a very important role in determining the L_{DEEP2} dependence. As discussed in Section 6.2, there is a competition between current flowing in the diffusion and current flowing in the silicide. The current will flow in such a way that a balance is established so that the total resistance is minimized. For example, if ρ_c is very large, current will flow in the diffusion for a longer distance before entering the silicide because the resistance to enter the silicide is large. On the other hand, if ρ_c is very small, the majority of the current will have entered the silicide at a short distance because the diffusion is more resistive for current flow than the silicide. It is this balance between ρ_c and $\rho^D{}_{sh}$ (sheet resistance of diffusion) that determines the length l_c at which most of the current will have entered the silicide. For L_{DEEP2} much longer than l_c, there is

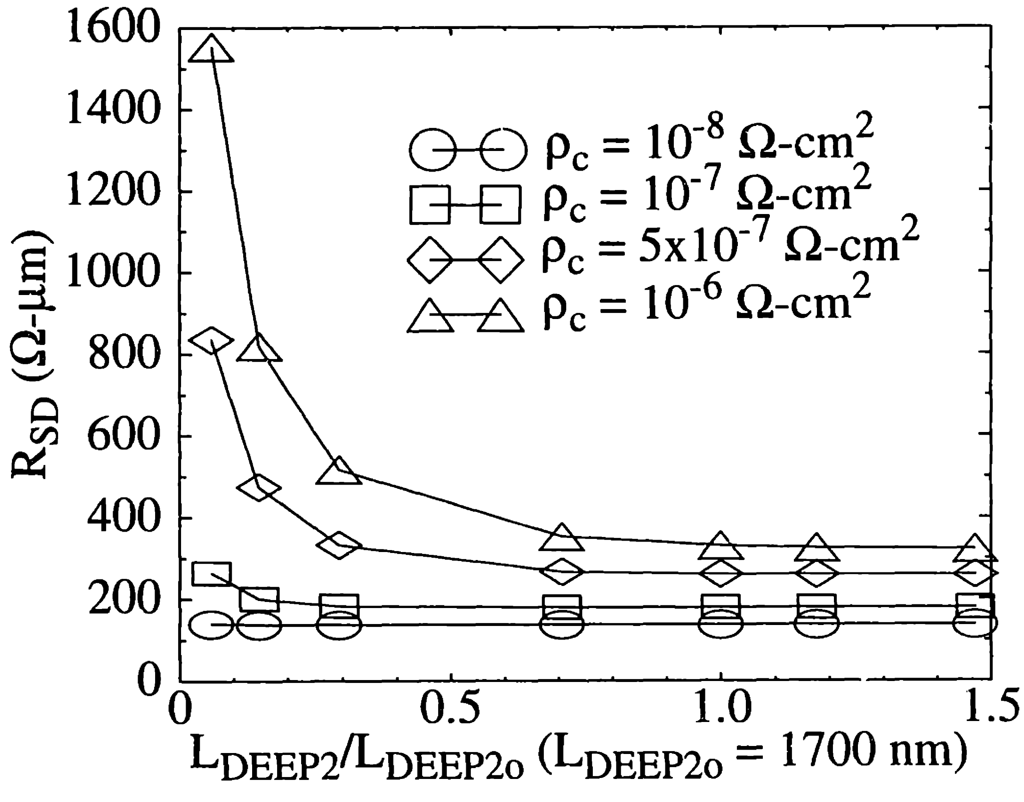

Figure 6.3.2.7: R_{SD} dependence on normalized silicide contact length L_{DEEP2}/L_{DEEP2o}.

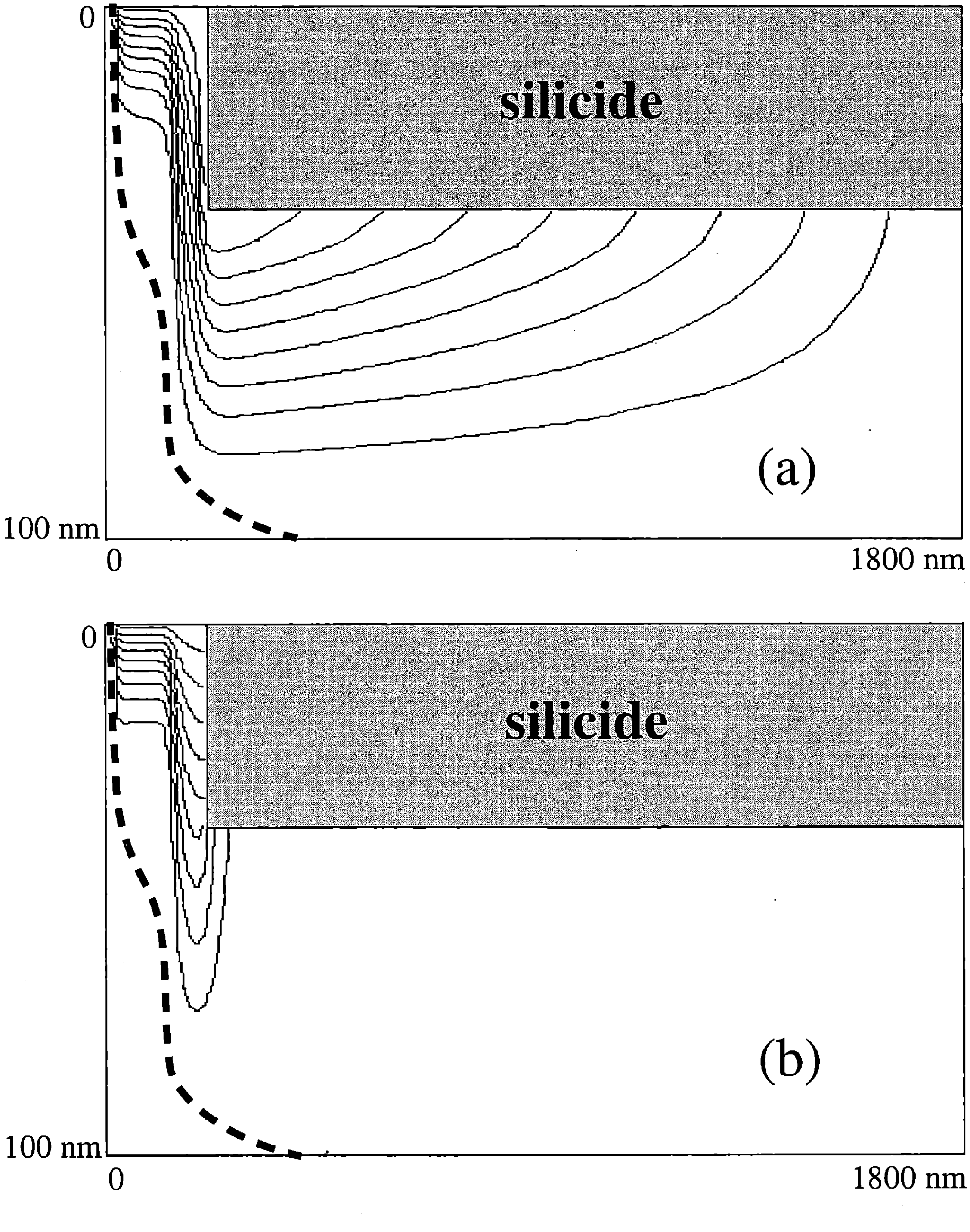

Figure 6.3.2.8: Simulated current flow lines for (a) high ρ_c ,and (b) low ρ_c. It can be seen that the current flow lines travel longer distances in the case of high ρ_c. Also, in the case of low ρ_c, more current flow lines terminate on the sidewall of the silicide. The approximate positions of the junctions are denoted by dashed lines.

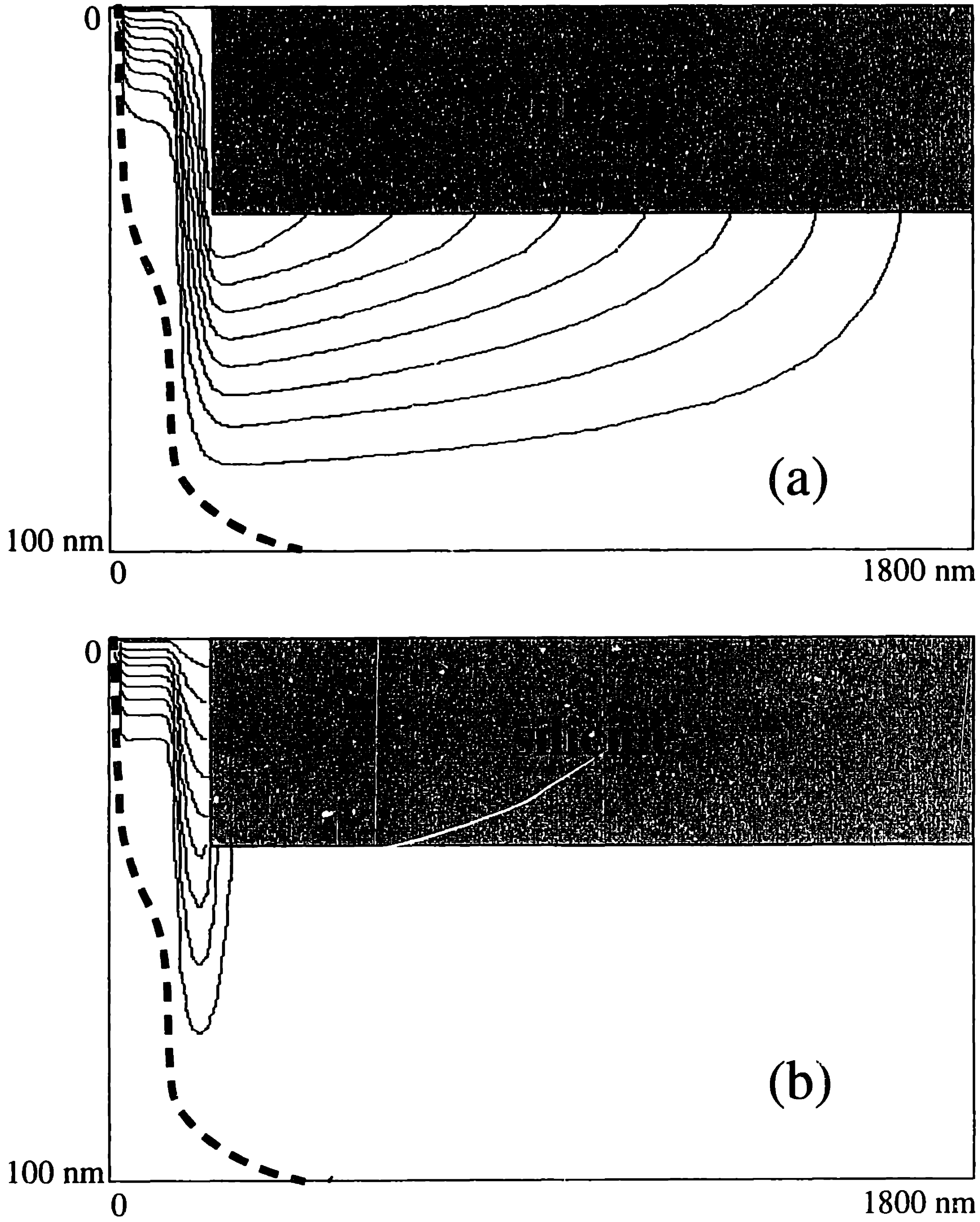

Figure 6.3.2.8: Simulated current flow lines for (a) high ρ_c ,and (b) low ρ_c. It can be seen that the current flow lines travel longer distances in the case of high ρ_c. Also, in the case of low ρ_c, more current flow lines terminate on the sidewall of the silicide. The approximate positions of the junctions are denoted by dashed lines.

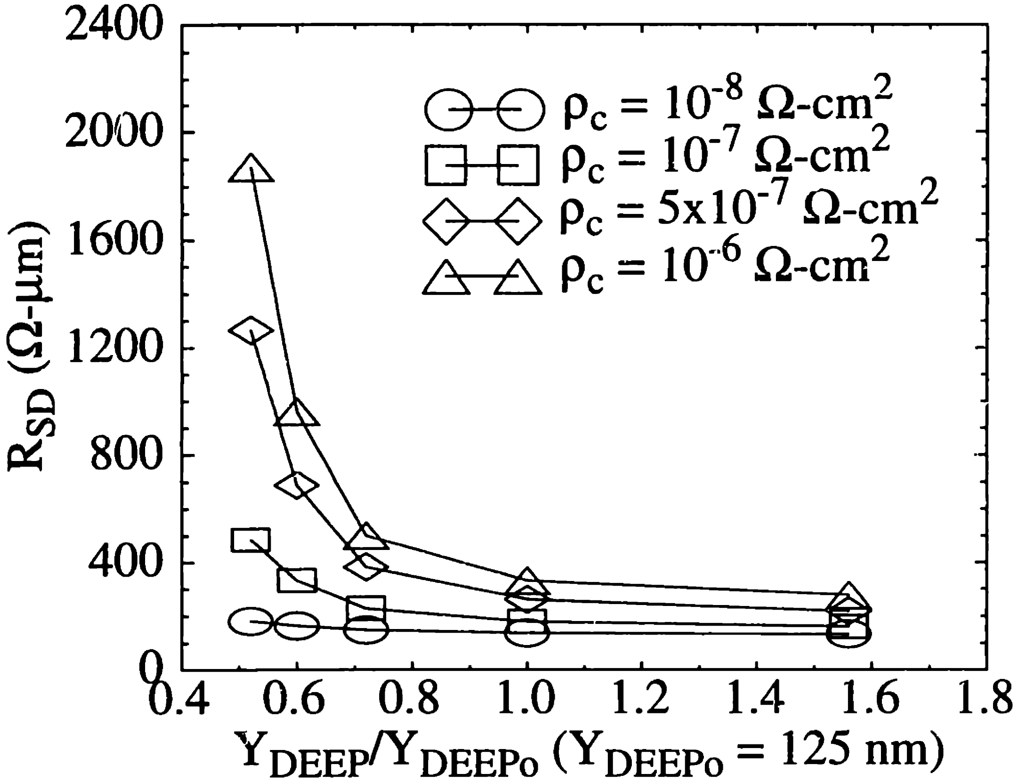

Figure 6.3.2.9: R_{SD} dependence on normalized deep junction depth Y_{DEEP}/Y_{DEEPo}.

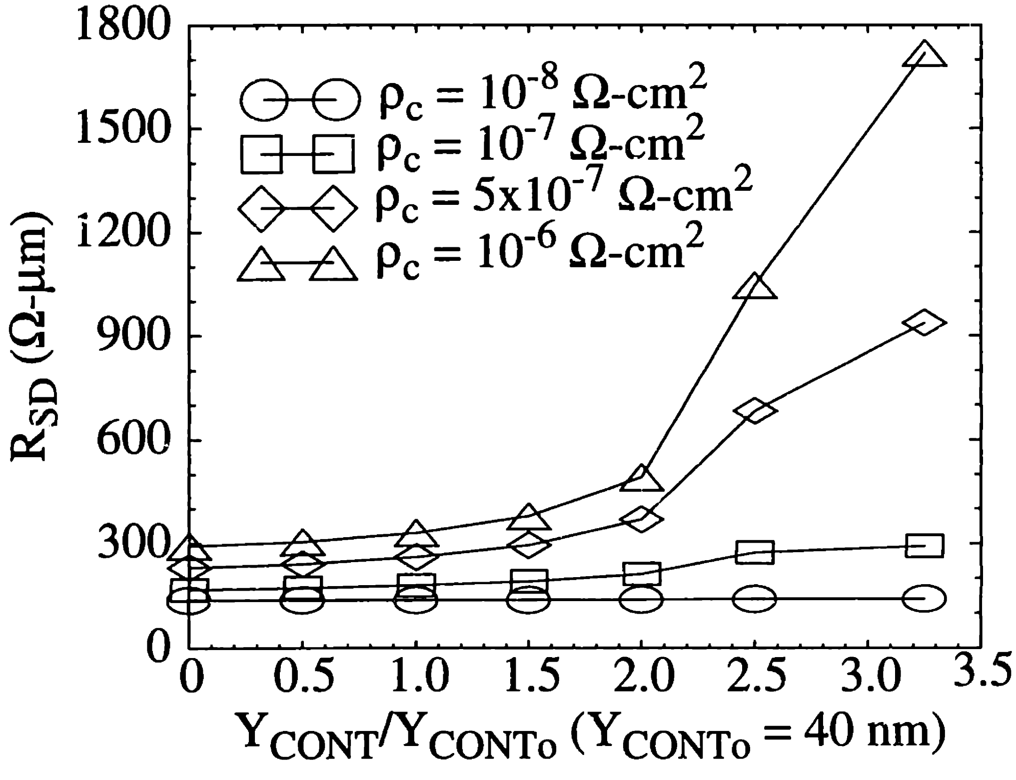

Figure 6.3.2.10: R_{SD} dependence on normalized silicide thickness Y_{CONT}/Y_{CONTo}.

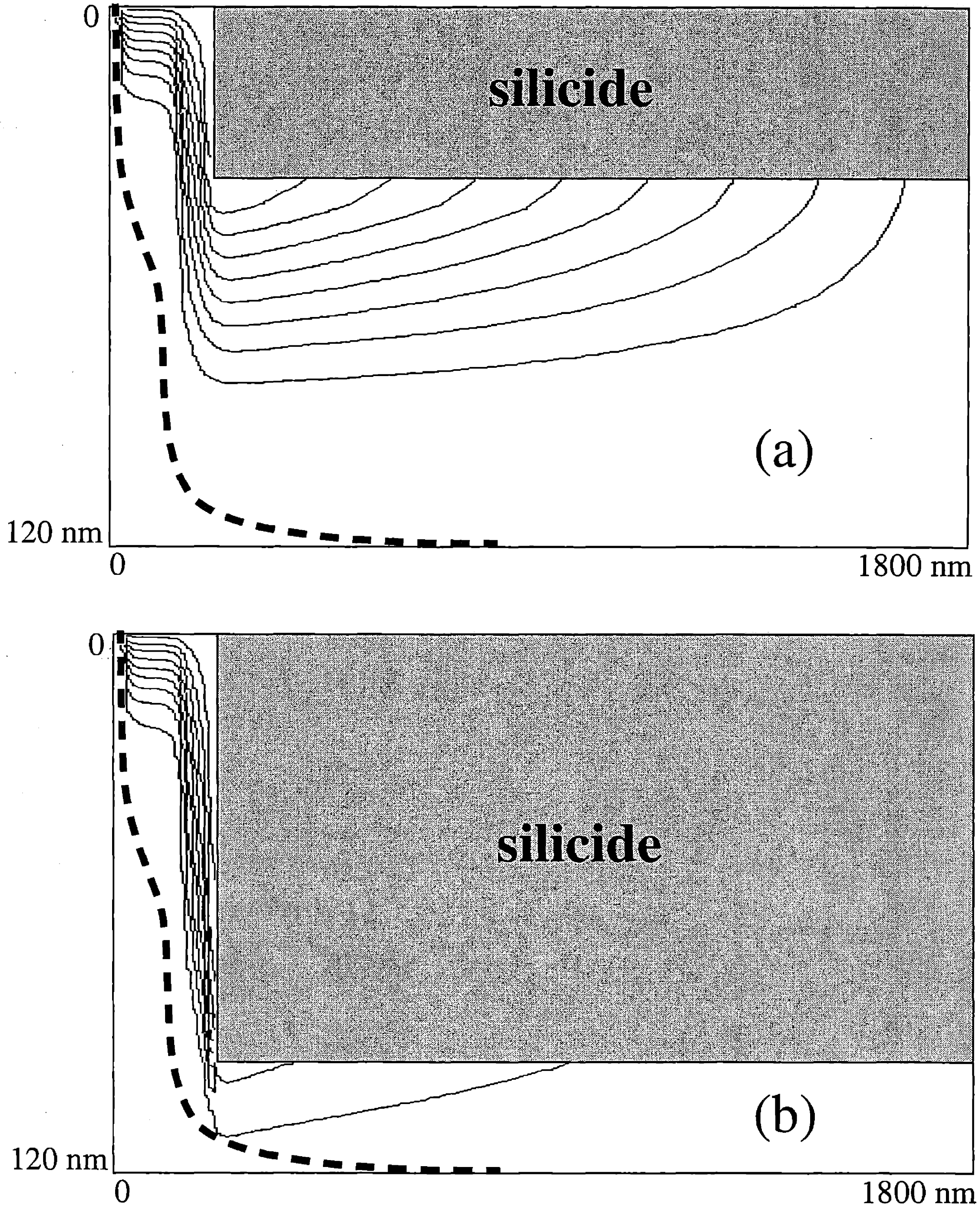

Figure 6.3.2.11: Simulated current flow lines for (a) $Y_{CONT} = 40$ nm, and (b) $Y_{CONT} = 100$ nm (~ Y_{DEEP}). The large silicide thickness makes the sheet resistance of the diffusion very high, forcing the current to enter the silicide at a short distance. The approximate positions of the junctions are denoted by dashed lines.

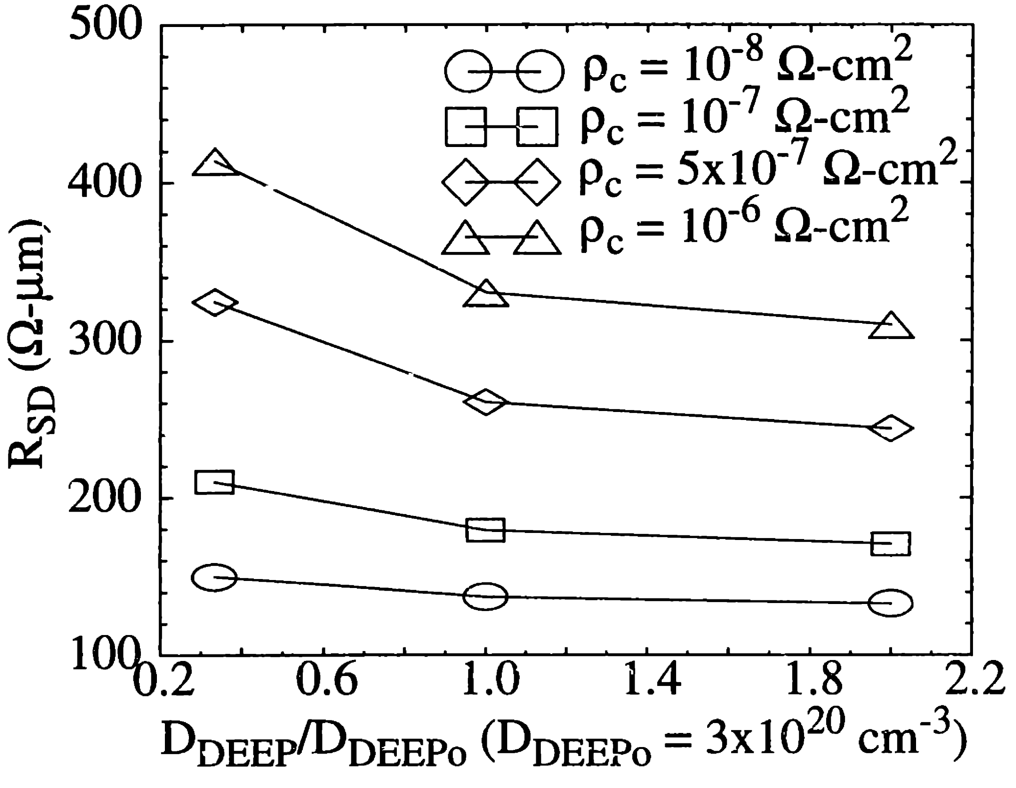

Figure 6.3.2.12: R_{SD} dependence on normalized deep junction peak doping concentration D_{DEEP}/D_{DEEPo}.

6.3.3 Effects of Accumulation Resistance R_{ac} and Spreading Resistance R_{sp}

The discussion up to this point has ignored the effects of accumulation in the source/drain junctions under the gate (i.e., accumulation resistance R_{ac}). Also, an ohmic contact has been used to mimic the inversion layer. Consequently the spreading resistance R_{sp} may not have been accounted for properly. In this section, these effects are quantified using simulations of a MOSFET structure that contains all the details of the source/drain junctions. Using simulated I-V data of 1 μm and 5 μm channel-length devices, R_{SD} is extracted using the "shift and ratio" method (see Appendix C). A correction to the R_{SD} values of Section 6.3.2, in order to account for these effects, is provided.

A schematic diagram of the simulation structure is shown in Fig. 6.3.3.1. Here, the source/drain junctions are identical to those discussed previously, with a channel region, a gate, and an oxide dielectric added (see Table 6.3.1.1 for the values of the junction parameters). In order to study the effect of accumulation, an overlap parameter δ (relative to the metallurgical junction as shown in the figure) is defined and its value is systematically varied.

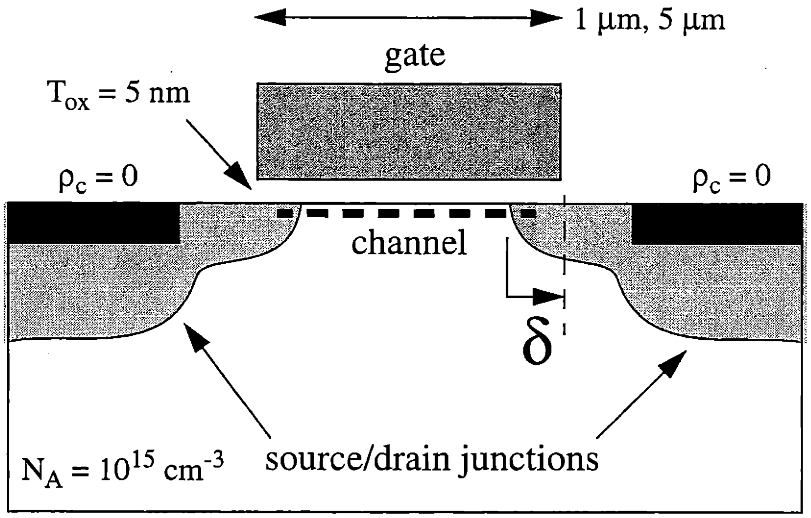

Figure 6.3.3.1: Schematic diagram of the simulation structure used to study accumulation resistance and spreading resistance effects.

The R_{SD} values extracted from simulated data, for various D_{EXT}, are shown in Table 6.3.3.1, with δ a variable. It can be seen that for $D_{EXT} = 2\text{x}10^{20}$ cm^{-3}, at $\delta > 30$ nm, R_{SD} does not depend on δ. On the other hand, for $\delta < 30$ nm, R_{SD} increases because of the high resistance in the regions of the channel that are not covered by the gate. The same trend can be observed for lower D_{EXT}, except that a larger δ is needed to achieve the lowest possible R_{SD}. Note that the channel does not end at the metallurgical junctions, since the source/drain junctions are not abrupt. In fact, parts of the source/drain junctions form parts of the channel, so that the channel-length is slightly longer than the distance between the metallurgical junctions (see the schematic of Fig. 6.3.3.1). Therefore, the regions of the source/drain junctions that are parts of the channel and are covered by the gate are in accumulation under normal operation. Since R_{sp} depends on the local carrier concentration where the current from the channel spreads into the source/drain junctions, the amount of overlap between the gate and the source/drain junctions can strongly affect R_{sp} [110]. For example, if δ is very small, inverted carriers in the channel will have to enter the low concentration region (i.e., tail of the doping profile) of the source/drain junctions, where the local ρ_{eff} is large [eqn. (6.2.1)], resulting in large R_{sp}. On the other hand, if δ is large enough so that the low doping concentration region is covered by the gate and is accumulated, then the current from the channel will spread into this region of low ρ_{eff}, and the R_{sp} will be lower. However, covering the high doping concentration region of the source/drain junctions by the gate will not reduce R_{sp} because the carrier concentration there is high enough so that accumulation effects are insignificant. This is why the R_{SD} values of Table 6.3.3.1 are independent of δ for large δ. Note that for $D_{EXT} = 2\text{x}10^{20}$ cm^{-3}, at $\delta = 35$ nm (i.e., 70 % of the 50 nm junction depth), accumulation effects can no longer improve the R_{SD} values. For $D_{EXT} = 5\text{x}10^{19}$ cm^{-3}, δ has to be greater than 50 nm before its value can no longer affect R_{SD}.

For the simulations in Section 6.3.2, the ohmic contact that mimics the inversion layer is positioned so that it is in direct contact with the position of peak doping concentration of the source/drain junctions. In a MOSFET, however, the current in the inversion layer usually spreads in a region having a different carrier concentration than the peak doping, depending on the inversion layer concentration, accumulated carrier concentration in the source/drain junctions, as well as the junction doping concentration). As a result, the R_{SD} values shown in Section 6.3.2 will need to be refined in order to accomodate these effects. From Table 6.3.3.1, the R_{SD} value for $D_{EXT} = 2x10^{20}$ cm^{-3}, and at large δ, is 158 Ω–µm, while the corresponding value obtained using an ohmic contact is 123 Ω–µm. This 35 Ω–µm additional resistance accounts for the difference in the assumed injection point and local carrier concentration where the current spreads when an ohmic contact is used. Therefore, in order to use the results of the previous section, and have R_{SD} values that more closely match those extracted using standard techniques, it is recommended that 35 Ω–µm of resistance be added to the R_{SD} values for $D_{EXT} = 2x10^{20}$ cm^{-3}. At $D_{EXT} = 5x10^{19}$ cm^{-3}, however, the R_{SD} value obtained using an ohmic contact is actually higher than that obtained using a full MOSFET structure, since accumulation effects can be significant for such low source/drain doping concentrations. The adjustments ΔR_{SD} needed for the R_{SD} values of Section 6.3.2 (provided δ is large), in order to account for accumulation/spreading resistance effects, are given in Table 6.3.3.1.

Table 6.3.3.1: R_{SD} values extracted from simulated data of the structure of Fig. 6.3.1.1. ΔR_{SD} denotes the values that need to be added to the R_{SD} values of Section 6.3.2 in order to account for accumulation/spreading resistance effects (provided δ is large).

	R_{SD} (Ω–µm)		
δ (nm)	D_{EXT} = $2x10^{20}$ cm^{-3}	D_{EXT} = 10^{20} cm^{-3}	D_{EXT} = $5x10^{19}$ cm^{-3}
10	399	518	700
20	179	242	375
30	159	198	300
40	158	195	290
50	158	195	286
60	158	194	286
70	158	194	285
ΔR_{SD}	+35	+1	-38

6.4 Identification of Most Important Contributions to Source/Drain Series Resistance R_{SD}

As discussed in Chapter 5, an accurate simulation of I-V characteristics at high current levels requires an accurately calibrated mobility model, and, especially at low V_{DS}, a realistic value of R_{ext} (or R_{SD}). Although R_{ext} may be extracted from device measurements, a reasonably accurate and reliable way to evaluate its value from simulations is also very important. Being able to accurately simulate R_{SD} for one generation of devices, it may also be possible to predict its behavior for the next generations. In practice, however, without extensive characterization, it may be difficult to accurately know every detail of a junction that is needed for simulation (e.g., spacer length, silicide thickness, deep junction doping profile ... etc.). Consequently, this section is focused on the identification of the most important contributions to R_{SD}.

As shown in section 6.3, all junction parameters contribute to R_{SD}, however, not all of them are significant. For example, with an allowance of 30% error in the value of a parameter, the parameters L_{DEEP1}, L_{DEEP2}, D_{DEEP}, and Y_{CONT} are clearly less significant in affecting R_{SD} compared to L_{EXT}, Y_{EXT}, D_{EXT}, and Y_{DEEP}. The most important parameter, nevertheless, is the specific contact resistivity ρ_c. Although its value can be measured using Kelvin resistor test structures [108-109], for the discussion here, an exact value is not sought. Instead, the widely accepted range of 10^{-7} to 10^{-6} Ω-cm^2 [95, 105-106] is used here. The same data of section 6.3 (with appropriate adjustments discussed in Section 6.3.3) are replotted in Figs. 6.4.1(a) and (b), which show, respectively, the simulated R_{SD} values versus normalized L_{EXT}, Y_{EXT}, D_{EXT}, and Y_{DEEP} for $\rho_c = 10^{-7}$ and 10^{-6} Ω-cm^2. It can be seen that the value of ρ_c has a very significant effect on R_{SD}. It also significantly affects the Y_{DEEP} dependence, since it is closely linked to the contact resistance. The dependence on other parameters, however, is not strongly affected by ρ_c. Moreover, except for Y_{DEEP}, all of the parameters that have large effects on R_{SD} (i.e., L_{EXT}, Y_{EXT}, and D_{EXT}) are related to the extension, and each of them is approximately equally important. Therefore, R_{SD} is most sensitive to ρ_c as well as the geometry and doping of the extension.

As an illustration, a simple methodology on the use of the results of this section is discussed. In Chapter 5, the external resistance R_{ext} that must be added to the simulation structure in order to match the low V_{DS} data of the STEP device is found to be 190 Ω-μm. The extension used in the simulation has a peak doping concentration of $2\text{x}10^{20}$ cm^{-3} and a length of 150 nm (no deep junction used, and $\rho_c = 0$). Using the $\rho_c = 0$ curve of Fig. 6.3.2.1, the extension of the simulation structure is found to have an R_{SD} of 140 + 35 Ω-μm = 175 Ω-μm (the 35 Ω-μm is the correction discussed in section 6.3.3). This is justified even though the simulation structure of Chapter 5 does not have the deep junction because when $\rho_c = 0$, the deep junction has very little effect on R_{SD}, which can be verified easily. Assuming an external parasitic resistance of 5 Ω (due to pads, probes, and silicide/metal contacts ... etc.), which amounts to 50 Ω-μm for the 10 μm width device, the total R_{SD} of the device is 175 + 190 - 50 Ω-μm = 315 Ω-μm, which is consistent with the experimental value [23]. This value is quite close to the center value of Fig. 6.4.1(b). If ρ_c is closer to 10^{-7} Ω-cm^2, from Fig. 6.4.1(a), it is possible that Y_{EXT}, D_{EXT}, or Y_{DEEP}, or a combination of them, are over-estimated (actual value smaller than the center value of Table 6.3.1.1), or L_{EXT} is under-estimated (actual value larger than the center value). Since Y_{EXT} is obtained from inverse modeling (Chapter 4), it is unlikely to be the cause. It is then D_{EXT} and/or Y_{DEEP} that are most likely over-estimated, or L_{EXT} under-estimated. Similarly, if ρ_c is closer to 10^{-6} Ω-cm^2, from Fig. 6.4.1(b), it is possible that L_{EXT} is over-estimated, or Y_{DEEP} is under-estimated, or a combination of both. In either case, it is interesting to see that, because of finite ρ_c, the deep junction depth Y_{DEEP} in fact strongly affects R_{SD}.

6.5 Conclusion

The total external parasitic series resistance R_{ext} discussed in Chapter 5 is explored in detail, with particular attention paid to the source/drain series resistance R_{SD}, which is the largest contribution to R_{ext}. In order to study the effect on R_{SD} due to different junction configurations, the junction is parameterized, and simulations performed to explore the dependence and sensitivity on these parameters. It is found that for a typical junction, as in the devices discussed in this thesis, the geometry and doping of the extension strongly affect R_{SD}. The deep junction (particularly the depth) also has a significant effect on R_{SD}.

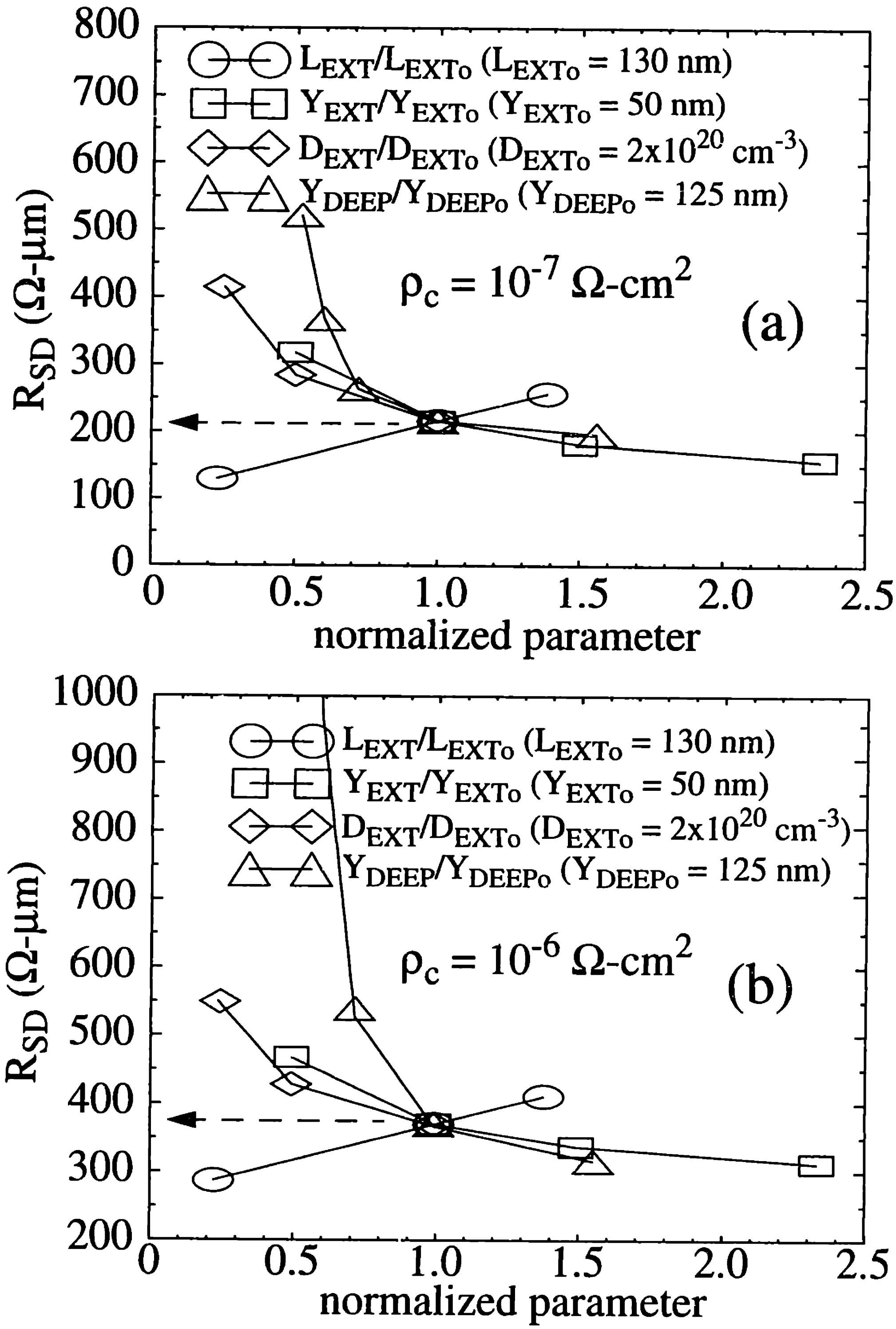

Figure 6.4.1: Simulated R_{SD} versus various normalized parameters of section 6.3 showing the relative sensitivity of L_{EXT}, Y_{EXT}, D_{EXT}, and Y_{DEEP}. Within the range of interest, the parameters L_{DEEP1}, L_{DEEP2}, D_{DEEP}, and Y_{CONT} are not very important in affecting the R_{SD} value. Note that the values shown here have already been corrected for the 35 Ω-μm difference discussed in section 6.3.3.

Chapter 7

Conclusion

This thesis started out with the discovery that useful information pertaining to the 2D distribution of dopants in sub-100 nm-channel-length MOSFETs can be obtained using I-V characteristics in the subthreshold region, a region of MOSFET operation that is often overlooked. Consequently, much of the work that followed was focused on developing a better understanding of device operation in this region, and more importantly, on developing a technique to harness the information that can be obtained, leading to useful applications including 2D doping profile characterization and calibration of transport models. A summary of the key aspects of the work is given below, followed by some recommendations for future research.

7.1 Summary

In this thesis a new inverse-modeling-based 2D doping profile characterization technique using subthreshold I-V characteristics for sub-100 nm MOSFETs is introduced. For a MOSFET in weak-inversion, the drain current is very sensitive to the 2D electrostatic potential distribution in the depletion region of the device channel, which in turn depends on the applied potential at the source, drain, gate, and substrate in ways that are strongly determined by the 2D dopant distribution. In subthreshold I-V inverse modeling, 2D doping profiles are obtained by matching a large set of simulated subthreshold I-V characteristics to the corresponding experimental data by optimizing for the doping profiles. Physically, the technique combines the matching of incremental changes in current and various biases (that allow the doping concentration near the depletion edge to be determined) and the absolute values of the current and biases (that allow the doping concentration within the depletion region to be determined). A custom non-linear least square optimizer implementing the Levenberg-Marquardt algorithm has been developed for the purpose of this work.

There are several key advantages and disadvantages of this technique. Advantages include: (1) in the subthreshold region, current changes exponentially with biases, and is very sensitive to the device electrostatics (i.e., doping concentration), so that uncertainties in transport properties such as mobility have little effect on the results; (2) since only subthreshold I-V characteristics are needed, the technique does not require the use of any special test structures; (3) the technique can be applied to devices having very short channel-lengths (sub-100 nm); (4) the technique is immune to parasitic capacitances, resistances, and noise; (5) results are insensitive to gate electrode area variations; (6) the channel-length is automatically obtained as part of the 2D doping profile; (7) the technique is non-destructive; and (8) subthreshold I-V characteristics are easily obtainable. For the disadvantages, the technique is sensitive to the dielectric (oxide) thickness, can only extract 2D doping profiles of short-channel devices, as well as the tremendous amounts of CPU time needed for the simulations. For devices having very thin dielectrics, proper care must be taken in order to obtain accurate results. For instance, quantum mechanical and polysilicon gate depletion effects in thin dielectrics must be modeled or accounted for properly. Moreover, more work is needed in the future to verify the accuracy of the 2D doping profiles obtained from inverse modeling, since a mature technique for directly measuring 2D doping profiles is lacking in the present moment.

A second goal of the thesis is to demonstrate that knowledge of the 2D doping profiles can be used to calibrate transport (mobility) models, leading to highly accurate predictive capabilities useful for device design. An inverse-modeling-based transport model calibration methodology is discussed. Lastly, parasitic effects such as source/drain series resistance, that are important for device design, and may have an impact on the accuracy in simulating high-current I-V characteristics, are discussed and analyzed in detail.

7.2 Recommendations for Future Work

The subthreshold I-V inverse modeling technique has been developed as a stand-alone technique for 2D doping profile characterization. It is useful, however, to combine this technique with the 2D C-V technique in order to obtain the best sensitivity in different regions of the device. For instance, the 2D C-V technique has good sensitivity in the source/drain regions, while the subthreshold I-V technique has good sensitivity in the channel region. Matching subthreshold I-V data and C-V data simultaneously may improve the accuracy of the results, and should be explored in the future.

Due to the large amounts of data to match, the CPU time needed for the technique is quite substantial. Consequently it is desirable to improve the optimization algorithm as well as to speed up the 2D device simulations. For the work in this thesis, the only device simulator used was MEDICI. A survey and evaluation of different available device simulators will be very useful.

For the work in this thesis, quantum mechanical and polysilicon gate depletion effects have been crudely approximated using an equivalent oxide thickness. More sophisticated models to account for these effects should be used in the future.

The 2D doping profiles in this thesis have been represented using a sum of 1D B-spline and 2D Gaussian functions. Although these functions were found to provide very good results, other functional forms may be used as well, and this should be explored.

Once direct techniques for measuring 2D doping profiles are mature and become available, the 2D doping profiles obtained from inverse modeling should be checked against those obtained using these techniques.

The calibration of transport models has been performed on a family of very similar devices. Experimental data on more devices are needed in order to test the applicability of the calibrated transport models. Process model calibration using the 2D doping profiles obtained from inverse modeling should also be explored.

References

[1] D.A. Antoniadis and I. Moskowitz, "Diffusion of indium in silicon inert and oxidizing ambients," *J. Appl. Phys.*, 53(12), pp. 9214-9216, 1982.

[2] D.A. Antoniadis and I. Moskowitz, "Diffusion of substitutional impurities in silicon at short oxidation times: An insight into point defect kinetics," *J. Appl. Phys.*, 53(10), pp. 6788-6796, 1982.

[3] C.S. Fuller and J.A. Ditzenberger, "Diffusion of Donor and Acceptor Elements in Silicon," *J. Appl. Phys.*, 27(5), pp. 544-553, 1956.

[4] I.C. Kizilyalli, "Diffusion parameters of indium for silicon process modeling," *J. Appl. Phys.*, 80(9), pp. 4944-4947, 1996.

[5] M.J. van Dort et al., "A high-resolution study of two-dimensional oxidation-enhanced diffusion in silicon," *IEEE IEDM Tech. Dig.*, p. 93-299, 1993.

[6] C. Rafferty, "Progress in Predicting Transient Diffusion," *IEEE SISPAD proceedings*, pp. 1-4, 1997.

[7] C.S. Rafferty et al., "Modeling Transient Diffusion following High Energy Implantation," *IEEE IEDM Tech. Dig.*, p. 96-791, 1996.

[8] J.M. Poate et al., "Ion Implantation and Transient Enhanced Diffusion," *IEEE IEDM Tech. Dig.*, p. 95-77, 1995.

[9] N.E.B. Cowern et al., "Mechanisms of implant damage annealing and transient enhanced diffusion in Si," *Appl. Phys. Lett.*, 65(23), pp. 2981-2983, 1994.

[10] N.E.B. Cowern et al., "Role of C and B clusters in transient diffusion of B in silicon," *Appl. Phys. Lett.*, 68(8), pp. 1150-1152, 1996.

[11] P.A. Packan and J.D. Plummer, "Transient diffusion of low-concentration B in Si due to ^{29}Si

implantation damage," *Appl. Phys. Lett.*, 56(18), pp. 1787-1789, 1990.

[12] D.J. Eaglesham et al., "Implantation and transient B diffusion in Si: The source of the interstitials," *Appl. Phys. Lett.*, 65(18), pp. 2305-2307, 1994.

[13] P.B. Griffin et al., "Species, Dose and Energy Dependence of Implant Induced Transient Enhanced Diffusion," *IEEE IEDM Tech. Dig.*, p. 93-295, 1993.

[14] A. Hofler et al., "A technology oriented model for transient diffusion and activation of boron in silicon," *J. Appl. Phys.*, 78(6), pp. 3671-3679, 1995.

[15] A. Cacciato et al., "Dislocation formation and B transient diffusion in C coimplanted Si," *J. Appl. Phys.*, 79(5), pp. 2314-2325, 1996.

[16] K.S. Jones et al., "Diffusion of ion implanted boron in preamorphized silicon," *Appl. Phys. Lett.*, 68(19), pp. 2672-2674, 1996.

[17] S. Crowder et al., "The Effect of Source/Drain Processing on the Reverse Short Channel Effect of Deep Sub-Micron Bulk and SOI nMOSFETs," *IEEE IEDM Tech. Dig.* p. 95-427, 1995.

[18] M. Orlowski, C. Mazure, and F. Lau, "Submicron Short Channel Effects Due to Gate Reoxidation Induced Lateral Interstitial Diffusion," *IEEE IEDM Tech. Dig.*, p. 87-632, 1987.

[19] C. Mazure and M. Orlowski, "Guidelines for Reverse Short-Channel Behavior," *IEEE Electron Device Lett.*, vol. EDL-10(12), pp. 556-558, 1989.

[20] C. Rafferty et al., "Explanation of Reverse Short Channel Effect by Defect Gradients," *IEEE IEDM Tech. Dig.*, p. 93-311, 1993.

[21] K. Nishi et al., "Evidence of Channel Profile Modification Due to Implantation Damage Studied by a New Method, and Its Implication to Reverse Short Channel Effects of NMOSFETs," *IEEE IEDM Tech. Dig.*, p. 95-993, 1995.

[22] C.G. Shahidi et al., "Indium Channel Implant for Improved Short-Channel Behavior of Submicrometer NMOSFET's," *IEEE Electron Device Lett.*, EDL-14(8), pp. 409-411, 1993.

[23] H. Hu et al., "A Study of Deep-Submicron MOSFET Scaling Based on Experiment and Simulation," *IEEE Tran. Electron Devices*, ED-42(4), pp. 669-677, 1995.

[24] C. Wann et al., "Channel Profile Optimization and Device Design for Low-Power High-Performance Dynamic-Threshold MOSFET," *IEEE IEDM Tech. Dig.*, p. 96-113, 1996.

[25] S. Venkatesan et al., "Device Drive Current Degradation Observed with Retrograde Channel Profiles," *IEEE IEDM Tech. Dig.*, p. 95-419, 1995.

[26] S.E. Thompson, P.A. Packan and M.T. Bohr, "Linear versus Saturated Drive Current: Tradeoffs in Super Steep Retrograde Well Engineering," *Symp. on VLSI Tech. Dig.*, pp. 154-155, 1996.

[27] F. Benistant et al., "A Fully Implanted Heavy Ion 0.10 μm gate length NMOS transistor with Gallium channel implantation and Indium pocket doping," *Abstract of ISDRS'95*, 1995.

[28] C-Y. Wei, J.M. Pimbley and Y. Nissan-Cohen, "Buried and Graded/Buried LDD Structures for Improved Hot-Electron Reliability," *IEEE Electron Device Lett.*, EDL-7(6), pp. 380-382, 1986.

[29] B. Yu et al., "Short-Channel Effect Improved by Lateral Channel-Engineering in Deep-Submicronmeter MOSFET's," *IEEE Tran. Electron Devices*, vol. ED-44(4), pp. 627-634, 1997.

[30] H. Hwang, D-H. Lee, and J.M. Hwang, "Degradation of MOSFETs Drive Current due to Halo Ion Implantation," *IEEE IEDM Tech. Dig.*, p. 96-567, 1996.

[31] S.M. Sze, "Physics of Semiconductor Devices," 2nd Edition, John Wiley and Sons, New York, 1981.

[32] N. Khalil et al., "A B-splines regression technique to determine one-dimensional MOS doping profiles," *Proc. ESSDERC '95*, pp. 191-4, 1995.

[33] A. Pieracci et al., "Extraction of Channel Doping Profile in DMOS Transistors," *IEEE IEDM Tech. Dig.*, p. 96-485, 1996.

[34] P.C. Zalm, "The application of dynamic SIMS in silicon semiconductor technology," *Philips Journal of Research*, vol.47, no.3-5, pp. 287-302, 1993.

[35] R.N. Kleiman et al., "Junction Delineation of 0.15 μm MOS Devices Using Scanning Capacitance Microscopy," *IEEE IEDM Tech. Dig.*, p. 97-691, 1997.

[36] N. Khalil et al., "The Extraction of Two-Dimensional MOS Transistor Doping via Inverse Modeling," *IEEE Electron Device Lett.*, vol. EDL-16 (1), pp. 17-19, 1995.

[37] G.J.L. Ouwerling, "A Problem-Specific Inverse Method For Two-Dimensional Doping Profile Determination From Capacitance-Voltage Measurements," *Solid State Electronics*, 34(2), pp. 197-214, 1991.

[38] G.J.L. Ouwerling, J.C. Staalenburg and M. Kleefstra, "Electrical Characterization of 2D Doping Profiles," *IEEE 1990 Int. Conf. on Microelectronic Test Structures*, vol. 3, pp. 3-8, 1990.

[39] G.J.L. Ouwerling, "Physical Parameter Extraction by Inverse Device Modeling: Application to One- and Two-Dimensional Doping Profiling," *Solid State Electronics*, 3(6), pp. 757-771, 1990.

[40] G.J.L. Ouwerling, "Non-Destructive Measurement of 2D Doping Profiles by Inverse Modeling," *Proc. of the Sixth Int. Nasecode Conf.*, pp. 534-539, 1989.

[41] Z.K. Lee, M.B. McIlrath, and D.A. Antoniadis, "Inverse Modeling of MOSFETs using I-V Characteristics in the Subthreshold Region," *IEEE IEDM Tech. Dig.*, p. 97-683, 1997.

[42] Z.K. Lee, M.B. McIlrath, and D.A. Antoniadis, "Two-Dimensional Doping Profile Characterization of MOSFETs by Inverse Modeling using I-V Characteristics in the Subthreshold Region," submitted to *IEEE Tran. Electron Devices*, 1998.

[43] MEDICI manual, version 2.3.1, Technology Modeling Associates, 1996.

[44] Y. Tsividis, "Operation and Modeling of the MOS Transistor," McGraw-Hill, New York, 1987.

[45] J. Brews, "Physics of the MOS Transistor, in D. Kahng (ed.), "Silicon Integrated Circuits," part A, Academic Press, New York, 1981.

[46] J.R. Brews, "A Charge-Sheet Model of the MOSFET," *Solid State Electronics*, vol. 21, pp. 345-355, 1978.

[47] E.S. Yang, "Microelectronic Devices," McGraw-Hill, New York, 1988.

[48] G. W. Taylor, "Subthreshold Conduction in MOSFET's", *IEEE Trans. Electron Devices*, vol. ED-25(3), pp. 337-350, 1978.

[49] J.R. Brews, "Subthreshold Behavior of Uniformly and Nonuniformly Doped Long-Channel MOSFET," *IEEE Trans. Electron Devices*, vol. ED-26(9), pp. 1282-1291, 1979.

[50] R.R. Troutman, "VLSI Limitations from Drain-Induced Barrier Lowering," *IEEE J. of Solid-State Circuits*, SC-14(2), pp. 383-391, 1979.

[51] R.R. Troutman, "VLSI Limitations from Drain-Induced Barrier Lowering," *IEEE Trans. Electron Devices*, ED-26(4), pp. 461-469, 1979.

[52] R.R. Troutman, "Subthreshold Design Considerations for Insulated Gate Field-Effect Transistor," *IEEE J. of Solid-State Circuits*, SC-9(2), pp. 55-60, 1974.

[53] D.R. Poole and D.L. Kwong, "Two-Dimensional Analytical Modeling of Threshold Voltages of Short-Channel MOSFET's," *IEEE Electron Device Lett.*, EDL-5(11), pp. 443-446, 1984.

[54] P-S. Lin and C-Y. Wu, "A New Approach to Analytically Solving the Two-Dimensional Poisson's Equation and Its Application in Short-Channel MOSFET Modeling," *IEEE Trans. Electron Devices*, vol. ED-34(9), pp. 1947-1956, 1987.

[55] M. Conti and C. Turchetti, "On the Short-Channel Theory for MOS Transistor," *IEEE Trans.*

Electron Devices, vol. ED-38(12), pp. 2657-2661, 1991.

[56] C.R. Viswanathan et al., "Threshold Voltage in Short-Channel MOS Devices," *IEEE Trans. Electron Devices*, vol. ED-32(5), pp. 932-940, 1985.

[57] T. Skotnicki, G. Merckel, and T. Pedron, "The Voltage-Doping Transformation: A New Approach to the Modeling of MOSFET Short-Channel Effects," *IEEE Electron Device Lett.*, EDL-9(3), pp. 109-112, 1988.

[58] S. Biesemans, S. Kubicek, and K. de Meyer, "New Current-Defined Threshold Voltage Model from 2D Potential Distribution Calculations in MOSFETs," *Solid State Electronics*, no. 1, pp. 43-48, 1996.

[59] Z. Liu et al., "Threshold Voltage Model for Deep-Submicrometer MOSFET's," *IEEE Trans. Electron Devices*, vol. ED-40(1), pp. 86-94, 1993.

[60] J.S. Fu, "Dominant Subthreshold Conduction Paths in Short-Channel MOSFET's," *IEEE Trans. Electron Devices*, vol. ED-31(4), pp. 440-447, 1984.

[61] K. Doganis and D.L. Scharfetter, "General Optimization and Extraction of IC Device Model Parameters," *IEEE Trans. Electron Devices*, vol. ED-30(9), pp. 1219-1228, 1983.

[62] R.R. Meyer and P.M. Roth, "Modified Damped Least Squares: an algorithm for nonlinear optimization," *Journal of the Institute of Mathematics and its Applications*, vol. 9, pp. 218-233, 1972.

[63] C.G. Broyden, "A class of methods for solving nonlinear simultaneous equations," *Mathematics of Computation*, vol. 19, pp. 577-593, 1965.

[64] E.V. Shikin and A.I. Plis, "Handbook on Splines for the User," CRC Press, New York, 1995.

[65] R. Rios and N.D. Arora, "Determination of Ultra-Thin Gate Oxide Thicknesses for CMOS Structures Using Quantum Effects," *IEEE IEDM Tech. Dig*, p. 94-613, 1994.

[66] R. Rios et al., "A Physical Compact MOSFET Model, Including Quantum Mechanical Effects, for Statistical Circuit Design Applications," *IEEE IEDM Tech. Dig*, p. 95-937, 1995.

[67] A. Spinelli, A. Benvinuti, and A. Pacelli, "Investigation of Quantum Effects in Highly-Doped MOSFET's by Means of a Self-Consistent 2D Model," *IEEE IEDM Tech. Dig*, p. 96-399, 1996.

[68] P. Vande Voorde et al., "Accurate Doping Profile Determination Using TED/QM Models Extensible to Sub-Quarter Micron nMOSFETs," *IEEE IEDM Tech. Dig*, p. 96-811, 1996.

[69] J.M. Shannon, "D.C. Measurement of the Space Charge Capacitance and Impurity Profile Beneath the Gate of an MOST," *Solid State Electronics*, vol. 14, pp.1099-1106, 1971.

[70] M.G. Buehler, "The D-C MOSFET Dopant Profile Method," *J. Electrochem. Soc.*, 127(3), pp. 701-704, 1980.

[71] M.G. Buehler, "Dopant profiles determined from enhancement-mode MOSFET dc measurements," *Appl. Phys. Lett.*, 31(12), pp. 848-851, 1977.

[72] See, for example, J.R Brews (ed: S.M. Sze), "The Sub-micron MOSFET" in "High-Speed Semiconductor Devices," John Wiley and Sons, New York, 1990.

[73] C. Huang et al., "An Accurate Gate Length Extraction Method for Sub-Quarter Micron MOSFET's," *IEEE Trans. Electron Devices*, vol. ED-43(6), pp. 958-964, 1996.

[74] Y. Taur et al., "A New "Shift and Ratio" Method for MOSFET Channel-Length Extraction," *IEEE Electron Device Lett.*, vol. EDL-13(5), pp. 267-269, 1992.

[75] K, Terada and H. Muta, "A New Method to Determine Effective MOSFET Channel Length," *Japn. J. of Appl. Phys.*, 18(5), pp. 953-959, 1979.

[76] S.C. Sun and J.D. Plummer, "Electron Mobility in Inversion and Accumulation Layers on Thermally Oxidized Silicon Surfaces," *IEEE Trans. Electron Devices*, vol. ED-27(8), pp. 1497-1508, 1980.

[77] S. Jallepalli et al., "Understanding The Differences In The Effective-Field Dependence of Electron and Hole Inversion Layer Mobilities," *IEEE IEDM Tech. Dig*, p. 96-391, 1996.

[78] C-L. Huang et al., "A New Technique for Measuring MOSFET Inversion Layer Mobility," *IEEE Trans. Electron Devices*, vol. ED-40(6), pp. 1134-1139, 1993.

[79] K. Yamaguchi, "A Mobility Model for Carriers in the MOS Inversion Layer," *IEEE Trans. Electron Devices*, vol. ED-30(6), pp. 658-1074, 1983.

[80] K. Yamaguchi, "Field-Dependent Mobility Model for Two-Dimensional Numerical Analysis of MOSFET's," *IEEE Trans. Electron Devices*, vol. ED-26(7), pp. 1068-1074, 1979.

[81] K.K. Thornber, "Relation of drift velocity to low-field mobility and high-field saturation velocity," *J. Appl. Phys.*, 51(4), pp. 2127-2136, 1980.

[82] M.R. Pinto, E. Sangiorgi, and J. Bude, "Silicon MOS Transconductance Scaling into the Overshoot Regime," *IEEE Electron Device Lett.*, EDL-14(8), pp. 375-378, 1993.

[83] T. Mizuno and R. Ohba, "Experimental Study of Carrier Velocity Overshoot in Sub-0.1 μm Devices - Physical Limitation of MOS Structures," *IEEE IEDM Tech. Dig*, p. 96-109, 1996.

[84] G. Masetti, M. Severi, and S. Solmi, "Modeling of Carrier Mobility Against Carrier Concentration in Arsenic-, Phosphorus-, and Boron-Doped Silicon," *IEEE Trans. Electron Devices*, vol. ED-30(7), pp. 764-769, 1983.

[85] T. Nishida and C-T Sah, "A Physically Based Mobility Model for MOSFET Numerical Simulation," *IEEE Trans. Electron Devices*, vol. ED-34(2), pp. 310-319, 1987.

[86] C. Sah et al., "The Scattering of Electrons by Surface Oxide Charges and by Lattice Vibrations at the Silicon-Silicon Dioxide Interface," *Surface Sci.*, vol. 32, pp. 561-575, 1972.

[87] S.A. Schwarz and S. E. Russek, "Semi-Empirical Equations for Electron Velocity in Silicon: Part I - Bulk," *IEEE Trans. Electron Devices*, vol. ED-30(12), pp. 1629-1633, 1983.

[88] S.A. Schwarz and S. E. Russek, "Semi-Empirical Equations for Electron Velocity in Silicon: Part II - MOS Inversion Layer," *IEEE Trans. Electron Devices*, vol. ED-30(12), pp. 1634-1639, 1983.

[89] C. Lombardi et al., "A Physically Based Mobility Model for Numerical Simulation of Nonplanar Devices," *IEEE Trans. Computer-Aided Design*, vol. 7(11), pp. 1164-1170, 1988.

[90] D.M. Caughey, and R.E. Thomas, "Carrier mobilities in silicon empirically related to doping and field,", *Proc. IEEE*, vol. 55, pp. 2192-2193, 1967.

[91] B. Meinerzhagen and W.L. Engl, "The Influence of the Thermal Equilibrium Approximation on the Accuracy of Classical Two-Dimensional Numerical Modeling of Silicon Submicrometer MOS Transistors," *IEEE Trans. Electron Devices*, vol. ED-35(5), pp. 689-697, 1988.

[92] P. Antognetti, C. Lombardi, and D. Antoniadis, "Use of Process and 2-D MOS Simulation in the Study of Doping Profile Influence on S/D Resistance in Short Channel MOSFET's," *IEEE IEDM Tech. Dig*, p. 81-574, 1981.

[93] A.H. Perera and J.P. Krusius, "Ultimate CMOS Density Limits: Measured Source Drain Resistance in Ultra Small Devices," *IEEE IEDM Tech. Dig*, p. 89-625, 1989.

[94] K.K. Ng and W.T. Lynch, "The Impact of Intrinsic Series Resistance on MOSFET Scaling," *IEEE Trans. Electron Devices*, vol. ED-34(3), pp. 503-511, 1987.

[95] Y. Taur et al., "Source-Drain Contact Resistance in CMOS with Self-Aligned $TiSi_2$," *IEEE Trans. Electron Devices*, vol. ED-34(3), pp. 575-580, 1987.

[96] G. A. Sai-Halasz, "Device-Grade Ultra-Shallow Junctions Fabricated with Antimony," *IEEE Electron Device Lett.*, EDL-7(9), pp. 534-536, 1986.

[97] G. Baccarani and G.A. Sai-Halasz, "Spreading Resistance in Submicron MOSFETs," *IEEE Electron Device Lett.*, EDL-4(2), pp. 27-29, 1983.

[98] A. Yagi and J. Frey, "Effects of Injection Resistance on the Performance of Very-Short-Channel MOSFETs," *IEEE Trans. Electron Devices*, vol. ED-31(12), pp. 1804-1808, 1984.

[99] K.K. Ng, R.J. Bayruns, and S.C. Fang, "The Spreading Resistance of MOSFET's," *IEEE Electron Device Lett.*, EDL-6(4), pp. 195-198, 1985.

[100] K.K. Ng and W.T. Lynch, "Analysis of the Gate-Voltage-Dependent Series Resistance of MOSFET's," *IEEE Trans. Electron Devices*, vol. ED-33(7), pp. 965-972, 1986.

[101] D.B. Scott, W.R. Hunter, and H. Shichijo, "A Transmission Line Model for Silicided Diffusions: Impact on the Performance of VLSI Circuits," *IEEE Trans. Electron Devices*, vol. ED-29(4), pp. 651-661, 1982.

[102] D.B. Scott et al., "Titanium Disilicide Contact Resistivity and Its Impact on 1-µm CMOS Circuit Performance," *IEEE Trans. Electron Devices*, vol. ED-34(3), pp. 562-573, 1987.

[103] K. Suzuki et al., "Source/Drain Contact Resistance of Silicided Thin-Film SOI MOSFET's," *IEEE Trans. Electron Devices*, vol. ED-41(6), pp. 1007-1012, 1994.

[104] K.K. Ng, "A Novel Technique to Measure the Contact Resistance of a MOSFET," *IEEE Trans. Electron Devices*, vol. ED-34(3), pp. 544-547, 1987.

[105] M. Tabesky et al., "Direct Silicidation of Co on Si by Rapid Thermal Annealing," *IEEE Trans. Electron Devices*, vol. ED-34(3), pp. 548-553, 1987.

[106] L.V.D. Hove et al., "A Self-Aligned $CoSi_2$ Interconnection and Contact Technology for VLSI Applications," *IEEE Trans. Electron Devices*, vol. ED-34(3), pp. 554-561, 1987.

[107] K. Maex, "Silicides for Integrated Circuits: $TiSi_2$ and $CoSi_2$," *Materials Science and Engineering*, vol. R11(2-3), 1993.

[108] R.L. Gillenwater et al., "The Effect of Lateral Current Spreading on the Specific Contact Resistivity in D-Resistor Kelvin Devices," *IEEE Trans. Electron Devices*, vol. ED-34(3), pp. 537-543, 1987.

[109] W.M. Loh et al., "Modeling and Measurement of Contact Resistances," *IEEE Trans. Electron Devices*, vol. ED-34(3), pp. 512-523, 1987.

[110] S. Thompson et al, "Source/Drain Extension Scaling for 0.1 µm and Below Channel Length MOSFETS," *Symp. on VLSI Tech. Dig.*, 1997.

Appendix A

Levenberg-Marquardt Optimization

A.1 Introduction

All of the inverse modeling work in this thesis relies on fitting simulated data to experimental data by modifying model parameters. This is accomplished by using the Levenberg-Marquardt optimization algorithm [61-62] that finds the values of the model parameters such that the RMS error (or least squares error) between the simulated data and experimental data is minimized. In this appendix, the algorithm is reviewed.

A.2 Optimization Method

The goal of the optimization algorithm is to minimize the least squares error given by the scalar objective function

$$F(\bar{p}) = \overline{h(\bar{p})}^T \overline{h(\bar{p})} = \sum_j h_j^2(\bar{p}) = \sum_j [x_j^{\mathrm{expt}} - x_j^{sim}(\bar{p})]^2, \qquad \text{(A.2.1)}$$

where $\bar{p}$, $\overline{h(\bar{p})}$, x_j^{expt}, and $x_j^{sim}(\bar{p})$ are, respectively, the model parameter vector, the error vector, the value of the jth experimental data point, and the value of the jth simulated data point (which depends on the model parameter vector). The minimum of $F(\bar{p})$ can be found by solving for $\bar{p}$ such that the coupled set of equations is satisified:

$$\nabla_{\bar{p}} F(\bar{p}) = \overline{g(\bar{p})} = 2J^T(\bar{p})\overline{h(\bar{p})} = \bar{0}. \tag{A.2.2}$$

Here, $\overline{g(\bar{p})}$ is the gradient of $F(\bar{p})$ with respect to $\bar{p}$, and $J^T(\bar{p})$ is the Jacobian matrix given by

$$J_{ij}(\bar{p}) = \frac{\partial h_i(\bar{p})}{\partial p_j}. \tag{A.2.3}$$

Applying Newton's method to eqn. (A.2.2), and with the explicit functional dependences dropped, one gets, for each iteration,

$$\left[2J^T J + 2\sum_j h_j \nabla^2 h_j\right]\overline{\delta p} = H\overline{\delta p} = -2J^T\bar{h}, \tag{A.2.4}$$

where $\overline{\delta p}$ is the search vector for each iteration update, such that $\bar{p}^i$ in the ith iteration is updated to

$$\bar{p}^{i+1} = \bar{p}^i + \overline{\delta p}^i. \tag{A.2.5}$$

The matrix in square brackets on the left-hand-side of eqn. (A.2.4) is the Hessian matrix which can also be written explicitly as

$$H_{ij} = \frac{\partial^2 F}{\partial p_i \partial p_j}. \tag{A.2.6}$$

Since H involves second order derivatives, which are very computationally expensive, different approaches to approximate it have been proposed, giving rise to the Newton-Gauss and Levenberg-Marquardt algorithms.

In the Newton-Gauss approximation, the second term on the left-hand-side of eqn. (A.2.4) is dropped, leaving

$$H \approx 2J^T J. \tag{A.2.7}$$

This is a good approximation when $\bar{p}$ is close to the minimum (i.e., when $\bar{h} \to \bar{0}$).

The Levenberg-Marquardt method approximates H by

$$H \approx 2J^T J + \lambda D, \tag{A.2.8}$$

where λ is a scalar parameter, and D is a diagonal matrix having components

$$D_{ij} = 2(J^T J)_{ij}\delta_{ij}. \tag{A.2.9}$$

During optimization, λ is initially chosen to be large, so that the search vector $\overline{\delta p}$ is forced to be parallel to the negative of the gradient (i.e., follows the direction of steepest descend). If this update leads to a reduction of the residual error, λ is reduced for the next iteration. Therefore, near the minimum, $\lambda \to 0$ so that the method approaches the Newton-Gauss approximation. On the other hand, if an update leads to divergence, λ is increased so that the search vector simply follows the direction of steepest descend. Details of the Levenberg-Marquardt algorithm is provided in the next section. Note that the Newton-Gauss method can be considered as a special case here, by setting $\lambda = 0$.

A.3 Complete Algorithm

The algorithm is summarized below. Parameters that can be changed by the user are shown in boldface. Their default values are given in section A.5.

(1) Initial guess for $\bar{p}^0$.

(2) Set λ = **lambda_init**.

(3) Solve $(2J^T J + \lambda D)\overline{\delta p} = -2J^T \bar{h}$.

(4) Check to make sure that the search vector is going downhill; otherwise increase λ and recalculate:

if $\bar{g} \bullet \overline{\delta p} = (2J^T \bar{h}) \bullet \overline{\delta p} > 0$, then $\lambda \to \lambda$ ***lambda_scale**; goto (3).

(5) Check if the norm of the search vector is within the user-specified range; otherwise scale it by an appropriate factor:

if $\|\overline{\delta p}\| > \left\|\overline{\delta p}^{allowed}\right\|$, scale $\overline{\delta p}$ so that $\|\overline{\delta p}\| = \left\|\overline{\delta p}^{allowed}\right\|$.

(6) $\overline{\delta p} \rightarrow \overline{\delta p} *$**damping_init**.

(7) $\bar{p} \rightarrow \bar{p} + \overline{\delta p}$.

(8) Check if $\bar{h}^T\bar{h}$ is smaller than that of the previous iteration; otherwise, scale $\overline{\delta p}$ by a factor and recalculate $\bar{h}^T\bar{h}$. Repeat the process until the condition is satisfied or the user-specified number of times that this is allowed is exceeded. i.e.,

```
    calculate $\bar{h}^T\bar{h}$
    if ($\bar{h}^T\bar{h} > \bar{h}^T\bar{h}$ of previous iteration), then
    repeat $\bar{p} \rightarrow \bar{p} - \overline{\delta p}$          (un-update)
           $\overline{\delta p} \rightarrow \overline{\delta p}$ *damping_factor
           $\bar{p} \rightarrow \bar{p} + \overline{\delta p}$          (update with new $\overline{\delta p}$)
           calculate $\bar{h}^T\bar{h}$
    until  ($\bar{h}^T\bar{h}$ is smaller than that of previous iteration)
           or (number of times in this loop > failcountMax)
           or (damping_min > damping_init * damping_factor^(number of times in
              loop));
    if (number of times in loop > failcountMax) or (damping_min > damping_init *
       damping_factor^(number of times in loop)), then
           $\lambda \rightarrow \lambda$ *lambda_fail
    else
           $\lambda \rightarrow \lambda$ /lambda_scale
```

(9) goto (3) for new iteration or stop if convergence criteria are met.

A.4 Other Issues

Evaluation of the Jacobian is normally accomplished by either forward difference or centered difference. In evaluating derivatives, the following scheme is used:

$$\frac{dy}{dx} \approx \frac{y(x+\Delta x) - y(x)}{\Delta x},$$

where Δx = abs(x * **fraction**) + **verysmall**. This process is very time consuming. However, it is possible to estimate the Jacobian without carrying out the finite differencing (which requires running new simulations), based on the results of two iterations. Broyden proposed a formula [63] that is found very useful:

$$J^{i+1} = J^i + \left[\frac{\bar{h}^i - \bar{h}^{i-1} - J^i(\bar{p}^i - \bar{p}^{i-1})^T}{\left\|\bar{p}^i - \bar{p}^{i-1}\right\|^2}\right](\bar{p}^i - \bar{p}^{i-1})^T. \tag{A.4.1}$$

Of course, this formula is only an approximation. Over-use will lead to non-convergence. The user can specify the number of iterations that this approximation is allowed before returning to a full forward or centered difference Jacobian evaluation by specifying the parameter **iter_restart_Jacobian.**

Typically, 3 to 15 iterations are needed. The amount of CPU time it takes varies greatly, depending on the number of model parameters, number of data points, as well as the grid size and convergence of the device simulations. For a 2D doping profile extraction, the amount of time needed normally ranges from 4 hours to 15 hours.

Termination of the optimization process occurs if one or more of the following conditions is satisfied: (1) $F(\bar{p}) = \bar{h}^T\bar{h}$ < **tolerance**; (2) number of iterations > **maxIt**; and (3) $F^i(\bar{p})/F^{i-1}(\bar{p})$ < **fractionImprove**. In practice, it is often condition (3) that leads to termination.

A.5 Default Parameters

The default parameters and their values are summarized below:

Table A.5.1: Default parameters used in optimizer.

fraction	1e-5
verysmall	1e-7
lambda_init	0.01
lambda_scale	4
lambda_fail	10
damping_init	1
damping_factor	0.5
damping_min	0.1
failcountMax	5
iter_restart_Jacobian	2
maxIt	25
fractionImprove	0.02

Appendix B

Miscellaneous Information

B.1 One-Dimensional B-Spline

One dimensional B-spline curves are useful for parameterizing the depthwise variation of doping concentration in MOSFETs. An elementary cubic B-spline curve, defined by four points $\overline{P_O}, \overline{P_1}, \overline{P_2}, \overline{P_3}$, is given by the parametric equation [64]:

$$\overline{R(t)} = \frac{(1-t)^3}{6}\overline{P_0} + \frac{3t^3 - 6t^2 + 4}{6}\overline{P_1} + \frac{-3t^3 + 3t^2 + 3t + 1}{6}\overline{P_2} + \frac{t^3}{6}\overline{P_3}, \qquad \text{(B.1.1)}$$

where $0 \le t \le 1$, $\overline{R(t)} = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$, and $\overline{P_0} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ ... etc. In a more compact form, eqn. (B.1.1) can be written as:

$$\overline{R(t)} = PMT, \qquad \text{(B.1.2)}$$

where

$$\bar{P} = \begin{bmatrix} \overline{P_0} \; \overline{P_1} \; \overline{P_2} \; \overline{P_3} \end{bmatrix} = \begin{bmatrix} x_0 & x_1 & x_2 & x_3 \\ y_0 & y_1 & y_2 & y_3 \end{bmatrix}, \qquad \text{(B.1.3)}$$

$$M = \begin{bmatrix} 1 & -3 & 3 & -1 \\ 4 & 0 & -6 & 3 \\ 1 & 3 & 3 & -3 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{B.1.4}$$

$$T = \begin{bmatrix} t^0 \\ t^1 \\ t^2 \\ t^3 \end{bmatrix}. \tag{B.1.5}$$

A more genaral composite B-spline curve, defined by the points $\overline{P_O}, \overline{P_1}, \overline{P_2}, \overline{P_3}, \ldots, \overline{P_m}$, with $m \geq 3$, is made up of fragments given by:

$$\overline{R^{(i)}(t)} = \left[\overline{P_{i-1}}\ \overline{P_i}\ \overline{P_{i+1}}\ \overline{P_{i+2}}\right] MT, \tag{B.1.6}$$

where $i = 1, \ldots, m-2$. The distinguishing properties of B-spline curves, amongst others (see [64]), include: (1) the curve does not interpolate the points; and (2) the change of one point causes the curve to change in a region near the changed point only. Note that B-spline curves generally do not start and end at the end points (since they do not interpolate the points). However, it is possible to force them to start and end closer to the end points by making the end points doubly or triply degenerate. For the work in this thesis, doubly degenerate end points are used.

B.2 Two-Dimensional Doping Profile

For the work in this thesis, the 2D doping profiles in the region of interest are represented by:

$$D(x, y) = \sum_i A_i \exp[-(x-B_i)^2/C_i^2] \exp[-(y-D_i)^2/E_i^2] + S(y;\bar{F}), \tag{B.2.1}$$

where S is a 1D B-spline curve, and the $A_i, B_i, C_i, D_i, E_i, \bar{F}$ are fitting parameters.

Appendix C

The "Shift and Ratio" Method for Effective Channel-Length Extraction

Channel-length is a very important parameter for gauging technologies. A reliable method for extracting such is necessary. Throughout this thesis, the "shift and ratio" (SAR) method [74] is used extensively for extracting the effective channel-length.

In addition to the device in question, the application of the SAR method also require a long-channel device whose channel-length is known. The basic assumptions of the method are: (1) I_D in the linear region varies inversely with the effective channel-length L_{eff} [as in eqn. (2.3.28)]; and (2) the linear region I-V characteristics of the long-channel device and the device in question have the same shape (i.e., can be described by the same functional form). Therefore, the total resistance of each device can be described by:

$$R_{tot,i} = \left(\frac{V_{DS}}{I_D}\right)_i = R_{ext} + L_{eff,i}F(V_{GS} - V_{t,i}), \qquad \text{(C.1)}$$

where the subscript indices denote the device, R_{ext} is the external parasitic resistance, and F is a function which describes the shape of the linear region I-V curve of the devices. Differentiating eqn. (C.1) eliminates R_{ext}, so that it does not affect the extraction results, leading to

$$S_i(V_{GS} - V_{t,i}) = \frac{dR_{tot,i}}{dV_{GS}} = L_{eff,i}G(V_{GS} - V_{t,i}). \qquad \text{(C.2)}$$

Defining a ratio H and a shift parameter δ, and ratioing the S_i gives

$$H(V_{GS}, \delta) = \frac{L_{eff,1} G(V_{GS} - V_{t,1})}{L_{eff,2} G(V_{GS} - V_{t,2} - \delta)}. \tag{C.3}$$

Finding an appropriate δ (i.e., $\delta = V_{t,1} - V_{t,2}$) allows G in the numerator and the denominator to cancel, so that H becomes the ratio of the L_{eff}'s of the two devices. If $L_{eff,1}$ is known (i.e., corresponds to that of the long-channel device), $L_{eff,2}$ can be calculated. The value of δ to be used in eqn. (C.3) is the value δ_m such that the variance σ^2 of H is minimized, i.e.,

$$\langle \sigma^2 \rangle = \langle H^2 \rangle - \langle H \rangle^2, \tag{C.4}$$

$$\langle H \rangle = \frac{\int_{\Delta V_{GS}} H(V_{GS}, \delta) dV_{GS}}{\int_{\Delta V_{GS}} dV_{GS}}, \tag{C.5}$$

$$\langle H^2 \rangle = \frac{\int_{\Delta V_{GS}} H^2(V_{GS}, \delta) dV_{GS}}{\int_{\Delta V_{GS}} dV_{GS}}. \tag{C.6}$$

The ratio of the L_{eff}'s is then given by:

$$\langle H \rangle |_{\delta_m} = \frac{L_{eff,1}}{L_{eff,2}}. \tag{C.7}$$

Once the above are found, eqn. (C.1) can be used to calculate R_{ext}, i.e.,

$$R_{ext} = \frac{\langle H \rangle |_{\delta_m} R_{tot,2}(V_{GS} - \delta_m) - R_{tot,1}(V_{GS})}{\langle H \rangle |_{\delta_m} - 1}, \tag{C.8}$$

where the subscript 1 refers to the long-channel device whose L_{eff} is known.

2997-15