DESIGN AND MODELING OF DEEP-SUBMICROMETER MOSFETS

by

Min-Chie Jeng

Memorandum No. UCB/ERL M90/90

2 October 1990

DESIGN AND MODELING OF DEEP-SUBMICROMETER MOSFETS

by

Min-Chie Jeng

Memorandum No. UCB/ERL M90/90

2 October 1990

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Design and Modeling of Deep-Submicrometer MOSFETs

Ph.D Min-Chie Jeng Department of EECS

Ping K. Ko
Committee Chairman

ABSTRACT

A photoresist ashing technique has been developed which, when used in conjunction with conventional optical lithography, permits controlled definition of the gate of deep-submicrometer MOSFETs. This technique can also be extended to other lithographic processes, such as e-beam and x-ray. Comprehensive studies based on the performance and hot-electron reliability have shown that the basic physics associated with deep-submicrometer devices is similar to that of their longer-channel counterparts. Therefore, existing device design guidelines and models can still be used with minor modifications. A set of design curves has been generated based on experimental results with various mechanisms under consideration. With these design curves, the trade-offs between device dimensions and power supply for a particular technology can be observed. The relative importance of each mechanism can also be identified.

A semi-empirical MOSFET drain current model accurate down to quarter-micron channels, suitable for digital as well as analog applications has been developed. Both the drain current and the output resistance are accurately modeled. The first derivative of the drain current equation is continuous from the subthreshold region to the strong-inversion region and from the linear region to the saturation region. for all biases. This model has been implemented in SPICE3. A parameter extraction system dedicated to the model was also developed.

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest appreciations to my research advisor Prof. Ping K. Ko. This dissertation would not have been possible without his guidance and support.

I also would like to thank my co-advisor Prof. Chenming Hu for his helpful discussion and valuable comments throughout this work.

I would like to thank Prof. Charles J. Stone of Statistics Department for being a member of my qualifying exam and for providing me basic statistical background, which is very helpful in extracting model parameters in this work.

I would like to thank Prof. David A. Hodges for his continuous encouragement and support of the BSIM project.

I am grateful to Prof. Bing J. Sheu of USC and Tony Fung for their kind help in the early stage of this work. I thank Dr. Albert T. Wu, James Chung, James E. Moon, Kataline Voros, Marilyn Kushner, Robin R. Rudell, and Tom Booth for their assistance in fabricating deep-submicrometer devices.

I also would like to thank my officemates, Peter M. Lee and Jon S. Duster, for providing me a friendly working environment and for proofreading this dissertation. I really enjoyed working with them.

Finally, I am indebted to my parents who always provide me strength and faith at the right time, and my beloved wife, Ya-Lee, for her patience and sacrifice during my graduate study at Berkeley.

This research was funded by the Semiconductor Research Corporation, the California State MICRO program and JSEP under contract F49620-84-C-0057.

Table of Contents

Chapter 1

INTRODUCTION

In the last decade, MOS devices have been miniaturized to achieve higher packing density, higher integration levels, and higher current drive. Recent advances in process technology [1.1-1.7] have made deep-submicrometer MOSFETs potential candidates for next generation ULSI designs. However, by decreasing channel length while maintaining the current power supply voltage, the electric fields in the device will further increase, causing the device characteristics to deviate from the long-channel behavior and also creating reliability problems.

The two high-field effects most pronounced on device performance are the mobility degradation due to vertical-field [1.8-1.11] and the carrier velocity saturation [1.12,1.13]. Both effects cause the MOSFET drain current drive to increase at a slower rate than that predicted by simple scaling theories. The threshold voltage shift and subthreshold swing are also larger at shorter channel lengths and high drain voltages [1.14-1.20] which make short-channel MOS transistors more difficult to turn off. Such parameter variations have a severe impact on worst case circuit design rules and pose serious problems in VLSI process control.

More consequences of the high electric fields in submicrometer devices are the hot-electron effects [1.21,1.22] due to impact ionization in the velocity saturation (pinch-off) region. The injection of energetic electrons released by impact ionization into the Si-SiO_2 interface generates interface traps that degrade the device characteristics, and results in long-term reliability problems [1.23-1.27]. The substrate current, which is composed of impact-ionization-generated holes, can overload the substrate-bias generator and causes snap-back and CMOS latch-up [1.28,1.29]. In addition to the performance and hot-electron reliability, which are two major concerns for the feasibility of deep-submicrometer MOSFETs in circuit applications, the increasing complexity in circuit designs and fabrication processes is another subject to consider in developing VLSI/ULSI systems.

Hot-electron effects and their related reliability issues together with the already complicated short-channel effects make deep-submicrometer MOS device design much more difficult than ever. From a device design point of view, fabricating devices with optimal performance and reliability requires a comprehensive understanding of the trade-offs among many factors such as device dimensions, device performance, parameter variations, and process complexity. From a circuit design point of view, to expedite VLSI system design and to reduce development overhead, it is necessary to start the circuit design in the early stages of technology development and to predict the circuit behavior as accurate as possible before the circuit is actually fabricated. However, previous reports on deep-submicrometer devices have focused on how to fabricate these devices without formulating any design guidelines, which makes optimal device design almost impossible. For circuit simulations, most existing MOSFET models are not accurate enough for the deep-submicrometer regime.

This dissertation provides a unified understanding of deep-submicrometer devices through experimental study of basic device characteristics and hot-electron effects. By investigating the effects of device parameter variations on various design constraints, different types of device design curves are obtained. An accurate MOSFET model is also developed based on an improved physical understanding of deep-submicrometer transistors.

1.1 Device design

Traditional electrostatic approaches to scaled device designs have been based on generic guidelines known as constant-field [1.30], constant-voltage, and quasi-constant-voltage [1.31] scaling laws. A summary of these scaling laws and the results are given in Table 1.1 [1.32]. The constant-field (CE) scaling law was first proposed by Denarrd. According to this scaling law, all the device parameters and the power supply are scaled by the same factor k so that the internal electric field strength and patterns are unchanged after scaling. However, because the CE scaling also proportionally reduces the power supply, it lacks TTL compatibility and also reduces the device current driving capability and signal-to-noise ratio.

To avoid the TTL compatibility problems, the constant-voltage (CV) and quasi-constant-voltage (QCV) scalings were proposed. Under these scaling laws, the device dimensions are also scaled by the same factor k as in the CE case, but the power supply is kept constant (CV) or scaled down by a factor of $\sqrt{k}$ (QCV). Although these non-constant field scaling laws are more practical and result in better device and circuit performance, the hot-electron effects are much more severe because the channel electric field in the velocity saturation region is increasing rapidly as the device channel length is reduced. For this reason, it has been generally recognized that as long as the power supply remains high for practical considerations, some type of hot-electron-resistant structure, like LDD, is needed for submicrometer MOS transistors.

In reality, however, some device parameters, such as the source and drain junction depths, are relatively unscalable for most technologies, and the power supply can not be easily scaled. All of these scaling laws are difficult to apply in practice. They are only used as conceptual guidelines for minimizing the short-channel and/or hot-electron effects. Practical scaling approaches should be developed based on device performance limitations and constraints as proposed by Masuda [1.33], Brews [1.34], and Shichijo [1.35] for near-micron devices. Because of technology advances, these design curves and conclusions are not applicable in the deep-submicrometer regime. More recent studies for 0.5μm devices were reported by Takeda [1.36] and Kakumu [1.37]. However, these studies are incomplete as only few design constraints were considered. For deep-submicrometer devices, more physical effects are becoming important and should be taken into considerations when developing design guidelines.

In the first part of this report, a comprehensive study of the performance and reliability constraints on the device dimensions and power supply of deep-submicrometer MOSFETs is presented. A set of design curves, extracted from experimental results, are developed based on the following considerations: short-channel and drain-induced-barrier-lowering effects, off-state leakage current, hot-electron reliability, time-dependent dielectric breakdown, current driving capability, voltage gain, and switching speed. Although these design curves are only for n-

channel non-LDD devices, the same methodology can still be applied to other technologies including p-channel and LDD devices.

1.2 Device Modeling

With increasing system complexity due to high-level integration, an efficient circuit simulator with accurate device models becomes an indispensable tool in VLSI/ULSI designs. A complete device model must be capable of predicting device characteristics for all operating modes over a wide range of device dimensions. Since models with underlying equations derived from semiconductor physics are more extendible to include new physics and suitable for process control and diagnosis, most early MOSFET models are physics-based models. However, with the ever decreasing device dimensions, an accurate model based fully on device physics is impossible to develop due to the 3-dimensional nature of small-geometry devices and other high-field effects. Even if it were feasible, the complicated equation forms involved in a fully physical model would have prohibited its usage for circuit simulation purposes.

Furthermore, a fully physics-oriented modeling approach usually makes the parameter extraction very difficult. The desire to achieve more accurate modeling and alleviate difficulties in parameter extraction created the need to add empirically-based parameters to the existing physical parameters. This type of model is categorized as a semi-empirical model. The semi-empirical model retains the basic functional form of fully physics-based models while replacing sophisticated equations by empirical equations with fitting parameters to account for small-geometry effects and minor process variations. Since semi-empirical models have the advantages of simplicity and computational efficiency, all models in circuit simulations to date, to a certain extent, have been semi-empirical models.

It has been shown that properly designed deep-submicron MOSFETs exhibit device characteristics similar to those of their longer-channel counterparts [1.38], but significant second-order effects due to previously negligible physical phenomena make existing drain current models unsatisfactory. Furthermore, many of the drain current models used for circuit

simulations are inadequate in modeling the output resistance and the weak-inversion characteristics, which are very important for analog applications. Since deep-submicron devices typically have thin gate oxides, the inversion-layer capacitance becomes comparable to the gate capacitance, which is an important factor to consider in circuit simulations. In order to bridge the gap between deep-submicrometer devices and circuit simulations, a MOSFET drain current model accurate down to quarter-micron channel lengths, suitable for digital as well as analog applications has been developed based on an improved physical understanding of deep-submicrometer MOS transistors.

1.3 Outline

Chapter 2 describes the fabrication process and some of the characterization procedures for the deep-submicrometer MOSFETs used in this study.

Chapter 3 describes some device characterization methods important to the short-channel devices.

Chapter 4 presents a set of design curves derived from experimental results based on a wide range of design considerations. These design curves provide comprehensive design guidelines for deep-submicrometer devices. The relative importance of various mechanisms is also identified.

Chapter 5 describes a deep-submicrometer MOSFET drain current model suitable for both digital and analog simulations. The basic algorithm and theory for parameter extraction are also briefly described.

Chapter 6 concludes this dissertation.

Device parameter	CE	CV	QCV
Power supply (V_{DD})	1/k	1	$1/\sqrt{k}$
Gate oxide (T_{ox})	1/k	$1/\sqrt{k}$	1/k
Channel length (L)	1/k	1/k	1/k
Channel width (W)	1/k	1/k	1/k
Junction depth (X_j)	1/k	1/k	1/k
Doping concentration (N_{SUB})	k	k	k
Threshold voltage (V_{th})	1/k	1	$1/\sqrt{k}$
Saturation current (I_{DSAT})	1/k	$\sqrt{k}$	1
Transconductance (g_m)	1	$\sqrt{k}$	$\sqrt{k}$
Output resistance (R_{out})	1	1/k	$1/k^{3/4}$
Unity gain frequency (f_T)	k	k^2	$k^{3/2}$
Power dissipation (P)	$1/k^2$	$\sqrt{k}$	$1/k^2$
Power density (P/WL)	1	$k^{5/2}$	$k^{3/2}$
Subthreshold swing (S)	1	1	1

Table 1.1 Results of various scaling laws.

1.4 References

[1.1] W. Fichtner, R.K. Watts, D.B. Fraser, R.L. Johnston, and S.M. Sze, "0.15 μm Channel-Length MOSFETs Fabricated Using E-Beam Lithography," IEDM Tech. Dig., pp.272-725, 1982.

[1.2] W. Fichtner, E.N. Fuls, R.L. Johnston, R.K. Watts, and W.W. Weick, "Optimized MOSFETs with Subquartermicron Channel Lengths," IEDM Tech. Dig., pp.384-387, 1983.

[1.3] T. Kobayashi, S. Horiguchi, and K. Kiuchi, "Deep-Submicron MOSFET Characteristics with 5nm Gate Oxide," IEDM Tech. Dig., pp.414-417, 1984.

[1.4] S. Horiguchi, T. Kobayashi, M. Oda, and K. Kiuchi, "Extremely High Transconductance (Above 500mS/mm) MOSFET with 2.5nm Gate Oxide," IEDM tech. Dig., pp.761-763, 1985.

[1.5] S.Y. Chou, H.I. Smith, and D.A. Antoniadis, "Sub-100-nm Channel-Length Transistors Fabricated Using X-Ray Lithography," J. of Vacuum Science Tech. B, vol. 4, no. 1, pp.253-255, Jan/Feb 1986.

[1.6] J. Chung, M.-C. Jeng, J.E. Moon, A.T. Wu, T.Y. Chan, P.K. Ko, and C. Hu, "Deep-Submicrometer MOS Device Fabrication Using a Photoresist-Ashing Technique," IEEE Electron Device Letters, vol. 9, no. 4, pp.186-188, April, 1988.

[1.7] G.A. Sai-Halasz, M.R. Wordeman, D.P. Kern, E. Ganin, S. Rishton, D.S. Zicherman, H. Schmid, M.R. Polcari, H.Y. Ng, P.J. Restle, T.H.P. Chang, and R.H. Dennard, "Design and Experimental Technology for 0.1 μm Gate-Length Low-Temperature Operation FET's," IEEE Electron Device Letters, vol. EDL-8, no. 10, Oct., 1987.

[1.8] S.C. Sun and J.D. Plummer, "Electron Mobility in Inversion and Accumulation Layers on Thermally Oxidized Silicon Surfaces, " IEEE Tran. on Electron Devices, vol. ED-27, p.1497, August, 1980.

[1.9] A.G. Sabnis and J.T. Clements, "Characterization of the Electron Mobility in the Inverted <100> Si Surface," IEDM Tech. Dig., pp.18-21, 1979

[1.10] M-S. Liang, J.Y. Choi, P.K. Ko, and C. Hu, "Inversion-Layer Capacitance and Mobility of Very Thin Gate-Oxide MOSFET's," IEEE Tran. on Electron Devices, vol. ED-33, p.409, March, 1986.

[1.11] S. Takagi, M. Iwase, and A. Toriumi, " On the Universality of Inversion-Layer Mobility in N- and P-Channel MOSFET's," IEDM Tech. Dig., pp.398-401, 1988.

[1.12] R.W. Coen and R.S. Muller, "Velocity of Surface Carriers in Inversion Layers on Silicon," Solid-State Electronics, vol. 23, p.35, 1980.

[1.13] J.-P. Leburton and G.E. Dorda, "v-E Dependence in Small-Sized MOS Transistors," IEEE Tran. on Electron Devices, vol. ED-29, no. 8, Aug. 1982.

[1.14] H.S. Lee, "An Analysis of the Threshold Voltage for Short-Channel IGFET's," Solid-State Electronics, vol. 16, pp.1407-1417, 1973.

[1.15] L.D. Yau, "A Simple Theory to Predict the Threshold Voltage of Short-Channel IGFET's," Solid-Stae Electronics, vol. 17, pp.1059-1063, 1974.

[1.16] T. Toyabe and S. Asai, "Analytical Models of Threshold Voltage and Breakdown Voltage of Short-Channel MOSFET's Derived from Two-Dimensional Analysis," IEEE J. Solid-State Circuits, vol. SC-14, p.375, April, 1979.

[1.17] K.N. Ratnakumar and J.D. Meindl, "Short-Channel MOST Threshold Voltage Model," IEEE J. Solid-State Circuits, vol. SC-17, p.937, Oct., 1982.

[1.18] H.C. Pao and C.T. Sah, "Effects of Diffusion Current on Characteristics of Metal-Oxide (Insulator) - Semiconductor Transistors, "Solid-State Electronics, vol. 9, p.927, 1966.

[1.19] R.M. Swanson and J.D. Meindl, "Ion-Implanted Complementary MOS Transistor in Low Voltage Circuits," IEEE J. Soilid-State Circuits, vol. SC-7, p.146, April, 1972.

[1.20] R.R. Troutman and S.N. Chakravarti, "Subthreshold Characteristics of Insulated-Gate Field Effect Transistors," IEEE Tran. on Circuit Theory, vol. CT-20, p.659, Nov., 1973.

[1.21] T.H. Ning, P.W. Cook, R.H. Dennard, C.M. Osburn, S.E. Schuster, and H.N. Yu," 1 μm MOSFET VLSI Technology, Part IV. Hot-Electron Design Constrains," IEEE Tran. on Electron Devices, vol. ED-26, pp.346-353, 1979.

[1.22] C. Hu, "Hot-Electron Effects in MOSFET's," IEDM Tech. Dig., pp.176-181, 1983.

[1.23] H. Gesch, J.P. Leburton, and G.E. Dorda, "Generation of Interface States by Hot-Electron Injection," IEEE Tran. on Electron Devices, vol. ED-29, p.913, 1982.

[1.24] E. Takeda and N. Suzuki, "An Empirical Model for Device Degradation due to Hot-Carrier Injection," IEEE Electron Device Letters, vol. EDL-4, pp.111-113, 1983.

[1.25] E. Takeda, A. Shimizu, and T. Hagiwara, "Role of Hot-Hole Injection in Hot-Carrier Effects and the Small Degraded Channel Region in MOSFET's," IEEE Electron Device Letters, vol.EDL-4, no. 9, Sep., 1983.

[1.26] F.C. Hsu and S. Tam, "Relationship between MOSFET Degradation and Hot-Electron Induced Interface-State Generation," IEEE Electron Device Letters, vol. EDL-5, p.50, 1984.

[1.27] K.-L. Chen, S.A. Saller, I.A. Groves, and D.B. Scott, "Reliability Effects on MOS Transistors due to Hot-Carrier Injection," IEEE Tran. on Electron Devices, vol. ED-32, pp.386-393, 1985.

[1.28] Y.W. Sing and B. Sudlow, "Modeling and VLSI Design Constraints of Substrate Current," IEDM Tech. Dig., p.31, 1975.

[1.29] J. Matsunaga, "Characterization of Two Step Impact Ionization and It's Influence on NMOS and PMOS VLSI's," IEDM Tech. Dig., p.736, 1980.

[1.30] R.H. Dennard, F.H. Gaensslen, H.-N. Yu, V.L. Rideout, E. Bassous, and A.R. LeBlanc, "Design of Ion-Implamted MOSFET's with Very Small Physical Dimen-

sions," IEEE J. Solid-State Circuits, vol. SC-9, no. 5, Oct., 1974.

[1.31] P.K. Chatterjee, W.R. Hunter, T.C. Holloway, and Y.T. Lin, "The Impact of Scaling Laws on the Choice of n-Channel or p-Channel for MOS VLSI," IEEE Electron Device Letters, vol. EDL-1, no. 10, Oct., 1980.

[1.32] C.K. Wang, "Switched Capacitor Signal Processing Circuits in Scaled Technologies," Ph.D Dissertation, Univ. of California, Berkeley, 1986.

[1.33] H. Masuda, M. Nakai, and M. Kubo, "Characteristics and Limitation of Scaled-Down MOSFET's Due to Two-Dimensional Field Effect," IEEE Tran. on Electron Devices, vol. ED-26, pp.980-986, June, 1979.

[1.34] J.R. Brews, W. Fichtner, E.H. Nicollian, and S.M. Sze, "Generalized Guide for MOSFET Miniaturization," IEEE Electron Device Letters, vol. EDL-1, no. 1, Jan., 1980.

[1.35] H. Shichijo, "A Re-examination of Practical Performance Limits of Scaled n-Channel and p-Channel MOS Devices for VLSI," Solid-State Electronics, vol. 26, no. 10, pp.969-986, 1983.

[1.36] E. Takeda, G.A.C. Jones, and H. Ahmed, "Constraints on the Application of 0.5-μm MOSFET's to ULSI Systems," IEEE Tran. on Electron Devices, vol. ED-32, no. 2, Feb., 1985.

[1.37] M. Kakumu, M. Kinugawa, K. Hashimoto, and J. Matsunaga, "Power Supply Voltage for Future CMOS VLSI in Half and Sub Micrometer," IEDM Tech. Dig., pp.399-402, 1986.

[1.38] M.-C. Jeng, J. Chung, A.T. wu, T.Y. Chan, J. Moon, G. May, P.K. Ko, and C. Hu, "Performance and Hot-Electron Reliability of Deep-Submicrometer MOSFETs," IEDM Tech. Digs., pp.710-713, 1987.

Chapter 2

DEVICE FABRICATION

The devices used in this study were n-channel non-LDD transistors fabricated using an NMOS technology with a photoresist-ashing technique [2.1] to define the gates of deep-submicrometer devices. Since most steps of this process are common to those of standard fabrication processes, only the major procedures are described. A complete process flow is given in Appendix A.

2.1 Fabrication process

The starting wafers have p-type substrates with 15-30 Ωμm bulk resistivity. A blanket boron (B11) implant of 1.5×10^{12} cm^{-2} at 70 KeV was used for both field and punchthrough controls. The active area was defined using LOCOS. The field oxide thickness of 2800 Å was grown in wet oxygen at 950 °C and annealed in nitrogen for 20 minutes at the same temperature. The enhancement threshold implant dose (B11 at 30KeV) were chosen to yield a long-channel threshold voltage around 0.65V for all gate oxide thicknesses. An array of depletion implant dose (As, 50KeV) were used for these wafers because of the difficulty in determining the threshold voltage due to severe short-channel effects in depletion-mode devices.

Various gate oxide thicknesses, 3.6, 5.6, 7.2, 8.6, and 15.6nm, were grown in dry oxygen at 800-900 °C, depending on the oxide thickness. Immediately after the gate oxidation, a layer of 2500 Å phosphorus-doped polysilicon was deposited using LPCVD. After the gate definition, which will be described in more detail in next section, the n^+ source/drain regions were implanted (As, 3×10^{15} cm^{-2}, 50KeV) with 0° inclination to avoid asymmetric device characteristics [2.2]. Then, a layer of 3000 Å undoped LTO was deposited at 450 °C and densified at 900 °C for 20 minutes in dry oxygen. After etching the contact hole, 2500 Å

phosphorus-doped polysilicon was deposited at 650 °C and activated in nitrogen at 900 °C for 15 minutes. This polysilicon served as a buffer layer to prevent aluminum from spiking through the source/drain region into the substrate. Finally, the contacting metal (Al with 2% Si) was sputtered and defined, followed by an etch of the polysilicon outside the contact area.

In order to minimize the junction depth, all of the subsequent thermal cycles after the source/drain implantation were limited to 900 °C or below, and the total amount of time required by these thermal cycles was less than 60 minutes. The junction depth was determined to be 0.18μm from spreading resistance technique. The lateral diffusion was estimated to be about 0.025μm from SEM pictures.

2.2 Photoresist-ashing technique

Because of the limited resolution of conventional optical lithography, e-beam direct writing and X-ray lithography have been the principal techniques used to fabricate deep-submicrometer devices [2.3,2.4]. However, both techniques are complicated and expensive. In addition, their impact on the long-term device reliability as a result of exposing the device to high-energy radiation has yet to be fully characterized.

In this study, a photoresist-ashing technique has been developed which, when used in conjunction with conventional g-line optical lithography, permits the controlled definition of the gates of deep-submicrometer devices. Although this technique does not help to improve the circuit layout design rules, it does provide an alternative, economical, and efficient means for device-level studies of deep-submicrometer MOSFETs. When this technique is applied to an existing process, it will improve the circuit performance because of the enhanced device current drive due to smaller channel length beyond lithography limits. Since most polymer-based resist materials are ashable with oxygen plasma, this photoresist-ashing technique can also be extended to supplement other lithographic process, such as those of e-beam and X-ray.

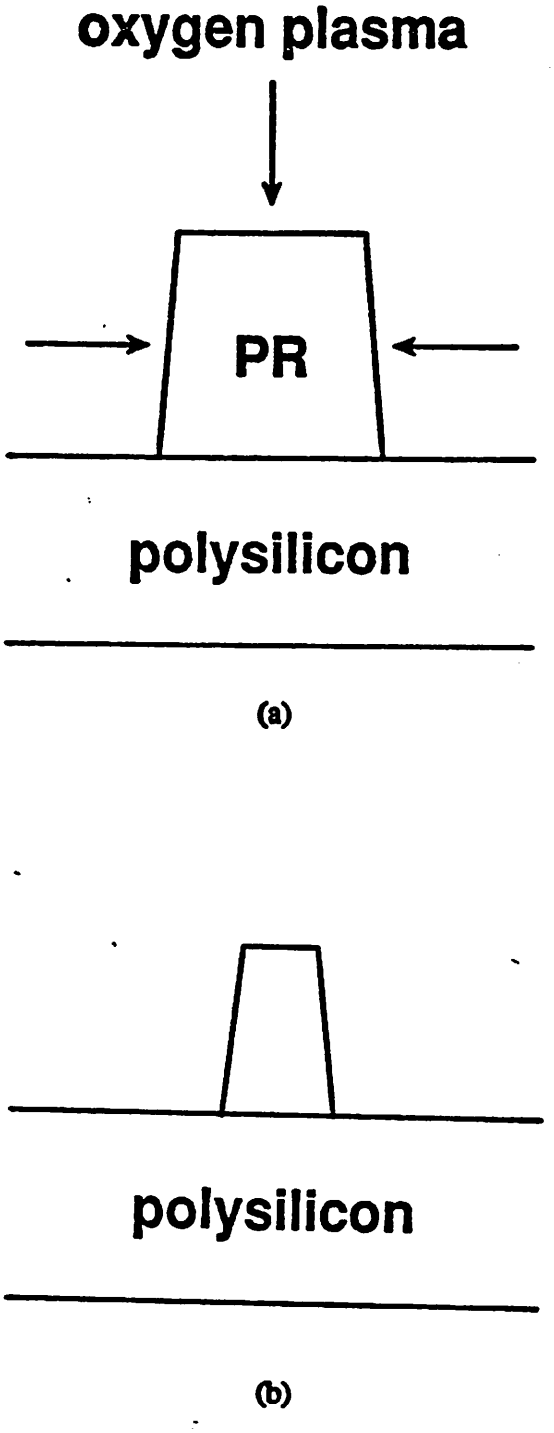


Fig. 2.1 A schematic diagram of the photoresist-ashing process, (a) before ashing, (b) after ashing.

The basic idea of this photoresist-ashing technique is very simple as is illustrated in Fig. 2.1. First, photoresist with near-micron size was defined using conventional optical lithography and developed (Fig. 2.1a). Then the wafers are isotropically etched in oxygen plasma at a calibrated rate until the designated pattern size is achieved (Fig. 2.1b). Since the left and right sides of the photoresist are etched at the same time, the horizontal dimension is etched at twice the rate of the vertical dimension. The photoresist after etching has an ultra-fine pattern but still with enough thickness to define the polysilicon gates.

2.2.1 Wafer preparation

Kodak 820 photoresist was spun at 4600 rpm for 25 seconds and soft-baked for 1 minute at 100 °C, resulting in a photoresist thickness of 1.1 μm before etching. Transistors gates with mask-level lengths ranging from 0.5 to 1.6 μm, with 0.1 μm increment, were defined using GCA 6200 10X wafer stepper (g-line, λ = 436nm), developed, and hard-baked at 120 °C for 15 minutes. Since the resolution of g-line optical lithography is only about 1μm, the photoresist patterns with lengths less than 1μ would not have sharp edges under nominal focus and exposure. To obtain consistent photoresist profiles and step coverage for all mask-level dimensions, a focus-exposure test using specially designed test patterns was performed on GCA 6200 wafer stepper before the wafers were exposed. By examining these photoresist test patterns using various focus-exposure combinations, optimal values were determined. This calibration procedure was the most critical step in the process. Depending on the condition of the light source, the optimal exposure and focus deviated as much as ±20% and ±5%, respectively, from their nominal values.

2.2.2 Etching process

Although this photoresist ashing process could have been done in any oxygen plasma etching system, the Technics-C plasma etcher was used in this study because it has been used in descuming the photoresist in the Micro-Electronics Fabrication Laboratory. The optimal etching condition for this purpose is still unknown; however, it was found that high controlla-

bility and uniformity could be achieved at an oxygen pressure of 300 mTorr and an RF power of 50W. A horizontal etch rate (per side) of 0.035μm/min and vertical etch rate of 0.04μ/min under these etching conditions were observed. The differing etch rates were due to a slight anisotropy of the system.

2.2.3 Experimental results

Fig. 2.2 shows SEM-measured gate length (L_{SEM}) versus ashing time for four different mask-level gate lengths (L_{mask}). The lateral etching rate was calculated from the slopes of these lines and the vertical etching rate was calculated from the photoresist thicknesses before and after etching using an Alpha-Step profiler. These parallel lines indicate that the etching rate was relatively constant during the process and is independent of the initial photoresist size and profile. In preparing these samples, an exposure about 15% under nominal exposure was determined to be the optimal exposure value. This under exposure explains why the photoresist length L_{SEM} is slightly larger than the mask-level length L_{mask} before ashing (ashing time = 0 min) in Fig. 2.2.

Due to the slow etch rate, this ashing process was easily controlled and reproduced. The integrity of the photoresist profile was also preserved throughout the ashing process. Fig. 2.3 displays the effective channel length L_{eff} as a function of L_{mask} for two different ashing times. These parallel lines suggest a consistent photoresist profile for all mask-level channel lengths that is independent of ashing time, which demonstrates that the correct focus and exposure values were used. The effective channel lengths, L_{eff}, were extracted using a capacitance technique [2.5]. Another independent method to derive L_{eff} which measures the resistance of the gate polysilicon lines also confirms the results in Figs. 2.2 and 2.3.

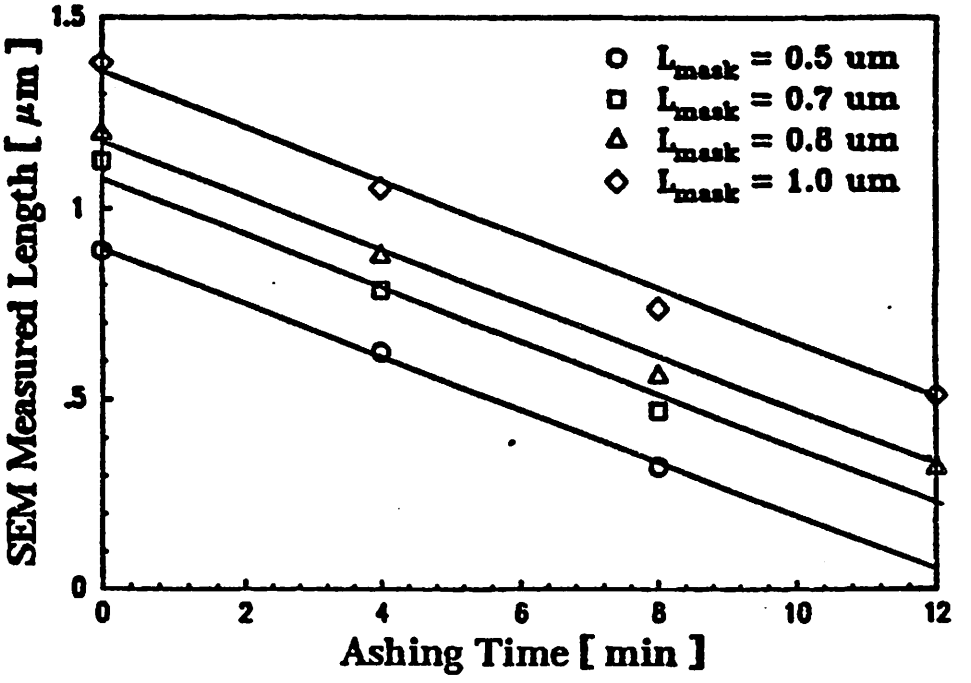


Fig. 2.2 SEM measured gate length versus ashing time for various mask-level channel lengths.

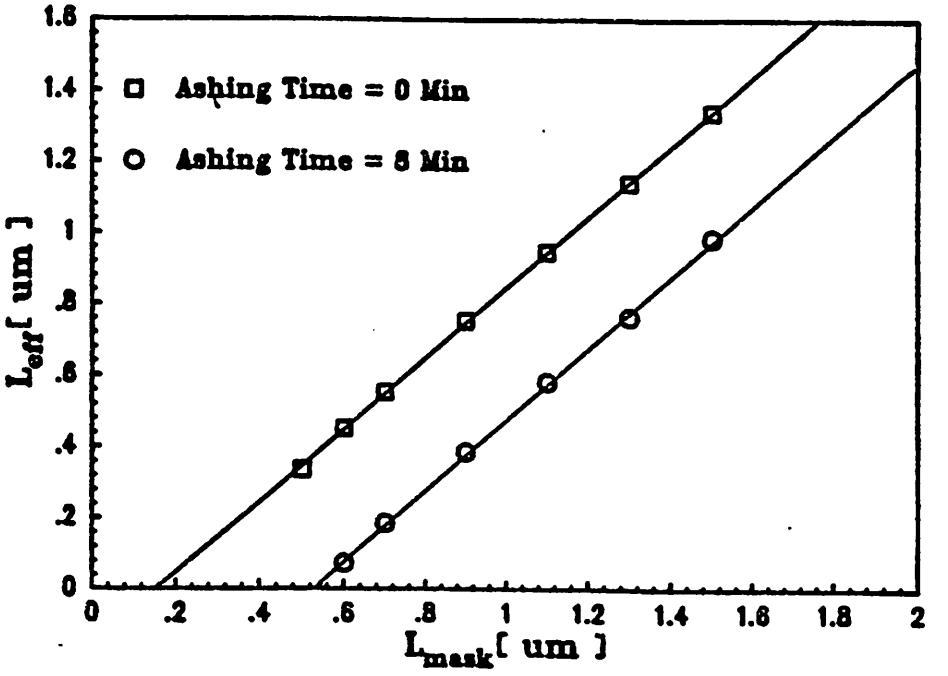


Fig. 2.3 Transistor effective channel length versus mask-level channel length before and after the photoresist-ashing process.

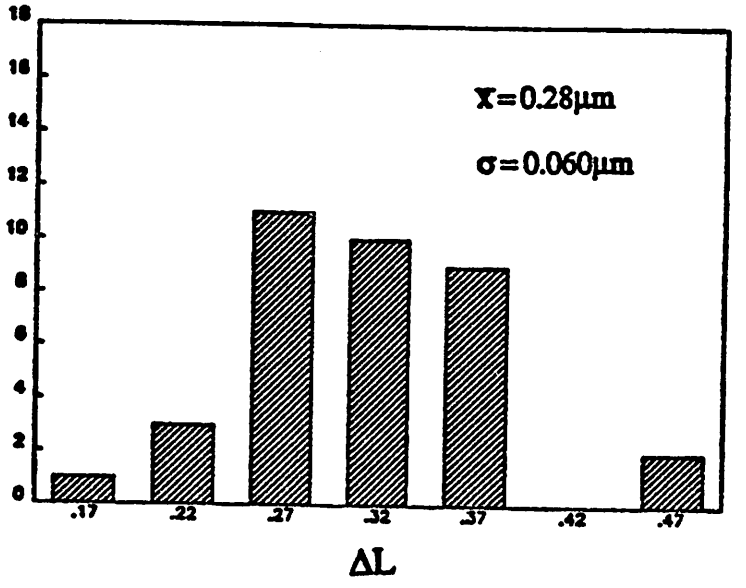
The uniformity of the effective channel length of the transistors across the wafer can be observed in Fig. 2.4 which shows the statistical spread of ΔL ($\equiv L_{mask} - L_{eff}$) of two wafers, one before and one after the ashing process. The standard deviations of ΔL for both cases were roughly the same, 0.06μm, revealing that this photoresist ashing technique did not introduce additional channel length variations to the process. It is believed that the nonuniformity in L_{eff} was inherent to the optical lithography system rather than being introduced by the ashing process.

Fig. 2.5 shows an SEM picture of the cross-section of a photoresist line after 8 minutes of ashing. The line width was originally 1 μm and reduced to 0.45 μm after ashing. Since the effective horizontal etch rate is higher than the vertical etch rate, the aspect ratio of the photoresist profile increases as the ashing process continues until the size of the photoresist reduces to about 0.2μm, which is roughly equal to the difference between the top and base width of the profile. Fig. 2.6 is an SEM picture of a photoresist-covered polysilicon line lying over alternated field and active regions showning the step coverage of the photoresist along the boundary of these two regions.

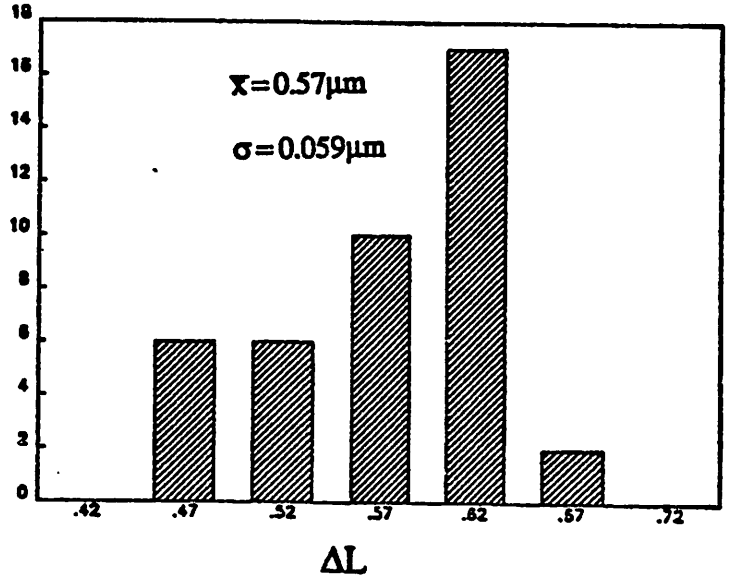
In order to get a 0.65V long-channel threshold voltage, V_{th0}, for all oxide thicknesses, different implant doses were used. Fig. 2.7 shows measured V_{th0} versus implant dose for several oxide thicknesses. The symbols represent measured data and the curves are calculated from the well-known expression for long-channel threshold voltage.

$$V_{th0} = V_{FB} + \phi_S + \frac{\sqrt{2q\epsilon_{si}N_{SUB}(\phi_S - V_{BS})}}{C_{ox}} \qquad (2.1)$$

where V_{FB} has an empirical value of -0.75V, ϕ_S is the surface potential, N_{SUB} is the average channel doping concentration derived from the substrate-bias effect. Fig. 2.8 shows a typical channel doping profile for this process. The depth of the channel implant is about 0.15μm. The experimental relationship between N_{SUB} and the implant dose D is shown in Fig. 2.9,



(a)



(b)

Fig. 2.4 Statistical spread of the effective channel length on a wafer, (a) without ashing, (b) with 8 minutes of ashing.

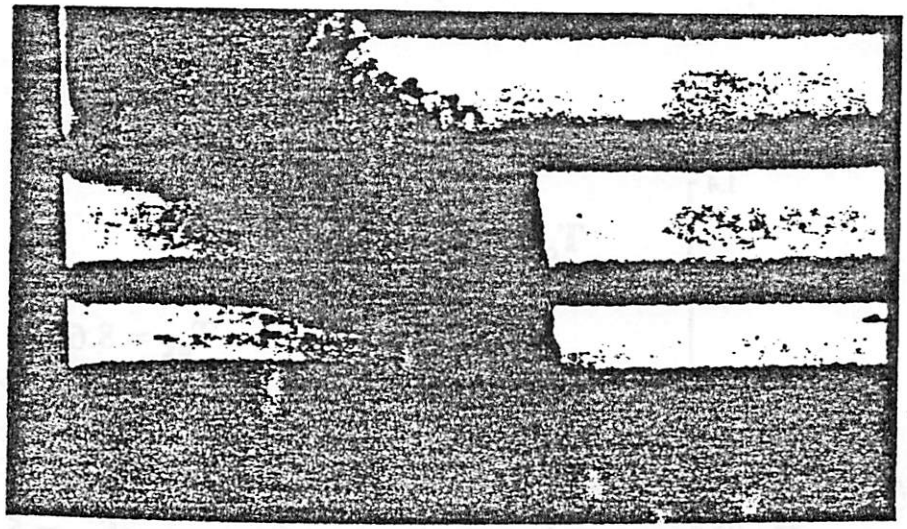


Fig. 2.5 SEM picture of the cross-section of a photoresist line after 8 minutes of ashing.

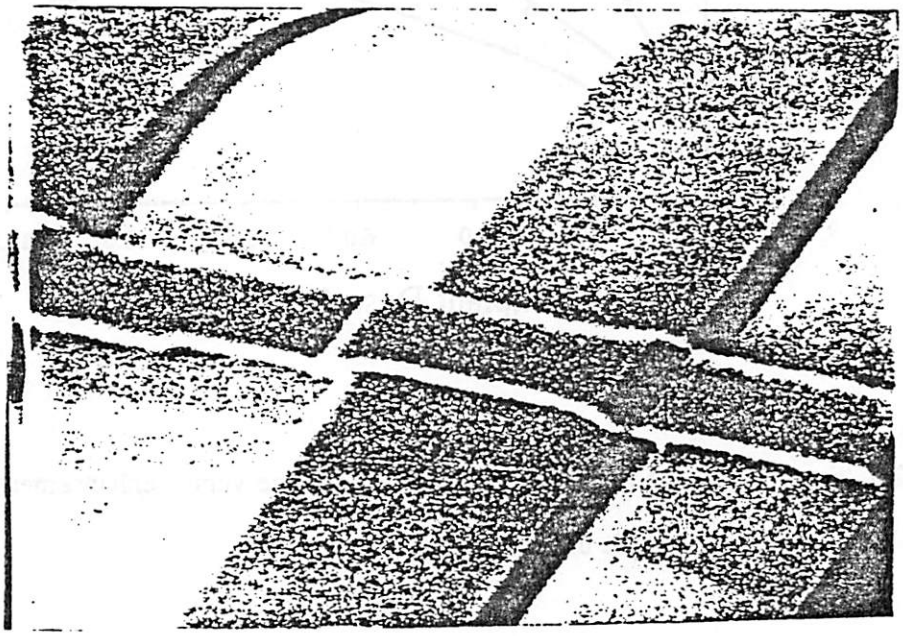


Fig. 2.6 SEM picture of a photoresist-covered polysilicon line lying over alternated field and active regions.

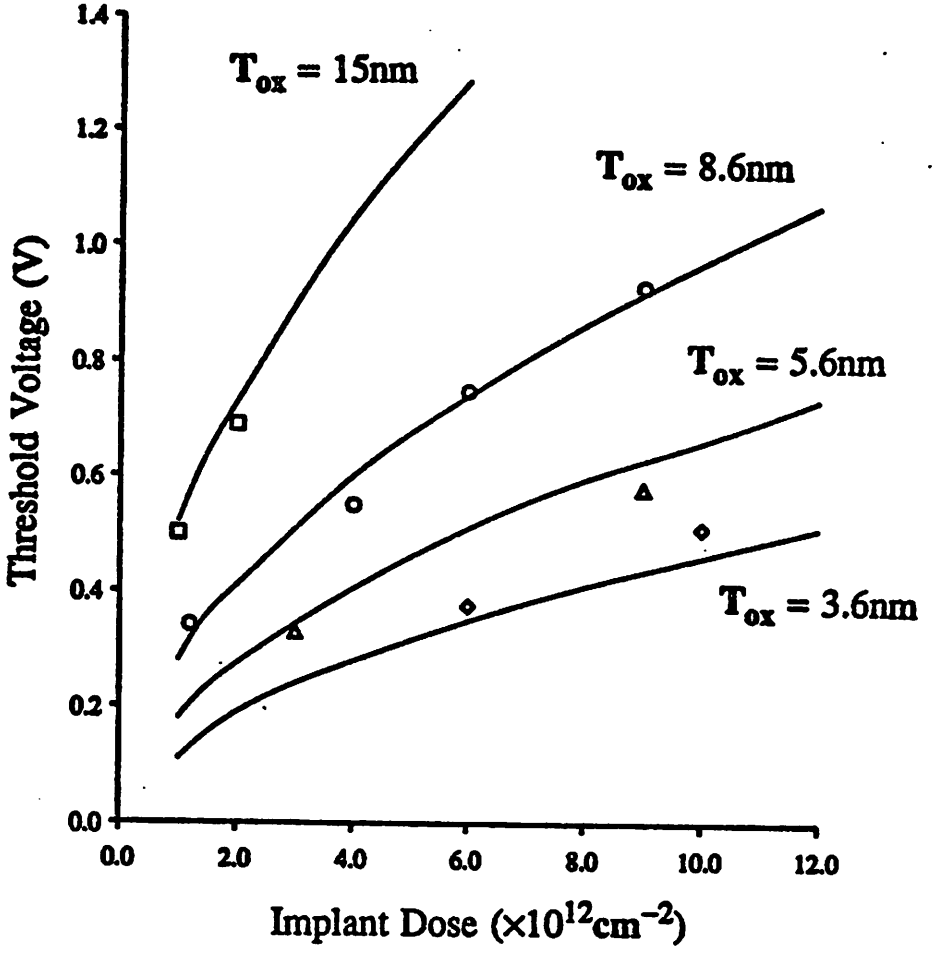


Fig. 2.7 Measured long-channel threshold voltage versus enhancement implant dose for various oxide thicknesses.

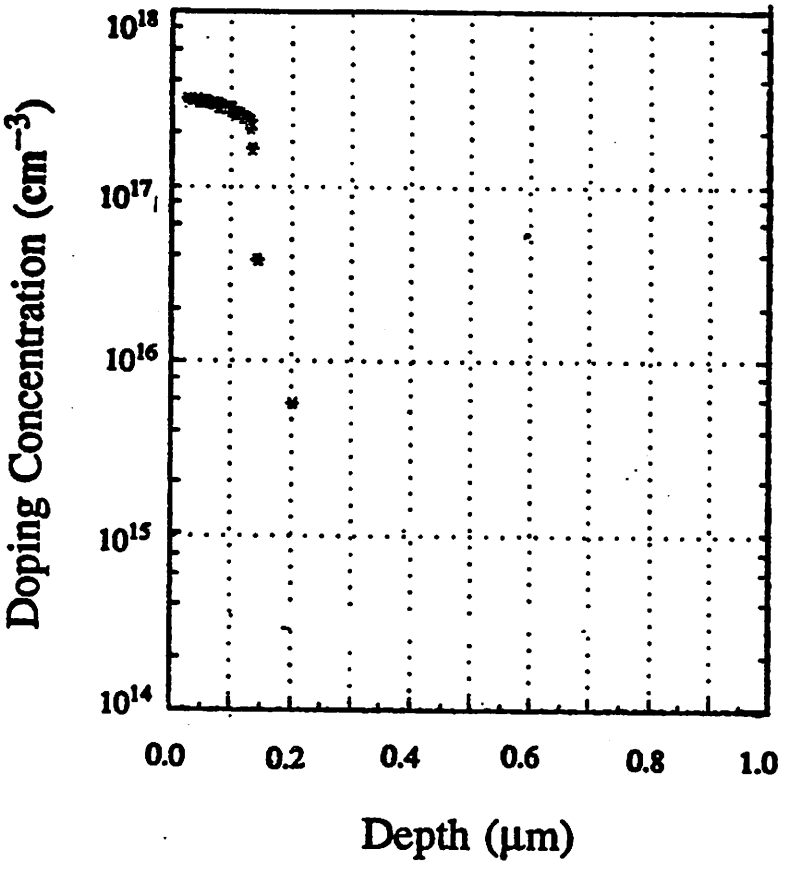


Fig. 2.8 Typical channel doping profile of this process.

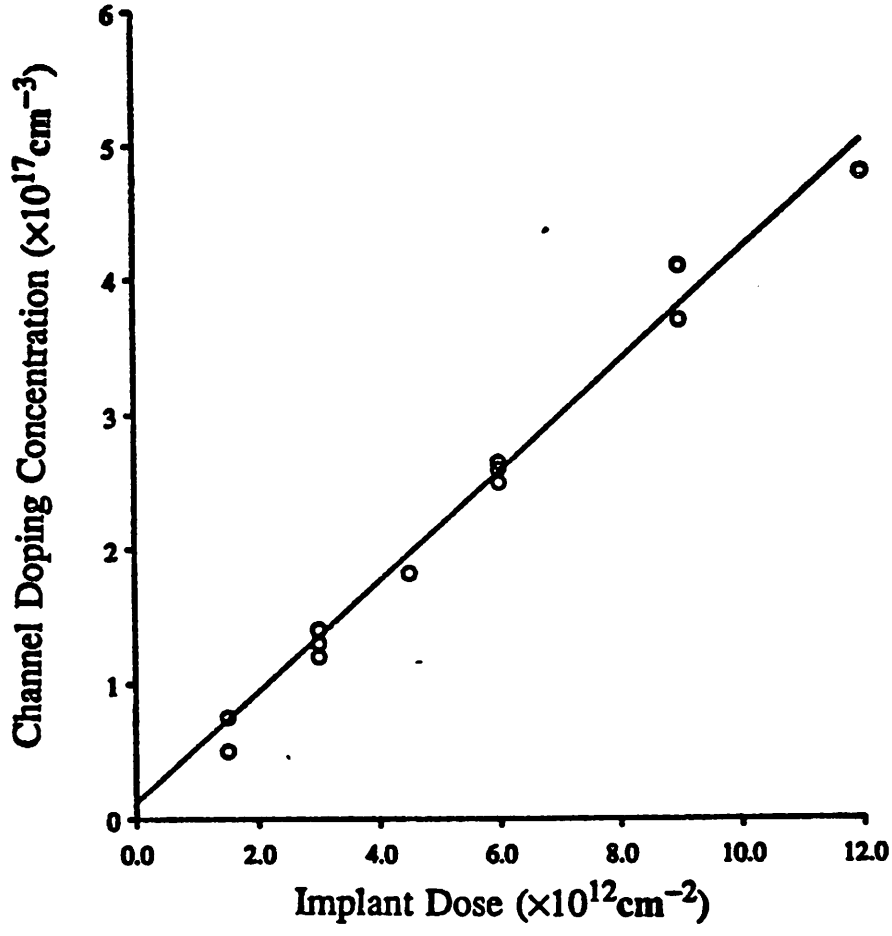


Fig. 2.9 Channel doping concentration versus enhancement implant dose.

where the symbols are measured data and the solid line is an empirical equation given by

$$N_{SUB} = 1.2\times10^{16} + 4.1\times10^{4} * D \qquad (2.2)$$

where D is the implant dose in cm^{-2}. With Figs. 2.7 and 2.9 or equations (2.1) and (2.2), the enhancement implant dose for any oxide thickness and any threshold voltage can be determined using this process.

2.3 Device characteristics

This photoresist-ashing technique has been successfully employed to fabricate n-channel non-LDD MOS transistors with effective channel length as small as 0.15μm. Excellent device characteristics were observed. Fig. 2.10 shows an SEM picture of a transistor cross section with 0.22 μm effective channel length. This transistor would have a 0.8μm effective channel length if the ashing process was not used. The junction depth is about 0.18μm measured from spreading resistance method and the lateral diffusion is about 0.05 μm. The strong-inversion and subthreshold characteristics of a transistor with 3.6nm gate oxide and 0.15μ effective channel length are shown in Fig. 2.11. The transconductance of this device is about 650mS/mm, which is among the highest reported at room temperature. More characteristics are shown in later chapters.

The output waveform of a 101-stage enhancement/depletion-type ring oscillator with one fan-in and one fan-out is shown in Fig. 2.12. This ring oscillator has 7.2nm gate oxide and 0.2μm effective channel length. The delay time is about 22ps/stage at a power supply of 3V which is also one of the fastest ever reported at room temperature for MOS technology.

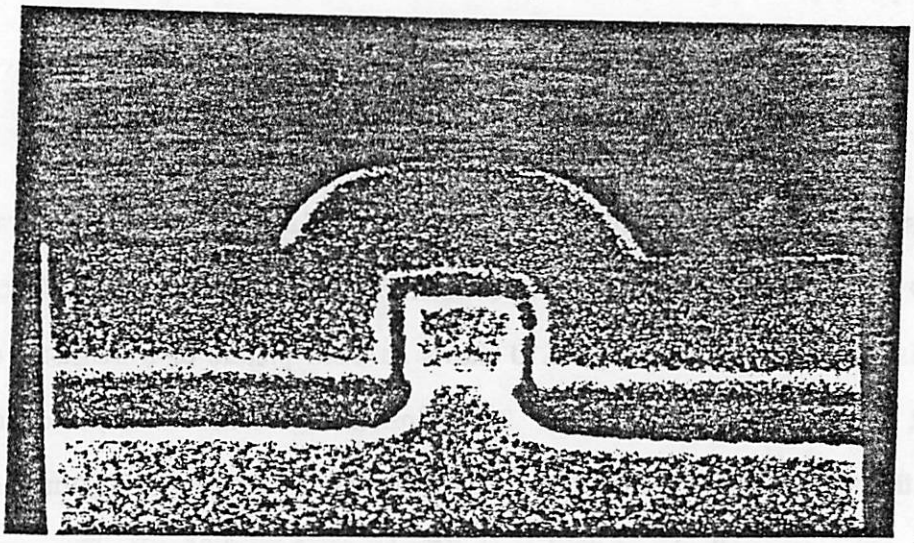
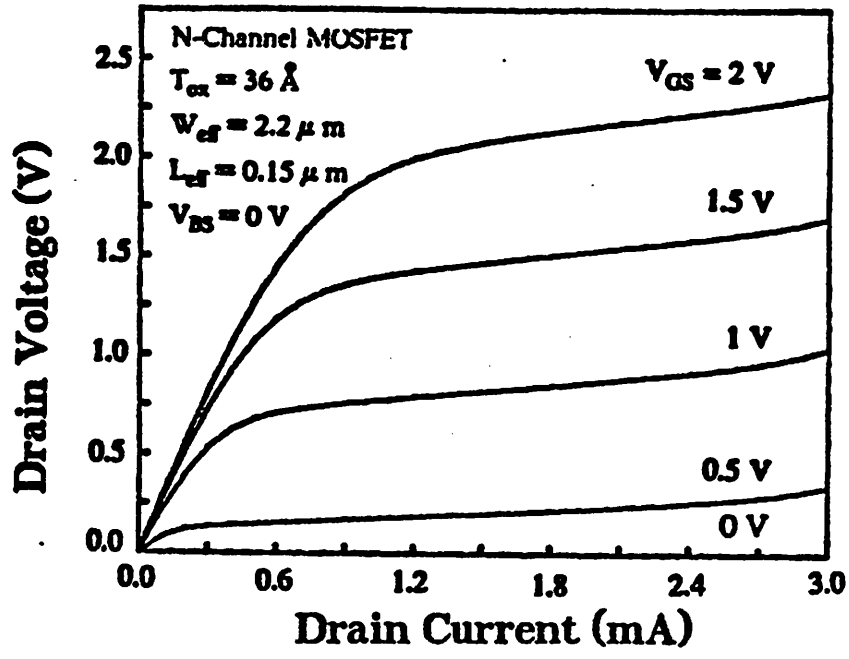
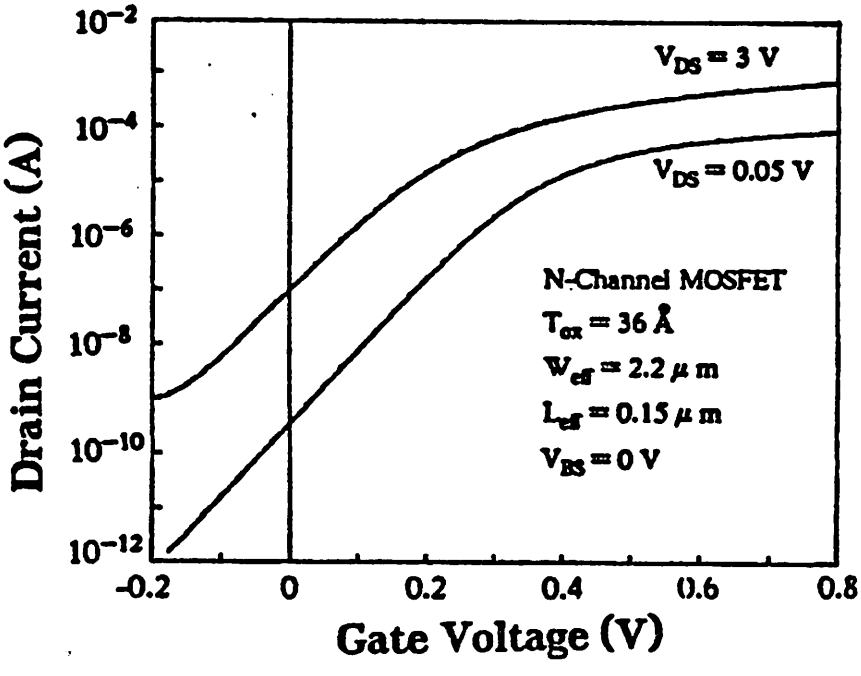


Fig. 2.10 SEM picture of a transistor cross section with 0.22μm effective channel length. The junction depth is 0.18μm and the source/drain lateral diffusion is about 0.05μm.

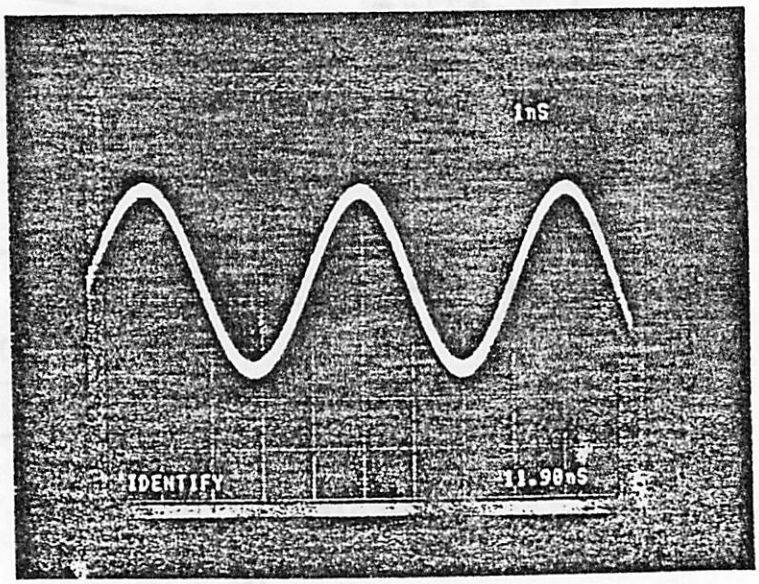


(a)



(b)

Fig. 2.11 Characteristics of a transistor with T_{ox} = 3.6nm and L_{eff} = 0.15μm, (a) strong-inversion, (b) subthreshold.



$T_{ox} = 72$ Å

$W_{eff} = 7\ \mu m$

$L_{eff} = 0.25\ \mu m$

Fan-in = 1

Fan-out = 1

$\tau_{delay} = 21$ ps/stage

Fig. 2.12 Output Waveform of an NMOS 101-stage enhancement/depletion type ring oscillator.

2.4 References

[2.1] J. Chung, M.-C. Jeng, J.E. Moon, A.T. Wu, T.Y. Chan, P.K. Ko, and C. Hu, "Deep-Submicrometer MOS Device Fabrication Using a Photoresist-Ashing Technique," IEEE Electron Device Letters, vol. 9, no. 4, pp.186-188, April, 1988.

[2.2] T.Y. Chan, A.T. Wu, P.K. Ko, C. Hu, and R. Razouk, "Asymmetrical Characteristics in LDD and Minimum-Overlap MOSFET's", IEEE Electron Device Letters, vol. EDL-7, No. 1, pp.16-19, Jan. 1986.

[2.3] W. Fichtner, R.K. Watts, D.B. Fraser, R.L. Johnston, and S.M. Sze, "0.15 μm Channel-Length MOSFETs Fabricated Using E-Beam Lithography," IEDM Tech. Dig., pp.272-725, 1982.

[2.4] S.Y. Chou, H.I. Smith, and D.A. Antoniadis, "Sub-100-nm Channel-Length Transistors Fabricated Using X-Ray Lithography," J. of Vacuum Science Tech. B, vol. 4, no. 1, pp.253-255, Jan/Feb 1986.

[2.5] B.J. Sheu and P.K. Ko, "A Capacitance Method to Determine Channel Lengths for Conventional and LDD MOSFET's," IEEE Electron Device Letters, vol. EDL-5, no. 11, Nov., 1984.

Chapter 3

PERFORMANCE AND HOT-ELECTRON RELIABILITY OF DEEP-SUBMICROMETER MOSFETS

Recent advances in process technology [3.1-3.7] have made deep-submicrometer MOSFETs potential candidates for next generation ULSI designs. However, the emphases of most previous reports have been to demonstrate the feasibility of fabricating these devices with little discussion of the physics. It is still unclear whether the basic physics associated with deep-submicrometer devices is the same as that of their longer-channel counterparts. The lack of physical understanding is one of the reasons preventing deep-submicrometer devices from being used in current VLSI system designs. One of the goals of this study is to establish an improved understanding of deep-submicrometer devices and to provide a basis for device design guidelines. Since performance and hot-electron reliability are the two major concerns in deep-submicrometer device designs, they are carefully studied in this chapter. More device characteristics are presented in the next chapter.

The effective channel length, L_{eff}, is probably the most important parameter among all MOSFET parameters. Since the device characteristics and even some other device parameters, .e.g the threshold voltage, are sensitive functions of the channel length, L_{eff} has been commonly used to identify a technology. In the deep-submicrometer regime, accurate determination of L_{eff} is more crucial, because incorrect determination of L_{eff} may lead to wrong conclusion or interpretation of a physical phenomenon such as velocity overshoot. From a circuit designer's point of view, using incorrect channel lengths in simulation may cause large errors between simulation results and actual circuit performance. Therefore, the first section of this chapter will be devoted to discussions of the various methods for extracting L_{eff} in this study.

Another important subject that should be included in deep-submicrometer study is the source/drain parasitic resistance R_{SD} effect. The voltage drop across R_{SD} effectively reduces

supply voltages and degrades the current driving capability of scaled devices. It was claimed one time that the parasitic resistance posed a limit in MOSFET scaling, but was proven wrong recently. Previous results based on near-micron technologies tend to overestimate the R_{SD} effect in the deep-submicrometer regime. In section 3.4, experimental studies of the parasitic resistance effect on deep-submicrometer device characteristics and circuit performance are presented. This updated results of the R_{SD} effect can help judge the cost/performance factor in MOSFET scaling and also provides some guidelines to technology developments.

3.1 Effective channel length (width) determination

Existing methods for determining MOSFET effective channel lengths can be divided into two categories: the resistance approach [3.8-3.14] and the capacitance approach [3.15-3.17]. The basic theory behind the resistance approach is based on the I_{DS} - V_{DS} relationship. Depending on the extraction procedures, some methods are sensitive to the parasitic resistance variations between devices [3.9] and some are sensitive to the I-V model used [3.10]. The capacitance approach is based on the measurement of the net capacitance under the inversion-layer region. The capacitance approach is more accurate because it is insensitive to R_{SD} and does not require an I-V model. However, the accuracy of capacitance methods diminishes when the gate area is reduced as in small-geometry devices, because the stray capacitance is comparable to the gate capacitance unless high resolution instruments are used. Most of these methods have been demonstrated on devices with channel length longer than 1 μm, but no study has been reported about their validity in the deep-submicrometer regime. In this section, two resistance methods and one capacitance method are examined and compared.

(A) Channel-resistance method

The channel-resistance method [3.8,3.9] is the most commonly used method in determining L_{eff} because of its simplicity and its ability to separate R_{SD} from the intrinsic channel resistance. The principle of this method is briefly described below. When an MOSFET is biased

in the linear region with a small drain voltage (e.g. 0.1V), the intrinsic channel resistance, R_{ch}, is given by

$$R_{ch}=\frac{L_{eff}}{\mu C_{ox} W_{eff}(V_{GS}-V_{th})} \tag{3.1}$$

where μ is the carrier mobility, $W_{eff} = W_{drawn} - \Delta W$, and $L_{eff} = L_{drawn} - \Delta L$.

The measured device resistance R_{meas} is equal to

$$R_{meas}=R_{SD}+R_{ch} \tag{3.2}$$

Therefore, plotting R_{meas} against L_{drawn} for a set of transistors with the same W_{eff} and same $V_{GS}-V_{th}$ results in a straight line, assuming μ is not channel length dependent. The slope of the line is inversely proportional to $V_{GS}-V_{th}$. All the lines with different $V_{GS}-V_{th}$ values will intersect at the same point as shown in Fig. 3.1. The x- and y-coordinate of the intersection give ΔL and R_{SD}, respectively. The accuracy of this method relies on the assumption that R_{SD} is the same for all devices with the same channel width. In reality, R_{SD} values may vary slightly between devices either due to process non-uniformity or introduced by contacting probes during measurement, but this assumption is still good as long as the R_{SD} variation is small compared to R_{ch}. Therefore, when applying this method to the deep-submicrometer regime, special care should be taken in probing devices (on-wafer measurements) or using devices with small channel widths. Running the device under high current levels for a couple of seconds before taking data usually can minimize R_{SD} variations. Since R_{ch} is a function of $V_{GS}-V_{th}$, the linearity of these straight lines, which determines the quality of the intersection point (how close these lines intersect), is also highly dependent on the same $V_{GS}-V_{th}$ value for every device. To minimize the effect of the uncertainty in V_{th} between devices, the minimum applied gate voltage should be 0.5V to 1V higher than V_{th}. When all these considerations are taken cares of, this method can be extended to extract L_{eff} down to 0.2μm or smaller as illustrated in Fig. 3.1.

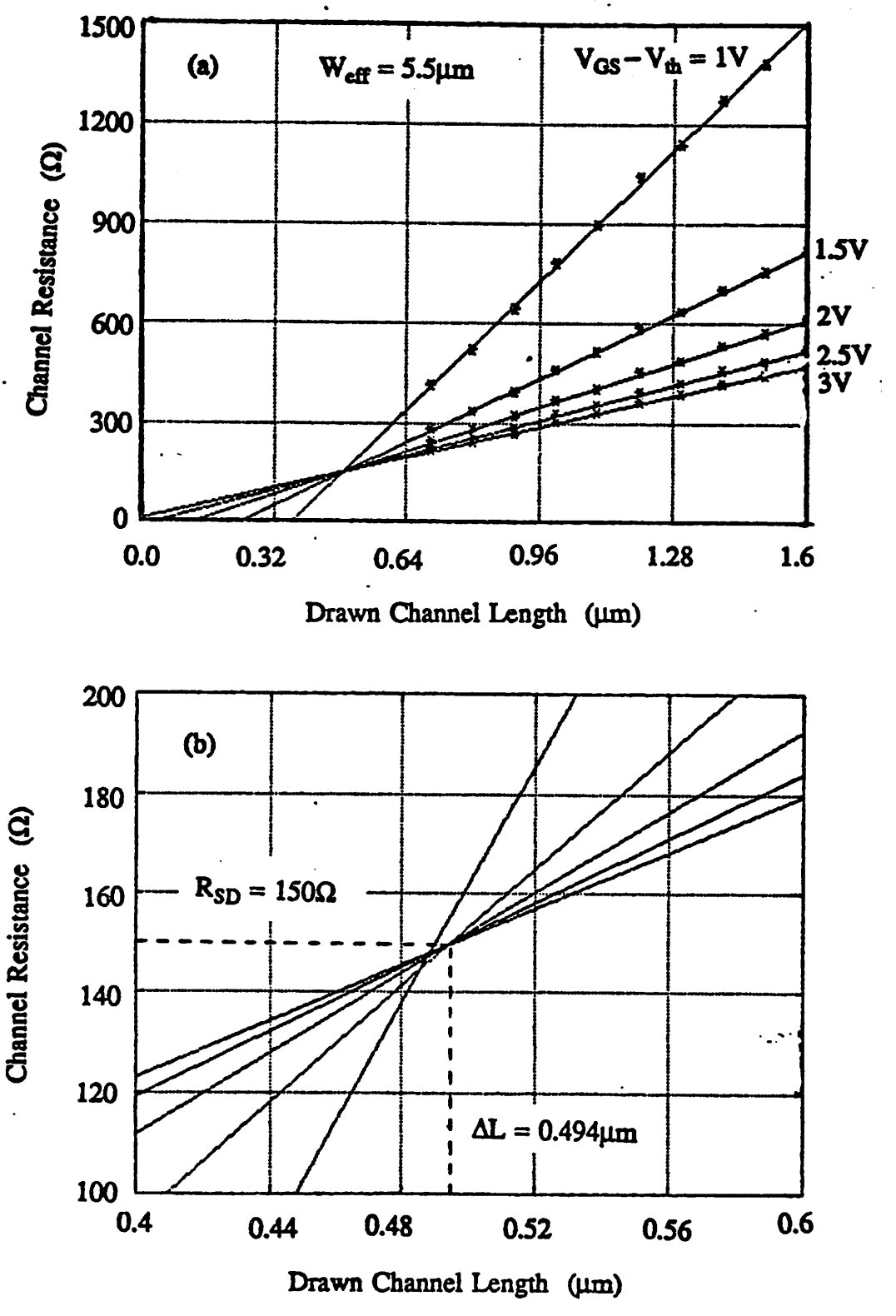


Fig. 3.1 (a) Measured channel resistance versus drawn channel length for various gate voltages, (b) enlarged area near the intersection.

As for the channel width determination, a similar channel-resistance approach does not work well as was pointed out by Ma [3.18], because R_{SD} varies with the channel width. However, it should be noted that the reciprocal of the slopes Gi (= $\mu C_{ox} W_{eff}(V_{GS}-V_{th})$) of the lines in Fig. 3.1 are linearly dependent on the channel width for a given V_{GS} - V_{th}. Therefore, ΔW can still be extracted by plotting Gi as a function of W_{drawn} as illustrated in Fig. 3.2. Each line in Fig. 3.2 corresponds to a particular $V_{GS}-V_{th}$ value and the x-intercept gives ΔW. Because of the bird's beak at the edge of the active region, the effective channel width is in general a function of the gate voltage. This result is reflect by the different intercepts for different $V_{GS}-V_{th}$ values in Fig. 3.2. Since ΔW for each V_{GS} can be obtained, the gate-voltage dependence of ΔW can also be obtained. The functional form of ΔW depends on the isolation technology used. The insert in Fig. 3.2 shows the extracted ΔW - $V_{GS}-V_{th}$ result for a LOCOS process.

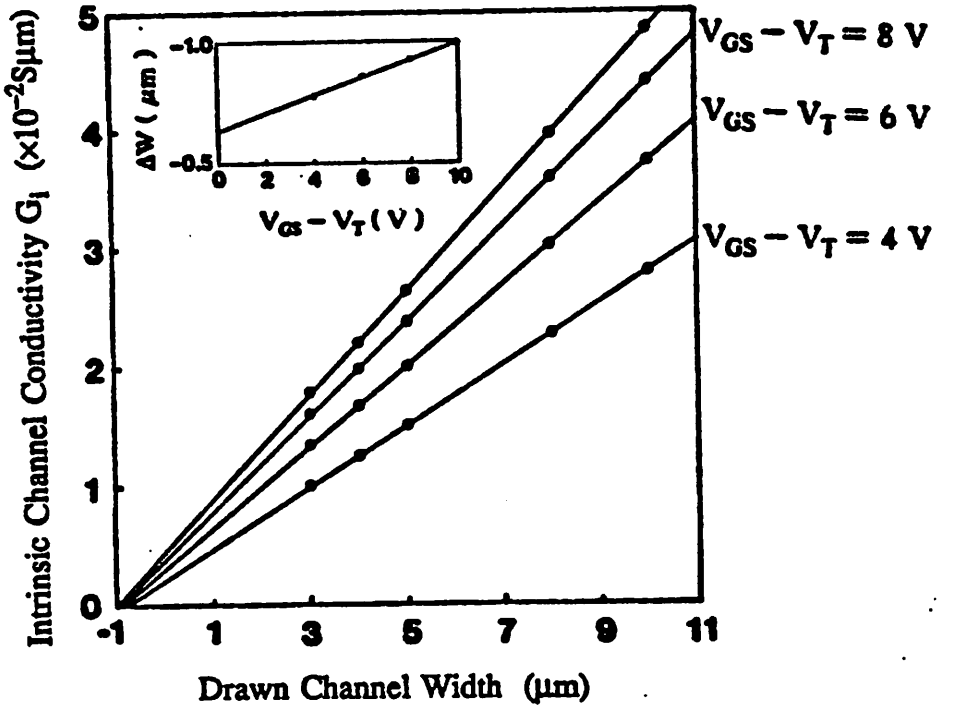


Fig. 3.2 Intrinsic channel conductivity (the reciprocal of the slopes in Fig. 3.1) versus drawn channel width. The insert shows the extracted ΔW as a function of $V_{GS}-V_{th}$.

(B) Modified Suciu's method

Unlike the channel-resistance method, this method is insensitive to the parasitic resistance of individual device, but an I-V model is required. The accuracy of this method is affected by the I-V model used. Since the mobility model used in the original method proposed by Suciu [3.10] is too crude to be applied to deep-submicrometer devices that typically have thin gate oxides, an improved I-V model is used. The basic theory of this modified method is described below.

The drain current of a MOSFET with R_{SD} effect included can be expressed as

$$I_{DS}=\frac{\beta_0(V'_{GS}-V_{th})V'_{DS}}{1+U_a(V'_{GS}-V_{th})+U_b(V'_{GS}-V_{th})^2} \tag{3.3}$$

where

$$\beta_0=\frac{\mu_0 C_{ox}(W_{drawn}-\Delta W)}{L_{drawn}-\Delta L} \tag{3.4}$$

$$V'_{GS}=V_{GS}-I_{DS}R_S \tag{3.5}$$

$$V'_{DS}=V_{DS}-I_{DS}R_{SD} \tag{3.6}$$

R_S is the parasitic resistance on the source side, U_a and U_b are coefficients of the mobility reduction due to vertical field. Note that the U_b term in (3.3) is the modified mobility term. The meanings of U_a and U_b are explained in chapter 5. When V_{DS} is small, Eq. (3.3) can be simplified and re-arranged as

$$R_{meas}V_{Gt}=\frac{1}{\beta_0}+(R_{SD}+\frac{U_a}{\beta_0})V_{Gt}+\frac{U_b}{\beta_0}V_{Gt}^2 \tag{3.7}$$

where $V_{Gt} = V_{GS} - V_{th}$. Since (3.7) has the form of "$y = a + b\,x + c\,x^{2}$", the coefficients β_0, $R_{SD}+U_a/\beta_0$, and U_b for each device can be extracted by fitting (3.7) through a least-square fit routine. This fitting procedure is similar to that shown in section 5.4.3. Since β_0 is proportional to W_{eff}/L_{eff}, ΔL can be obtained from the x-intercept in the plot of $1/\beta_0$ versus L_{drawn} for fixed channel width as shown in Fig. 3.3. Similarly, ΔW can be obtained from β_0 versus

W_{drawn} plots, but ΔW's extracted from this method represents an "averaged" value that does not show any gate-voltage dependence. When all β_0s are extracted, R_{SD} and U_a can be obtained by plotting $R_{SD}+U_a/\beta_0$ versus $1/\beta_0$. The devices used in Fig. 3.3 are identical to those used in Fig. 3.1. The ΔL's derived from both methods are very similar.

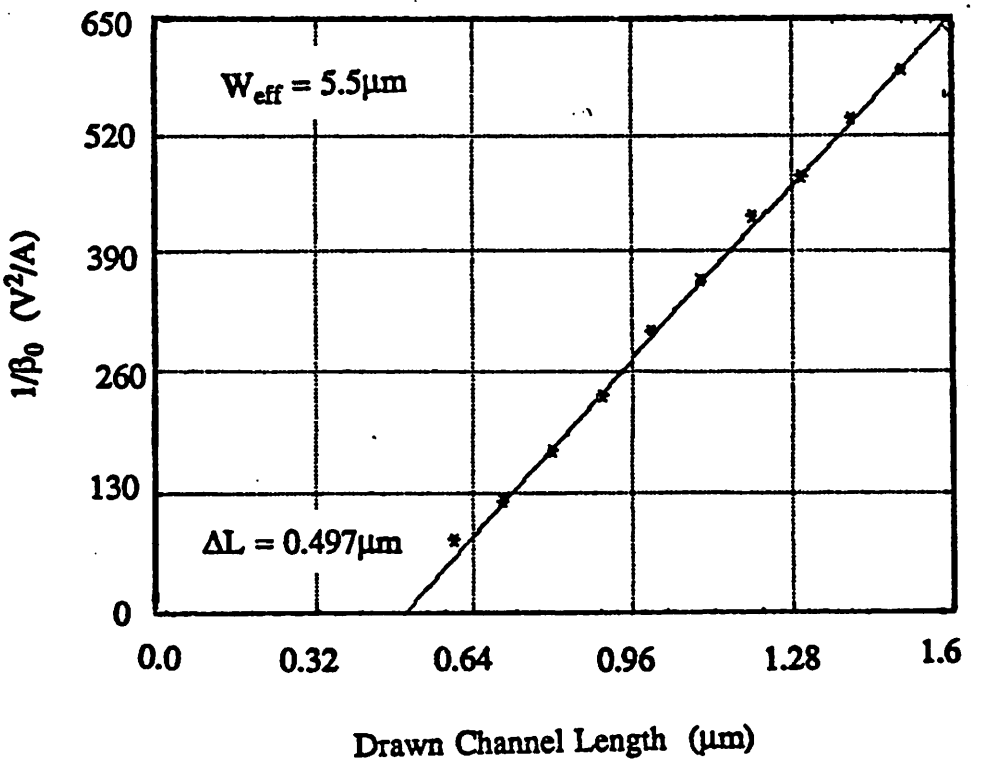
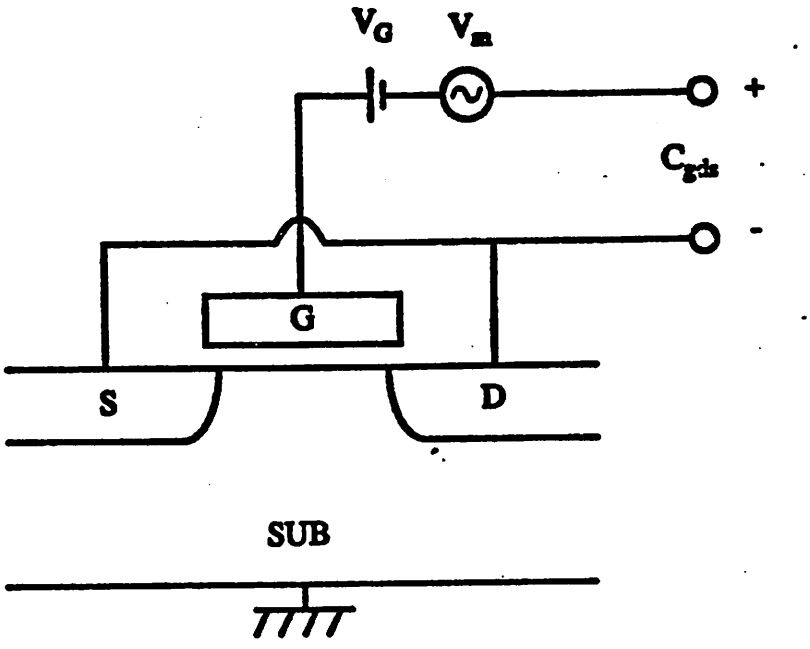


Fig. 3.3 Reciprocal of the channel conductance versus drawn channel length.

(C) Capacitance method

For longer channel devices, the capacitance method is the most accurate method among all methods, but the measurement instruments required are not widely available in typical automated characterization systems. Furthermore, the parasitic resistance can not be extracted. Therefore, the capacitance method is a good means to justify the accuracy of a resistance method and is often used when high accuracy in L_{eff} is required. In this study, the capacitance method is also used to confirm the results of the two resistance methods. A schematic diagram of the capacitance method is shown in Fig. 3.4. The device is first biased in the accumulation region and the source-drain to gate capacitance, C_{gds}, is measured. This capacitance (indicated by C_1 in Fig. 3.4) is composed of the overlap capacitance and any stray capacitance of the system. Then, the device is driven into the strong-inversion region and C_{gds} is measured again (indicated by C_2). C_2 is larger than C_1 by $C_{ox}(W_{drawn}-\Delta W)(L_{drawn}-\Delta L)$. By plotting the difference between C_1 and C_2 against L_{drawn} or W_{drawn}, ΔL and ΔW can be extracted from the x-intercept as shown in Fig. 3.5, where the same devices in Fig. 3.1 and 3.3 are used again.

Generally speaking, the resistance methods require simpler equipment and work better for narrower channel width and thicker gate oxide devices and the capacitance method works better for wider channel width and thinner gate oxide devices. These two types of methods serve as complementary to each other. If care is taken, all three methods discussed here can be applied to the deep-submicrometer regime and the extracted ΔL's agree within 0.01μm. In this study, most of the effective channel lengths were simultaneously determined by the two resistance methods and were frequently checked by the capacitance method to ensure high accuracy and high confidence.



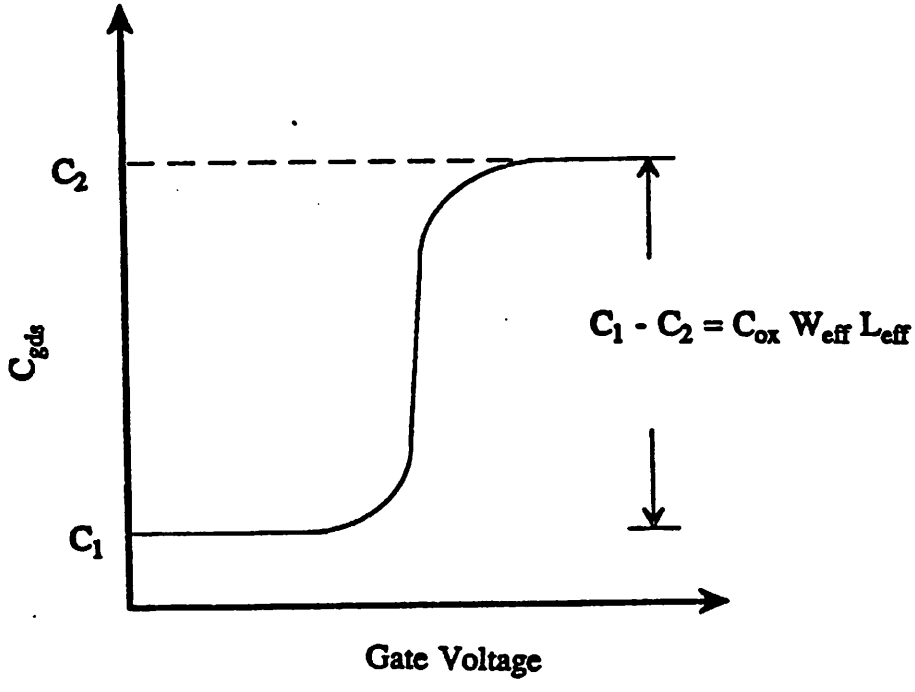


Fig. 3.4 (a) A schematic diagram of the measurement setup for the capacitance method, (b) measured C_{gds} versus gate voltage.

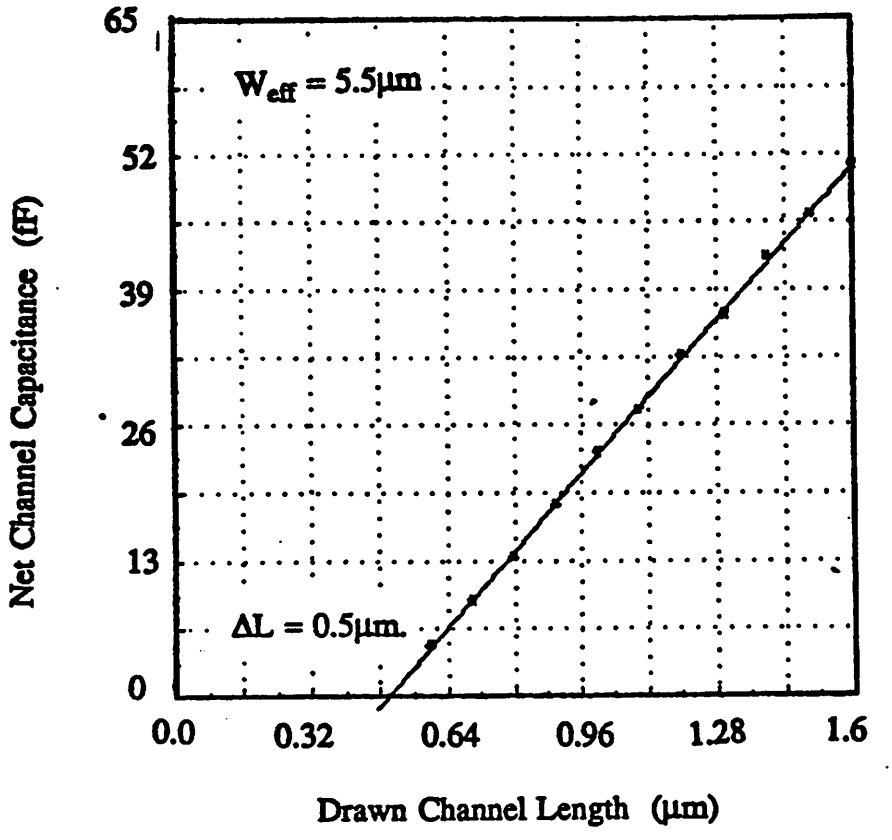


Fig. 3.5 Measured net channel capacitance versus drawn channel length.

3.2 Short-channel effect

The short-channel effect is one of the major concerns in MOSFET scaling, because device parameter variations caused by the short-channel effect poses difficulty in both process control and circuit design. In this section, two most short-channel-sensitive parameters, threshold voltage and subthreshold swing, are examined.

3.2.1 Threshold voltage

Threshold voltage shift ΔV_{th} due to source/drain charge sharing and drain-induced-barrier lowering (DIBL) is the most commonly observed short-channel effect in MOSFETs and has been widely used as an indicator for measuring the extent of the short-channel effect for a

given technology. Therefore, a comprehensive study of the short-channel effect on threshold voltage is a necessary step to optimal device designs. However, a complete characterization of the threshold voltage over a wide range of technologies requires a huge amount of devices and measurements. Various threshold voltage models were developed to supplement this study in predicting future technologies and are used in circuit simulations.

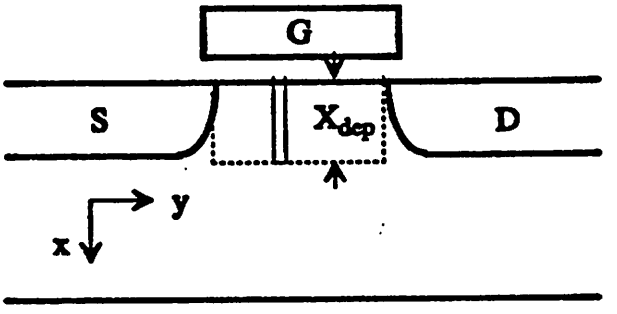
Three approaches have been generally adopted in modeling the short-channel threshold voltage, namely, the charge partitioning [3.19,3.20], the numerical analysis [3.21], and the 2-D analytical approach [3.22-3.25]. Recently, threshold voltage models derived from 2-D analytical solutions of Poisson's equation in the depletion region have become more favorable since ΔV_{th} expressions obtained from this approach show an exponential dependence on L_{eff} which agrees better with experimental results. However, because of the different approximations used for the boundary conditions in deriving the models, the model parameter values vary from paper to paper. Usually parameter values can only be obtained from characterization of physical devices. Furthermore, these simple exponential V_{th} models fail to explain the accelerated V_{th} reduction at shorter channel lengths and tend to underestimate the short-channel effect.

In this section, a short-channel threshold voltage model is derived based on a quasi two-dimensional approach, which has been successfully applied to model the MOSFET substrate current and other hot-electron phenomena [3.26-3.28]. When the device channel length is much longer than the characteristic length (defined later), this model agrees in functional form with those in [3.22-3.25]. At shorter channel lengths, this model predicts a faster increase in ΔV_{th} and are more accurate than other models.

(A) the model

Applying Gauss's law to a rectangular box of height X_{dep} and length Δy in the depletion region as illustrated in Fig. 3.6a, Eq. (3.8) can be derived [3.26-3.28].

$$\frac{X_{dep}}{\eta}\frac{dE_s(y)}{dy}+\frac{\varepsilon_{ox}}{\varepsilon_{si}}\frac{V_{GS}-V_{FB}-V_s(y)}{T_{ox}}=\frac{q}{\varepsilon_{si}}N_{SUB}X_{dep} \tag{3.8}$$



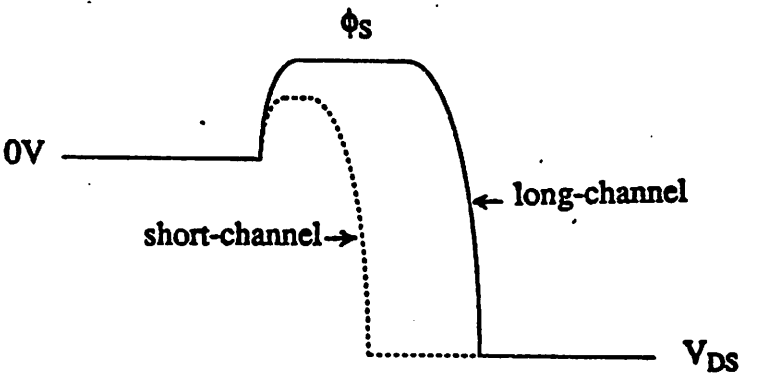


Fig. 3.6 (a) A MOSFET cross section showing the depletion region and the Gaussian box. The depletion region is assumed to be uniform across the channel. (b) The energy diagram of the surface potential from the source to the drain for both a long-channel and a short-channel devices.

where X_{dep} is the depletion layer thickness equal to

$$X_{dep}=\sqrt{\frac{2\varepsilon_{si}(\phi_S-V_{BS})}{qN_{SUB}}} \tag{3.9}$$

ϕ_S is the surface potential at which the threshold voltage is defined, and $E_s(y)$, $V_s(y)$ are the lateral electric field and the channel potential at the Si-SiO_2 interface, respectively. The first term on the left-hand side of (3.8) is equal to the lateral electric field in the channel, $E_y(x,y)$, integrated over the vertical side of the box (see Fig. 3.6a). The non-uniform distribution of $E_y(x,y)$ along the x-direction is taken into account by a fitting parameter η [3.27]. The second term is equal to the vertical electric field at position y on the top side of the box and the term on the right-hand side of (3.8) is equal to the depletion charge density in the box. The solution to (3.8) under the boundary conditions: $V_s(0) = V_{bi}$ at the source and $V_s(L) = V_{DS}+V_{bi}$ at the drain, is given by

$$V_s(y)=D+(V_{bi}+V_{DS}-D)\frac{\sinh(y/l)}{\sinh(L/l)}+(V_{bi}-D)\frac{\sinh(L-y/l)}{\sinh(L/l)} \tag{3.10}$$

where $D = V_{GS}-V_{th0}+\phi_S$, V_{th0} is the textbook long-channel threshold voltage given by $V_{FB}+\phi_S+qN_{SUB}X_{dep}/C_{ox}$, V_{bi} is the built-in potential between the source/drain junction and the substrate, and l is the characteristic length defined as

$$l=\sqrt{\frac{\varepsilon_{si}T_{ox}X_{dep}}{\varepsilon_{ox}\eta}} \tag{3.11}$$

η has an empirical value between 0.5 and 1.5 depending on the device structure and process technology. For most technologies, $L >> l$ and (3.10) can be approximated as

$$V_s(y)=D+(V_{bi}+V_{DS}-D)e^{-L/l}e^{y/l}+(V_{bi}-D)e^{-y/l} \tag{3.12}$$

The channel potential $V_s(y)$ has a minimum at

$$y_{min}=0.5l\ \ln[\frac{(V_{bi}+V_{DS}-D)}{(V_{bi}-D)}e^{L/l}] \tag{3.13}$$

and the minimum channel potential $V_{s,min}$ is given by

$$V_{s,min} = D + 2\sqrt{(V_{bi} - V_{DS} - D)(V_{bi} - D)}e^{-L/l} \tag{3.14}$$

The channel potentials for a long- and a short-channel device for a given gate and drain voltages are plotted in Fig. 3.6b. As the channel length is reduced, the minimum channel potential decreases as shown in Fig. 3.6b. When the minimum channel potential is equal to ϕ_S, the corresponding gate voltage, which is defined as the threshold voltage $V_{th}(L)$ can be calculated.

$$\begin{aligned}\Delta V_{th}(L) &= V_{th0} - V_{th}(L) \\ &= (4R + 2V_{DS})e^{-L/l} + 2\sqrt{V_{DS}^2 e^{-2L/l} + R(R + V_{DS})e^{-L/l}}\end{aligned} \tag{3.15}$$

where $R = V_{bi} - \phi_S$. For $V_{DS} << R$, Eq. (3.15) reduces to

$$\Delta V_{th}(L) = (V_{bi} - \phi_S)(2e^{-L/2l} + 4e^{-L/l}) \tag{3.16}$$

When $L > 5l$, the second exponential term in (3.16) can be neglected and (3.16) reduces to a form similar to those given in [3.22,3.23]. For very short channel lengths, the accelerated V_{th} reduction can be explained by the second exponential term in (3.16).

(B) experimental results

Fig. 3.7 shows threshold voltage versus effective channel length at several drain and substrate biases. The symbols are measured data and the curves are the model. In general, devices with thicker gate oxide and high substrate bias exhibit more threshold reduction due to larger characteristic length l according to (3.11). The accelerated V_{th} reduction phenomenon at shorter channel length can be observed in Fig. 3.8 where ΔV_{th} is plotted against L_{eff} in logarithmic scale. Note that the measured data deviates from the simple exponential expression when L_{eff} is smaller than about $5l$ which translates to $\Delta V_{th} \approx 0.1V$. Since most ΔV_{th} data are taken around 0.1V, this slope-increasing behavior at shorter channel length is important in accurately modeling V_{th}. Without taking into account this accelerated V_{th} reduction at shorter channel length, it would lead to an underestimation of the short-channel effect or result in incorrect extraction of l.

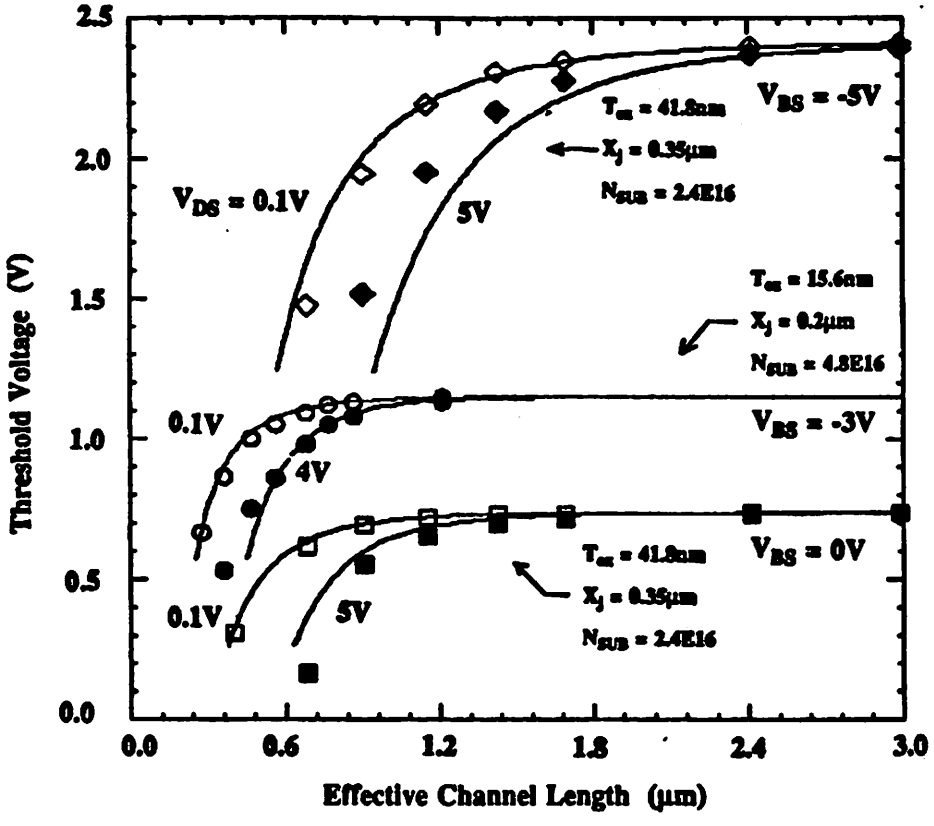


Fig. 3.7 Threshold voltage versus effective channel length at various drain and substrate biases for several technologies. The symbols are measured data and the curves are the model.

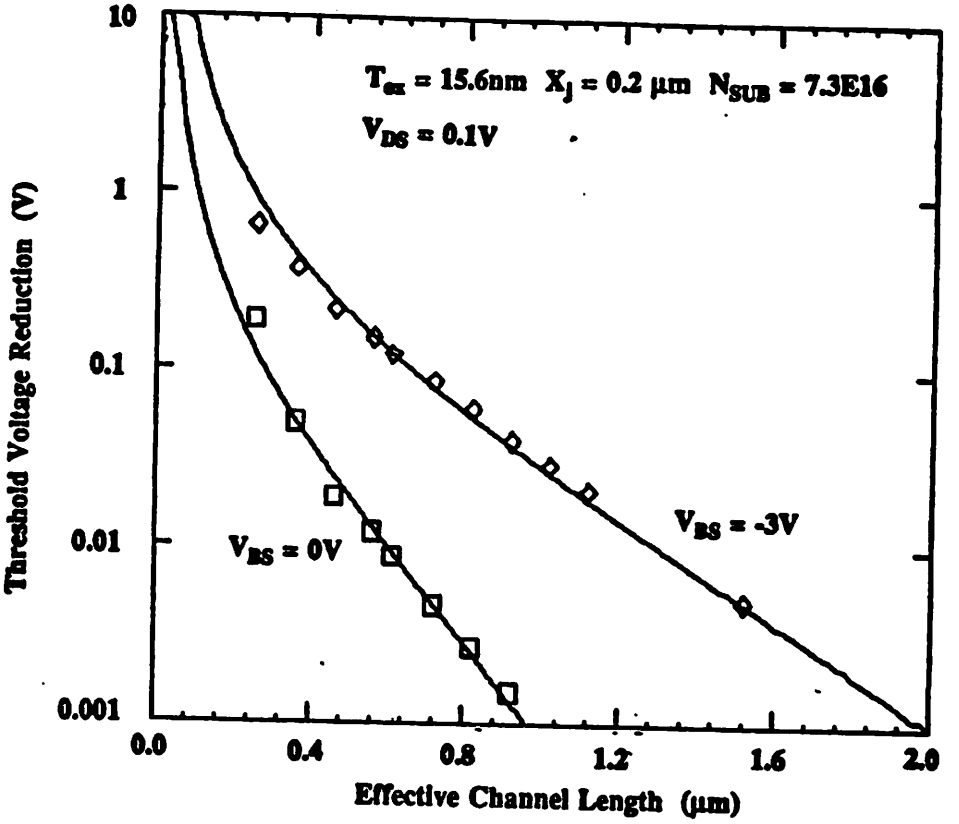


Fig. 3.8 Threshold voltage reduction (ΔV_{th}) versus effective channel length in logarithmic scale. At short channel lengths ΔV_{th} deviates from the simple exponential function as seen by the superlinear behavior. The symbols are measured data and the curves are the model.

Although l calculated from (3.11) has the correct order of magnitude, exact values of l need to be characterized from physical devices because of the unknown parameter η. Extracted l versus the depletion layer thickness for several different technologies are shown in Fig. 3.9. The different X_{dep}'s in Fig. 3.9 for a given technology corresponding to different substrate biases. These straight lines with similar slopes suggests that l is proportional to $X_{dep}^{2/3}$ relatively independent of technologies.

Both this model and those from 2-D analysis indicate that the source/drain charge-sharing and the DIBL effects are basically caused by the same mechanism, namely the channel potential lowering, however, these two effects are usually distinguished for easy explanation. The source/drain charge-sharing usually refers to the ΔV_{th} measured at low drain voltage while DIBL refers to the ΔV_{th} induced by the drain voltage only. Fig. 3.10 plots V_{th} versus V_{DS} for various channel lengths to show the DIBL effect. In other models the DIBL effect is usually approximated by a linear function of V_{DS}, but the linear model fails to explain the faster V_{th} reduction at low V_{DS} as shown in Fig. 3.10. This phenomenon was also observed in [3.23-3.25,3.29]. But this non-linear V_{th} - V_{DS} behavior can still be predicted by this model as shown by the solid curves in Fig. 3.10. At large V_{DS}, this model approaches to a linear function of V_{DS}, while at low V_{DS} it approximately reduces to a square-root function according to (3.16).

Masuda [3.29] found empirically that, for longer channel lengths, the measured V_{th} - V_{DS} curves all intercept at the same point, but not for shorter channel lengths. This observation can also be qualitatively explained by this model if we draw straight lines to best fit the curves in Fig. 3.10 as illustrated by the dashed lines. For longer channel lengths, the V_{th} - V_{DS} curves can be well approximated by straight lines and the x-intercept of these asymptotes can be derived from (3.15) But for for smaller channel lengths, a large portion of the V_{th} - V_{DS} curves are not straight. Therefore, trying to draw lines to best fit the curves (or data) for short channel lengths would result in lines steeper than their asymptotes.

The same approach can also be applied to LDD structures, but the boundary conditions should be modified to be suitable for the n^-/p junctions. Since the built-in potential of an n^-/p junction is smaller than that of an n^+/p, LDD devices generally show less V_{th} shift than non-LDD devices. The voltage drop across the n^- region also decreases the effective drain voltage applied to the channel and reduces the DIBL effect.

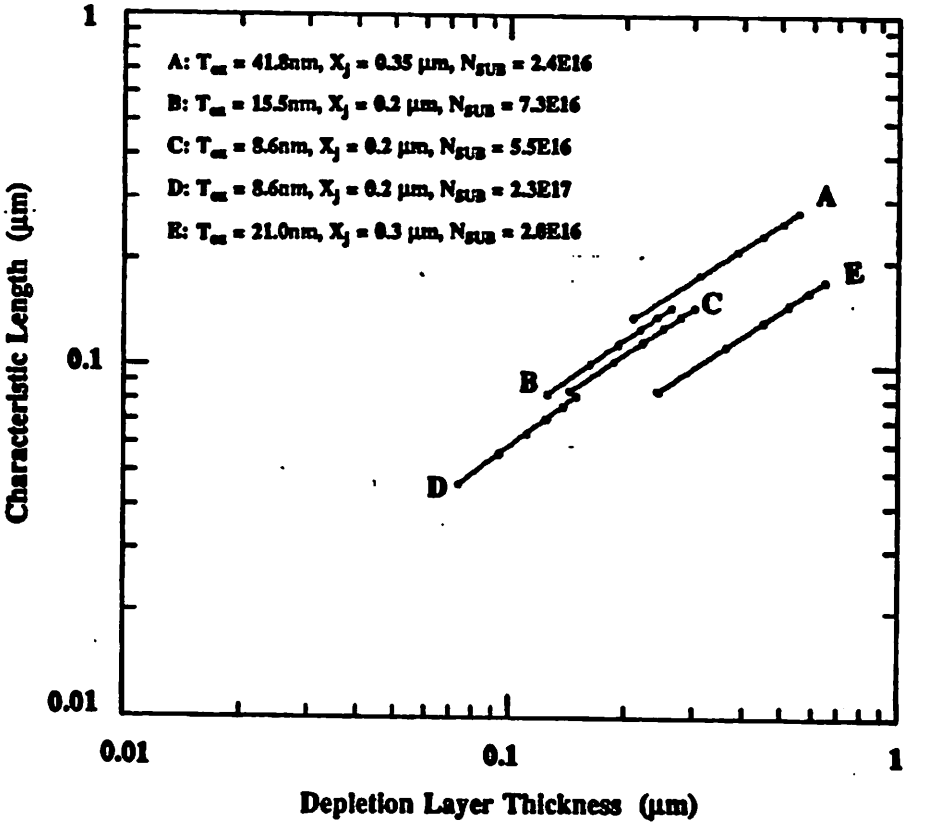


Fig. 3.9 Measured characteristic length versus depletion layer thickness for different technologies.

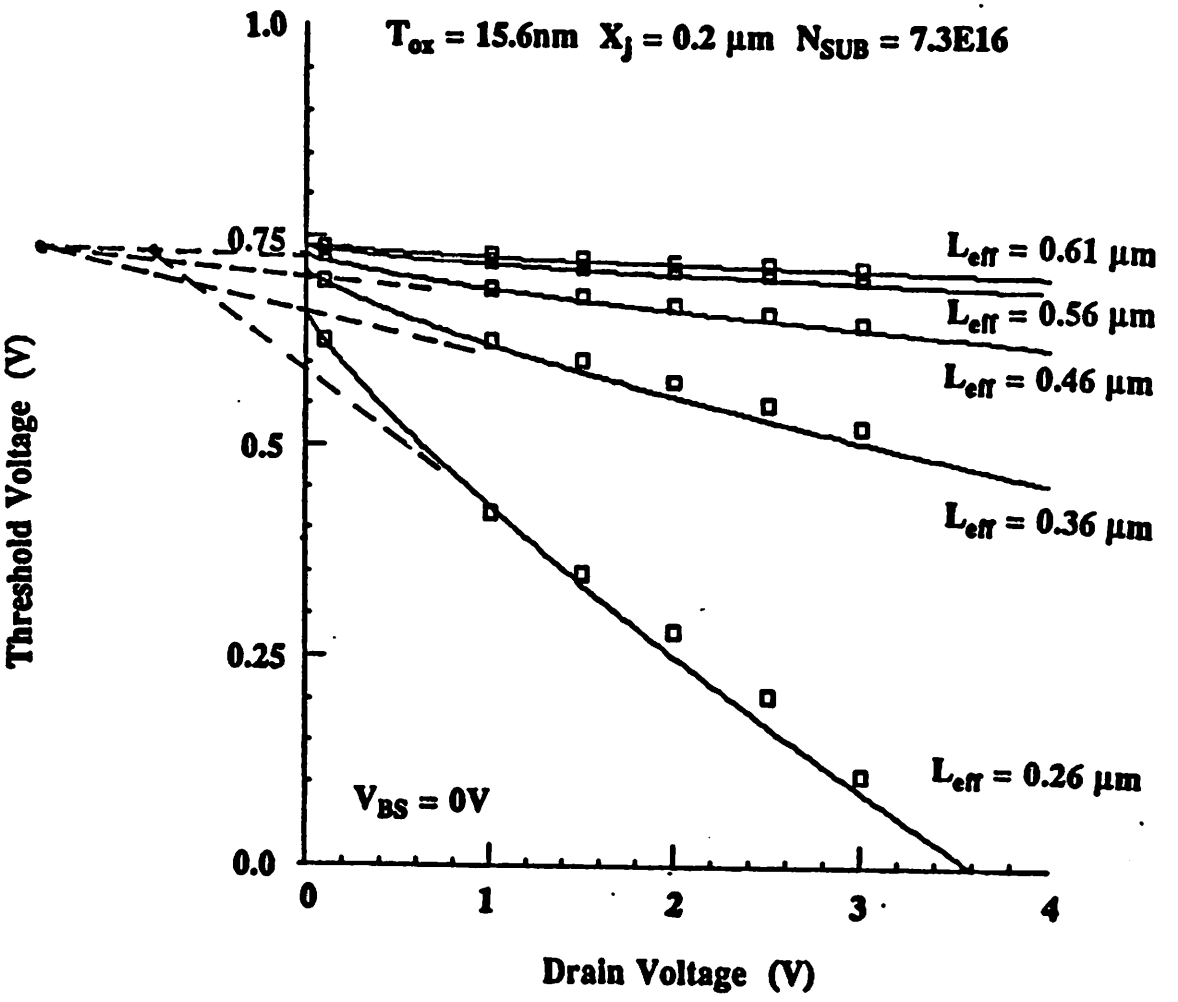
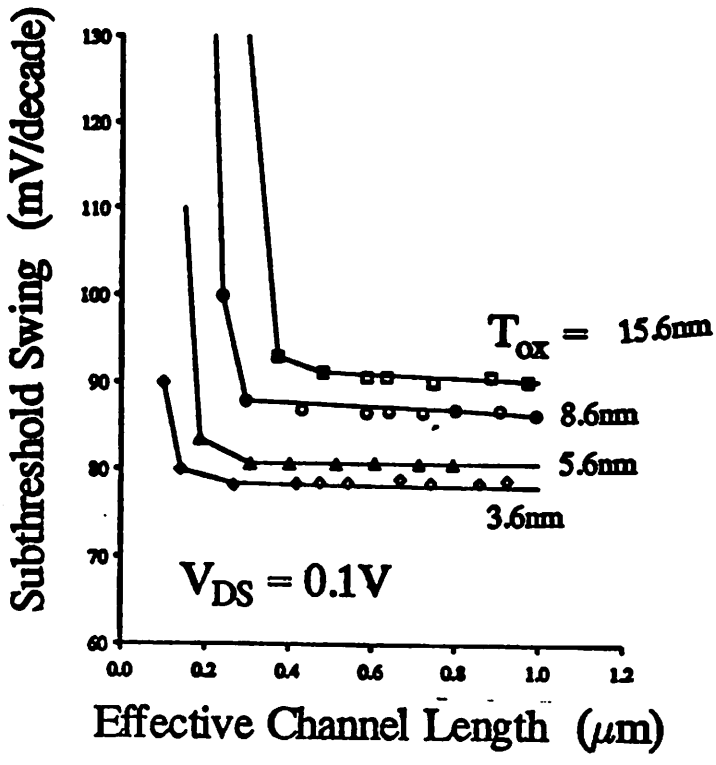


Fig. 3.10 Threshold voltage versus drain voltage. The symbols are measured data, the solid curves are the model, and the dashed lines are the best linear fit to the curves.



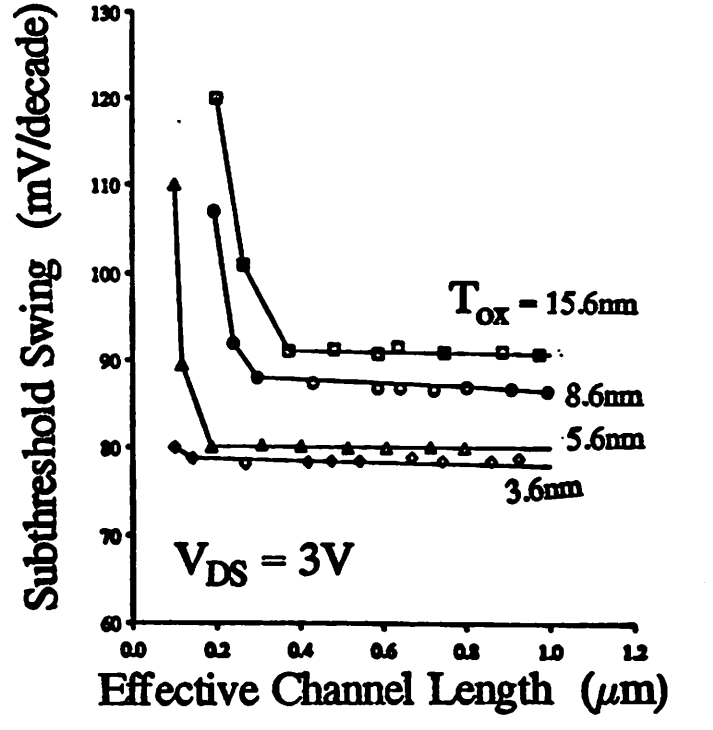


Fig. 3.11 Subthreshold swing versus effective channel length for four oxide thicknesses. (a) V_{DS} = 0.1V, (b) V_{DS} = 3V.

3.2.2 Subthreshold swing

The increase of subthreshold swing, S, at shorter channel lengths [3.30] is another factor causing short-channel devices to be more difficult to turn off. Therefore, the subthreshold swing also serves as an alternative way to monitor the extent of short-channel effects. The subthreshold swing versus L_{eff} for this process at a low and a high drain voltage are shown in Fig. 3.11a and 3.11b, respectively. Unlike the threshold voltage, subthreshold swing is fairly constant even when ΔV_{th} starts to show up, but suddenly increases to a large value when the device is near punchthrough. Also, the subthreshold swing is less sensitive to the drain voltage than V_{th}. Theoretical value of the subthreshold swing is given by

$$S = V_{tm}\frac{C_{ox}+C_D}{C_{ox}} \qquad [3.17]$$

where C_D is the depletion-layer capacitance. The subthreshold swing decreases as the gate oxide thickness is reduced as predicted by (3.17). For thin gate oxides, higher channel doping concentrations are required to maintain a 0.65V threshold voltage which also increases C_D. This explain why the long-channel subthreshold swings for 3.6nm and 5.6nm gate oxide devices are very close.

3.3 Current driving capability

The improved current drive of short channel devices is one of the motivations for MOSFET scaling. Because of the carrier velocity saturation effect, drain saturation current increases only sublinearly with $1/L_{eff}$ in the submicrometer regime but the design and fabrication overheads increase drastically with reducing the channel length. Therefore, a quantitative study of the current driving capability of deep-submicrometer devices is another important procedure in optimal device designs. When the channel length is smaller than 0.2μm and the power supply is not proportionally scaled (3V or higher), the velocity saturation region extends into a substantial fraction of the channel and a considerable portion of electrons in the velocity saturation

region move with a velocity higher than the saturation velocity v_{sat}. It was claimed that the current driving capability of deep-submicrometer devices would be enhanced by this velocity-overshoot effect [3.31,3.32]. A straight forward way to examine whether the current driving capability of deep-submicrometer devices is enhanced or affected by new carrier transport mechanisms is to compare experimental data with existing physical models. The drain current model used here was developed by Ko [3.33], improved by Toh [3.34], and has been successfully applied to devices with channel length longer than 1μm. Some of the model equations are listed below.

$$I_{DSAT}=W_{eff}v_{sat}C_{ox}(V_{GS}-V_{th}-V_{DSAT}) \tag{3.18}$$

and

$$g_{msat}=\frac{\partial I_{DSAT}}{\partial V_{GS}} \tag{3.19}$$

where g_{msat} is the saturation transconductance,

$$V_{DSAT}=\frac{E_cL_{eff}(V_{GS}-V_{th})}{E_cL_{eff}+(V_{GS}-V_{th})} \tag{3.20}$$

$$E_c=\frac{2v_{sat}}{\mu_{eff}} \tag{3.21}$$

$$\mu_{eff}=\frac{\mu_0}{1+(E_S/E_0)^n} \tag{3.22}$$

and E_S is the effective vertical field in the channel that can be approximated by

$$E_S=\frac{1}{2\varepsilon_{si}}[2Q_B+C_{ox}(V_{GS}-V_{th})] \tag{3.23}$$

Q_B is the depletion bulk charge, E_0 = 0.67MV/cm, n = 1.6, μ_0 = 670cm^2/V sec, and v_{sat} = 8×10^{16}cm/sec.

The measured drain saturation current I_{DSAT}, saturation transconductance g_{msat} and the model for an array of devices with channel length down to 0.15μm are shown in Fig. 3.12.

The same model parameters were used for all device dimensions and oxide thicknesses. The symbols in Fig. 3.12 are measured data and the solid curves are the model. The data shown in Fig. 3.12 have been corrected for the source/drain parasitic resistance (≈ 30Ω/side). A comprehensive study of the source/drain parasitic resistance effect on device performance is given in the next section. The inversion-layer capacitance effect [3.35], which is more important for thin-oxide devices at low gate bias, was also not included in these equations.

The well-behaved trends of I_{DSAT}, and g_{msat} and the good agreement between measured data and the model indicate that the basic physics of deep-submicrometer devices is rather well understood. Although Monte Carlo simulations show the existence of velocity overshoot in the velocity saturation region, it has little effect on the MOSFET current driving capability, at least down to 0.15μ channel length. This observation also coincides with the conclusion of another independent study [3.36], which used an improved mobility model (extended drift-diffusion model) to simulate the velocity overshoot effect in the velocity saturation region. According to their simulations, the velocity overshoot effect is of little importance to the MOSFET current driving capability for devices with channel length longer than 0.06μm.

Since the basic physics in deep-submicrometer devices is essentially unchanged, the basic framework of most existing drain current models can be kept without major modifications. The drain current model used in this section, while simplistic in formulation, still provides good physical as well as quantitative understanding of the current performance of MOS devices down to the deep-submicrometer regime and can serve as a means for process control and diagnosis.

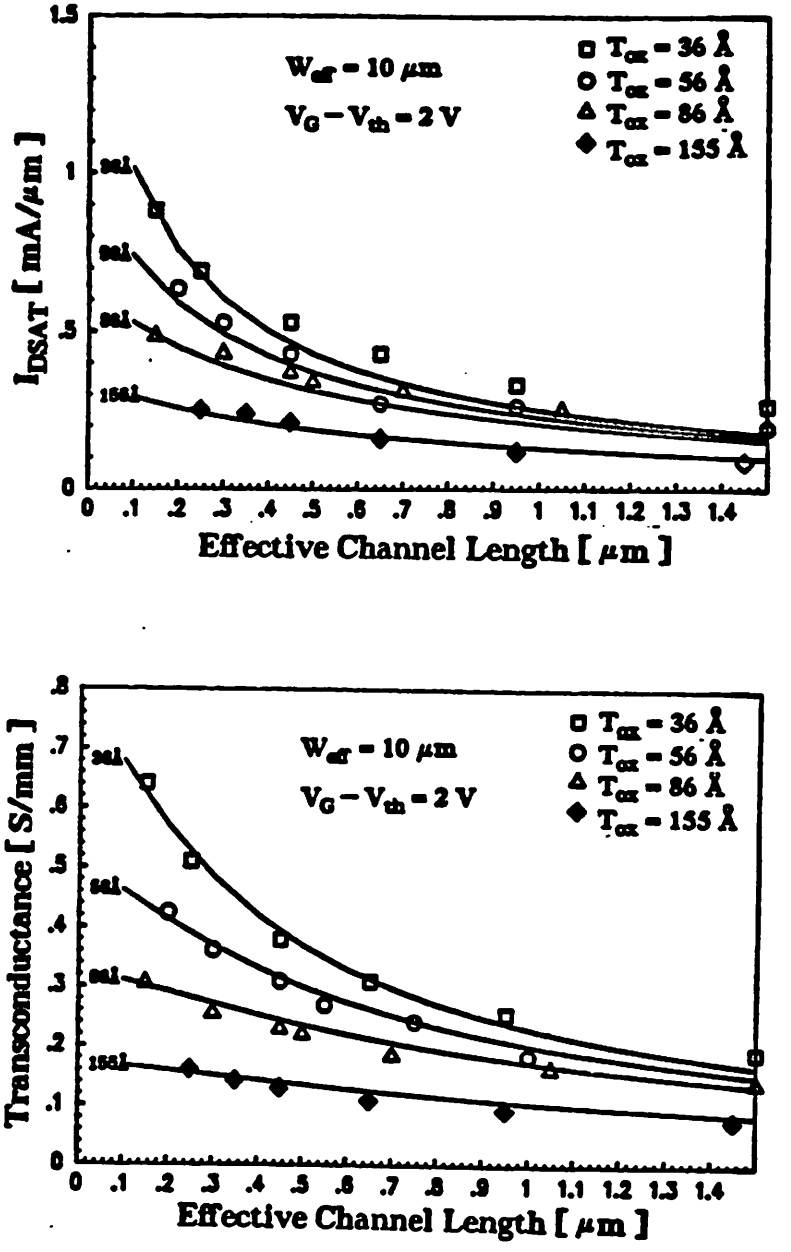


Fig. 3.12 (a) Measured drain saturation current at $V_{GS}-V_{th}$ versus effective channel length for various oxide thicknesses, (b) measured transconductance versus effective channel length. The data have been corrected, to the first order, for the parasitic resistance effect.

3.4 Source/drain parasitic resistance effect

The parasitic source/drain resistance is one of the device parameters that can not be proportionally scaled. As MOSFET channel lengths are scaled down to the deep-submicrometer regime, device performance reduction due to parasitic source/drain resistance (R_{sd}) becomes an important factor to consider in MOSFET scaling [3.37-3.40]. A quantitative study of R_{sd} effects is essential, since it can provide guidelines for both MOSFET scaling and contact technology development.

It was claimed that as the device channel length is scaled below 0.5 μm, the current drive and transconductance starts to decrease rather than increase with the reduction of the channel length [3.38] implying that the parasitic resistance poses a limit on MOSFET scalability. But this statement has been shown to be incorrect because of the recent improvements in device technology. Previous reports [3.37-3.39] on this subject, based on near-micron technologies, also may not be applicable to the deep-submicrometer regime. More recently, Ng and Lynch [3.40], using computer simulations, studied the R_{sd} effects in the deep-submicrometer regime but with only little experimental results. In this section, experimental studies of the R_{sd} effects on deep-submicrometer n-channel non-LDD MOSFETs is presented. The reduction in drain currents and ring oscillator speed for various channel lengths and R_{sd} values is examined. The effect of salicide technologies on device performance is also discussed and projections of the ultimate achievable device performance are given.

3.4.1 Experimental procedure

Intrinsic Device Performance Measurement Procedure: In order to determine the amount of performance reduction due to R_{sd}, the following calibration procedure was performed. The drain current in the linear (I_{DLIN}) and saturation (I_{DSAT}) regions and the maximum saturation transconductance (g_{msat}) were measured. In Fig. 3.13 the measured I_{DSAT} is plotted against R_{SD}; different R_{SD} values were achieved by attaching external resistors (R_{ext}), equally

divided between the source and drain, to each device, i.e. $R_{SD} = R_{SD0} + R_{ext}$. The circles indicate measured data. The solid lines represent the simple physical drain current model described in section 3.3. A calibration constant (in the range between 1.0 to 1.1 for all devices) is multiplied to the model to best fit the measured data for each channel length. To obtain higher accuracy, parasitic resistance effects were included in the drain current model through iteration; parasitic resistance-induced body effect, which was neglected in [3.40], was also included in the calculation. The theoretical drain currents at $R_{SD} = 0$ are taken as the intrinsic current ($I_{DLIN,i}$ and $I_{DSAT,i}$). The percent drain current reduction from the intrinsic value as a function of R_{SD} is given by the alternated curves.

3.4.2 Experimental results

(A) Saturation region: Fig. 3.14 shows I_{DSAT} versus L_{eff} for a power supply of 3.3V. The symbols indicate measured data; the curves are the calculated intrinsic ($R_{SD} = 0$) drain current obtained in the manner shown in Fig. 3.13 and the corresponding current derating, $I_{dsat}/I_{dsat,i}$. Because of the slightly different parasitic resistance between wafers, R_{SD} values were adjusted to be about 600Ωμm for all oxide thicknesses using external resistors. Similar results were also obtained for the transconductance. We observed that the current (transconductance) derating decreases as L_{eff} and/or T_{ox} decreases because the debiasing (source follower) effect of R_{SD} is stronger as I_{DS} (g_{msat}) increases. However, the derating is still about 87% even at $L_{eff} = 0.2\mu m$, if R_{sd} is kept at 600Ωμm.

(B) Linear region: Fig. 3.15 shows I_{DLIN} versus L_{eff} at $V_{DS} = 0.1V$ and $V_{GS} = 3.3V$. The drain current derating in the linear region is significantly lower than that in the saturation region. The derating can be as low as 50% at $L_{eff} = 0.2\mu m$. This is because in the linear region R_{sd} reduces the current through both the effective V_{GS} and V_{DS} while in the saturation region it only reduces the current through the effective V_{GS}. Also, the current derating is less sensitive to T_{ox} than in the saturation region. This is because I_{DLIN} is less sensitive to T_{ox} due to the transverse-field-induced mobility reduction than I_{DSAT}, which is mainly determined by

carrier saturation velocity that is insensitive to the transverse field.

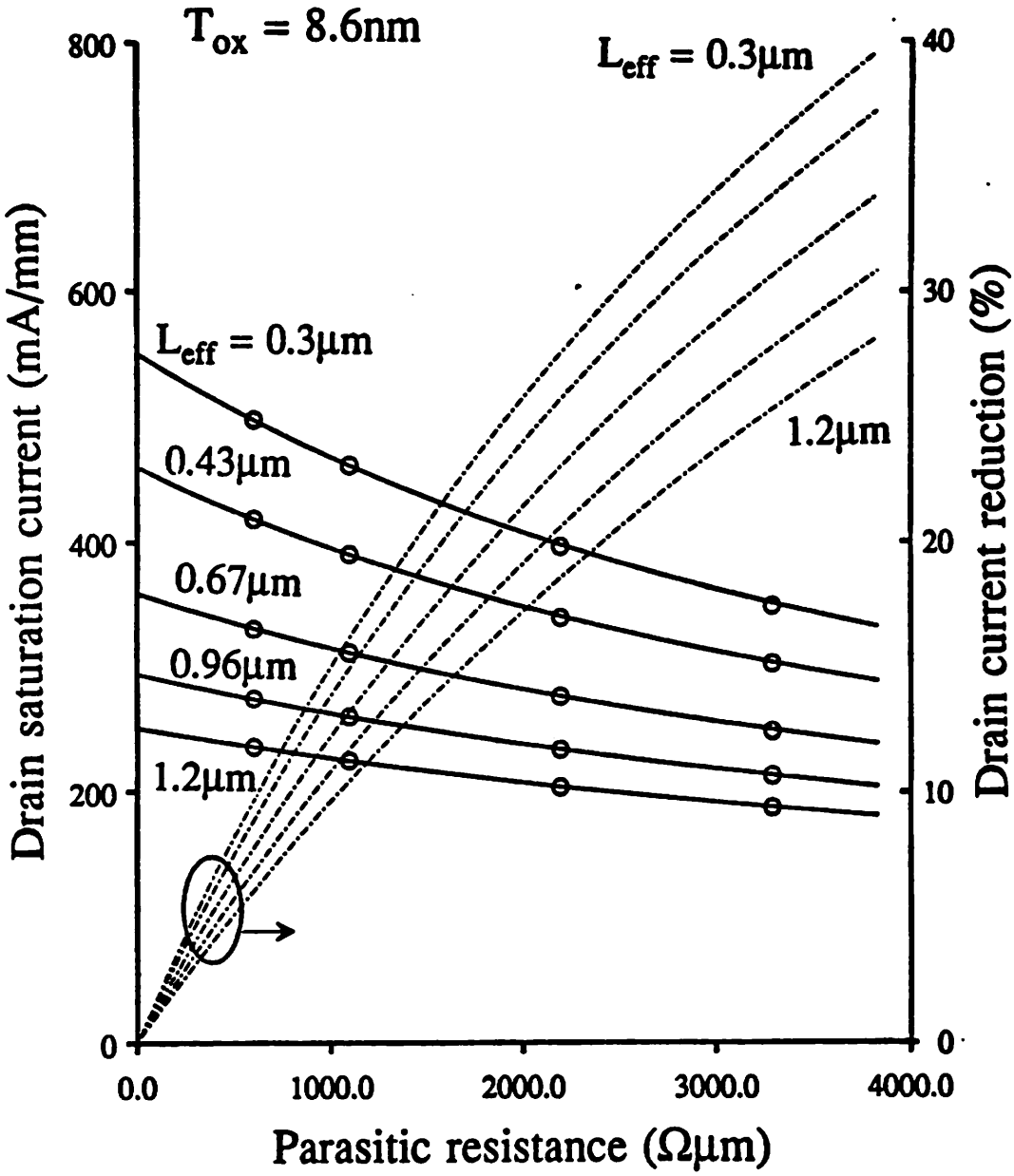


Fig. 3.13 Drain saturation current versus parasitic resistance. The circles are measured data and the solid curve are the results of the calibrated model. The dashed lines indicate the percentage drain current reduction.

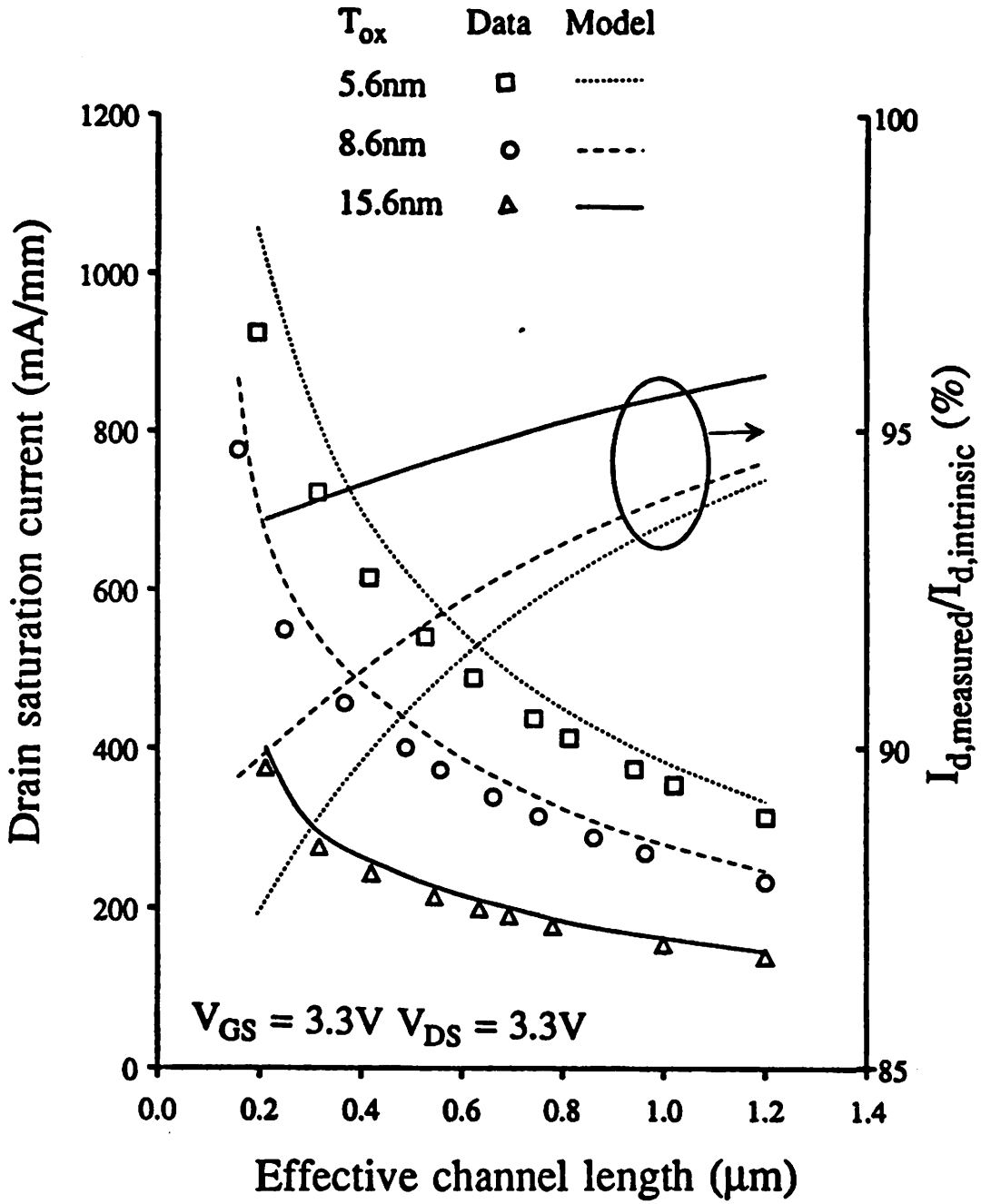


Fig. 3.14(a) Drain saturation current and the derating versus effective channel length. The symbols are measured drain current and the curves are their corresponding intrinsic values and deratings.

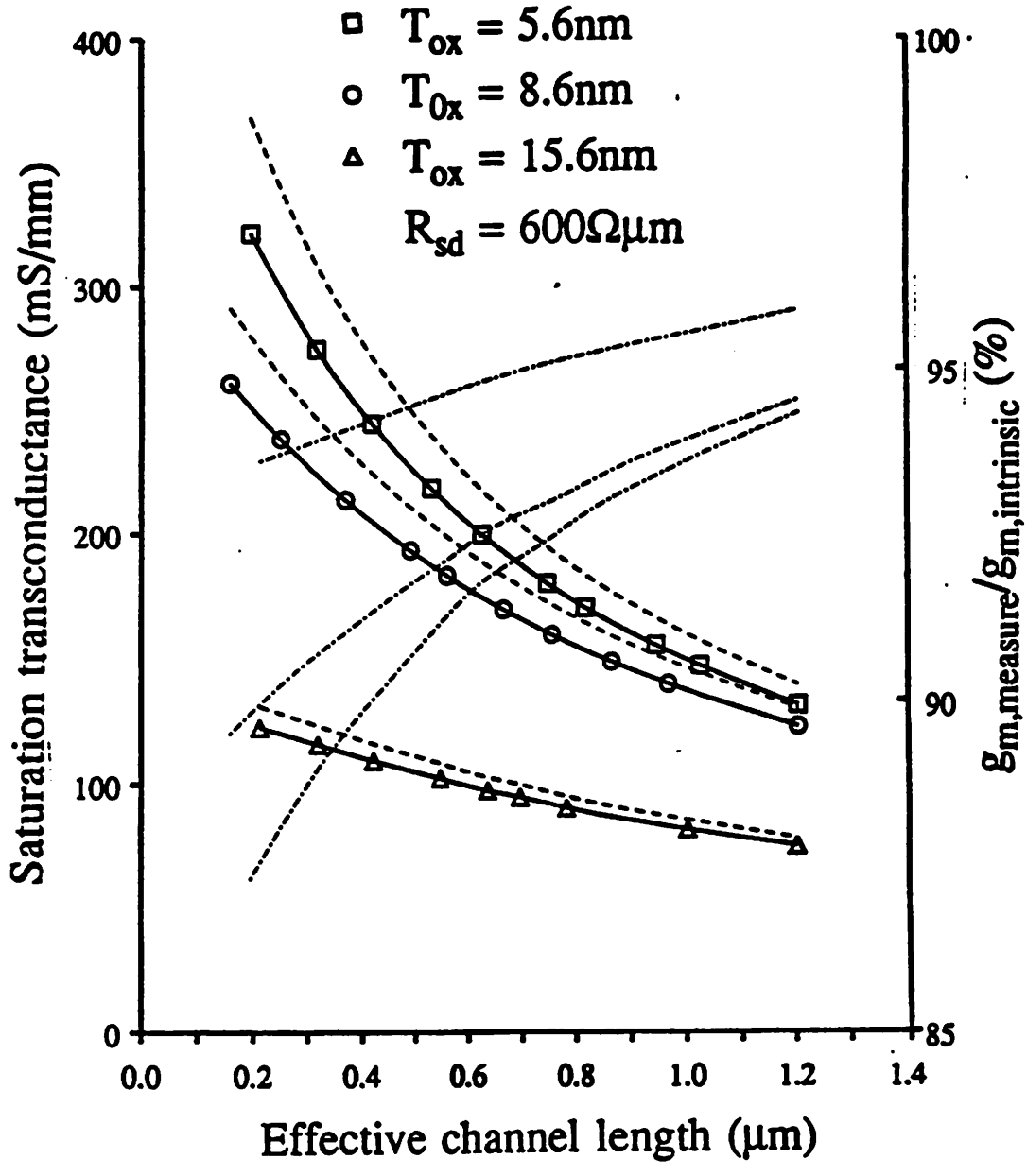


Fig. 3.14(b) Saturation transconductance and the derating versus effective channel length. The symbols are measured data and the curves are their corresponding intrinsic values and deratings.

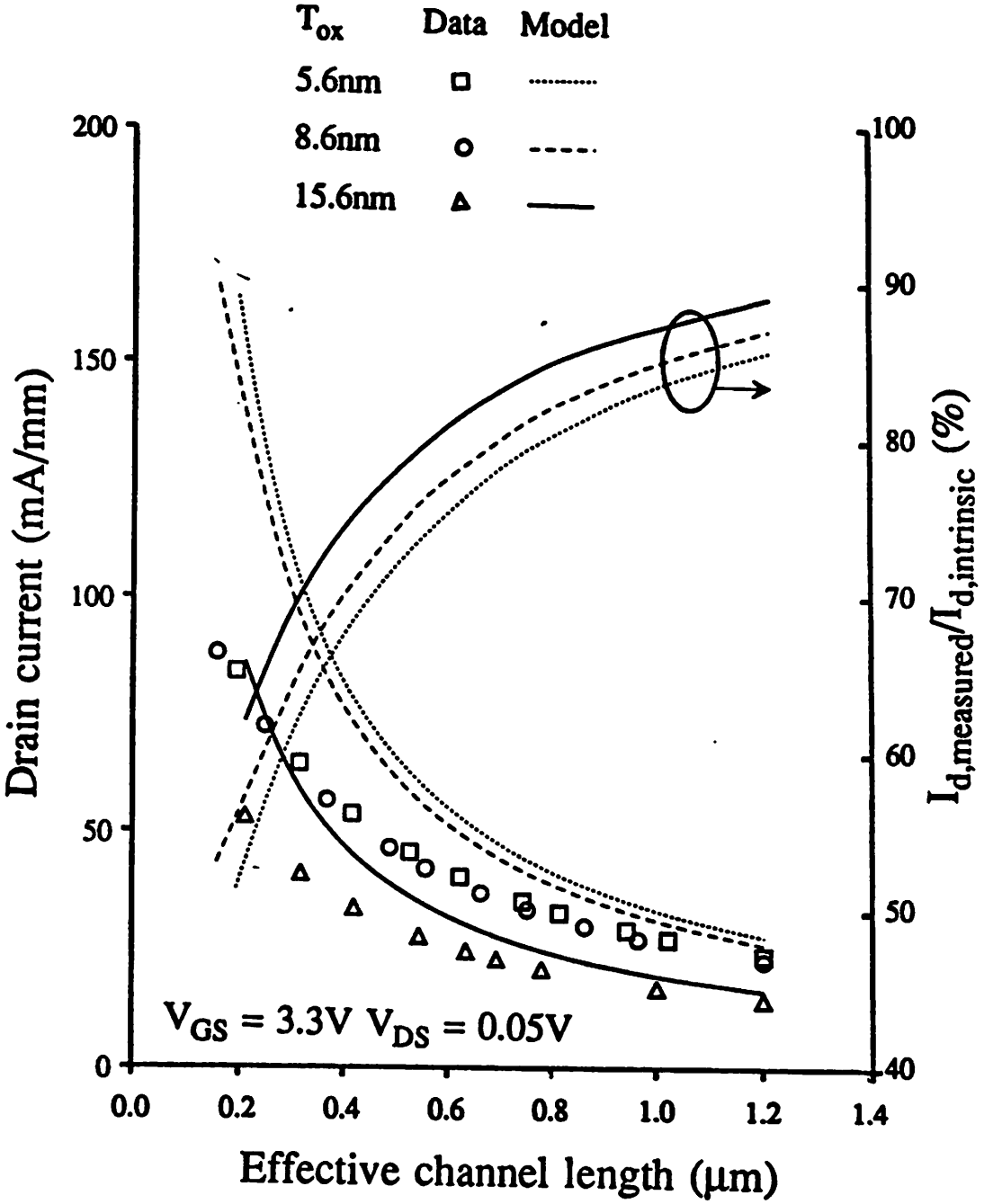


Fig. 3.15 Linear region drain current and the derating versus effective channel length. The symbols are measured data and the curves are their corresponding intrinsic values and deratings.

(C) Switching Speed: Depending on the circuit configuration, the circuit speed derating should lie between the derating of I_{DLIN} and I_{DSAT} shown in Fig. 3.14 and Fig. 3.15. Fig. 3.16 shows SPICE simulated delay time (τ) per stage of CMOS ring oscillators versus R_{SD} for T_{ox} = 8.6nm. The widths of the n- and p-channel devices are 15μm and 30μm, respectively. The loading capacitance is 0.1pF on each stage. The simulation results show that τ increases roughly as a linear function of R_{sd} with a slope about 10 - 40% per kΩμm for T_{ox} between 5.6nm and 15.6nm, and L_{eff} between 0.25μm and 1.5μm. At 600 Ωμm and for T_{ox} = 8.6nm, the speed derating is about 60% for L_{eff} = 0.2μm and 85% for L_{eff} = 1μm. One must conclude that the speed derating is closer to the I_{DLIN} derating than the I_{DSAT} derating. This conclusion differs from that drawn in [3.40].

3.4.3 Discussion

R_{SD} is usually divided into four components, namely, the contact (R_{co}), the diffusion sheet (R_{sh}), the spreading (R_{sp}), and the accumulation (R_{ac}) resistances as shown in Fig. 3.17 [3.40]. In this study, there is a polysilicon film between Al and Si, and R_{SD} is typically 500 ~ 600 Ωμm. The contact-to-gate spacing is about 0.7 μm. With similar contact-to-gate spacing, an R_{SD} value of 300Ωμm (R_{co} = 100, R_{sh} = 50, R_{sp} + R_{ac} = 150) can be achieved with conventional contact technologies [3.40]. This R_{SD} value corresponds to 95% I_{DSAT} derating, 79% I_{DLIN} derating, and 85% speed derating for L_{eff} = 0.3μm and T_{ox} = 8.6nm (see Fig. 3.18). Because of the hot-carrier effects, LDD structure will likely be used in scaled MOSFETs and could introduce an additional 100 - 400 Ωμm of R_{sd} depending on the device design and bias condition [3.41]. This LDD resistance would further derate I_{DSAT} by about 2-5%, I_{lin} by 9-17%, and the speed by 5-15%. Finally, the maximum benefit of employing a salicide technology can be estimated by assuming that the salicide technology totally eliminates R_{co} and R_{SD}. This would increase I_{DSAT}, I_{DLIN} and the speed by about 2.5%, 12%, and 7.5 for both the LDD and non-LDD devices.

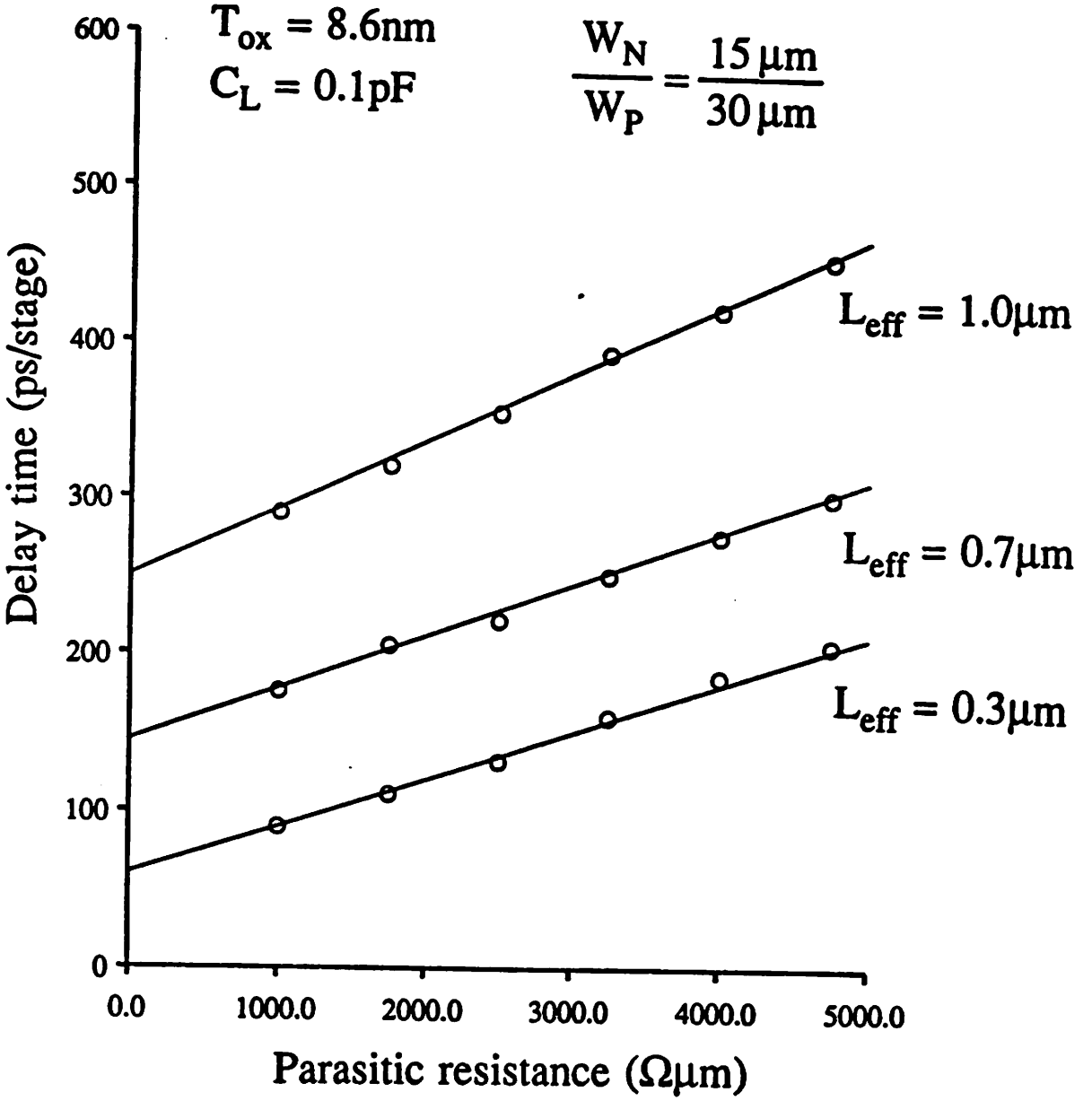


Fig. 3.16 SPICE simulated ring oscillator delay time versus parasitic resistance for several channel lengths.

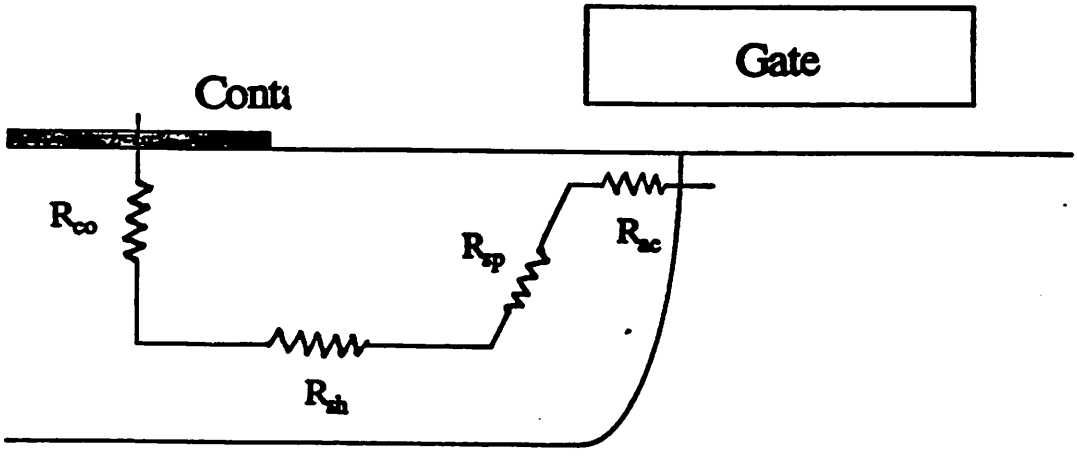


Fig. 3.17 A schematic diagram showing the various components of the parasitic resistance.

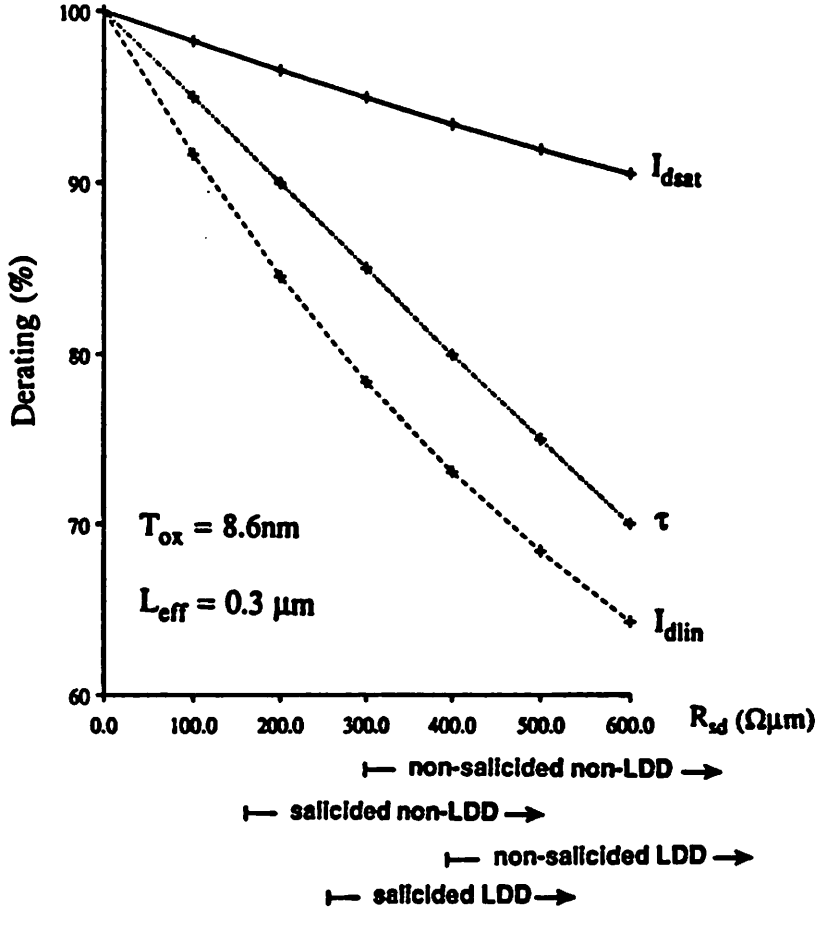


Fig. 3.18 Derating versus parasitic resistance. The projected parasitic resistance effect on various technologies are indicated.

3.5 Hot-electron effects

It is well known that the hot-electron induced device degradation poses circuit reliability problems. Therefore, device degradation is the most discussed topic among all hot-electron effects. Because of the extremely high electric filed in the channel due to the unscaled power supply, hot-electron effects are more severe in deep-submicrometer devices and have gained more attention than ever in MOS scaling. In this section, the substrate current and the device lifetime are discussed.

3.5.1 Substrate current

When carriers pass through the velocity saturation region, electron-hole pairs are generated by impact ionization. The holes are collected by the substrate and constitutes the substrate current I_{SUB}. Since many hot-electron phenomena, including device degradation, have close correlations with I_{SUB}, the substrate current is widely used to monitor the hot-electron effects. In section 3.3, we have shown that the basic physics of deep-submicrometer devices is unchanged. In this section, this statement will be demonstrated again from the hot-electron point of view. A commonly used substrate current model [3.42] in the literature is used for this purpose. The model equations are summarized below.

$$I_{SUB} = I_{DS}\left(\frac{\alpha_i}{\beta_i}\right)\left(\frac{E_{max}}{l}\right)e^{\frac{-\beta_i}{E_{max}}} \tag{3.24}$$

where

$$E_{max} \approx \frac{V_{DS} - V_{DSAT}}{l} \tag{3.25}$$

α_i and β_i are impact ionization coefficients whose typical values are $2E6cm^{-1}$ and 1.7E7V/cm, respectively ($1E7cm^{-1}$ and 3.7E6V/cm for PMOS). For long channel and thick gate oxide devices, an empirical expression for l has been observed [3.43,3.44].

$$l = 0.22T_{ox}^{1/3}X_j^{1/3} \tag{3.26}$$

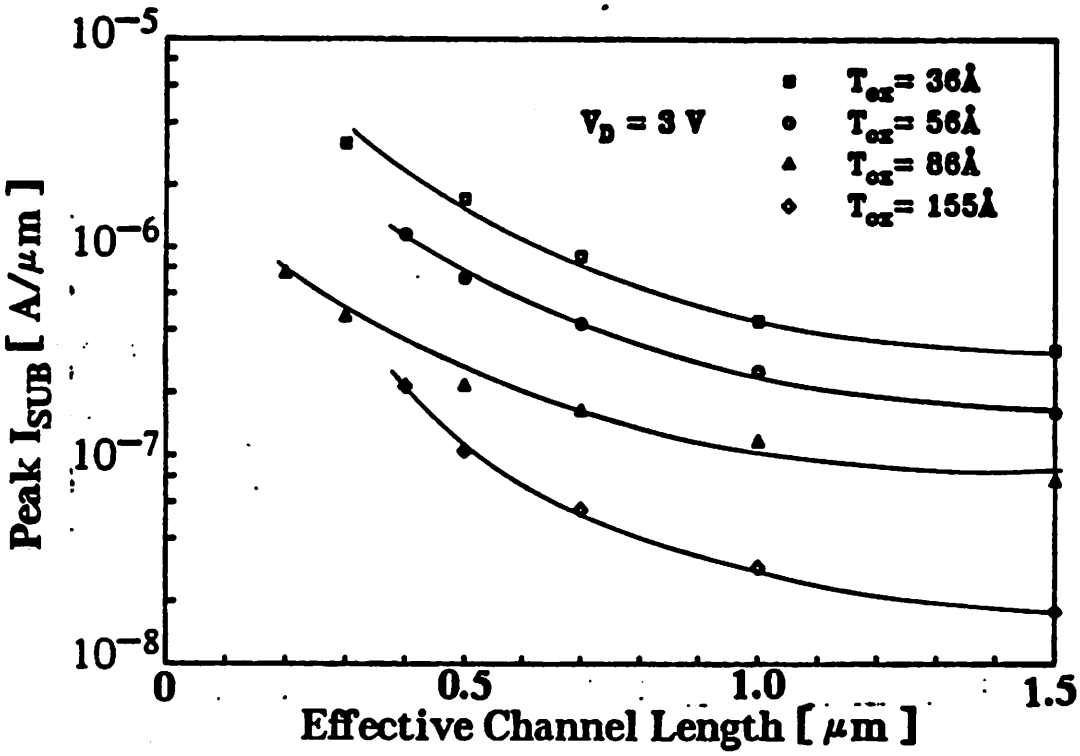


Fig. 3.19 Peak substrate current versus effective channel length at V_{DS} = 3V for four oxide thicknesses.

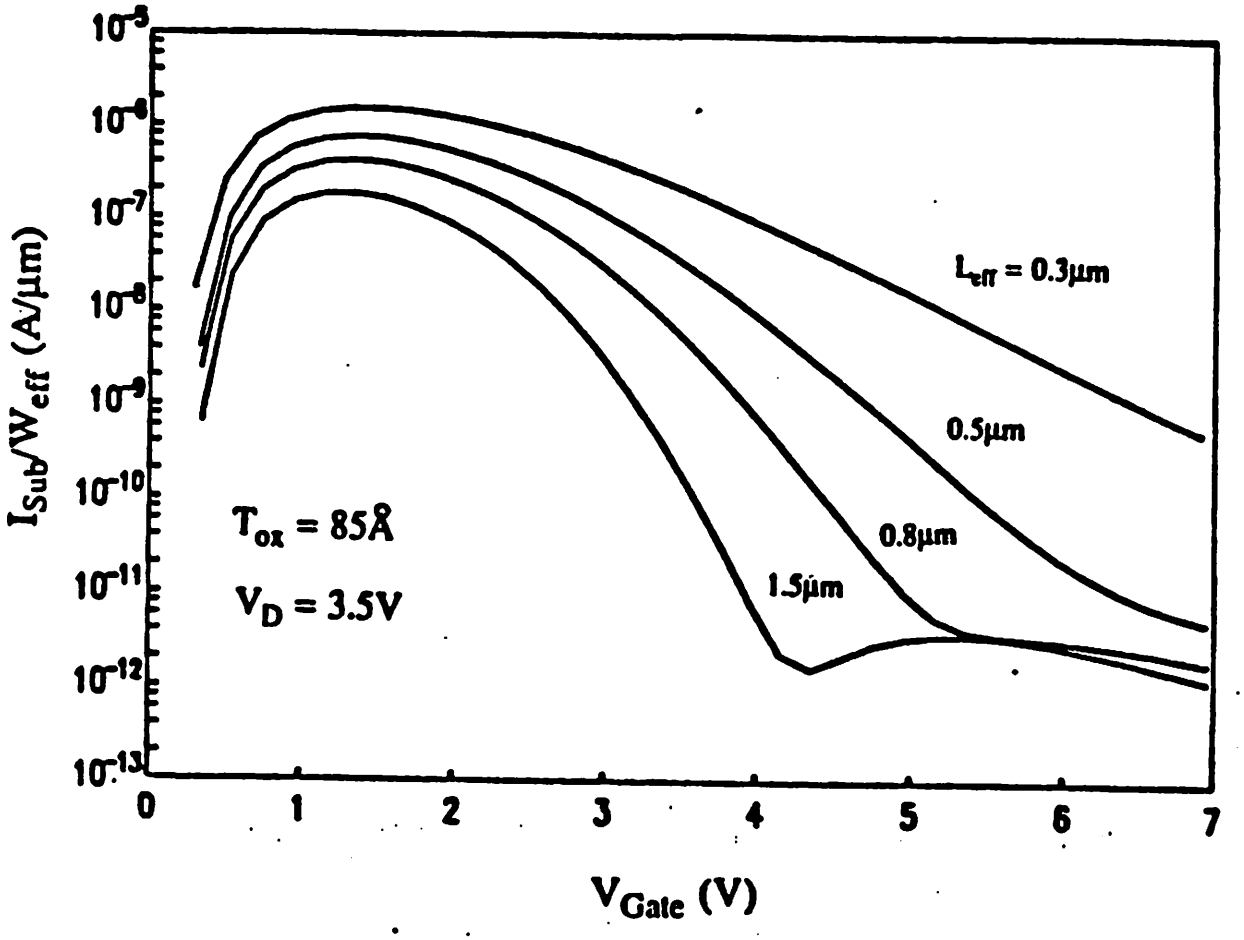


Fig. 3.20 Substrate current characteristics for different channel lengths. As the channel length decreases, the gate voltage control over the substrate current reduces.

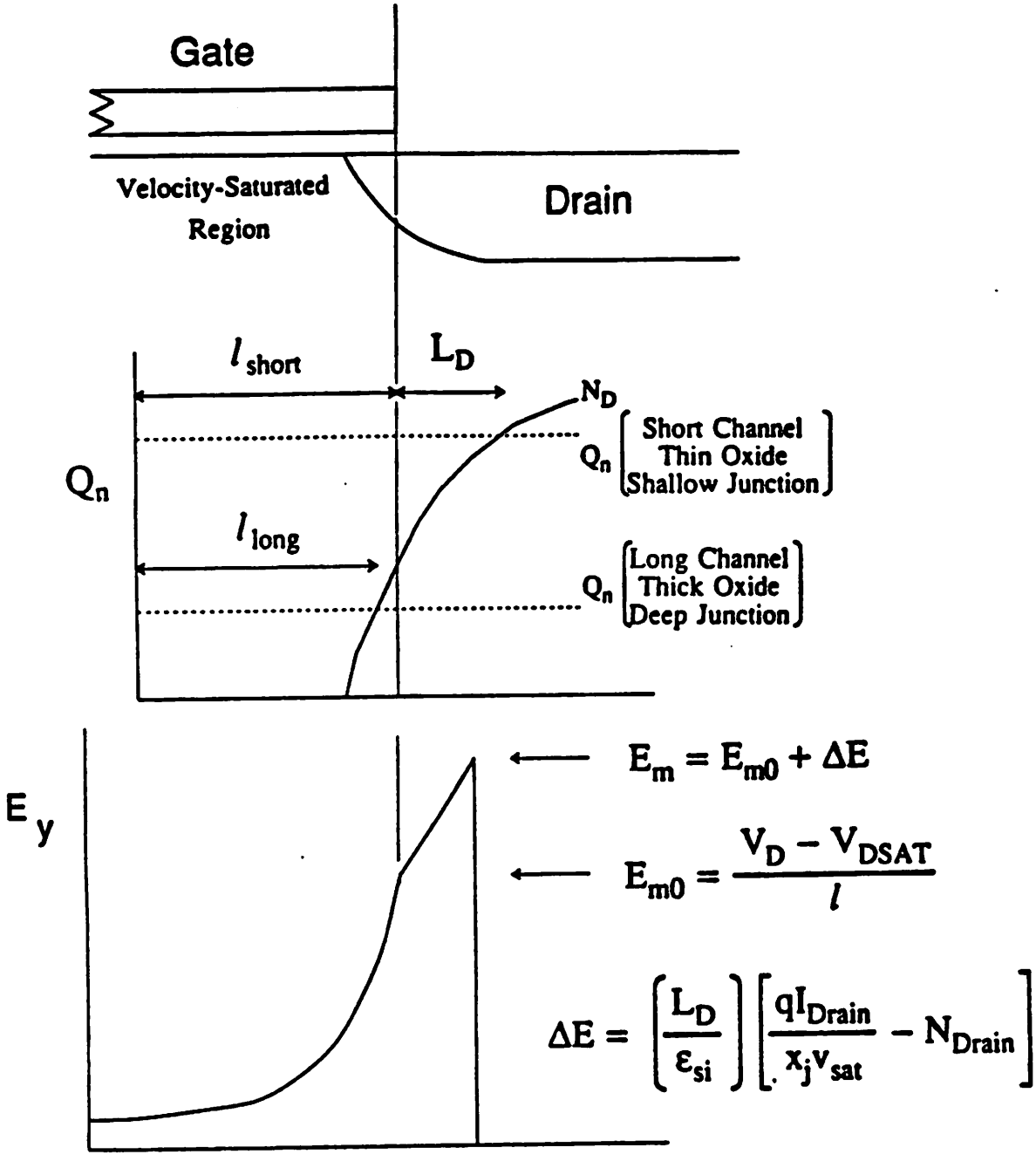


Fig. 3.21 Schematic diagrams showing the current-crowding induced weak gate control effect [3.48]. (a) cross section of a MOS transistor, (b) mobile charge density as a function of the channel position, (c) the electric field as a function of the channel position.

where T_{ox} and X_j are in centimeters.

The peak substrate current versus effective channel length at V_{DS} = 3V for several oxide thicknesses is shown in Fig. 3.19. The substrate current peak is higher for a device with shorter channel length and thinner gate oxide as predicted by (3.24)-(3.26). Fig. 3.20 shows I_{SUB} - V_{GS} characteristics for T_{ox} = 8.6nm at various channel lengths. At longer channel lengths (L_{eff} > 1.0μm), the shape of the I_{SUB} characteristics can be modeled. However, as the channel length is reduced, the substrate current becomes less sensitive to the gate voltage that can not be explained by the model.

Several modifications [3.45-3.47] to the I_{SUB} model have been proposed to describe the deviation of measured I_{SUB} from the simple theory. These modifications are usually implemented by making the impact ionization coefficients functions of applied voltages through some empirical expressions. The physical basis for them is the non-local impact ionization effect and the non-equilibrium conditions in the high-field region near the drain.

Another proposal is the so-called "current-crowding induced weak gate control" [3.48] which can be schematically explained in Fig. 3.21, where the cross section and the doping profile of the drain of a MOSFET are shown. When the drain current is small, as in long and thick gate oxide devices, the mobile charge density required to carry the drain current in the velocity saturation region is negligible compared to the drain doping concentration. Therefore, the boundary of the velocity saturation region is very close to the edge of the drain junction. The peak channel electric field is inside the velocity saturation region and can be approximated by (3.24). For short-channel and thin gate oxide devices, the drain current is large and the mobile charge density in the velocity saturation region is comparable to the doping concentration of the drain region. Therefore, the velocity saturation region extends into the drain and the peak channel electric field also occurs inside the drain. As a result of the extra depletion region in the drain, the peak electric field is higher than that given in (3.24) by ΔE.

$$\Delta E = \frac{L_D}{\varepsilon_{si}}\left(\frac{qI_{DS}}{X_j v_{sat}} - N_D\right) \tag{3.27}$$

where N_D is the average drain doping concentration, L_D is the length of the weak gate-controlled region. The magnitude of L_D depends on the drain structure, usually in the range of 0 - 10nm. Detailed description of this current-crowding induced weak gate-control can be found in [3.48].

More informative figures of I_{SUB} are I_{SUB}/I_{DS} versus $1/(V_{DS}-V_{DSAT})$ plots as shown in Fig. 3.22. The straight lines in Fig. 3.22 suggests that the basic physical mechanism for hot-electron effects still prevail in deep-submicrometer devices. According to the hot-electron model, the slopes of the lines ($\propto l$) in Fig. 3.22 are independent of the channel length and have a one-third power dependence on gate oxide thickness. However, it is found that when the effective channel length is smaller than about 0.5μm, the slope (l) decreases with L_{eff} (see Fig.3.22a). We suspect this channel length dependence of l have to do with the encroachment of the linear region into the velocity saturation region as the channel length decreases. It is also found that, when the gate oxide thickness is smaller than about 15nm, l is very weakly dependent on T_{ox} (see Fig. 3.22b). One explanation to this weak gate oxide dependence of l is the finite depth of the current path in the velocity saturation region. In deriving (3.21)-(3.23), it is assumed that the impact ionization occurs at the Si-SiO_2 interface. In the velocity saturation region, the actual drain current path, and thereby the peak impact ionization, is at about 10-30nm below the interface. Therefore, an effective gate oxide thickness T'_{ox}, which consists of T_{ox} and the current depth should be used in (3.23). When the gate oxide thickness is comparable to or smaller than the current depth, T'_{ox} is limited by the current depth and l becomes a weaker function of T_{ox}. An empirical expression for l is determined to be

$$l = 1.7\times10^{-2} T_{ox}^{1/8} X_j^{1/3} L_{eff}^{1/5} \qquad \text{for } L_{eff}<0.5\mu m \text{ and } T_{ox}<15nm \qquad (3.28)$$

where all quantities have the units of cm.

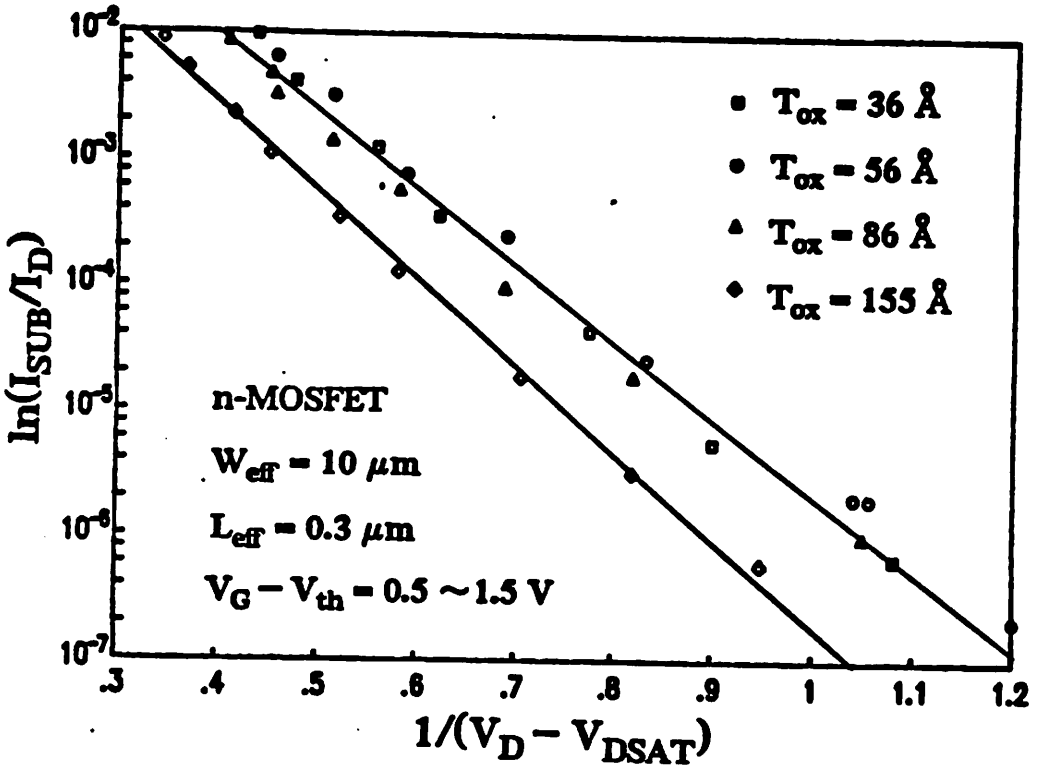


Fig. 3.22(a) $\log(I_{SUB}/I_{DS})$ versus $1/(V_{DS}-V_{DSAT})$ plots for L_{eff} = 0.3μm and four oxide thicknesses. The slopes of these lines are proportional to the impact ionization coefficient.

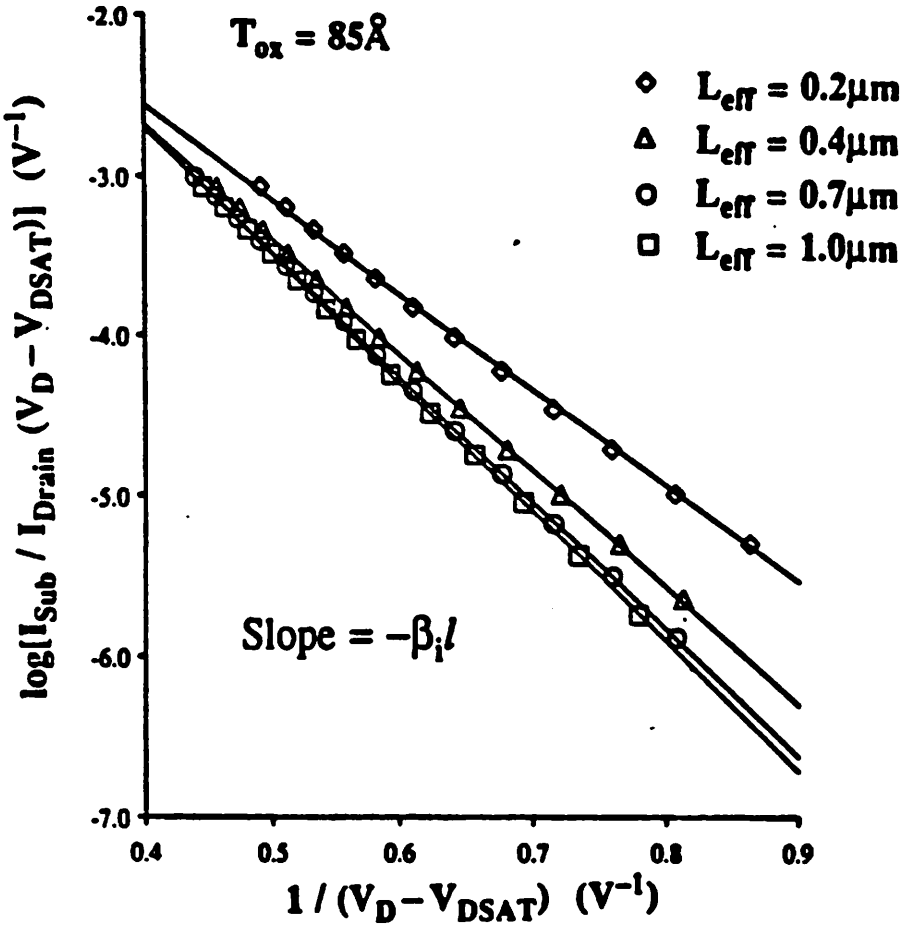


Fig. 3.22(b) $\log(I_{SUB}/I_{DS})$ versus $1/(V_{DS} - V_{DSAT})$ plots for T_{ox} = 8.5nm and several channel lengths. The slopes of these lines are proportional to the impact ionization coefficient.

3.5.2 Device lifetime

A detrimental effect of the high channel electric field is the injection of energetic electrons into the Si-SiO_2 interface that generates interface traps and results in device degradation [3.49,3.50]. How to reduce hot-electron device degradation has been the goal of many hot-electron studies. A common quantity to measure the immunity of a device to the hot-electron effect is the device lifetime, which is usually defined as 3% (sometimes 10%) forward drain current change in the linear region after hot-electron stress [3.42]. Previous studies on near-micron devices showed that the device degradation is technology dependent and is relatively independent of the channel length under the same stress conditions [3.51]. However, in the submicrometer regime, the effect of device degradation on the device performance is more prominent as indicated by the strong channel-length-dependent device lifetime shown in Fig. 3.23 [3.52]. Similar results are also observed for other oxide thicknesses. This channel length dependence of lifetime can be qualitatively explained in Fig. 3.24. If we assume that the hot-electron created damage (the dark region) is independent of the channel length for the same amount of stress (I_{SUB} * time = constant), then the ratio of the damaged interface area to the total channel area increases as the channel length is reduced and the device lifetime decreases because the relative amount of degradation increases.

A useful variation of Fig. 3.24 which provides direct device design guidelines is shown in Fig. 3.25, where the extrapolated maximum supply voltage to ensure a 10-year device lifetime for 8.6nm gate oxide is plotted against the channel length. As a result of shorter lifetimes, the maximum supply voltage is smaller for short channel devices. For a quarter-micron device with 8.6nm gate oxide, the maximum supply voltage is about 2.5V, suggesting that some kind of hot-electron-resistant structures are still needed even if the power supply is lowered to 3.3V.

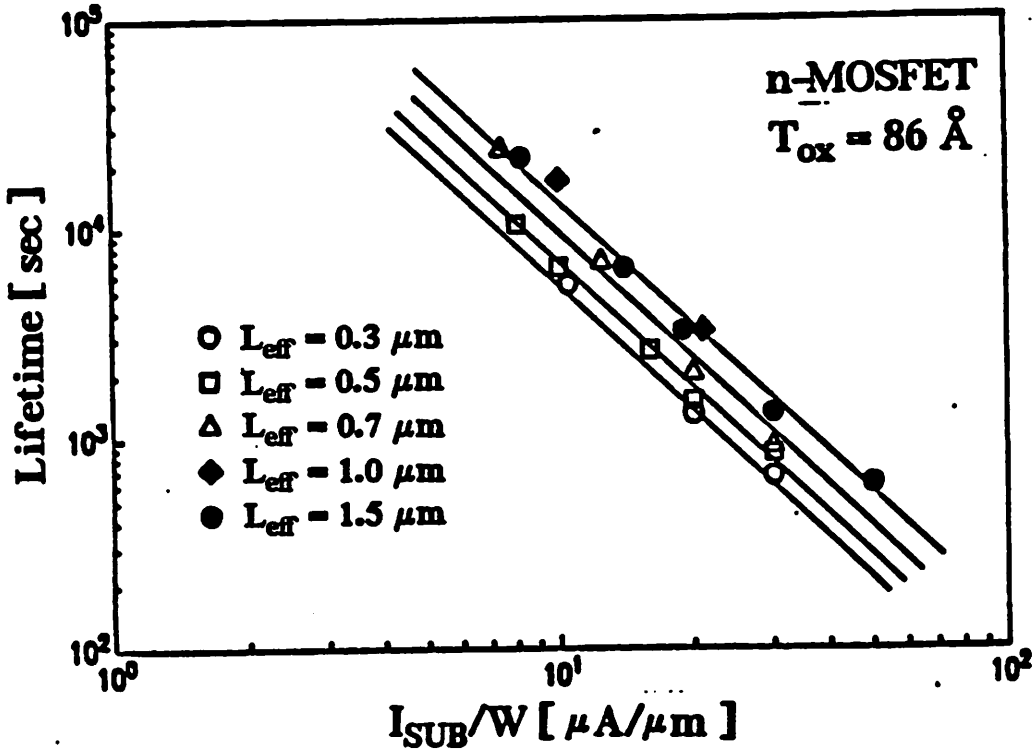
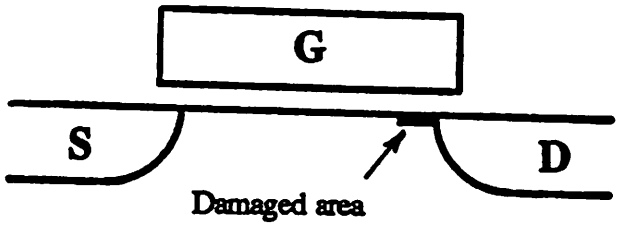
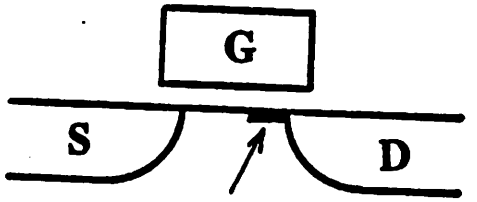


Fig. 3.23 Device lifetime versus substrate current for various channel lengths. The device lifetime is defined as 3% forward drain current degradation in the linear region.



Substrate

Long Channel



Substrate

Short Channel

Fig. 3.24 A schematic diagram explaining the channel length dependency of the device lifetime.

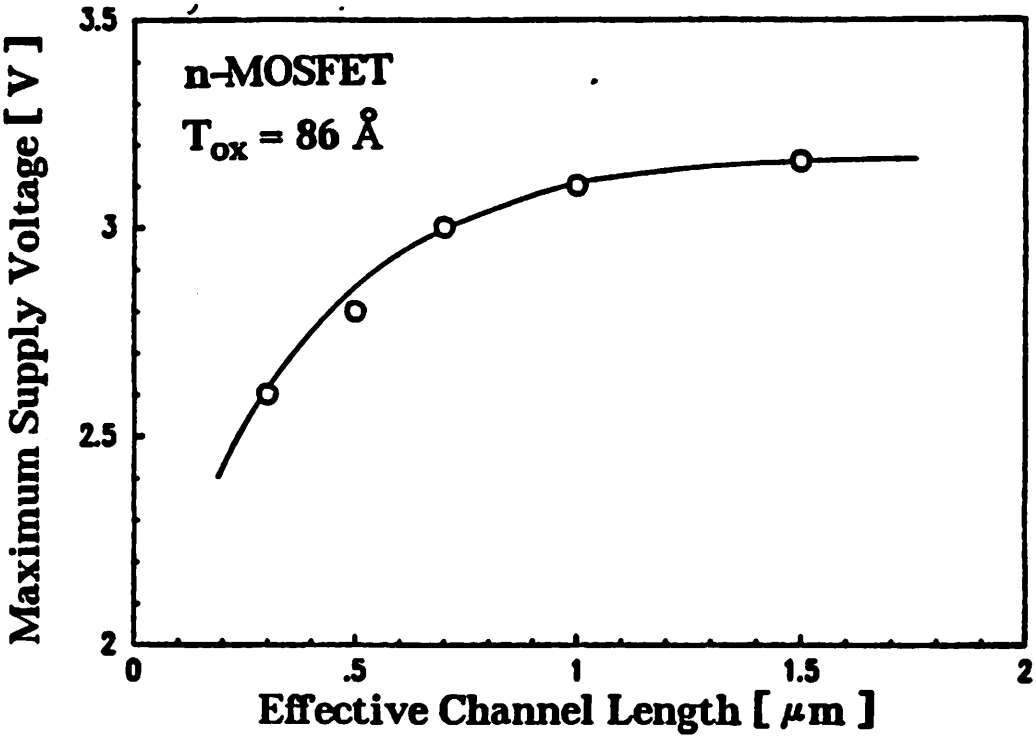


Fig. 3.25 Maximum supply voltage to ensure a 10-year device lifetime versus effective channel length for 8.6nm gate oxide. The device lifetime is defined as 3% forward drain current degradation in the linear region.

3.6 References

[3.1] W. Fichtner, R.K. Watts, D.B. Fraser, R.L. Johnston, and S.M. Sze, "0.15 μm Channel-Length MOSFETs Fabricated Using E-Beam Lithography," IEDM Tech. Dig., pp.272-725, 1982.

[3.2] W. Fichtner, E.N. Fuls, R.L. Johnston, R.K. Watts, and W.W. Weick, "Optimized MOSFETs with Subquartermicron Channel Lengths," IEDM Tech. Dig., pp.384-387, 1983.

[3.3] T. Kobayashi, S. Horiguchi, and K. Kiuchi, "Deep-Submicron MOSFET Characteristics with 5nm Gate Oxide," IEDM Tech. Dig., pp.414-417, 1984.

[3.4] S. Horiguchi, T. Kobayashi, M. Oda, and K. Kiuchi, "Extremely High Transconductance (Above 500mS/mm) MOSFET with 2.5nm Gate Oxide," IEDM tech. Dig., pp.761-763, 1985.

[3.5] S.Y. Chou, H.I. Smith, and D.A. Antoniadis, "Sub-100-nm Channel-Length Transistors Fabricated Using X-Ray Lithography," J. of Vacuum Science Tech. B, vol. 4, no. 1, pp.253-255, Jan/Feb 1986.

[3.6] J. Chung, M.-C. Jeng, J.E. Moon, A.T. Wu, T.Y. Chan, P.K. Ko, and C. Hu, "Deep-Submicrometer MOS Device Fabrication Using a Photoresist-Ashing Technique," IEEE Electron Device Letters, vol. 9, no. 4, pp.186-188, April, 1988.

[3.7] G.A. Sai-Halasz, M.R. Wordeman, D.P. Kern, E. Ganin, S. Rishton, D.S. Zicherman, H. Schmid, M.R. Polcari, H.Y. Ng, P.J. Restle, T.H.P. Chang, and R.H. Dennard, "Design and Experimental Technology for 0.1 μm Gate-Length Low-Temperature Operation FET's," IEEE Electron Device Letters, vol. EDL-8, no. 10, Oct., 1987.

[3.8] K. Terada and H. Muta, "A New Method to Determine Effective MOSFET Channel Length," Japanese J. Applied Phys., vol. 18, no. 5, pp.953-959, May, 1979.

[3.9] J.G.J. Chern, P. Vhang, R.F. Motta, and N. Godinho, "A New Method to Determine

MOSFET Channel Length," IEEE Electron Device Letters, vol. EDL-1, no. 9, pp.170-173, Sep., 1980.

[3.10] P.I. Sucis and R.L Johnston, "Experimental Derivation of the Source and Drain Resistance of MOS Transistors," IEEE Tran. on Electron Devices, vol. ED-27, no. 9, p.1846, Sep., 1980.

[3.11] K.L Peng and M.A. Afromowitz, "An Improved Method to Determine MOSFET Channel Length," IEEE Electron Device Letters, vol. EDL-3, no. 12, p.360, Dec., 1982.

[3.12] K.L. Peng, S.-Y. Oh, M.A. Afromowitz, and J.L. Moll, "Basic Parameter Measurement and Channel Broadening Effect in the Submicrometer MOSFET," IEEE Electron Device Letters, vol. EDL-5, no. 11, p.473, Nov., 1984.

[3.13] J. Whitfield, "A Modification on 'An Improved Method to Determine MOSFET Channel Length'," IEEE Electron Device Letters, vol. EDL-6, no. 3, p.109, March, 1985.

[3.14] M.J. Thoma and C.R. Westgate, "A New AC Measurement Technique to Accurately Determine MOSFET Constants," IEEE Tran. on Electron Devices, vol. ED-31, no. 9, p.1113, Sep., 1984.

[3.15] E.J. Korma, K. Visser, J. Snuder, and J.F. Verwey, "Fast Determination of the Effective Channel Length and the Gate Oxide Thickness in Polycrystalline Silicon MOSFET's," IEEE Electron Device Letters, vol. EDL-5, no. 9, p.368, Sep., 1984.

[3.16] B.J. Sheu and P.K. Ko, "A Capacitance Method to Determine Channel Lengths for Conventional and LD MOSFET's" IEEE Electron Device Letters, vol. EDL-5, no. 11, p.491, Nov., 1984.

[3.17] J. Scsrpulla, T.C. Mele, and J.P. Krusius, "Accurate Criterion for MOSFET Effective Gate Length Extraction Using the Capacitance Method," IEDM Tech. Dig., pp.722-725, 1987.

[3.18] Y.-R. Ma and K.-L. Wang, "A New Method to Electrically Determine Effective MOSFET Channel Width," IEEE Tran. on Electron Devices, vol. ED-29, no. 12, p.1825,

Dec., 1982.

[3.19] H.S. Lee, "An Analysis of the Threshold Voltage for Short-Channel IGFET's," Solid-State Electronics, vol. 16, pp.1407-1417, 1973.

[3.20] L.D. Yau, "A Simple Theory to Predict the Threshold Voltage of Short-Channel IGFET's," Solid-Stae Electronics, vol. 17, pp.1059-1063, 1974.

[3.21] J.A. Greenfield and R.W. Dutton, "Nonplanar VLSI Device Analysis Using the Solution of Poisson's Equation," IEEE Tran. on Electron Devices, vol. ED-27, no. 8, pp.1520-1532, Aug. 1980.

[3.22] T. Toyabe and S. Asai, "Analytical Models of Threshold Voltage and Breakdown Voltage of Short-Channel MOSFET's Derived from Two-Dimensional Analysis," IEEE J. Solid-State Circuits, vol. SC-14, p.375, April, 1979.

[3.23] K.N. Ratnakumar and J.D. Meindl, "Short-Channel MOST Threshold Voltage Model," IEEE J. Solid-State Circuits, vol. SC-17, p.937, Oct., 1982.

[3.24] D.R. Poole and D.L Kwong, "Two-Dimensional Analysis Modeling of Threshold Voltage of Short-Channel MOSFET's" IEEE Electron Device Letters, vol. EDL-5, pp.443-446, Nov. 1984.

[3.25] J.D. Kendall and A.R. Boothroyd, "A Two-Dimensional Analysis Threshold Voltage Model for MOSFET's" with Arbitrary Doped Substrate," IEEE Electron Device Letters, vol. EDL-7, pp.401-403, July 1986.

[3.26] Y.A. El-Mansy and A.R. Boothroyd, "A simple two-dimensional model for IGFET operation in the saturation region," IEEE Tran. Electron Devices, vol. ED-24, p. 254, Mar. 1977.

[3.27] T.Y. Chan, P.K. Ko, and C. Hu, "Dependence of channel electric field on device scaling," IEEE Electron Device Letters, vol. EDL-6, p. 551, Oct. 1985.

[3.28] K.W. Terrill, C. Hu, and P.K. Ko, "An analytical model for the channel electric field in MOSFET with graded-drain structure," IEEE Electron Device Letters, vol. EDL-5,

p. 440, Nov. 1984.

[3.29] H. Masuda, M. Nakai, and M. Kubo, "Characteristics and Limitations of Scaled-Down MOSFET's Due to Two-Dimensional Field Effect," IEEE Tran. on Electron Devices, vol. ED-26, p.105, June 1979.

[3.30] R.R. Troutman and S.N. Chakravarti, "Subthreshold Characteristics of Insulated-Gate Field Effect Transistors," IEEE Tran. on Circuit Theory, vol. CT-20, p.659, Nov., 1973.

[3.31] S.E. Laux and M.V. Fischetti, "Monte-Carlo Simulation of Submicrometer Si-n MOSFET at 77 and 300 K," IEEE Electron Device Letters, vol. EDL-9, p.467, 1988.

[3.32] G.A. Sai-Halasz, M.R. Wordeman, D.P. Kern, S. Rishton, and E. Ganin, "High Transconductance and Velocity Overshoot in NMOS Devices at the 0.1-μm Gate Length Level," IEEE Electron Device Letters, vol. EDL-9, p.464, 1988.

[3.33] P.K. Ko, "Hot-Electron Effects in MOSFETs," Ph.D Dissertation, Univ. of California, Berkeley, 1982.

[3.34] K.Y. Toh, P.K. Ko, and R.G. Meyer, "An engineering model for short-channel MOS devices," IEEE Jour. of Solid-State Circuits, p.950, Aug. 1988.

[3.35] G. Baccarani, M.R. Wordeman, and R.H. Dennard, "Generalized Scaling Theory and Its Application to a 1/4 Micrometer MOSFET Design," IEEE Tran. on Electron Devices, vol. ED-31, no. 4, pp.452-462, April 1984.

[3.36] W. Hansch and H. Jacobs, "Enhanced Transconductance in Deep-Submicrometer MOSFET," IEEE Electron Device Letters, vol. EDL-10, no. 7, p.285, July 1989.

[3.37] Y. El-Mansy, "MOS device and technology constraints in VLSI, "IEEE Tran. on Electron Devices, vol. ED-29, no. 4, p. 567, 1982.

[3.38] P.K. Chatterjee, W.R. Hunter, T.C. Holloway, and Y.T. Lin, "The impact of scaling laws on the choice of n-channel or p-channel for MOS VLSI," IEEE Electron Device Lett., vol. EDL-1, no. 10, p. 220, 1980.

[3.39] H. Shichijo, "A re-examination of practical performance limits of scaled n-channel and p-channel MOS devices for VLSI," Solid-State Electron, vol. 26, no. 10, p. 969, 1983.

[3.40] K.K. Ng and W.T. Lynch, "The impact of intrinsic series resistance on MOSFET scaling," IEEE Tran. on Electron Device, vol. ED-34, no. 3, p. 503, 1987.

[3.41] K. Mayaram, J. Lee, and C. Hu, "A model for the electric field in lightly doped drain structures," IEEE Tran. on Electron Devices, ED-34, vol. 7, p. 1509, 1987

[3.42] C. Hu, S. Tam, F.-C. Hsu, P.K. Ko, T.Y. Chan, and K.W. Terrill, "Hot-Electron-Induced MOSFET Degradation - Model, Monitor, and Improvement," IEEE Tran. on Electron Devices, vol. ED-32, p.375, Feb. 1985.

[3.43] T.Y. Chan, Electron Device Letters, vol. EDL-6, p.551, Oct. 1985.

[3.44] M. Kakumu, IEDM Tech. Digs., p.399, 1986.

[3.45] C.G. Hwang, R.W. Dutton, J.M. Higman, and K. Hess, "Accurate Modeling of Impact Ionization Effect in Submicron MOSFET," IEEE Tran. on Electron Devices, vol. ED-34, no. 11, p.2385, Nov. 1987.

[3.46] 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 Tran. on Electron Devices, vol. ED-35, no. 5, pp.689-697, May 1988.

[3.47] M. Tomizawa, K. Yokoyama, and A. Yoshii, "Nonstationary Carrier Dynamics in Quarter-Micron Si MOSFET's," IEEE Tran on Computer-Aided Design, vol. CAD-7, no. 2, pp.254-258, 1988.

[3.48] J. Chung, M.-C. Jeng, G. May, P.K. Ko, and C. Hu, "Hot-Electron Currents in Deep-Submicrometer MOSFETs," IEDM Tech. Digs., p.200, 1988.

[3.49] F.C. Hsu and S. Tam, "Relationship between MOSFET Degradation and Hot-Electron Induced Interface-State Generation," IEEE Electron Device Letters, vol. EDL-5, p.50, 1984.

[3.50] H. Gesch, J.-P. Leburton, and G.E. Dorda, "Generation of Interface States by Hot Hole Injection in MOSFET's," IEEE Tran. on Electron Devices, vol. ED-29, no.5, p.913, May 1982.

[3.51] T.Y. Chan, P.K. Ko, and C. Hu, "A Simple Method to Characterize Substrate Current in MOSFET's," IEEE Electron Device Letters, vol. EDL-5, no. 12, p.505, 1984.

[3.52] M.-C. Jeng, J. Chung, A.T. Wu, T.Y. Chan, J. Moon, G. May, P.K. Ko, and C. Hu, "Performance and Hot-Electron Reliability of Deep-Submicrometer MOSFET's," IEDM Tech. Digs., p.710, 1987.

Chapter 4

DEEP-SUBMICROMETER MOSFET DESIGN

Although deep-submicrometer MOSFETs with excellent characteristics have been demonstrated and the basic device physics of these devices has been shown to be essentially unchanged, deep-submicrometer devices are still restricted to device level studies because no design guidelines are available. For longer-channel devices, previous studies [4.1-4.3] have proposed design guidelines based mainly on the threshold voltage shift due to short-channel effects, subthreshold current, and hot-electron reliability considerations, but the different trade-offs between reducing oxide thickness, channel length, and power supply are still not clear. This chapter attempts to provide comprehensive design guidelines for MOSFETs in the deep-submicrometer regime by investigating a wide range of performance and reliability constraints on device dimensions and power supply. The mechanisms examined in this study are: short-channel and drain-induced-barrier-lowering (DIBL) effects, the punchthrough and gate-induced drain leakage (GIDL) [4.4-4.6] currents, hot-electron reliability, time-dependent dielectric breakdown (TDDB) [4.7-4.9], current-driving capability, voltage gain, and switching speed. Using this set of performance and reliability constraints, design curves are developed based on measurements of n-channel non-LDD deep-submicrometer devices. The relative importance of each mechanism for a given technology and design criteria is compared. The five basic parameters in MOS scaling are: effective channel length L_{eff}, oxide thickness T_{ox}, power supply V_{DD}, junction depth X_j, and channel doping concentration N_{SUB}. For most technologies, X_j is relatively constant compared to other parameters. Once T_{ox} and the threshold voltage is determined, N_{SUB} is fixed. Therefore, only L_{eff}, T_{ox}, and V_{DD} are considered in this study. X_j is fixed to about 0.18μm and N_{SUB}'s are adjusted such that the long-channel threshold voltages for all oxide thicknesses are around 0.65V. The design considerations included are divided into two categories. One sets device limitations and the other sets performance constraints.

4.1 Device limitations

4.1.1 Threshold voltage shift

Fig. 4.1a and 4.1b show the threshold voltage shift due to short-channel and DIBL effects. As mentioned in section 3.2 that these two effects are essentially one, they are separated here for easy description. In Fig. 4.1b, only T_{ox} = 8.6nm data are shown. Similar results are also observed for other oxide thicknesses. The threshold voltage shift ΔV_{th} is defined as the difference between the measured threshold voltage at a given drain voltage and its corresponding lon-channel value (V_{th0}) at a drain voltage of 50mV. Although the threshold voltage model derived in section 3.2 showed that ΔV_{th} deviates from a simple exponential expression for ΔV_{th} > 0.1V, straight lines are drawn to fit measured data for simplicity. The dashed lines in Fig. 4.1-4.8 demarcates the performance and reliability criteria (Table 4.1) used in this study to obtain the design curves (Fig. 4.9-4.11); the arrows indicate the acceptable regions. As an example, for T_{ox} = 8.6nm and V_{DD} = 3V, the minimum allowable L_{eff} in the circuit is about 0.28μm purely based on the threshold voltage shift consideration.

4.1.2 Off-state leakage current

The off-state leakage current is also sensitive to the short-channel effects and was used as one of the criteria for MOSFET miniaturization [4.2]. As shown in the insert of Fig. 4.2a, off-state leakage current is composed of two main components: punchthrough current (I_{PT}) and gate-induced drain leakage current (I_{GIDL}). The punchthrough current is the leakage current between the source and the drain. The gate-induced drain leakage current is the drain-to-substrate leakage due to band-to-band tunneling between n^+ and p regions. I_{PT} increases with decreasing channel length because of the threshold voltage reduction and the increase in subthreshold swing. I_{GIDL} is, however, independent of L_{eff} and is determined by T_{ox} and the power supply used. In Fig. 4.2, the off-state leakage currents were measured at V_{th0} - 0.6V for all device dimensions and drain voltages. A gate voltage of V_{th0} - 0.6V was used to elim-

inate any effect caused by the variations in threshold voltage between different gate oxide thicknesses. The punchthrough current dominated regions are indicated by open symbols; the GIDL current dominated regions are indicated by closed symbols. The current level of the experimental data is clamped at a lower bound of 0.5pA/μm due to limits in the measurement resolution.

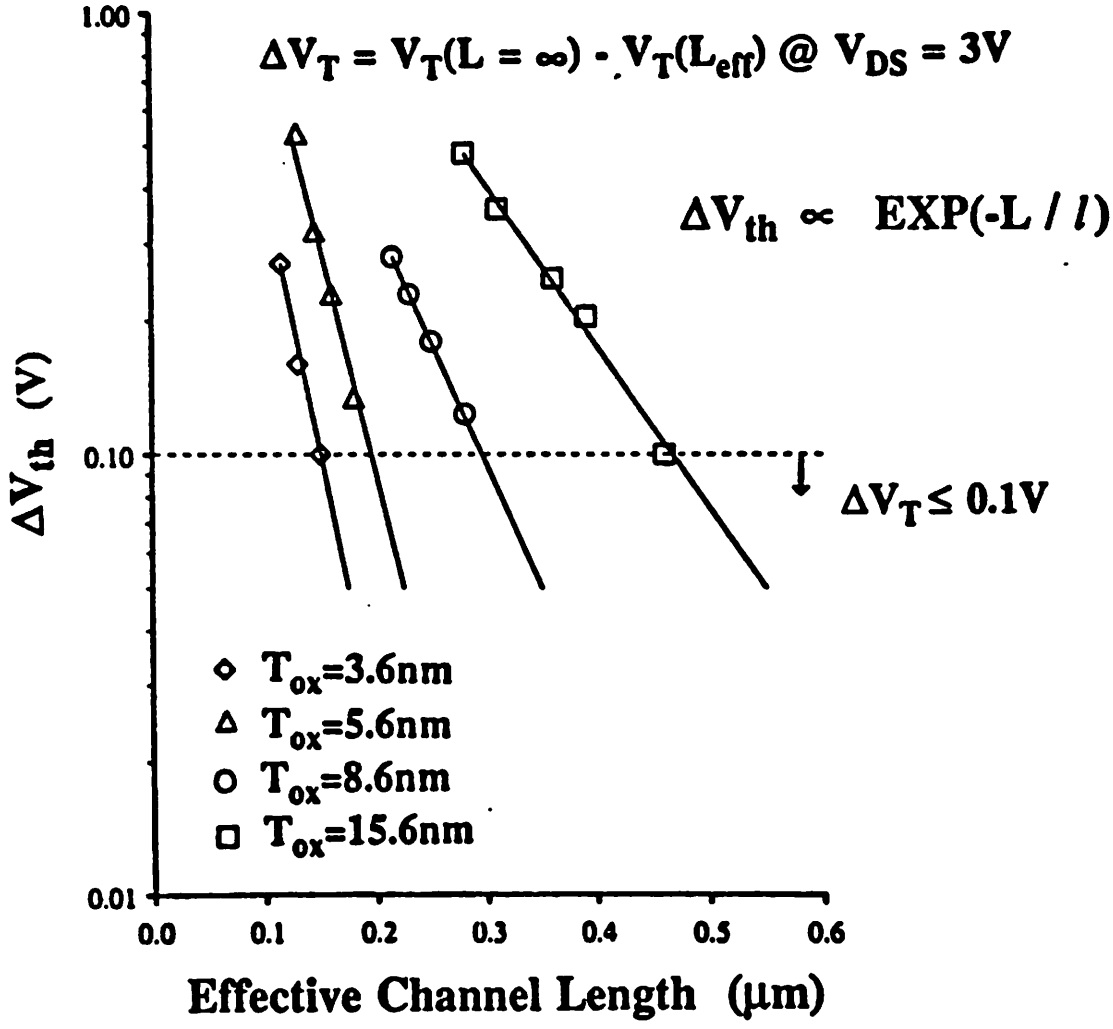


Fig. 4.1(a) Threshold voltage reduction (ΔV_{th}) versus effective channel length at V_{DS} = 3V for four oxide thicknesses.

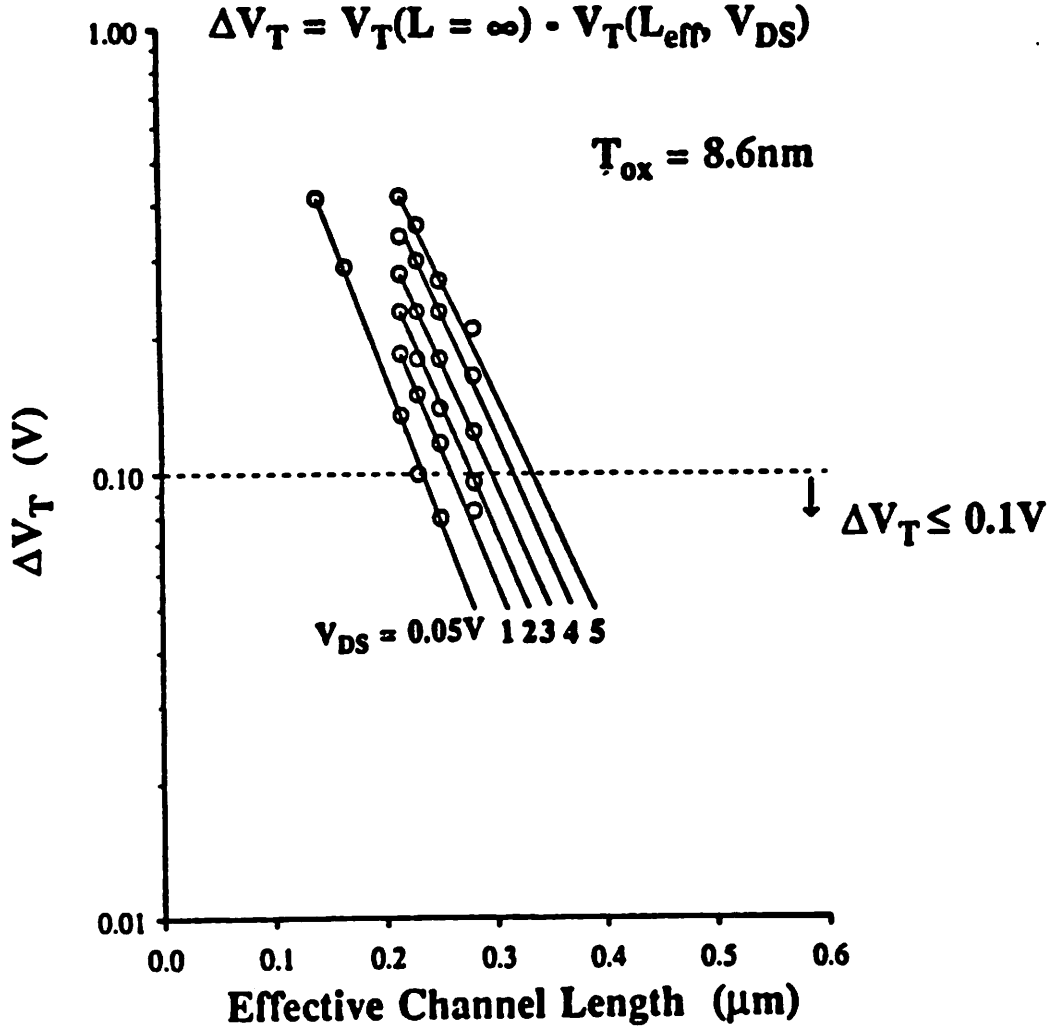


Fig. 4.1(b) Threshold voltage reduction (ΔV_{th}) versus effective channel length for T_{ox} = 8.6nm at different drain voltages.

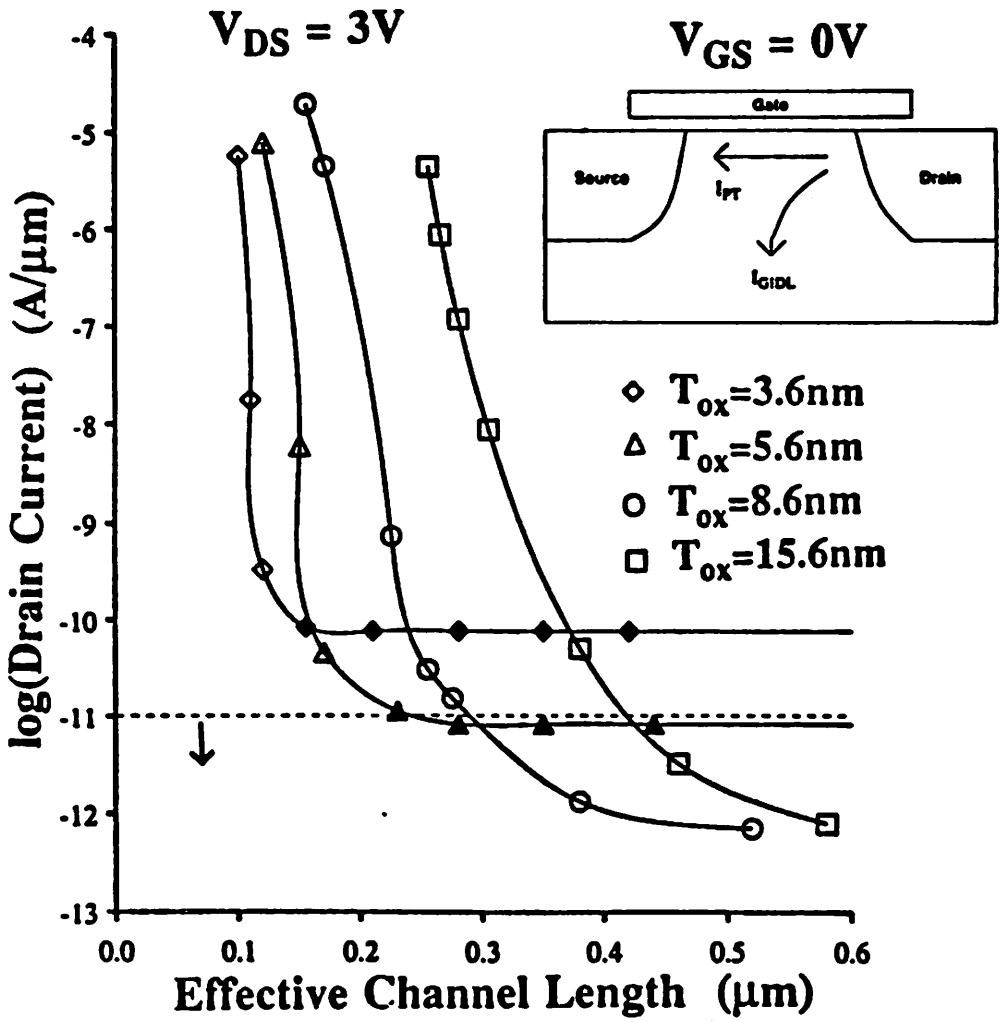


Fig. 4.2(a) Off-state leakage current versus effective channel length measured at V_{DS} = 3V for four oxide thicknesses. The off-state leakage current has two components: punchthrough current and gate-induced drain leakage current. The insert shows the different paths for these two components.

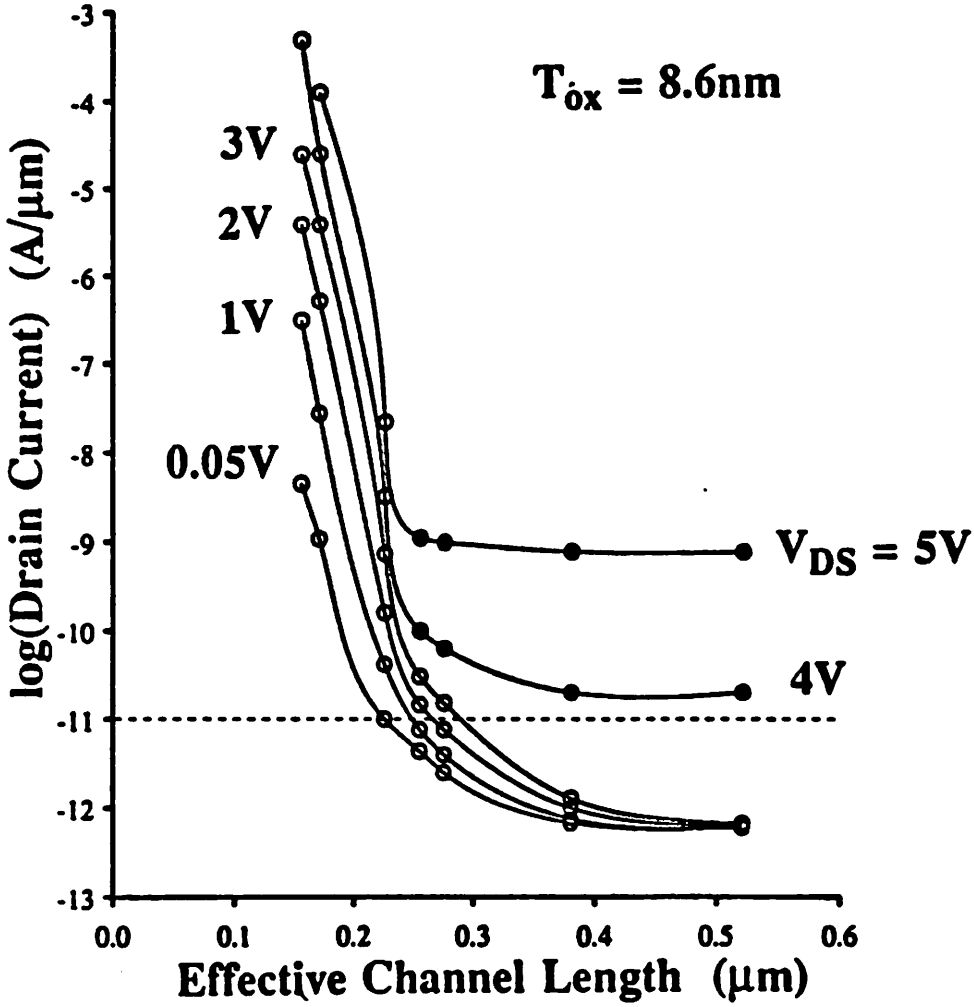


Fig. 4.2(b) Off-state leakage current versus effective channel length for T_{ox} = 8.6nm at different drain voltages.

4.1.3 Hot-electron reliability

Recent studies showed that digital circuits are fairly robust to hot-electron effects [4.10]. Therefore, the definition of device lifetime in section 3.5.1 (3% drain current change in the linear region) is very tight for most applications. In this chapter, a more relaxed definition, 10% drain current reduction in the linear region, is used for device lifetime. Fig. 4.3 displays the extrapolated maximum allowable power supply voltage to ensure a 10-year device lifetime [4.11] as a function of channel length for four oxide thicknesses. For a given substrate current, thinner oxide devices exhibit less degradation than those with thicker oxides [4.12]. However, for a given drain bias, thinner gate oxide devices also exhibit greater peak substrate current than those with thicker oxides (see Fig. 3.18). These two counteracting trends explain why 8.6nm gate oxide devices show a slightly smaller minimum channel length than those of 5.6nm and 15.6nm gate oxides devices at a power supply of 3V. According to Fig. 4.3, at a power supply of 3.3V and with effective channel length larger than 0.5μm, LDD may not be needed.

4.1.4 Breakdown voltage

Fig. 4.4 shows breakdown voltage versus effective channel length for different oxide thicknesses. The breakdown voltage is defined as the minimum voltage of the c-shaped breakdown curve shown in the insert. Comparing Fig. 4.3 and Fig. 4.4, it is found that the breakdown voltage is about 1V to 2V higher than the maximum allowable power supply set by hot-electron requirements. Although under normal operations the breakdown will not be a limiting mechanism in MOS scaling, it sets an upper bound to the burn-in voltage.

4.1.5 Time-dependent dielectric breakdown

Based on a defect-density model, a technique to predict oxide breakdown statistics has been developed [4.9]. Plotted in Fig. 4.5 is the maximum allowable supply voltage to ensure 10-year lifetime at 125°C versus oxide thickness for two defect densities. Because oxide quality is a sensitive function of the device fabrication process, the oxide reliability results used in this study should be viewed as a rough approximation only. Other fabrication technologies can

yield a higher quality oxide with a lower defect-density than is observed in this study (1.0 cm^{-2}). Listed in Table 1 is the oxide reliability criterion used in Fig. 4.9-4.11.

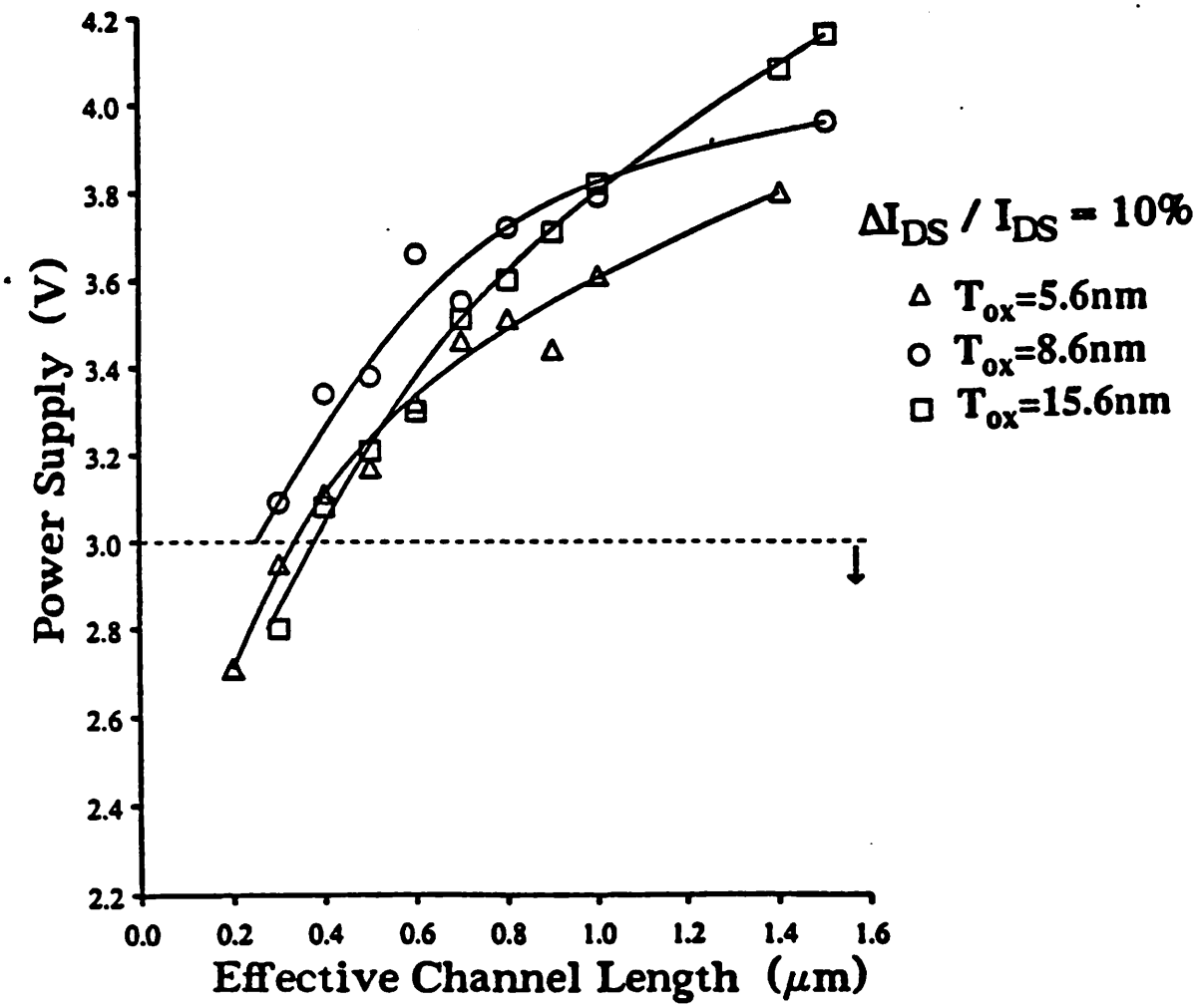


Fig. 4.3 Maximum allowable power supply to ensure 10-year device lifetime due to hot-electron effects versus effective channel length for several oxide thicknesses. The device lifetime is defined as 10% forward drain current degradation in the linear region. The dashed line indicates the criterion used to obtained the design curves.

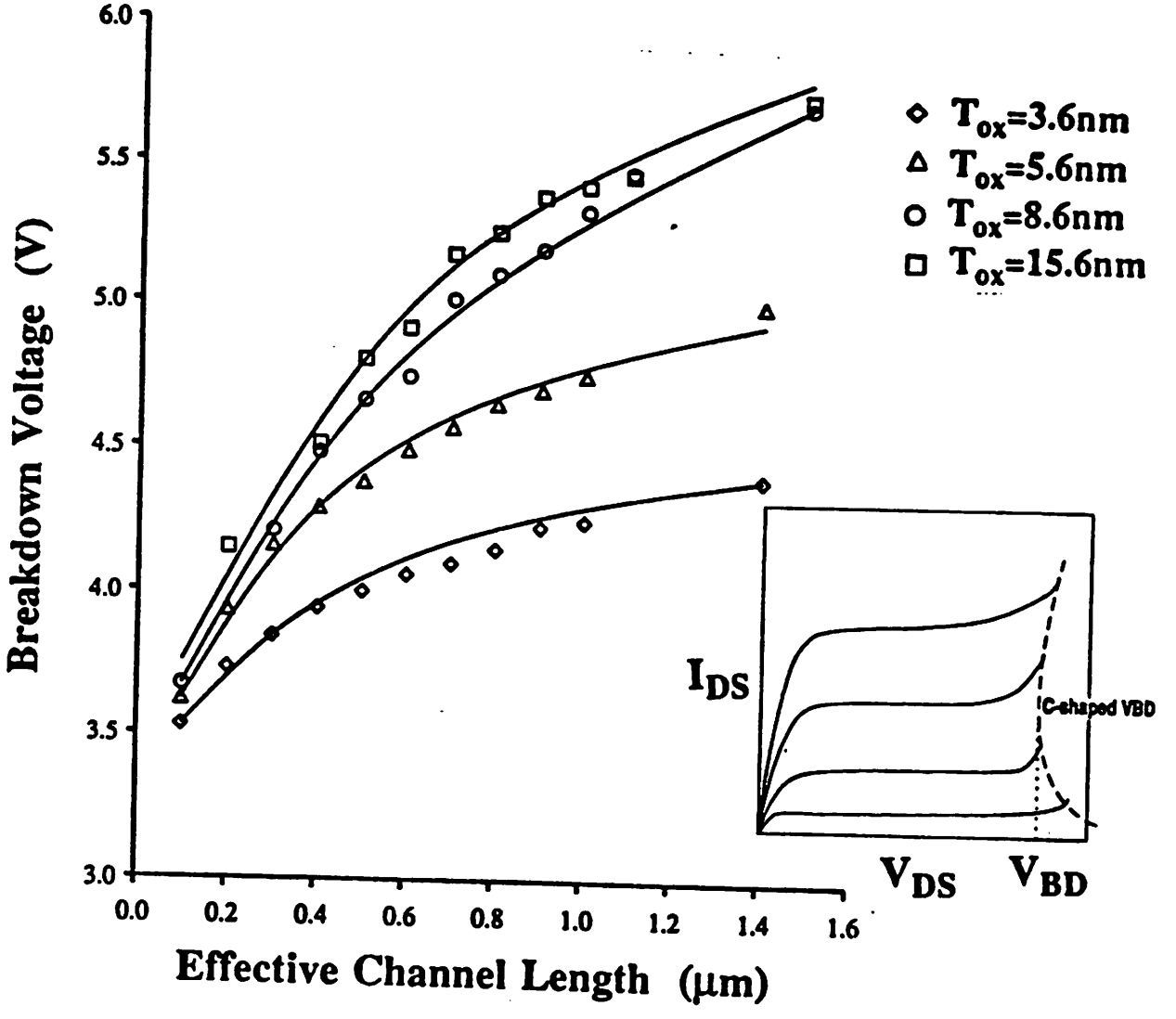


Fig. 4.4 Breakdown voltage versus effective channel length for several oxide thicknesses.

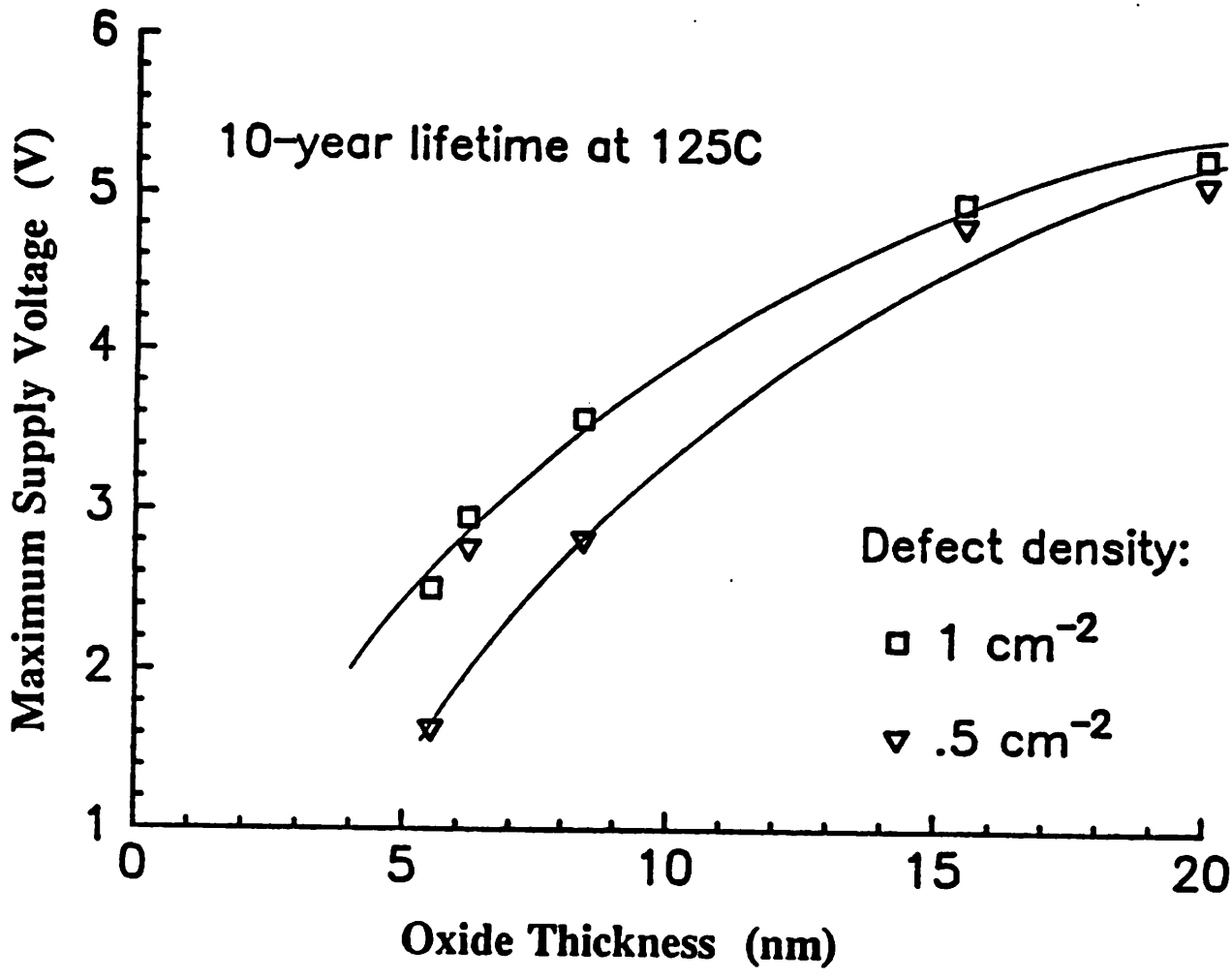


Fig. 4.5 Maximum allowable power supply to ensure 10-year device lifetime due to time-dependent dielectric breakdown versus oxide thickness for two defect densities.

4.2 Performance constraints

4.2.1 Current-driving capability:

Fig. 4.6 shows drain saturation current I_{DSAT} versus effective channel length. The drain saturation current is measured at V_{GS} = 3V and V_{DS} = 3V as shown in the insert. As expected, the current-driving capability for a given gate oxide increases as the channel length decreases. However, because of mobility degradation due to high vertical fields, the current-driving capability tends to saturate at very thin gate oxides unless the channel length is very small such that all carriers in the channel are moving with the saturation velocity. The high channel doping concentration required to achieve the required threshold voltage for thin oxide devices also degrades carrier mobility. The sharp increase in I_{DSAT} at very short-channel lengths is mainly caused by the threshold voltage reduction due to short-channel effects.

4.2.2 Voltage gain

In Fig. 4.7, the peak voltage gain (solid lines) measured near V_{GS} = 0V and the gain (alternated lines) measured at $V_{GS}-V_T$ = 0.3V, where most analog circuits are biased, are plotted for various device dimensions. The voltage gain is defined as $g_m R_{out}$, where g_m is the measured transconductance and R_{out} is the output resistance. Since both g_m and R_{out} are higher for thinner gate oxide devices, the voltage gain increases as oxide thickness decreases. The sharp decrease of the gain at very short channel lengths is caused by bulk punchthrough which significantly reduces R_{out}.

4.2.3 Switching speed

Because no CMOS circuits were available, the switching speed studies were achieved through simulations on CMOS ring oscillator delay time. To ensure high confidence, a MOSFET model accurate down to quarter-micron channel length was used in the simulation. This model is described in chapter 5. Fig. 4.8 shows SPICE simulated delay time of an 11-stage CMOS ring oscillator with a 0.1pF load capacitor on each stage for different oxide thicknesses,

channel lengths, and power supplies. The channel width is 15μm for n-channel and 30μm for p-channel devices. The overlap between the gate and the source (drain) is 0.05μm. As oxide thickness decreases, the gate capacitance eventually becomes larger than the load capacitance. However, because I_{DSAT} tends to saturate at thiner gate oxide (see Fig. 4.6), the capacitance charging rate does not increase as rapidly as the gate capacitance. These two mechanisms explain why the delay time does not continue to decrease with diminishing oxide thickness in Fig. 4.8a. Because the drain current saturates at larger gate voltage, the delay time also saturates at larger power supply.

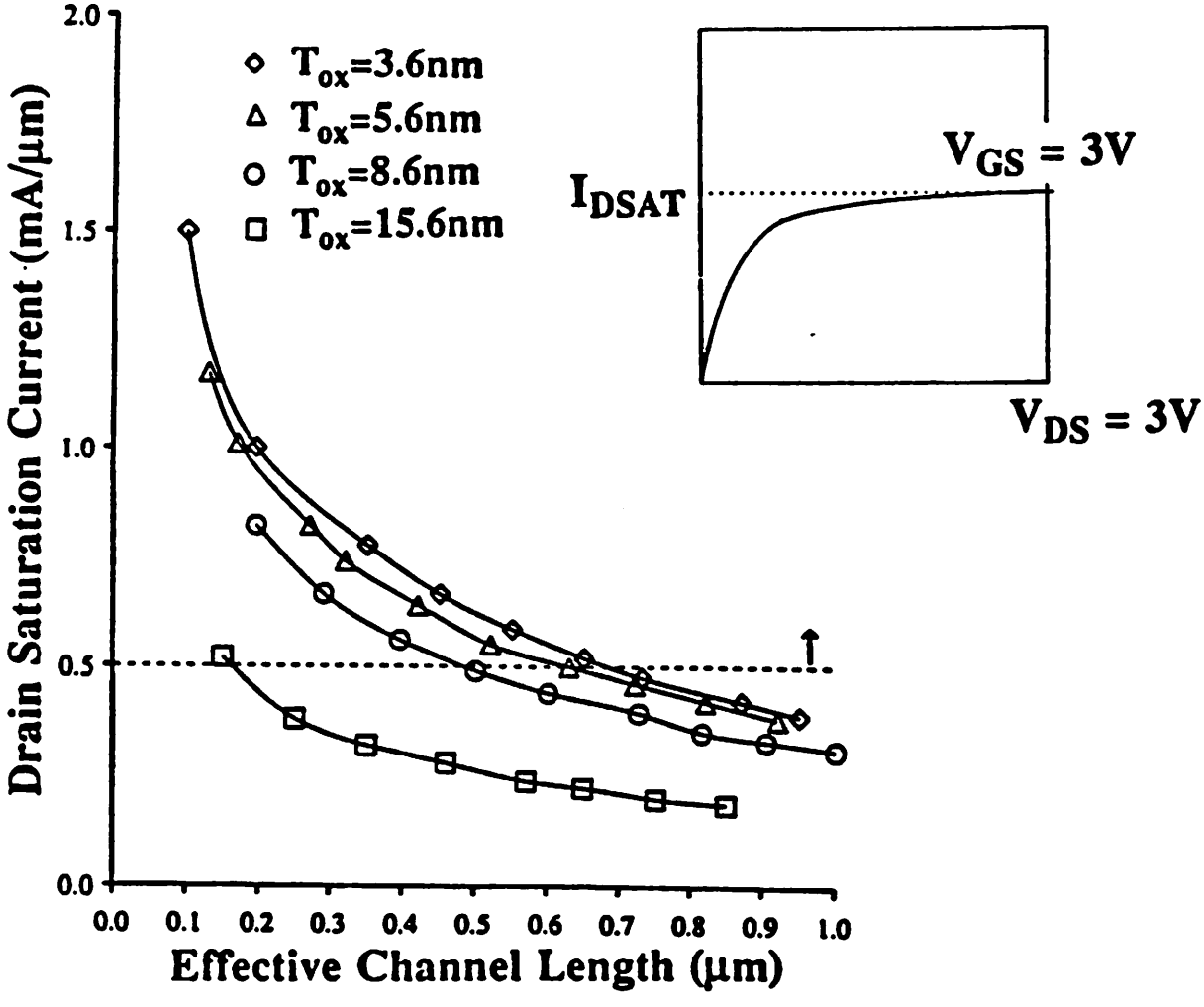


Fig. 4.6 Drain saturation current versus effective channel length for four oxide thicknesses. The insert shows the bias conditions when the current was measured.

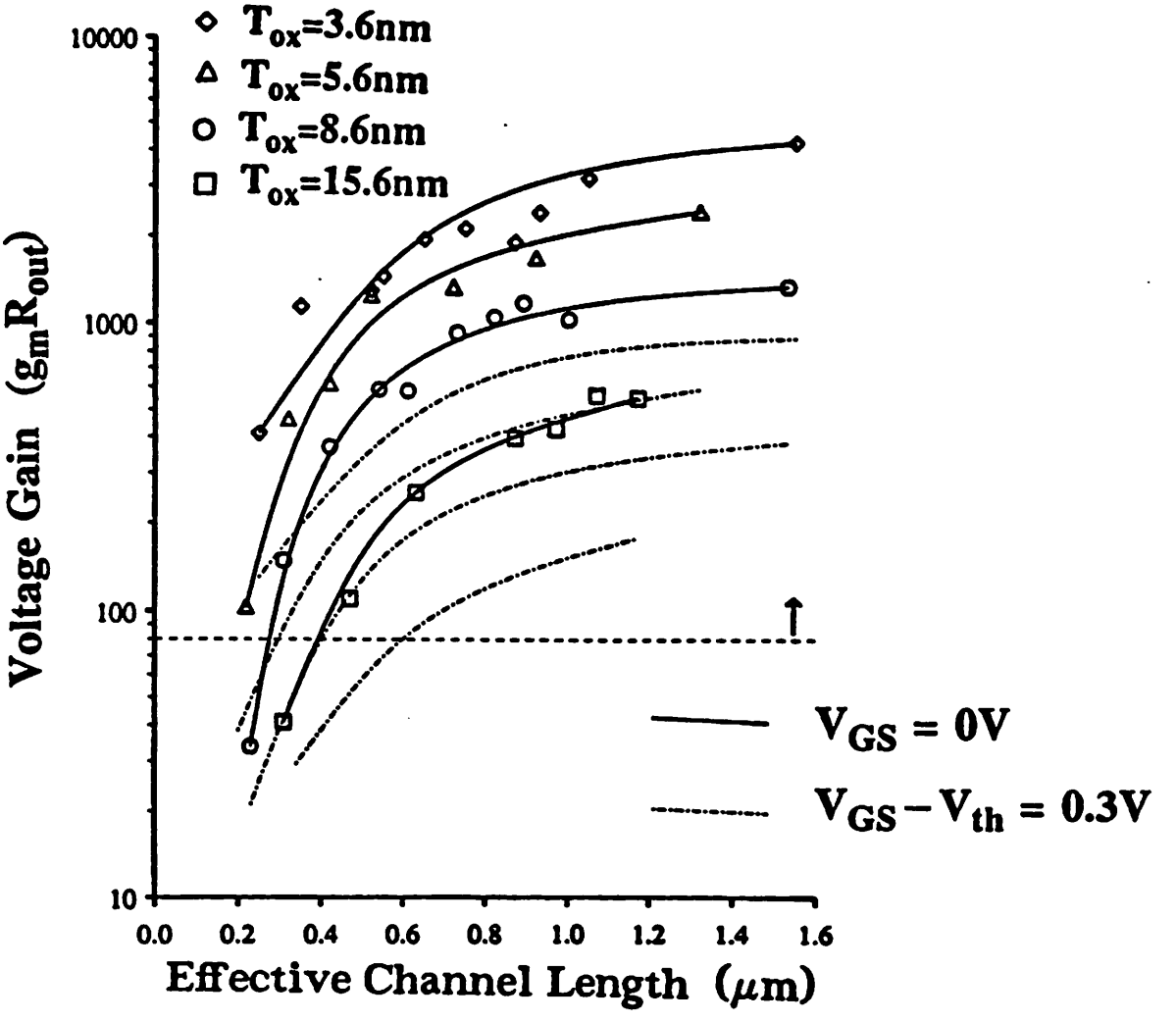


Fig. 4.7 Single stage voltage gain (g_mR_{out}) versus effective channel length for four oxide thicknesses. The solid lines indicate the maximum available gain and the alternated lined are the gain measured at $V_{GS}-V_{th}$ = 0.3V.

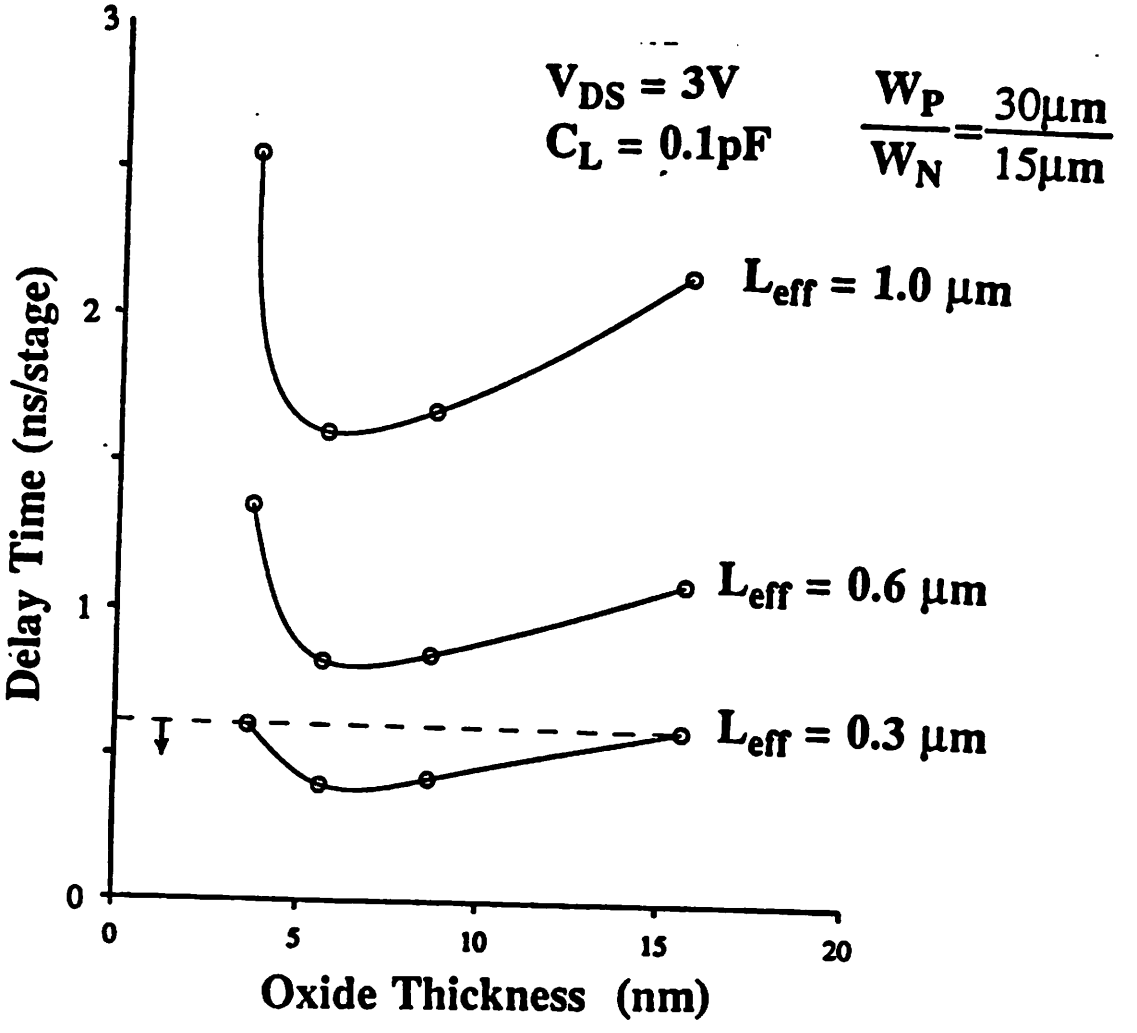


Fig. 4.8(a) SPICE simulated CMOS ring oscillator delay time operated at a power supply of 3V versus oxide thickness for several channel lengths. The load capacitance is 0.1pF on every stage.

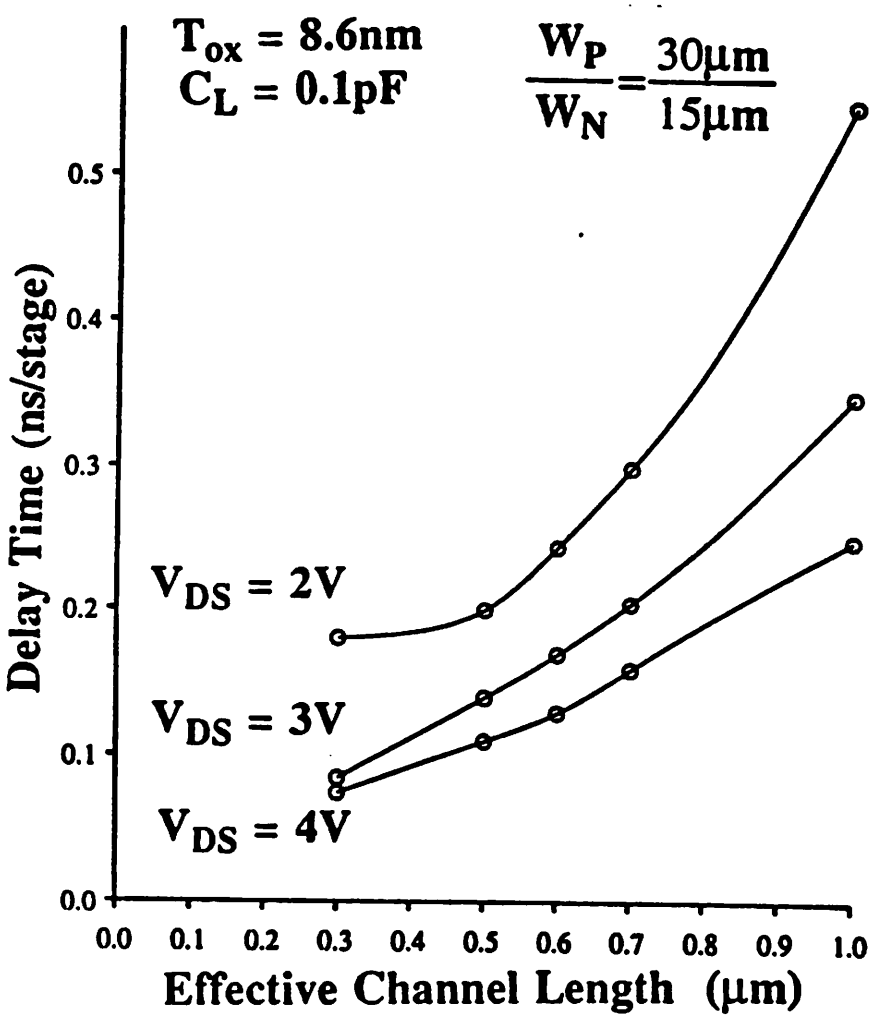


Fig. 4.8(b) SPICE simulated CMOS ring oscillator delay time versus effective channel length for T_{ox} = 8.6nm operated at different supply voltages.

4.3 Design guidelines

Based on the experimental results presented in sections 4.1 and 4.2, various design curves were developed. As mentioned before, since oxide thickness, channel length, and supply voltage are the key design parameters in this study, three types of design curves are provided for maximum flexibility (Fig. 4.9-4.11). Each type of curve fixes one parameter while varying the other two. The intersection of these performance and reliability curves (shaded area) indicates the region of allowable device dimensions and/or power supply for both digital and analog applications under some design specifications. Table 4.1 summarizes the meanings of the symbols in Fig. 4.9-4.11 and lists the performance and reliability criteria used in developing these design curves.

TABLE I		
Symbol	Description	Criterion
ΔV_T	Threshold voltage shift	$\leq$ 0.1V
I_{PT}	Punchthrough current at V_{GS} = 0V	$\leq$ 10pA/μm
G	Voltage Gain at $V_{GS}-V_T$ = 0.3V	$\geq$ 80
I_{GIDL}	Gate-induced-drain-leakage current	$\leq$ 10pA/μm
τ_{HE}	Lifetime (10% I_{DS} degradation)	$\geq$ 10 years
τ_d	delay time	$\leq$ 120ps/stage
I_{DSAT}	Drain saturation current	$\geq$ 0.5mA/μm
TDDB	Time Dependent Dielectric Breakdown on 5mm^2 area for 10 years	$\leq$ 1%

Table 4.1 Design criteria for design curves.

4.3.1 Oxide thickness versus channel length

Fig. 4.9 shows design curves where the optimal oxide thickness for this technology is plotted versus channel length for a power supply of 3V. All the curves corresponding to the constant contours of different design considerations use the criteria listed in Table 4.1. For example, the curve marked by ΔV_{th} was obtained from Fig. 4.1. The gate oxide and effective channel length combinations along this curve will give 0.1V threshold voltage shift. The arrows indicate the acceptable regions. Devices with T_{ox} and L_{eff} in the acceptable region have less threshold voltage shift than 0.1V. But the channel length can not be too long due to I_{DSAT} and switching speed requirements, which set upper limits to device dimensions. The intersection of all acceptable regions forms design windows (shaded regions). Because of different design requirements for analog and digital circuits, different bounds (different windows) are used for these two applications. The breakdown curve is not included in Fig. 4.9-4.11 because it is not a limiting mechanism under normal device operation.

According to the design windows, the minimum gate oxide thickness at this supply voltage is limited to 5.6nm by the gate-induced drain leakage current, and may be limited by the time-dependent dielectric breakdown for technologies with less robust oxide. For digital applications, depending upon the oxide thickness, the minimum channel length is determined by either the threshold voltage shift or by the hot-electron reliability criterion; the minimum allowable channel length is found to be 0.26μm at T_{ox} = 7.8nm. The largest channel length is about 0.45μm limited by the switching speed requirement. For analog applications, the minimum channel length is about 0.31μm at T_{ox} = 6.3nm limited by the voltage gain requirement. It should be kept in mind that the minimum (maximum) device dimensions mentioned here refer to the "worst case" conditions. For example, if the channel length variation for a given process is ±0.1μm, then a minimum channel of 0.26μm implies a nominal channel length of 0.36μm. The same argument also applies to the oxide thickness. Another advantage of these design curves is that the relative importance of each mechanism can be identified for any device dimensions which makes design trade-offs very clear and provides a direction for future tech-

nology development.

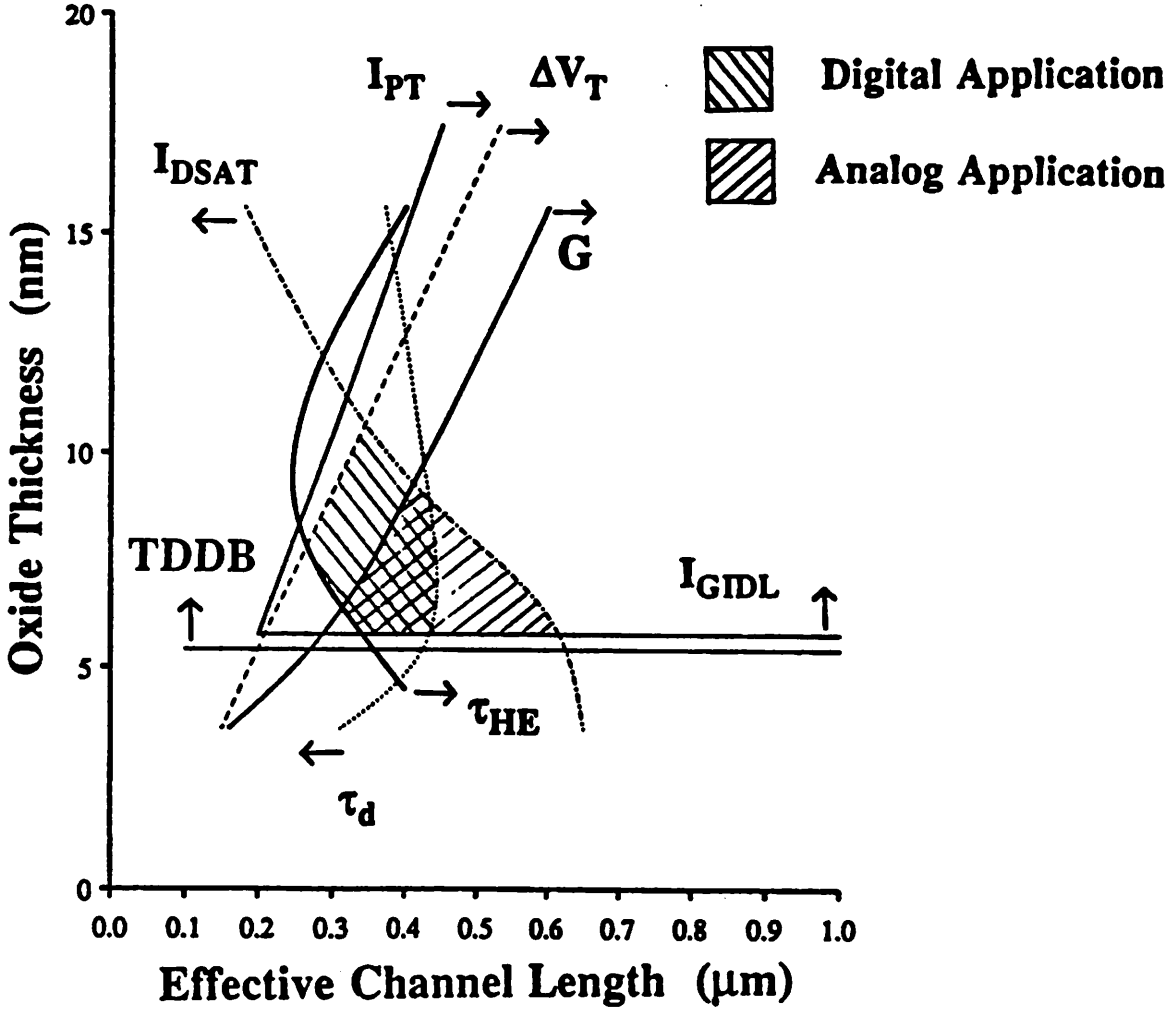


Fig. 4.9 Design curves for a power supply of 3V. Different curves correspond to the constant contour of each design consideration with the criteria listed in Table 4.1. The shaded areas indicate the allowable regions.

4.3.2 Power supply versus channel length

Fig. 4.10 shows design curves where the suitable power supply for this technology is plotted against channel length for T_{ox} = 8.6nm. Again, each curve corresponds to the constant contour of a design criterion listed in Table 4.1 and shaded regions are the allowable design windows. At this oxide thickness, the maximum power supply voltage is limited by the hot-electron reliability while the minimum power supply voltage is limited by the switching speed requirement. The minimum allowable channel lengths is about 0.28μm for digital applications limited by the threshold voltage shift, and is about 0.36μm for analog applications limited by the voltage gain. These values are roughly independent of the power supply voltage because the peak voltage gain is independent of power supply and short-channel effect is much more sensitive to the channel length than to the power supply. The maximum channel length is about 0.48μm. At a power supply of 3.3V, hot-electron reliability does not pose a problem to devices with channel length longer than 0.4 μm implying that LDD may not be needed, but with the burn-in consideration, longer channel length or LDD may still be necessary.

4.3.3 Power supply versus oxide thickness

The last type of curves is the design curves for L_{eff} = 0.3μm. Fig. 4.11 shows power supply versus oxide thickness of each design consideration for L_{eff} = 0.3μm. At this channel length, the maximum power supply is limited to about 3V due to hot-electron reliability, no matter what oxide thickness is used. The minimum power supply is determined by the speed requirement, about 2V at T_{ox} = 4.0nm. The maximum T_{ox} is about 9nm for digital applications and 6.5nm for analog applications due to voltage gain requirement.

4.3.4 Junction depth

Although the junction depth has been fixed at 0.18μm in this study, with slight modifications, the design curves in Fig. 4.9-4.11 can be extended to other junction depths. For example, if the junction depth is decreased, short-channel and DIBL effects and punchthrough currents would diminish. However, hot-electron reliability would degrade due to the increase

in the peak channel electric field.

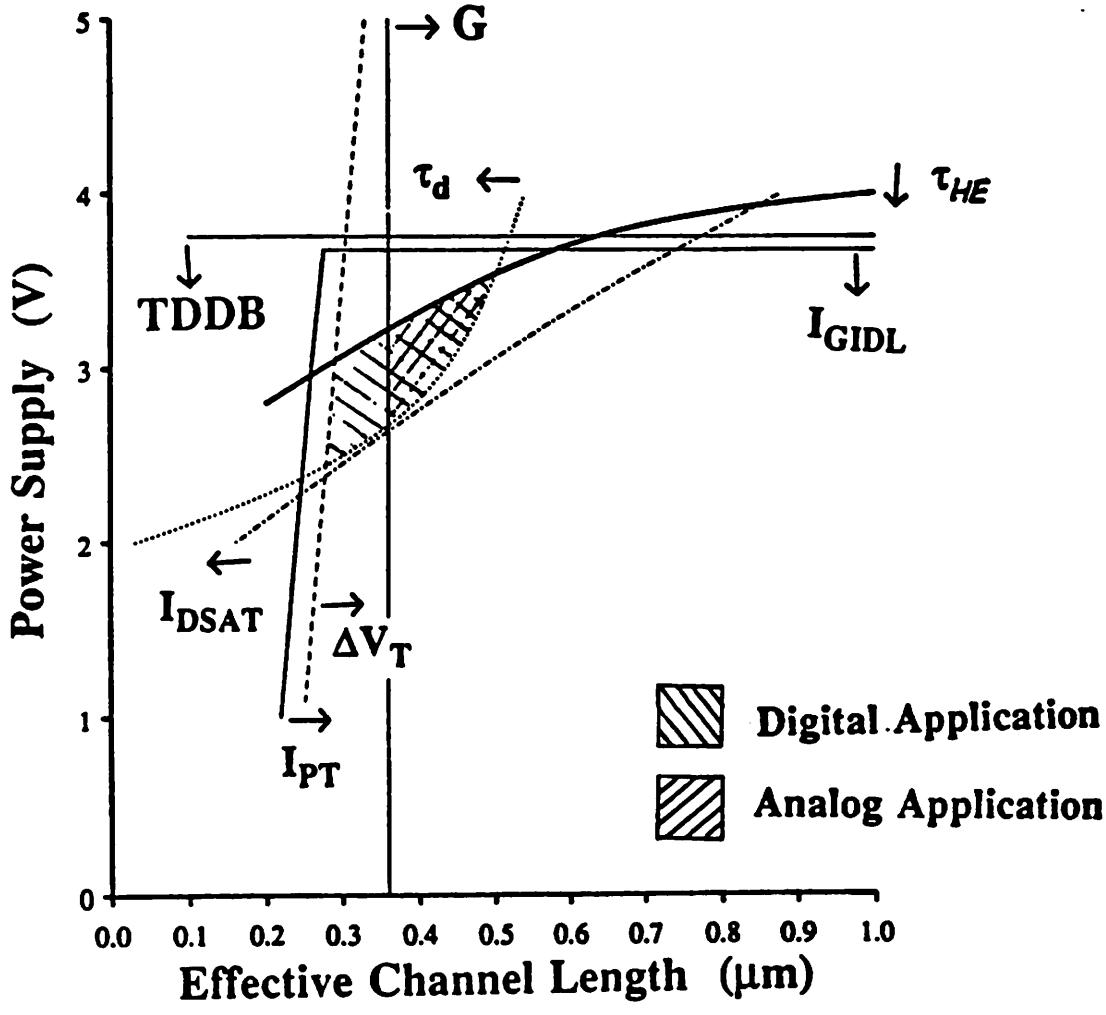


Fig. 4.10 Design curves for T_{ox} = 8.6nm. Different curves correspond to the constant contour of each design consideration with the criteria listed in Table 4.1. The shaded areas indicate the allowable regions.

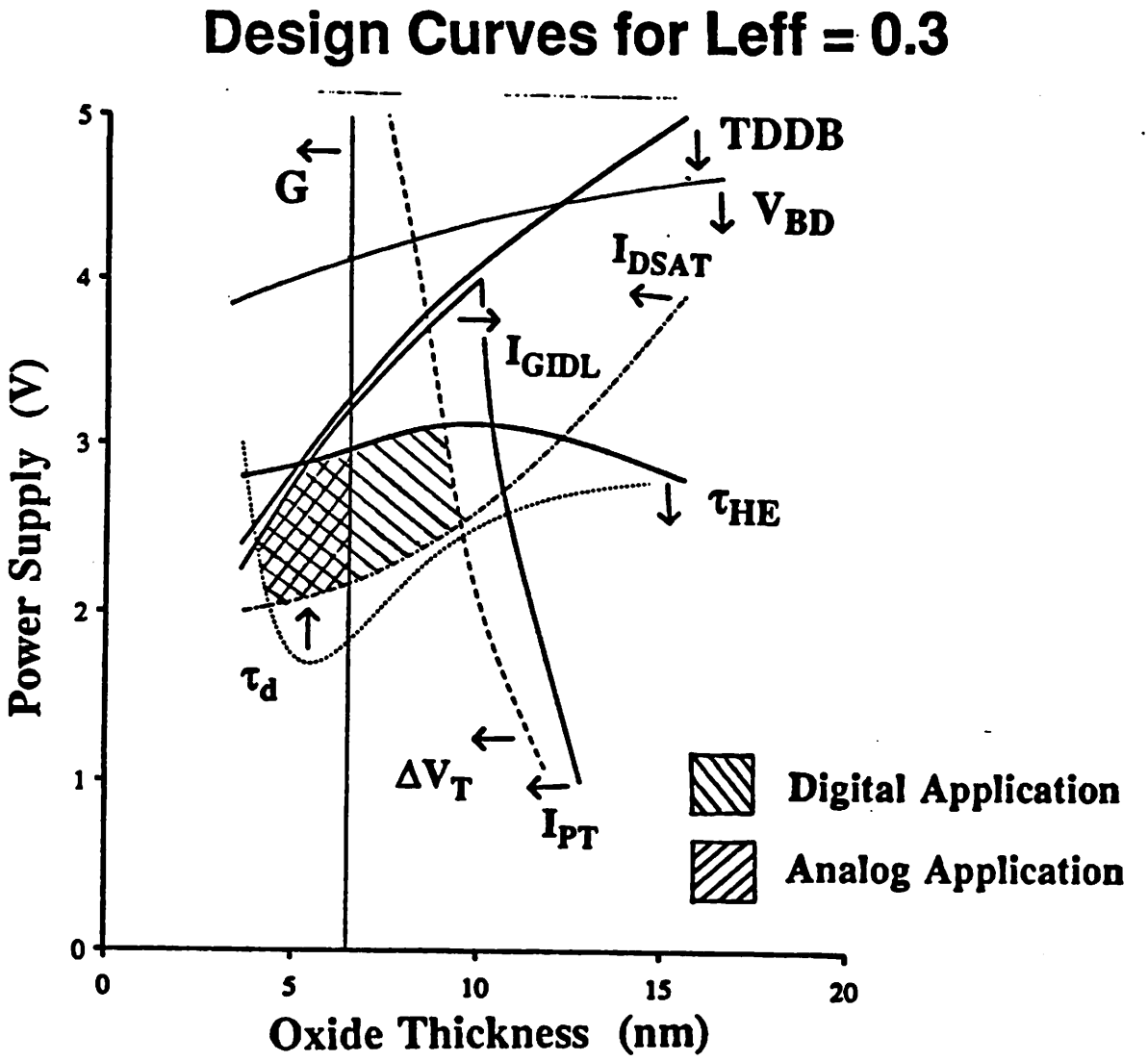


Fig. 4.11 Design curves for L_{eff} = 0.3μm. Different curves correspond to the constant contour of each design consideration with the criteria listed in Table 4.1. The shaded areas indicate the allowable regions.

4.3.5 Other power supply and device dimensions

Design curves for other power supply, T_{ox}, and L_{eff} values can be obtained with the same approach. Similar design curves as those in Fig. 4.9 for a power supply of 3.3V are shown in Fig. 4.12. Since the device lifetime is more sensitive to the power supply than other mechanisms (see Fig. 4.1-4.8), this fact is reflected by the large shift on the τ_{HE} cureve in Fig. 4.12 compared to that in Fig. 4.9. The design windows are smaller and shift toward longer channel and thicker oxide directions as expected.

4.3.6 Other technologies

The same methodology used to derive design curves shown in Figs. 4.9-4.11 can also be extended to any technology, including p-channel and LDD devices. For example, with LDD devices, short channel, DIBL, and GIDL effects would be less severe and the hot-electron lifetime would be longer. However, current-driving capability and gain would decrease due to the increase in source/drain resistance. Therefore, the design windows in Fig. 4.9 will move toward the lower left, i.e., shorter channel length and thinner oxide as expected. The LDD effects on Fig. 4.10 and 4.11 can also be analogized.

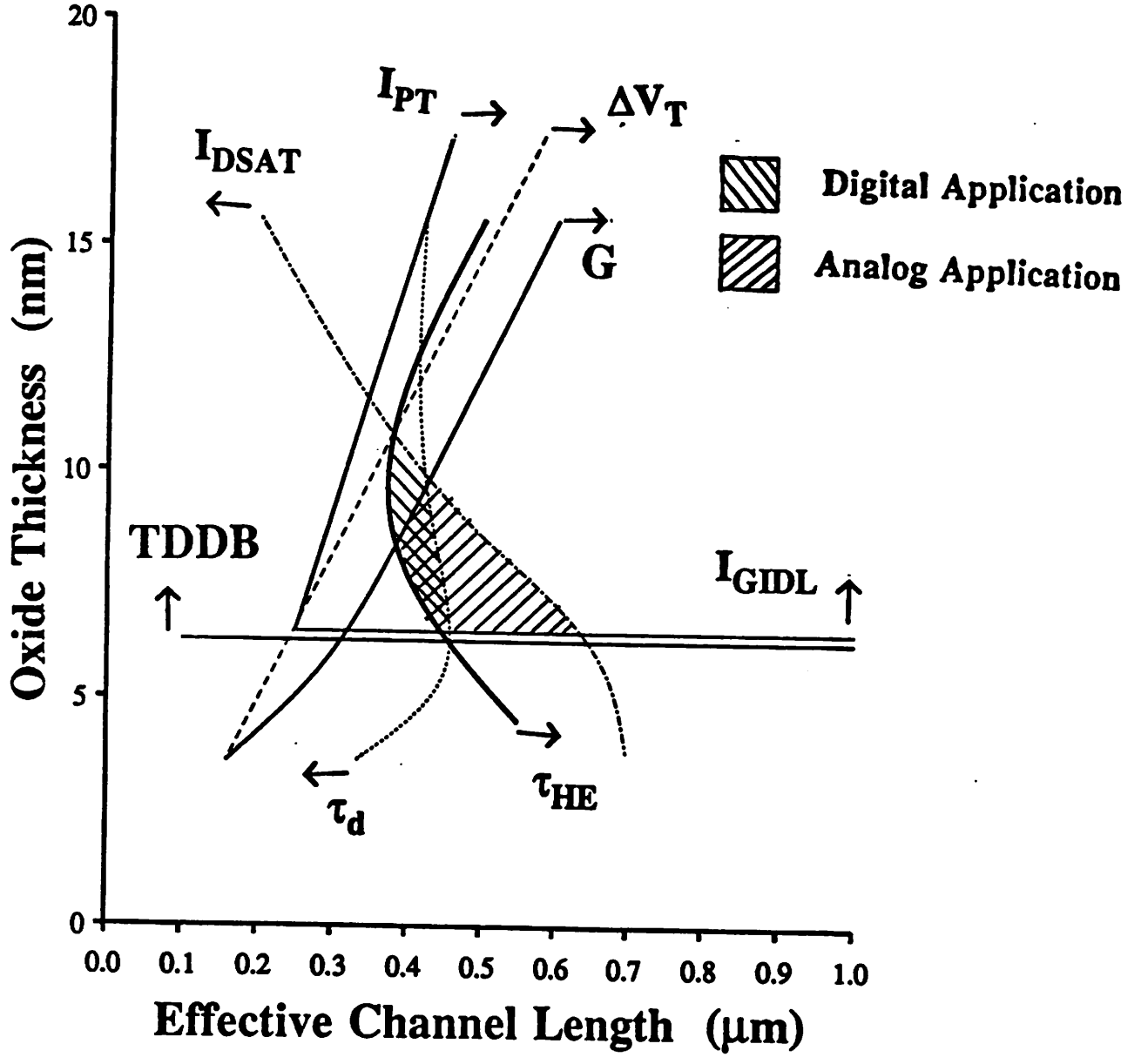


Fig. 4.12 Design curves for a power supply of 3.3V.

4.4 References

[4.1] H. Masuda, M. Nakai, and M. Kubo, "Characteristics and Limitation of Scaled-Down MOSFET's Due to Two-Dimensional Field Effect," IEEE Tran. on Electron Devices, vol. ED-26, pp.980-986, June, 1979.

[4.2] J.R. Brews, W. Fichtner, E.H. Nicollian, and S.M. Sze, "Generalized Guide for MOSFET Miniaturization," IEEE Electron Device Letters, vol. EDL-1, no. 1, Jan., 1980.

[4.3] H. Shichijo, "A Re-examination of Practical Performance Limits of Scaled n-Channel and p-Channel MOS Devices for VLSI," Solid-State Electronics, vol. 26, no. 10, pp.969-986, 1983.

[4.4] T.Y. Chan, J. Chen, P.K. Ko, and C. Hu, "The Impact of Gate-Induced Drain Leakage Current on MOSFET Scaling," IEDM Tech. Digs., p.718, 1987.

[4.5] C. Chang and J. Lien, "Corner-Field Induced Drain Leakage in Thin Oxide MOSFETs," IEDM Tech. Digs., p.714, 1987.

[4.6] H. Sasaki, M. Saitoh, and K. Hashimoto, "Hot-Carrier Induced Drain Leakage Current in N-Channel MOSFET," IEDM Tech. Digs., p.726, 1987.

[4.7] M.H. Woods and B.L. Euzent, "Reliability in MOS Integrated Circuits," IEDM Tech. Digs., p.50, 1984.

[4.8] J. Lee, I.C. Chen, and C. Hu, "Modeling and Characterization of Gate Oxide Reliability," IEEE Tran. on Electron Devices, 1989.

[4.9] R. Moazzami, J. Lee, I.C. Chen, and C. Hu, "Projecting the Minimum Acceptable Oxide Thickness fir Time-Dependent Dielectric Breakdown," IEDM Tech. Digs., p.710, 1988.

[4.10] P.M. Lee, M.M. Kuo, K. Seki, P.K. Ko, and C. Hu, "Circuit Aging Simulator (CAS)," IEDM Tech. Digs., pp.134-137, 1988.

[4.11] C. Hu, S. Tam, F.-C. Hsu, P.K. Ko, T.Y. Chan, and K.W. Terrill, "Hot-Electron-Induced MOSFET Degradation - Model, Monitor, and Improvement," IEEE Tran. on Electron Devices, vol. ED-32, p.375, Feb. 1985.

[4.12] M. Yoshida, IEDM Tech. Digs., p.254, 1985.

Chapter 5

A DEEP-SUBMICROMETER MOSFET MODEL FOR ANALOG/DIGITAL CIRCUIT SIMULATIONS

In addition to high current drive, another major advantage of scaled-down devices is the reduced device area which allows higher integration levels. With the dramatic increase in the number of transistors per chip, the circuit complexity and the fabrication cost also increase proportionally. In order to speed up the VLSI/ULSI system design and to reduce costs, it has become necessary to start the circuit design in the early stages of technology development and to predict circuit behavior before the circuit is actually fabricated, both of which require intensive use of circuit simulators. Since the device characteristics of small-geometry devices are highly sensitive to parameter variations, optimal circuit designs become even more difficult to create than before. Therefore, an accurate and computationally efficient drain current model for deep-submicrometer MOSFETs becomes extremely crucial and indispensable in developing future system designs.

In this chapter, a MOSFET drain current model suitable to predict small geometry effects for size as small as quarter-micron channel length, for digital as well as analog applications is described. The basic framework of this model is based on the recent improved physical understanding of deep-submicrometer MOS devices. In developing this model, meticulous care has been taken in retaining the basic functional form of fully physical models while improving model accuracy and computational efficiency. The ease of parameter extraction was also a major consideration. In addition to the effects commonly included in the MOSFET drain current equation, it is found that the inversion-layer capacitance effect [5.1,5.2], hot-electron induced output resistance degradation [5.3,5.4], and source/drain parasitic resistance effect [5.5-5.7] are also important factors to consider in deep-submicrometer MOSFET modeling. A model considering all of these effects has been implemented in SPICE3 [5.8-5.10]. Some of

the simulation results are presented. The model parameter extraction algorithm and the measurement system are also briefly discussed [5.11].

5.1 General properties of MOSFET modeling

5.1.1 Semi-empirical nature

Because of stringent requirements, fully physics-based models are seldom used in circuit analysis either due to too complicated model equations or poor accuracy. Most MOSFET drain current models for circuit simulations are, to a certain extent, semi-empirical in nature [5.2]. In semi-empirical models, only the basic functional form of equations derived from device physics are kept to describe the general MOSFET behavior. Higher-order physical effects are incorporated through empirical equations with fitting parameters to achieve better accuracy and computational efficiency. The drawbacks of the semi-empirical approach are, however, the existence of non-physical parameters and the increasing size of the parameter set. With the computer capacity of today, an increase in the number of parameters is usually not a problem as long as all the parameters can be automatically extracted. But the existence of non-physical parameters and the associated parameter redundancy problems are vital to the optimization process during parameter extraction [5.12,5.13]. Non-physical parameter values are also difficult to interpolate or extrapolate for devices with different sizes. Therefore, a careful selection of model equations and parameter extraction strategy is important during the model development.

5.1.2 Accuracy

Due to tight specifications in circuit designs, MOSFET models for circuit simulations usually require high accuracy under all bias conditions. Much work has been done to the MOSFET drain current models. Important physical phenomena such as velocity saturation, mobility degradation, and some short-channel effects previously discussed are already included in most sophisticated models. For digital applications, existing drain current models are generally considered adequate for MOSFETs down to near-micron channel length. But for analog

applications, where output resistance is as important as the drain current, existing models are still far from satisfactory. For deep-submicrometer devices, the extra considerations are the source/drain parasitic resistance and the inversion-layer capacitance effects, both of which are not included in most MOSFET models.

5.1.3 Computational efficiency

Since the model equations are usually evaluated thousands of times during simulation for most VLSI circuits, it is essential that device models used in circuit simulators be as computationally efficient as possible. The computational efficiency of a model is reflected by three properties: simplicity, explicitness, and continuity of the model equations and their derivatives. The first two properties directly reduce the equation evaluation time and the last one reduces the number of iterations required in simulation and helps the program to converge. While most MOSFET models have explicit expressions for the drain current, discontinuities at transition points from strong-inversion to subthreshold, and from triode to saturation regions usually exist in the model equations and their derivatives. With the increasing complexity in VLSI circuits, convergence requirements have become more strict. The continuity property has been a crucial consideration in future MOSFET modeling.

5.1.4 Ease of parameter extraction

For modern MOSFET models that typically have a rather larger set of parameters, ease of parameter extraction is another essential property of MOSFET models, since the applicability of a MOSFET model in circuit simulation is highly dependent on how easily and accurately the model parameters can be extracted. Usually, MOSFET models and the parameter extraction system are developed in sequence which implicitly embeds difficulty to the parameter extraction and limits the potential capability of a model. This is one of the reasons for the weak link between device characterization and circuit simulation. A useful MOSFET model should be developed with ease of parameter extraction in consideration.

5.2 BSIM - Berkeley Short-Channel IGFET Model

As pointed out in section 3.3, the basic physics involved in deep-submicrometer MOSFETs is similar to that of their micron-sized counterparts and existing 1μ MOSFET models are extendible to the deep-submicrometer regime with only the need for minor modifications. Since the existing BSIM drain current model [5.14,5.15] (BSIM1) has all the favorable properties listed in section 5.1, it provides a good basis for developing a deep-submicrometer MOSFET model. This section briefly discusses the BSIM approach and formulation. Some results and problems of the BSIM1 model are also reviewed. More detailed description of the BSIM history and models can be found in [5.16].

5.2.1 The BSIM approach

BSIM is a generic name of an integrated system for circuit designs. It is composed of a group of models such as the drain current model and the substrate current model, a parameter extraction system, and a circuit simulator such as SPICE. BSIM models are physically meaningful and mathematically compact, and since they were developed based on comprehensive studies of device physics, most of the important physical effects are included. Except for the very basic physical equations, the rest of the model equations were empirically determined to achieve high accuracy and maximize the computational efficiency. There are no default parameter values and all model parameters are automatically extracted from physical devices with the associated parameter extraction system. This approach links circuit designs with process technologies, which makes simulation results more realistic and also eliminates any problems caused by slight process variations.

5.2.2 BSIM1 review

The present BSIM drain current model has been shown to be adequate in modeling MOSFETs with one micron channel length, but the accuracy begins to degrade when the channel length is reduced because the second-order effects of some physical phenomena become more severe and others are not properly implemented. Without mentioning the success that

BSIM1 has accomplished in the past, this section will first briefly describe the BSIM1 formulation, and then discuss some limitations and problems that BSIM1 has when applied to the deep-submicrometer regime.

The drain current equations of BSIM1 are similar to those of a textbook model, but with a better mobility model. The bulk-charge effect is also included and simplified. The major model equations of BSIM1 are listed in (5.1)-(5.12). A complete list of BSIM1 model equations and the meanings of each parameters are given in Appendix B.

(A) Threshold voltage:

$$V_{th} = V_{FB} + \phi_S + K_1\sqrt{\phi_S - V_{BS}} - K_2(\phi_S - V_{BS}) - \eta V_{DS} \tag{5.1}$$

where V_{FB} is the flat-band voltage, ϕ_S is the surface potential at which the threshold voltage voltage is defined, K_1 is the body-effect coefficient (equivalent to the parameter γ used in most text books), K_2 accounts for the non-uniform channel doping effect, and η is the drain-induced-barrier-lowering coefficient [5.17].

(B) Linear drain current:

$$I_{DS} = \frac{\beta_0(V_{GS} - V_{th} - \frac{a}{2}V_{DS})V_{DS}}{[1 + U_a(V_{GS} - V_{th})](1 + V_{DS}/E_c L_{eff})} \tag{5.2}$$

where β_0 is the conductance coefficient given by

$$\beta_0 = \frac{\mu_0 C_{ox} W_{eff}}{L_{eff}} \tag{5.3}$$

$$a = 1 + \frac{g K_1}{2\sqrt{\phi_S - V_{BS}}} \tag{5.4}$$

$$g = 1 - \frac{1}{1.744 + 0.8364(\phi_S - V_{BS})} \tag{5.5}$$

U_a is the mobility degradation coefficient due to the vertical field, E_c is the critical field for the velocity saturation effect, a and g are derived from the simplified bulk-charge effect.

(C) Saturation drain current:

$$I_{DS}=\frac{\beta_0}{1+U_a(V_{GS}-V_{th})}\frac{(V_{GS}-V_{th})^2}{2aK} \tag{5.6}$$

where

$$K=\frac{1+V_c+\sqrt{1+2V_c}}{2} \tag{5.7}$$

$$V_c=\frac{(V_{GS}-V_{th})}{aE_cL_{eff}} \tag{5.8}$$

and the drain saturation voltage is given by

$$V_{DSAT}=\frac{V_{GS}-V_{th}}{a\sqrt{K}} \tag{5.9}$$

(D) Subthreshold drain current:

$$I_{subth}=\frac{I_{EXP}I_{lim}}{I_{EXP}+I_{lim}} \tag{5.10}$$

where

$$I_{EXP}=\beta_0(V_{tm})^2e^{1.8}e^{(\frac{V_{GS}-V_{th}}{nV_{tm}})}[1-e^{(\frac{-V_{DS}}{V_{tm}})}] \tag{5.11}$$

$$I_{lim}=\frac{\beta_0}{2}(3V_{tm})^2 \tag{5.12}$$

V_{tm} is the thermal voltage given by kT/q, n is the subthreshold swing coefficient, I_{EXP} is the current in the subthreshold region, and I_{lim} is used to limit I_{subth} in the strong-inversion region. Some of the parameters in (5.1)-(5.12) are body-bias and/or drain-bias dependent.

Although BSIM1 includes many of the important physical effects, these effects are only corrected to the first order. Therefore, the minimum valid device dimensions of BSIM1 are limited to $L_{eff} \approx 1\mu m$ and $T_{ox} \approx 15nm$. The channel length modulation effect (or the output resistance) is empirically modeled by the β_0 parameter using a quadratic function of V_{DS} (see Appendix B or [5.18]). This approach is generally adequate in modeling the drain current as shown by the good agreement between the measured and calculated I_{DS} - V_{DS} characteristics in Fig. 5.1a, but does not correctly predict the output resistance as shown in Fig. 5.1b, where the

corresponding output resistance of Fig. 5.1a is plotted. The inability to model the output resistance makes BSIM1 less suitable for analog applications than for digital. Furthermore, the transition from the subthreshold region to the strong-inversion region is achieved by summing up the currents calculated in both regions. This approach adds I_{lim} to the strong-inversion drain current and results in a constant current offset in the strong-inversion region as can shown in Fig. 5.2. For analog applications, where most MOSFETs are operated at low current levels, this current offset may cause large errors in simulation results. Also, an empirical constant $e^{1.8}$ is used in (5.11) to account for the slight difference in threshold voltage between strong-inversion and subthreshold regions. In reality however, it is found that this threshold voltage offset is technology dependent, varying with process, and is a function of device dimensions. An example is shown in Fig. 5.3. Using a fixed number for the offset may also introduce large simulation errors related to the off-state leakage such as DRAM refresh time.

5.3 The BSIM2 model

The deep-submicrometer model, BSIM2, was developed based on BSIM1 but with the aforementioned problems of BSIM1 in mind. In this section, the detailed derivation of the BSIM2 model is described. BSIM2 has been successfully used to model the drain current and output resistance of MOSFETs with gate oxide thicknesses as thin as 3.6nm and channel lengths as small as 0.2μm. The drain current equations and their first derivatives are continuous throughout the bias ranges. An improved parameter extraction algorithm and system for BSIM2 were also developed [5.11] and are discussed in section 5.4.

5.3.1 Physical effects included

Based on recent physical understanding of deep-submicrometer MOSFETs discussed in chapter 3, it is found that the important effects that should be included in MOSFET modeling are:

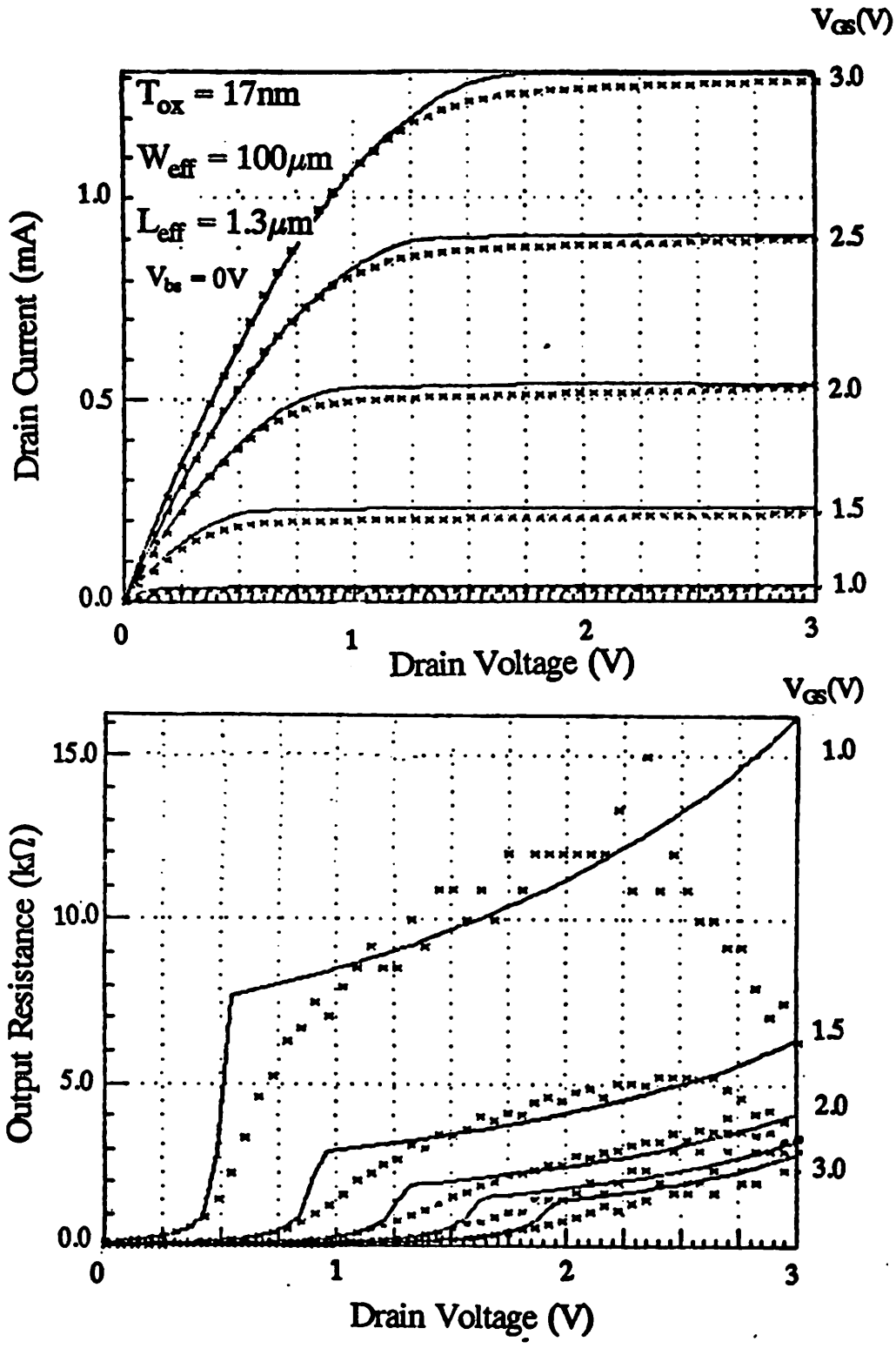
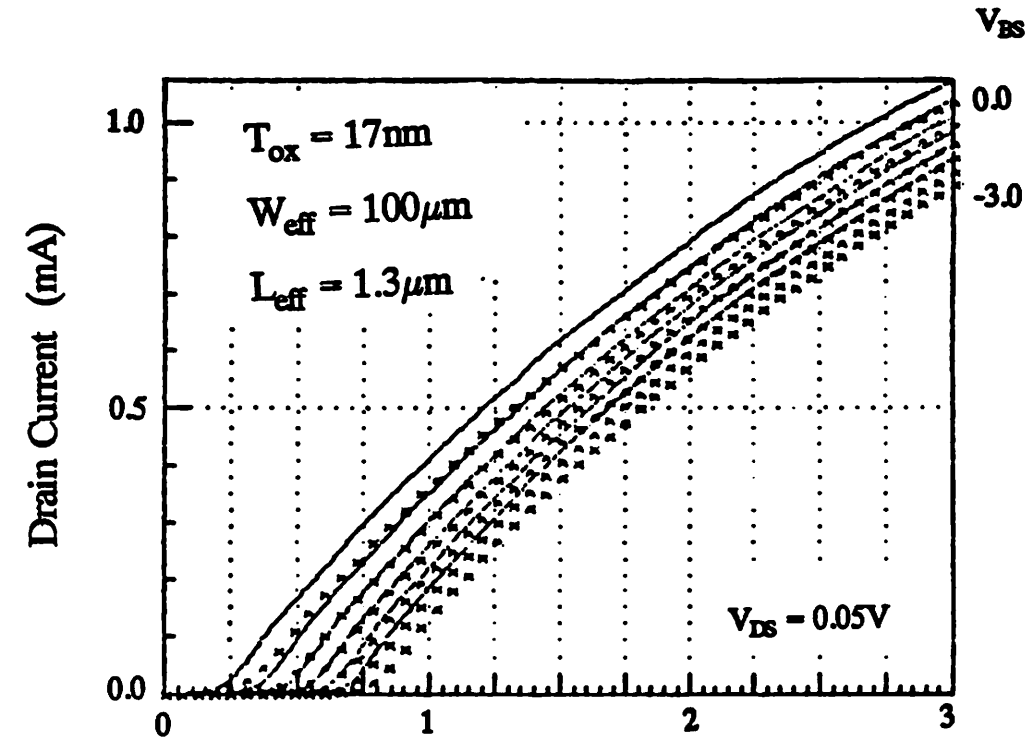


Fig. 5.1 BSIM1 modeling results. "x"s are measured data and the solid lines are the BSIM1 model. (a) I_{DS} - V_{DS}, (b) the corresponding output resistance of the same device.



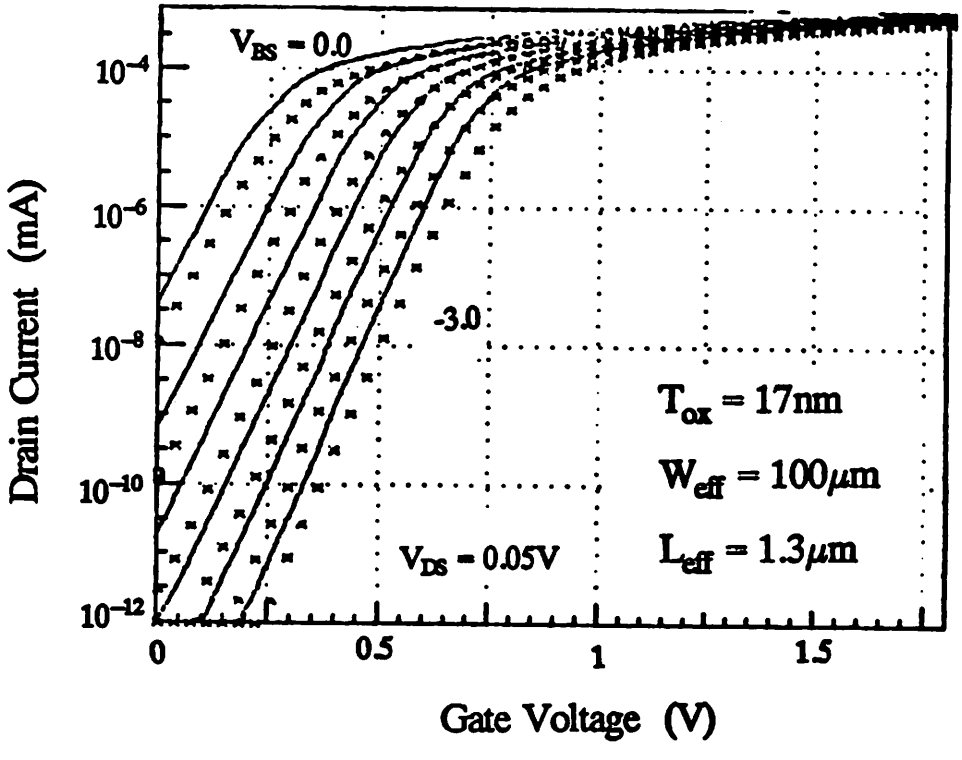


Fig. 5.2 I_{DS} - V_{GS} modeling results of BSIM1. "x"s are measured data and the solid lines are the BSIM1 model.

Fig. 5.3 BSIM1 modeling results in the subthreshold region. "x"s are measured data and the solid lines are the BSIM1 model.

1. mobility reduction due to the vertical field.
2. carrier velocity saturation.
3. drain-induced barrier lowering.
4. source/drain charge sharing.
5. non-uniform channel doping.
6. channel length modulation.
7. subthreshold conduction.
8. source/drain parasitic resistance.
9. hot-electron-induced output resistance reduction.
10. inversion-layer capacitance.

Except for the last three effects, most of these were already included in the BSIM1 model. In the following sections, these effects were re-examined based on deep-submicrometer MOSFETs considerations and their implementation in BSIM2 is described.

5.3.2 Strong-inversion region

(A) Threshold voltage

It has been shown in chapter 3 that the threshold voltage of properly designed deep-submicrometer MOSFETs does not exhibit severe short-channel effects until the devices are near punchthrough. Therefore, the threshold voltage model, Eq.(5.1), used in BSIM1 which already includes most of the important short-channel effects such as the source/drain charge-sharing, non-uniform channel doping, and drain-induced-barrier-lowering, is retained in BSIM2. A typical threshold voltage for a quarter-micron n-channel MOS transistor is shown Fig. 5.4. The asterisks are measured data and the solid curve is calculated from (5.1). In BSIM1, the drain-induced-barrier-lowering coefficient η is empirically expressed as a linear function of V_{DS} and V_{BS}. But in BSIM2, the dependence of η on V_{DS}, η_D, is removed since it does not agree with physical principle and may cause negative output resistance at low current levels if η_D has a wrong sign.

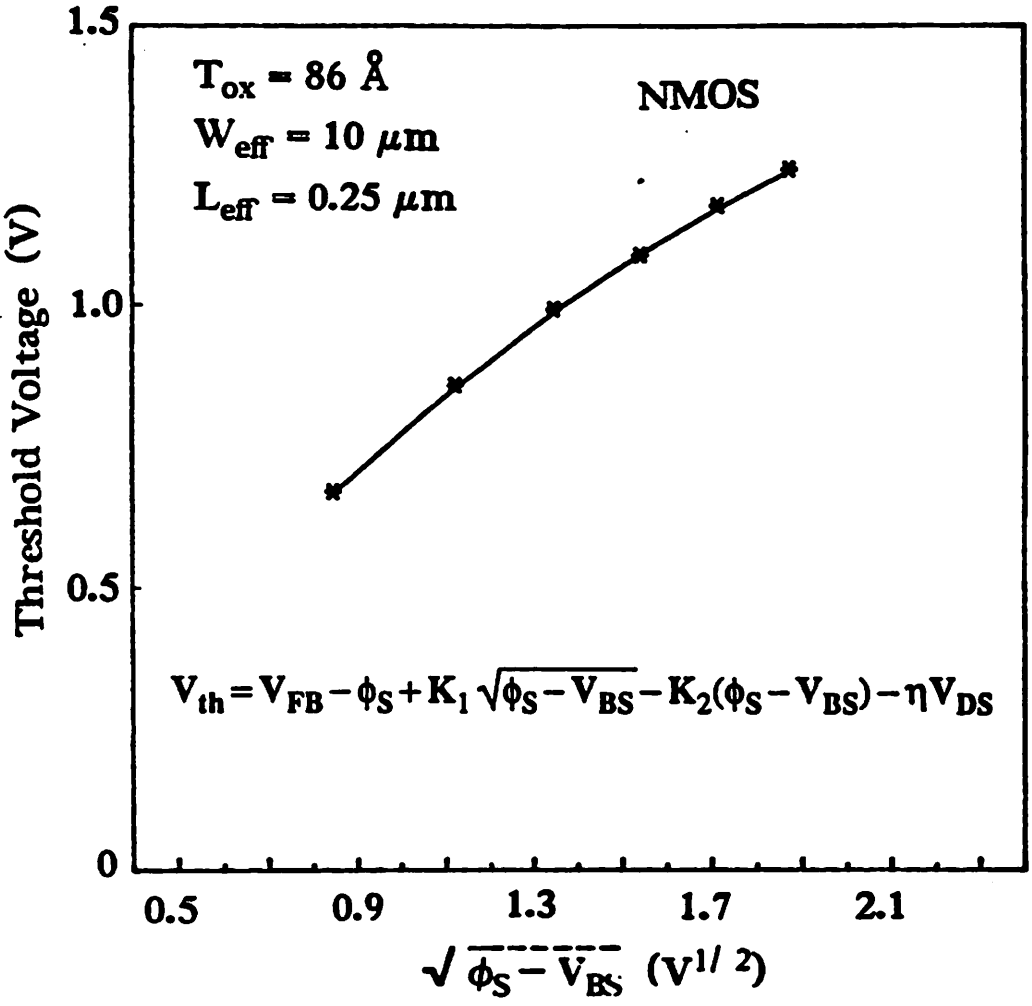


Fig. 5.4 Threshold voltage versus body bias for a quarter micron n-channel MOSFET. The asterisks are measured data and the solid curve is the model.

(B) Velocity saturation

The accuracy of a drain current model is directly affected by how the velocity saturation effect is implemented, especially for deep-submicrometer devices in which the channel electric field is high and the drain current quickly saturates. The relationship between the carrier velocity and the channel electric field has been studied by various groups [5.19,5.20]. One of the results is shown in Fig. 5.5 [5.19]. The dots are measured data and the curves represent different velocity models which will be described below. The most commonly used carrier velocity model is (5.13), because it leads to simple analytical drain current equations.

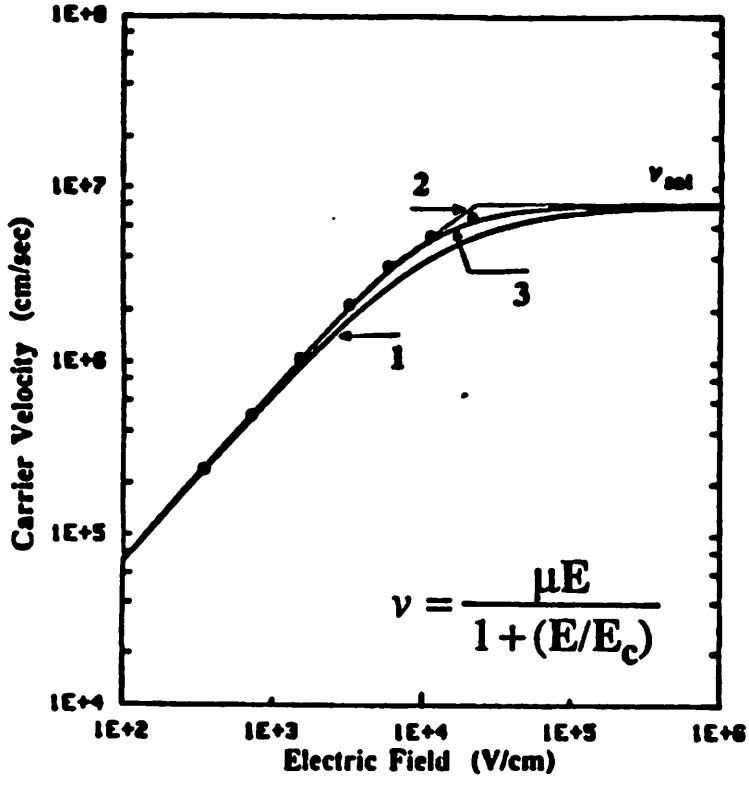
$$v = \frac{\mu}{1+(E/E_c)} \tag{5.13}$$

where $E_c = E_{c0} \equiv v_{sat}/\mu$, μ is the mobility, and v_{sat} is the carrier saturation velocity. The result of (5.13) is indicated by curve-1 in Fig. 5.5. However, this model underestimates the carrier velocity in the low-field region as can be seen in Fig. 5.5. A second model, the so-called "two-section" model [5.21], was proposed, to improve the accuracy of (5.13). In the two-section model, a larger critical field of $E_c = 2E_{c0}$ is used in (5.13) when $E < E_c$ and sets $v = v_{sat}$ when $E > E_c$. The result is plotted as curve-2. Although this model can achieve better accuracy, it results in a discontinuity in the first derivative of the drain current equation at V_{DSAT}. To retain the high accuracy and yet to avoid the discontinuity problem, a compromised solution is used in BSIM2.

$$E_c = E_{c0}\left[1+\frac{E_{cD}(V_{DS}-V_{DSAT})^2}{V_{DSAT}^2}\right] \qquad \text{if } V_{DS} \le V_{DSAT}$$

$$= E_{c0} \qquad \text{if } V_{DS} > V_{DSAT} \tag{5.14}$$

where the critical field E_c starts from a larger value of $E_{c0}(1 + E_{cD})$ in the low-field region as in the two-section model and smoothly changes to E_{c0} at $V_{DS} = V_{DSAT}$ as in (5.13). E_{cD} is a fitting parameter. The result of (5.14) is shown in curve-3 in Fig. 5.5. The quadratic function used in (5.14) is to keep the first derivative of the drain current equation continuous at V_{DSAT}.



1. $E_c = E_{c0} \equiv \dfrac{v_{sat}}{\mu}$

2. $E_c = 2E_{c0} \qquad E \le E_c$

   $v = v_{sat} \qquad E > E_c$

3. $$E_c = E_{c0}\left[1+\frac{(V_{DS}-V_{DSAT})^2}{V_{DSAT}^2}\right]$$

   $$= E_{c0} \qquad V_{DS} > V_{DSAT}$$

Fig. 5.5 Carrier velocity versus longitudinal channel electric field. The dots are measured electron velocity [5.19] and the solid curves are different models.

(C) Mobility reduction due to vertical field

It has been shown that the carrier mobility can be expressed as a universal function of the effective vertical electric field for a wide range of oxide thickness and channel doping concentration [5.22-5.24].

$$\mu=\frac{\mu_0}{1+(E_{eff}/E_o)^n} \tag{5.15}$$

where μ_0, E_o, and n are constants whose values can be found in [5.24], E_{eff} is the effective vertical electric field in the inversion layer given by

$$E_{eff}=\frac{1}{\epsilon_{si}}(Q_B+\frac{Q_n}{2}) \tag{5.16}$$

Q_B and Q_n are the bulk and inversion charge density, respectively. However, Eq.(5.15) is seldom used in circuit analysis for several reasons. First, Eq.(5.15) does not have a desirable functional form for circuit simulation purposes because of the power function in the denominator. Table 5.1 lists the relative evaluation time required for various functions based on 10^6 calculations in a SUN SPARC-1 station. The evaluation time of a power function is about 3 times that of an exponential and 120 times that of a simple arithmetic operation. Secondly, Q_n in the subthreshold region can not be explicitly expressed as a function of the terminal voltages, therefore it can not be directly used in most circuit simulators and results in difficulty in parameter extraction. Thirdly, the parasitic source/drain resistance effect also causes measured mobility to deviate from (5.15). A more widely accepted mobility model is (5.17), which can be considered as a first-order approximation of (5.15).

$$\mu=\frac{\mu_0}{1+U_a(V_{GS}-V_{th})} \tag{5.17}$$

For deep-submicrometer devices with thin gate oxides, the vertical electric field is too large which causes the first-order approximation of (5.15) to become inadequate as shown in Fig. 5.6, where the electron mobility for 8.6nm gate oxide devices is plotted against $V_{GS}-V_{th}$. The dots are measured data and the dashed line is the best fit of (5.17). A simple amendment to

improve the accuracy is to add higher-order terms to the denominator of (5.17). It is found that the inclusion of the second-order term only, (5.18), is enough for most thin gate oxide devices under normal bias conditions and the result is given by the solid curve in Fig. 5.6.

$$\mu=\frac{\mu_0}{1+U_a(V_{GS}-V_{th})+U_b(V_{GS}-V_{th})^2} \tag{5.18}$$

The use of (5.18) also provides BSIM2 with the capability to model the non-monotonic mobility behavior with the vertical field at low temperatures [5.25].

(D) Source/drain parasitic resistance

The effect of the source/drain resistance on device performance has been discussed in section 3.3. In section 3.3, these parasitic resistances were treated as external components to the MOSFETs, but from a circuit simulation point of view, adding extra elements to each transistor would greatly increase the circuit size and slow down the simulation speed. Furthermore, extracting intrinsic MOSFET parameters from extrinsic device characteristics requires special care during the measurement and optimization procedures, which would greatly complicate the parameter extraction process. It can be shown [5.26] that with a proper selection of the model equations, the parasitic resistance effect can be lumped into the mobility term. In BSIM2, the parasitic resistance effect is incorporated in (5.15) and (5.18). Therefore, when the mobility parameters are extracted from test devices, the source/drain parasitic resistance effect is automatically included in this model.

(E) Drain current

In BSIM1, the velocity saturation and mobility reduction due to vertical field effects were put together through multiplication of (5.15) and (5.17), resulting in a product term in the denominator of (5.2). With such formulation, the saturation velocity decreases as the gate voltage increases which does not agree with the physical observation that the carrier saturation velocity is constant at a given temperature, independent of the vertical field [5.19,5.20,5.27]. A more physical approach, which sums the two effects together, is adopted in BSIM2 and the

resulting drain current for the linear region is given in (5.19).

Function	Total time (msec)	Comment
EMPTY FOR LOOP	916	for(i=1;i<10^6,++i) {}
IF	817	if (a) {}
ASSIGNMENT	567	a = b
PLUS	383	a + b
INCREMENT	1067	++a
MINUS	383	a - b
DECREMENT	1050	--a
MULTIPLICATION	400	a * b
DIVISION	2650	a / b
MODULUS	3167	A % B
EXP(a)	16299	e^a
POW(a,b)	50506	a^b
SQRT(a)	1867	$\sqrt{a}$
LOG(a)	4283	ln(a)
LOG10(a)	16949	log(a)
FABS(a)	1300	| a |
ACOS(a)	6367	$\cos^{-1}(a)$
ASIN(a)	13133	$\sin^{-1}(a)$
ATAN(a)	11533	$\tan^{-1}(a)$
ATAN2(a,b)	268189	$\tan^{-1}(b/a)$
CEIL(a)	4350	[a+1]
FLOOR(a)	4233	[a]
SIN(a)	14600	sin(a)
COS(a)	12583	cos(a)
TAN(a)	23966	tan(a)
SINH(a)	38915	sinh(a)
COSH(a)	26332	cosh(a)
TANH(a)	35249	tanh(a)
AND	1767	a && b
OR	900	a || b
NOT	800	!a
EQUAL	900	a == b
NOT EQUAL	750	a != b
LARGER OR EQUAL TO	883	a >= b
SMALLER OR EQUAL TO	900	a <= b
LARGER THAN	733	a > b
SMALLER THAN	783	a < b

Table 5.1 Relative evaluation time required for various mathematical functions in C language based on 10^6 calculations from a SUN SPARK-1 station.

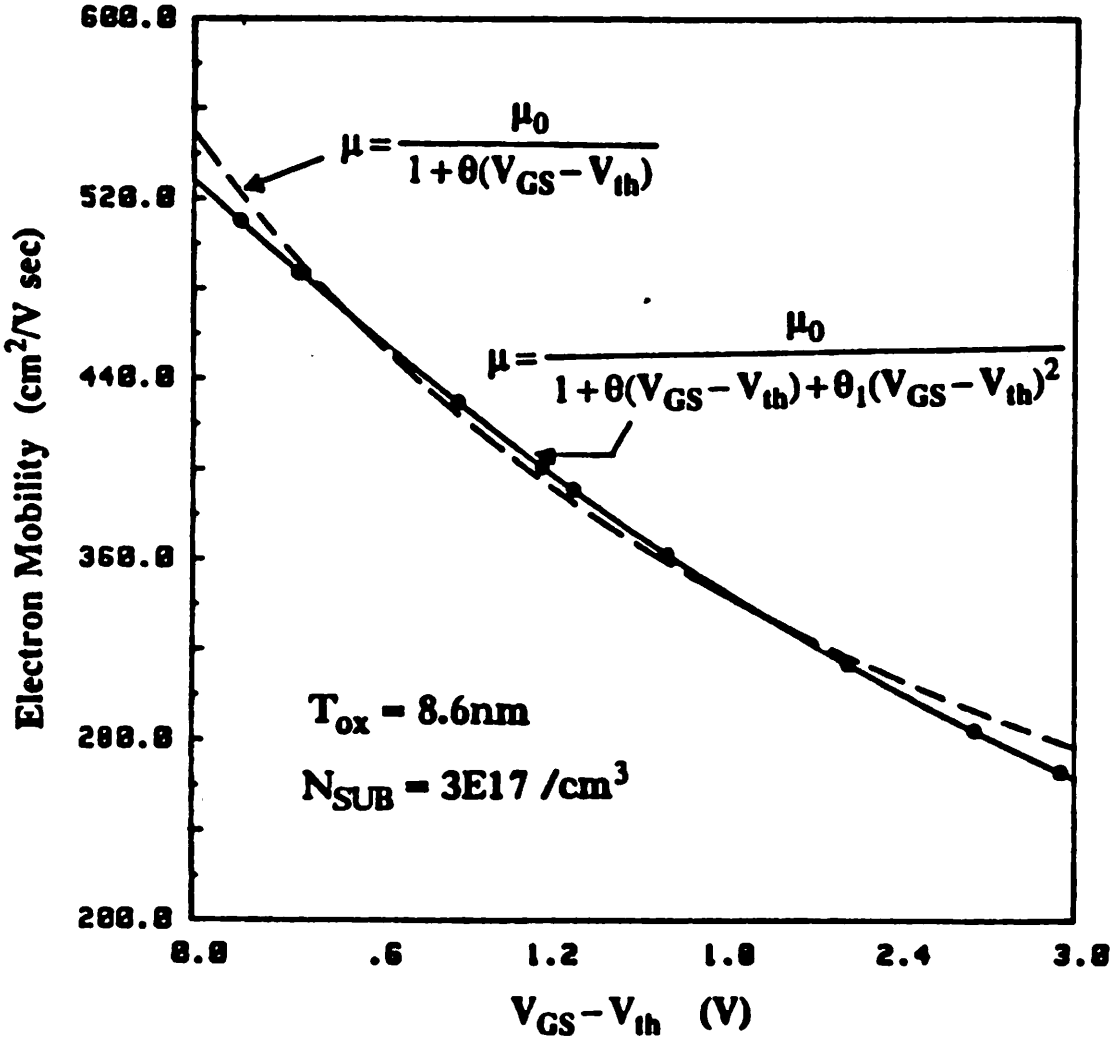


Fig. 5.6 Low-field electron mobility versus gate voltage. The dots are measured data [5.22] and the curves are the best fit using different equations.

$$I_{DS}=\frac{\beta_0(V_{GS}-V_{th}-\frac{a}{2}V_{DS})V_{DS}}{1+U_a(V_{GS}-V_{th})+U_b(V_{GS}-V_{th})^2+V_{DS}/E_cL_{eff}} \tag{5.19}$$

where E_c is given in (5.15). As a matter of fact, when the channel length is longer than one micron, either using a product term or a summation term in the denominator of (5.19) has very little difference, as far as the accuracy is concerned. This is shown in Fig. 5.7 where simulated drain currents for both approaches are compared to measured data for a device with T_{ox} = 8.6nm and L_{eff} = 1.5μm. The dots are measured data and the curves are simulations. The difference between the two simulation results is negligible. However, as the channel length is reduced, velocity saturation effects become more important and simulations from the summation approach becomes more accurate than the multiplication approach as shown in Fig. 5.8, where measured and simulated I_{DS} - V_{GS} characteristics for a quarter-micron MOSFET are compared.

Thus far the drain current formulation of BSIM2 in the strong-inversion region is similar to that of BSIM1 listed in (5.2)-(5.9) except for some minor modifications. Since the denominator in (5.19) is different from that in (5.2), the expression for V_c in (5.8) should be changed to (5.20) accordingly.

$$V_c=\frac{(V_{GS}-V_{th})}{aE_cL_{eff}[1+U_a(V_{GS}-V_{th})+U_b(V_{GS}-V_{th})^2]} \tag{5.20}$$

(F) Modeling results

Fig. 5.9 shows measured and calculated I_{DS} - V_{GS} characteristics in the linear region for an n-channel MOSFET with L_{eff} = 0.25μm and T_{ox} = 8.6nm. The asterisks are measured data and the solid curves are simulation results. The dashed curve along V_{BS} = 0V shows the result if (5.17) is used instead of (5.18) in the drain current equation, (5.19), exemplifying the importance of the second-order term of the mobility reduction effect in (5.18). Fig. 5.10 shows measured and calculated I_{DS} - V_{GS} characteristics in the saturation region for the same device. Similar figures for a p-channel MOSFET with L_{eff} = 0.4μm and T_{ox} = 7.5nm are shown in Fig.

5.11 and 5.12. More modeling results for different device dimensions and technologies can also be found in [5.11].

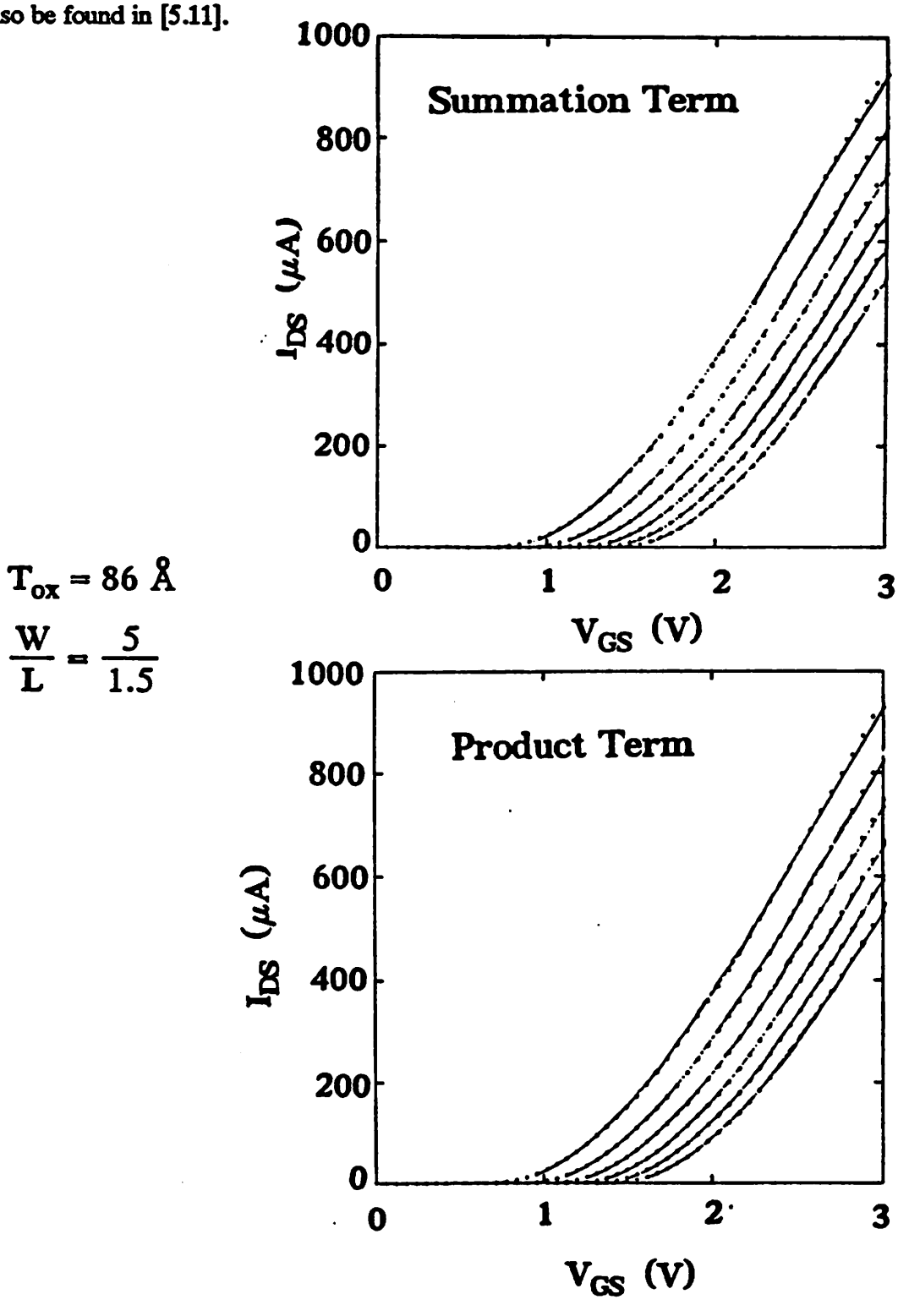
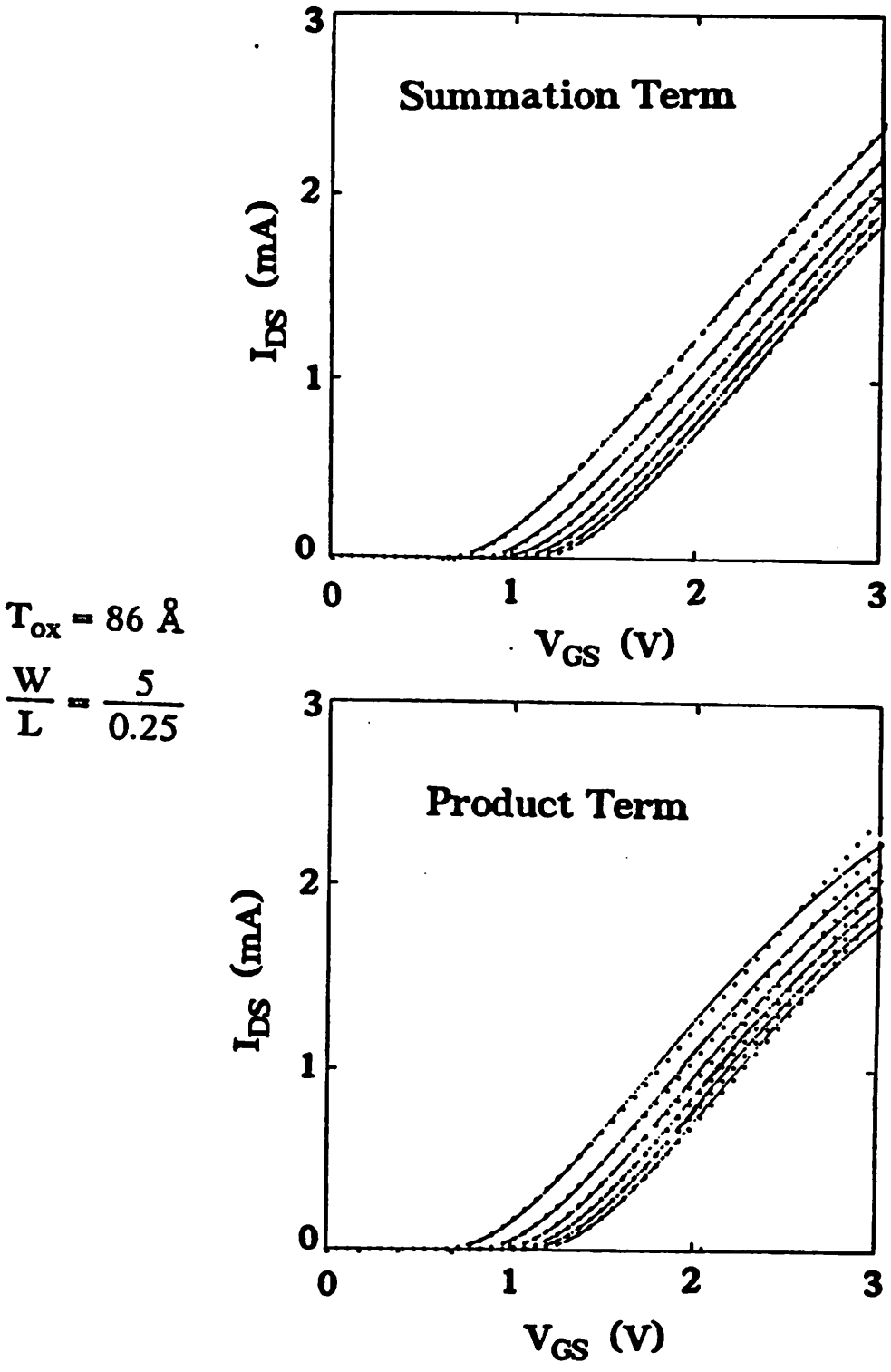


Fig. 5.7 Comparison of modeling results between a summation term and a product term in the denominator of Eq.(5.19) for a "longer" channel device.



T_{ox} = 86 Å

$\frac{W}{L} = \frac{5}{0.25}$

Fig. 5.8 Comparison of modeling results between a summation term and a product term in the denominator of Eq.(5.19) for a quarter-micron device.

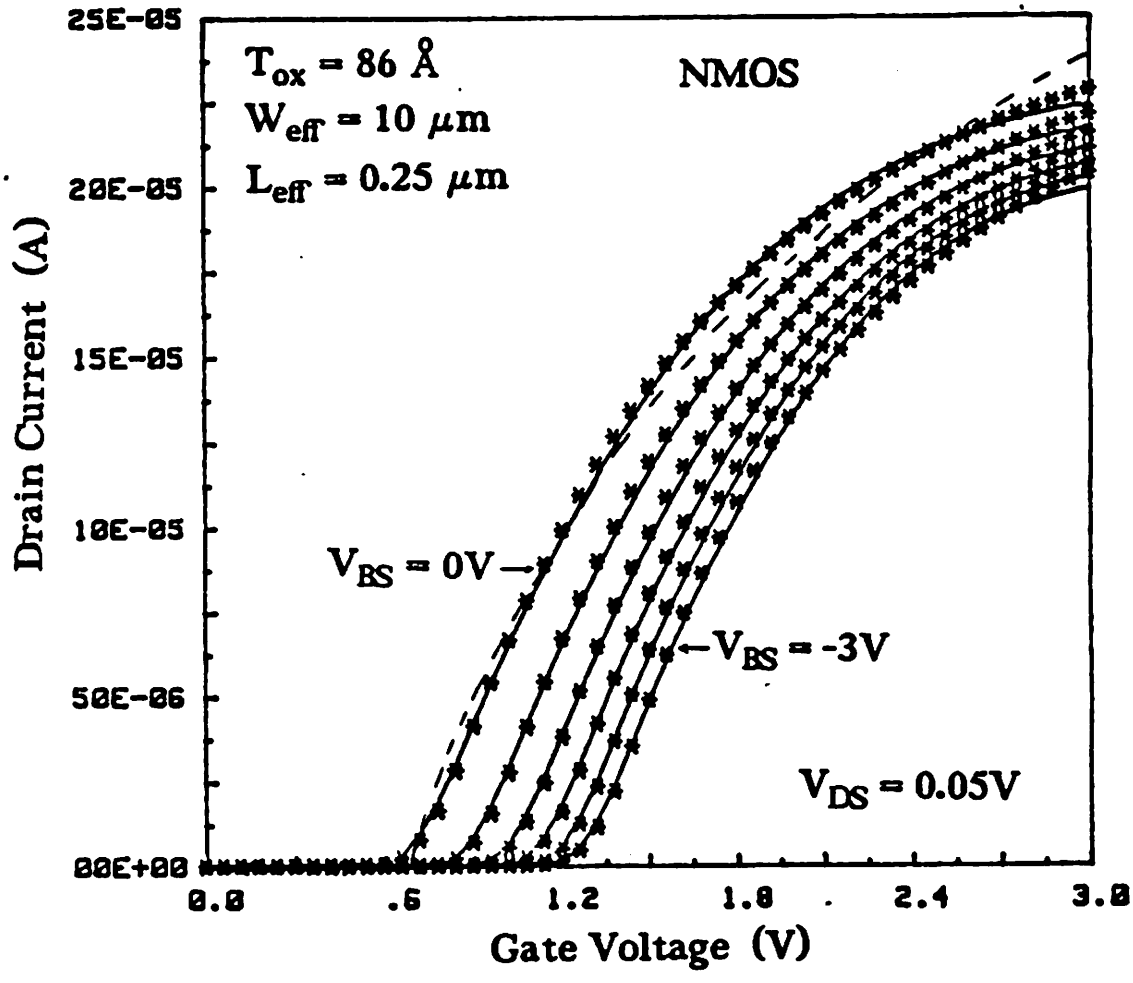


Fig. 5.9 BSIM2 modeling results in the linear region for an n-channel device. The asterisks are measured data and the solid lines are the model.

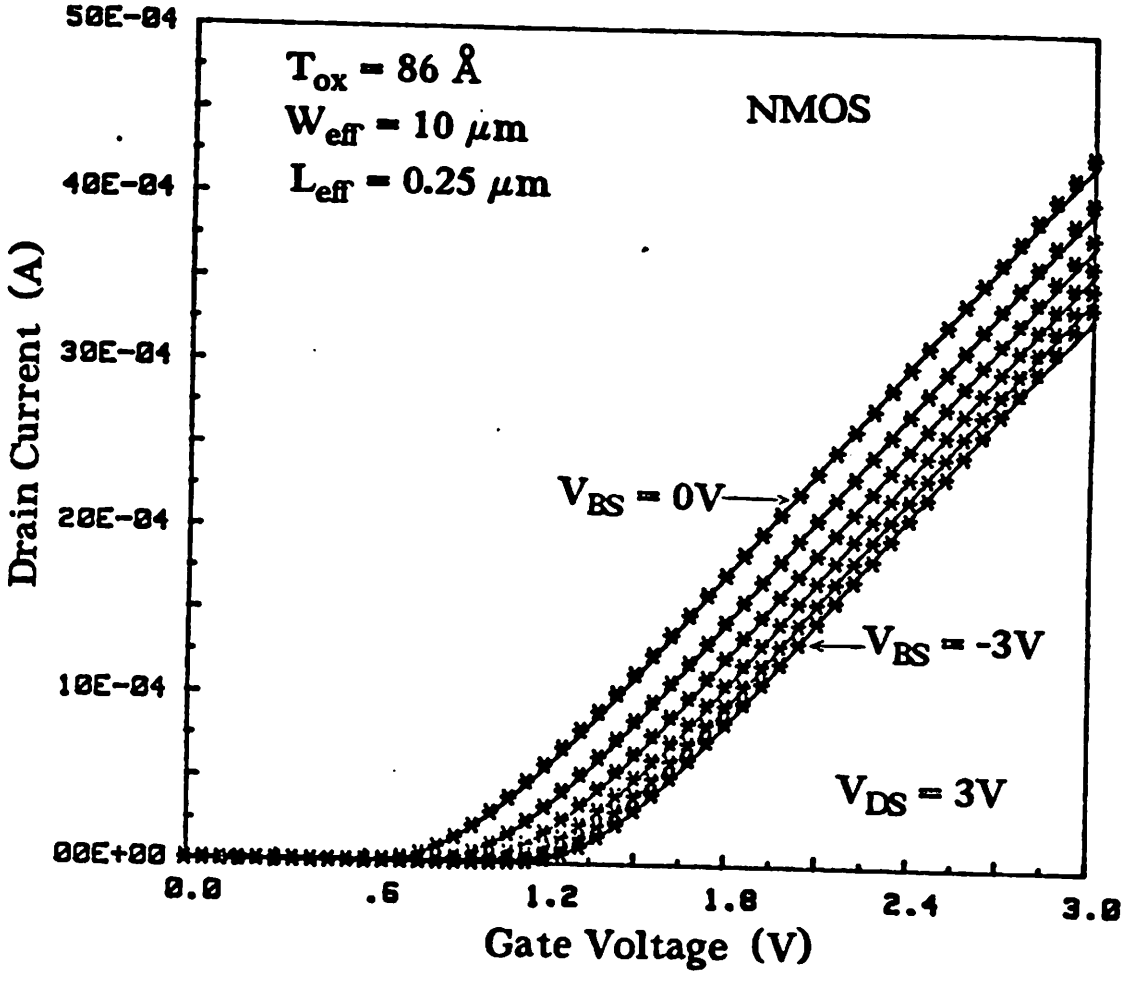


Fig. 5.10 BSIM2 modeling results in the saturation region for an n-channel device. The asterisks are measured data and the solid lines are the model.

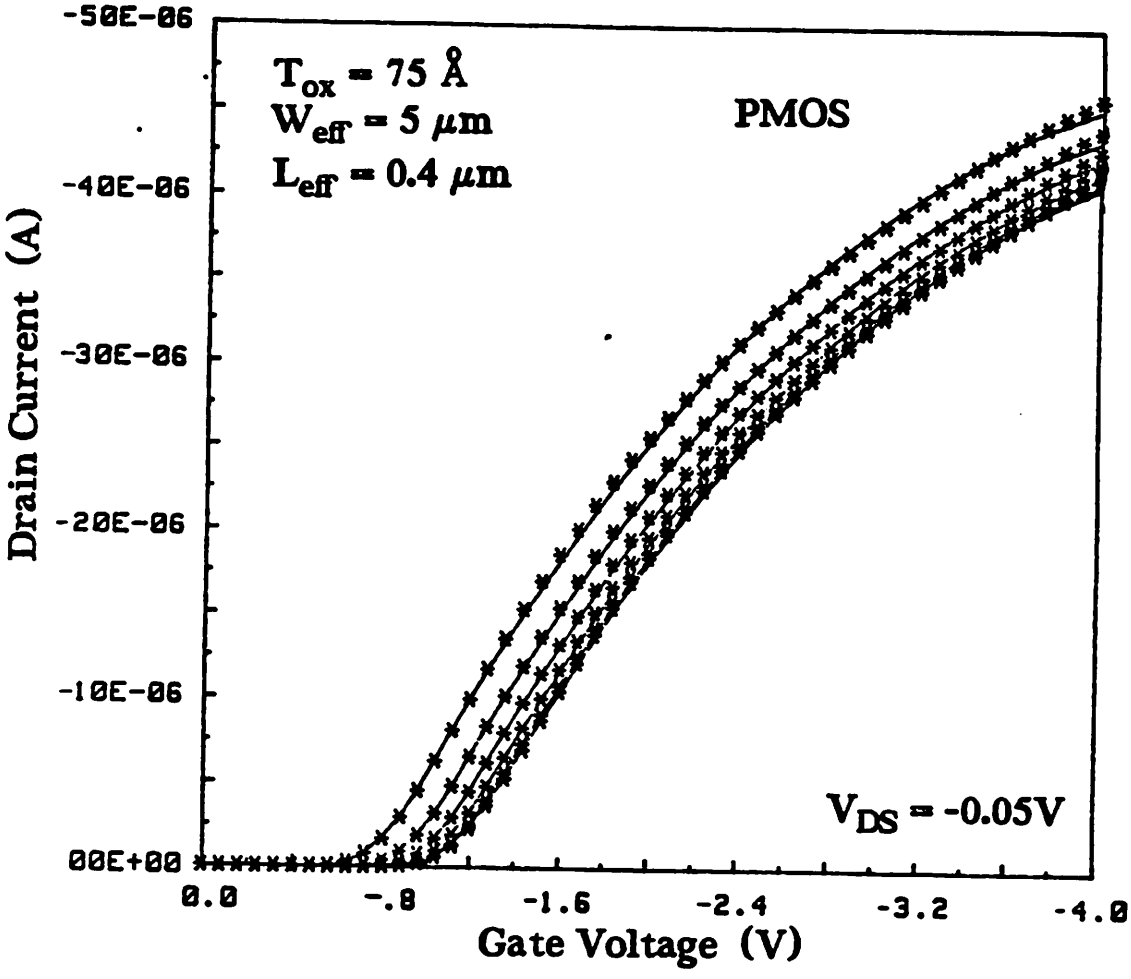


Fig. 5.11 BSIM2 modeling results in the linear region for a n-channel device. The asterisks are measured data and the solid lines are the model.

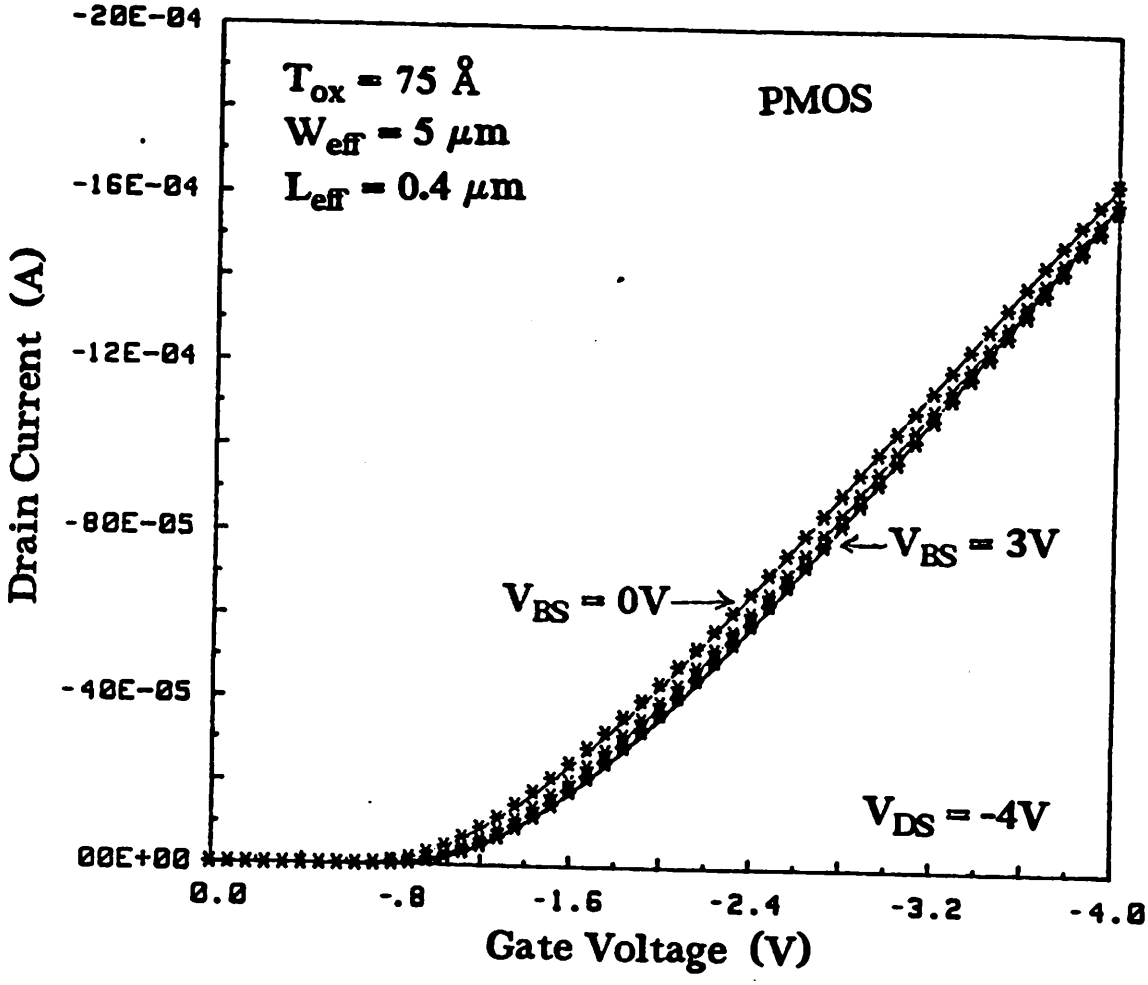


Fig. 5.12 BSIM2 modeling results in the saturation region for a p-channel device. The asterisks are measured data and the solid lines are the model.

5.3.2 Subthreshold region

The drain current in the subthreshold region is dominated by the diffusion current [5.27]. Using the charge-sheet approximation [5.28], the subthreshold current of a MOSFET can be expressed as

$$I_{subth}=\frac{3}{2}\beta_0(V_{tm})^2\frac{C_d}{C_{ox}}\frac{e^{\phi_s-2\phi_B}}{V_{tm}}(1-e^{\frac{V_{DS}}{V_{tm}}}) \tag{5.21}$$

where $\phi_B = V_{tm}\ln(N_{SUB}/n_i)$, C_d is the depletion-layer capacitance give by

$$C_d=\sqrt{\frac{q\epsilon_{si}N_{SUB}}{2\phi_S}} \tag{5.22}$$

A quick derivation of (5.21) is given in Appendix C. However, Eq.(5.21) can not be directly used unless ϕ_S can be expressed as an explicit function of the terminal voltages. A relationship between the gate voltage, V_{GS}, and ϕ_S has been derived in [5.27], but it is too complicated to be used in circuit analysis. In BSIM2, a simple approximation is used which will be described below. Fig. 5.13 shows a plot of ϕ_S versus V_{GS} - V_{FB} calculated from two-dimensional analysis. It is found that ϕ_S in the subthreshold region can be accurately approximated as a linear function of V_{GS} with slope nV_{tm} and y-intercept $-V_o$ as drawn by the dashed in in Fig. 5.13.

$$\phi_S\approx\frac{(V_{GS}-V_o)}{nV_{tm}}V_o \tag{5.23}$$

where n is the subthreshold swing coefficient given by

$$n=1+\frac{C_d}{C_{ox}} \tag{5.24}$$

and

$$V_o\approx\frac{2\phi_BC_d}{C_d+C_{ox}} \tag{5.25}$$

A detailed derivation of (5.23)-(5.25) can also be found in Appendix C. Substituting (5.23)

into (5.21), the subthreshold current can be rewritten as

$$I_{subth}=\frac{3}{2}\beta_0(V_{tm})^2\frac{C_d}{C_{ox}}\frac{e^{\frac{V_{GS}-V_{FB}-2\phi_B-nV_o}{nV_{tm}}}}{nV_{tm}}[1-e^{\frac{V_{DS}}{V_{tm}}}] \quad (5.26)$$

Eq. (5.26) can further be simplified to yield

$$I_{subth}=\beta_0(V_{tm})^2\frac{e^{\frac{V_{GS}-V_{th}-V_{offset}}{nV_{tm}}}}{nV_{tm}}[1-e^{\frac{V_{DS}}{V_{tm}}}] \quad (5.27)$$

where all the missing terms in (5.26) are incorporated into the fitting parameter V_{offset}, which accounts for the threshold voltage offset between strong-inversion and subthreshold regions.

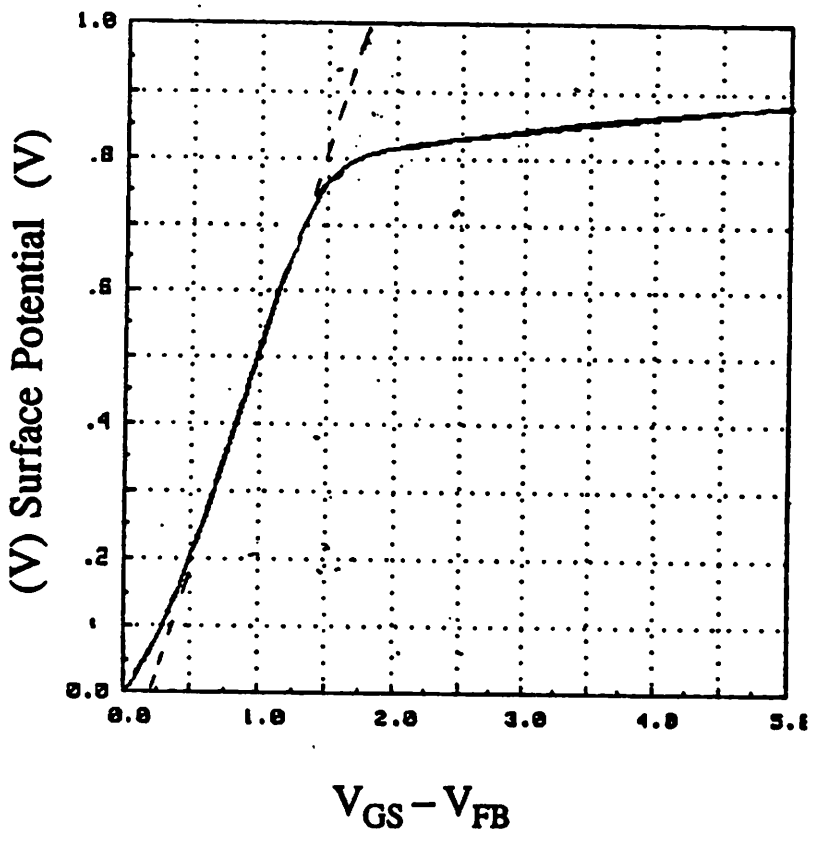


Fig. 5.13 Surface potential versus gate voltage. The dashed line is a linear approximation of the surface potential in the subthreshold region.

5.3.3 Transition region

(A) Inversion-layer capacitance effect

When the gate oxide thickness becomes comparable to the depletion-layer thickness as in in most deep-submicrometer devices, the inversion-layer capacitance effect is no longer negligible and has to be included in MOSFET modeling especially for low gate voltage operations. Fig. 5.14 shows calculated inversion charge density as a function of the gate voltage for a T_{ox} = 8nm device. The straight line in the same figure represents the usual linear approximation given by $C_{ox}(V_{GS}-V_{th})$. Because of the inversion-layer capacitance effect, the inversion charge density deviates from its linear approximation when the gate voltage is near the threshold voltage. This deviation increases as the gate oxide thickness decreases. Although the drain current equations in both the strong-inversion and the subthreshold regions can be derived, there is no simple analytical expression for the drain current near the threshold voltage.

In this model, a transition region between the strong-inversion and the subthreshold regions is determined. This transition region is marked in Fig. 5.15. A cubic spline function of V_{GS} is created for this region to account for the inversion-layer capacitance effect. The upper and lower bound of the transition region can be calculated from the equivalent circuit shown in Fig. 5.16. The effective gate capacitance C'_{ox} of a MOSFET looking into the gate terminal is equal to C_{ox} in series with C_{inv}.

$$C'_{ox}=\frac{dQ_n}{dV_{GS}}=\frac{C_{ox}C_{inv}}{C_{ox}+C_{inv}} \tag{5.28}$$

where C_{inv} is defined as $dQ_n/d\phi_S$ and can be expressed as (5.29) in the subthreshold region (see Appendix C).

$$C_{inv}=C_d\exp^{\frac{\phi_S-2\phi_B}{V_{tm}}} \tag{5.29}$$

Using the same approach used in the derivation of (5.27), C_{inv} can be approximated as

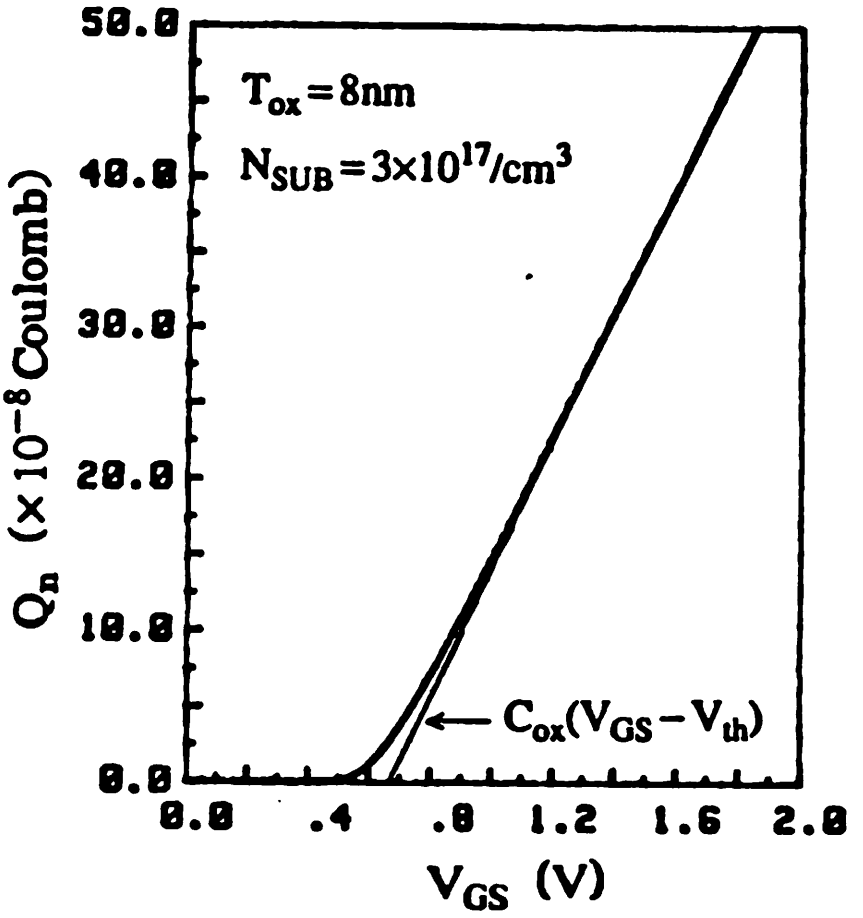


Fig. 5.14 Channel inversion charge density versus gate voltage calculated from two-dimensional analysis. The straight line is the usual approximation given by $C_{ox}(V_{GS}-V_{th})$.

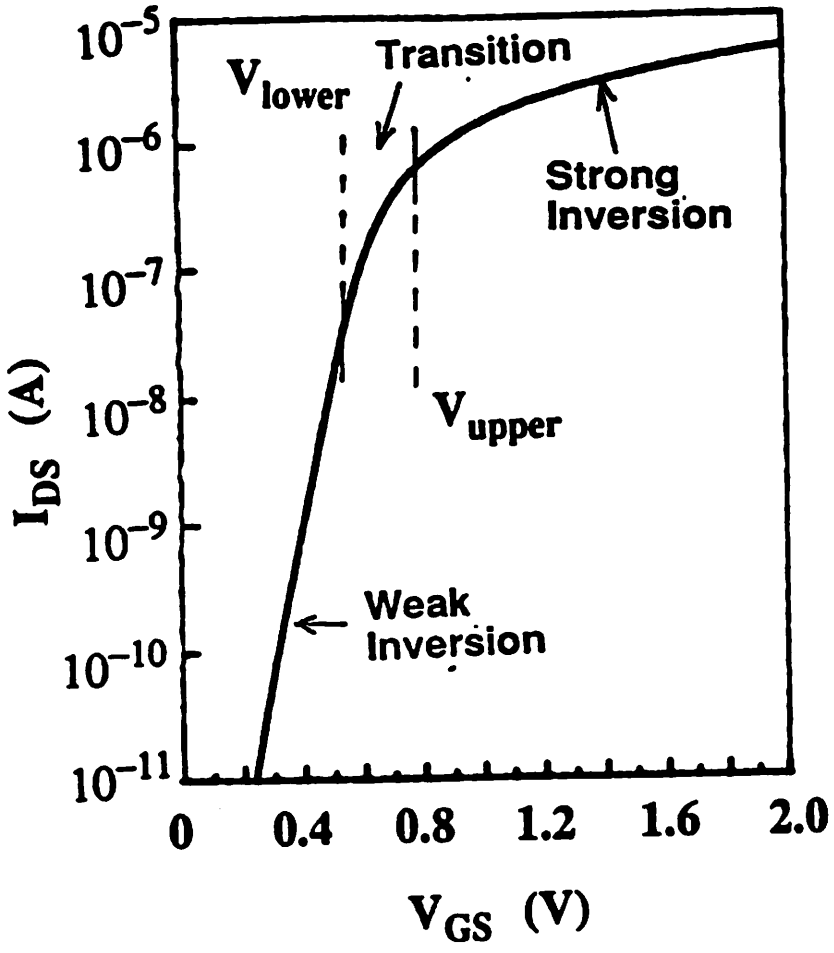


Fig. 5.15 Drain current versus gate voltage in logarithmic scalei illustrating the smooth transition from subthreshold region to the strong inversion region.

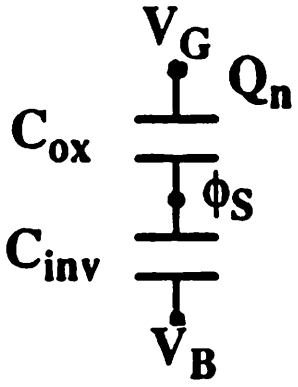


Fig. 5.16 Equivalent capacitance circuit of a MOSFET looking into the gate.

$$C_{inv} \approx C_d e^{\frac{V_{GS}-V_{th}-V_{offset}}{nV_{tm}}} \tag{5.30}$$

In the strong-inversion region, C_{inv} is much larger than C_{ox}, therefore, $C'_{ox} \approx C_{ox}$. This is where the linear approximation $Q_n = C_{ox}(V_{GS} - Vth)$ is valid. In the subthreshold region however, C_{ox} is much larger than C_{inv}. This is where the diffusion current dominates. By comparing the relative magnitudes of C_{ox} and C_{inv} along V_{GS}, the the lower bound V_{G1} and upper bound V_{G2} of the transition region can be determined. For example, if we define the lower bound to be at which $C_{ox} = 100C_{inv}$, then V_{G1} will be

$$V_{G1} = V_{th} - V_{offset} + nV_{tm}\ln(\frac{C_{ox}}{100C_d}) \tag{5.31}$$

Similarly, the upper bound, defined at $C_{inv} = 100C_{ox}$, can also be calculated.

$$V_{G2} = V_{th} + nV_{tm}\ln(\frac{100C_{ox}}{C_d}) \tag{5.32}$$

The width of this transition region ($= V_{G2} - V_{G1}$) is about 0.2 - 0.3V depending on the oxide thickness.

An effective gate voltage V'_{GS} is created in this region by using a cubic spline function of V_{GS}.

$$V'_{GS} = C_0 + C_1 V_{GS} + C_2 V_{GS}^2 + C_3 V_{GS}^3 \tag{5.33}$$

where the coefficients C_i's are to be determined from the boundary conditions. The drain current equation used in the transition region is the same as that in strong-inversion region except V'_{GS} is used instead of V_{GS}. This cubic spline function also serves as a means to acquire a smooth transition from the subthreshold region to the strong-inversion region. The reason to use V'_{GS} for the cubic spline function in this region rather than using the drain current directly is to simplify the boundary conditions. This approach avoids the need to determine whether the linear region or the saturation region drain current equations should be used at the upper bound.

(B) Boundary conditions

The boundary conditions for this cubic spline function are chosen so that the drain current and its first derivative are continuous at both bounds.

At the lower bound, $V_{GS} = V_{G1}$, the boundary conditions are

$$V'_{GS} = \sqrt{2aK}V_{tm}e^{\frac{V_{G1}-V_{offset}}{2nV_{tm}}}(1-\exp^{\frac{-V_{DS}}{V_{tm}}})^{1/2}$$

$$\frac{dV'_{GS}}{dV_{GS}} = \frac{V_{G1}}{2nV_{tm}} \tag{5.34}$$

and at the upper bound, $V_{GS} = V_{G2}$, the boundary conditions are

$$V'_{GS} = V_{G2}$$

$$\frac{dV'_{GS}}{dV_{GS}} = 1 \tag{5.35}$$

Substituting (5.34) and (5.35) into (5.33), the coefficients of the cubic spline function, C_i's can be determined. The solutions for these coefficients are given in Appendix D.

(C) Modeling results

The measured and calculated subthreshold characteristics of an n- and p-channel devices are shown in Fig. 5.17 and Fig. 5.18. In Fig. 5.17, three regions marked A, B, and C for the $V_{BS} = 0V$ curve indicate the subthreshold, transition, and strong-inversion regions where the three different drain current equations were used. A smooth transition between the regions is observed.

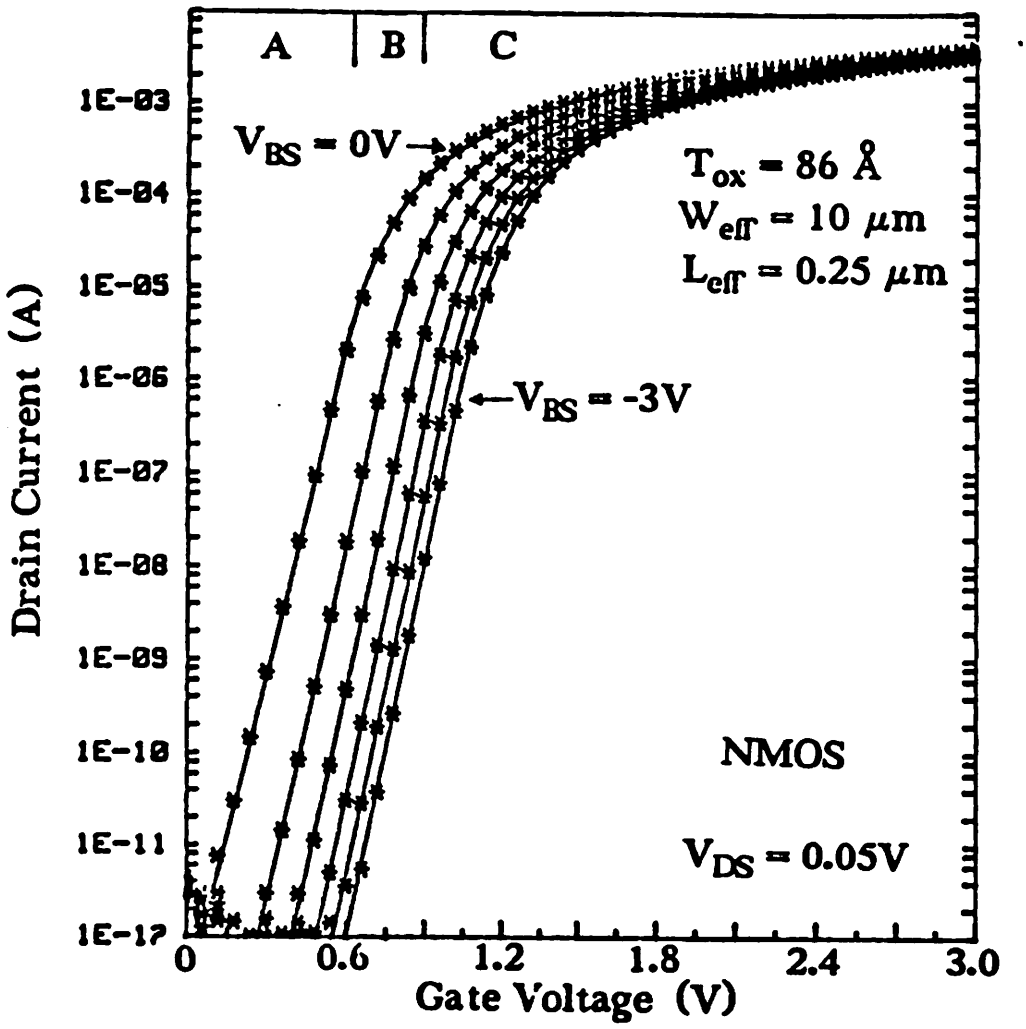


Fig. 5.17 BSIM2 modeling results in the subthreshold region for an n-channel device. The asterisks are measured data and the solid lines are the model.

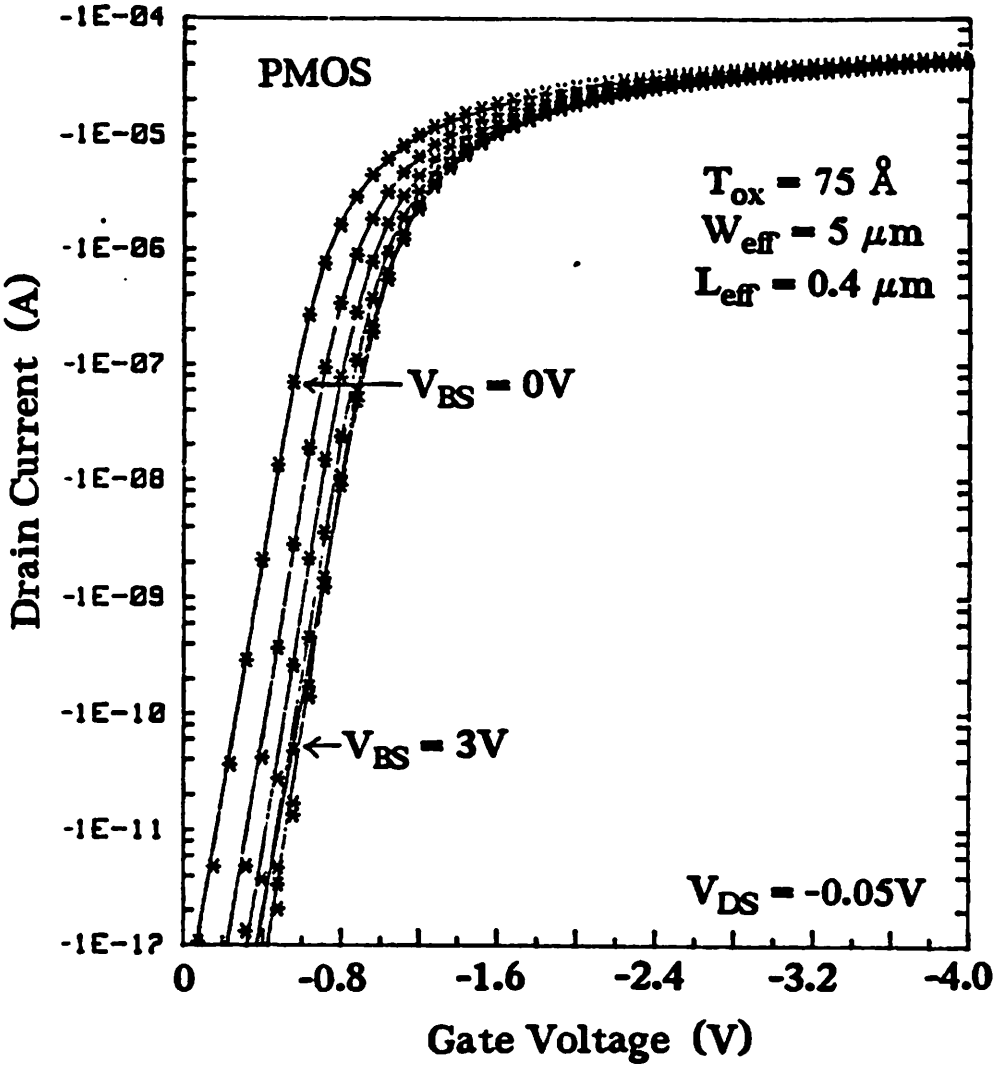


Fig. 5.18 BSIM2 modeling results in the subthreshold region for a p-channel device. The asterisks are measured data and the solid lines are the model.

5.3.4 Output resistance modeling

Although the output resistance R_{out} is one of the most important parameters in analog circuit designs, few MOSFET models can simulate R_{out} accurately. Typical output resistance characteristic of a short-channel MOSFET is shown in Fig. 5.19. The shape of the output resistance in the saturation region can be attributed to three mechanisms: drain-induced barrier lowering, channel length modulation, and hot-electron induced output resistance reduction. In most MOSFET models, the drain-induced-barrier lowering effect is included in the threshold voltage. In BSIM2 this effect is modeled by the η parameter, therefore in this section, only the other two effects will be discussed.

(A) Channel length modulation

When the drain voltage is larger than V_{DSAT}, the velocity saturation region extends toward the source, which effectively reduces the channel length and results in a non-zero channel conductance (or finite R_{out}) in the saturation region [5.29,5.30]. A schematic diagram of the channel length modulation is shown in Fig. 5.20.

Channel length modulation is the dominant mechanism affecting R_{out} when V_{DS} is near V_{DSAT}. This effect is usually implemented in the model through the β_0 parameter. Because of the reduced channel length, the channel conductance coefficient, β_0, increases as V_{DS} increases.

$$\beta_0' = \beta_0\left(\frac{1}{1-\Delta L_{CLM}/L_{eff}}\right) \approx \beta_0\left(1+\frac{\Delta L_{CLM}}{L_{eff}}\right) \tag{5.36}$$

where ΔL_{CLM} is the channel length reduction due to the channel length modulation effect. According to a quasi-2D analysis [5.31,5.32], ΔL_{CLM} is zero when the drain voltage is less than V_{DSAT} and increases logarithmically with V_{DS} after V_{DSAT}. A qualitative result of the quasi-2D analysis is plotted by the solid curve in Fig. 5.21. In practice, however, it is found empirically that the combination of a hyperbolic tangent function and a quadratic function is more accurate and appropriate to model R_{out} when V_{DS} is near V_{DSAT}.

$$\beta_0' = \beta_0 + \beta_1 \tanh\left(\frac{\beta_2 V_{DS}}{V_{DSAT}}\right) + \beta_3 V_{DS} - \beta_4 V_{DS}^2 \tag{5.37}$$

where β_0 and β_s are the conductance coefficients extracted at linear and saturation regions, respectively, β_2, β_3, and β_4 are fitting parameters, and $\beta_1 = \beta_s - (\beta_0 + \beta_3 V_{DD} - \beta_4 V_{DD}^2)$. A qualitative sketch (5.37) is plotted by the dashed curve in Fig. 5.21. In (5.37), the conductance coefficient β starts to increase from $V_{DS} = 0$ which is different from the quasi-2D analysis. Eq. (5.37) also eliminates the discontinuity problem at V_{DSAT}. The effect of each parameter in (5.37) on the shape of R_{out} is indicated in Fig. 5.22 which also reveals the means in which these parameters can be extracted from measured data.

(B) Hot-electron-induced output resistance reduction

As the drain voltage increases beyond V_{DSAT}, the peak electric field in the velocity saturation region increases sharply and electron-hole pairs are generated due to impact ionization [5.33,5.34]. The holes generated are collected by the substrate is referred to as the substrate current I_{SUB}. When the substrate current flows through the substrate, it slightly forward biases the source junction with respect to the substrate because of the ohmic voltage drop V'_{BS} ($= I_{SUB}R_{SUB}$). This positive body bias (for NMOS) reduces the threshold voltage causing the drain current to increase and in turn degrades the output resistance [5.3]. This process is depicted in Fig. 5.23. Therefore, to the first-order approximation, the resultant drain saturation current I'_{DSAT} due to the hot-electron effects can be expressed as

$$I'_{DSAT} = I_{DSAT}(1 + CI_{SUB}R_{SUB}) \tag{5.38}$$

where I_{DSAT} is the drain saturation current without hot-electron effect given by (5.6), R_{SUB} is the effective resistance of the substrate, and C is a constant. Substituting the substrate current equation, (3.24), into (5.38), Eq. (5.38) can be rewritten as

$$= I_{DSAT}\left[1 + A_i e^{\left(\frac{-B_i}{V_{DS} - V_{DSAT}}\right)}\right] \tag{5.39}$$

where A_i and B_i are impact ionization coefficients. Under normal bias conditions, the second term in the brackets is much smaller than one, hence (5.39) has very little effect on the magnitude of the calculated drain current, but it has significant effect on the output resistance as will

be shown later.

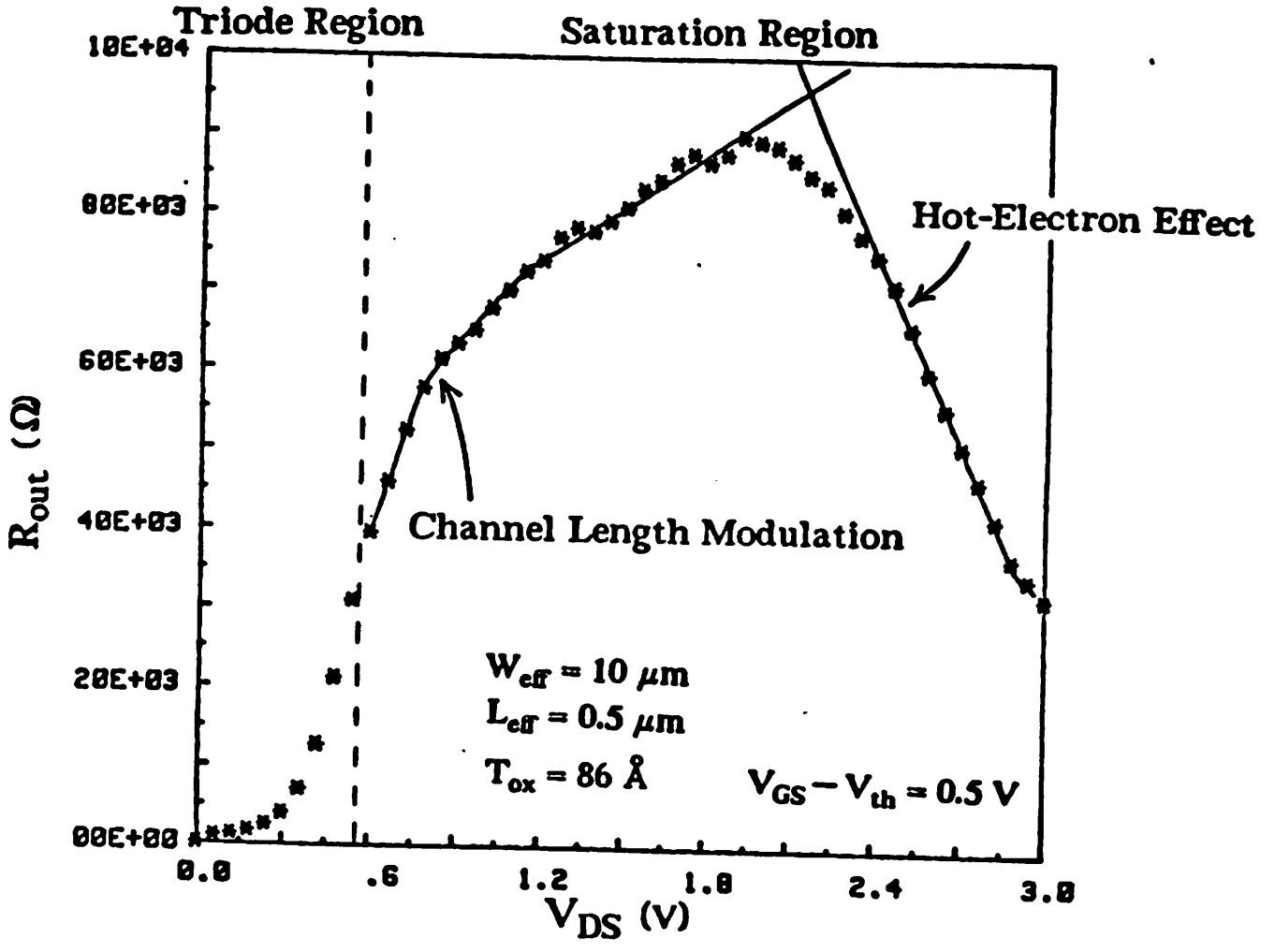
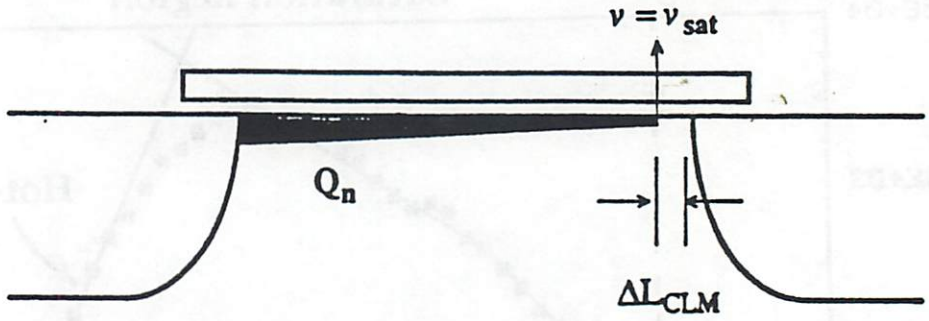


Fig. 5.19 Typical output resistance characteristics of a short-channel MOSFET.



$$L_{eff} \rightarrow L_{eff} - \Delta L_{CLM}$$

$$\beta_0 \rightarrow \beta_0 \left(\frac{L_{eff}}{L_{eff} - \Delta L_{CLM}} \right)$$

Fig. 5.20 A schematic diagram showing the channel length modulation effect.

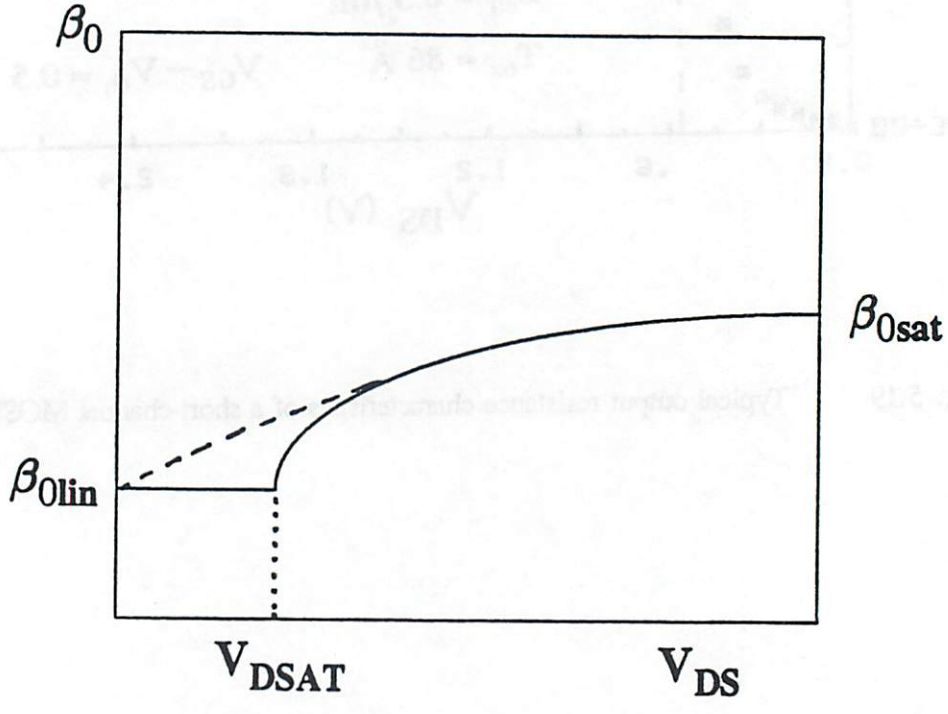


Fig. 5.21 Qualitative plots of the conductance coefficient versus drain voltage. The solid curve is the result of a quasi 2-D analysis and the dashed curve is the BSIM2 model.

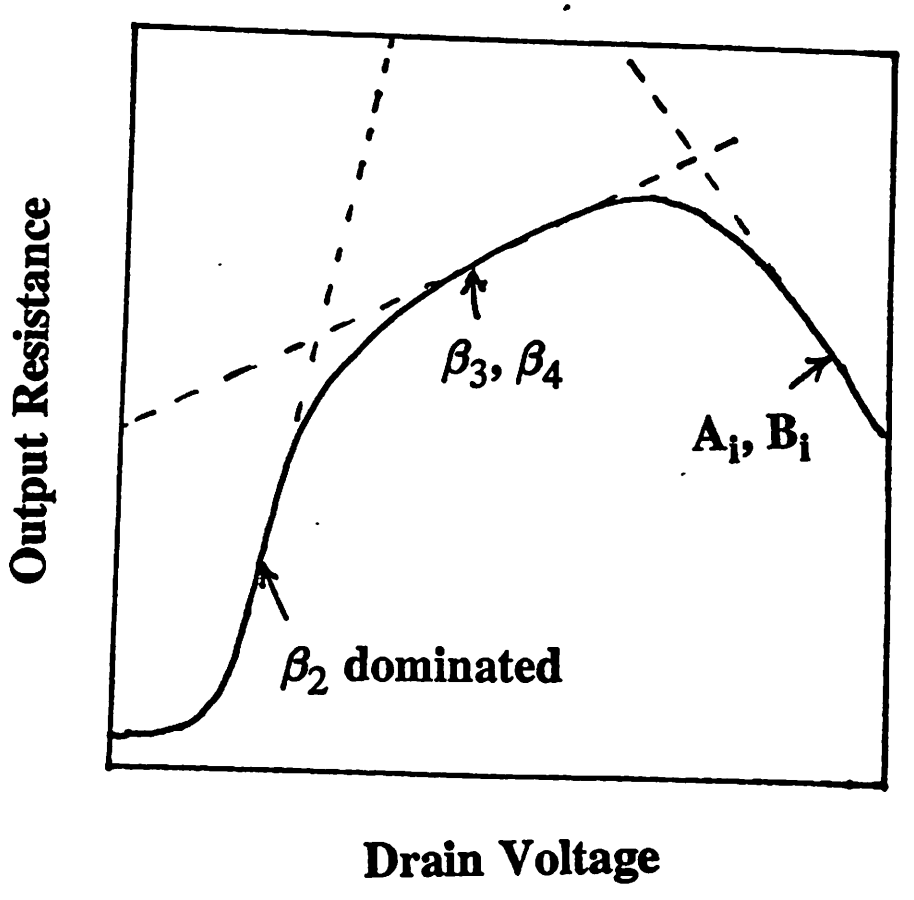
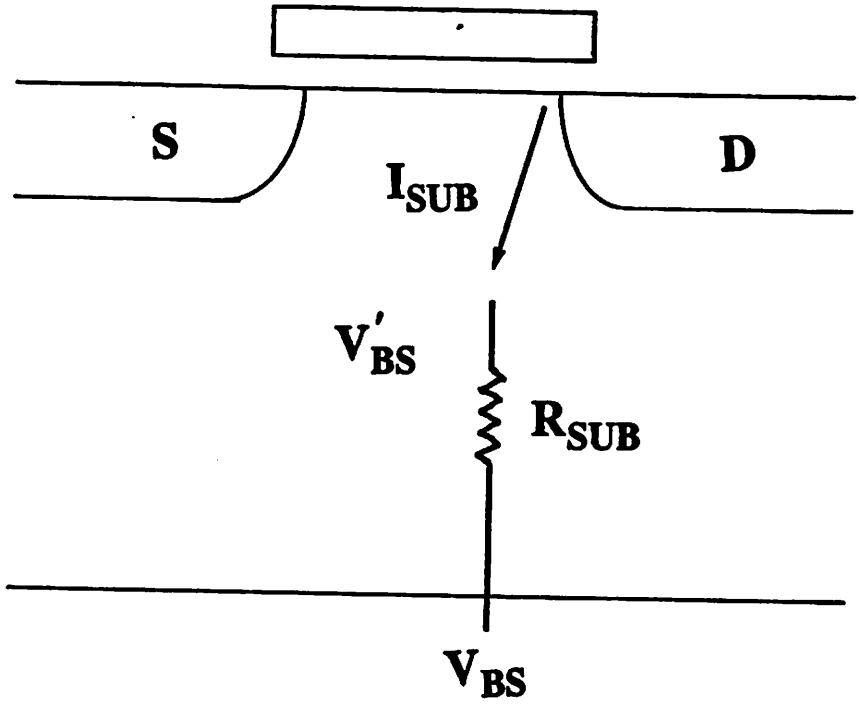


Fig. 5.22 Output resistance versus drain voltage showing the effects of various BSIM2 parameters on the shape of the output resistance.



$$V'_{BS} = V_{BS} + I_{SUB} R_{SUB}$$

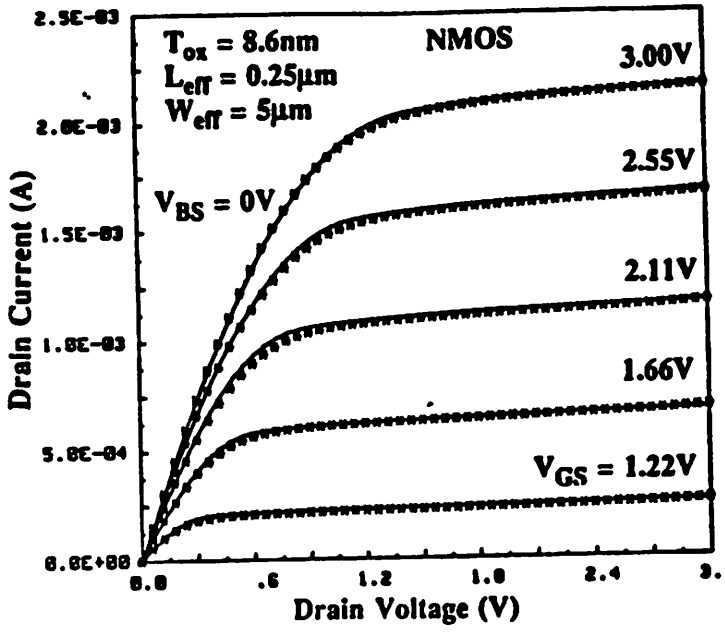
Fig. 5.23 A schematic diagram showing the process of output resistance reduction due to hot-electron effects.

(C) Modeling results

Fig. 5.24a shows measured and calculated I_{DS} - V_{DS} characteristics of a quarter-micron n-channel MOSFET. The corresponding output resistance is shown in Fig. 5.24b. The accurate I_{DS} and R_{out} modeling results make BSIM2 highly suitable for both digital and analog applications.

As illustrated in the design curves developed in chapter 4, hot-electron reliability has become the major concern in deep-submicrometer device and circuit design. To predict the device lifetime and/or aged circuit behavior due to hot-electron effects have also become one of the design steps in VLSI/ULSI systems. Recently, various activities in this area have been reported [5.35-5.38]. More programs are expected to be developed in the future. In these programs, model parameters from both fresh and hot-electron stressed transistors are usually required. Therefore, a MOSFET model for the future should serve well for this purpose. To test BSIM2's capability in modeling stressed devices, the device used in Fig. 5.24 was purposely degraded by hot-electron stress to generate a threshold voltage shift of 0.22 volts, then model parameters were re-extracted from this device. The modeling results are shown in Fig. 5.25. Again, very good agreement between measured and calculated I_{DS} and R_{out} are observed which proves that BSIM2 is also a potential candidate for circuit aging related simulations.

Since the LDD structure has become common in current technologies, a MOSFET model will not be useful if it fails to model LDD devices. BSIM2's capability in this aspect is verified in Fig. 5.26 where measured and calculated I_{DS} and R_{out} for an LDD transistor are shown. Finally, the modeling results for a non-LDD p-channel transistor are shown in Fig. 5.27.



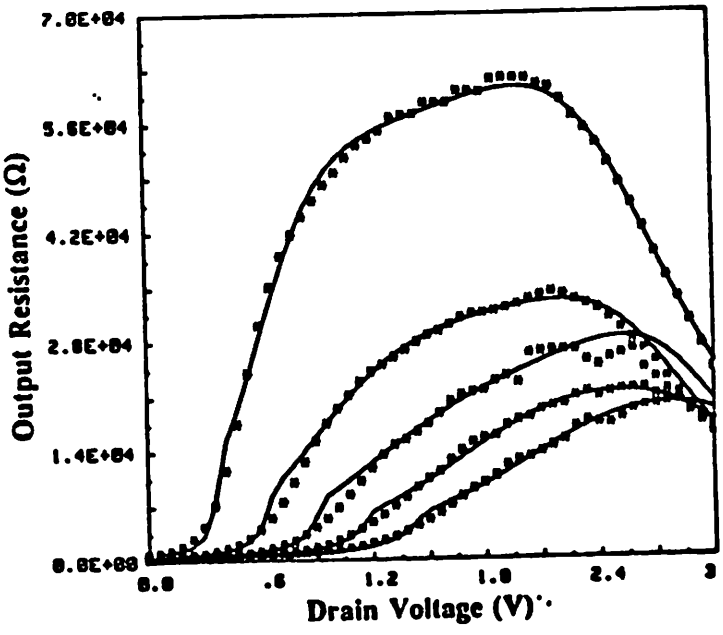


Fig. 5.24 (a) BSIM2 I_{DS} - V_{DS} modeling results for a quarter-micron n-channel device, (b) the corresponding output resistance. The asterisks are measured data and the solid lines are the model.

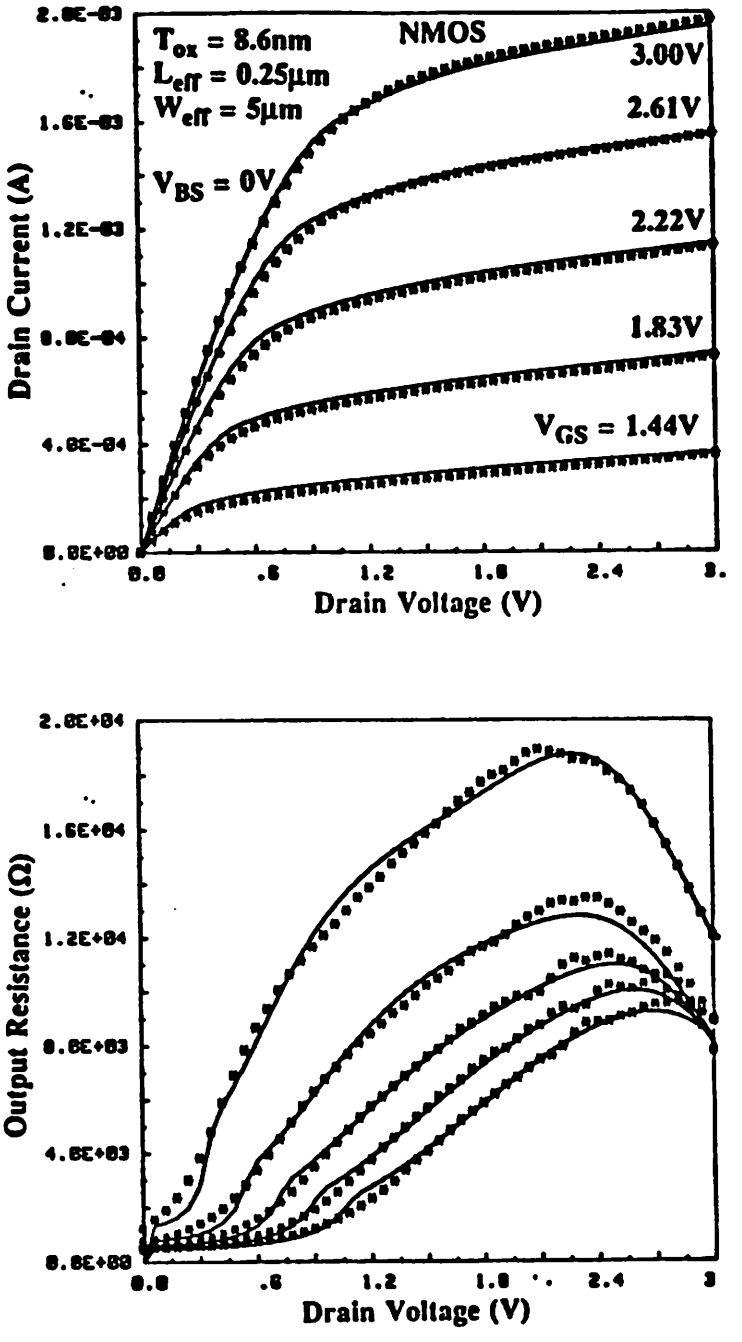


Fig. 5.25 (a) BSIM2 I_{DS} - V_{DS} modeling results for a quarter-micron n-channel device after hot-electron stress. The threshold voltage shift is 0.22V. (b) the corresponding output resistance. The asterisks are measured data and the solid lines are the model.

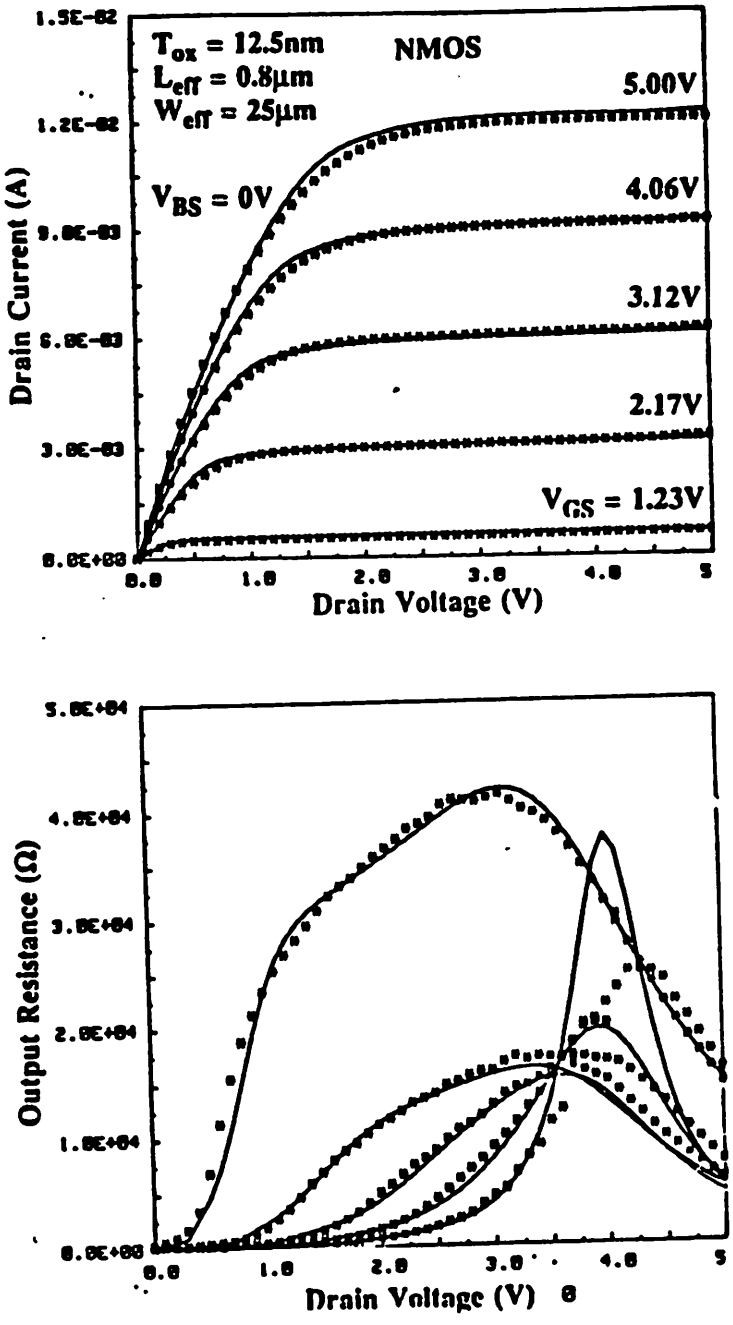


Fig. 5.26 (a) BSIM2 I_{DS} - V_{DS} modeling results for an n-channel LDD device, (b) the corresponding output resistance. The asterisks are measured data and the solid lines are the model.

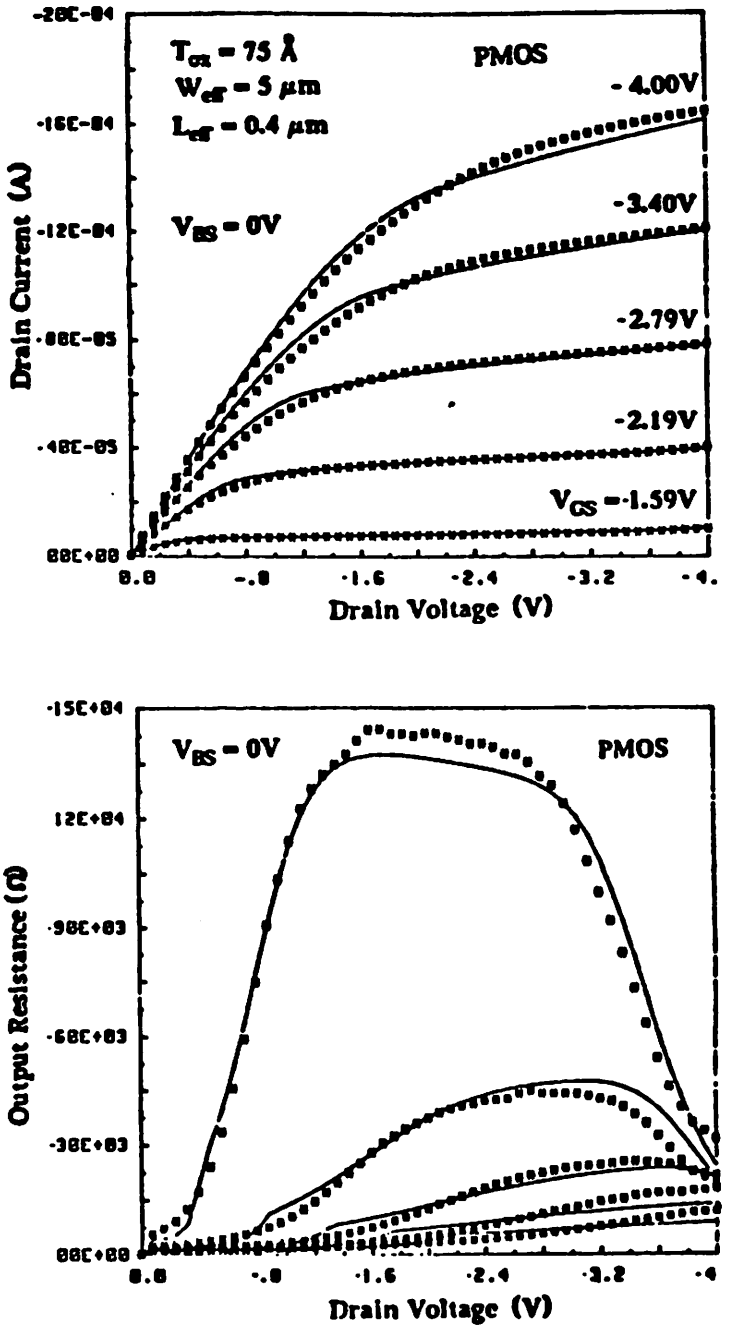


Fig. 5.27 (a) BSIM2 I_{DS} - V_{DS} modeling results for a p-channel device, (b) the corresponding output resistance. The asterisks are measured data and the solid lines are the model.

5.3.5 Bias-dependent parameters

So far, 18 BSIM2 parameters have been described. Some of these parameters are found to be slightly bias-dependent which are approximated by linear functions of biases. All of these parameters, along with the equations demonstrating their bias-dependencies are listed below.

1. V_{FB} (5.1)
2. ϕ_S (5.1)
3. K_1 (5.1)
4. K_2 (5.1)
5. $\eta = \eta_0 + \eta_B V_{BS}$ (5.1)
6. $\beta_0 = \beta_{00} + \beta_{0B} V_{BS}$ (5.19)
7. $\beta_s = \beta_{s0} + \beta_{sB} V_{BS}$ (5.37)
8. $\beta_2 = \beta_{20} + \beta_{2B} V_{BS} + \beta_{2G} V_{GS}$ (5.37)
9. $\beta_3 = \beta_{30} + \beta_{3B} V_{BS} + \beta_{3G} V_{GS}$ (5.37)
10. $\beta_4 = \beta_{40} + \beta_{4B} V_{BS} + \beta_{4G} V_{GS}$ (5.37)
11. $U_a = U_{a0} + U_{aB} V_{BS}$ (5.18)
12. $U_b = U_{b0} + U_{bB} V_{BS}$ (5.18)
13. $U_1 = U_{10} + U_{1B}$ (5.14)
14. U_{1D} (5.14)
15. $n = n_0 + n_B / \sqrt{1+(\phi_S - V_{BS})} + n_D V_{DS}$ (5.27)
16. $V_{offset} = V_{offset0} + V_{offsetB} V_{BS} + V_{offsetD} V_{DS}$ (5.27)
17. $A_i = A_{i0} + A_{iB}$ (5.39)
18. $B_i = B_{i0} + B_{iB}$ (5.39)

Note that the bias-dependencies of parameters 5 and 15 are different from those in BSIM1. They are more physical now. E_{c0} and E_{cD} in (5.14) have been replaced by U_1 and U_{1D},

respectively as in BSIM1, where $U_1 = 1/E_{c0}L_{eff}$.

A complete list of the BSIM2 model equations, the associated expressions for transconductance, g_m, output resistance, R_{out}, and body (back-gate) transconductance, g_{bs}, for all regions, and the meaning of each parameter are given in Appendix D.

5.3.6 Size-independent parameters

The parameters extracted from a test device only pertain to that particular device size. The parameter set for each device size is referred to as a "parameter file". Several size-dependent "parameter files" can be processed to generate a set of size-independent parameters called a "process file". The equation used in BSIM1 to generate the "process file" is given in (5.40).

$$P(L_i, W_i) = P_0 + \frac{P_L}{L_i} + \frac{P_W}{W_i} \tag{5.40}$$

where $P(L_i, W_i)$ is a parameter for a particular effective length L_i and width W_i, P_0, P_L, and P_W are the size-independent parameters. A schematic diagram of this procedure is shown in Fig. 5.28. The size-independent parameters are generated by fitting parameter files with different device dimensions (L_i's and W_i's) to (5.40). Once P_0, P_L, and P_W are known, the model parameters for a device with any channel length and width can be calculated from (5.40) by replacing L_i and W_i with the desired dimensions. Detailed description of this procedure can be found in [5.11].

Although (5.40) does not work well when the range of the device dimensions is wide, for example $L_{max}/L_{min} > 10$, it is still kept in BSIM2 for the time being. If a wide range of device dimensions has to be used in a circuit design, breaking the process file into two or more process files with smaller device dimension ranges is recommended. Studies on improving (5.40) are underway.

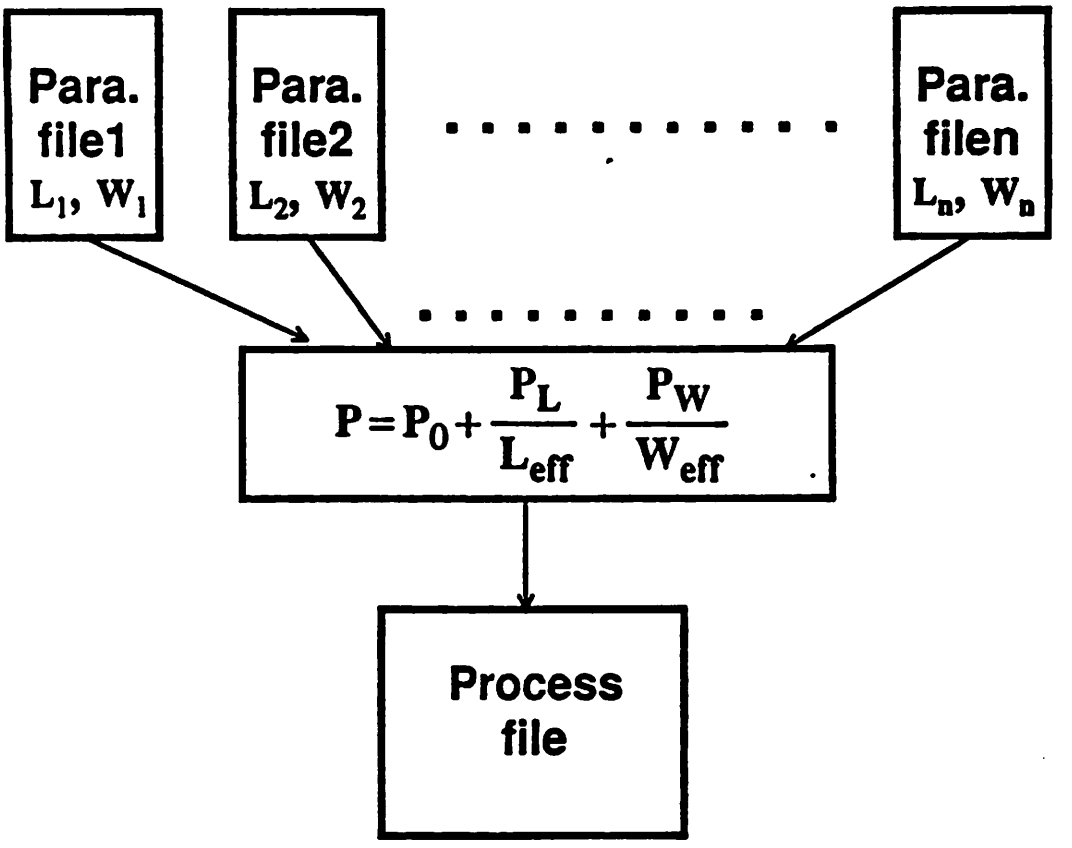


Fig. 5.28 A schematic diagram illustrating how a process file is generated.

5.4 Parameter extraction for BSIM2

An IBM PC-based integrated system has been developed for automated extraction of BSIM2 parameters. In this section, only a general overview of the system is described. A complete description of the parameter extraction system, user's guide, and examples are given in [5.11].

5.4.1 Automated parameter extraction system

(A) System configuration

A schematic diagram of the system hardware in shown in Fig. 5.29. This system consists of three major parts: an IBM PS/2 (model 50 or higher) or a PC-AT computer with a VGA graphics card running under DOS 3.0 or higher (DOS 4.0 or higher to enable VGA screen dump), an HP4145 parametric analyzer, and a manual probe station. An IOtech GP488/2 interface board is required in the computer to communicate with the HP4145 parametric analyzer. The extraction program is written in Microsoft C version 5.1 with modified IOtech Personal488/2 modules. The executable code is about 310KB.

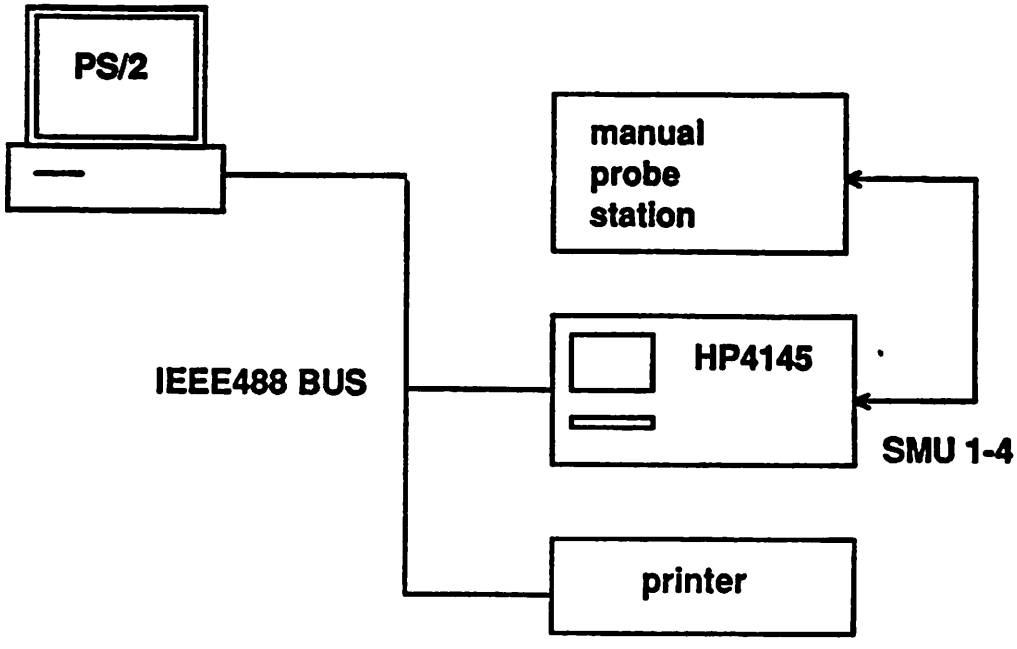
(B) System operation

Fig. 5.30 shows the flowchart of the parameter extraction program. The program is modular and menu-driven for easy operation and future modification. The only required user-supplied inputs are the device dimensions, die location, supply voltages, and the SMU (Stimulus-Measurement Unit) connections for the HP4145 parameter analyzer. The functions of the parameter extraction system are grouped into two categories: single device mode and multiple device mode. In the single device mode, parameter files are extracted from each device. In the multiple device mode, a process file is created from user selected parameter files. In either mode, calculated I-V characteristics can be displayed together with measured results for comparison. Device I-V data can be either measured directly from physical devices through the HP4145 parametric analyzer or read back from a hard/floppy disk. If the I-V data are to

be measured from a device, certain device functionality tests are performed before the formal extraction procedures begin to determine the device type and screen out devices with problematic characteristics. Four sets of I-V data are required to extract all the parameters. They are: I_{DS} - V_{GS} in the linear region (V_{DS} = 0.1V), I_{DS} - V_{GS} in the saturation region (V_{DS} = V_{DD}), and two I_{DS} - V_{DS}'s at two different substrate biases (V_{BS} = 0 and V_{BS} = -V_{DD}). The first two data sets are used to extract drain-current related parameters, (5.1) and (5.13)-(5.35), and the last two data sets are used to extract output-resistance related parameters, (5.37)-(5.39). The playback feature of the program then allows the users to check the quality of the extracted parameters for each device. This extraction process is repeated every time the single device model operation is executed. The total extraction time for each device is about 20 - 30 seconds without measurement and about 2 minutes with measurement on a PS/2 Model 50 computer. After a few devices have been extracted, a process file may be created.

5.4.2 Extraction algorithms

The most commonly used technique in parameter extraction is the nonlinear global optimization [5.39-5.42]. Although global optimization will give the minimum average error between calculated and measured results, the extracted parameter values may not be physically meaningful which makes interpolation or extrapolation of parameters for other device dimensions very difficult and unreliable. Also, the optimization processes are usually slow, since the computation time increases drastically with the number of parameters. Most of all, if some parameters are mutually correlated in the model, the optimization process becomes difficult to converge or may result in non-unique solutions. Various auxiliary methods such as Levenberg-Marquardt method [5.43,5.44], modified Gauss method [5.45,5.46], and the steepest descent method [5.47], etc., are incorporated into the global optimization process to expedite the convergence or to minimize the effect of parameter redundancy. Therefore, parameter extraction programs adopting the global optimization technique are always quite complicated.



Hardware Platform

- IBM model 50 (or higher)
- IOtech GP488/2 interface board
- HP4145B parametric analyzer
- VGA graphics capability
- Parallel Printer

Fig. 5.29 A schematic diagram of the BSIM2 parameter extraction system hardware configuration.

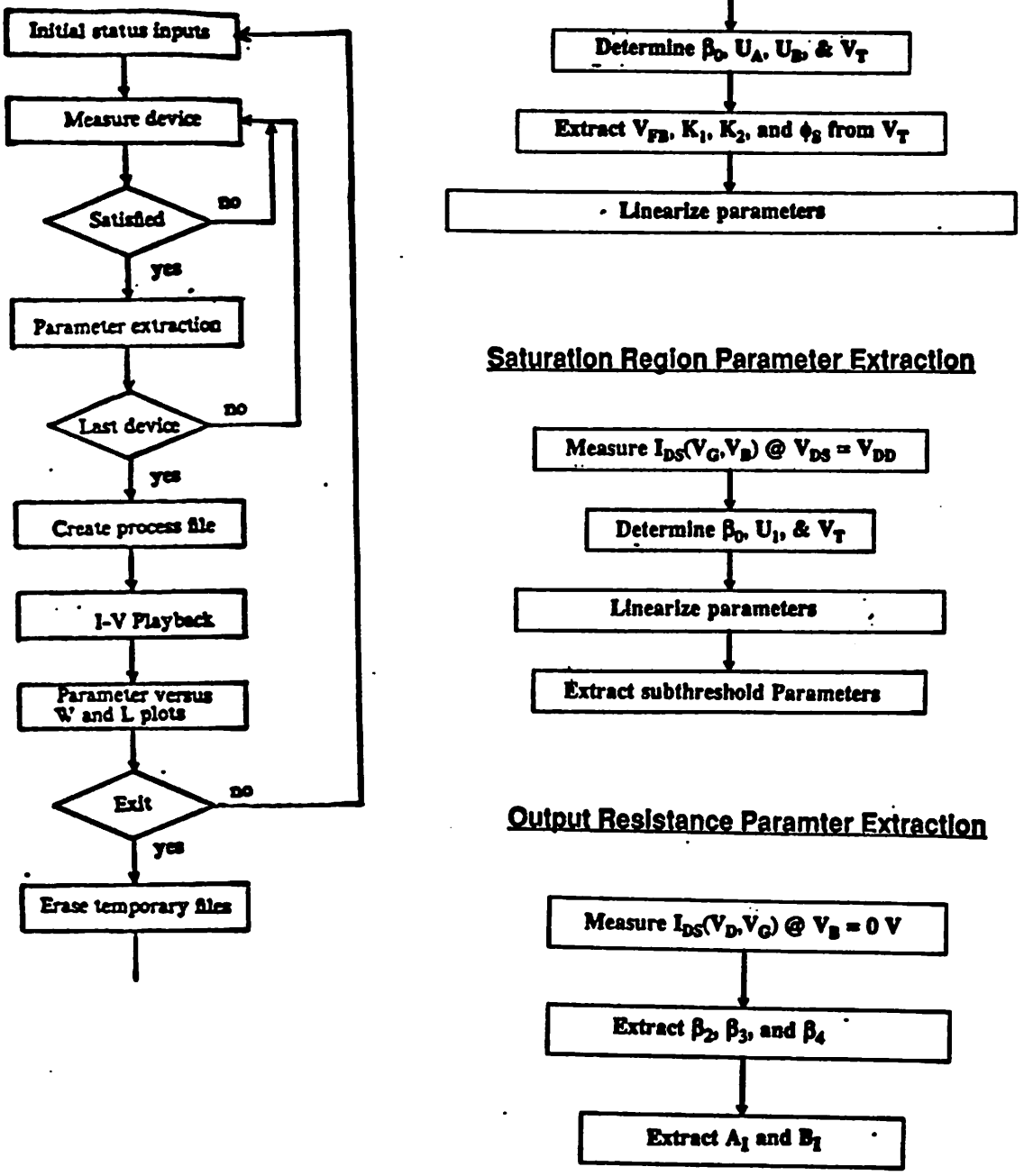


Fig. 5.30 Flowchart of the BSIM2 parameter extraction program.

Due to the mathematically compact functional forms of the BSIM2 model, a local optimization technique together with a physics-based parameter extraction method can be employed to extract BSIM2 parameters. Because no sophisticated algorithms are needed with this approach, the parameter extraction program can be easily implement and the parameters extracted are also more physical and reliable than those from global optimizations. With local optimization, only two or three parameters are extracted at a time under a certain bias condition and the optimization process is repeated to cover all operation regions until all of the parameters are extracted. Since only two or three parameters are optimized each time, the optimization process is very fast. The non-convergence and non-uniqueness problems usually do not exist.

In the BSIM2 parameter extraction program, the only optimization process is a combination of Newton-Raphson's iteration and a linear least-square fit routine with two or three variables. The flowchart of the optimization process is shown in Fig. 5.31. The model equations are first arranged in a form suitable for Newton-Raphson's iteration as shown in (5.41).

$$f(P_{1o},P_{2o},P_{3o})-f(P_1^{(m)},P_2^{(m)},P_3^{(m)})=\frac{\partial f}{\partial P_1}\Delta P_1^{(m)}+\frac{\partial f}{\partial P_2}\Delta P_2^{(m)}+\frac{\partial f}{\partial P_3}\Delta P_3^{(m)} \qquad (5.41)$$

where f() is the function to be optimized, P_1, P_2, and P_3 are the parameters to be extracted, P_{1o}, P_{2o}, and P_{3o} stand for the true parameters that we are looking for, $P_1^{(m)}$, $P_2^{(m)}$, and $P_3^{(m)}$ represent the parameter values after the m^{th} iteration. In real case, the function f() may be the drain current equation or its variational form, and $f(P_{1o},P_{2o},P_{3o})$ would be measured I-V data. To make (5.41) ready for the linear least-square fit routine (a form of $y=a+bx_1+cx_2$), both sides of (5.41) are first divided by $\partial f/\partial P_1$. Then the measured I-V data are fitted to (5.41) and the increments of parameters for next iteration, $\Delta P_i^{(m)}$s, are determined. The parameter values for the $(m+1)^{th}$ iteration are given by

$$P_i^{(m+1)}=P_i^{(m)}+\Delta P_i^{(m)} \qquad i=1,2,3 \qquad (5.42)$$

This procedure is repeated until all ΔP_i's are smaller than some pre-determined values, at this point the optimization process is considered converged.

The procedures for extracting parameters β_0, U_a, and U_b are describe below as an example showing how the optimization process is physically applied. First of all, the drain current equation in the linear region (5.19) is re-arranged in such a form suitable for (5.41).

$$f(\beta_0, U_a, U_b) = \beta_0 \frac{G(1 + U_a(V_{GS} - V_{th}) + U_b(V_{GS} - V_{th})^2)}{V_{GS} - V_{th} - \frac{a}{2}V_{DS}} = 0 \tag{5.43}$$

where $G = I_{DS}/V_{DS}$ is the measured channel conductance. The U_1V_{DS} term in the denominator of (5.19) has been dropped in (5.43), since V_{DS} is small (0.1V) compared to V_{GS}. Substituting (5.43) into (5.41), the equation used in the Newton-Raphson's iteration can be obtained.

$$0 - f(\beta_0^{(m)}, U_a^{(m)}, U_b^{(m)}) = \Delta\beta_0^{(m)} + \frac{\partial f}{\partial U_a}\Delta U_a^{(m)} + \frac{\partial f}{\partial U_b}\Delta U_b^{(m)} \tag{5.44}$$

where

$$\frac{\partial f}{\partial U_a} = \frac{G(V_{GS} - V_{th})}{V_{GS} - V_{th} - \frac{a}{2}V_{DS}} \tag{5.45}$$

and

$$\frac{\partial f}{\partial U_b} = \frac{G(V_{GS} - V_{th})}{V_{GS} - V_{th} - \frac{a}{2}V_{DS}} \tag{5.46}$$

During each iteration, the same measured I_{DS} - V_{GS} data are fitted through (5.44) to calculate $\Delta\beta_0^{(m)}$, $\Delta U_a^{(m)}$, and $\Delta U_b^{(m)}$ for next iteration. The iteration is terminated when the increments of all three parameters are less than 0.01% of their current values or when a pre-set maximum iteration number is reached.

For each substrate voltage, a set of β_0, U_a, and U_b are extracted. These parameter values may exhibit slight substrate bias dependence and are fitted through a linear equation as listed in section 5.3.5 to extract the substrate-bias dependent parameters.

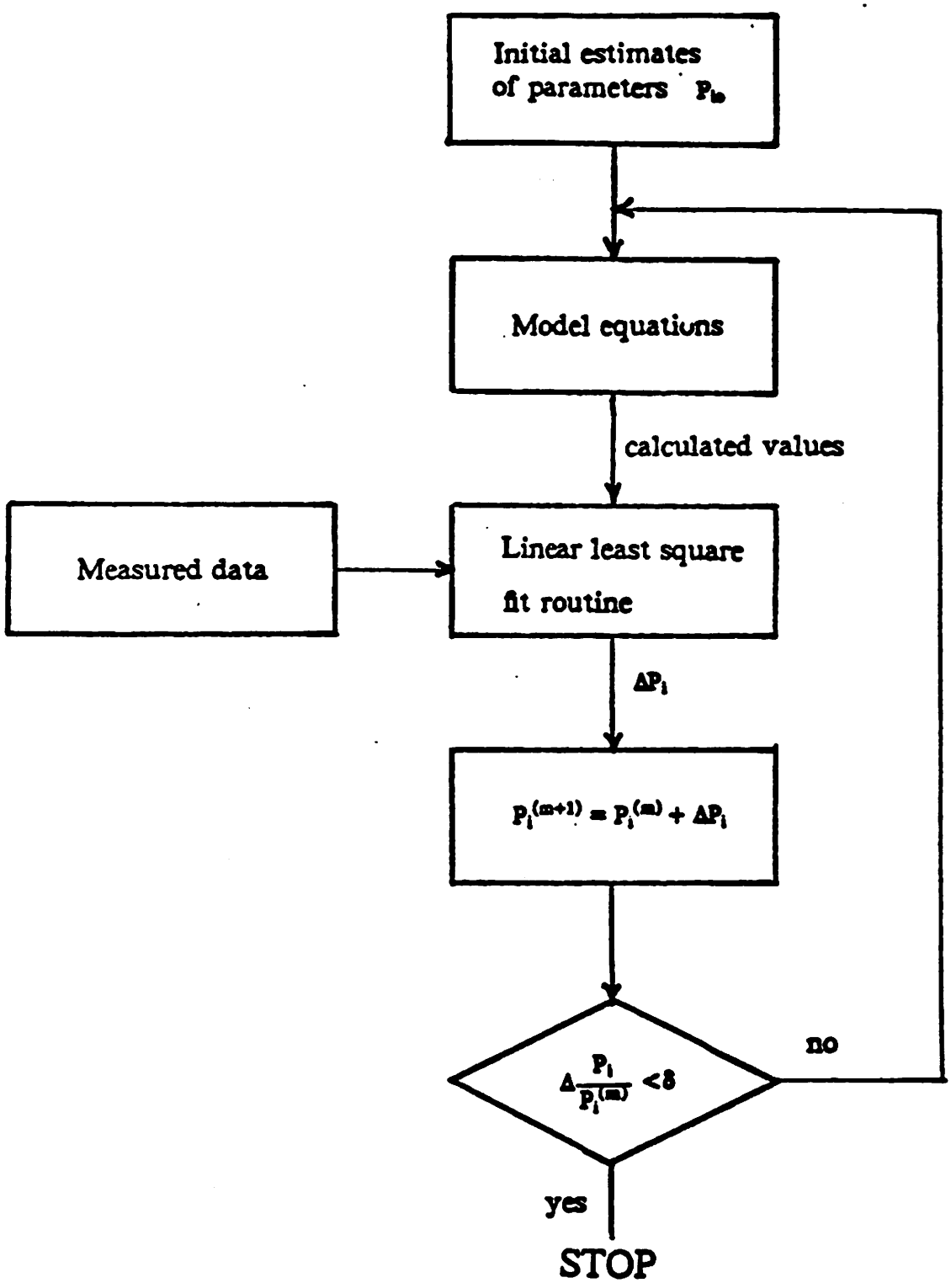


Fig. 5.31 Flowchart of the optimization process in the BSIM2 parameter extraction program.

5.5 SPICE simulation results

The BSIM2 model has been implemented in SPICE3.C1. A program list of the BSIMevaluate.c module, which evaluates the BSIM2 model equations in SPICE3, is given in Appendix E. An example of SPICE input deck with BSIM2 model parameters is shown in Fig. 5.32. BSIM2 model has been used to simulate an NMOS enhancement-depletion type ring oscillator with T_{ox} = 8.6nm and L_{eff} = 0.25μm. The simulated delay time is 24ps/stage, very close to the measured data of 22ps/stage.

The accuracy of the BSIM2 model has been demonstrated in section 5.3. More results can be found in [5.11]. In this section, the computational efficiency of the model is discussed. Table 5.2 shows the comparison of SPICE simulation times required for BSIM2, level-1, and level-2 models on two typical circuits. The device dimensions used in the simulation were T_{ox} = 25nm and L_{eff} = 2μm for level-1 and level-2 models; T_{ox} = 8.6nm and L_{eff} = 0.25, 0.5, and 0.7μm for BSIM2. The parameters for level-1 and level-2 models are typical values for 2-μ technology provided by MOSIS.

The first simulation is a DC analysis of five I_{DS} - V_{DS} characteristics. Although the evaluation time of BSIM2 model itself is slower than that of level-1 model and comparable to that of level-2 model, the total computation times required for BSIM2 in this simulation is similar to that of level-1 model and twice as fast as that of level-2 model. This is because the number of iterations needed in the simulation for BSIM2 is less than those of the other two models due to BSIM2's smoother functional form. The total number of iterations for each model are also shown in Table 5.2. The second simulation is the transient analysis of a 15-stage CMOS ring oscillator. Similar results were also observed. More simulations on analog circuits are underway.

```
AN 11-STAGE CMOS RING OSCILLATOR WITH TOX=86A AND LEFF=0.3UM
.SUBCKT INV 1 2 3
C1 3 0 0.1P
M1 3 2 1 1 PMEN L=0.3U W=30U AD=120P AS=75P PD=36U PS=6U
M2 3 2 0 0 NMEN L=0.3U W=15U AD=60P AS=37.5P PD=23U PS=6U
.ENDS INV
X1 21 2 3 INV
X2 1 3 4 INV
X3 1 4 5 INV
X4 1 5 6 INV
X5 1 6 7 INV
X6 1 7 8 INV
X7 1 8 9 INV
X8 1 9 10 INV
X9 1 10 11 INV
X10 1 11 12 INV
X11 1 12 2 INV
VDD1 1 0 4.0
VDD 21 0 PULSE(0 4.0 0 0.5N 1N 500N)
.TRAN 60P 60N
.PLOT TRAN V(2)
.PRINT TRAN V(2)
.OPTIONS VNTOL=1E-5 ABSTOL=1E-9 ITL4=50
+ CPTIME=1E4 RELTOL=0.01 CHGTOL=1E-12 PIVTOL=1E-29 ITL1=500 ITL2=500
.OPT ACCT
.WIDTH OUT=80
.model NMEN nmos level = 4
+ vfb = -1.0 lvfb = 0.0 wvfb = 0.0
+ phi = 0.889 lphi = 0.0 wphi = 0.0
+ k1 = 0.93 lk1 = 0.0 wk1 = 0.0
+ k2 = 0.125 lk2 = 0.0 wk2 = 0.0
+ eta0 = 0.017 leta0 = 0.0 weta0 = 0.0
+ etab = -0.007 letab = 0.0 wetab = 0.0
+ mu0 = 327.3 dl = 0.0 dw = 0.0
+ mu0b = -8.42 lmu0b = 0.0 wmu0b = 0.0
+ mus0 = 431.8 lmus0 = 0.0 wmus0 = 0.0
+ musb = -7.4 lmusb = 0.0 wmusb = 0.0
+ mu30 = 15.1 lmu30 = 0.0 wmu30 = 0.0
+ mu3b = -1.34 lmu3b = 0.0 wmu3b = 0.0
+ mu3g = -2 lmu3g = 0.0 wmu3g = 0.0
+ mu20 = 2.37 lmu20 = 0.0 wmu20 = 0.0
+ mu2b = 0.09 lmu2b = 0.0 wmu2b = 0.0
+ ua0 = .443 lua0 = 0.0 wua0 = 0.0
+ uab = -0.025 luab = 0.0 wuab = 0.0
+ ub0 = 0.075 lub0 = 0.0 wub0 = 0.0
+ ubb = -0.0076 lubb = 0.0 wubb = 0.0
+ u10 = 0.18 lu10 = 0.0 wu10 = 0.0
+ u1b = 0.00014 lu1b = 0.0 wu1b = 0.0
+ n0 = 1.125 ln0 = 0.0 wn0 = 0.0
+ nb = 0.35 lnb = 0.0 wnb = 0.0
+ nd = -0.017 lnd = 0.0 wnd = 0.0
+ vof0 = 1.16 lvof0 = 0.0 wvof0 = 0.0
+ vofb = -0.034 lvofb = 0.0 wvofb = 0.0
+ vofd = -0.069 lvofd = 0.0 wvofd = 0.0
+ ai0 = 332.68 lai0 = 0.0 wai0 = 0.0
+ aib = 108.55 laib = 0.0 waib = 0.0
```

```
+ bi0 = 24.62 lbi0 = 0.0 wbi0 = 0.0
+ bib = 2.92 lbib = 0.0 wbib = 0.0
+ vghigh = 0.232 lvghigh = 0.0 wvghigh = 0.0
+ vglow = -0.114 lvglow = 0.0 wvglow = 0.0
+ tox = 8.6e-3 temp = 27 vdd = 3 vgg = 4 vbb = -3
+ cgdo = 2.0e-10 cgso = 2.0e-10 cgbo = 5.0e-11
+ xpart = 0
+ rsh = 10 cj = 0.0002 cjsw = 1.0e-10
+ js = 5e-5 pb = 0.7 pbsw = 0.8
+ mj = 0.5 mjsw = 0.33 wdf = 0
+ dell = 0
.model PMEN pmos level = 4
+ vfb = -1.0 lvfb = 0.0 wvfb = 0.0
+ phi = 0.889 lphi = 0.0 wphi = 0.0
+ k1 = 0.93 lk1 = 0.0 wk1 = 0.0
+ k2 = 0.125 lk2 = 0.0 wk2 = 0.0
+ eta0 = 0.017 leta0 = 0.0 weta0 = 0.0
+ etab = -0.007 letab = 0.0 wetab = 0.0
+ mu0 = 131 dl = 0.0 dw = 0.0
+ mu0b = -3.42 lmu0b = 0.0 wmu0b = 0.0
+ mus0 = 173.8 lmus0 = 0.0 wmus0 = 0.0
+ musb = -3 lmusb = 0.0 wmusb = 0.0
+ mu30 = 6. lmu30 = 0.0 wmu30 = 0.0
+ mu3b = -0.6 lmu3b = 0.0 wmu3b = 0.0
+ mu3g = -0.8 lmu3g = 0.0 wmu3g = 0.0
+ mu20 = 2.37 lmu20 = 0.0 wmu20 = 0.0
+ mu2b = 0.09 lmu2b = 0.0 wmu2b = 0.0
+ ua0 = .443 lua0 = 0.0 wua0 = 0.0
+ uab = -0.025 luab = 0.0 wuab = 0.0
+ ub0 = 0.075 lub0 = 0.0 wub0 = 0.0
+ ubb = -0.0076 lubb = 0.0 wubb = 0.0
+ u10 = 0.13 lu10 = 0.0 wu10 = 0.0
+ u1b = 0.0001 lu1b = 0.0 wu1b = 0.0
+ n0 = 1.125 ln0 = 0.0 wn0 = 0.0
+ nb = 0.35 lnb = 0.0 wnb = 0.0
+ nd = -0.017 lnd = 0.0 wnd = 0.0
+ vof0 = 1.16 lvof0 = 0.0 wvof0 = 0.0
+ vofb = -0.034 lvofb = 0.0 wvofb = 0.0
+ vofd = -0.069 lvofd = 0.0 wvofd = 0.0
+ ai0 = 0 lai0 = 0.0 wai0 = 0.0
+ aib = 0 laib = 0.0 waib = 0.0
+ bi0 = 0 lbi0 = 0.0 wbi0 = 0.0
+ bib = 0 lbib = 0.0 wbib = 0.0
+ vghigh = 0.232 lvghigh = 0.0 wvghigh = 0.0
+ vglow = -0.114 lvglow = 0.0 wvglow = 0.0
+ tox = 8.6e-3 temp = 27 vdd = 3 vgg = 4 vbb = -3
+ cgdo = 2.0e-10 cgso = 2.0e-10 cgbo = 5.0e-11
+ xpart = 0
+ rsh = 0 cj = 0.0002 cjsw = 1.0e-10
+ js = 5e-5 pb = 0.7 pbsw = 0.8
+ mj = 0.5 mjsw = 0.33 wdf = 0
+ dell = 0
.END
```

Fig. 5.32 An example of SPICE3 input deck for an 11-stage CMOS ring oscillator with BSIM2 parameters.

SPICE Simulation Results

Measured delay time: 22ps/stage
Simulated delay time: 24ps/stage
(NMOS E-D ring. osc., Tox=8.6nm, Leff=0.25um)

Model	Level 1	Level 2	BSIM2	BSIM2	BSIM2
Tox (A)	250	250	86	86	86
Leff (um)	2.0	2.0	0.25	0.5	0.7
Circuit	DC Analysis, Id-Vd characteristics				
Time (s)	1.51	2.80	1.55	1.49	1.52
Iterations	113	111	71	71	71
Circuit	TRAN. Analysis, 15-stage ring osc.				
Time (s)	185.4	-----	173.7	142.7	152.6
Iterations	576	-----	435	342	369

Table 5.2 Comparison of SPICE simulation times between the BSIM2 model and level-1, level-2 models.

5.6 References

[5.1] G. Baccarani, M.R. Wordeman, and R.H. Dennard, "Generalized Scaling Theory and Its Application to a 1/4 Micrometer MOSFET Design," IEEE Tran. on Electron Devices, vol. ED-31, no. 4, pp.452-462, April 1984.

[5.2] P.K. Ko, "MOS Device Modeling for Circuit Simulation," IEDM Tech. Digs., p. 488, 1985.

[5.3] F. Hsu, P. K. Ko, S. Tam, C. Hu, and R. S. Muller, "An Analytical Breakdown Model for Short-Channel MOSFET's," IEEE Tran. on Electron Devices, vol. ED-29, p.1735, Nov., 1982.

[5.4] M.-C. Jeng, P.K. Ko, and C. Hu, "A Deep-Submicrometer MOSFET Model for Analog/Digital Circuit Simulations," IEDM Tech. Digs., pp.114-117, 1988.

[5.5] P.K. Chatterjee, W.R. Hunter, T.C. Holloway, and Y.T. Lin, "The impact of scaling laws on the choice of n-channel or p-channel for MOS VLSI," IEEE Electron Device Lett., vol. EDL-1, no. 10, p. 220, 1980.

[5.6] H. Shichijo, "A re-examination of practical performance limits of scaled n-channel and p-channel MOS devices for VLSI," Solid-State Electron, vol. 26, no. 10, p. 969, 1983.

[5.7] K.K. Ng and W.T. Lynch, "The impact of intrinsic series resistance on MOSFET scaling," IEEE Tran. on Electron Device, vol. ED-34, no. 3, p. 503, 1987.

[5.8] T.L. Quarles, "Analysis of Performance and Concergence Issues for Circuit Simulation," Univ. of California, Berkeley, Memorandum no. UCB/ERL M89/42, 1989.

[5.9] T.L. Quarles, "The SPICE3 Implementation Guide," Univ. of California, Berkeley, Memorandum no. UCB/ERL M89/44, 1989.

[5.10] T.L. Quarles, "SPICE3 Version 3C1 Users Guide," Univ. of California, Berkeley, Memorandum no. UCB/ERL M89/46, 1989.

[5.11] J. Duster, "Parameter Extraction for BSIM2," Master Thesis, University of California, Berkeley, 1990.

[5.12] S.J. Wang, J.Y. Lee, and C.Y. Chang, "An Efficient and Reliable Approach for Semiconductor Device Parameter Extraction," IEEE Tran. on Computer-Aided Design, vol. CAD-5, no. 1, pp. 170-179, Jan. 1986.

[5.13] C.F. Machala, P.C. Pattnaik, and P. Yang, "An Efficient Algorithm for the Extraction of Parameters with High Confidence from Nonlinear Models," Electron Device Letters, vol. EDL-7, no. 4, pp.214-218, 1986.

[5.14] B.J. Sheu, D.L. Sharfetter, P.K. Ko, and M.-C. Jeng, "BSIM: Berkeley Short-Channel IGFET MODEL for MOS Transistors," IEEE J. of Solid-State Circuits, vol. SC-22, no. 4, Aug. 1987.

[5.15] A.H.-C. Fung, "A Subthreshold Conduction Model for BSIM," Univ. of california, Berkeley, ERL memo. no. UCB/ERL M85/22, 1985.

[5.16] B.J. Sheu, D.L. Scharfetter, and H.C. Poon, "Compact Short-Channel IGFET Model," Univ. of California, Berkeley, ERL memo. no. UCB/ERL M84/20, 1984.

[5.17] C.S. Chao and L.A. Akers, "Drain-Voltage Effects on the Threshold Voltage of a Small-Geometry MOSFET," Solid-State Electronics, vol. 26, no. 9, pp.851-860, 1983.

[5.18] M.-C. Jeng, B.J. Sheu, and P.K. Ko, "BSIM Parameter Extraction - Algorithm and Users Guide," Univ. of California, Memo. no. UCB/ERL M85/79, 1985.

[5.19] R.W. Coen and R.S. Muller, "Velocity of Surface Carriers in Inversion Layers on Silicon," Solid-State Electronics, vol. 23, p.35, 1980.

[5.20] J.-P. Leburton and G.E. Dorda, "v-E Dependence in Small-Sized MOS Transistors," IEEE Tran. on Electron Devices, vol. ED-29, no. 8, Aug. 1982.

[5.21] C.G. Sidini, P.K. Ko, and J. Moll, "The Effect of High Fields on MOS Device and Circuit Performance," IEEE Tran. on Electron Devices, vol. ED-31, no. 10, pp. 1386-1393, Oct. 1984.

[5.22] S.C. Sun and J.D. Plummer, "Electron Mobility in Inversion and Accumulation Layers on Thermally Oxidized Silicon Surfaces, " IEEE Tran. on Electron Devices, vol. ED-27, p.1497, August, 1980.

[5.23] A.G. Sabnis and J.T. Clements, "Characterization of the Electron Mobility in the Inverted <100> Si Surface," IEDM Tech. Dig., pp.18-21, 1979

[5.24] M-S. Liang, J.Y. Choi, P.K. Ko, and C. Hu, "Inversion-Layer Capacitance and Mobility of Very Thin Gate-Oxide MOSFET's," IEEE Tran. on Electron Devices, vol. ED-33, p.409, March, 1986.

[5.25] H. Hartstein, A.B. Fowler, and M. Albert, "Temperature Dependence of Scattering in the Inversion Layer," Surface Sci., vol. 98, p.181, 1980.

[5.26] P.K. Ko, "Hot-Electron Effects in MOSFETs," Ph.D. Dissertation, University of California, Berkeley, 1982.

[5.27] S.M. Sze, "Physics of Semiconductor Devices," 2nd Edition , Wiley-Interscience, 1981.

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

[5.29] D. Frohman-Bentchkowsky and A. S. Grove, "Conductance of MOS Transistors in Saturation," IEEE Tran. on Electron Devices, vol. ED-16, p.108, Jan., 1969.

[5.30] T. L. Chiu and C. T. Sah, "Correlation Experiments with a Two-Section-Model Theory of the Saturation Drain Conductance of MOS Transistors," Solid-State Electronics, vol. 11, p.1149, Dec., 1968.

[5.31] Y. A. El-Mansy and A. R. Boothroyd, "A simple Two-Dimensional Model for IGFET Operation in the Saturation Region," IEEE Tran. on Electron Devices, vol. ED-24, p. 254, Mar., 1977.

[5.32] P. K. Ko, R. S. Muller, and C. Hu, "A Unified Model for Hot-Electron Currents in MOSFETs," 1981 IEDM Tech. Dig., p.600.

[5.33] T.H. Ning, P.W. Cook, R.H. Dennard, C.M. Osburn, S.E. Schuster, and H.N. Yu," 1 μm MOSFET VLSI Technology, Part IV. Hot-Electron Design Constrains," IEEE Tran. on Electron Devices, vol. ED-26, pp.346-353, 1979.

[5.34] C. Hu, "Hot-Electron Effects in MOSFET's," IEDM Tech. Dig., pp.176-181, 1983.

[5.35] Y.Leblebici, S.M. Kang, C.T. Sah, and N,Nishida, "Modeling and Simulation of Hot-Electron Effects for VLSI Reliability," ICCAD Tech. Digs., pp.252-255, 1987.

[5.36] P.M. Lee, M.M. Kuo, K. Seki, P.K. Ko, and C. Hu, "Circuit Aging Simulator (CAS)," IEDM Tech. Digs., pp.134-137, 1988.

[5.37] J.H. Chern and P. Yang, "Reliability modeling tools," Wafer Level Reliability Workshop, October 1988.

[5.38] W.-J. Hsu and B.J. Sheu, "Computer-Aided VLSI Reliability Studies", Wafer Level Reliability Workshop, October 1988.

[5.39] W. Maes, K.M. De Meyer, and L.H. Dupas, "SIMPAR: A Versatile Technology Independent Parameter Extraction Program Using a New Optimized Fit-Strategy," IEEE Tran. on Computer-Aided Design, vol. CAD-5, no. 2, p.214, April, 1986.

[5.40] C.J.B. Spanos and S.W. Director, "Parameter Extraction for Statistical IC Process Characterization," IEEE Tran. on Computer-Aided Design, vol. CAD-5, p.66, 1986.

[5.41] P. Yang, P.K. Chatterjee, "An Optimal Parameter Extraction Program for MOSFET Models," IEEE Tran. on Electron Devices, vol. ED-30, no. 9, p.1214, 1983.

[5.42] D.E. Eard and K. Doganis, "Optimizaed Extraction of MOS Model Parameters," IEEE Tran. on Computer-Aided Design, vol. CAD-1, no. 4, p.163, 1982.

[5.43] K. Levenberg, "A Method for the Solution of Certain Nonlinear Problems in Least Squares," Quart. Appll. Maths., vol. 2, pp.164-168, 1944.

[5.44] D.W. Marquardt, "An algorithm for Least-Square Estimation of Nonlinear Parameters," J. Sco. Indust. Appl. Maths., vol. 11, pp.431-441, 1963.

[5.45] D. Jacobs, "State of the Art in Numerical Analysis," Academic, 1977.

[5.46] G. Dahlquist and A. Bjorck, "Automatic Computation: Numerical Methods," Englwood Cliffs, N.J., Prentice Hall, 1974.

[5.47] R. Fletcher and M.M.D. Powell, "A Rapid Convergent Decent Method for Minimization," The Computer Journal, vol. 6, pp.163-168, 1963.

Chapter 6

CONCLUSION

High current drive, high packing density, and high integration levels have been the motivations in MOS scaling. Due to the resolution of optical lithography, the minimum MOSFET channel length was limited to the range of near micron. With the advent of X-ray lithography and E-beam direct-write technique, MOSFETs with channel length as small as 0.1μm have been demonstrated. However, these techniques are very expensive and are incompatible with existing technologies. In this work, a photoresist ashing technique has been developed which, when used in conjunction with conventional g-line optical lithography, permits the controlled definition of the gates of deep-submicrometer devices. Since most polymer-based resist material are ashable with oxygen plasma, this photoresist-ashing technique can also be extended to supplement other lithographic processes, such as those of e-beam and X-ray.

As the device dimensions are scaled below one micron, short-channel effects and other second-order effects become more prominent. To design and model deep-submicrometer devices, the physics of these devices has to be re-examined and understood. Studies based on the performance and reliability have shown that the basic physics associated with deep-submicrometer devices is similar to that of their long-channel counterparts. Therefore, existing design guidelines and models can still be applied with the need only for minor modifications.

In this work, various design curves for deep-submicrometer devices have been generated from experimental results based on the following considerations: short-channel and drain-induced-barrier-lowering effects, punchthrough and gate-induced drain leakage currents, hot-electron reliability, time-dependent-dielectric breakdown, current-driving capability, voltage gain, and switching speed. It is found that for an effective channel length of 0.3μm, the maximum allowable power supply is 3V, which implies that some kind of hot-electron-resistant structure is still desirable for 0.5μm technology with 3.3V power supply.

With the dramatic increase in the number of transistors per chip, the circuit complexity and the fabrication cost also increase proportionally. In order to speed up the VLSI/ULSI system design and to reduce costs, it has become necessary to start the circuit design in the early stages of technology development and to predict circuit behavior before the circuit is actually fabricated, both of which require intensive use of circuit simulators. Since the device characteristics of small-geometry devices are highly sensitive to parameter variations, optimal circuit designs become even more difficult to create than before. Therefore, an accurate and computationally efficient drain current model for deep-submicrometer MOSFETs becomes extremely crucial and indispensable in developing future system designs.

In this work, a MOSFET drain current model suitable to predict small geometry effects for size as small as quarter-micron channel length, for digital as well as analog applications is developed. In developing this model, meticulous care has been taken in retaining the basic functional form of fully physical models while improving model accuracy and computational efficiency. The ease of parameter extraction was also a major consideration. In addition to the effects commonly included in the MOSFET drain current equation, it is found that the inversion-layer capacitance effect, hot-electron induced output resistance degradation, and source/drain parasitic resistance effect are also important factors to consider in deep-submicrometer MOSFET modeling. This model has been implemented in SPICE3. A parameter extraction system dedicated to this model was also developed.

APPENDIX A

PROCESS FLOW

SUB-MICROMETER NMOS PROCESS
(modified from MICROLAB CMOS PROCESS version 3.0)
single poly-Si, single metal

Step 0: Starting Wafers: 15-30Ω-cm, p-type, <100>,
scribed SUBC1 to SUBC10.
Control Wafers: PSUB (p-type)
Measure bulk resistivity (Ω-cm) of prime wafers on Sonogage.
[NOTE]: Only carry out HF dips where indicated.

Step 1: Initial Oxidation: target oxide = 200angstrom

1.1 TCA clean furnace tube.

1.2 Standard clean wafers, include PSUB control:
piranha clean for 10 min, spin-dry.

1.3 Dry oxidation at 950 degreeC:
30 min dry O_2
20 min dry N_2
Measured oxide = on PSUB control.
(Rework if oxide <50angstrom).

Step 2: Field Implant (Blanket Implant): boron (B11), 1.5E12/cm^2, 70 KeV
Include PSUB control (no photoresist).

Step 3: Locos Pad Oxidation/Nitride Deposition:
target thickness = 250angstrom SiO_2 + 1000angstrom Si_3N_4

3.1 TCA clean furnace tube.

3.2 Remove all oxide in 10:1 HF until wafers dewet (including PSUB).
Measure sheet resistance R_s of PSUB control on Prometrix.

3.3 Standard clean wafers.

3.4 Dry oxidation at 950 degreeC:
45 min dry O_2
20 min dry N_2 anneal.
a) Measured oxide = on PSUB control.

(Rework if oxide <180angstrom).
b) Strip oxide off PSUB control in BHF.

3.5 Deposit 1000angstrom of Si-nitride immediately:
Deposition time = 22 min, temperature = 800 degreeC.
Include PSUB control.
Measured $T_{nitride}$ =
(Make additional deposition if $T_{nitride}$ <800angstrom)
Save PSUB control for Step 7.

Step 4: Active Area Photo Mask: ND (emulsion-cf)
Dehydrate, HMDS, Spin, expose, develop, descum, hard bake.

Step 5: Nitride Etch: Technics-C plasma etcher
[NOTE]: 1. Rotate wafers to insure uniform nitride etch.
2. Wet-etch oxide. (Check for dewet in field regions).

5.1 Remove photoresist and piranha clean wafers.

Step 6: LOCOS Oxidation: target oxide = 3000angstrom)

6.1 TCA clean furnace tube.

6.2 Standard clean wafers.
No HF dip.

6.3 Wet oxidation at 950 degreeC:
5 min dry O_2
1 hour 25 min wet O_2
5 min dry O_2
20 min N_2 anneal
Measured oxide = on a device wafer in the field area.
(Stop and consult if oxide <2500angstrom and check wafer uniformity; stop and consult if oxide variation >500angstrom)

Step 7: Nitride Removal (include PSUB control)

7.1 Oxide dip in 25:1 HF for 1 min.

7.2 Etch nitride off in hot phosphoric acid: 145 degreeC, 60 min.
End point detection by dewet of PSUB control.

7.3 Dip off sacrificial oxide (dewet) in 25:1 HF for 1 min.

Step 8: Sacrificial Oxide Growth: target oxide = 200angstrom

8.1 TCA clean furnace tube.

8.2 Standard clean wafers. No HF dip, including PSUB control.

8.3 Dry oxidation at 950 degreeC:
30 min dry O_2
20 min N_2 anneal
a) Measured oxide = on PSUB control
b) Do not include PSUB control in Steps 9 to 12.

Step 9: Enhancement Implant Mask: NG (chrome-df)
Dehydrate, HMDS, spin, align, expose, develop, descum, hardbake.
[NOTE]: The exposure should be increased by 25% to compensate for the dark field.

Step 10: Enhancement Implant: Boron (B11) at 30 KeV,
9 splits in implant dose.

Wafer SUBC-	1	2	3	(T_{ox} = 25angstrom)
Dose (*1E12/cm^2)	6.0	10	16	

Wafer SUBC-	4	5	6,10	(T_{ox} = 50angstrom)
Dose (*1E12/cm^2)	3.0	6.0	9.0	

Wafer SUBC-	7	8	9	(T_{ox} = 75angstrom)
Dose (*1E12/cm^2)	2.0	4.0	8.0	

Remove photoresist and piranha clean wafers after implant.

Step 11: Depletion Implant Mask: NI (chrome-df)
Dehydrate, HMDS, spin, align, expose, develop, descum, hardbake.
[NOTE]: The exposure should be increased by 25% to compensate for the dark field.

Step 12: Depletion Implant: Arsenic at 50 KeV,
9 splits in implant dose.

Wafer SUBC-	1	2	3	(T_{ox} = 25angstrom)
Dose (*1E12/cm^2)	35	30	25	

Wafer SUBC-	4	5	6,10	(T_{ox} = 50angstrom)
Dose (*1E12/cm^2)	20	14	12	

Wafer SUBC-	7	8	9	
Dose (*1E12/cm^2)	13	10	7.0	(T_{ox} = 75angstrom)

Remove resist and piranha clean wafers after implant.

Step 13: Gate Oxidation/Poly-Si Deposition:
target = 25angstrom SiO_2 + 2500angstrom poly-Si for wafers SUBC1-3
target = 50angstrom SiO_2 + 2500angstrom poly-Si for SUBC4- and SUBC10
target = 75angstrom SiO_2 + 2500angstrom poly-Si for wafers SUBC7-9

13.1 TCA clean furnace tube; reserve poly-Si deposition tube.

13.2 Standard clean wafers,
include new monitor wafers TOX1,TOX2 and TOX3.

13.3 Dip off sacrificial oxide (dewet) in 25:1 HF (approx. 1 min).

13.4a Dry oxidation at 800 degreeC: Wafers: SUBC1-3,TOX1
target T_{ox} = 25angstrom
5 min dry O_2
10 min N_2 anneal.
Measure T_{ox} = on TOX1 control.

13.5a Immediately after oxidation deposit 2500angstrom of phosphorous-doped poly-Si.
time = 1 hour 15 min, temperature= 650 degreeC
[NOTE]: Do not include TOX1 control; include a new control with 1000angstrom thermal SiO_2 on it.
Measure T_{poly} =
stop and consult if T_{poly} <2000angstrom or T_{poly} >3000angstrom)

13.4b Dry oxidation at 800 degreeC: SUBC4-6,SUBC10,TOX2
target T_{ox} = 50angstrom
30 min dry O_2
10 min N_2 anneal.
Measure T_{ox} = on TOX2 control.

13.5b Immediately after oxidation deposit 2500angstrom of phosphorous-doped poly-Si.
time = 1 hour 15 min, temperature = 650 degreeC
[NOTE]: Do not include TOX2 control; include a new control with 1000angstrom thermal SiO_2 on it.
Measure T_{poly} =
stop and consult if T_{poly} <2000angstrom or T_{poly} >3000angstrom)

13.4c Dry oxidation at 900 degreeC: SUBC7-9,TOX3
target T_{ox} = 75angstrom
7 min dry O_2
20 min N_2 anneal.
Measure T_{ox} = on TOX3 control.

13.5c Immediately after oxidation deposit 2500angstrom of phosphorous-doped poly-Si.
time = 1 hour 15 min, temperature = 650 degreeC
[NOTE]: Do not include TOX3 control; include a new control with 1000angstrom thermal SiO_2 on it.
Measure T_{poly} =
stop and consult if T_{poly} <2000angstrom or T_{poly} >3000angstrom)

Step 14: Gate Definition Mask: NP (emulsion-cf)
Dehydrate, HMDS, spin, align, expose, develop.
1. Right before exposure, do a focus-exposure test for GCA wafer stepper to determine the best focus and exposure.
2. Expose and develop.
3. Photoresist ashing.
4. 25:1 BHF dip.
5. Etch poly-Si in LAM.

Step 15: Reoxidation: target oxide = 200angstrom on poly-Si.

15.1 TCA clean furnace tube.

15.2 Standard clean wafers, include controls, TOX2 and TOX3.
[NOTE]: No dip in HF after piranha.

15.3 Dry oxidation at 950 degreeC: all wafers & TOX2 and TOX3.
15 min dry O_2
10 min N_2 anneal.
Measure oxide thickness:
T_{ox}(TOX2) = T_{ox}(TOX3) =

Step 16: N^+ Source/Drain Implant

16.1 Implant Arsenic at 0degree inclination, 50 KeV, 3E15/cm^2, including TOX2, and TOX3 controls.

Step 17: N^+ S/D Reoxidation and Anneal: target oxide = 400angstrom on poly-Si

---

17.1 TCA clean furnace tube.

17.2 Standard clean wafers, include TOX2 and TOX3 control

[NOTE]: No dip in HF after piranha.

17.3 Dry oxidation at 925 degreeC: all wafers & TOX2 and TOX3.
15 min dry O_2
10 min N_2 anneal.
Measure oxide thickness:
T_{ox}(TOX2) = T_{ox}(TOX3) =

17.4 Strip TOX2 and TOX3 controls and measure sheet resistance (Ω/square) on Prometrix.
Save all controls in "completed controls" box.

Step 18: Reflow Glass: target oxide = 3000angstrom

18.1 Standard clean wafers (NOTE: No HF dip).
Include only one new, PSG control.

18.2 Deposit undoped LTO: including PSG control.
Layers: 3000angstrom undoped LTO.
time = (approx) 15 min total (check current deposition rates)
temperature = 450 degreeC.
Measure T_{PSG} = on PSG control.

18.3 Densify glass at 925 degreeC:
include one PSG control.
20 min dry O_2.

Step 19: Contact Photo Mask: NC (chrome-df)
Dehydrate, HMDS, spin, align, expose, HAND develop, descum, hardbake.
The exposure should be increased by 25% to compensate for the dark field.
Hand develop 90 seconds.
One part Microposit developer; one part water.

Step 20: Contact Etch:

20.1 Dry plasma etch in LAM-2
Inspect thoroughly.

Step 21: Buffer Doped-Poly Deposition:

21.1 Remove photoresist and piranha clean.

21.2 Standard clean wafers: piranha clean for 5 min, followed by dip in 25:1 for 15 sec.
[NOTE]: use fresh 25/1 HF solution.

21.3 Immediately after spin dry, deposit 2000angstrom of

phosphorous-doped poly-Si.
time = 1 hour, temperature = 650 degreeC
Include a new control with 1000angstrom thermal SiO_2 on it.
Measure T_{poly} =

21.4 N^+ poly activation:
Anneal wafers in N_2 at 900 degreeC for 15 min.

Step 22: Back Side Etch

22.1 Spin photoresist (front side), do not expose; hard bake.

22.2 Spin photoresist again, and hard bake.

22.3 Etch back side of wafers as follows:
a) Wet etch poly-Si (buffer poly-Si thickness).
b) Etch off PSG in BHF.
c) Wet etch poly-Si (gate poly-Si thickness).
d) Final dip in BHF until back dewets.

22.4 Remove photoresist in O_2 plasma: 5-7 min at 300 watts, followed by piranha clean wafers.

22.5 Do a 20 sec 25:1 HF dip just before metallization.

Step 23: First Metallization: target thickness = 6000angstrom
Sputter Al with 2% Si on all wafers.

Step 24: First Metal Photo Mask: NM (emulsion-cf)

24.1 Spin Hunt WX-235 resist, expose, develop, descum.
[NOTE]: 1. No HMDS step.
2. No hard bake.

24.2 Wet etch Al. (Wet wafers first in DI water.)

24.3 Wet etch buffer poly-Si, visual end point detection.

24.4 Remove resist with acetone (no piranha!)

24.5 Rinse wafers in DI water for 20 minutes, spin dry.

24.6 Probe test devices.

Step 25: Back Side Metalization

25.1 Spin photoresist (front side), do not expose; hard bake.

25.2 Dip in BHF until back dewets.

25.3 Hardbake at 120 degreeC, 10 min.

25.4 Right after hardbake, sputter 6000angstrom Al with 2% Si on back side of all wafers.

Step 26: Sintering: 400 degreeC for 20 min in forming gas.

Step 27: End of Process

APPENDIX B

BSIM1 MODEL

List of Parameters

1. V_{FB} flat-band voltage
2. ϕ_S surface potential
3. K_1 body-effect coefficient
4. K_2 non-uniform channel doping coefficient
5. η_0 value of η extracted at $V_{BS} = 0$ and $V_{DS} = V_{DD}$
6. η_B sensitivity of η to V_{BS}
7. η_D sensitivity of η to V_{DS}
8. β_Z value of β_0 extracted at $V_{BS} = 0$ and $V_{DS} = 0$
9. β_{ZB} sensitivity of β_0 to V_{BS} at $V_{DS} = 0$
10. β_S value of β_0 extracted at $V_{BS} = 0$ and $V_{DS} = V_{DD}$
11. β_{SB} sensitivity of β_0 to V_{BS} at $V_{DS} = V_{DD}$
12. β_{SD} sensitivity of β_0 to V_{DS} at $V_{DS} = V_{DD}$
13. U_{a0} value of U_a extracted at $V_{BS} = 0$
14. U_{aB} sensitivity of U_a to V_{BS}
15. U_{1Z} value of U_1 extracted at $V_{BS} = 0$ and $V_{DS} = V_{DD}$
16. U_{1B} sensitivity of U_1 to V_{BS}
17. U_{1D} sensitivity of U_1 to V_{DS}
18. n_0 value of n extracted at $V_{BS} = 0$ and $V_{DS} = 0$
19. n_B sensitivity of n to V_{BS}
20. n_D sensitivity of n to V_{DS}

(A) Threshold voltage:

$$V_{th} = V_{FB} + \phi_S + K_1\sqrt{\phi_S - V_{BS}} - K_2(\phi_S - V_{BS}) - \eta V_{DS} \tag{B.1}$$

(B) Linear region:

$$I_{DS0} = \frac{\beta_0(V_{GS} - V_{th} - \frac{a}{2}V_{DS})V_{DS}}{[1 + U_a(V_{GS} - V_{th})](1 + U_1 V_{DS})} \tag{B.2}$$

where $U_1 = 1/E_c L_{eff}$ and

$$\beta_0 = \frac{\mu_0 C_{ox} W_{eff}}{L_{eff}}$$

$$a = 1 + \frac{gK_1}{2\sqrt{\phi_S - V_{BS}}} \tag{B.3}$$

$$g = 1 \frac{1}{1.744 + 0.8364(\phi_S - V_{BS})} \tag{B.4}$$

(C) Saturation region:

$$I_{DS0} = \frac{\beta_0(V_{GS} - V_{th})^2}{[1 + U_a(V_{GS} - V_{th})]2aK} \tag{B.5}$$

$$V_{DSAT} = \frac{V_{GS} - V_{th}}{a\sqrt{K}} \tag{B.6}$$

$$K = \frac{1 + V_c + \sqrt{1 + 2V_c}}{2} \tag{B.7}$$

$$V_c = \frac{U_1(V_{GS} - V_{th})}{a} \tag{B.8}$$

(D) Subthreshold region:

$$I_{subth} = \frac{I_{EXP} I_{limit}}{I_{EXP} + I_{limit}} \tag{B.9}$$

$$I_{EXP} = \beta_0 V_{tm}^2 e^{1.8} e^{\frac{V_{GS} - V_{th}}{nV_{tm}}} \left[1 - e^{\frac{-V_{DS}}{V_{tm}}}\right] \tag{B.10}$$

$$I_{limit}=\beta_0\frac{(3V_{tm})^2}{2} \tag{B.11}$$

(E) Total drain current:

$$I_{DS}=I_{DS0}+I_{subth} \tag{B.12}$$

(F) Bias-dependent parameters:

$$\eta=\eta_0+\eta_B V_{BS}+\eta_D(V_{DS}-V_{DD}) \tag{B.12}$$

$$U_a=U_{a0}+U_{aB}V_{BS} \tag{B.13}$$

$$U_1=U_{1Z}+U_{1B}V_{BS}+U_{1D}(V_{DS}-V_{DD}) \tag{B.14}$$

$$\beta_0=\beta_1(\frac{V_{DS}}{V_{DD}}-1)^2+\beta_2(2-\frac{V_{DS}}{V_{DD}})\frac{V_{DS}}{V_{DD}}+\beta_{SD}V_{DS}(\frac{V_{DS}}{V_{DD}}-1) \tag{B.15}$$

$$\beta_1=\beta_Z+\beta_{ZB}V_{BS} \tag{B.16}$$

$$\beta_2=\beta_S+\beta_{SB}V_{BS} \tag{B.17}$$

$$n=n_0+n_B V_{BS}+n_D V_{DS} \tag{B.18}$$

APPENDIX C

SUBTHRESHOLD CONDUCTION

The drain current in the subthreshold region is dominated by the diffusion current given by

$$I_{D,diff} = -D_n \frac{dQ}{dy} \approx D_n \frac{Q(S) - Q(D)}{L_{eff}} \tag{C.1}$$

where Q(S) and Q(D) are the inversion-charge density at the source and the drain. D_n is the diffusion constant given by $D_n = V_{tm}\mu_0$.

$$Q(S) = qN_{SUB}\, e^{\frac{\phi(x) - 2\phi_B}{V_{tm}}} \tag{C.2}$$

$$Q(D) = qN_{SUB}\, e^{\frac{\phi(x) - 2\phi_B - V_{DS}}{V_{tm}}} \tag{C.3}$$

$$\phi(x) = \phi_S\left(1 - \frac{x}{x_d}\right) \tag{C.4}$$

$$\phi_B = V_{tm} \ln\left(\frac{N_{SUB}}{n_i}\right) \tag{C.5}$$

$$x_d = \sqrt{\frac{2\varepsilon_{si}\phi_S}{qN_{SUB}}} \tag{C.6}$$

Integrating (C.1) over the depletion depth, the subthreshold current can be derived.

$$I_{subth} = \int_0^{x_d} I_{D,diff}\, dx$$

$$= \beta_0 (V_{tm})^2 \frac{C_d}{C_{ox}} e^{\frac{\phi_S - 2\phi_B}{V_{tm}}} \left[1 - e^{\frac{-V_{DS}}{V_{tm}}}\right] \tag{C.7}$$

where

$$C_d \approx \sqrt{\frac{q\varepsilon_{si}N_{SUB}}{2\phi_S}} \tag{C.8}$$

In deriving (C.7), $\phi(x)$ has been approximated by $\phi_S(1-\frac{2x}{x_d})$ and integrated from 0 to $x_d/2$.

The relationship between the gate voltage V_{GS} and the surface potential ϕ_S is given by

$$V_{GS}-V_{FB}=\frac{Q_n}{C_{ox}}+\phi_S \tag{C.9}$$

where the inversion-charge Q_n in the subthreshold region can be approximated by [5.27]

$$Q_n=\sqrt{2q\varepsilon_{si}N_{SUB}\phi_S} \tag{C.10}$$

A plot of ϕ_S versus V_{GS} is shown in Fig. 5.13. In the subthreshold region ($\phi_S < 2\phi_B$), ϕ_S can be approximated by a linear function of V_{GS} as shown by the dashed line in Fig. 5.13.

$$\phi_S \approx \frac{(V_{GS}-V_{FB})}{n}-V_o \tag{C.11}$$

where

$$n=\frac{dV_{GS}}{d\phi_S}=1+\frac{1}{C_{ox}}\sqrt{\frac{q\varepsilon_{si}N_{SUB}}{2\phi_S}}$$

$$\approx 1+\frac{C_d}{C_{ox}} \qquad \text{when } \phi_S \approx 2\phi_B \tag{C.12}$$

Substituting (C.12) into (C.11) and comparing with (C.9) and (C.10), V_o can solved.

$$V_o \approx \frac{\phi_S C_d}{C_d+C_{ox}} \approx \frac{2\phi_B C_d}{C_d+C_{ox}} \tag{C.13}$$

APPENDIX D

BSIM2 MODEL

List of Parameters

Given parameters:

T_{ox}	oxide thickness
T	temperature
V_{DD}	drain supply voltage
V_{GG}	gate supply voltage
V_{BB}	body supply voltage
V_{G1}	lower bound of the transition region
V_{G2}	upper bound of the transition region

Extracted parameters:

V_{FB}	flat-band voltage
ϕ_S	surface potential
K_1	body-effect coefficient
K_2	non-uniform channel doping coefficient
η	drain-induced-barrier-lowering coefficient
η_0	value of η extracted at $V_{BS} = 0$
η_B	sensitivity of η to V_{BS}
β_0	conductance coefficient
β_{00}	value of β_0 extracted at $V_{BS} = 0$ and $V_{DS} = 0$
β_{0B}	sensitivity of β_0 to V_{BS} at $V_{DS} = 0$
β_{S0}	value of β_0 extracted at $V_{BS} = 0$ and $V_{DS} = V_{DD}$
β_{SB}	sensitivity of β_S to V_{BS} at $V_{DS} = V_{DD}$
β_3	linear empirical parameter in β_0 expression
β_{30}	value of β_3 extracted at $V_{BS} = 0$ and $V_{GS} = 0$
β_{3B}	sensitivity of β_3 to V_{BS}
β_{3G}	sensitivity of β_3 to V_{GS}
β_4	quadratic empirical parameter in β_0 expression
β_{40}	value of β_4 extracted at $V_{BS} = 0$ and $V_{GS} = 0$
β_{4B}	sensitivity of β_4 to V_{BS}
β_{4G}	sensitivity of β_4 to V_{GS}
β_2	empirical parameter in β_0 expression
β_{20}	value of β_2 extraxted at $V_{BS} = 0$ and $V_{GS} = 0$
β_{2B}	sensitivity of β_2 to V_{BS}
β_{2G}	sensitivity of β_2 to V_{GS}
U_a	first-order parameter of vertival field effect
U_{a0}	value of U_a extracted at $V_{BS} = 0$

U_{aB}	sensitivity of U_a to V_{BS}
U_b	second-order parameter of vertical field effect
U_{b0}	value of U_b extracted at $V_{BS} = 0$
U_{bB}	sensitivity of U_b to V_{BS}
U_1	velocity saturation coefficient
U_{1S0}	value of U_1 extracted at $V_{BS} = 0$ and $V_{DS} = V_{DD}$
U_{1SB}	sensitivity of U_1 to V_{BS}
U_{1D}	sensitivity of U_1 to V_{DS}
n	subthreshold swing coefficient
n_0	value of n when $V_{BS} = \infty$ and $V_{DS} = 0$
n_B	sensitivity of n to V_{BS}
n_D	sensitivity of n to V_{DS}
V_{offset}	V_{th} offset in the subthreshold region
$V_{offset0}$	value of V_{offset} extracted at $V_{BS} = 0$ and $V_{DS} = 0$
$V_{offsetB}$	sensitivity of V_{offset} to V_{BS}
$V_{offsetD}$	sensitivity of V_{offset} to V_{DS}
A_i	pre-exponential parameter of R_{out} degradation due to high field
A_{i0}	value of A_i extracted at $V_{BS} = 0$
A_{iB}	sensitivity of A_i to V_{BS}
B_i	exponential parameter of R_{out} degradation due to high field
B_{i0}	value of B_i extracted at $V_{BS} = 0$
B_{iB}	sensitivity of B_i to V_{BS}

Model equations:

(A) Threshold Voltage:

$$V_{th}=V_{FB}+\phi_S+K_1\sqrt{\phi_S-V_{BS}}-K_2(\phi_S-V_{BS})-\eta V_{DS} \quad (D.1)$$

$$a=1+\frac{g K_1}{2\sqrt{\phi_S-V_{BS}}} \quad (D.2)$$

$$g=1\frac{1}{1.744+0.8364(\phi_S-V_{BS})} \quad (D.3)$$

(B) Drain saturation voltage:

$$V_{DSAT}=\frac{V_{GS}-V_{th}}{a\sqrt{K}} \quad (D.4)$$

$$K=\frac{1+V_c+\sqrt{1+2V_c}}{2} \quad (D.5)$$

$$V_c=\frac{U_{1S}(V_{GS}-V_{th})}{aU_{vert}} \quad (D.6)$$

(C) Linear region:

(C.1) drain current:

$$I_{DS}=\frac{\beta_0(V_{GS}-V_{th}-\frac{a}{2}V_{DS})V_{DS}}{U} \quad (D.7)$$

$$\beta_0=\frac{\mu_0 C_{ox} W_{eff}}{L_{eff}} \quad (D.8)$$

$$U=U_{vert}+U_1 V_{DS} \quad (D.9)$$

$$U_{vert}=1+U_a(V_{GS}-V_{th})+U_b(V_{GS}-V_{th})^2 \quad (D.10)$$

(C.2) bias-dependent parameters:

$$\beta_0=\beta_{0in}+\beta_1\tanh(\frac{\beta_2 V_{DS}}{V_{DSAT}})+\beta_3 V_{DS}-\beta_4 V_{DS}^2 \quad (D.11)$$

$$\beta_{0lin} = \beta_{00} + \beta_{0B} V_{BS} \quad \text{(D.12)}$$

$$\beta_1 = \beta_S - (\beta_{0lin} + \beta_3 V_{DD} + \beta_4 V_{DD}^2) \quad \text{(D.13)}$$

$$\beta_S = \beta_{S0} + \beta_{SB} V_{BS} \quad \text{(D.14)}$$

$$\beta_2 = \beta_{20} + \beta_{2B} V_{BS} + \beta_{2G} V_{GS} \quad \text{(D.15)}$$

$$\beta_3 = \beta_{30} + \beta_{3B} V_{BS} + \beta_{3G} V_{GS} \quad \text{(D.16)}$$

$$\beta_4 = \beta_{40} + \beta_{4B} V_{BS} + \beta_{4G} V_{GS} \quad \text{(D.17)}$$

$$\eta = \eta_0 + \eta_B V_{BS} \quad \text{(D.18)}$$

$$U_a = U_{a0} + U_{aB} V_{BS} \quad \text{(D.19)}$$

$$U_b = U_{b0} + U_{bB} V_{BS} \quad \text{(D.20)}$$

$$U_1 = U_{1S} \left[1 - \frac{U_{1D} (V_{DS} - V_{DSAT})^2}{V_{DSAT}^2} \right] \qquad \text{if } V_{DS} < V_{DSAT}$$

$$= U_{1S} \qquad \text{if } V_{DS} > V_{DSAT} \quad \text{(D.21)}$$

where

$$U_1 \equiv 1/E_c L_{eff}$$

$$U_{1S} = U_{1S0} + U_{1SB} V_{BS} \quad \text{(D.22)}$$

(C.3) transconductance (g_m):

$$g_m = \frac{dI_{DS}}{dV_{GS}} = \frac{I_{DS}}{\beta} \frac{d\beta}{dV_{GS}} + \beta V_{DS} \quad \text{(D.23)}$$

$$\frac{d\beta}{dV_{GS}} = \frac{1}{U} \left(\frac{d\beta_0}{dV_{GS}} - \beta \frac{dU}{dV_{GS}} \right) \quad \text{(D.24)}$$

$$\frac{dU}{dV_{GS}}=\frac{dU_{vert}}{dV_{GS}}=\frac{dU_{vert}}{dV_{gt}}=U_a+2U_b(V_{GS}-V_{th}) \tag{D.25}$$

$$\frac{d\beta_0}{dV_{GS}}=\beta_1\,\mathrm{sech}^2\left(\frac{\beta_2 V_{DS}}{V_{DSAT}}\right)\left(\frac{V_{DS}}{V_{DSAT}}\beta_{2G}-\frac{\beta_2 V_{DS}}{V_{DSAT}^2}\frac{dV_{DSAT}}{dV_{GS}}\right)$$
$$+\tanh\left(\frac{\beta_2 V_{DS}}{V_{DSAT}}\right)+\beta_{3G}V_{DS}-\beta_{4G}V_{DS}^2 \tag{D.26}$$

$$\frac{dV_{DSAT}}{dV_{GS}}=\frac{1}{a\sqrt{K}}-\frac{1}{2}\frac{V_{DSAT}}{K}\frac{dK}{dV_{GS}} \tag{D.27}$$

$$\frac{dK}{dV_{GS}}=\frac{dK}{dV_c}\frac{dV_c}{dV_{GS}} \tag{D.28}$$

$$\frac{dK}{dV_c}=\frac{1}{2}\left(1+\frac{1}{\sqrt{1+2V_c}}\right) \tag{D.29}$$

$$\frac{dV_c}{dV_{GS}}=\frac{U_{1S}}{aU_{vert}}-\frac{U_{1S}(V_{GS}-V_{th})}{aU_{vert}^2}\frac{dU_{vert}}{dV_{GS}}$$

$$=\frac{V_c}{(V_{GS}-V_{th})}-\frac{V_c}{U_{vert}}\frac{dU_{vert}}{dV_{GS}} \tag{D.30}$$

(C.4) output conductance (g_{ds}):

$$g_{ds}=\frac{1}{R_{out}}=\frac{dI_{DS}}{dV_{DS}}=\frac{I_{DS}}{\beta}\frac{d\beta}{dV_{DS}}+\frac{I_{DS}}{V_{DS}}+\beta V_{DS}\left(\eta-\frac{a}{2}\right) \tag{D.31}$$

$$\frac{d\beta}{dV_{DS}}=\frac{1}{U}\left(\frac{d\beta_0}{dV_{DS}}-\beta\frac{dU}{dV_{DS}}\right) \tag{D.32}$$

$$\frac{dU}{dV_{DS}}=\frac{dU_{vert}}{dV_{DS}}+U_1+V_{DS}\frac{dU_1}{dV_{DS}} \tag{D.33}$$

$$\frac{dU_{vert}}{dV_{DS}}=[U_a+2U_b(V_{GS}-V_{th})]\eta \tag{D.34}$$

$$\frac{dU_1}{dV_{DS}}=U_{1S}\left[\frac{-2U_{1d}(V_{DS}-V_{DSAT})}{V_{DSAT}^2}\left(1-\frac{dV_{DSAT}}{dV_{DS}}\right)+\frac{(V_{DS}-V_{DSAT})^2}{V_{DSAT}^3}\frac{dV_{DSAT}}{dV_{DS}}\right]$$

$$=2U_{1S}U_{1d}\frac{(V_{DS}-V_{DSAT})}{V_{DSAT}^2}\left(\frac{V_{DS}}{V_{DSAT}}\frac{dV_{DSAT}}{dV_{DS}}-1\right) \tag{D.35}$$

$$\frac{dV_{DSAT}}{dV_{DS}}=\frac{V_{DSAT}}{(V_{GS}-V_{th})}\eta-\frac{V_{DSAT}}{2K}\frac{dK}{dV_{DS}} \tag{D.36}$$

$$\frac{dV_{th}}{dV_{DS}}=-\eta \tag{D.37}$$

$$\frac{dK}{dV_{DS}}=\frac{dK}{dV_c}\frac{dV_c}{dV_{DS}} \tag{D.38}$$

$$\frac{dV_c}{dV_{DS}}=\frac{U_{1S}}{aU_{vert}}\eta-\frac{V_c}{U_{vert}}[U_a+2U_b(V_{GS}-V_{th})]\eta$$

$$=V_c\left(\frac{1}{(V_{GS}-V_{th})}-\frac{1}{U_{vert}}\frac{dU_{vert}}{dV_{GS}}\right)\eta \tag{D.39}$$

$$\frac{d\beta_0}{dV_{DS}}=\beta_3-2\beta_4V_{DS}+\beta_1\,\mathrm{sech}^2\left(\frac{\beta_2V_{DS}}{V_{DSAT}}\right)\left(\frac{\beta_2}{V_{DSAT}}-\frac{\beta_2V_{DS}}{V_{DSAT}^2}\frac{dV_{DSAT}}{dV_{DS}}\right) \tag{D.40}$$

(C.5) body- (back-gate)-Transconductance (g_{mbs}):

$$g_{mbs}=\frac{dI_{DS}}{dV_{BS}}=\frac{I_{DS}}{\beta}\frac{d\beta}{dV_{BS}}-\beta V_{DS}\left(\frac{dV_{th}}{dV_{BS}}+\frac{V_{DS}}{2}\frac{da}{dV_{BS}}\right) \tag{D.41}$$

$$\frac{da}{dV_{BS}}=\frac{K_1}{2\sqrt{\phi_S-V_{BS}}}\frac{dg}{dV_{BS}}+\frac{gK_1}{4(\phi_S-V_{BS})^{3/2}}$$

$$=(a-1)\left[\frac{1}{g}\frac{dg}{dV_{BS}}+\frac{1}{2(\phi_S-V_{BS})}\right] \tag{D.42}$$

$$\frac{dg}{dV_{BS}}=\frac{-0.8364}{(1.744+0.8364(\phi_S-V_{BS}))^2}=-0.8364(1-g)^2 \tag{D.43}$$

$$\frac{d\beta}{dV_{BS}}=\frac{1}{U}\left(\frac{d\beta_0}{dV_{BS}}-\beta\frac{dU}{dV_{BS}}\right) \tag{D.44}$$

$$\frac{dU}{dV_{BS}}=\frac{dU_{vert}}{dV_{BS}}+V_{DS}\frac{dU_1}{dV_{BS}} \tag{D.45}$$

$$\frac{dU_{vert}}{dV_{BS}} = \frac{dU_{vert}}{dV_{GS}} \frac{dV_{th}}{dV_{BS}} + (V_{GS} - V_{th}) U_{aB} + (V_{GS} - V_{th})^2 U_{bB} \quad \text{(D.46)}$$

$$\frac{dV_{th}}{dV_{BS}} = \frac{K_1}{2\sqrt{\phi_S - V_{BS}}} + K_2 - \eta_B V_{DS} \quad \text{(D.47)}$$

$$\frac{d\beta_0}{dV_{BS}} = \beta_{0linB} + \beta_{3B} V_{DS} - \beta_{4B} V_{DS}^2 + \tanh\left(\frac{\beta_2 V_{DS}}{V_{DSAT}}\right) \frac{d\beta_1}{dV_{BS}}$$
$$+ \beta_1 \operatorname{sech}^2\left(\frac{\beta_2 V_{DS}}{V_{DSAT}}\right)\left(\frac{V_{DS}}{V_{DSAT}} \beta_{2B} - \frac{\beta_2 V_{DS}}{V_{DSAT}^2} \frac{dV_{DSAT}}{dV_{BS}}\right) \quad \text{(D.48)}$$

$$\frac{d\beta_1}{dV_{BS}} = \beta_{satB} - \beta_{3B} V_{DD} + \beta_{4B} V_{DS}^2 - \beta_{0linB} \quad \text{(D.49)}$$

$$\frac{dV_{DSAT}}{dV_{BS}} = \frac{V_{DSAT}}{a} \frac{da}{dV_{BS}} - \frac{V_{DSAT}}{(V_{GS} - V_{th})} \frac{dV_{th}}{dV_{BS}} - \frac{V_{DSAT}}{2K} \frac{dK}{dV_{BS}} \quad \text{(D.50)}$$

$$\frac{dK}{dV_{BS}} = \frac{dK}{dV_c} \frac{dV_c}{dV_{BS}} \quad \text{(D.51)}$$

$$\frac{dV_c}{dV_{BS}} = \frac{V_c}{U_1} U_{1satB} - \frac{U_{1S}}{a U_{vert}} \frac{dV_{th}}{dV_{BS}} - \frac{V_c}{U_{vert}} \frac{dU_{vert}}{dV_{BS}} - \frac{V_c}{a} \frac{da}{dV_{BS}} \quad \text{(D.52)}$$

$$\frac{dU_1}{dV_{BS}} = U_{1satB}\left[1 - \frac{U_{1d}(V_{DS} - V_{DSAT})^2}{V_{DSAT}^2}\right] \quad \text{(D.53)}$$
$$+ \frac{2U_{1S} U_{1d} V_{DS} (V_{DS} - V_{DSAT})}{V_{DSAT}^3} \frac{dV_{DSAT}}{dV_{BS}} \quad \text{(D.53)}$$

(D) Saturation region:

(D.1) drain current:

$$I_{DS} = \frac{\beta (V_{GS} - V_{th})^2}{2aK} FR \quad \text{(D.54)}$$

$$FR = \left[1 + A_i e^{\frac{-Bi}{(V_{DS} - V_{DSAT})}}\right] \quad \text{(D.55)}$$

$$\beta = \frac{\beta_0}{U_{vert}} = \frac{\beta_0}{1 + U_a (V_{GS} - V_{th}) + U_b (V_{GS} - V_{th})^2} \quad \text{(D.56)}$$

(D.2) bias-dependent parameters:

$$A_i = A_{i0} + A_{iB} V_{BS} \tag{D.57}$$

$$B_i = B_{i0} + B_{iB} V_{BS} \tag{D.58}$$

(D.3) transconductance (g_m):

$$g_m = \frac{dI_{DS}}{dV_{GS}} = \frac{I_{DS}}{\beta}\frac{d\beta}{dV_{GS}} + \frac{2I_{DS}}{V_{GS}-V_{th}} - \frac{I_{DS}}{K}\frac{dK}{dV_{GS}} + \frac{I_{DS}}{FR}\frac{dFR}{dV_{GS}} \tag{D.59}$$

$$\frac{d\beta}{dV_{GS}} = \frac{1}{U_{vert}}\left(\frac{d\beta_0}{dV_{GS}} - \beta\frac{dU_{vert}}{dV_{GS}}\right) \tag{D.60}$$

$$\frac{dFR}{dV_{GS}} = -A_i e^{\frac{-Bi}{(V_{DS}-V_{DSAT})}} \frac{B_i}{(V_{DS}-V_{DSAT})^2}\frac{dV_{DSAT}}{dV_{GS}}$$

$$= (1-FR)\frac{B_i}{(V_{DS}-V_{DSAT})^2}\frac{dV_{DSAT}}{dV_{GS}} \tag{G.61}$$

(D.4) output conductance (Gds):

$$g_{ds} = R_{out} = \frac{dI_{DS}}{dV_{DS}} = \frac{I_{DS}}{\beta}\frac{d\beta}{dV_{DS}} + \frac{2I_{DS}}{V_{GS}-V_{th}}\eta - \frac{I_{DS}}{K}\frac{dK}{dV_{DS}} + \frac{I_{DS}}{FR}\frac{dFR}{dV_{DS}} \tag{D.62}$$

$$\frac{d\beta}{dV_{DS}} = \frac{1}{U_{vert}}\left(\frac{d\beta_0}{dV_{DS}} - \beta\frac{dU_{vert}}{dV_{DS}}\right) \tag{D.63}$$

$$\frac{dFR}{dV_{DS}} = A_i e^{\frac{-Bi}{(V_{DS}-V_{DSAT})}} \frac{B_i}{(V_{DS}-V_{DSAT})^2}\left(1 - \frac{dV_{DSAT}}{dV_{DS}}\right)$$

$$= (FR-1)\frac{B_i}{(V_{DS}-V_{DSAT})^2}\left(1 - \frac{dV_{DSAT}}{dV_{DS}}\right) \tag{D.64}$$

(D.5) body- (back-gate)-transconductance (g_{mbs}):

$$g_{mbs} = \frac{dI_{DS}}{dV_{BS}} = \frac{I_{DS}}{\beta}\frac{d\beta}{dV_{BS}} - \frac{2I_{DS}}{V_{GS}-V_{th}}\frac{dV_{th}}{dV_{BS}} - \frac{I_{DS}}{K}\frac{dK}{dV_{BS}} \tag{D.65}$$

$$- \frac{I_{DS}}{a}\frac{da}{dV_{BS}} + \frac{I_{DS}}{FR}\frac{dFR}{dV_{BS}} \tag{D.65}$$

$$\frac{d\beta}{dV_{BS}} = \frac{1}{U_{vert}}\left(\frac{d\beta_0}{dV_{BS}} - \beta\frac{dU_{vert}}{dV_{BS}}\right) \tag{D.66}$$

$$\frac{dFR}{dV_{BS}} = -A_i e^{\frac{-Bi}{(V_{DS}-V_{DSAT})}} \frac{B_i}{(V_{DS}-V_{DSAT})^2}\frac{dV_{DSAT}}{dV_{BS}}$$

$$= (1-FR)\frac{B_i}{(V_{DS}-V_{DSAT})^2}\frac{dV_{DSAT}}{dV_{BS}} \tag{D.67}$$

(E) Transition region:

(E.1) effective gate capacitance:

$$C'_{ox} = \frac{C_{ox}C_{inv}}{C_{ox}+C_{inv}} \tag{D.68}$$

where

$$C_{inv} \approx C_d e^{\frac{(\phi_S - 2\phi_B)}{V_{tm}}}$$

$$\approx C_d e^{\frac{(V_{GS}-V'_{th})}{nV_{tm}}} \tag{D.69}$$

$$C_d = \sqrt{\frac{q\varepsilon_{si}N_{SUB}}{2\phi_B}} \tag{D.70}$$

$$\phi_B = V_{tm}\ln\left(\frac{N_{SUB}}{n_i}\right) \tag{D.71}$$

$$V'_{th} = V_{th} - V_{offset} \tag{D.72}$$

(E.2) bias-dependent parameters:

$$n = n0 + \frac{nB}{\sqrt{\phi_S - V_{BS}}} + nD\,V_{DS} \tag{D.73}$$

$$V_{offset} = V_{offset0} + V_{offsetB}\,V_{BS} + V_{offsetD}\,V_{DS} \tag{D.74}$$

(E.3) upper bound:

$$V_{G2} = V_{th} + n V_{tm} \ln\left(\frac{10 C_{ox}}{C_d}\right) \tag{D.75}$$

(E.4) lower bound:

$$V_{G1} = V_{th} + V_{offset} + n V_{tm} \ln\left(\frac{C_{ox}}{10 C_d}\right) \tag{D.76}$$

(E.5) subthreshold current:

$$I_{DS} = \beta (V_{tm})^2 e^{\frac{(V_{GS} - V_{th} + V_{offset})}{n V_{tm}}} \left[1 - e^{\frac{-V_{DS}}{V_{tm}}}\right] \tag{D.77}$$

(E.6) cubic spline function:

$$V'_{GS} = C_0 + C_1 (V_{GS} - V_{th}) + C_2 (V_{GS} - V_{th})^2 + C_3 (V_{GS} - V_{th})^3 \tag{D.78}$$

boundary conditions at V_{G2}:

$$V'_{GS} = V_{G2} \equiv RA \tag{D.79}$$

$$\frac{dV'_{GS}}{dV_{GS}} = 1 \tag{D.80}$$

boundary conditions at V_{G1}:

$$V'_{GS} = V_{th} + V_{offset} + \sqrt{2aK}\, V_{tm} e^{\frac{(V_{G1} - V_{th} - V_{offset})}{2n V_{tm}}} \equiv RC \tag{D.81}$$

$$\frac{dV'_{GS}}{dV_{GS}} = \frac{V_{G1} - V_{th} - V_{offset}}{2n V_{tm}} \equiv RD \tag{D.82}$$

(E.7) solutions of C_i's in (G.78)

$$C_1 = \frac{1}{\Delta} \begin{vmatrix} 1 & 2V_{G2} & 3V_{G2}^2 \\ R_D & 2V_{G1} & 3V_{G1}^2 \\ R_C - R_A + V_{G2} & V_{G1}^2 + V_{G2}^2 & 2V_{G2}^3 + V_{G1}^3 \end{vmatrix} \tag{D.83}$$

$$C_2 = \frac{1}{\Delta} \begin{vmatrix} 1 & 1 & 3V_{G2}^2 \\ 1 & R_D & 3V_{G1}^2 \\ V_{G1} & R_C - R_A + V_{G2} & 2V_{G2}^3 + V_{G1}^3 \end{vmatrix} \tag{D.84}$$

$$C_3=\frac{1}{\Delta}\begin{bmatrix}1 & 2V_{G2} & 1\\ 1 & 2V_{G1} & R_D\\ V_{G1} & V_{G1}^2+V_{G2}^2 & R_C-R_A+V_{G2}\end{bmatrix} \tag{D.85}$$

$$C_0=R_A-C_1V_{G2}-C_2V_{G2}^2-C_3V_{G2}^3 \tag{D.86}$$

where

$$\Delta=\begin{bmatrix}1 & 2V_{G2} & 3V_{G2}^2\\ 1 & 2V_{G1} & 3V_{G1}^2\\ V_{G1} & V_{G1}^2+V_{G2}^2 & 2V_{G2}^3+V_{G1}^3\end{bmatrix} \tag{D.87}$$

APPENDIX E

PROGRAM LIST OF "BSIMevaluate.c" IN SPICE3

```
#include "prefix.h"
#include <stdio.h>
#include <math.h>
#include "util.h"
#include "CKTdefs.h"
#include "BSIMdefs.h"
#include "TRANdefs.h"
#include "CONST.h"
#include "suffix.h"

/* This routine evaluates the drain current, its derivatives and the
 * charges associated with the gate,bulk and drain terminal
 * using the BSIM (Berkeley Short-Channel IGFET Model) Equations.
 */
void
BSIMevaluate(vds,vbs,vgs,here,model,gmPointer,gdsPointer,gmbsPointer,
        qgPointer,qbPointer,qdPointer,cggbPointer,cgdbPointer,cgsbPointer,
        cbgbPointer,cbdbPointer,cbsbPointer,cdgbPointer,cddbPointer,
        cdsbPointer,cdrainPointer,vonPointer,vdsatPointer,ckt)

    register CKTcircuit  *ckt;
    register BSIMmodel   *model;
    register BSIMinstance *here;
    double vds,vbs,vgs;
    double *gmPointer;
    double *gdsPointer;
    double *gmbsPointer;
    double *qgPointer;
    double *qbPointer;
    double *qdPointer;
    double *cggbPointer;
    double *cgdbPointer;
    double *cgsbPointer;
    double *cbgbPointer;
    double *cbdbPointer;
    double *cbsbPointer;
    double *cdgbPointer;
    double *cddbPointer;
    double *cdsbPointer;
    double *cdrainPointer;
    double *vonPointer;
```

```
    double *vdsatPointer;

{
    double gm,gds,gmbs;
    double qg,qb,qd;
    double cggb,cgbb,cgdb,cgsb;
    double cbgb,cbbb,cbdb,cbsb;
    double cdgb,cdbb,cddb,cdsb;
    double ua,ub,u1,u1inv;
    double duadVbs,dubdVbs,du1dVbs;
    double eta;
    double detadVbs;
    double Ai,Bi;
    double daidVbs,dbidVbs;
    double Vdlimit,Vglimit,Vblimit;
    double Vp,Vpinv;
    double sqrtvp,sqrtvpinv;
    double Von,Vth,Vth0;
    double dvthdVbs,dvth0dVbs,dvthdVds;
    double Vgsminvth,Vgsminvth0;
    double Vg1,Vg2;
    double Vof,Vtm,Vcom;
    double Vgeff,Vgeffinv;
    double Vc,Vcnew,sqrt2vc;
    double dvcdVbs,dvcdVgs,dvcdVds,dvcnewdVbs;
    double kk,kkinv,kknew;
    double dkkdVbs,dkkdVgs,dkkdVds,dkkdVc,dkknewdVbs;
    double sqrtkk,sqrtkkinv;
    double Vdsat,Vdsatinv,Vdsat0;
    double dvdsatdVbs,dvdsatdVds,dvdsatdVgs;
    double Ids,Ids0;
    double FR,FRinv;
    double dfrdVgs,dfrdVds,dfrdVbs;
    double Uvert,Uvertinv;
    double duvertdVds,duvertdVgs,duvertdVbs;
    double Usatvel;
    double dusatveldVds,dusatveldVbs;
    double Utot,Utotinv;
    double dutotdVds,dutotdVbs,dutotdVgs;
    double G,dgdVbs;
    double A,Ainv;
    double dadVbs;
    double alphax;
    double dalphaxdVbs;
    double n;
    double betalin,betasat,betatmp;
    double dbetatmpdVbs,dbetatmpdVgs;
```

```
double beta2,beta3,beta4;
double dbeta2dVgs,dbeta3dVgs,dbeta4dVgs;
double beta0,beta,betainv;
double dbeta0dVds,dbeta0dVgs,dbeta0dVbs;
double tanh,sqrsech;
double Con1,Con2,Con3,Con4;
double sqrvg1,sqrvg2,cubvg1,cubvg2;
double tmp,tmp1,tmp2,tmp3,tmp4,tmp5,tmp6,tmp7,tmp8,tmp9,tmp10,tmp11;
double tmp12,tmp13,tmp14,tmp15,tmp16,tmp17,tmp18,tmp19,tmp20,tmp21;
double tmp22,tmp23,dtmp9dVbs,dtmp9dVds,dtmp15dVbs;
double Coeffa,Coeffb,Coeffc,Coeffd,delta;
double Vgb,Vgb_Vfb,WLCox,sqrvdd;
int ChargeComputationNeeded;
int tmp24,i;

if( (ckt->CKTmode & (MODEAC | MODETRAN)) ||
      ((ckt->CKTmode & MODETRANOP) && (ckt->CKTmode & MODEUIC)) ||
      (ckt->CKTmode & MODEINITSMSIG) ) {
   ChargeComputationNeeded = 1;
} else {
   ChargeComputationNeeded = 0;
}
if (vbs <= 2.0 * here->BSIMvbb)
  {
   Vblimit = 2.0 * here->BSIMvbb;
   /* the dependence of parameters on Vbs is limited to 2Vbb */
   detadVbs = 0.0;
   duadVbs = 0.0;
   dubdVbs = 0.0;
   du1dVbs = 0.0;
   daidVbs = 0.0;
   dbidVbs = 0.0;
  }
else
  {
   Vblimit = vbs;
   detadVbs = here->BSIMetaB;
   duadVbs = here->BSIMuaB;
   dubdVbs = here->BSIMubB;
   du1dVbs = here->BSIMu1B;
   daidVbs = here->BSIMaiB;
   dbidVbs = here->BSIMbiB;
  }
if (vgs >= 2.0 * here->BSIMvgg)
  {
   Vglimit = 2.0 * here->BSIMvgg;
   /* the dependence of parameters on Vgs is limited to 2Vgg */
```

```
    dbeta2dVgs = 0.0;
    dbeta3dVgs = 0.0;
    dbeta4dVgs = 0.0;
   }
  else
   {
    Vglimit = vgs;
    dbeta2dVgs = here->BSIMbeta2G;
    dbeta3dVgs = here->BSIMbeta3G;
    dbeta4dVgs = here->BSIMbeta4G;
   }

  eta = here->BSIMeta0 + here->BSIMetaB * Vblimit;
  if ( eta <= 0.0 )
   {
    eta = 0.0;
    detadVbs = 0.0;
   }
  else if ( eta > 0.5 )
   {
    eta = 0.5;
    detadVbs = 0.0;
   }

  if ( vds >= 2.0 * here->BSIMvdd )
   {
    Vdlimit = 2.0 * here->BSIMvdd;
    dvthdVds = 0.0;
   }
  else
   {
    Vdlimit = vds;
    dvthdVds = -eta;
   }
  if ( vbs <-0.5 )
   {
    Vp = here->BSIMphi - vbs;
   }
  else
   {
    Vp = here->BSIMphi;
   }
  Vpinv = 1.0 / Vp;
  sqrtvp = sqrt(Vp);
  sqrtvpinv = 1.0 / sqrtvp;
  Vth = here->BSIMvfb + here->BSIMphi + here->BSIMk1 *
        sqrtvp - here->BSIMk2 * Vp - eta * Vdlimit;
```

```
Von = Vth;
dvthdVbs = here->BSIMk2 - detadVbs * Vdlimit - 0.5 * here->BSIMk1 *
        sqrtvpinv;
Vgsminvth = vgs - Vth;
G = 1.0 - 1.0 / (1.744 + 0.8364 * Vp);
dgdVbs = -0.8364 * (1.0 - G) * ( 1.0 - G);
A = 1.0 + 0.5 * G * here->BSIMk1 * sqrtvpinv;
Ainv = 1.0 / A;
dadVbs = 0.25 * here->BSIMk1 * sqrtvpinv * (2.0 * dgdVbs + G * Vpinv);
Vg1 = here->BSIMvghigh;
Vg2 = here->BSIMvglow;
Vof = here->BSIMvof0 + here->BSIMvofB * Vblimit + here->BSIMvofD * Vdlimit;
Vtm = 8.625e-5 * (here->BSIMtemp + 273); /* should be moved*/
n = here->BSIMn0 + here->BSIMnB * sqrtvpinv + here->BSIMnD * Vdlimit;
tmp18 = 2.0 * n * Vtm;
tmp19 = 1.0 / tmp18;
if (Vgsminvth > Vg1)
  {
   Vgeff = Vgsminvth;
  }
else if (Vgsminvth < Vg2)
  {
   tmp1 = Vgsminvth * tmp19;
   if (tmp1 <-15.0)
    {
     Vgeff = sqrt(2 * A) * Vtm * exp(0.5 * Vof - 15.0);
    }
   else
    {
     Vgeff = sqrt(2 * A) * Vtm * exp(0.5 * Vof + tmp1);
    }
  }
else
  {
   Con1 = Vg1;
   Con2 = 1.0;
   Con3 = sqrt(2.0 * A) * Vtm * exp(0.5 * Vof + Vg2 * tmp19);
   Con4 = Con3 * tmp19;
   sqrvg1 = Vg1 * Vg1;
   sqrvg2 = Vg2 * Vg2;
   cubvg1 = Vg1 * sqrvg1;
   cubvg2 = Vg2 * sqrvg2;
   tmp1 = sqrvg1 + sqrvg2;
   tmp2 = 2.0 * cubvg1 + cubvg2;
   tmp3 = Con3 - Con1 + Con2 * here->BSIMvghigh;
   delta = 2.0 * Vg2 * tmp2 + 3.0 * sqrvg1 * tmp1 + 6.0 * Vg1
       * cubvg2 - 6.0 * sqrvg1 *sqrvg2 - 2.0 * tmp2 * Vg1
```

```
            - 3.0 * sqrvg2 * tmp1;
        delta = 1.0 / delta;
        Coeffb = 2.0 * Con2 * Vg2 * tmp2 + 3.0 * sqrvg1 * tmp1 * Con4
            + 6.0 * Vg1 * sqrvg2 * tmp3 - 6.0 * Vg2 * sqrvg1 * tmp3
            - 2.0 * Vg1 * Con4 * tmp2 - 3.0 * sqrvg2 * Con2 * tmp1;
        Coeffb = Coeffb * delta;
        Coeffc = Con4 * tmp2 + 3.0 * sqrvg1 * tmp3 + 3.0 * cubvg2 *Con2
            - 3.0 * sqrvg1 * Vg2 * Con4 - Con2 * tmp2 - 3.0
            * sqrvg2 * tmp3;
        Coeffc = Coeffc * delta;
        Coeffd = 2.0 * Vg2 * tmp3 + Con2 * tmp1 + 2.0 * Vg1 * Vg2
            * Con4 - 2.0 * sqrvg2 * Con2 - 2.0 * Vg1 * tmp3 - Con4
            * tmp1;
        Coeffd = Coeffd * delta;
        Coeffa = Con1 - Con2 * Vg1 + Coeffc * sqrvg1 + 2.0 * Coeffd
            * cubvg1;
        Vgeff = Coeffa + Coeffb * Vgsminvth + Coeffc * Vgsminvth
            * Vgsminvth + Coeffd * Vgsminvth * Vgsminvth * Vgsminvth;
      }
    Vgeffinv = 1.0 / Vgeff;
    ua = here->BSIMua0 + here->BSIMuaB * Vblimit;
    ub = here->BSIMub0 + here->BSIMubB * Vblimit;
    Uvert = 1.0 + ua * Vgeff + ub * Vgeff * Vgeff;
    duvertdVgs = ua + 2.0 * ub * Vgeff;
    duvertdVds = duvertdVgs * dvthdVds;
    duvertdVbs = Vgeff * duadVbs + Vgeff * Vgeff * dubdVbs -
            duvertdVgs * dvthdVbs;
    Uvert = MAX(Uvert,0.5);
    Uvertinv = 1.0 / Uvert;
    u1 = here->BSIMu10 + here->BSIMu1B * Vblimit;
    u1 = MAX(u1,1e-5);
    u1inv = 1.0 / u1;
    Vc = u1 * Vgeff * Ainv * Uvertinv;
    dvcdVgs = Vc * Vgeffinv - Vc * Uvertinv * duvertdVgs;
    dvcdVds = Vc * dvthdVds * ( Uvertinv * duvertdVgs - Vgeffinv );
    dvcdVbs = Vc * u1inv * du1dVbs - Vc * Vgeffinv * dvthdVbs
        - Vc * Ainv * dadVbs - Vc * Uvertinv * duvertdVbs;
    sqrt2vc = sqrt(1 + 2.0 * Vc);
    dkkdVc = 0.5 * (1.0 + 1.0 / sqrt2vc);
    kk = 0.5 * ( 1.0 + Vc + sqrt2vc);
    kkinv = 1.0 / kk;
    sqrtkk = sqrt(kk);
    sqrtkkinv = 1.0 / sqrtkk;
    dkkdVgs = dkkdVc * dvcdVgs;
    dkkdVbs = dkkdVc * dvcdVbs;
    dkkdVds = dkkdVc * dvcdVds;
    Vdsat = Vgeff * Ainv * sqrtkkinv;
```

```
Vdsat = MAX(Vdsat,1e-5);
Vdsatinv = 1.0 / Vdsat;
dvdsatdVbs = - Vdsat * Ainv * dadVbs - Vdsat * Vgeffinv * dvthdVbs
        - 0.5 * Vdsat * kkinv * dkkdVbs;
dvdsatdVgs = Ainv * sqrtkkinv - 0.5 * Vdsat * kkinv * dkkdVgs;
dvdsatdVds = -Vdsat * Vgeffinv * dvthdVds - 0.5 * Vdsat * kkinv *
          dkkdVds;

betalin = here->BSIMbeta0 + here->BSIMbeta0B * Vblimit;
betasat = here->BSIMbetas0 + here->BSIMbetasB * Vblimit;
sqrvdd = here->BSIMvdd * here->BSIMvdd;
beta2 = here->BSIMbeta20 + here->BSIMbeta2B * Vblimit +
       here->BSIMbeta2G * Vglimit;
beta3 = here->BSIMbeta30 + here->BSIMbeta3B * Vblimit
     + here->BSIMbeta3G * Vglimit;
beta4 = here->BSIMbeta40 + here->BSIMbeta4B * Vblimit
     + here->BSIMbeta4G * Vglimit;
betatmp = betasat - betalin - beta3 * here->BSIMvdd + beta4 *
         sqrvdd;
dbetatmpdVbs = here->BSIMbetasB - here->BSIMbeta0B
           - here->BSIMbeta3B * here->BSIMvdd
           + here->BSIMbeta4B * sqrvdd;
dbetatmpdVgs = - dbeta3dVgs * here->BSIMvdd
           + dbeta4dVgs * sqrvdd;
tmp7 = beta2 * vds * Vdsatinv;
if ( tmp7 >20.0)
 {
   tanh = 1.0;
   sqrsech = 0.0;
 }
else if ( tmp7 <-20.0 )
 {
   tanh = -1.0;
   sqrsech = 0.0;
 }
else
 {
   tmp = exp(tmp7);
   tmp6 = tmp * tmp;
   tanh = (tmp6 - 1) / (tmp6 + 1);
   sqrsech = 4.0 * tmp6 / ((tmp6 + 1) * (tmp6 + 1));
 }
beta0 = betalin + betatmp * tanh + beta3 * vds - beta4 * vds * vds;
dbeta0dVgs = - betatmp * sqrsech * Vdsatinv * (tmp7 * dvdsatdVgs
         - here->BSIMbeta2G * vds) + dbeta3dVgs * vds
         - dbeta4dVgs * vds * vds + tanh * dbetatmpdVgs;
dbeta0dVds = beta3 + betatmp * sqrsech * ( beta2 * Vdsatinv
```

```
        - tmp7 * Vdsatinv * dvdsatdVds ) - 2 * beta4 * vds;
  dbeta0dVbs = here->BSIMbeta0B + dbetatmpdVbs * tanh +
            here->BSIMbeta3B * vds + betatmp * sqrsech * ( vds *
           Vdsatinv * here->BSIMbeta2B - tmp7 * Vdsatinv *
           dvdsatdVbs) - here->BSIMbeta4B * vds * vds;

  tmp22 = vds - Vdsat;
  if ( vds <= Vdsat )
   {
    tmp20 = (1.0 - here->BSIMu1D * tmp22 * tmp22 * Vdsatinv * Vdsatinv);
    Usatvel = u1 * tmp20;
    dusatveldVds = u1 * tmp22 * Vdsatinv * Vdsatinv * 2.0 *
              here->BSIMu1D * (vds * Vdsatinv * dvdsatdVds - 1);
    dusatveldVbs = tmp20 * du1dVbs + u1 * vds * tmp22 * 2.0
              * here->BSIMu1D * Vdsatinv * Vdsatinv
              * Vdsatinv * dvdsatdVbs;
    Utot = Uvert + Usatvel * vds;
    Utotinv = 1.0 / Utot;
    dutotdVds = duvertdVds + Usatvel + vds * dusatveldVds;
    dutotdVgs = duvertdVgs;
    dutotdVbs = duvertdVbs + vds * dusatveldVbs;
    beta = beta0 * Utotinv;
    betainv = 1.0 / beta;
    tmp21 = Vgeff - 0.5 * A * vds;
    Ids = beta * tmp21 * vds;
    gm = beta * vds + Ids * betainv * Utotinv * (dbeta0dVgs
      - beta * dutotdVgs);
    gds = Ids * betainv * Utotinv * (dbeta0dVds - beta
       * dutotdVds) + beta * tmp21 - beta * vds * (dvthdVds
       + 0.5 * A);
    gmbs = Ids * betainv * Utotinv * (dbeta0dVbs - beta * dutotdVbs)
        - beta * vds * (dvthdVbs + 0.5 * vds * dadVbs);
   }
  else
   {
    beta = beta0 * Uvertinv;
    betainv = 1.0 / beta;
    Ids0 = 0.5 * beta * Vgeff * Vgeff * Ainv * kkinv;
    Ai = here->BSIMai0 + here->BSIMaiB * Vblimit;
    Bi = here->BSIMbi0 + here->BSIMbiB * Vblimit;
    tmp = MAX(tmp22,1e-4);
    tmp4 = Bi / tmp;
    if ( tmp4 <50.0 )
      {
       tmp5 = exp(-tmp4);
       FR = 1.0 + Ai * tmp5;
       tmp23 = (1 - FR) * tmp4 / tmp;
```

```
        FRinv = 1.0 / FR;
        dfrdVgs = tmp23 * dvdsatdVgs;
        dfrdVds = -tmp23 * (1 - dvdsatdVds);
        dfrdVbs = tmp23 * dvdsatdVbs;
        Ids = Ids0 * FR;
       }
     else
       {
        FR = 1.0;
        FRinv = 1.0;
        dfrdVgs = 0.0;
        dfrdVds = 0.0;
        dfrdVbs = 0.0;
        Ids = Ids0;
       }
     gm = Ids * betainv * Uvertinv * (dbeta0dVgs - beta * duvertdVgs)
       + 2.0 * Ids * Vgeffinv - Ids * kkinv * dkkdVgs + Ids
       * FRinv * dfrdVgs;
     gds = Ids * betainv * Uvertinv * (dbeta0dVds - beta
        * duvertdVds) - 2.0 * Ids * Vgeffinv * dvthdVds
        - Ids * kkinv * dkkdVds + Ids * FRinv * dfrdVds;
     gmbs = Ids * betainv * Uvertinv * (dbeta0dVbs - beta
        * duvertdVbs) - 2.0 * Ids * Vgeffinv * dvthdVbs
        - Ids * Ainv * dadVbs - Ids * kkinv * dkkdVbs
        + Ids * FRinv * dfrdVbs;
   }

ChargeComputation:

   /* Some Limiting of DC Parameters */
   gm = MAX(gm,0.0);
   gds = MAX(gds,1.0e-20);
   gmbs = MAX(gmbs,0.0);

   WLCox = model->BSIMCox *
      (here->BSIMl - model->BSIMdeltaL * 1.e-6) *
      (here->BSIMw - model->BSIMdeltaW * 1.e-6) * 1.e4;   /* F */
   if( here->BSIMchannelChargePartitionFlag >1 )
    {
     ChargeComputationNeeded = 1;
    }

   if( ! ChargeComputationNeeded )
    {
     qg  = 0.0;
     qd = 0.0;
     qb = 0.0;
```

```
    cggb = 0.0;
    cgsb = 0.0;
    cgdb = 0.0;
    cdgb = 0.0;
    cdsb = 0.0;
    cddb = 0.0;
    cbgb = 0.0;
    cbsb = 0.0;
    cbdb = 0.0;
    goto finished;
  }

if( (here->BSIMchannelChargePartitionFlag == 1)
  {
/*0/100 partitioning for drain/source charges in saturation region*/

    Vgb = vgs - vbs ;
    Vgb_Vfb = Vgb - here->BSIMvfb;
    Vg1 = 0.2;
    Vg2 = -0.15;
    Vth0 = here->BSIMvfb + here->BSIMphi + here->BSIMk1
        * sqrtvp;
    Vgsminvth0 = vgs - Vth0;
    Vdsat0 = Vgsminvth0 * Ainv;
    Vdsat0 = MAX(Vdsat0,0);

    if( Vgb_Vfb <0.0)   /* Accumulation Region */
     {
       qg = WLCox * Vgb_Vfb;
       qb = - qg;
       qd = 0.0;
       cggb = WLCox;
       cgdb = 0.0;
       cgsb = 0.0;
       cgbb = - cggb;
       cbgb = - WLCox;
       cbdb = 0.0;
       cbsb = 0.0;
       cbbb = cggb;
       cdgb = 0.0;
       cddb = 0.0;
       cdsb = 0.0;
       cdbb = 0.0;
       goto finished;
     }
    else if ( Vgsminvth0 <= Vg2 )   /* Subthreshold Region */
     {
```

```
        tmp8 = sqrt(1.0 + 4.0 * Vgb_Vfb / (here->BSIMk1
            * here->BSIMk1));
        qg = 0.5 * WLCox * here->BSIMk1 * here->BSIMk1
          * (-1.0 + tmp8);
        qb = -qg;
        qd = 0.0;
        cggb = WLCox / tmp8;
        cgbb = - cggb;
        cgdb = cgsb = 0.0;
        cbbb = cggb;
        cbgb = - cggb;
        cbdb = cbsb = 0.0;
        cdgb = cddb = cdsb = cdbb = 0.0;
        goto finished;
       }
     else if ( Vgsminvth0 <Vg1 )
      {
       tmp8 = sqrt(1.0 + 4.0 * (Vgb_Vfb - Vgsminvth0 + Vg2)
          / (here->BSIMk1 * here->BSIMk1));
       Con1 = 0.5 * WLCox * here->BSIMk1 * here->BSIMk1
          * (-1.0 + tmp8);
       tmp9 = sqrt(1.0 + 4.0 * (Vgb_Vfb - Vgsminvth0)
          / (here->BSIMk1 * here->BSIMk1));
       Con2 = 0.5 * WLCox * here->BSIMk1 * here->BSIMk1
          * (-1.0 + tmp9);
       dvth0dVbs = - here->BSIMk1 * 0.5 * sqrtvpinv;
       if( vds <Vdsat0 )      /* triode region */
        {
          alphax = A;
          dalphaxdVbs = dadVbs;
          tmp9 = Vg1 - 0.5 * alphax * vds;
          tmp9 = MAX(tmp9, 1e-9);
          tmp2 = 1.0 / tmp9;
          dtmp9dVbs = -dvth0dVbs - 0.5 * vds * dalphaxdVbs;
          dtmp9dVds = - 0.5 * alphax;
          tmp10 = alphax * vds;
          tmp1 = tmp10 * tmp2;
          tmp11 = vds * tmp2;
          tmp12 = tmp10 * tmp11 * tmp2;
          tmp14 = 1.0 - alphax;
          tmp13 = tmp14 * vds;
          Con3 = WLCox * (Vg1 + Vth0 - here->BSIMvfb
             - here->BSIMphi - 0.5 * vds + 0.08333 * vds * tmp1);
          tmp20 = Con1 - Con2;
          tmp21 = Con3 - Con2;
          delta = 1.0 / (Vg2 * Vg1 * Vg1 - Vg1 * Vg2 * Vg2);
          Coeffb = (tmp20 * Vg1 * Vg1 - tmp21 * Vg2 * Vg2) * delta;
```

```
Coeffc = (Vg2 * tmp21 - Vg1 * tmp20) * delta;
qg = Con2 + Coeffb * Vgsminvth0 + Coeffc * Vgsminvth0
  * Vgsminvth0;
Con3 = - WLCox * ( - Vth0 + here->BSIMvfb + here->BSIMphi
    + 0.5 * tmp13 - 0.08333 * tmp13 * tmp1);
tmp21 = Con3 - Con2;
Coeffb = (tmp20 * Vg1 * Vg1 - tmp21 * Vg2 * Vg2) * delta;
Coeffc = (Vg2 * tmp21 - Vg1 * tmp20) * delta;
qb = - (Con2 + Coeffb * Vgsminvth0 + Coeffc * Vgsminvth0
  * Vgsminvth0);
tmp20 = Vg1 - Vg2;
tmp21 = 1.0 / (tmp20 * tmp20);
tmp22 = Vgsminvth0 - Vg2;
tmp23 = tmp22 * tmp22;
tmp19 = tmp21 * tmp23;
qd = - WLCox * (0.5 * Vg1 - 0.75 * tmp10
  + 0.125 * tmp10 * tmp1) * tmp19;
cggb = WLCox / tmp8 * (1.0 - tmp19) + WLCox * (1.0
    - 0.08333 * tmp12) * tmp19;
cgdb = WLCox * ( - 0.5 + 0.16667 * alphax * tmp11
    - 0.08333 * tmp12 * dtmp9dVds) * tmp19;
cgbb = - WLCox / tmp8 * (1.0 - tmp19) + WLCox * 0.08333
    * (vds * tmp11 * dalphaxdVbs - tmp12 * dtmp9dVbs)
    * tmp19;
cgsb = - (cggb + cgdb + cgbb);
cbgb = - WLCox / tmp8 * (1.0 - tmp19) + WLCox * 0.08333
    * tmp12 * tmp14 * tmp19;
cbdb = WLCox * tmp14 * (0.5 - 0.16667 * alphax * tmp11
    + 0.08333 * tmp12 * dtmp9dVds) * tmp19;
cbbb = WLCox / tmp8 * (1.0 - tmp19) - WLCox * (dvth0dVbs
    + 0.5 * vds * dalphaxdVbs + 0.08333 * vds * tmp11
    * (1.0 - 0.5 * alphax) *dalphaxdVbs - 0.08333
    * tmp14 * tmp12 * dtmp9dVbs) * tmp19;
cbsb = - (cbgb + cbdb + cbbb);
cdgb = - WLCox * (0.5 - 0.125 * alphax * tmp12) * tmp19;
cddb = WLCox * (0.75 * alphax - 0.25 * alphax * alphax
    * tmp11 + 0.125 * alphax * tmp12 * dtmp9dVds)
    * tmp19;
cdbb = WLCox * (0.5 * dvth0dVbs + 0.75 * vds
    * dalphaxdVbs - 0.25 * alphax * vds * tmp11
    * dalphaxdVbs + 0.125 * alphax * tmp12
    * dtmp9dVbs) * tmp19;
cdsb = - (cdgb + cddb + cdbb);
goto finished;
}
else
{
```

```
            alphax = A;
            dalphaxdVbs = dadVbs;
            tmp9 = 1.0 / (3.0 * alphax);
            tmp10 = Vg1 * tmp9;
            tmp11 = tmp10 / alphax;
            Con3 = WLCox * (Vth0 + Vg1 - here->BSIMvfb
               - here->BSIMphi - tmp10);
            tmp20 = Con1 - Con2;
            tmp21 = Con3 - Con2;
            delta = 1.0 / (Vg2 * Vg1 * Vg1 - Vg1 * Vg2 * Vg2);
            Coeffb = (tmp20 * Vg1 * Vg1 - tmp21 * Vg2 * Vg2) * delta;
            Coeffc = (Vg2 * tmp21 - Vg1 * tmp20) * delta;
            qg = Con2 + Coeffb * Vgsminvth0 + Coeffc * Vgsminvth0
              * Vgsminvth0;
            Con3 = - WLCox * (here->BSIMvfb + here->BSIMphi - Vth0
               + (1.0 - alphax) * tmp10);
            tmp21 = Con3 - Con2;
            Coeffb = (tmp20 * Vg1 * Vg1 - tmp21 * Vg2 * Vg2) * delta;
            Coeffc = (Vg2 * tmp21 - Vg1 * tmp20) * delta;
            qb = - (Con2 + Coeffb * Vgsminvth0 + Coeffc * Vgsminvth0
              * Vgsminvth0);
            tmp20 = Vg1 - Vg2;
            tmp21 = 1.0 / (tmp20 * tmp20);
            tmp22 = Vgsminvth0 - Vg2;
            tmp23 = tmp22 * tmp22;
            tmp19 = tmp21 * tmp23;
            qd = 0.0;
            cggb = WLCox / tmp8 * (1.0 - tmp19) + WLCox * (1.0
               - tmp9) * tmp19;
            cgdb = 0.0;
            cgbb = - WLCox / tmp8 * (1.0 - tmp19) + WLCox * (tmp9
               * dvth0dVbs + tmp11 * dalphaxdVbs) * tmp19;
            cgsb = - (cggb + cgdb + cgbb);
            cbgb = -WLCox / tmp8 * (1.0 - tmp19) + WLCox * (tmp9
               - 0.33333) * tmp19;
            cbdb = 0.0;
            cbbb = WLCox / tmp8 * (1.0 - tmp19) - WLCox * ((0.66667
               + tmp9) * dvth0dVbs + tmp11 * dalphaxdVbs) * tmp19;
            cbsb = - (cbgb + cbdb + cbbb);
            cdgb = 0.0;
            cddb = 0.0;
            cdbb = 0.0;
            cdsb = 0.0;
            goto finished;
          }
        }
      else if( vds <Vdsat0 )     /* triode region */
```

```
{
  alphax = A;
  dalphaxdVbs = dadVbs;
  tmp9 = Vgsminvth0 - 0.5 * alphax * vds;
  tmp9 = MAX(tmp9, 1e-9);
  tmp2 = 1.0 / tmp9;
  dtmp9dVbs = -dvth0dVbs - 0.5 * vds * dalphaxdVbs;
  dtmp9dVds = - 0.5 * alphax;
  tmp10 = alphax * vds;
  tmp1 = tmp10 * tmp2;
  tmp11 = vds * tmp2;
  tmp12 = tmp10 * tmp11 * tmp2;
  tmp14 = 1.0 - alphax;
  tmp13 = tmp14 * vds;
  qg = WLCox * (vgs - here->BSIMvfb - here->BSIMphi
    - 0.5 * vds + 0.08333 * vds * tmp1);
  qb = WLCox * ( - Vth0 + here->BSIMvfb + here->BSIMphi
    + 0.5 * tmp13 - 0.08333 * tmp13 * tmp1);
  qd = - WLCox * (0.5 * Vgsminvth0 - 0.75 * tmp10
    + 0.125 * tmp10 * tmp1);
  cggb = WLCox * (1.0 - 0.08333 * tmp12);
  cgdb = WLCox * ( - 0.5 + 0.16667 * alphax * tmp11
     - 0.08333 * tmp12 * dtmp9dVds);
  cgbb = WLCox * 0.08333 * (vds * tmp11 * dalphaxdVbs
     - tmp12 * dtmp9dVbs);
  cgsb = - (cggb + cgdb + cgbb);
  cbgb = WLCox * 0.08333 * tmp12 * tmp14;
  cbdb = WLCox * tmp14 * (0.5 - 0.16667 * alphax * tmp11
     + 0.08333 * tmp12 * dtmp9dVds);
  cbbb = - WLCox * (dvth0dVbs + 0.5 * vds * dalphaxdVbs
     + 0.08333 * vds * tmp11 * (1.0 - 0.5 * alphax)
     * dalphaxdVbs - 0.08333 * tmp14 * tmp12 * dtmp9dVbs);
  cbsb = - (cbgb + cbdb + cbbb);
  cdgb = - WLCox * (0.5 - 0.125 * alphax * tmp12);
  cddb = WLCox * (0.75 * alphax - 0.25 * alphax * alphax
     * tmp11 + 0.125 * alphax * tmp12 * dtmp9dVds);
  cdbb = WLCox * (0.5 * dvth0dVbs + 0.75 * vds * dalphaxdVbs
     - 0.25 * alphax * vds * tmp11 * dalphaxdVbs
     + 0.125 * alphax * tmp12 * dtmp9dVbs);
  cdsb = - (cdgb + cddb + cdbb);
  goto finished;
 }
else if( vds >= Vdsat0 )   /* saturation region   */
 {
  alphax = A;
  dalphaxdVbs = dadVbs;
  tmp9 = 1.0 / (3.0 * alphax);
```

```
        tmp10 = Vgsminvth0 * tmp9;
        tmp11 = tmp10 / alphax;
        qg = WLCox * (vgs - here->BSIMvfb - here->BSIMphi - tmp10);
        qb = WLCox * (here->BSIMvfb + here->BSIMphi - Vth0
          + (1.0 - alphax) * tmp10);
        qd = 0.0;
        cggb = WLCox * (1.0 - tmp9);
        cgdb = 0.0;
        cgbb = WLCox * (tmp9 * dvth0dVbs + tmp11 * dalphaxdVbs);
        cgsb = - (cggb + cgdb + cgbb);
        cbgb = WLCox * (tmp9 - 0.33333);
        cbdb = 0.0;
        cbbb = - WLCox * ((0.66667 + tmp9) * dvth0dVbs
            + tmp11 * dalphaxdVbs);
        cbsb = - (cbgb + cbdb + cbbb);
        cdgb = 0.0;
        cddb = 0.0;
        cdbb = 0.0;
        cdsb = 0.0;
        goto finished;
    }

    goto finished;
 }
else    /* ChannelChargePartionFlag  < 1 */
 {
/*40/60 partitioning for drain/source charges in saturation region*/
    tmp24 = tmp24 - 20;
    Vgb  =  vgs - vbs ;
    Vgb_Vfb  =  Vgb - here->BSIMvfb;
    Vg1 = 0.2;
    Vg2 = -0.15;
    Vth0 = here->BSIMvfb + here->BSIMphi + here->BSIMk1
        * sqrtvp;
    Vgsminvth0 = vgs - Vth0;
    Vdsat0 = Vgsminvth0 * Ainv;
    Vdsat0 = MAX(Vdsat0,0);

    if( Vgb_Vfb <0.0)    /* Accumulation Region */
     {
        qg = WLCox * Vgb_Vfb;
        qb = - qg;
        qd = 0.0;
        cggb = WLCox;
        cgdb = 0.;
        cgsb = 0.;
        cgbb = - cggb;
```

```
        cbgb = - WLCox;
        cbdb = 0.;
        cbsb = 0.;
        cbbb = cggb;
        cdgb = 0.;
        cddb = 0.;
        cdsb = 0.;
        cdbb = 0.0;
        goto finished;
      }
    else if ( Vgsminvth0 <= Vg2 )    /* Subthreshold Region */
     {
      tmp8 = sqrt(1.0 + 4.0 * Vgb_Vfb / (here->BSIMk1
          * here->BSIMk1));
      qg = 0.5 * WLCox * here->BSIMk1 * here->BSIMk1
        * (-1.0 + tmp8);
      qb = -qg;
      qd = 0.0;
      cggb = WLCox / tmp8;
      cgbb = - cggb;
      cgdb = cgsb = 0.0;
      cbbb = cggb;
      cbgb = - cggb;
      cbdb = cbsb = 0.0;
      cdgb = cddb = cdsb = cdbb = 0.0;
      goto finished;
     }
    else if ( Vgsminvth0 <Vg1 )    /* Subthreshold Region */
     {
      tmp8 = sqrt(1.0 + 4.0 * (Vgb_Vfb - Vgsminvth0 + Vg2)
         / (here->BSIMk1 * here->BSIMk1));
      Con1 = 0.5 * WLCox * here->BSIMk1 * here->BSIMk1
         * (-1.0 + tmp8);
      tmp9 = sqrt(1.0 + 4.0 * (Vgb_Vfb - Vgsminvth0)
         / (here->BSIMk1 * here->BSIMk1));
      Con2 = 0.5 * WLCox * here->BSIMk1 * here->BSIMk1
         * (-1.0 + tmp9);
    if( vds <Vdsat0 )      /* triode region */
     {
      alphax = A;
      dalphaxdVbs = dadVbs;
      tmp9 = Vg1 - 0.5 * alphax * vds;
      tmp9 = MAX(tmp9, 1e-9);
      tmp2 = 1.0 / tmp9;
      dtmp9dVbs = -dvth0dVbs - 0.5 * vds * dalphaxdVbs;
      dtmp9dVds = - 0.5 * alphax;
      tmp10 = alphax * vds;
```

```
tmp1 = tmp10 * tmp2;
tmp11 = vds * tmp2;
tmp14 = 1.0 - alphax;
tmp13 = tmp14 * vds;
tmp12 = tmp10 * tmp11 * tmp2;
tmp15 = alphax * tmp11 / tmp9;
tmp16 = tmp10 * tmp10;
tmp17 = 0.1667 * Vg1 * Vg1;
tmp18 = 0.125 * tmp10 * Vg1;
tmp19 = 0.025 * tmp16;
tmp20 = tmp17 - tmp18 + tmp19;
tmp21 = 2.0 * tmp15 / tmp9 * tmp20;
Con3 = WLCox * (Vg1 + Vth0 - here->BSIMvfb - here->BSIMphi
    - 0.5 * vds + 0.08333 * vds * tmp1);
tmp22 = Con1 - Con2;
tmp23 = Con3 - Con2;
delta = 1.0 / (Vg2 * Vg1 * Vg1 - Vg1 * Vg2 * Vg2);
Coeffb = (tmp22 * Vg1 * Vg1 - tmp23 * Vg2 * Vg2) * delta;
Coeffc = (Vg2 * tmp23 - Vg1 * tmp22) * delta;
qg = Con2 + Coeffb * Vgsminvth0 + Coeffc * Vgsminvth0
  * Vgsminvth0;
Con3 = - WLCox * ( - Vth0 + here->BSIMvfb + here->BSIMphi
    + 0.5 * tmp13 - 0.08333 * tmp13 * tmp1);
tmp23 = Con3 - Con1;
Coeffb = (tmp22 * Vg1 * Vg1 - tmp23 * Vg2 * Vg2) * delta;
Coeffc = (Vg2 * tmp23 - Vg1 * tmp22) * delta;
qb = - (Con2 + Coeffb * Vgsminvth0 + Coeffc * Vgsminvth0
  * Vgsminvth0);
tmp22 = Vg1 - Vg2;
tmp22 = 1.0 / (tmp22 * tmp22);
tmp23 = Vgsminvth0 - Vg2;
tmp23 = tmp23 * tmp23 * tmp22;
qd = - WLCox * (0.5 * Vg1 - 0.5 * tmp10
  + tmp15 * tmp20) * tmp23;
cggb = WLCox / tmp8 * (1.0 - tmp23) + WLCox * (1.0
    - 0.08333 * tmp12) * tmp23;
cgdb = WLCox * ( - 0.5 + 0.16667 * alphax * tmp11
    - 0.08333 * tmp12 * dtmp9dVds) * tmp23;
cgbb = - WLCox / tmp8 * (1.0 - tmp23) + WLCox * 0.08333
    * (vds * tmp11 * dalphaxdVbs - tmp12 * dtmp9dVbs)
    * tmp23;
cgsb = - (cggb + cgdb + cgbb);
cbgb = - WLCox / tmp8 * (1.0 - tmp23) + WLCox * 0.08333
    * tmp12 * tmp14 * tmp23;
cbdb = WLCox * tmp14 * (0.5 - 0.16667 * alphax * tmp11
    + 0.08333 * tmp12 * dtmp9dVds) * tmp23;
cbbb = WLCox / tmp8 * (1.0 - tmp23) - WLCox * (dvth0dVbs
```

```
            + 0.5 * vds * dalphaxdVbs + 0.08333 * vds * tmp11
            * (1.0 - 0.5 * alphax) * dalphaxdVbs - 0.08333
            * tmp14 * tmp12 * dtmp9dVbs) * tmp23;
        cbsb = - (cbgb + cbdb + cbbb);
        cdgb = - WLCox * (0.5 + tmp15 * (0.3333 * Vg1
            - 0.125 * tmp10) - tmp21) * tmp23;
        cddb = WLCox * (0.5 * alphax + tmp21 * dtmp9dVds
            - alphax * tmp2 * tmp2 * (tmp17 - 2.0 * tmp18
            + 3.0 * tmp19)) * tmp23;
        cdbb = WLCox * (0.5 * dvth0dVbs + 0.5 * vds * dalphaxdVbs
            + tmp21 * dtmp9dVbs - tmp11 / tmp9 * (tmp17
            * dalphaxdVbs - 0.3333 * alphax * Vg1 * dvth0dVbs
            - 2.0 * tmp18 * dalphaxdVbs + 0.125 * alphax * tmp10
            * dvth0dVbs + 3.0 * tmp19 * dalphaxdVbs)) * tmp23;
        cdsb = - (cdgb + cddb + cdbb);
        goto finished;
      }
    else if( vds >= Vdsat0 )    /* saturation region   */
      {
        alphax = A;
        dalphaxdVbs = dadVbs;
        tmp9 = 1.0 / (3.0 * alphax);
        tmp10 = Vg1 * tmp9;
        tmp11 = tmp10 / alphax;
        Con3 = WLCox * (Vth0 + Vg1 - here->BSIMvfb - here->BSIMphi
            - tmp10);
        tmp20 = Con1 - Con2;
        tmp21 = Con3 - Con2;
        delta = 1.0 / (Vg2 * Vg1 * Vg1 - Vg1 * Vg2 * Vg2);
        Coeffb = (tmp20 * Vg1 * Vg1 - tmp21 * Vg2 * Vg2) * delta;
        Coeffc = (Vg2 * tmp21 - Vg1 * tmp20) * delta;
        qg = Con2 + Coeffb * Vgsminvth0 + Coeffc * Vgsminvth0
            * Vgsminvth0;
        Con3 = - WLCox * (here->BSIMvfb + here->BSIMphi - Vth0
            + (1.0 - alphax) * tmp10);
        tmp21 = Con3 - Con2;
        Coeffb = (tmp20 * Vg1 * Vg1 - tmp21 *  Vg2 * Vg2) * delta;
        Coeffc = (Vg2 * tmp21 - Vg1 * tmp20) * delta;
        qb = - (Con2 + Coeffb * Vgsminvth0 + Coeffc * Vgsminvth0
            * Vgsminvth0);
        tmp20 = Vg1 - Vg2;
        tmp21 = 1.0 / (tmp20 * tmp20);
        tmp22 = Vgsminvth0 - Vg2;
        tmp23 = tmp22 * tmp22 * tmp21;
        qd = - WLCox * 0.266667 * Vg1 * tmp23;
        cggb = WLCox / tmp8 * (1.0 - tmp23) + WLCox * (1.0 - tmp9)
            * tmp23;
```

```
        cgdb = 0.0;
        cgbb = - WLCox / tmp8 * (1.0 - tmp23) + WLCox * (tmp9
          * dvth0dVbs + tmp11 * dalphaxdVbs) * tmp23;
        cgsb = - (cggb + cgdb + cgbb);
        cbgb = - WLCox / tmp8 * (1.0 - tmp23) + WLCox * (tmp9
          - 0.33333) * tmp23;
        cbdb = 0.0;
        cbbb = WLCox / tmp8 * (1.0 - tmp23) - WLCox * ((0.66667
          + tmp9) * dvth0dVbs + tmp11 * dalphaxdVbs) * tmp23;
        cbsb = - (cbgb + cbdb + cbbb);
        cdgb = - WLCox * 0.266667 * tmp23;
        cddb = 0.0;
        cdbb =  WLCox * 0.266667 * dvth0dVbs * tmp23;
        cdsb =  - (cdgb + cddb + cdbb);
        goto finished;
      }
      goto finished;
    }
    else if( vds <Vdsat0 )      /* triode region */
     {
      alphax = A;
      dalphaxdVbs = dadVbs;
      tmp9 = Vgsminvth0 - 0.5 * alphax * vds;
      tmp9 = MAX(tmp9, 1e-9);
      tmp2 = 1.0 / tmp9;
      dtmp9dVbs = - dvth0dVbs - 0.5 * vds * dalphaxdVbs;
      dtmp9dVds = - 0.5 * alphax;
      tmp10 = alphax * vds;
      tmp1 = tmp10 * tmp2;
      tmp11 = vds * tmp2;
      tmp14 = 1.0 - alphax;
      tmp13 = tmp14 * vds;
      tmp12 = tmp10 * tmp11 * tmp2;
      tmp15 = alphax * tmp11 * tmp2;
      tmp16 = tmp10 * tmp10;
      tmp17 = 0.16667 * Vgsminvth0 * Vgsminvth0;
      tmp18 = 0.125 * tmp10 * Vgsminvth0;
      tmp19 = 0.025 * tmp16;
      tmp20 = tmp17 - tmp18 + tmp19;
      tmp21 = 2.0 * tmp15 * tmp2 * tmp20;
      qg = WLCox * (vgs - here->BSIMvfb - here->BSIMphi
        - 0.5 * vds + 0.08333 * vds * tmp1);
      qb = WLCox * ( - Vth0 + here->BSIMvfb + here->BSIMphi
        + 0.5 * tmp13 - 0.08333 * tmp13 * tmp1);
      qd = - WLCox * (0.5 * Vgsminvth0 - 0.5 * tmp10
        + tmp15 * tmp20);
      cggb = WLCox * (1.0 - 0.08333 * tmp12);
```

```
        cgdb = WLCox * ( - 0.5 + 0.16667 * alphax * tmp11
           - 0.08333 * tmp12 * dtmp9dVds);
        cgbb = WLCox * 0.08333 * (vds * tmp11 * dalphaxdVbs
           - tmp12 * dtmp9dVbs);
        cgsb = - (cggb + cgdb + cgbb);
        cbgb = WLCox * 0.08333 * tmp12 * tmp14;
        cbdb = WLCox * tmp14 * (0.5 - 0.16667 * alphax * tmp11
           + 0.08333 * tmp12 * dtmp9dVds);
        cbbb = - WLCox * (dvth0dVbs + 0.5 * vds * dalphaxdVbs
           + 0.08333 * vds * tmp11 * (1.0 - 0.5 * alphax)
           * dalphaxdVbs - 0.08333 * tmp14 * tmp12 * dtmp9dVbs);
        cbsb = - (cbgb + cbdb + cbbb);
        cdgb = - WLCox * (0.5 + tmp15 * (0.3333 * Vgsminvth0
           - 0.125 * tmp10) - tmp21);
        cddb = WLCox * (0.5 * alphax + tmp21 * dtmp9dVds
           - alphax * tmp2 * tmp2 * (tmp17 - 2.0 * tmp18
           + 3.0 * tmp19));
        cdbb = WLCox * (0.5 * dvth0dVbs + 0.5 * vds * dalphaxdVbs
            + tmp21 * dtmp9dVbs - tmp11 / tmp9 * (tmp17
           * dalphaxdVbs - 0.3333 * alphax * Vgsminvth0
           * dvth0dVbs - 2.0 * tmp18 * dalphaxdVbs + 0.125
           * alphax * tmp10 * dvth0dVbs + 3.0 * tmp19
           * dalphaxdVbs));
        cdsb = - (cdgb + cddb + cdbb);
        goto finished;
      }
    else if( vds >= Vdsat0 )    /* saturation region   */
      {
       alphax = A;
       dalphaxdVbs = dadVbs;
       tmp9 = 1.0 / (3.0 * alphax);
       tmp10 = Vgsminvth0 * tmp9;
       tmp11 = tmp10 / alphax;
       qg = WLCox * (vgs - here->BSIMvfb - here->BSIMphi - tmp10);
       qb = WLCox * (here->BSIMvfb + here->BSIMphi - Vth0
         + (1.0 - alphax) * tmp10);
       qd = - WLCox * 0.266667 * Vgsminvth0;
       cggb = WLCox * (1.0 - tmp9);
       cgdb = 0.0;
       cgbb = WLCox * (tmp9 * dvth0dVbs + tmp11 * dalphaxdVbs);
       cgsb = - (cggb + cgdb + cgbb);
       cbgb = WLCox * (tmp9 - 0.33333);
       cbdb = 0.0;
       cbbb = - WLCox * ((0.66667 + tmp9) * dvth0dVbs
          + tmp11 * dalphaxdVbs);
       cbsb = - (cbgb + cbdb + cbbb);
       cdgb = - WLCox * 0.266667;
```

```
            cddb = 0.0;
            cdbb = WLCox * 0.266667 * dvth0dVbs;
            cdsb = - (cdgb + cddb + cdbb);
            goto finished;
        }
    }

finished:       /* returning Values to Calling Routine */

    *gmPointer = gm;
    *gdsPointer = gds;
    *gmbsPointer = gmbs;
    *qgPointer = qg;
    *qbPointer = qb;
    *qdPointer = qd;
    *cggbPointer = cggb;
    *cgdbPointer = cgdb;
    *cgsbPointer = cgsb;
    *cbgbPointer = cbgb;
    *cbdbPointer = cbdb;
    *cbsbPointer = cbsb;
    *cdgbPointer = cdgb;
    *cddbPointer = cddb;
    *cdsbPointer = cdsb;
    *cdrainPointer = MAX(Ids, 1.0e-50);
    *vonPointer = Von;
    *vdsatPointer = Vdsat;
}
```