AD A044851

Research and Development Technical Report

ECOM-75-1344-F

COMPUTER AIDED ENGINEERING OF SEMICONDUCTOR INTEGRATED CIRCUITS

prepared by

David P. Kennedy
Principal Investigator
University of Florida
Electron Device Research Center
College of Engineering
Gainesville, Florida 32611

DDC
RECEIVED
OCT 4 1977
A

August 31, 1977

Final Report for Period 1 January 1976 - 31 August 1977

Distribution Statement
Approved for Public Release, Distribution Unlimited

Prepared for

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Va. 22209

AD No.
DDC FILE COPY

ECOM

US ARMY ELECTRONICS COMMAND FORT MONMOUTH, NEW JERSEY 07703

The findings in this report are not
to be construed as an official Department
of the Army position unless so designated
by other authorized documents.

Citation of manufacturers' or trade names
does not constitute an official endorsement
or approval of the use thereof.

Destroy this report when it is no longer
needed. Do not return it to the originator.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (18) ECOM-75-1344-F (19)	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) Computer Aided Engineering of Semiconductor Integrated ~~Circuits~~ Circuits,		5. TYPE OF REPORT & PERIOD COVERED (9) Final Report. 1 Jan. 76 — 31 Aug. 77
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) David P. Kennedy, Principal Investigator		8. CONTRACT OR GRANT NUMBER(s) (15) DAAB07-75-C-1344 ARPA Order-2985
9. PERFORMING ORGANIZATION NAME AND ADDRESS Electron Device Research Center, Univ. of Fla., Dept. of Elec. Engrg., Gainesville, FL 32611		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) IT1 61101 A 00A (17) 00
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency (ARPA), 1400 Wilson Blvd., Arlington, VA 22209		12. REPORT DATE Aug. 1977
		13. NUMBER OF PAGES 373
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) U.S. Army Electronics Command, ATTN: DRSEL-TL-IG, Ft. Monmouth, NJ 07703		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

(11) 31 Aug 77

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(12) 384P.

18. SUPPLEMENTARY NOTES

This research was sponsored by the Defense Advanced Research Projects Agency (DARPA) pursuant to ARPA Order No. 2985 dated 18 February 1975, Program Code W15P8W.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

MODELING, Semiconductor and Devices
SEMICONDUCTOR Processing
PROCESS Modeling of Semiconductors

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The objectives of this program are to remove the empiricism associated with the design and manufacturing of custom integrated circuits for military applications and to reduce the cost of these circuits by devising improved computer-aided engineering techniques. Efforts of research covered by this report are Part I, Semiconductor Device Modeling conducted by the University of Florida and, Part II, Integrated Circuit Process Modeling conducted by Stanford University. Part I of the report deals →

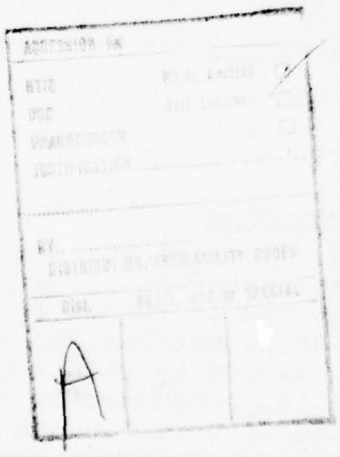
DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

409438

with: (a) A One-Dimensional Mathematical Model of MOSFET Operation; (b) The Inversion Layer Carrier Mobility in an MOSFET; (c) An Integral Equation-Relaxation Procedure for the Determination of Equilibrium Potentials in Semiconductor Devices; (d) A Test Pattern Model using Monte Carlo Methods of Analysis; (e) Equivalent Circuit Studies; and (f) A Two-Dimensional Model for MOSFET Operation. Part II deals with (a) Effects of Implantation Damage on Impurity Profiles in Annealed Si and Calculation of Range Profiles and Recoil Implantation in Multi-Layered Media; (b) Thermal Oxidation (a joint effort by Stanford University Integrated Circuits Laboratory, and Fairchild Camera and Instrument Corp. Research and Development Laboratory); (c) Silicon Epitaxy; and (d) A Mathematical Model of Impurity Diffusion (conducted by Louisiana State University).



SUMMARY

Final Report on Computer Engineering of Integrated Semiconductor Circuits

Custom integrated semiconductor circuits are often needed in the electronics industry. This need is particularly important for military applications. Military requirements involve a vast number of different, highly specialized, electronic systems, although seldom are these systems called for in large quantity. This combination produces an economically prohibitive situation. The large initial cost for designing integrated circuits, in conjunction with a small production requirement, creates an unreasonably large cost per IC chip. The source of this difficulty lies in empirical, and costly, engineering techniques used during both IC design, and production start-up. Empirical techniques are used during IC design to solve problems for which there is inadequate basic understanding. Similar empirical techniques are also used during production start-up to overcome problems that presently cannot be solved by design engineers.

An important source of this problem lies in the lack of adequate models for design and development of IC structures. These models are needed in three different areas: (1) for accurately predicting the physical characteristics of an integrated structure arising from numerous different fabrication processes; (2) for accurately predicting the electrical characteristics of semiconductor devices from their physical and geometrical properties; and (3) for accurately predicting the consequence of inadvertent fabrication process variables upon the ultimate electrical characteristics of a monolithic structure. An important goal for this program is to develop new models in these three areas; models that can be used in conjunction with existing computer methods for integrated circuit design and development.

Programs for model development have been initiated in two of the three above named areas. The University of Florida, Gainesville,

Florida, has undertaken development of semiconductor device models for IC design. Part I of this report outlines technical progress in this area during the total months of the program. Similarly, Stanford University, Stanford, California, has undertaken a program of process model development. Part II of this report outlines technical progress in this area during the total months of the program.

Table of Contents

		Page
Summary		iii
	Part I	
	Semiconductor Device Modeling (Univ. of Florida)	
Illustrations		1
Introduction to Part I		7
Chapter I	One-Dimensional Mathematical Model of MOSFET Operation	10
II	Inversion Layer Carrier Mobility in an MOSFET	118
III	An Integral Equation-Relaxation Procedure for the Determination of Equilibrium Potentials in Semiconductor Devices	143
IV	Test Pattern Model using Monte Carlo Methods of Analysis	163
V	Equivalent Circuit Studies	184
VI	A Two-Dimensional Model for MOSFET Operation	217
	Part II	
	Integrated Circuit Process Modeling (Stanford Univ.)	
Illustrations		v
Tables		viii
Introduction to Part II		1
Chapter I	Ion Implantation	4
II	Thermal Oxidation	37
III	Silicon Epitaxy	73
Appendix A		93
Appendix B		96

Part I

Semiconductor Device Modeling

University of Florida
Gainesville, Florida

Illustrations

Figure	Title	Page
1.1	Calculated inversion carrier distribution in a MOSFET	20
1.2	Calculated mobile carrier flux distribution in a MOSFET	21
1.3	Qualitative illustration of gate-induced electric field	22
1.4	Calculated potential distribution in a MOSFET (triode operation)	23
1.5	Calculated potential distribution in a MOSFET (saturation)	24
1.6	Calculated V-A Characteristics of a Long-Channel MOSFET	27
1.7	Calculated V-A Characteristics of a Long-Channel MOSFET	28
1.8	Calculated mobile carrier distribution in a MOSFET	29
1.9	Calculated V-A Characteristics of a Short-Channel MOSFET	30
1.10	Calculated V-A Characteristics of a Short-Channel MOSFET	31
1.11	Calculated Contours of Constant Source-Drain Electric Field	34
1.12	Calculated contour of constant (1000 V/cm) electric field (V_G = constant)	35
1.13	Calculated contour of constant (1000 V/cm) electric field (V_D = constant)	37
1.14	Source-junction potential barrier reduction	38
1.15	Calculated edge of source space-charge layer (V_D = 0.025)	39
1.16	Calculated edge of source space-charge layer (V_D = 1.0)	40
1.17	Calculated edge of source space-charge layer (V_D = 4.0)	41
1.18	Calculated gate threshold voltage of a MOSFET	43

Figure	Title	Page
1.19	Illustrative model of a MOSFET	45
1.20	Qualitative illustrations of Q_i and Q_{io}	55
1.21	One-dimensional calculations of Q_{io}	58
1.22	Calculated Q_i in a MOSFET (strong inversion)	63
1.23	Calculated dQ_i/dV_s in a MOSFET (strong inversion)	64
1.24	Calculated Q_i in a MOSFET (weak inversion)	65
1.25	Calculated dQ_i/dV_s in a MOSFET (weak inversion)	66
1.26	Calculated separation parameter, λ_o, in a MOSFET	68
1.27	Assumed boundary conditions for $V_s(x)$ in a MOSFET	70
1.28	Calculated magnitude of E_o vs V_o in a MOSFET	75
1.29	Calculated magnitudes of V_{sc} and V_o vs V_G	78
1.30	Calcualted ratio I_{drift}/I_{diff} vs V_o	82
1.31	Calculated transfer characteristic for a MOSFET ($N_A=2x10^{15}$; L=10μm)	89
1.32	Calculated transfer characteristic for a MOSFET ($N_A=2x10^{15}$; L=5μm)	90
1.33	Calculated transfer characteristic for a MOSFET ($N_A=2x10^{15}$; L=4μm)	91
1.34	Calculated transfer characteristic for a MOSFET ($N_A=2x10^{16}$; L=10μm)	92
1.35	Calculated transfer characteristic for a MOSFET ($N_A=2x10^{16}$; L=5μm)	93
1.36	Calculated transfer characteristic for a MOSFET ($N_A=2x10^{16}$; L=3μm)	94
1.37	Calculated transfer characteristic for a MOSFET ($N_A=2x10^{15}$; L=10μm)	95
1.38	Calculated transfer characteristic for a MOSFET ($N_A=2x10^{15}$; L=5μm)	96

Figure	Title	Page
1.39	Calculated transfer characteristic for a MOSFET ($N_A=2x10^{15}$; L=4μm)	97
1.40	Calculated transfer characteristic for a MOSFET ($N_A=2x10^{16}$; L=10μm)	98
1.41	Calculated transfer characteristic for a MOSFET ($N_A=2x10^{16}$; L=5μm)	99
1.42	Calculated transfer characteristic for a MOSFET ($N_A=2x10^{16}$; L=3μm)	100
1.43	Calculated drain characteristic for a MOSFET in strong inversion	102
1.44	Calculated drain characteristic for a MOSFET in weak inversion	103
1.45	Calculated drain characteristic for a MOSFET ($N_A=2x10^{16}$; L=3μm)	104
1.46	Calculated inversion charge distribution ($N_A=2x10^{16}$; L=3μm)	106
1.47	Calculated $V_s(x)$ in a MOSFET ($N_A=2x10^{16}$; L=3μm)	107
1.48	Calculated $V_s(x)$ near source junction ($N_A=2x10^{16}$)	108
1.49	Source junction potential barrier reduction ($N_A=2x10^{16}$; L=3μm)	109
1.50	Calculated contours of constant source-drain electric field in a short-channel MOSFET. $N_A = 2x10^{16}cm^{-3}$, t_{ox}=1000Å, L=3μm. V_G=5 volts, V_D=0.5 volts.	111
1.51	Calculated contours of constant source-drain electric field in a short-channel MOSFET. $N_A=2x10^{16}cm^{-3}$, t_{ox}=1000Å, L=3μm. V_G=5 volts, V_D=2 volts.	112
2.1	Mathematical model of an MOS structure.	122
2.2	Calculated electron energy levels in the surface potential well.	127

Figure	Title	Page
2.3	Comparison between quantum mechanical solution and traditional solution of the inversion layer carrier distribution.	129
2.4	Comparison between quantum mechanical solution and traditional solution of the inversion layer carrier distribution.	130
2.5	Inversion layer carrier distributions for three locations between the source and drain of an MOSFET.	132
3.1	Optimum Relaxation Parameter K as a Function of Width W for a Constant Gradient Junction.	150
3.2	Number of Iterations Required for $\varepsilon<10^{-3}$ as a Function of the Constant Relaxation Parameter K.	151
3.3	Calculated Potential Distribution for a Zero Biased, Constant Gradient, P-N Junction. Integration Width of 1000Å.	152
3.4	Calculated Potential Distribution for a Zero Biased, Constant Gradient, P-N Junction. Integration Width of 2000Å.	153
3.5	Calculated Potential Distribution for a Zero Biased, Constant Gradient, P-N Junction. Integration Width of 3000Å.	154
3.6	Calculated Potential Distribution for a Zero Biased, Constant Gradient, P-N Junction. Integration Width of 4000Å.	155
3.7	Potential Distribution in a Forward and Reverse Biased, Constant Gradient, P-N Junction for W=1200Å.	157
3.8	Generated Variable Relaxation Parameter $K_n(x)$ for a Constant Gradient Junction with W=3000Å.	159
3.9	Equilibrium Potential for a Constant Gradient P-N Junction Using a Variable Relaxation Parameter. Here W=3000Å and V_A=0.	160
4.1	Flowchart of computer program for sheet resistance calculations.	176

Figure	Title	Page
4.2	Predeposition Sheet Resistance Frequency Distribution.	178
4.3	C_o Frequency Distribution.	180
4.4	Drive-In Sheet Resistance Frequency Distribution.	181
4.5	W_{ox} Frequency Distribution.	182
5.1	The charging current $(i_J)_C$ at terminal J produces the accumulated charge dq_J	191
5.2	General equivalent circuit between each pair of terminals at an n-terminal electronic device.	194
5.3	The MOSFET divided into intrinsic and extrinsic parts.	197
5.4	Equivalent-circuit model of the intrinsic MOSFET.	201
5.5(a)	Functional dependence of capacitors and transcapacitors.	211
5.5(b)	Functional dependence of capacitors and transcapacitors.	212
5.6	Example of a circuit illustrating significance of the model.	214
6.1	(a) Cross-section of an n-channel MOSFET. (b) Rectangular region modeled. The metallurgical junctions between source-substrate and drain-substrate are regarded as planar and at the left and right borders of the region modeled.	219
6.2	Graded lattice of points used in the finite-difference equations dealt with in the computer model.	220
6.3	Boundary conditions for: (a) $\psi(x,y)$, (b) $\theta(x,y)$, and (c) $n(x,y)$.	226
6.4	Layout of the interlaced ψ and θ lattices used in the finite-difference equations.	230
6.5	Five-point "star" of ψ-lattice points used in the finite-difference equation for ψ.	232
6.6	Boundary conditions on ψ at the oxide-silicon interface are enforced through an appropriate choice for the location of the interface, leading to special finite-difference equations applicable for the two ψ-lattice rows bracketing the interface.	235

Figure	Title	Page
6.7	The charge term Q_p in the finite-difference equation for ψ is obtained by integrating the electron density $n(\xi,\eta)$ over the four quadrants shown.	240
6.8	Five point "star" of θ-lattice points used in the finite-difference equation for θ.	244

Introduction - Part I

Semiconductor Device Modeling

At the University of Florida, efforts have been directed toward the development of mathematical models that accurately predict the electrical characteristics of semiconductor devices. This effort has been directed toward three different aspects of the modeling problem: (1) rigorous two-dimensional models that are based upon an accurate description of the underlying physical mechanisms; (2) one-dimensional models that offer the accuracy needed for engineering purposes, yet require little computer time; and, (3) equivalent circuit representations of device operation that are applicable for circuit analysis purposes. In addition, development has also been initiated on mathematical models for test patterns to be used during IC fabrication process evaluation.

Included in this program is some research needed to further enhance our modeling capabilities. For example, research has been undertaken on topics relating to physical mechanisms of bipolar and MOS transistor operation that are known to be inadequately, or inaccurately, described by available modeling techniques. In addition, a research effort has been directed toward studies of integral methods of analysis, as a potentially better approach to the solution of two-dimensional boundary-value problems encountered in semiconductor device modeling.

For these two-dimensional models, initial work has been directed toward the MOSFET. Resulting from this effort is a two-dimensional model for this semiconductor device that can be used on an IBM 370 computer. Some details of this computer program are described in this report. More complete details are contained in a supplementary report which has been devised to serve as a Users Manual.

A one-dimensional model for MOSFET operation has been brought to completion during this contract period. This model has been tested extensively against a two-dimensional model, and the results are satisfactory. Complete details concerning this model are included in this report.

In MOSFET operation, the mobility of inversion layer carriers undergo important changes, with a change of applied gate voltage. This difficulty results from scattering at the oxide-semiconductor interface, which represents one side of the potential well in which the inversion layer resides. Resulting from our studies is a mathematical analysis of this physical situation, which is in adequate agreement with experiment for engineering purposes.

A statistical model has been developed for the sheet resistance measurements associated with a two-step boron diffusion process into silicon. This model utilizes Monte Carlo methods to approximate the process variations encountered during semiconductor device fabrication. From direct comparisons between experiment and theory, it is shown that this mathematical model can yield important information concerning inadvertent process variabilities associated with IC manufacturing.

Initial studies were undertaken on mathematical methods that are suitable for the two-dimensional transient analysis of semiconductor devices. Effort has been directed toward converting Poisson's equation to a Fredholm integral equation and, thereafter, solving this integral by iterative means, using the Picard technique. The purpose of this effort is to find a method whereby transient studies could be undertaken on semiconductor device operation, without requiring an unreasonable amount of computer time.

A new equivalent-circuit model has been developed for the large-signal, transient response of the MOSFET. The model represents the MOSFET operating as a four-terminal device and provides a more complete description of the currents flowing at each of the four terminals than was available previously. This description incorporates two physical mechanisms not accounted for in past work: (1) the non-reciprocity of the displacement currents occurring in the four-terminal device; and (2) the apportionment between source and drain of the currents flowing to change the inversion charge in the channel. In the derivation of the new model, an approach

is used that applies not only to the MOSFET but to all electronic devices described by charge control. The relation of this approach to the indefinite admittance matrix of circuit theory offers advantages in the modeling of devices having more than three terminals.

University of Florida
D. P. Kennedy
U. H. Kurzweg
F. A. Lindholm
A. D. Sutherland
M. Zahn

Chapter I
One-Dimensional Mathematical Model of MOSFET Operation

R. F. Motta and D. P. Kennedy

		Page
1.0	Introduction	11
2.0	Two-Dimensional Mechanisms in MOSFET Operation	17
	2.1 Electric Current Saturation	18
	2.2 Channel Termination in a MOSFET	19
	2.3 Velocity Saturation in a MOSFET	25
	2.4 Depletion Charge Distribution	32
3.0	A Continuity Theory for MOSFET Operation	44
	3.1 General Mathematical Development	44
	3.2 Physical Interpretation for $\lambda(V_s)$	47
	3.3 Volt-Ampere Characteristics of a MOSFET	48
	3.4 The Saturation Mode of Operation	51
	3.5 Calculation of Q_{io} in a MOSFET	52
	3.6 Calculation of λ_o in a MOSFET	54
4.0	A Revised MOSFET Theory	56
	4.1 Subthreshold Correction for Q_{io}	57
	4.2 Revised Calculation of λ_o	61
	4.3 Revised Calculation of $V_s(x)$	69
	4.4 A Modified Treatment of Current Saturation	74
	4.5 The Volt-Ampere Characteristics of a MOSFET	79
5.0	Evaluation of the Revised MOSFET Theory	84
	5.1 MOSFET Volt-Ampere Characteristics	86
	5.2 Short-Channel Efforts	101
6.0	Conclusions	110
7.0	List of Symbols	115

Chapter I

One-Dimensional Mathematical Model of MOSFET Operation

R. F. Motta and D. P. Kennedy

1.0 Introduction

Early theories of MOSFET operation [1-4] contain simplifications and approximations similar to those found in Shockley's theory of the unipolar transistor [5]. The purpose of these simplifications was to achieve mathematical tractability in analyzing this semiconductor structure and, for the device fabrication technology at that time, these theories produced calculated results in satisfactory agreement with experiment.

Following this early work, other researchers further developed and refined the theory of MOSFET operation. The physical mechanisms associated with this device were studied in detail, each investigation yielding an improved agreement between theory and experiment. At the present time numerous textbooks [6-15] offer discussions on this topic. In addition, the technical literature contains a multitude of papers covering a wide range of specialized studies of the physical and electrical properties of a MOSFET [16,17].

---

[1] H.K.J. Ihantola, *Stanford Electronics Laboratory, Technical Report No. 1661-1* (1961).

[2] S. R. Hofstein and F. P. Heiman, *Proc. IEEE*, 51, 1190 (1963).

[3] C. T. Sah, *IEEE Transactions on Electron Devices*, ED-11, 324 (1964).

[4] H.K.J. Ihantola and J. L. Moll, *Solid-State Electronics*, 7, 423 (1964).

[5] W. Shockley, *Proceedings of the IRE*, 40, 1365 (1952).

[6] T. J. Wallmark and H. J. Johnson, *Field Effect Transistors, Physics, Technology and Applications*, Prentice-Hall, NJ (1966).

[7] R. H. Crawford, *MOSFET in Circuit Design*, McGraw-Hill Company, NY (1967).

[8] A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley and Sons, Inc., NY (1967).

Concurrent with improvements in MOSFET theory, rapid changes took place in fabrication techniques for this semiconductor device. For example, integrated-circuit technologies were developed whereby high packing-density logic structures (e.g., CMOS) were used extensively in the design of complex electronic systems. These technologies utilized new and refined photolithographic methods to fabricate very short-channel MOSFET structures. In addition, TTL compatibility was achieved with the development of low voltage MOSFET devices. With these technological advances came an important recognition: these early theories of MOSFET operation sometimes yield results that are in poor agreement with experiment.

This situation initiated new studies into the physical mechanisms of MOSFET operation [18-59]. These studies clearly show that the simplifying assumptions and approximations of early MOSFET theory are not always applicable to the analysis of modern MOSFET devices. This inapplicability becomes particularly evident when dealing with the weak inversion mode of MOSFET operation and/or very short-channel structures.

---

[9] S. M. Sze, *Physics of Semiconductor Devices*, Wiley-Interscience, NY (1969).

[10] R.S.C. Cobbold, *Theory and Applications of Field-Effect Transistors*, Wiley-Interscience, NY (1970).

[11] D. J. Hamilton, F. A. Lindholm, and A. H. Marshak, *Principles and Applications of Semiconductor Device Modeling*, Holt, Rinehart and Winston, Inc., NY (1971).

[12] A. Many, Y. Goldstein, and N. B. Grover, *Semiconductor Surfaces*, American Elsevier Publishing Company, NY (1971).

[13] W. N. Carr and J. P. Mize, *MOS/LSI Design and Application*, McGraw-Hill Book Company, NY (1972).

[14] W. M. Penney (ed.), *MOS Integrated Circuits*, Van Nostrand Reinhold Company, NY (1972).

[15] P. Richman, *MOS Field-Effect Transistors and Integrated Circuits*, Wiley-Interscience, NY (1973).

[16] H. F. Storm, *IEEE Transactions on Electron Devices*, ED-14, 710 (1967); also ED-16, 957 (1969).

[17] E. S. Schlegel, *IEEE Transactions on Electron Devices*, ED-14, 728 (1967); also ED-15, 951 (1968).

[18] S. R. Hofstein and G. Warfield, *IEEE Transactions on Electron Devices*, ED-12, 129 (1965).

[19] V.G.K. Reddi and C. T. Sah, *IEEE Transactions on Electron Devices*, ED-12, 139 (1965).

[20] C. T. Sah and H. C. Pao, *IEEE Transactions on Electron Devices*, ED-13, 393 (1966).

[21] H. C. Pao and C. T. Sah, *Solid-State Electronics*, 9, 927 (1966).

[22] T. L. Chiu and C. T. Sah, *Solid-State Electronics*, 11, 1149 (1968).

[23] H. W. Loeb, R. Andrew, and W. Love, *Electronics Letters*, 4, 352 (1968).

[24] D. Frohman-Bentchkowsky, *Proc. IEEE*, 56, 217 (1968).

[25] D. Frohman-Bentchkowsky and A. S. Grove, *IEEE Transactions on Electron Devices*, ED-16, 108 (1969).

[26] G. A. Armstrong and J. A. Magowan, *Electronics Letters*, 6, 313 (1970).

[27] G. Baum and H. Beneking, *IEEE Transactions on Electron Devices*, ED-17, 481 (1970).

[28] D. P. Smith, *Stanford Electronics Laboratory, Technical Report No. 4825-4* (1971).

[29] D. P. Smith and J. G. Linvill, *IEEE International Solid-State Circuits Conference*, Philadelphia, PA, 40 (1971).

[30] R. M. Swanson and J. D. Meindl, *IEEE Journal of Solid-State Circuits*, SC-7, 146 (1972).

[31] M. B. Barron, *Solid-State Electronics*, 15, 293 (1972).

[32] B. Hoeneisen and C. A. Mead, *IEEE Transactions on Electron Devices*, ED-19, 382 (1972).

[33] J. Laur and T. S. Jayadevaiah, *Electronics Letters*, 8, 482 (1972).

[34] G. Merckel, J. Borel, and N. Z. Cupcea, *IEEE Transactions on Electron Devices*, ED-19, 681 (1972).

[35] A. Popa, *IEEE Transactions on Electron Devices*, ED-19, 774 (1972).

[36] D. P. Kennedy and P. C. Murley, *IBM Journal of Research and Development*, 17, 2 (1973).

[37] R. R. Troutman and S. N. Chakravarti, *IEEE Transactions on Circuit Theory*, CT-20, 659 (1973).

[38] R. R. Troutman, *IEEE Journal of Solid-State Circuits*, SC-9, 55 (1974).

[39] R. R. Troutman, *IEEE Transactions on Electron Devices*, ED-22, 1049 (1975).

[40] H. C. Poon, L. D. Yau, R. L. Johnston and D. Beecham, *International Electron Devices Meeting*, Washington, DC, 156 (1973).

[41] H. S. Lee, *Solid-State Electronics*, 16, 1407 (1973).

[42] R. C. Varshney, *Electronics Letters*, 9, 602 (1973).

[43] L. D. Yau, *Solid-State Electronics*, 17, 1059 (1974).

[44] Y. A. El-Mansy, *Modeling of Insulated-Gate Field-Effect Transistors*, Ph.D. Dissertation, Carleton University (1974).

[45] Y. A. El-Mansy and A. R. Boothroyd, *IEEE International Electron Devices Meeting*, Washington, DC, 35, (1974).

[46] H. Katto and Y. Itoh, *Solid-State Electronics*, 17, 1283 (1974).

[47] T. Masuhara, J. Etoh, and M. Nagata, *IEEE Transactions on Electron Devices*, ED-21, 363 (1974).

[48] R. M. Swanson, *Complementary MOS Transistors in Micropower Circuits*, Ph.D. Dissertation, Stanford University (1975).

[49] R. F. Motta and D. P. Kennedy, *Government Microcircuit Applications Conference (GOMAC)*, Lake Buena Vista, FL (1976 -- in preparation).

[50] J. E. Schroeder and R. S. Muller, *IEEE Transactions on Electron Devices*, ED-15, 954 (1968).

[51] M. B. Barron, *Stanford Electronics Laboratories, Technical Report No. 5501-1* (1969).

[52] G. A. Armstrong, J. A. Magowan, and M. D. Ryan, *Electronics Letters*, 5, 406 (1969).

[53] D. Vandorpe and N. H. Xuong, *Electronics Letters*, 7, 47 (1971).

[54] D. Vandorpe, J. Borel, G. Merckel, and P. Saintot, *Solid-State Electronics*, 15, 547 (1972).

[55] D. P. Kennedy, *Mathematical Simulation of the Effects of Ionizing Radiation on Semiconductors*, Final Report AFCRL-72-0257 (1972).

[56] M. S. Mock, *Solid-State Electronics*, 16, 601 (1973).

[57] F. H. De La Moneda, *IEEE Transactions on Circuit Theory*, CT-20, 666 (1973).

[58] G. D. Hachtel and M. H. Mack, *IEEE International Solid-State Circuits Conference*, Philadelphia, PA, 110 (1973).

[59] P. E. Cottrell and E. M. Buturla, *IEEE International Electron Devices Meeting*, Washington, DC, 51 (1975).

Some of these new studies developed and utilized two-dimensional computer solutions to obtain insight into MOSFET operation [50-59]. These studies have aided in understanding limitations of early MOSFET theory. Furthermore, in many cases such studies have also provided a basis for the development of new and improved MOSFET models [26,28,31,36,37,49].

Recent studies have shown that source-drain electric current continuity is not realized in early theories of MOSFET operation [36,60]. Specifically, there is no mechanism (either implied or stated) whereby electric current continuity can be attained. These theoretical treatments (and extensions of these treatments) have insufficient degrees of freedom whereby the requirement of electric current continuity could be imposed upon the approximating mathematical equations. In short, most available mathematical treatments of MOSFET operation assume that electric current continuity is realized, without imposing this characteristic as a requirement.

From a study of rigorous two-dimensional computer calculations, electric current continuity is, indeed, realized in the source-drain channel of a MOSFET, although this necessary situation is attained through mechanisms not previously discussed in the technical literature [55]. Namely, the total source-drain electric current (I_D) could be described as a sum of a drift component (I_{drift}) and a diffusion component (I_{diff}),

$$I_D = I_{drift} + I_{diff} \quad , \tag{1.1}$$

where the ratio of these current components changes continuously between the source and the drain. Briefly, near the drain junction I_D arises primarily from drift, and the diffusion component is negligible. In regions near the drain junction (removed from the drain space-charge layer) this drain current arises primarily from diffusion, with a substantial reduction of drift mechanisms.

---

[60] D. P. Kennedy and F. A. Lindholm, *Electron Device Research Center, Final Report No. HDL-CR-75-193-1,* (1975).

From this computer solution it became obvious that diffusion is an important mechanism associated with charge transport in the inversion layer of a MOSFET. It was also obvious that electric current continuity is not accurately described in any theory of MOSFET operation without explicitly including in this theory a source-drain electric current arising from both drift and diffusion.

In a previous study this aspect of MOSFET theory was addressed [36], and a modified theory was proposed that assures source-drain electric current continuity in this semiconductor device. It was shown that electric current continuity arises through two-dimensional mechanisms that are fundamental to MOSFET operation. Through these two-dimensional mechanisms an additional degree of freedom is realized in the operation of this semiconductor device and, thereby, electric current continuity is attained.

Furthermore, this new theory of MOSFET operation was reduced to a simplified one-dimensional system of equations that are suitable for engineering purposes. It was shown that this simplified theory is adequate for engineering purposes when applied to long-channel MOSFET structures (channel length >10μm), but that this theory fails for other important situations. Namely, this theory fails to adequately describe the volt-ampere characteristics of short-channel MOSFET structures (channel length $<\approx 7$ μm). Further, this theory fails to adequately describe the volt-ampere characteristics of a MOSFET structure in its weak inversion mode of operation, regardless of channel length.

Thus, we have the main purpose and direction of the present research effort: to identify and eliminate inadequacies associated with this previous theory of MOSFET operation that render it inapplicable to these aforementioned situations.

As an aid in this task of model development, a computer program was available for the two-dimensional analysis of MOSFET operation. The mathematical equations solved by this computer program represent general mechanisms associated with hole-electron transport in semiconductor material; these equations are derived from solid-state physics, and they are not unique to the device

under consideration. Further, the electrical properties predicted by this computer program agree with experiment, particularly in areas where traditional theory fails: short-channel structures and weak inversion operation. Thus, in the present study we used this computer program to study physical mechanisms of MOSFET operation that could not be determined through laboratory experiment. In a computational sense, we could "see" inside the transistor and accurately evaluate the dominant mechanisms influencing its electrical characteristics.

Using results of these computer calculations, in section 2 we discuss important two-dimensional mechanisms of MOSFET operation that are relevant to the present study. Section 3 of this chapter gives an overview of an earlier theory of MOS transistor operation that satisfies the requirement for electric current continuity in this semiconductor device. With an aim of extending the applicability of this earlier theory, we introduce our refinements and revisions in section 4. Through comparisons with our two-dimensional computer calculations and with two other widely used MOS transistor models, in section 5 we make an extensive evaluation of this revised theory. Finally, in section 6, we summarize conclusions drawn from the present research and offer recommendations for further research in this area.

2.0 Two-Dimensional Mechanisms in MOSFET Operation

Because the gate electrode and the drain junction are geometrically perpendicular, MOS transistor operation contains inherent two-dimensional mechanisms. For this reason, a rigorous analysis of this semiconductor device entails a self-consistent solution of an equation for electric current continuity and Poisson's equation in two spatial dimensions. No closed-form solution for this problem has been found.

With insight gained from two-dimensional computer calculations of MOSFET operation, in the present study we developed a simplified one-dimensional theory for this semiconductor device. In sections 3 and 4 we will present this theory and show that it implicitly takes into account some two-dimensional aspects of MOS transistor operation.

From our computer solution for this MOSFET problem, numerous two-dimensional mechanisms become clear. With the aid of these computer calculations, in the present section we will illustrate many of these mechanisms. Moreover, using these illustrations we will identify important physical mechanisms that produce "short-channel effects" in this semiconductor structure. Later, in section 5, we will further utilize these illustrations to assess limitations in our one-dimensional MOSFET theory.

2.1 Electric Current Saturation

According to traditional MOSFET theory, electric current saturation is attributable to a channel pinch-off mechanism [1]. Implied by this theory is that current saturation is associated with two different electrostatic potential distributions; i.e.: 1) one distribution between the source and a pinch-off point, and 2) another distribution between this pinch-off point and the drain. This concept of MOSFET operation has undergone significant modification by various authors [19,25,26,28].

Another viewpoint attributes electric current saturation to constriction of the source-drain channel in the vicinity of a pinch-off point. Pao and Sah viewed this constriction as a "bottleneck" in which diffusion is an important mechanism of carrier transport [21]. This viewpoint was corroborated by two-dimensional numerical solutions of MOSFET operation, calculated by Barron [51], and by Vandorpe and Xuong [53]. In contrast, another two-dimensional solution of this MOSFET problem shows no evidence to support this traditional concept of a pinch-off mechanism [36,55].

From our two-dimensional solution Fig. 1.1 shows calculated inversion carrier distributions in a MOSFET biased in the saturation mode. This figure shows a notable modification of channel configuration in the vicinity of the drain. However, in contrast to the traditional pinch-off concept, significant inversion carriers reside in this drain region. Further, an associated calculation of the mobile carrier flux distribution (Fig. 1.2) shows that saturation mode operation produces an increase of channel width. These illustrations place in question the concept of channel pinch-off.

2.2 Channel Termination in a MOSFET

For MOSFET drain voltages in excess of the gate voltage, the potential difference between gate-electrode and oxide-semiconductor interface (V_G-V_S) changes polarity at some point between the source and drain. This polarity change produces a reversal in the direction of the gate-induced electric field (Fig. 1.3). Between this point and the source, the gate-induced field produces an accumulation of inversion carriers along the semiconductor surface--thereby forming a conductive channel. Conversely, between this point and the drain, the gate-induced electric field forces mobile carriers away from this surface. (This component of electric field terminates on ions in the drain island.) Between this channel termination point and the drain, a negligible density of inversion carriers resides at the semiconductor surface (Fig. 1.1).

Figures 1.4 and 1.5 show rigorous two-dimensional calculations of electrostatic potential within a MOSFET substrate. Fig. 1.4 shows the potential distribution corresponding to a triode mode bias condition. Conversely, in Fig. 1.5 this MOSFET is biased well into saturation mode operation. Far from the gate electrode, regions of the drain depletion layer exhibit a potential distribution that is well approximated by one-dimensional abrupt p-n junction theory. In contrast, near the oxide-semiconductor inter-

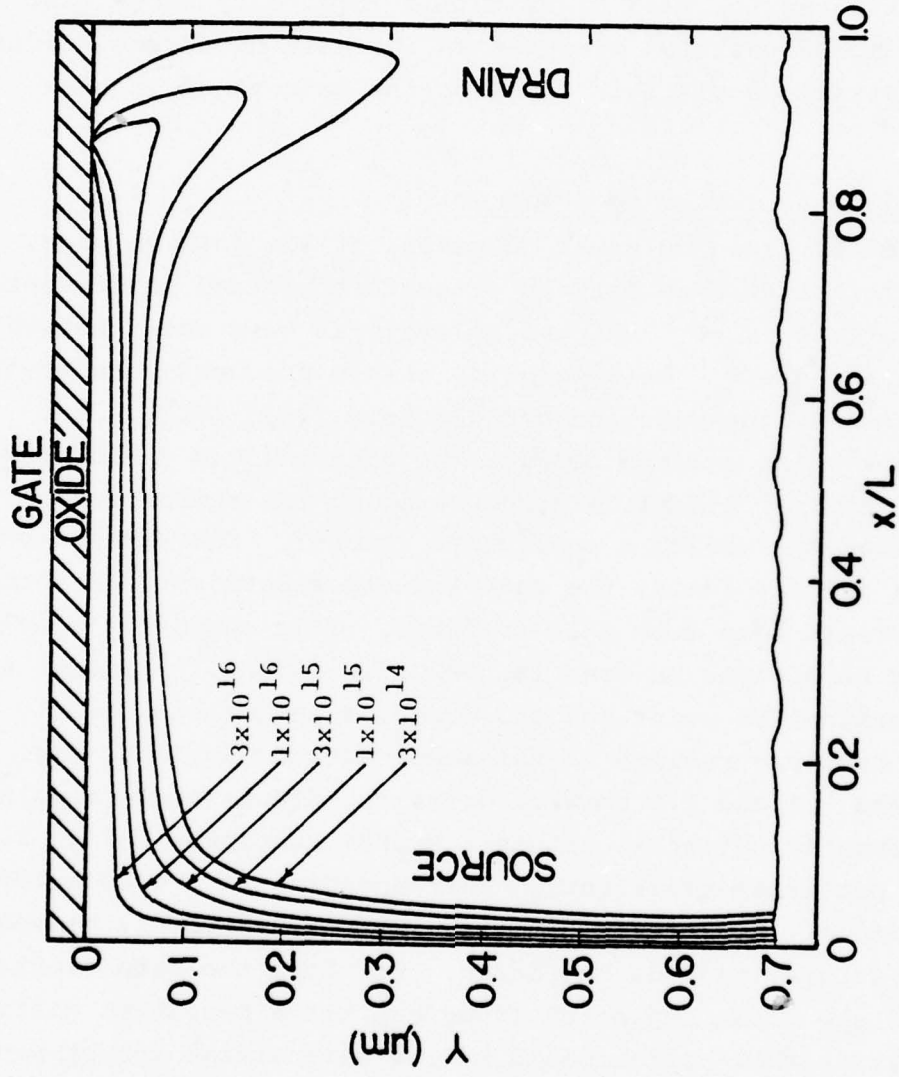


Fig. 1.1. Calculated inversion carrier distribution in a MOSFET, assuming a constant carrier mobility. V_G = 3 volts, V_D = 6 volts [55].

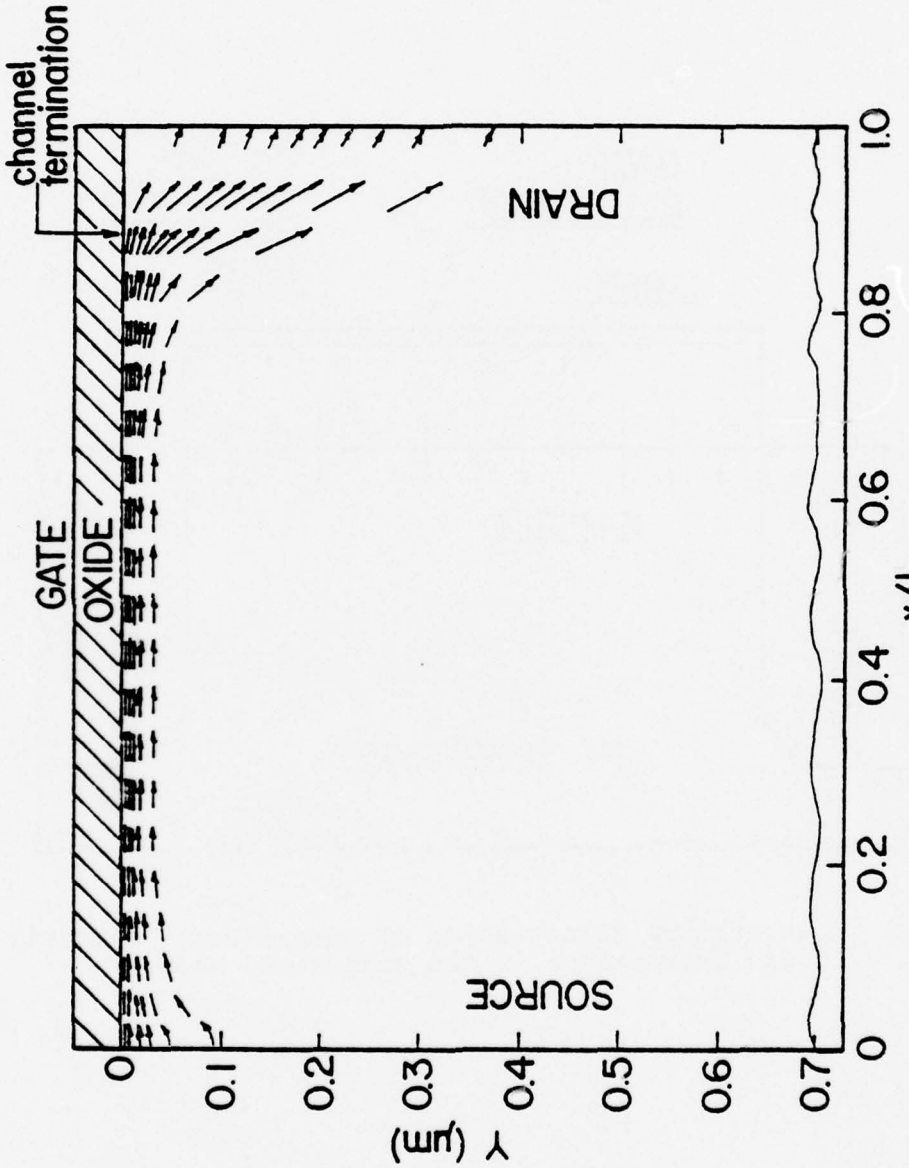


Fig. 1.2. Calculated mobile carrier flux distribution in a MOSFET, assuming a constant carrier mobility. V_G = 3 volts, V_D = 6 volts [55].

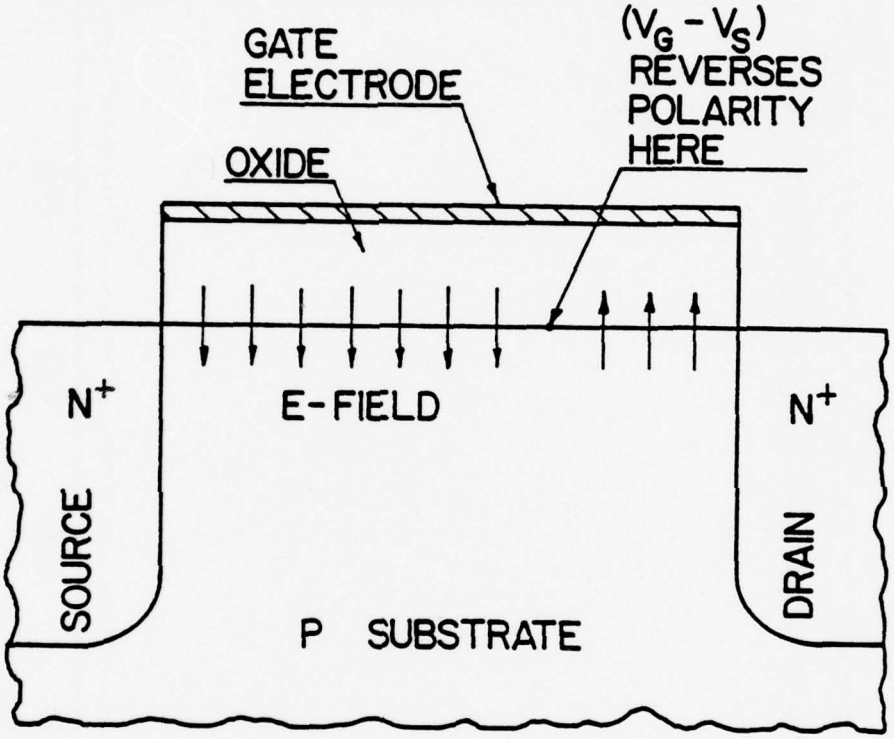


Fig. 1.3. Qualitative illustration of gate-induced electric-field orientation in the saturation mode.

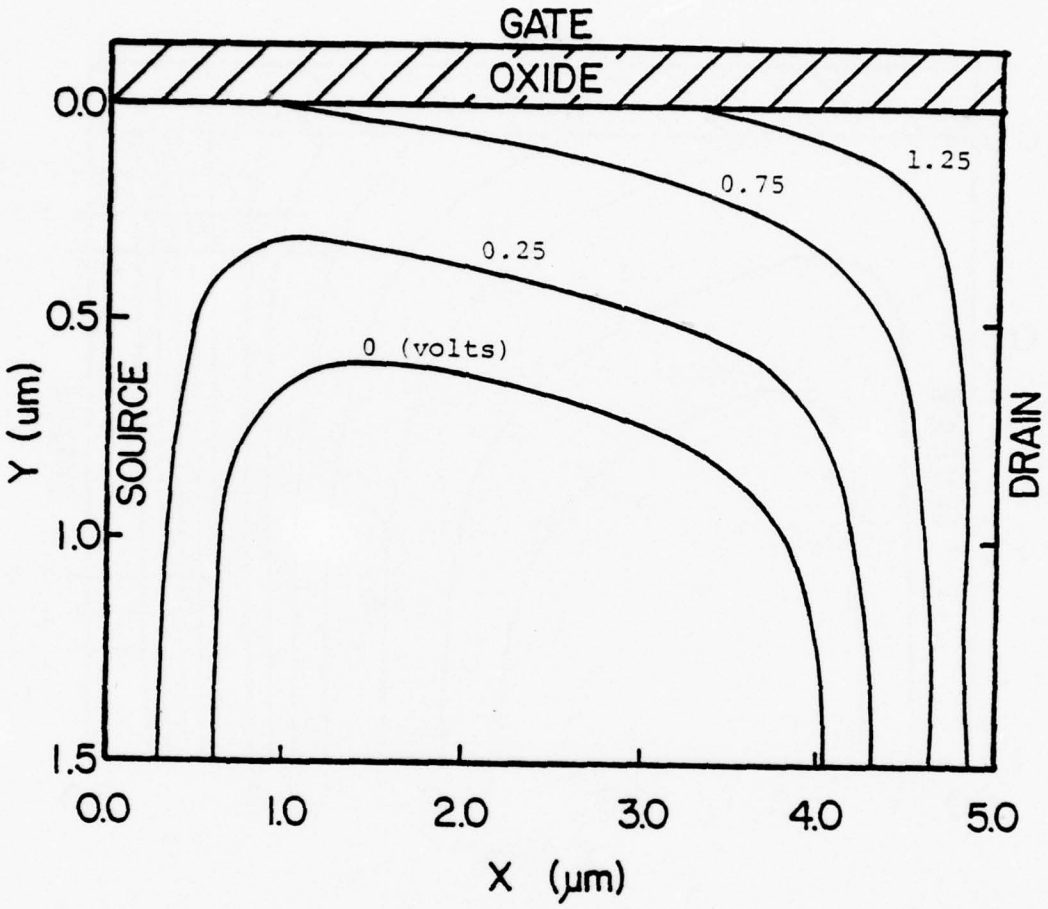


Fig. 1.4. Calculated potential distribution in a MOSFET, biased in the triode mode. $N_A = 2 \times 10^{15} cm^{-3}$, $t_{ox} = 1000A°$, $L = 5\mu m$. $V_G = 3.75$ volts, $V_D = 1.75$ volts [36].

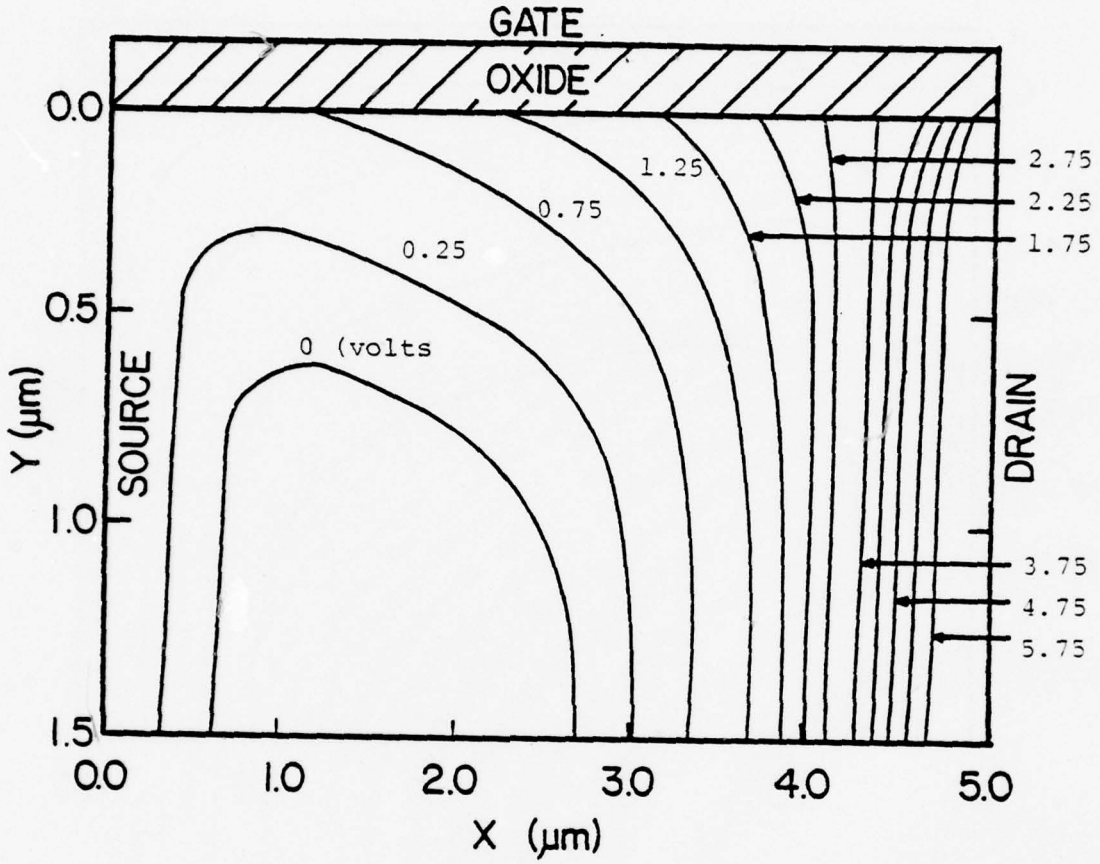


Fig. 1.5. Calculated potential distribution in a MOSFET biased well into the saturation mode. $N_A = 2 \times 10^{15} cm^{-3}$, $t_{ox} = 1000A°$, $L = 5\mu m$. $V_G = 3.75$ volts, $V_D = 6.75$ volts [36].

face, depletion charges are shared between the gate electrode and the drain junction. In these regions the equipotential contours are parallel to neither the gate electrode nor the drain junction.

Substantial differences can be observed between the potential distributions for triode mode and saturation mode operation (Figs. 1.4 and 1.5, respectively). These differences arise from a characteristic of the gate-induced electric field: in triode operation, this field is unidirectional between source and drain. In contrast, for bias into deep saturation, this field reverses direction at a point where surface potential equals gate voltage. Fig. 1.5 shows an equipotential line parallel to the drain junction at this point of field reversal. Clearly, the potential difference between the drain and this equipotential line (V_D-V_G) is less than the applied drain voltage; therefore, this point of channel termination must always reside within this drain depletion layer. It is emphasized that this point of channel termination carries no implication of a pinch-off or channel constriction mechanism. Within both the inversion channel and the adjacent drain depletion layer, a substantial source-drain electric field forces mobile carriers towards the drain junction.

2.3 Velocity Saturation in a MOSFET

Rigorous two-dimensional calculations of MOSFET operation show that two fundamentally different mechanisms are capable of producing electric current saturation. In a long-channel MOSFET deep saturation is a consequence of the potential distribution in the drain depletion layer -- between channel termination and the drain junction. Throughout this depletion layer, the longitudinal potential distribution is essentially determined by unneutralized impurity ions associated with the drain junction. Therefore, following the onset of channel termination, any further increase in applied drain voltage produces an increase in voltage between the point of channel termination and the drain; and the voltage

across the active channel (between the source and channel termination) remains essentially constant. In contrast, electric current saturation in a short-channel MOSFET can result from carrier velocity saturation within the source-drain channel.

Computer calculations show that velocity saturation has a negligible influence on the volt-ampere characteristics of a long-channel MOSFET; this situation is seen through a comparison of Figs. 1.6 and 1.7. In a long-channel MOSFET the electric fields are sufficient to produce velocity saturation only within the drain space-charge layer -- a region where the mobile carrier trajectory is two-dimensional, Fig. 1.2.

This two-dimensional trajectory is a consequence of a large electric field component E_x forcing carriers toward the drain island, and a small electric field component E_y forcing carriers away from the oxide-semiconductor interface. The relative magnitude of these electric field components imply velocity saturation in a direction parallel to the oxide-semiconductor interface (x-axis), and little (or no) velocity saturation *perpendicular* to this interface. As a consequence, velocity saturation produces a greater amount of two-dimensional carrier flow (compare Figs. 1.2 and 1.8), with little (or no) influence upon the total drain current. In short, the drain space-charge region remains an almost infinite sink for mobile carriers, despite the presence of velocity saturation. Thus, velocity saturation produces little (or no) change in the volt-ampere characteristics of a long-channel MOSFET.

Contrasting with this long-channel situation, velocity saturation of mobile carriers has a profound influence upon the volt-ampere characteristics of a very short-channel MOSFET (compare Figs. 1.9 and 1.10). Here we find that velocity saturation in a short-channel MOSFET produces a lower output conductance in regions of electric current saturation. In addition, velocity saturation produces electric current saturation at a drain voltage substantially lower than would be predicted from the traditional theory of MOSFET operation.

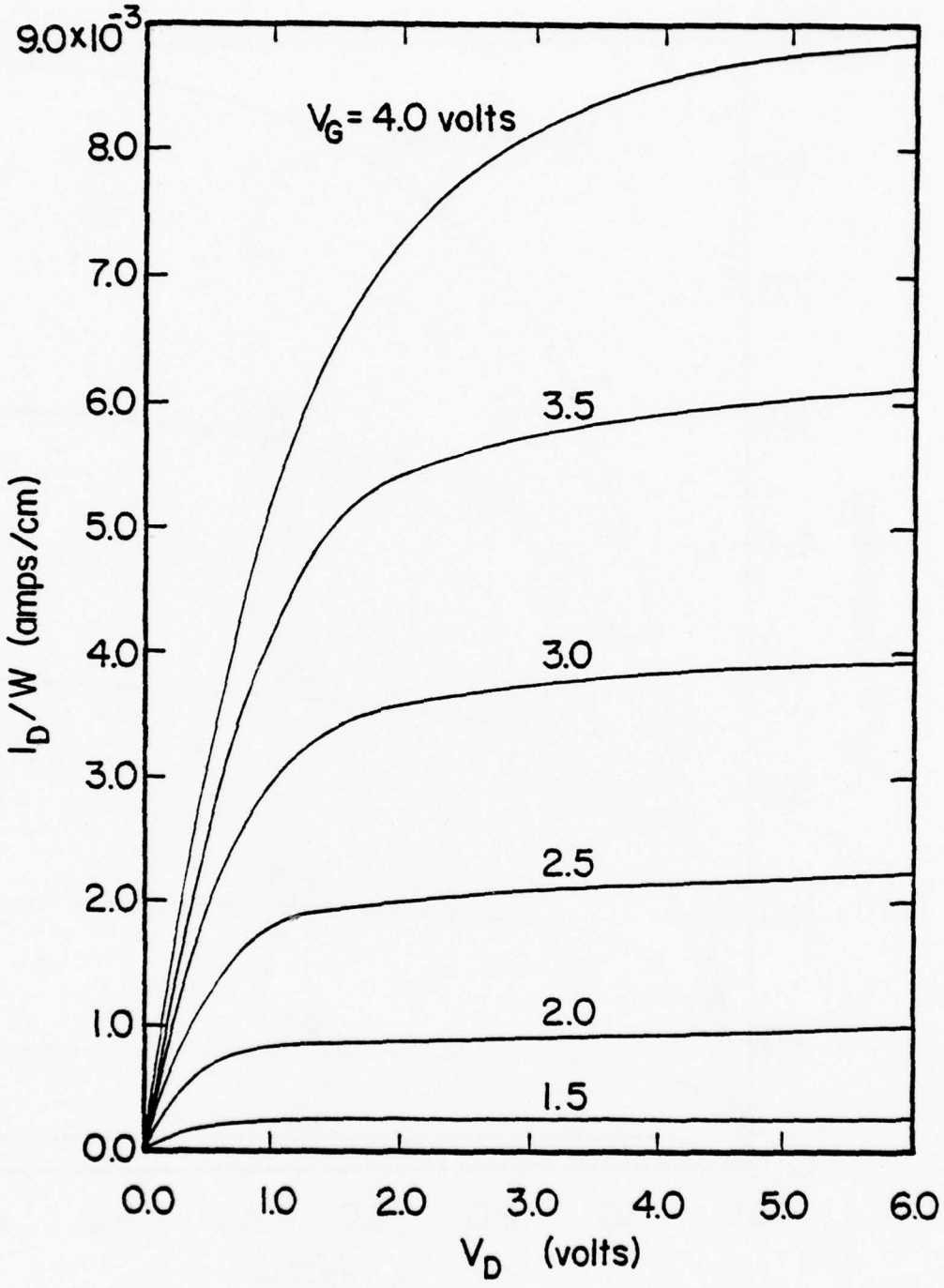


Fig. 1.6. Calculated volt-ampere characteristics of a long-channel MOSFET, assuming a constant carrier mobility. $N_A = 2x10^{15}cm^{-3}$, t_{ox} = 1000A°, L = 10μm [55].

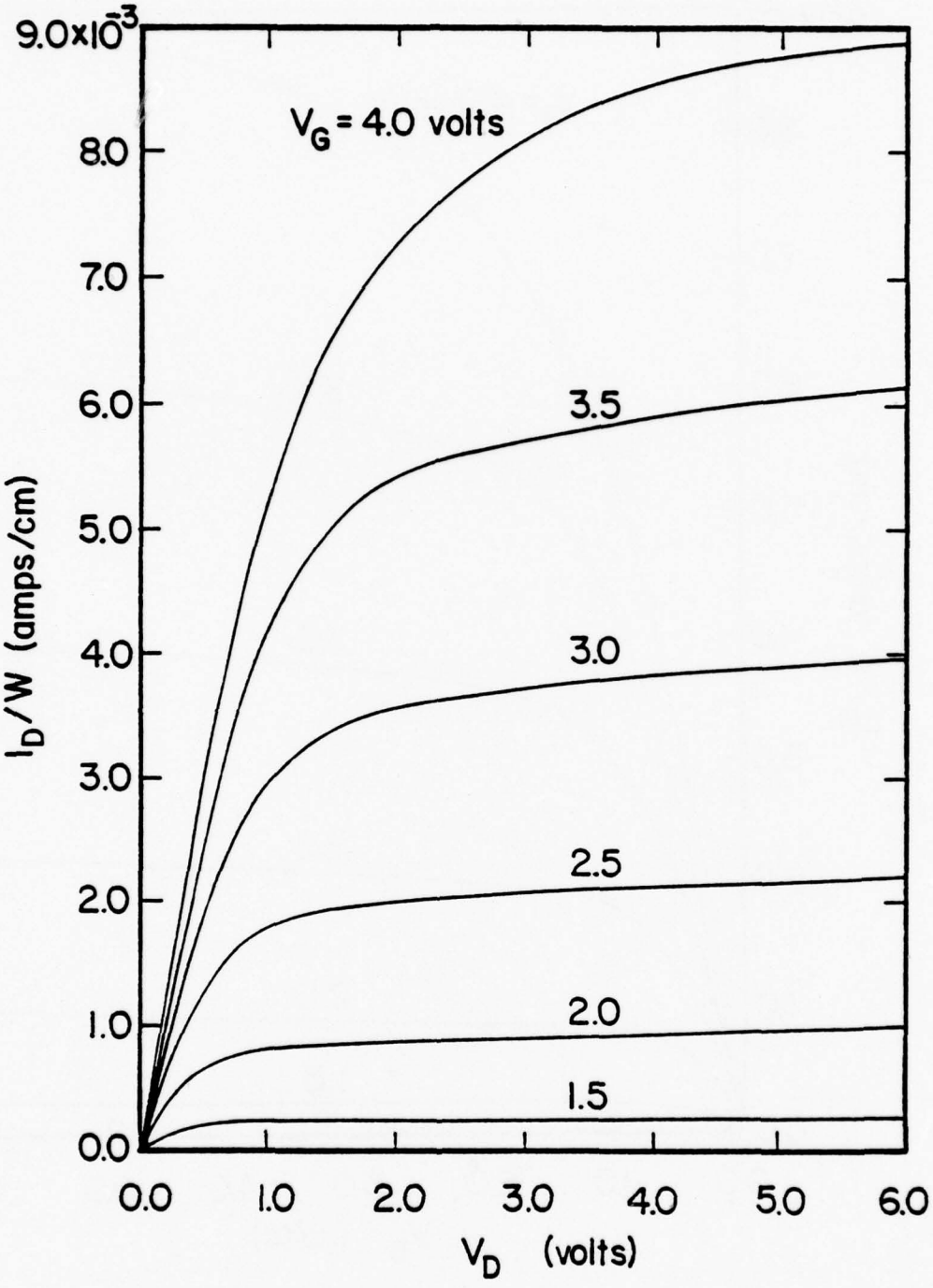


Fig. 1.7. Calculated volt-ampere characteristics of a long-channel MOSFET, assuming a field-dependent carrier mobility. $N_A = 2x10^{15}cm^{-3}$, $t_{ox} = 1000A°$, L = 10μm [55].

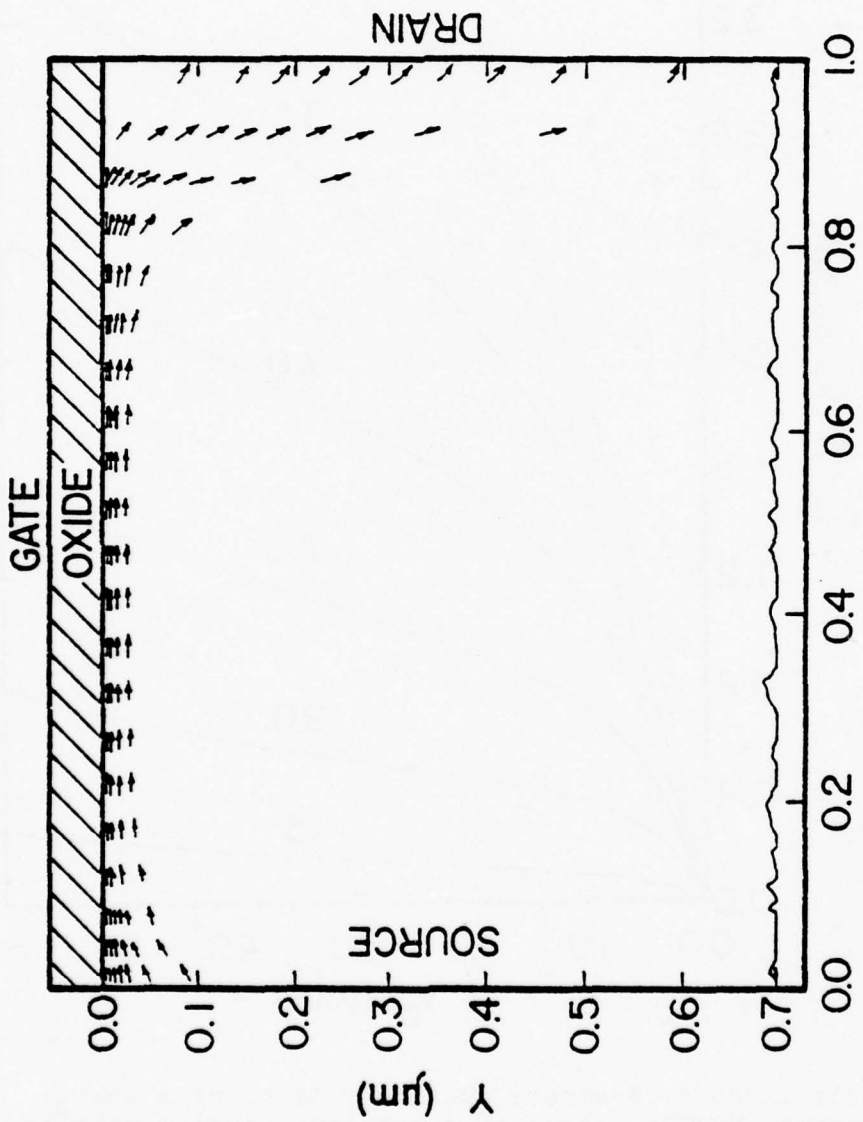


Fig. 1.8. Calculated mobile carrier flux distribution in a MOSFET, assuming a field-dependent carrier mobility. V_G = 3 volts, V_D = 6 volts [55].

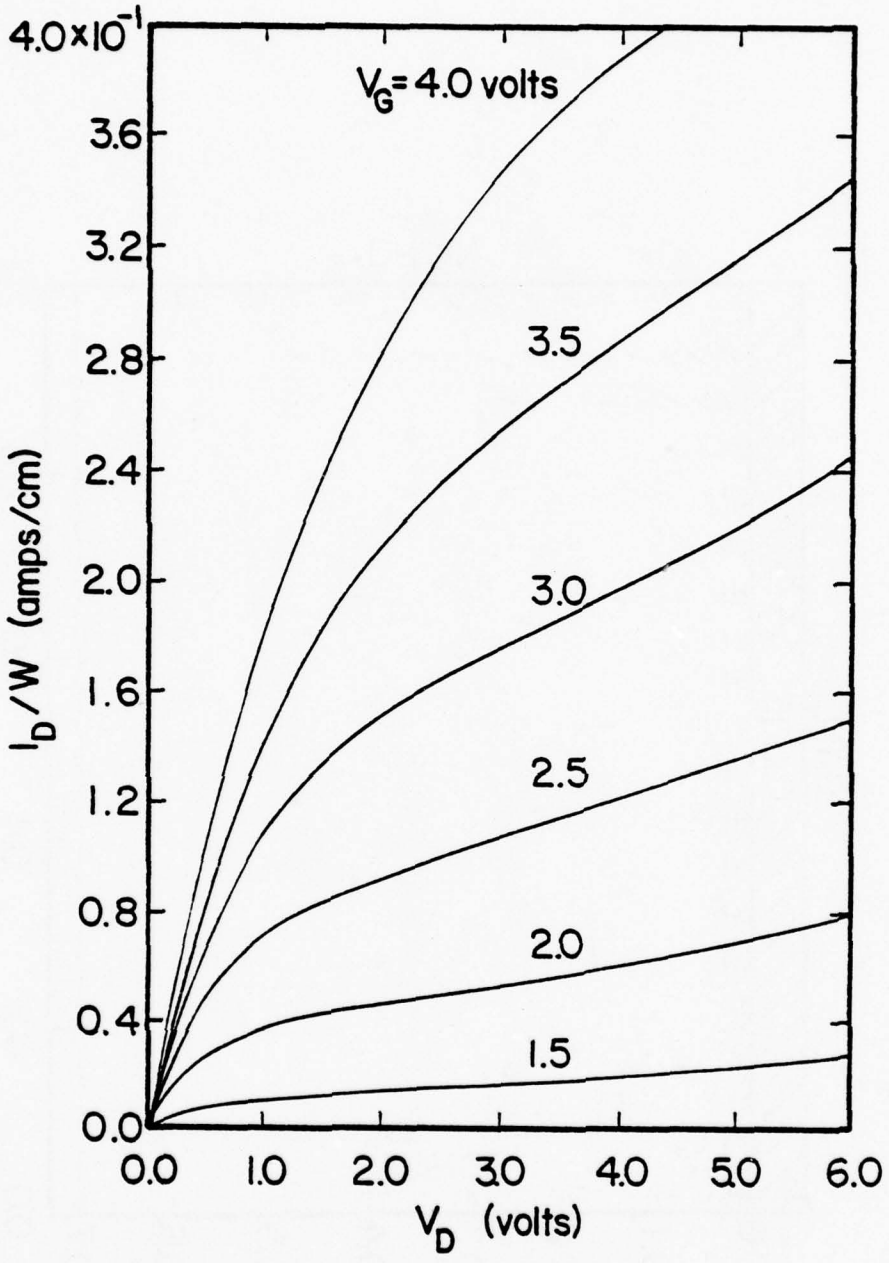


Fig. 1.9. Calculated volt-ampere characteristics of a short-channel MOSFET, assuming a constant carrier mobility. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 288A°$, $L = 1\mu m$ [55].

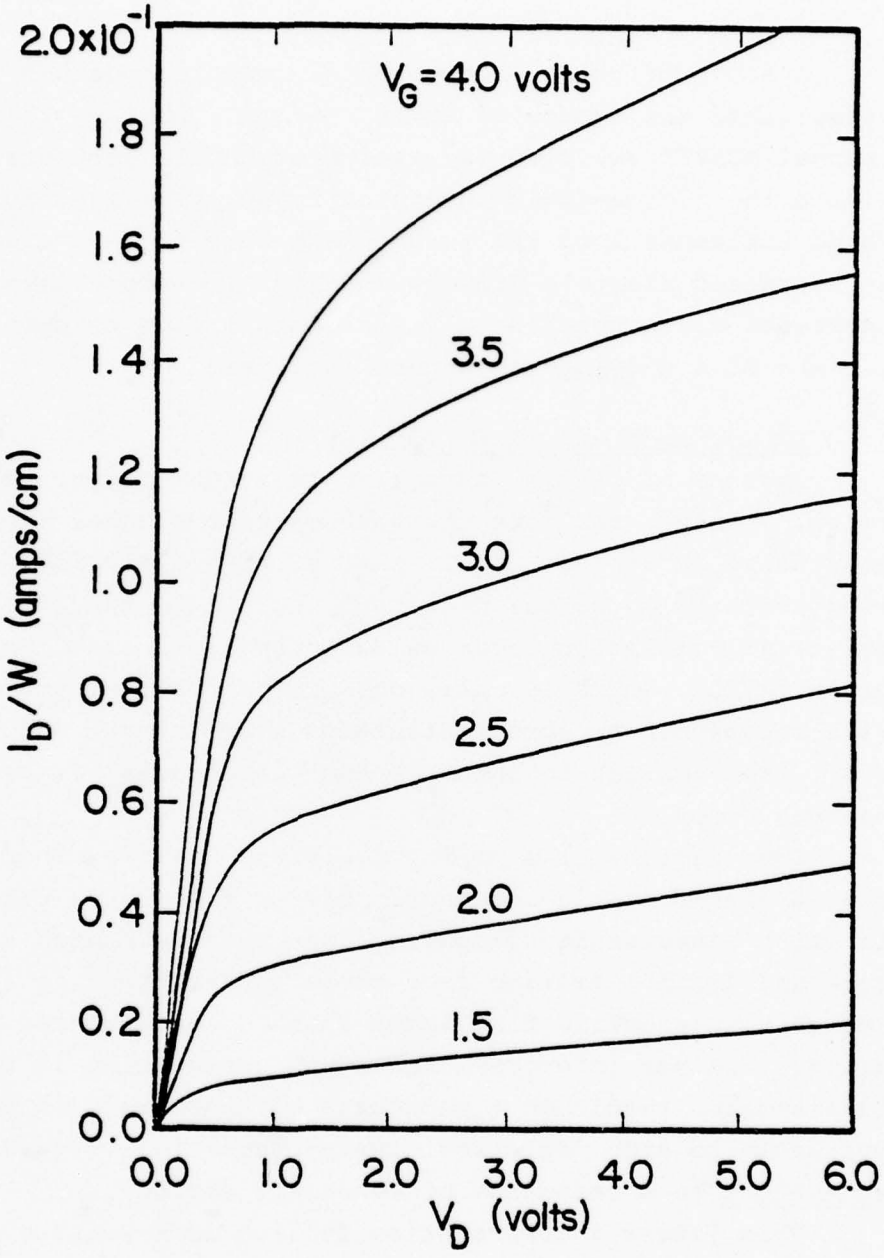


Fig. 1.10. Calculated volt-ampere characteristics of a short-channel MOSFET, assuming a field-dependent carrier mobility. $N_A = 2x10^{16}cm^{-3}$, $t_{ox} = 288A°$, L = 1um [55].

A study of these short-channel calculations has readily established the source of these changes. Namely, in a short-channel MOSFET velocity saturation of mobile carriers is found within the source-drain inversion layer, a region having a profound influence upon the total drain current. As a consequence, an increased electric field within the inversion layer (due to an increased drain voltage) has a minimum influence upon the drain current of a very short-channel structure.

2.4 Depletion Charge Distribution

Regions of carrier depletion in a MOSFET play two important roles: 1) they insulate the source-drain channel from a highly conductive substrate; and 2) they significantly influence the electrostatic potential distribution. Elementary MOSFET theory treats all depletion layer calculations on a one-dimensional basis [4,20]. At best, this one-dimensional treatment may provide a *reasonable approximation for long*-channel structures--however, when applied to short-channel structures, it results in serious errors.

Some regions of a MOSFET depletion layer exhibit two-dimensional electrostatic interactions; for example, a region exists in which unneutralized impurity ions must be shared between the gate and drain. Arising from these interactions is a substantial decrease in electric flux terminating on the source, gate, and drain. Another interpretation of this situation is that these interactions result in a reduction of unneutralized impurity ions available to each depletion layer--hence, this situation is equivalent to a reduction of substrate doping.

This latter interpretation follows from Poisson's equation for these depletion regions,

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = \frac{-qN_A}{\kappa_s \varepsilon_o} \quad , \tag{2.1}$$

where the x-axis and the y-axis are in the direction from source-to-drain and the direction from surface-to-substrate, respectively. Since $\vec{E} = -\text{grad}(V)$, this expression can also be written in the form:

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} = \frac{qN_A}{\kappa_s \varepsilon_o} \quad . \tag{2.2}$$

The specific manner in which ions are shared between the gate and drain is unknown. However, for steady-state operation, ion densities N_x and N_y can be defined such that $N_A = N_x + N_y$, where

$$N_x \equiv \frac{\kappa_s \varepsilon_o}{q} \frac{\partial E_x}{\partial x} \tag{2.3a}$$

$$N_y \equiv \frac{\kappa_s \varepsilon_o}{q} \frac{\partial E_y}{\partial y} \quad . \tag{2.3b}$$

From this interpretation, ion-sharing becomes an important mechanism: ions N_x that are electrostatically associated with the drain junction do not contribute to the gate depletion charge. Similarly, ions N_y that are electrostatically associated with the gate electrode do not contribute to the drain space-charge layer.

Figure 1.11 qualitatively establishes the degree of gate and drain interaction in a MOSFET. In this illustration we show contours of constant source-drain electric field component E_x. From Gauss' law, the magnitude of this electric field component is directly proportional to the integrated substrate electrostatic charge contributing to E_x. This illustration shows that a non-negligible amount of the substrate electrostatic charge contributes to E_x and, therefore, this charge cannot contribute to the gate-induced electric field -- the electric field component upon which elementary MOSFET theory is based [4,20].

To illustrate this ion-sharing mechanism, Fig. 1.12 shows two-dimensional calculations of a MOSFET depletion layer. In these calculations, the depletion layer edge was arbitrarily defined to be the locus at which majority carrier density equals

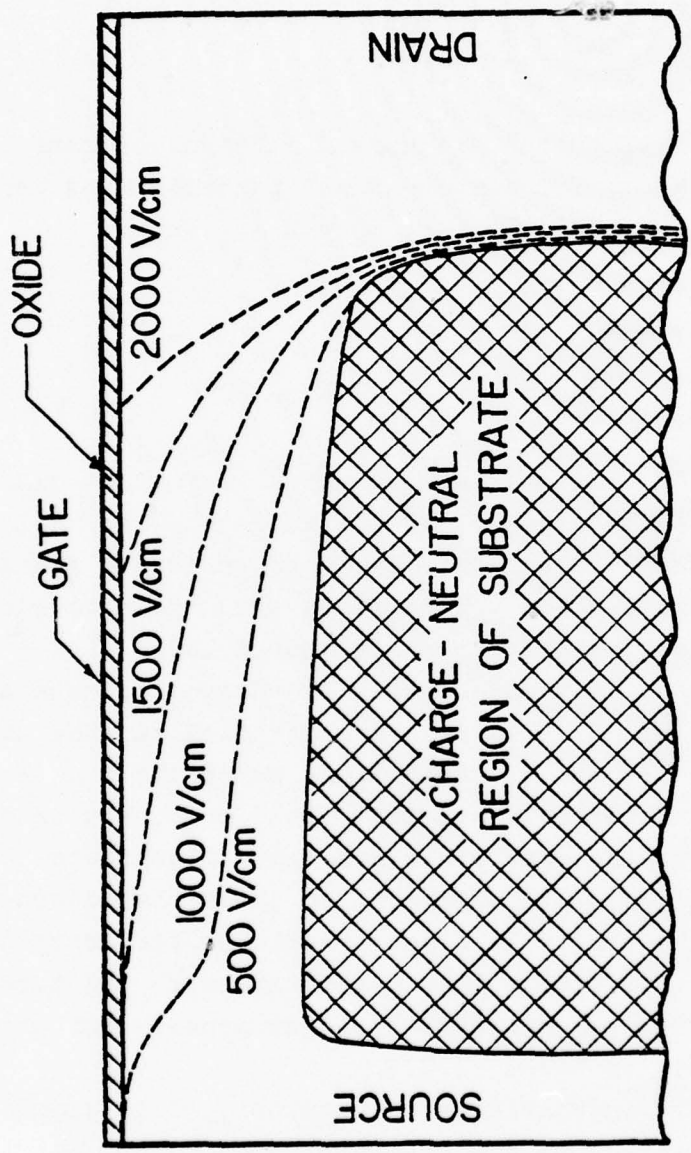


Fig. 1.11. Calculated contours of constant source-drain electric field in a MOSFET. $N_A = 2 \times 10^{15} cm^{-3}$, $t_{ox} = 1000A°$, $L = 10\mu m$ [36].

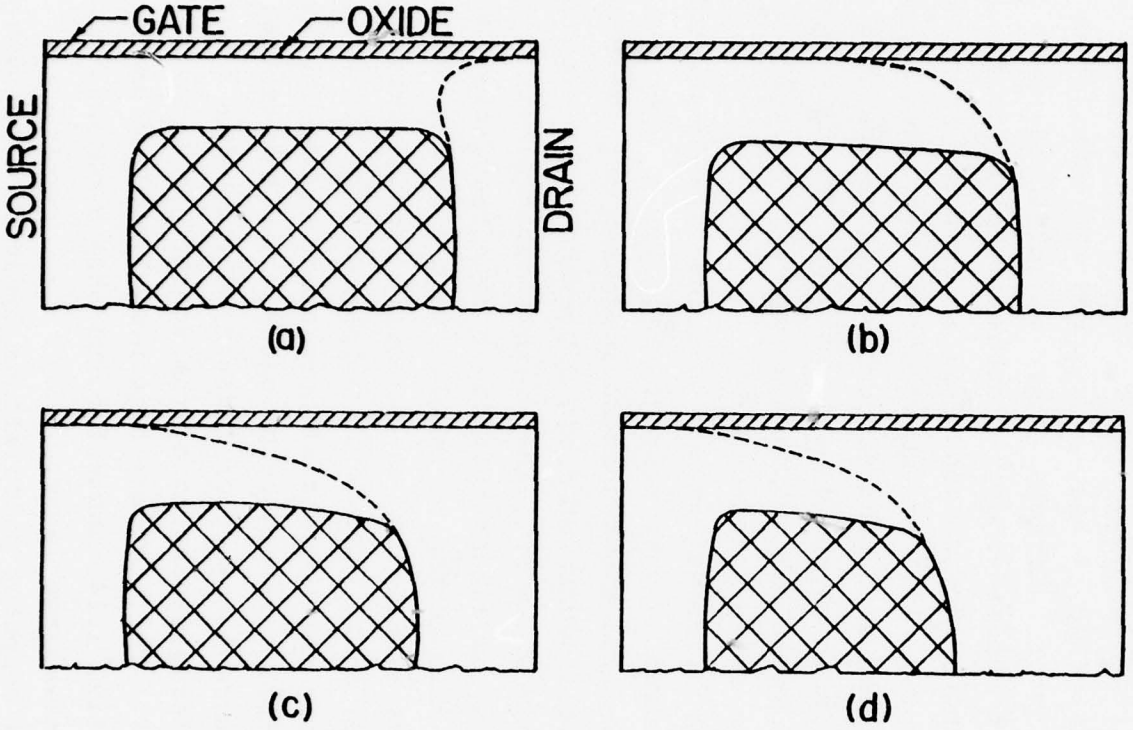


Fig. 1.12. Calculated contour of constant (1000 V/cm) source-drain electric field (broken line) in a MOSFET. $N_A = 2 \times 10^{15} cm^{-3}$, $t_{ox} = 1000A°$, $L = 5\mu m$. $V_G = 2$ volts and a) $V_D = 0$ volts, b) $V_D = 0.5$ volt, c) $V_D = 1$ volt, d) $V_D = 3$ volts. (Cross-hatched areas represent charge-neutral substrate [55].)

half the substrate impurity concentration. These figures show a contour of constant (1000 V/cm) source-drain electric field for a constant applied gate voltage. As the drain voltage is increased (from 0 volts to 3 volts), Fig. 1.12 qualitatively illustrates an expansion of the drain depletion layer towards the source. From Fig. 1.12, this expansion is most pronounced in the vicinity of the oxide-semiconductor interface, where a substantial part of the substrate electrostatic charge produces an electric field perpendicular to the oxide-semiconductor interface. Thus, the expansion shown in Fig. 1.12 is equivalent to a reduced electrostatic charge and, hence, a reduced number of unneutralized impurity ions available to the gate electrode. For this reason, the gate depletion layer behaves like a region of variable impurity ion density, due to drain interaction.

Additional calculations confirm that this "reach-through" mechanism is dependent on gate voltage (Fig. 1.13). For a fixed drain voltage, Fig. 1.13 shows that an increased gate voltage produces an increased degree of drain depletion-layer expansion.

In a short-channel MOSFET, the drain-induced electric field can reach-through to an extent sufficient to forward-bias the source junction. This mechanism is illustrated in Fig. 1.14, which clearly shows a reduction in potential barrier height at the source when applied drain voltage is increased from 0.025 volts to 4.0 volts. Although this potential barrier reduction is relatively small (0.0115 volts), this reduction can yield a substantial increase in source-drain current. Further indication that this situation is equivalent to forward-biasing of the source junction is shown in Figs. 1.15-1.17. For the same sequence of applied drain voltages used in Fig. 1.14, these figures show a notable contraction of the source space-charge layer.

The foregoing computer calculations identify important two-dimensional mechanisms producing threshold modulation in short-channel structures [57]. Specifically, ion-sharing between the gate and drain produces a two-dimensional electric field distribu-

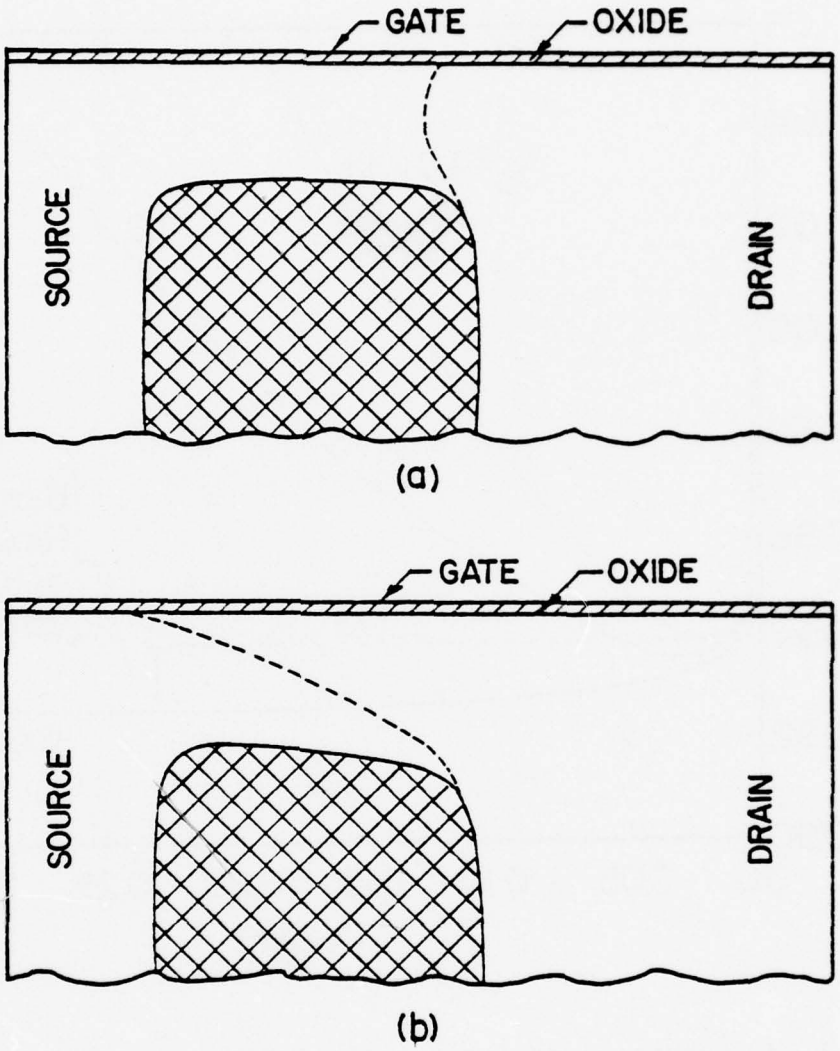


Fig. 1.13. Calculated contour of constant (1000 V/cm) source-drain electric field in a MOSFET. $N_A = 2 \times 10^{15} cm^{-3}$, $t_{ox} = 1000A°$, $L = 5\mu m$. $V_D = 5$ volts and a) $V_G = 1$ volt, b) $V_G = 2$ volts [55].

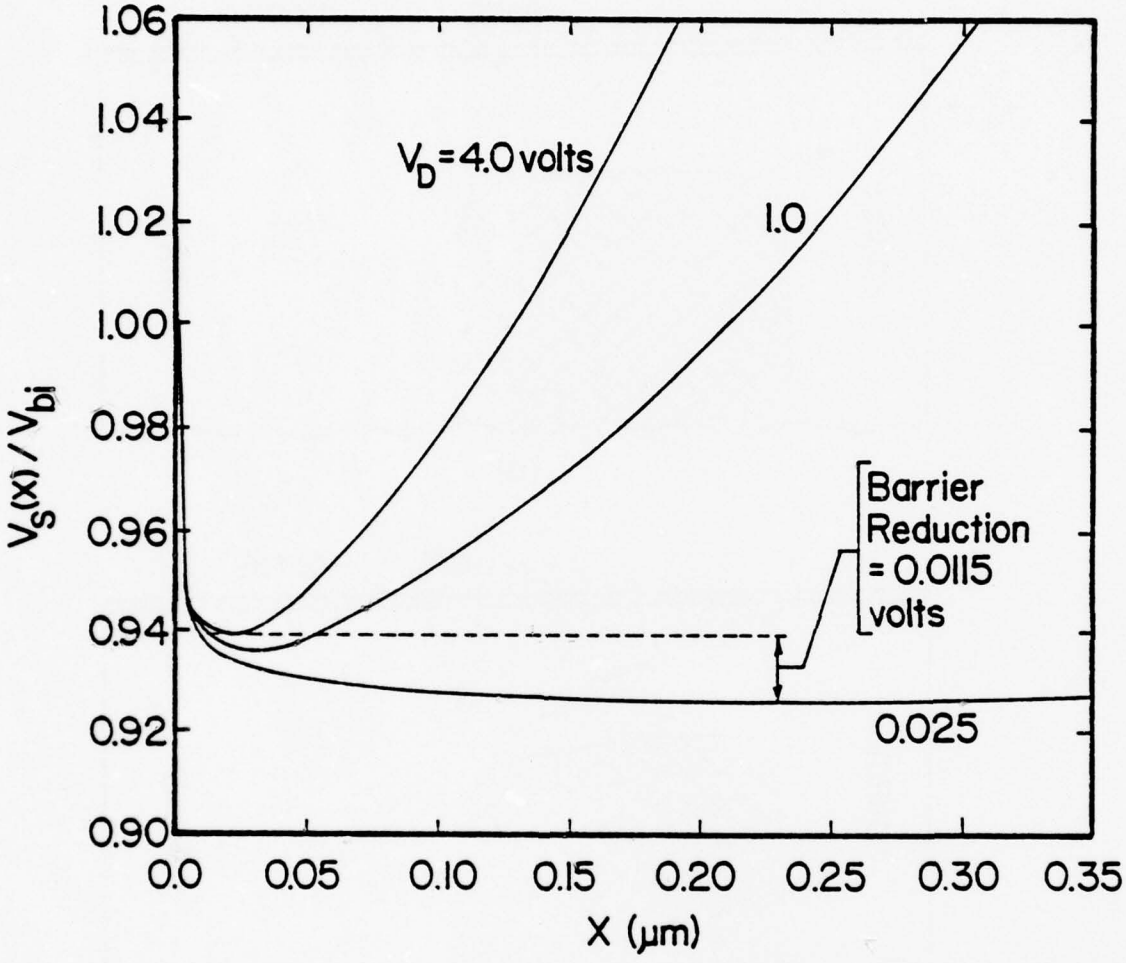


Fig. 1.14. Source-junction potential-barrier reduction (arising from reach-through) in a short-channel MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, t_{ox} = 1000A°, L = 1μm. V_G = 4 volts.

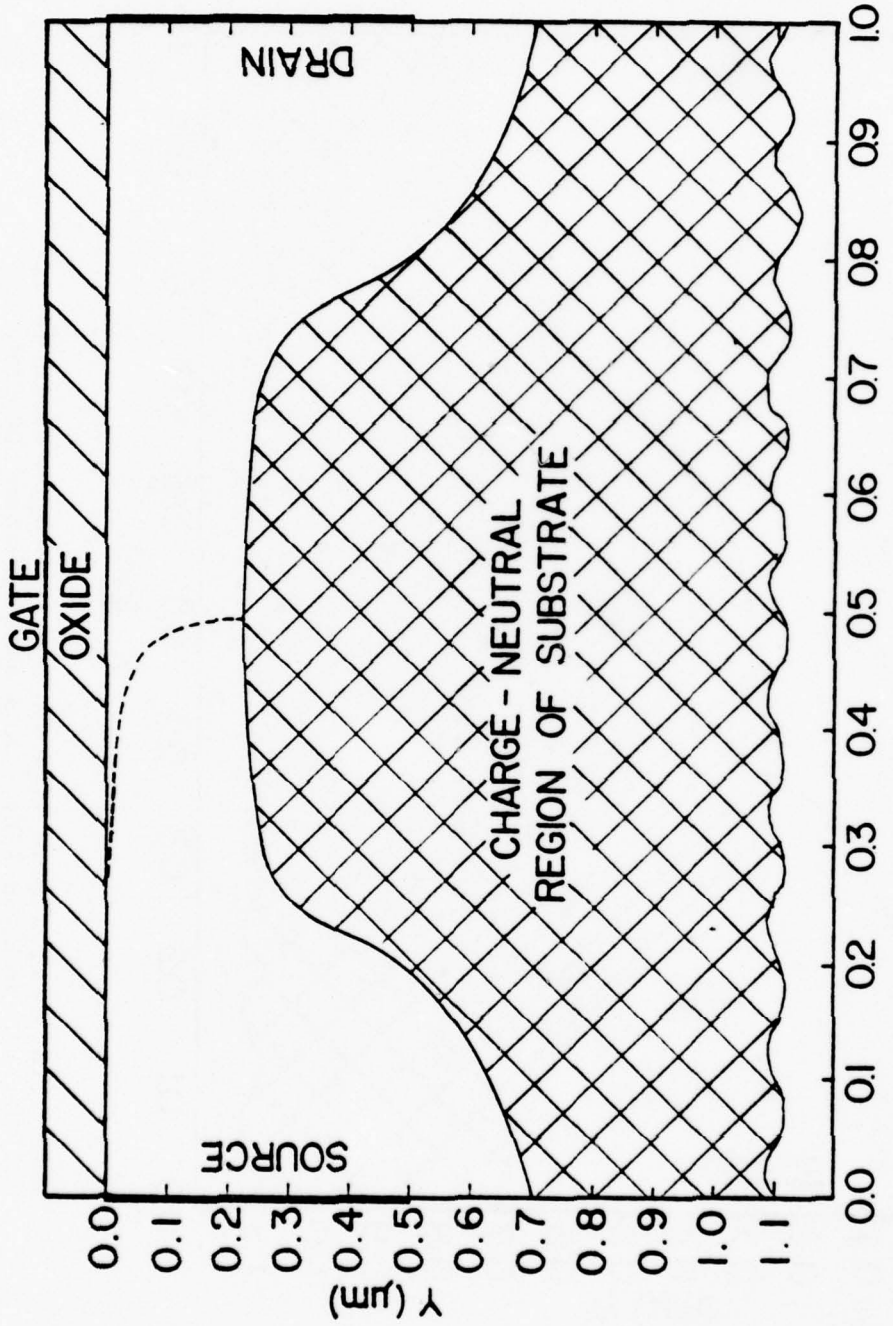


Fig. 1.15. Calculated edge of source space-charge layer (broken line) in a short-channel MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, $L = 1\mu m$. $V_G = 4$ volts, $V_D = 0.025$ volts.

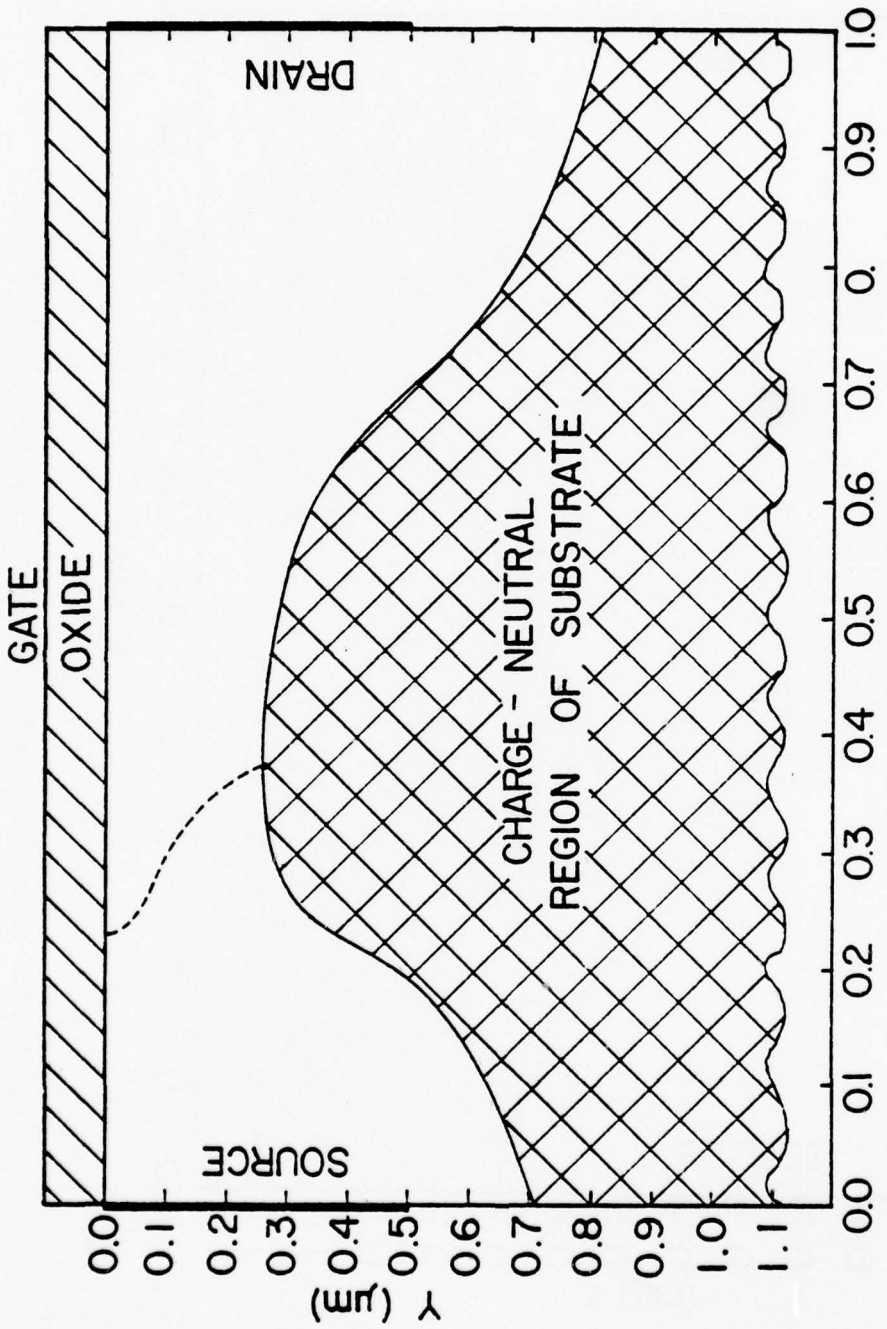


Fig. 1.16. Calculated edge of source space-charge layer (broken line) in a short-channel MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, $L = 1\mu m$. $V_G = 4$ volts, $V_D = 1$ volt.

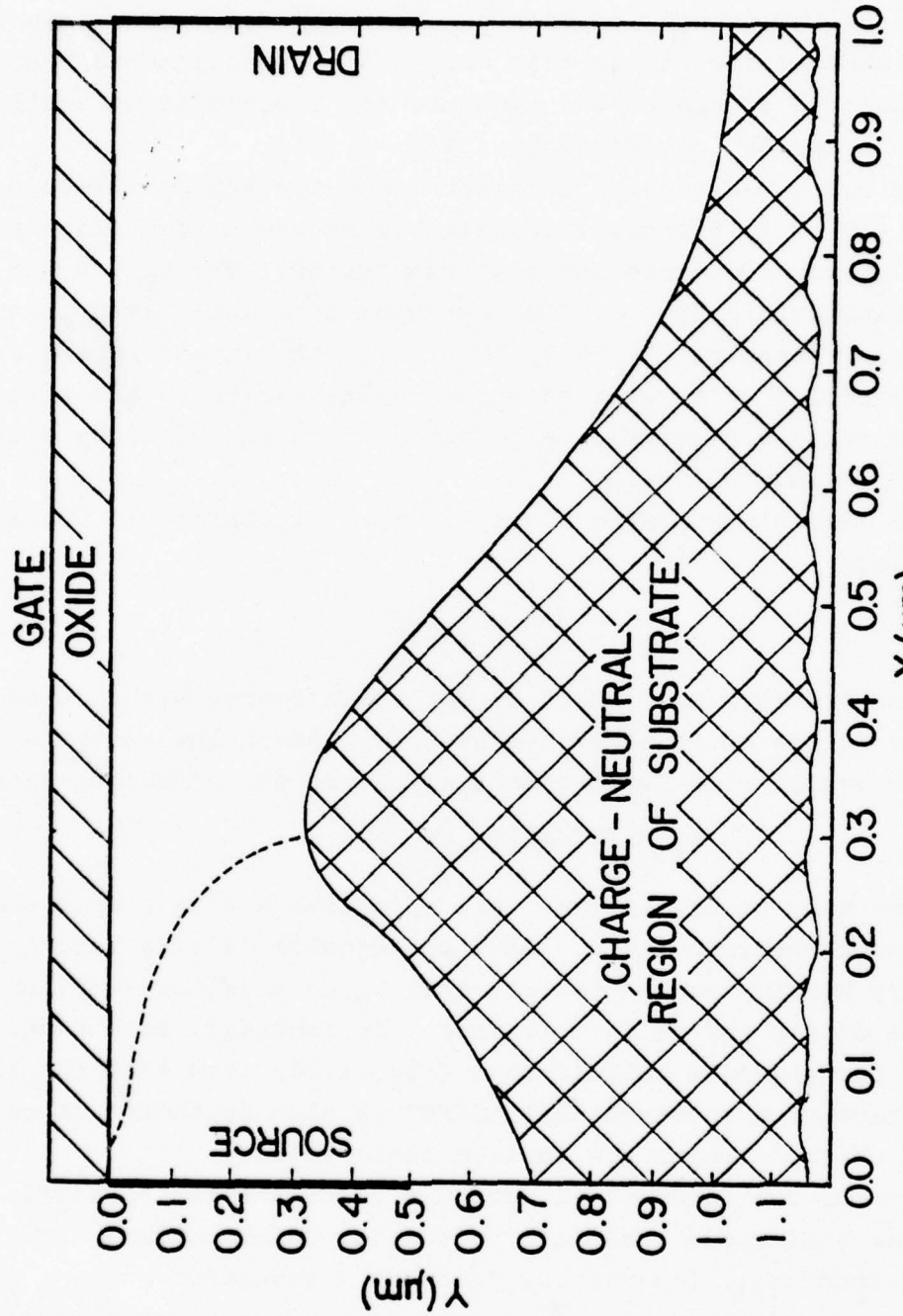


Fig. 1.17. Calculated edge of source space-charge layer (broken line) in a short-channel MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, $L = 1\mu m$. $V_G = 4$ volts, $V_D = 4$ volts.

tion that cannot be adequately described on an elementary one-dimensional basis. Although ion-sharing exists in all MOSFET structures, in short-channel structures this mechanism is encountered near the source junction to the extent required for a significant modification of the threshold voltage (see Fig. 1.18).

From our two-dimensional computations for strong inversion operation, Fig. 1.18 shows calculated gate-threshold voltage of a MOSFET (V_T) for two different bias conditions. For V_D = 0.025 volt, we determined V_T in Fig. 1.18 by extrapolating calculated channel conductance versus gate voltage (V_G), from the linear region of device operation [8]. When this device was biased in the saturation mode (V_D = 5.0 volts), we obtained V_T in Fig. 1.18 by extrapolating $\sqrt{I_D}$ versus V_G [7].

From elementary MOSFET theory [8,9], this threshold voltage has the form

$$V_T = 2\phi_F + \frac{Q_D}{C_{ox}} \quad , \tag{2.4}$$

where Q_D represents the "bulk" or depletion charge within this structure at the onset of strong inversion operation and ϕ_F represents the equilibrium Fermi potential in the semiconductor substrate

$$\phi_F = (kT/q)\log_e(N_A/ni) \quad . \tag{2.5}$$

This elementary theory assumes that Q_D arises entirely from the gate-induced electric field. As a consequence of this assumption, elementary MOSFET theory predicts that V_T is a unique function of substrate doping and oxide thickness. In contrast, both experiments [57] and our rigorous calculations (Fig. 1.18) show that the threshold voltage of a short-channel MOSFET is also dependent on metallurgical channel length and applied drain voltage.

From our two-dimensional computer calculations, physical mechanisms that cause threshold modulation become clear. In short-channel structures, ion-sharing produces a situation where a significant fraction of the depletion charge (Q_D) arises from the drain-

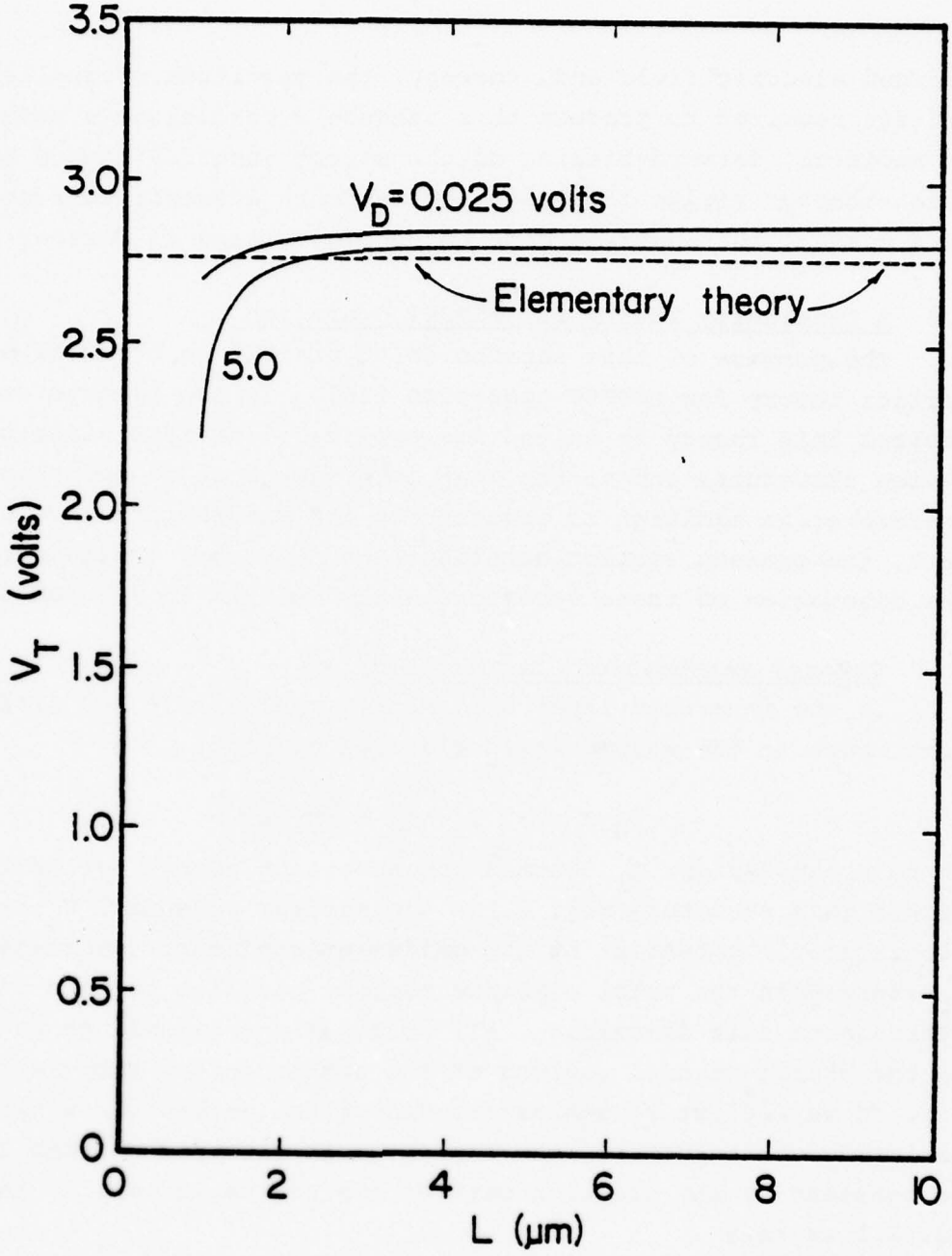


Fig. 1.18. Calculated gate-threshold voltage of a MOSFET, assuming a constant carrier mobility. $N_A = 2 \times 10^{16} cm^{-3}$, t_{ox} = 1000A°, μ = 270 $cm^2/V\text{-sec}$.

induced electric field and, thereby, the magnitude of applied gate voltage required to produce this threshold condition is reduced. In addition, forward-biasing of the source junction caused by reach-through yields increased source-drain electric current and, as a result, the apparent gate threshold voltage is further reduced.

3.0 A Continuity Theory for MOSFET Operation

The purpose of this section is to present an overview of an earlier theory for MOSFET operation [36]. In the present study we revised this theory to extend its applicability to a wider range of device structures and to the weak inversion mode of operation. Therefore, in addition to acquainting the reader with this earlier work, the present section constitutes a necessary preliminary to the discussion of these revisions which follows in section 4.0.

3.1 General Mathematical Development

In the inversion layer of a MOSFET, both drift and diffusion contribute to the source-drain electric current:

$$I_D = -Wu \left[Q_i \frac{dV_s}{dx} - \frac{kT}{q} \frac{dQ_i}{dx} \right] \quad . \qquad (3.1)$$

In this expression, Q_i denotes the inversion charge per unit area within this structure Fig. 1.19, the surface potential V_s is the electrostatic potential at the oxide-semiconductor interface, and I_D represents the total electric current parallel to this interface. (Throughout this discussion, all voltages are assumed to be referenced to the charge neutral regions of the semiconductor substrate.)

If we neglect recombination-generation mechanisms within this semiconductor structure, the source-drain electric current I_D must be constant at any location between the source and drain; thus from Eq. 3.1 we have

$$0 = \frac{dI_D}{dx} = Wu \left[Q_i \frac{d^2V_s}{dx^2} + \frac{dQ_i}{dx} \frac{dV_s}{dx} - \frac{kT}{q} \frac{d^2Q_i}{dx^2} \right] \quad . \qquad (3.2)$$

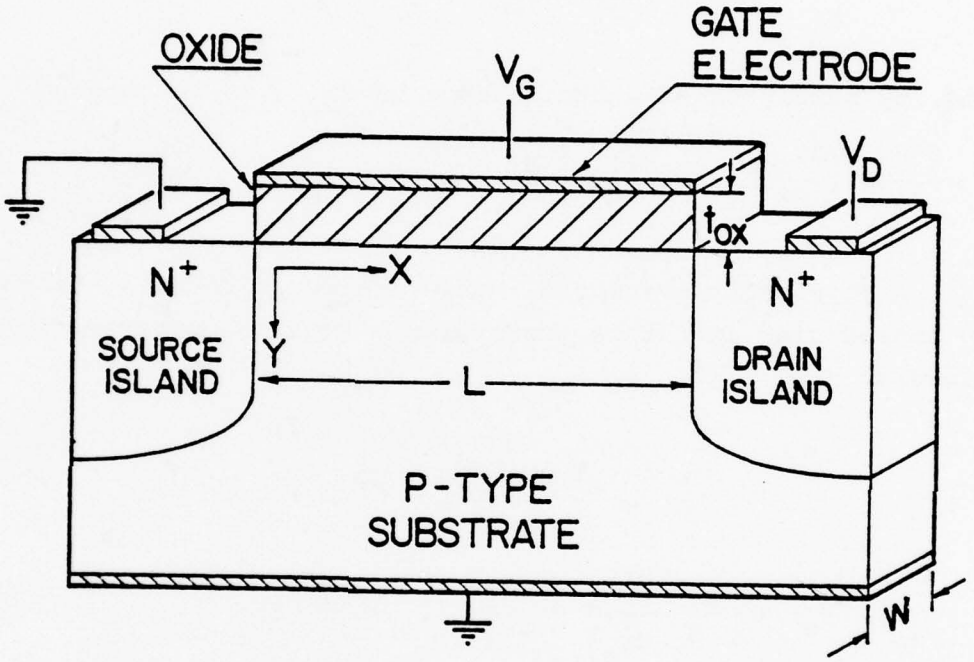


Fig. 1.19. Illustrative model of a MOSFET.

From the chain rule of differentiation,

$$\frac{dQ_i}{dx} = \frac{dQ_i}{dV_s} \cdot \frac{dV_s}{dx} \tag{3.3a}$$

$$\frac{d^2Q_i}{dx^2} = \frac{dQ_i}{dV_s}\frac{d^2V_s}{dx^2} + \frac{d^2Q_i}{dV_s^2}\left(\frac{dV_s}{dx}\right)^2 , \tag{3.3b}$$

and, by making these substitutions in Eq. 3.2, we obtain

$$\frac{d^2V_s}{dx^2}\left(\frac{dV_s}{dx}\right)^{-2} = \left[\frac{kT}{q}\frac{d^2Q_i}{dV_s^2} - \frac{dQ_i}{dV_s}\right]\left[Q_i - \frac{kT}{q}\frac{dQ_i}{dV_s}\right]^{-1} . \tag{3.4}$$

A physically meaningful modification of Eq. 3.4 is realized by introducing into this expression a separation parameter $\lambda(V_s)$, yielding

$$\lambda(V_s) = \frac{d^2V_s}{dx^2}\left(\frac{dV_s}{dx}\right)^{-2} \tag{3.5a}$$

$$\lambda(V_s) = \left[\frac{kT}{q}\frac{d^2Q_i}{dV_s^2} - \frac{dQ_i}{dV_s}\right]\left[Q_i - \frac{kT}{q}\frac{dQ_i}{dV_s}\right]^{-1} . \tag{3.5b}$$

Therefore,

$$\frac{d^2V_s}{dx^2} - \lambda(V_s)\left(\frac{dV_s}{dx}\right)^2 = 0 \tag{3.6a}$$

$$\frac{d^2Q_i}{dV_s^2} + \left[\lambda(V_s) - \frac{q}{kT}\right]\frac{dQ_i}{dV_s} - \left(\frac{q}{kT}\right)\lambda(V_s)Q_i = 0 . \tag{3.6b}$$

From Appendix A, Eqs. 3.6a and 3.6b have the solution equations

$$x = x_o - \frac{1}{E_o} \int_{V_o}^{V_s} \exp[\Omega(\xi)]d\xi \tag{3.7a}$$

$$Q_i(V_s) = Q_{io} e^{qV_s/kT} \left\{ e^{-qV_o/kT} - \left[\frac{q}{kT} - \frac{1}{Q_{io}} \left(\frac{dQ_i}{dV_s} \right)_{V_o} \right] \right.$$
$$\left. \cdot \int_{V_o}^{V_s} \exp\left[- \frac{q\xi}{kT} + \Omega(\xi) \right] d\xi \right\}, \tag{3.7b}$$

where

$$\Omega(\xi) = - \int_{V_o}^{\xi} \lambda(\eta)d\eta \quad . \tag{3.8}$$

These solution equations provide the foundation for our calculation of the volt-ampere characteristics for a MOSFET.

Through algebraic manipulations, it can readily be proved that if, indeed, a solution is found for $\lambda(V_s)$, the magnitudes of $V_s(x)$ and $Q_i(V_s)$ will always yield electric current continuity between the source and drain. This proof is given in Appendix B.

3.2 A Physical Interpretation of the Separation Parameter, $\lambda(V_s)$

The physical significance of the separation parameter, $\lambda(V_s)$, becomes clear from the derivations given in Appendix B. Therein, it is shown that the source-drain electric current is constant at any location within this semiconductor device if Eq. 3.6 is satisfied. Further, from Appendix B we have

$$\lambda(V_s) = \frac{I_{diff}}{I_D} \left[\frac{q}{kT} - \frac{d}{dV_s} \log_e \left(\frac{dQ_i/dx}{dV_s/dx} \right) \right] \quad . \tag{3.9}$$

From Eq. 3.9 , it is evident that the separation parameter $\lambda(V_s)$ produces a modification of the proportions of drift current and diffusion current necessary to yield constant source-drain electric current at all locations within this semiconductor device.

From a two-dimensional computer solution for this problem, it was found that $\lambda(V_s)$ always attains a magnitude of q/kT in the limit of weak inversion [36]. From Eq. 3.9 , this limit implies that all source-drain electric current is attributable to diffusion. Furthermore, from this computer calculation it was found that $\lambda(V_s)$ can be adequately approximated by [36]

$$\lambda(V_s) = \frac{\lambda_o}{1-A(V_s-V_o)} \quad , \tag{3.10}$$

where $\lambda_o \equiv \lambda(V_o)$. Assuming $\lambda(V_s)$ approaches this same limit at the point of channel termination (where $V_s=V_G$) we have

$$A = \frac{1-(kT/q)\lambda_o}{(V_G-V_o)} \quad . \tag{3.11}$$

Therefore, substituting Eq. 3.11 into Eq. 3.10, we obtain

$$\lambda(V_s) = \frac{(V_G-V_o)\lambda_o}{(V_G-V_s) + (kT/q)(V_s-V_o)\lambda_o} \quad . \tag{3.12}$$

3.3 The Volt-Ampere Characteristics of a MOSFET

From the implicit solution equations for $V_s(x)$ and $Q_i(V_s)$ (Eq. 3.7a and Eq. 3.7b , respectively) inconjunction with the approximate relation for $\lambda(V_s)$, Eq. 3.12, we can readily obtain an expression for the volt-ampere characteristics of a MOSFET. From Eq. 3.7a we obtain

$$V_s = V_o + \frac{1}{A}\left\{1 - \left[1 - (\lambda_o + A)E_o x\right]^{\frac{A}{\lambda_o+A}}\right\} \quad , \tag{3.13}$$

where E_o is the magnitude of source-drain electric field at the source end of this structure:

$$E_o = -\left(\frac{dV_s}{dx}\right)_{x=0} \quad . \tag{3.14}$$

In a similar manner, from Eq. 3.7b we obtain

$$Q_i(V_s) = Q_{io} e^{q(V_s-V_o)/kT} \left\{ 1 - \left[1 - \frac{kT}{q} \frac{1}{Q_{io}} \left(\frac{dQ_i}{dV_s}\right)_{V_o} \right] \right.$$

$$\left. \cdot \; Z_o^{\lambda_o/A} \exp(Z_o) \left[\Gamma\left(1 + \frac{\lambda_o}{A}, Z_o\right) - \Gamma\left(1 + \frac{\lambda_o}{A}, Z\right) \right] \right\} , \tag{3.15}$$

where

$$Z_o = -\frac{q}{kT} \cdot \frac{1}{A} \tag{3.16a}$$

$$Z = Z_o \left[1 - A(V_s - V_o) \right] \quad . \tag{3.16b}$$

For the range of variables encountered in this boundary value problem, it was proposed that Eq. 3.15 could be adequately approximated by the relation [36]

$$Q_i(V_s) \approx Q_{io} \left[1 - A(V_s - V_o) \right]^{\lambda_o/A} \quad . \tag{3.17}$$

This system of equations has been constrained to yield a divergence free electric current. Therefore, the magnitude of this current (from Eq. 3.1),

$$I_D = -W\mu Q_i \frac{dV_s}{dx} \left[1 - \frac{kT}{q} \frac{1}{Q_i} \frac{dQ_i}{dV_s} \right] , \tag{3.18}$$

can be calculated at any location along the source-drain channel. Selecting for this calculation the source end of this structure, Eq. 3.18 has the form

$$I_D = -W\mu Q_{io} \left(\frac{dV_s}{dx}\right)_{x=0} \left[1 - \frac{kT}{q} \frac{1}{Q_{io}} \left(\frac{dQ_i}{dV_s}\right)_{V_o}\right], \qquad (3.19)$$

where $V_o \equiv V_s(0)$. Substituting Eq. 3.14 into Eq. 3.15 yields

$$I_D = W\mu Q_{io} E_o \left[1 - \frac{kT}{q} \frac{1}{Q_{io}} \left(\frac{dQ_i}{dV_s}\right)_{V_o}\right] . \qquad (3.20)$$

From Eq. 3.17 we obtain the relation

$$\left(\frac{dQ_i}{dV_s}\right)_{V_o} = -\lambda_o Q_{io} \quad . \qquad (3.21)$$

Further, from Eq. 3.13 we have, upon requiring that $V_s=V_D$ when x=L,

$$E_o = \frac{1}{(\lambda_o+A)L} \left\{1 - \left[1 - A(V_D-V_o)\right]^{\frac{\lambda_o+A}{A}}\right\} . \qquad (3.22)$$

Thus, after substituting Eqs. 3.21 and 3.22 into Eq. 3.20, we have an expression for the volt-ampere characteristics of a MOSFET

$$I_D = \frac{W\mu Q_{io}}{(\lambda_o+A)L} \left[1 + \frac{kT}{q} \lambda_o\right] \left\{1 - \left[A(V_D-V_o)\right]^{\frac{\lambda_o+A}{A}}\right\} . \qquad (3.23)$$

It should be noted that Eq. 3.23 contains only two unknown parameters Q_{io} and λ_o. Both of these parameters arise at the source end of this structure where two-dimensional mechanisms are minimal; thus, these parameters can be evaluated on a one-dimensional basis.

Before determining these parameters, we first consider an application of Eq. 3.23 to the weak inversion mode of operation. From Eq. 3.11, as $\lambda_o \to q/kT$ in weak inversion the parameter A becomes small and, therefore, Eq. 3.23 has the limit

$$\lim_{\lambda_o \to q/kT} (I_D) \approx \frac{2WDQ_{io}}{L} \left[1 - e^{-q(V_D-V_o)/kT}\right] . \qquad (3.24)$$

Clearly, Eq. 3.24 shows an exponential saturation of the source-drain electric current with an increase of drain voltage. Thus, from the concept of electric current continuity within the inversion layer of a MOSFET, we have an expression for the current saturation in weak inversion that qualitatively agrees with both experimental observation [30,31] and other recent theoretical studies [30,31,37].

3.4 The Saturation Mode of Operation

Equation 3.23 and its weak inversion limit, Eq. 3.24, are applicable only when channel length L is a known quantity. This situation exists for triode mode operation. Assuming that electric current saturation arises when we have channel termination (i.e., when $V_D=V_G$), the source-drain inversion layer maintains a total voltage of V_G-V_o across its length in this mode of operation; however, the length of this inversion layer becomes a function of drain voltage and decreases with an increase of V_D [36]. Thus, in electric current saturation Eq. 3.23 has the form

$$I_D = \frac{W\mu Q_{io}}{(\lambda_o+A)L_c(V_D)}\left[1+\frac{kT}{q}\lambda_o\right]\left\{1-\left[1-A(V_G-V_o)\right]^{\frac{\lambda_o+A}{A}}\right\}, \quad (3.25)$$

where $L_c(V_D)$ is a voltage-dependent channel length.

Assuming that the substrate region of this MOSFET has a homogeneous impurity atom density, we apply to this calculation the depletion layer theory of abrupt asymmetrical p-n junctions. Thus, it was proposed that the drain junction space-charge layer extension into the substrate has the approximate form

$$W_D^2 = \frac{2\kappa_s\varepsilon_o}{qN_A}\,V_D \quad , \quad (3.26)$$

where N_A represents the substrate impurity ion density. Further, from this depletion layer theory, we have a voltage drop of V_D-V_G across a distance of ΔL, where

$$\Delta L = W_D - \left[W_D^2 - \frac{2\kappa_s \varepsilon_o}{qN_A} (V_D - V_G) \right]^{1/2} . \qquad (3.27)$$

Thus, if we subtract this distance (ΔL) from the metallurgical source-drain channel length (L), we obtain

$$L_c(V_D) = L - \left\{ W_D - \left[W_D^2 - \frac{2\kappa_s \varepsilon_o}{qN_A} (V_D - V_G) \right]^{1/2} \right\} . \qquad (3.28)$$

In current saturation, $L_c(V_D)$ represents the electrical channel length across which we have a total voltage of $(V_G - V_o)$. Thus, Eq. 3.25 in conjunction with Eq. 3.28 establishes the drain current when $V_D > V_G$.

3.5 Calculation of Q_{io} in a MOSFET

Before Eq. 3.7b and, hence, Eq. 3.23 can be used for a quantitative evaluation of the volt-ampere characteristics of a MOSFET, it is necessary to calculate a magnitude of Q_{io} -- the value of Q_i at the source end of this semiconductor structure (where $V_s = V_o$). From rigorous two-dimensional computer calculations of MOSFET operation, it was known that electrostatic interactions between the gate and drain are at a minimum near the source junction [36]. Therefore, it was presumed that a one-dimensional calculation of inversion charge is applicable in this region.

From elementary MOSFET theory [1], the total electrostatic charge (Q_T) within this semiconductor structure (i.e., both inversion charge and depletion charge) is quantitatively determined by the gate-induced electric field at this interface,

$$\kappa_s \varepsilon_o E_y(x,0) = -Q_T . \qquad (3.29)$$

For simplicity, in Eq. 3.29 we neglect electrostatic charges that may reside at the semiconductor surface, i.e., surface states.

For additional simplification, we now also neglect: 1) the difference in work function that may exist between the gate electrode and the semiconductor material, and 2) electrostatic charges that may reside within the gate oxide, i.e., oxide states. Based upon these simplifying assumptions continuity of electric field at the oxide-semiconductor interface requires that

$$E_y(x,0) = \frac{\kappa_i}{\kappa_s}\left(\frac{V_G - V_s}{t_{ox}}\right) \quad . \tag{3.30}$$

In Eq. 3.30, V_G is the applied gate voltage and t_{ox} represents the gate oxide thickness. Substituting Eq. 3.30 into Eq. 3.29 results in the following expression for the total gate-induced electrostatic charge:

$$Q_T = \frac{-\kappa_i \varepsilon_o}{t_{ox}}(V_G - V_s) \quad . \tag{3.31}$$

In one spatial dimension, the surface potential (V_s) within this structure arises in conjunction with an electrostatic charge (Q_T) due to unneutralized impurity ions and mobile electrons. From a one-dimensional solution of Poisson's equation in the semiconductor substrate, mobile electrons accumulate in an extremely thin layer at the oxide-semiconductor interface [61]. When the applied gate voltage is sufficiently positive, the density of electrons near the semiconductor surface can exceed the density of holes (majority carriers in the bulk), thereby forming an inversion layer. Thus, the mobile electrons within such a layer constitute an inversion charge (Q_i). Elementary MOSFET theory assumed that the corresponding distribution of unneutralized impurity ions is basically equivalent to the depletion charge (Q_D) in a one-sided abrupt p-n junction [62] and, therefore,

---

[61] C.G.B. Garrett and W. H. Brattain, *Physical Review*, 99, 376 (1955).

[62] W. Shockley, *Bell System Technical Journal*, 28, 435 (1959).

$$Q_D = -\sqrt{2\kappa_s \varepsilon_o q N_A V_s} \quad . \tag{3.32}$$

Fig. 1.20 qualitatively illustrates such one-dimensional distributions of Q_i and Q_D in a MOSFET.

The total electrostatic charge (Q_T) within this structure is comprised of the inversion charge (Q_i) and the depletion charge (Q_D),

$$Q_T = Q_i + Q_D \quad . \tag{3.33}$$

Therefore, by substituting Eqs. 3.31 and 3.32 into Eq. 3.33 and then solving for Q_i, we obtain

$$Q_i(V_s) = -C_{ox}(V_G - V_s) + \sqrt{2\kappa_s \varepsilon_o q N_A V_s} \quad , \tag{3.34}$$

where

$$C_{ox} \equiv \frac{\kappa_i \varepsilon_o}{t_{ox}} \tag{3.35}$$

represents the static capacitance of the gate oxide.

To obtain source-drain electric current continuity in a MOSFET, it was shown that $Q_i(V_s)$ must satisfy a relation (Eq. 3.7b) which differs significantly from this one-dimensional form of elementary MOSFET theory (Eq. 3.34). However, from the foregoing discussion, it was presumed Eq. 3.34 provides an adequate approximation for $Q_i(V_s)$ near the source end of this structure and, therefore,

$$Q_{io} = Q_i(V_o) = -C_{ox}(V_G - V_o) + \sqrt{2\kappa_s \varepsilon_o q N_A V_o} \quad . \tag{3.36}$$

3.6 Calculation of λ_o in a MOSFET

In this continuity theory of MOSFET operation, source-drain electric current is given by Eq. 3.23. This equation contains two unknown parameters (Q_{io} and λ_o) which must be evaluated in order to calculate this current. In the preceding section, we found an expression for Q_{io} (Eq. 3.36), by evaluating $Q_i(V_s)$ (from elementary MOSFET theory) at $V_s = V_o$. Herein we will derive an expression for

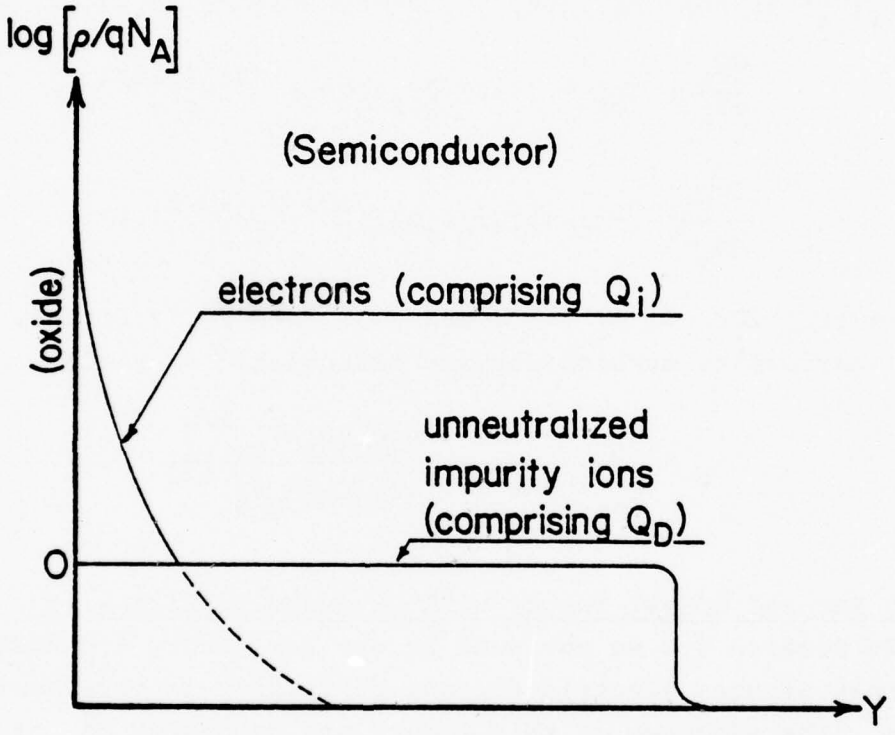


Fig. 1.20. Qualitative illustration showing one-dimensional distributions of Q_i and Q_D in a MOSFET.

λ_o, the remaining parameter needed for calculating the source-drain electric current.

To obtain an expression for λ_o, we will utilize a general relation for $\lambda(V_s)$ given by Eq. 3.5b. Since $\lambda_o \equiv \lambda(V_o)$ represents the magnitude of $\lambda(V_s)$ at the source end of this semiconductor structure, it was proposed that Eq. 3.34 could be used to approximate $Q_i(V_s)$ in this vicinity. Differentiating Eq. 3.34, we obtain

$$\frac{dQ_i}{dV_s} = C_{ox} + (1/2)(2\kappa_s\varepsilon_o qN_A)^{1/2}(V_s)^{-1/2} \qquad (3.37)$$

$$\frac{d^2Q_i}{dV_s^2} = -(1/4)(2\kappa_s\varepsilon_o qN_A)^{1/2}(V_s)^{-3/2} \quad . \qquad (3.38)$$

Substituting Eqs. 3.34, 3.37, and 3.38 into Eq. 3.5b and, thereafter, performing straightforward manipulations, yields

$$\lambda_o = \frac{C_{ox} + (1/2)(2\kappa_s\varepsilon_o qN_A/V_o)^{1/2}}{C_{ox}(V_G-V_o) - (2\kappa_s\varepsilon_o qN_A V_o)^{1/2}} \quad . \qquad (3.39)$$

4.0 A Revised MOSFET Theory with Extended Applicability

In Section 3.0 we reviewed an earlier theory for MOSFET operation that assures electric current continuity in this semiconductor device. The adequacy of this theory was demonstrated for long-channel structures operating in the strong inversion mode [36]. However, although implicitly containing the correct physics, this earlier theory failed to yield satisfactory results for the weak inversion mode and/or for short-channel structures. The main direction of the present research has been to alleviate these difficulties and, thereby, develop a MOSFET theory with wider applicability. Herein we propose a revised continuity theory for MOS transistor operation. In the following section, we demonstrate the broad range of device structures and bias conditions for which this revised theory applies.

4.1 Subthreshold Correction for Q_{io} in a MOSFET

To obtain source-drain electric current continuity in a MOSFET, it has been shown that the inversion charge $Q_i(V_s)$ must satisfy a relation differing significantly from that given by a one-dimensional form of Gauss' law. This relation is given by Eq. 3.7b. However, before Eq. 3.7b and, hence, Eq. 3.23 can be used for a quantitative evaluation of the volt-ampere characteristics of a MOSFET, it is necessary to calculate a magnitude for Q_{io} -- the value of Q_i at the source end of this semiconductor structure (where $V_s=V_o$).

From previous discussions, it was shown that electrostatic interactions between the gate and drain are at a minimum near the source junction. For this reason, it was presumed that a one-dimensional form of Gauss' law is applicable in this region of the structure. Furthermore, borrowing from elementary MOSFET theory, an inherently simple approximate form of this law (Eq. 3.36) was adopted to calculate Q_{io}. In the present research, we found Eq. 3.36 to be inapplicable to the weak inversion mode of MOSFET operation. Specifically, for weak inversion (where $V_o \lesssim 2\phi_F$) Eq. 3.36 yields totally incorrect values for Q_{io}.

For purposes of illustration, Fig. 1.21 shows a comparison between the calculated values of Q_{io}, using Eq. 3.36 suitably modified to eliminate the gate voltage V_G, and using a rigorous one-dimensional solution [9]. In Eq. 3.36 we replace the term representing the total electrostatic charge by an exact one-dimensional solution for Poisson's equation:

$$Q_T = -C_{ox}(V_G-V_o) = -\sqrt{2}\left(\frac{\kappa_s\varepsilon_o}{L_D}\right)\left(\frac{kT}{q}\right)$$

$$\cdot\left\{e^{-\beta V_o} + \beta V_o - 1 + \left(\frac{n_i}{N_A}\right)^2\left[e^{\beta V_o} - 1\right]\right\}^{1/2}, \qquad (4.1)$$

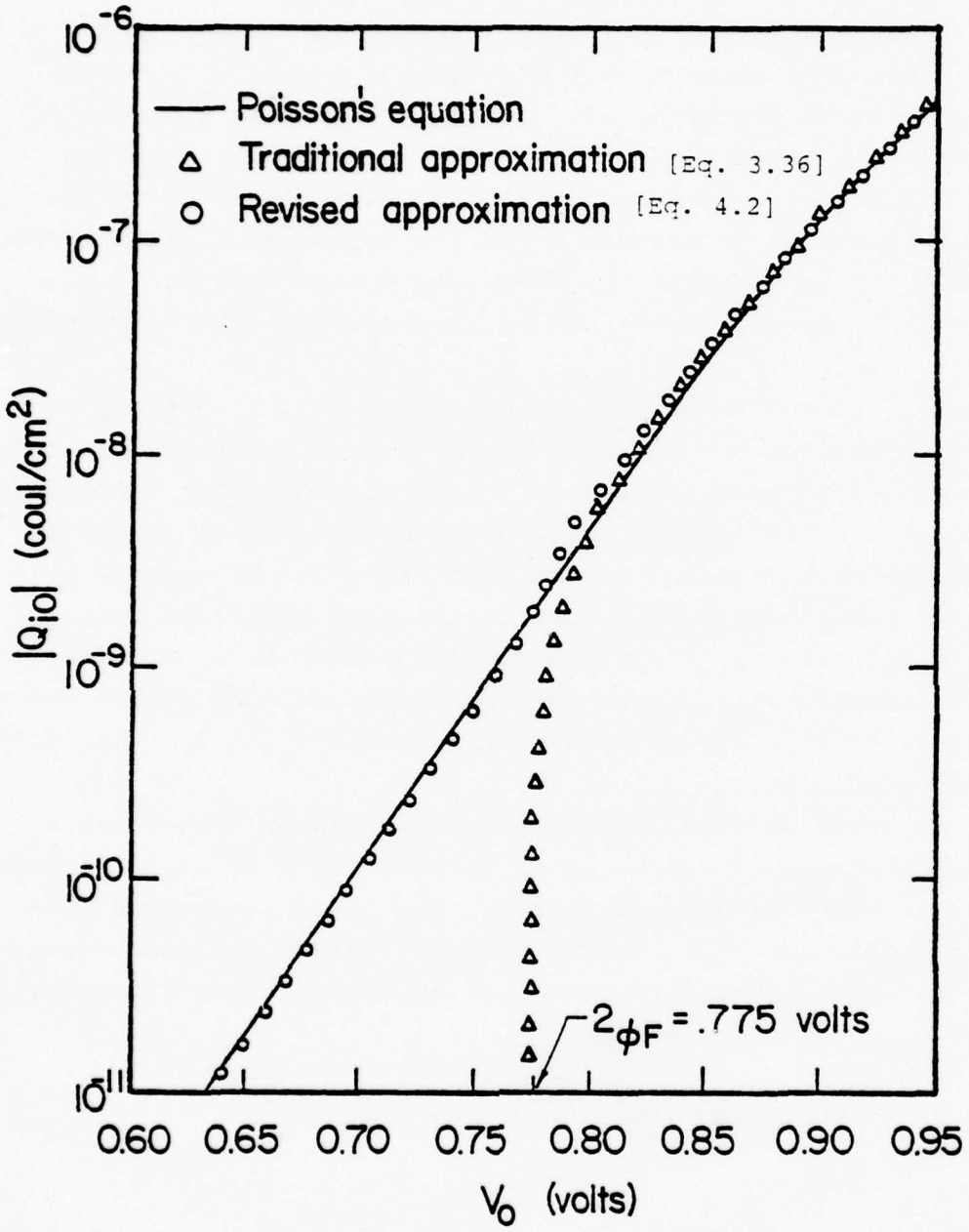


Fig. 1.21. One-dimensional calculations of inversion charge (Q_{io}) at the source end of a MOSFET. $N_A = 5 \times 10^{16} cm^{-3}$.

where L_D represents the extrinsic Debye length in this semiconductor material [9]. Therefore, Eq. 3.36 has the form

$$Q_{io} = -\sqrt{2}\left(\frac{\kappa_s \varepsilon_o}{L_D}\right)\left(\frac{kT}{q}\right)\left\{e^{-\beta V_o} + \beta V_o - 1 + \left(\frac{n_i}{N_A}\right)^2\left[e^{\beta V_o} - 1\right]\right\}^{1/2} + \sqrt{2\kappa_s \varepsilon_o q N_A V_o} \quad . \tag{4.2}$$

From Fig. 1.21, Eq. 4.2 (hence, Eq. 3.36) is shown to yield values of Q_{io} that are adequate only for the strong inversion mode of MOSFET operation. At approximately the surface threshold voltage ($V_o \approx 2\phi_F$) Eq. 3.36 yields a Q_{io} of zero, and for weak inversion operation ($V_o < 2\phi_F$) this approximation yields a <u>negative</u> magnitude for Q_{io}.

The source of this error has been identified: it arises from an application of the depletion approximation in evaluating the depletion charge (Q_D) in a MOSFET (Eq. 3.32). This depletion approximation over-estimates (slightly) the total depletion charge, due to an elimination of mobile carriers in the mathematical model. We recall (from Eq. 3.33) that we calculated the inversion charge (Q_i) as the difference between the total electrostatic charge (Q_T) and Q_D:

$$Q_i = Q_T - Q_D \quad . \tag{4.3}$$

Clearly, this slight error in Q_D produces little difficulty for strong inversion conditions (when $Q_T >> Q_D$); however, under weak inversion conditions $Q_T \approx Q_D$ in Eq. 4.3 and, therefore, a small error in the calculated value of Q_D produces a large error in the calculated value of Q_i.

A substantial increase in accuracy is obtained through a modification of the elementary relation for Q_i (Eq. 3.36). This improvement arises from replacing the traditional depletion charge term (obtained from the depletion layer theory of an abrupt asymmetrical p-n junction) by an approximation derived from an exact solution of Poisson's equation [9].

As in elementary MOSFET theory [1], for the depletion charge calculation we assume that the voltage V_s is attributable only to ionized impurity atoms. Implicit in this assumption is an inversion layer width that is small, relative to the depletion layer width. Thus, from this assumption we can neglect the contribution of inversion carriers in this exact solution of Poisson's equation and obtain an approximate expression for the depletion charge:

$$Q_D(V_s) = -\sqrt{2}\left(\frac{\kappa_s \varepsilon_o}{L_D}\right)\left(\frac{kT}{q}\right)\left[e^{-\beta V_s} + \beta V_s - 1\right]^{1/2} . \qquad (4.4)$$

Thus (substituting Eqs. 3.31, 3.35, and 4.4 into Eq. 4.3), we have

$$Q_i(V_s) = -C_{ox}(V_G - V_s) + \sqrt{2}\left(\frac{\kappa_s \varepsilon_o}{L_D}\right)\left(\frac{kT}{q}\right)\left[e^{-\beta V_s} + \beta V_s - 1\right]^{1/2} . \qquad (4.5)$$

From the foregoing discussion, we presume Eq. 4.5 yields an adequate approximation for $Q_i(V_s)$ near the source end of this semiconductor structure. In particular, when $V_s = V_o$, this equation yields

$$Q_{io} \equiv Q_i(V_o) = -C_{ox}(V_G - V_o) + \sqrt{2}\left(\frac{\kappa_s \varepsilon_o}{L_D}\right)\left(\frac{kT}{q}\right) \cdot \left[e^{-\beta V_o} + \beta V_o - 1\right]^{1/2} . \qquad (4.6)$$

From Fig. 1.21, this revised approximate relation for the inversion charge in a MOSFET yields results that are in satisfactory agreement with a rigorous solution of this problem. Clearly, this new approximation offers adequate agreement throughout the range from extremely weak inversion to strong inversiong.

It should be noted that by utilizing Eq. 4.6 in the present theory of MOSFET operation we encounter an additional complication. Namely this theory yields the magnitude of Q_{io} and, hence, I_D,

explicitly in terms of V_o and implicitly in terms of V_G. From Eq. 4.1 we have

$$V_G = V_o + \frac{\sqrt{2}}{C_{ox}} \left(\frac{\kappa_s \varepsilon_o}{L_D}\right)\left(\frac{kT}{q}\right) \cdot \left\{ e^{-\beta V_o} + \beta V_o - 1 + \left(\frac{n_i}{N_A}\right)^2 \left[e^{\beta V_o} - 1 \right] \right\}^{1/2} \quad . \qquad (4.7)$$

Clearly, this problem represents only a minor algebraic complication.

4.2 Revised Calculation of λ_o in a MOSFET

In the present theory of MOSFET operation, λ_o represents the magnitude of our separation parameter, $\lambda(V_s)$, at the source end of this semiconductor structure (where $V_s = V_o$). Thus, from Eq. 3.5b, λ_o has the qualitative form

$$\lambda_o = \left\{ \frac{\frac{kT}{q}\left(\frac{d^2 Q_i}{dV_s^2}\right)_{V_o} - \left(\frac{dQ_i}{dV_s}\right)_{V_o}}{Q_{io} - \frac{kT}{q}\left(\frac{dQ_i}{dV_s}\right)_{V_o}} \right\} \quad . \qquad (4.8)$$

In Section 3.0, we derived a quantitative expression for λ_o (Eq. 3.39) from an earlier theory for MOSFET operation. This derivation utilized a relation for $Q_i(V_s)$ taken from elementary MOSFET theory (Eq. 3.34). From discussions in the preceding section, Eq. 3.34 is inapplicable for the weak inversion mode of operation and, therefore, Eq. 3.39 for λ_o contains this same limitation.

From a rigorous one-dimensional solution of Poisson's equation, we obtained an expression for $Q_i(V_s)$ near the source end of this structure (Eq. 4.5). Using this expression to evaluate Q_{io}, we

obtained satisfactory results throughout both weak and strong inversion modes of MOSFET operation. Herein, we will utilize this relation (Eq. 4.5) to derive a new quantitative expression for λ_o.

A substantial degree of mathematical simplification results if we neglect the term $e^{-\beta V_s}$ in Eq. 4.5; thereby

$$Q_i(V_s) \approx -C_{ox}(V_G - V_s) + \sqrt{2}\left(\frac{\kappa_s \varepsilon_o}{L_D}\right)\left(\frac{kT}{q}\right)\left[\beta V_s - 1\right]^{1/2} . \quad (4.9)$$

We can neglect this term when

$$V_s >> (kT/q)\ [1 - \exp(-qV_s/kT)] \ , \quad (4.10)$$

which is clearly the situation in all cases of practical interest. From Eq. 4.9, by differentiating with respect to V_s, we have

$$\frac{dQ_i}{dV_s} = C_{ox} + \frac{1}{\sqrt{2}}\left(\frac{\kappa_s \varepsilon_o}{L_D}\right)\left(\frac{kT}{q}\right)^{1/2}\left[V_s - \frac{kT}{q}\right]^{-1/2} \quad (4.11)$$

$$\frac{d^2Q_i}{dV_s^2} = -\frac{1}{2\sqrt{2}}\left(\frac{\kappa_s \varepsilon_o}{L_D}\right)\left(\frac{kT}{q}\right)^{1/2}\left[V_s - \frac{kT}{q}\right]^{-3/2} . \quad (4.12)$$

To assess the adequacy of using these approximations to calculate λ_o, we made comparisons between Eqs. 4.9, 4.11, and 4.12 and a rigorous nonequilibrium solution of Poisson's equation in one spatial dimension (see Appendix C).

These comparisons are illustrated in Figs. 1.22 and 1.23 for strong inversion operation, and in Figs. 1.24 and 1.25 for weak inversion operation. From Figs. 1.22 through 1.25 we have satisfactory agreement between our approximations for Q_i and for dQ_i/dV_s and this rigorous one-dimensional solution of Poisson's equation. Moreover, in Figs. 1.23 and 1.25, we have a qualitative illustration

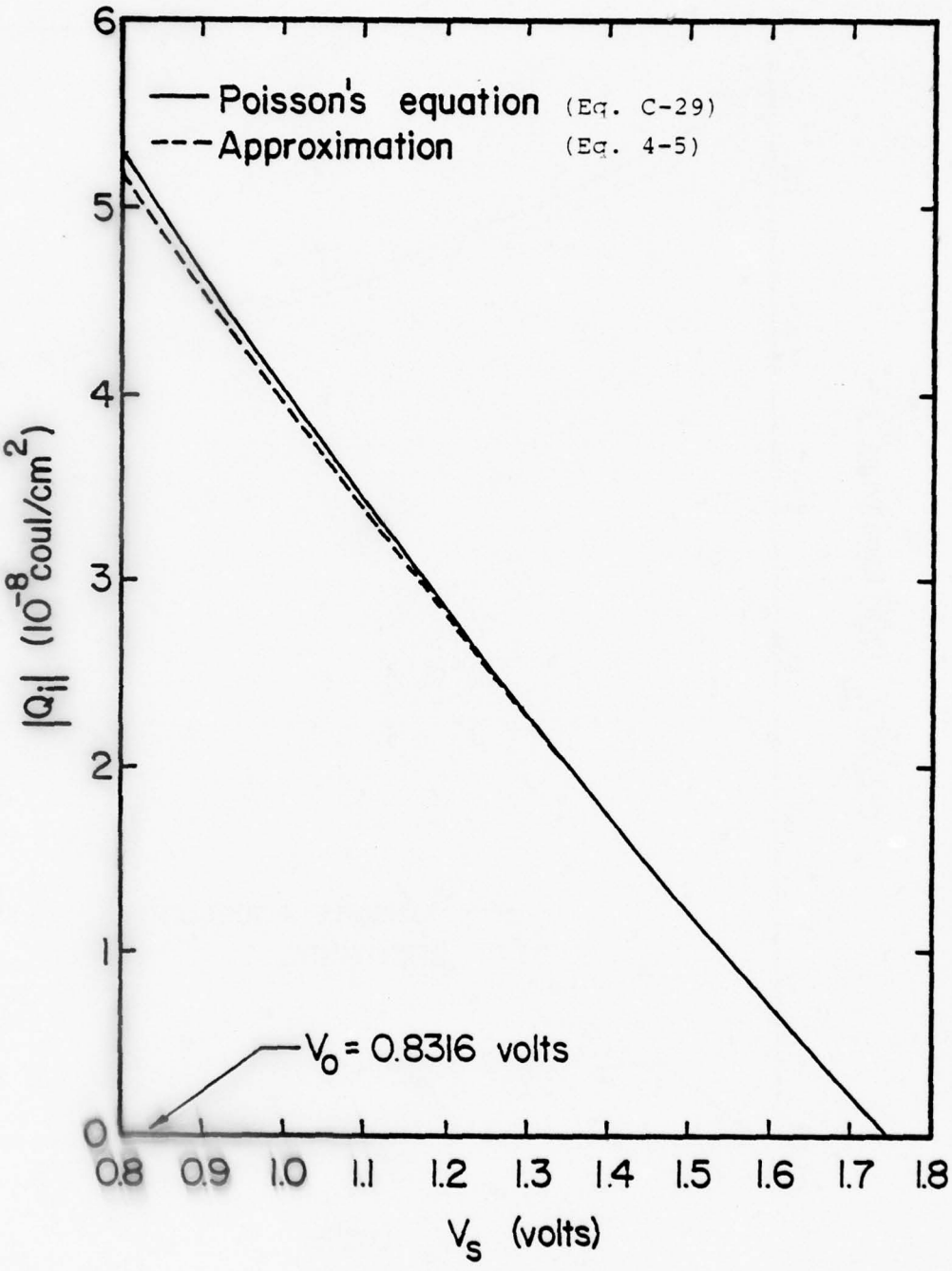


Fig. 1.22. Calculated inversion charge distribution in a MOSFET for strong inversion operation. $N_A = 2 \times 10^{16}$ cm^{-3}, $t_{ox} = 2000$A°, $V_G = 8$ volts.

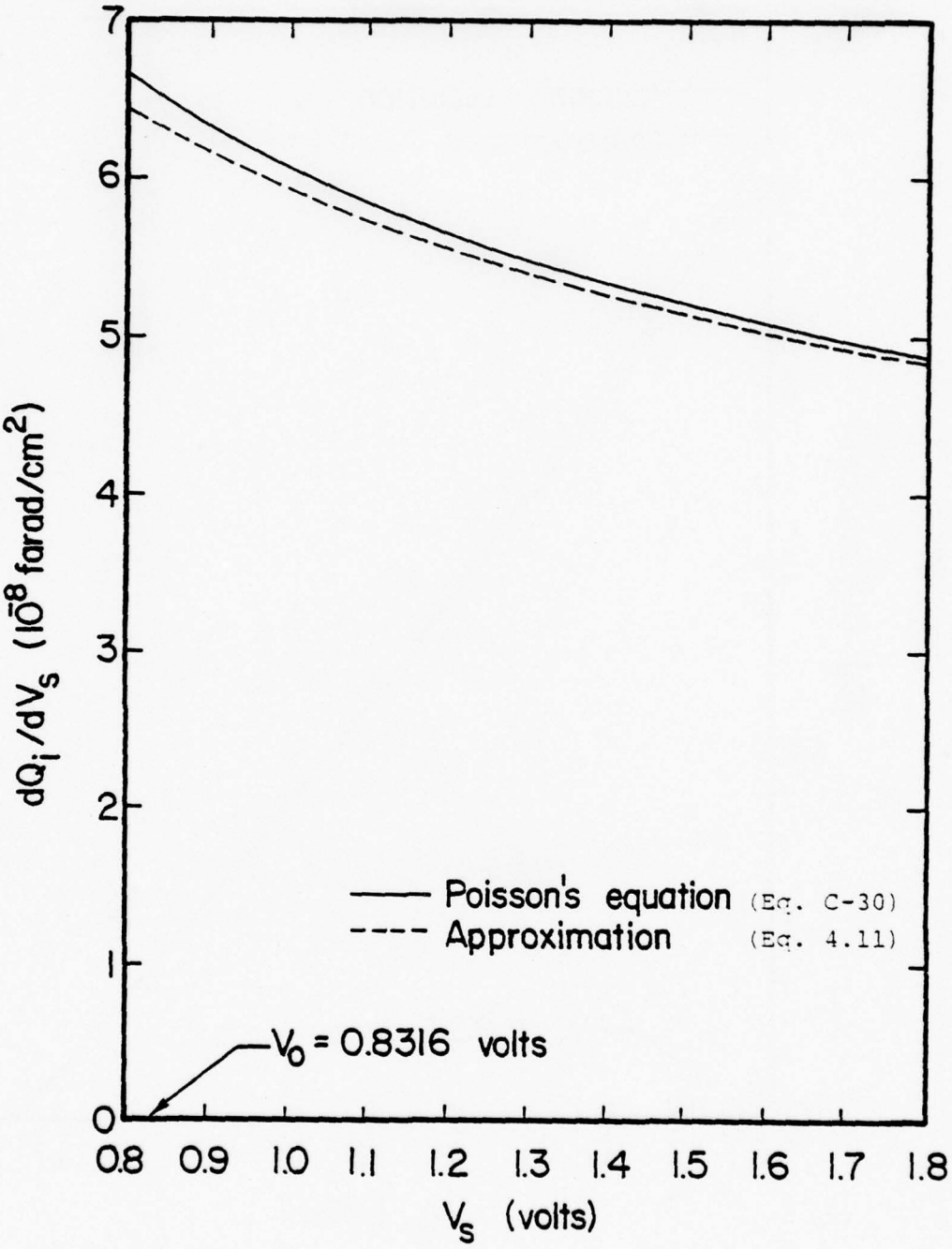


Fig. 1.23. Calculated dQ_i/dV_s in a MOSFET for strong inversion operation. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 2000A°$, $V_G = 8$ volts.

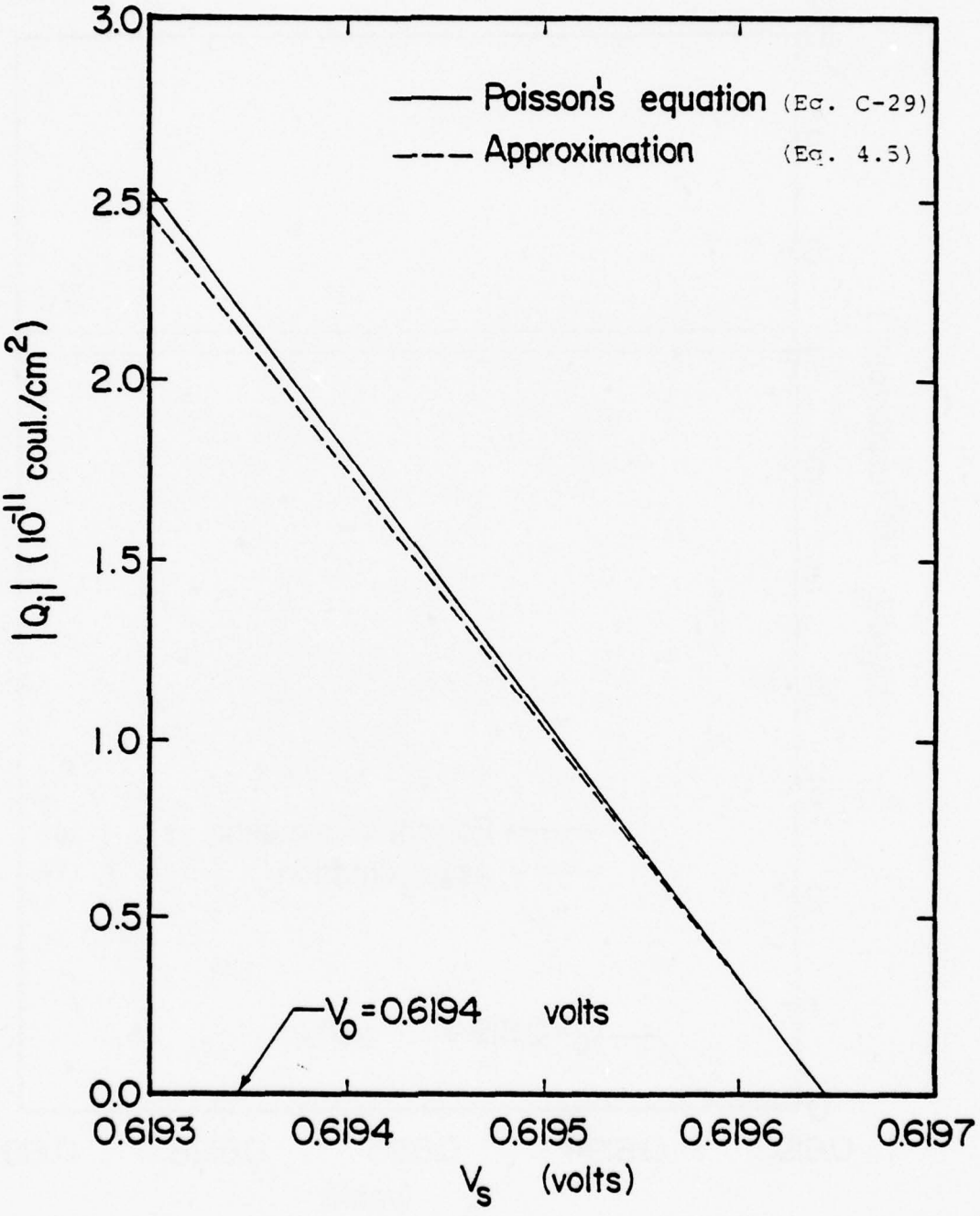


Fig. 1.24. Calculated inversion charge distribution in a MOSFET for weak inversion operation. $N_A = 2\times10^{16}cm^{-3}$, t_{ox} = 2000A°, V_G = 4.3 volts.

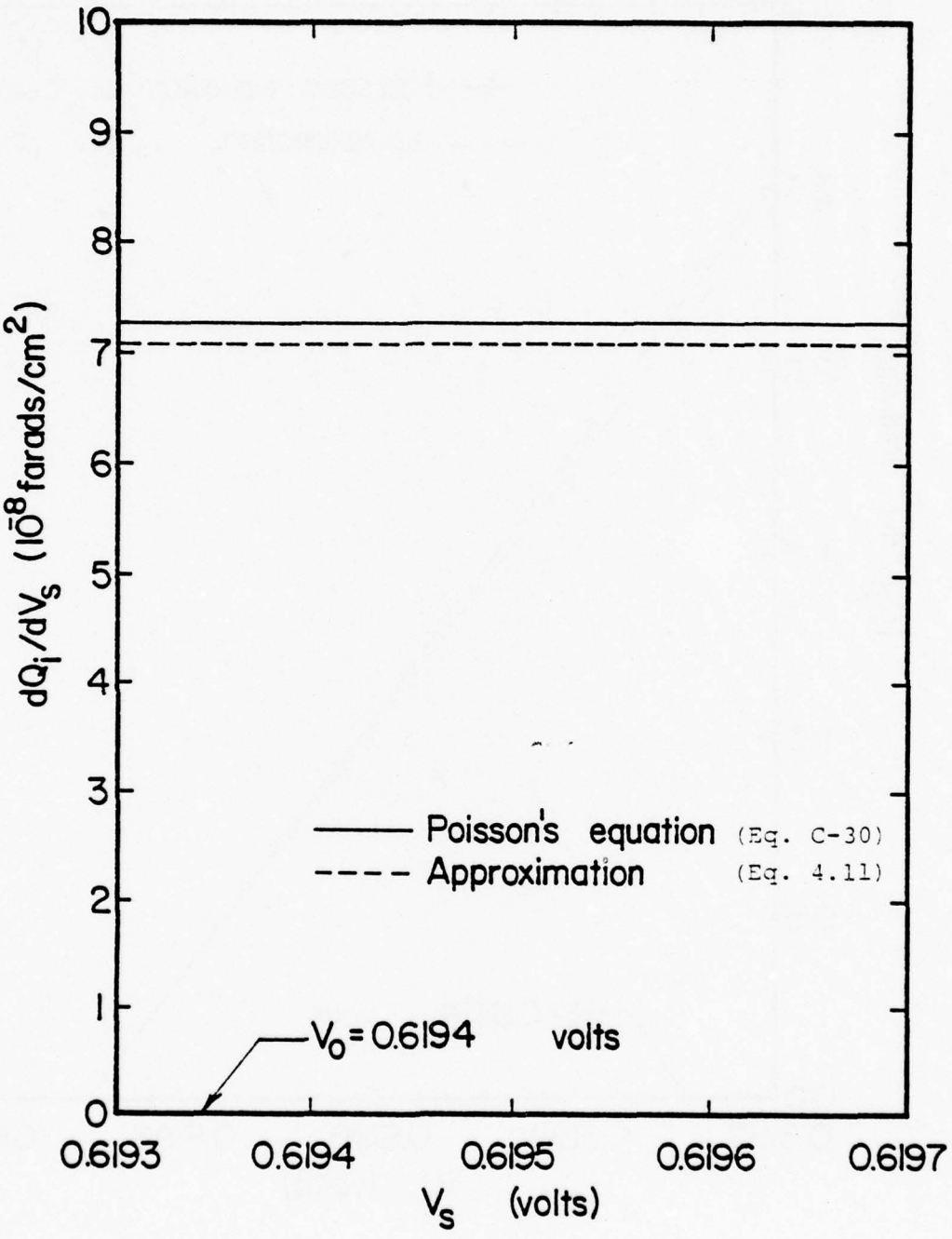


Fig. 1.25. Calculated dQ_i/dV_s in a MOSFET for weak inversion operation. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 2000A°$, $V_G = 4.3$ volts.

that good agreement is attained between our approximation for d^2Q_i/dV_s^2 and this rigorous solution. From this agreement, we presume Eqs. 4.9, 4.11, and 4.12 provide adequate approximations for our calculation of λ_o (Eq. 4.8).

Evaluating Eqs. 4.9, 4.11, and 4.12 at $V_s=V_o$ and, thereafter, substituting these results into Eq. 4.8, we obtain

$$\lambda_o = \left\{ \frac{C_{ox} + \frac{1}{2}\left(\frac{2\kappa_s\varepsilon_o qN_A}{V_o - kT/q}\right)^{1/2}\left[1 + \frac{kT/q}{2(V_o - kT/q)}\right]}{C_{ox}(V_G - V_o + \frac{kT}{q}) - \left[V_o - \frac{3}{2}\left(\frac{kT}{q}\right)\right]\left(\frac{2\kappa_s\varepsilon_o qN_A}{V_o - kT/q}\right)^{1/2}} \right\}. \quad (4.13)$$

For $V_o >> 3/2(kT/q)$, Eq. 4.13 has the approximate form

$$\lambda_o \approx \frac{C_{ox} + \frac{1}{2}\left[2\kappa_s\varepsilon_o qN_A/(V_o - \frac{kT}{q})\right]^{1/2}}{C_{ox}(V_G - V_o + \frac{kT}{q}) - \left[V_o - \frac{3}{2}\left(\frac{kT}{q}\right)\right]\left(\frac{2\kappa_s\varepsilon_o qN_A}{V_o - kT/q}\right)^{1/2}} \quad (4.14)$$

where V_o is given implicitly by Eq. 4.7.

For the weak inversion mode, Eq. 4.7 has the approximate form

$$V_G \approx V_o + \frac{\sqrt{2\kappa_s\varepsilon_o qN_A(V_o - \frac{kT}{q})}}{C_{ox}} . \quad (4.15)$$

Upon substituting Eq. 4.15 into Eq. 4.14 and, thereafter, simplifying the result through straightforward algebraic manipulations, we obtain $\lambda_o = q/kT$. Thus, our new approximate expression for λ_o yields the correct magnitude for this separation parameter in the limit of weak inversion. This limiting behavior is also demonstrated by calculations (using Eq. 4.14) shown in Fig. 1.26.

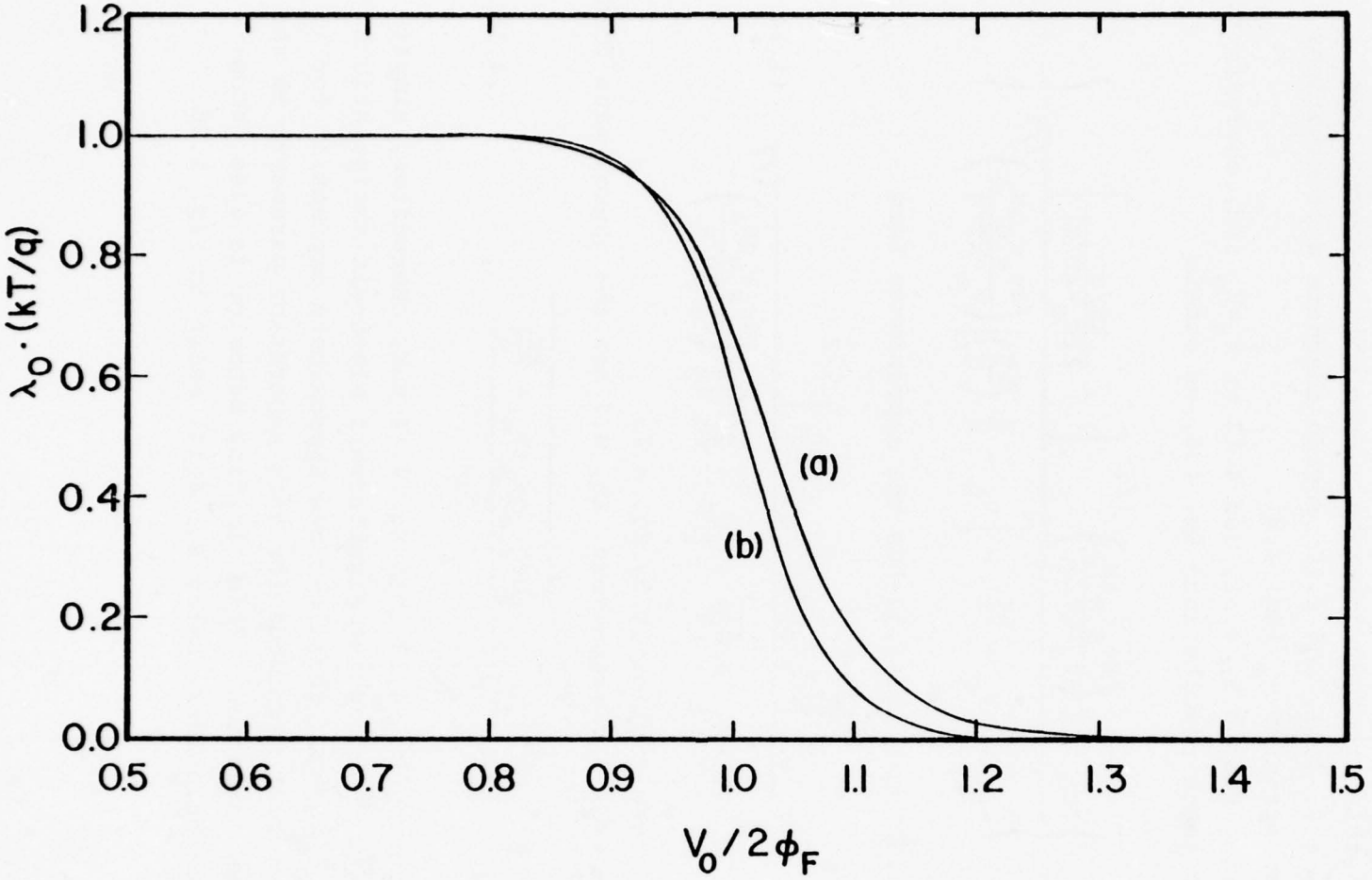


Fig. 1.26. Calculated separation parameter, λ_o, in a MOSFET. t_{ox} = 1000A° and a) $N_A = 2 \times 10^{15} cm^{-3}$, b) $N_A = 2 \times 10^{16} cm^{-3}$.

4.3 Revised Calculation of $V_s(x)$ in a MOSFET

From two-dimensional computer calculations for MOSFET operation, the voltage distribution within this semiconductor structure becomes modified near the source and drain metallurgical junctions. In an earlier study, it was proposed that these modifications result from source and drain junction space-charge layer interaction with the inversion layer. Further, through comparisons between this earlier theory and rigorous two-dimensional computer calculations, it was shown that this interaction produced a slight difference between the physical source-to-drain junction distance (L) in a MOSFET and the effective distance (L_{eff}) in this theory [36]. However, this interaction mechanism is not fully understood, and it is not explicitly included in the present MOSFET theory. For this reason, we place particular emphasis on the selection of boundary conditions in this mathematical analysis.

The foregoing discussion implies our viewpoint on this aspect of MOSFET operation: near the source and drain metallurgical junctions, regions exist that are not readily described by a simple one-dimensional theory; however, these regions act essentially as infinite sinks for electron transport through the source-drain inversion layer. For this reason, we presume that this inversion layer constitutes the active region of an MOS transistor and, therefore, contains the dominant physical mechanisms influencing the volt-ampere characteristics of this semiconductor device.

Figure 1.27a qualitatively illustrates this concept of MOS transistor operation. In this figure, we presume junction space-charge layer interactions with the source-drain inversion layer take place in regions 1 and 3; between these regions, we have the active source-drain inversion layer (region 2). Furthermore, Fig. 1.27b shows a qualitative potential distribution, $V_s(x)$, corresponding to Fig. 1.27a. (For this purpose of illustration, we have deliberately exaggerated Fig. 1.27b to show characteristics of $V_s(x)$ relevant to the present discussion.)

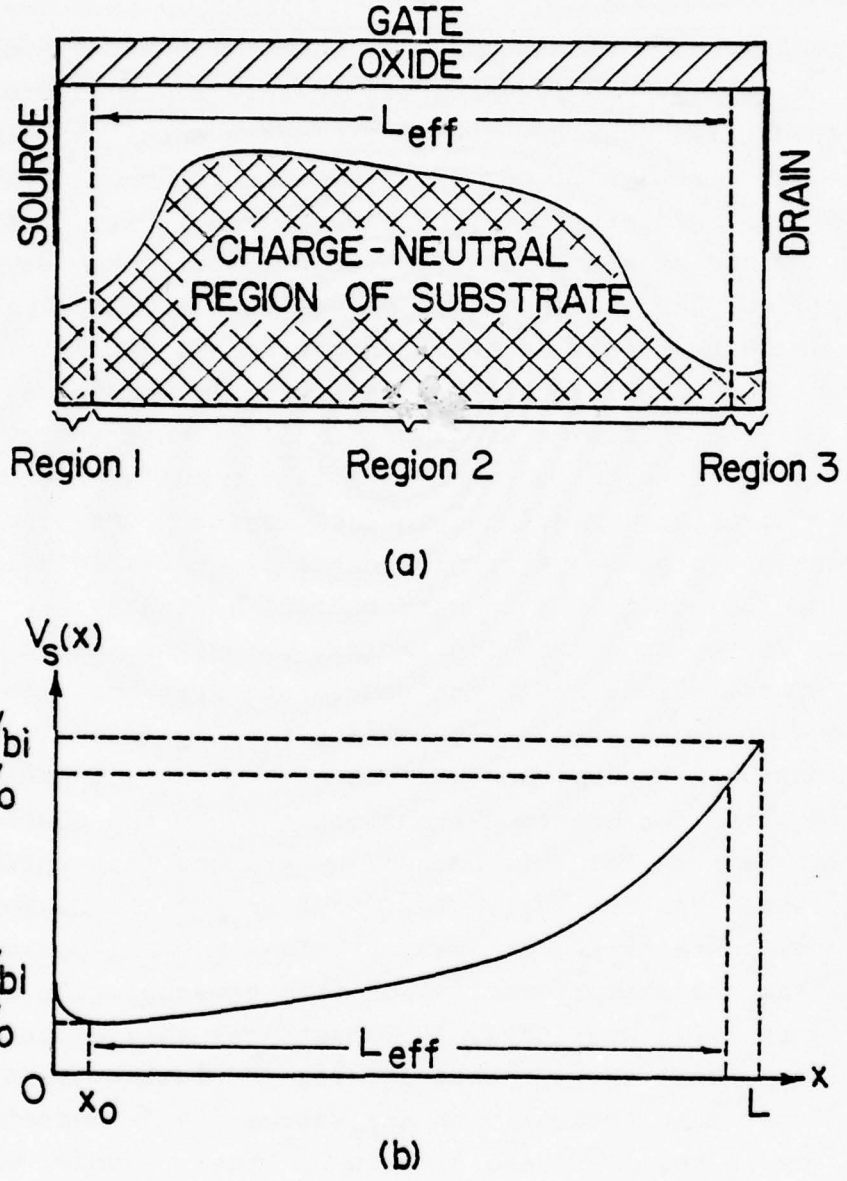


Fig. 1.27. Qualitative illustration showing assumed boundary conditions for $V_s(x)$ in a MOSFET.

In Fig. 1.27b, x_o represents a location near the source end of this semiconductor structure where electrostatic interactions between the gate electrode and the drain-and-source junctions are at a minimum. For this reason, we presumed that a one-dimensional calculation (Eq. 4.7) adequately approximates the magnitude of V_s at this location and, therefore, $V_s(x_o)=V_o$. In the present theory, we also assume that x_o is negligible in comparison to the total source-drain distance (L) and, therefore, $V_s(o)=V_o$. (This approximation is implicit in the earlier MOSFET theory of Section 3.0.)

To obtain an explicit solution for $V_s(x)$ in the present theory, it is also necessary to calculate E_o, the magnitude of source-drain electric field near the source junction. In an earlier treatment of this problem (see Section 3.0), it was assumed that $V_s=V_D$ when x=L, where V_D represents the applied drain-source voltage; however, this choice yields a nonphysical result -- from Eq. 3.22, we would have a nonzero source-drain electric field, E_o, even when $V_D=0$.

At the drain metallurgical junction (see Fig. 1.27b), $V_s=V_D+V_{bi}$, where V_{bi} represents the built-in potential of the drain junction. Thus, at the boundary between regions 2 and 3 (see Fig. 1.27a), we would expect this surface potential to attain a mganitude approaching this boundary condition. An exact calculation of this potential would require a two-dimensional solution of Poisson's equation whtin this semiconductor structure. Instead, we assume $V_s \approx V_D+V_o$ at this boundary and, in the following discussion, we will justify this assumption.

Since $V_s(x)$ is a rapidly increasing function near the drain junction, little error is introduced if we assume $V_s=V_D+V_o$ when x=L and, therefore, Eq. 3.13 yields

$$E_o = \frac{1}{(\lambda_o+A)L}\left\{1 - (1-AV_D)^{\frac{\lambda_o+A}{A}}\right\}. \qquad (4.16)$$

From Eq. 4.16, $E_o=0$ when $V_D=0$ and, thus, we have a self-consistent solution for this component of electric field.

Let us now consider an application of Eq. 4.16 to the weak inversion mode of MOSFET operation. From Eq. 3.11, as $\lambda_o \rightarrow q/kT$ in weak inversion the parameter A becomes small and, therefore, Eq. 4.16 has the limit

$$\text{Lim}(E_o) \approx \frac{(kT/q)}{L} [1-\exp(-qV_D/kT)] \quad . \qquad (4.17)$$

$$\lambda_o \rightarrow q/kT$$

In weak inversion, Eq. 4.17 predicts that E_o attains a limiting magnitude of $(kT/q)/L$ when $V_D \gtrapprox 5kT/q$. Contrasting with this result, throughout this mode of MOSFET operation our two-dimensional computer calculations show that E_o decreases exponentially with a decrease in applied gate voltage.

The source of this discrepancy between our theory and these computer calculations becomes apparent when we consider the basis upon which this theory was constructed. Briefly, the present theory satisfies a fundamental requirement of electric current continuity within this semiconductor structure, yet there is no guarantee that Poisson's equation will be satisfied. Indeed, our solution for $V_s(x)$ in a MOSFET (Eq. 3.13) contains an implicit assumption that this surface potential arises entirely from mobile electrons residing in the source-drain inversion layer; in contrast, our two-dimensional computer calculations show that V_s arises from the total electrostatic charge (due to both mobile electrons and unneutralized impurities) within this semiconductor structure.

At present, we have no rigorous method for amending our MOSFET theory to include the influence of these unneutralized impurities on $V_s(x)$. However, from a physical viewpoint, we have the following situation: at the oxide-semiconductor interface in strong inversion, the electron density greatly exceeds the substrate impurity level ($n_s >> N_A$) and, therefore, V_s arises essentially from these

mobile electrons; in weak inversion ($n_s << N_A$), these unneutralized impurities effectively shield the electrons and, thereby, modify this surface potential. Thus, intuitively we would expect that the degree of this shielding depends qualitatively on the ratio of mobile electrons (n_s) to the total number of electrostatic changes ($n_s + N_A$).

From this line of reasoning, we devised a semi-empirical correction to Eq. 4.16,

$$E_o = \frac{1}{(\lambda_o+A)L}\left(\frac{n_{so}}{n_{so}+2N_A}\right)\left\{1-(1-AV_D)^{\frac{\lambda_o+A}{A}}\right\} \quad , \qquad (4.18)$$

where the factor of 2 (times N_A) was selected to yield agreement with our two-dimensional computer calculations. In Eq. 4.18, n_{so} represents the surface density of electrons near the source end of this semiconductor structure (where $V_s=V_o$), hence

$$n_{so} \equiv n_s(V_o) = \frac{n_i^2}{N_A}\exp(+qV_o/kT) \quad . \qquad (4.19)$$

For strong inversion operation, when $n_{so} >> N_A$, Eq. 4.18 simplifies to our previous result (Eq. 4.16). However, for weak inversion operation, when $n_{so} << N_A$, Eq. 4.18 has the approximate form

$$\underset{\lambda_o \to q/kT}{\mathrm{Lim}}(E_o) \approx \frac{1}{2}\left(\frac{n_i}{N_A}\right)^2 \exp(+qV_o/kT) \cdot \frac{(kT/q)}{L}\left[1-\exp(-qV_D/kT)\right] \quad . \qquad (4.20)$$

To a first order approximation, it has been shown that V_o depends linearly on V_G in this weak inversion mode of operation [30,31]. Thus, in qualitative agreement with our two-dimensional computer calculations, Eq. 4.20 implicitly predicts an exponential dependence

of E_o on applied gate voltage. Furthermore, quantitative comparisons between Eq. 4.18 (hence, also its weak inversion limit, Eq. 4.20) and these computer calculations show that this relation adequately predicts E_o for a wide range of substrate doping (see Fig. 1.28).

4.4 A Modified Treatment of Electric Current Saturation

In a previous study [36], the MOSFET theory outlined in section 3.0 was shown to yield satisfactory results when applied to a long channel structure (L=10μm) with lightly doped substrate ($N_A=2x10^{15}cm^{-3}$) and moderately thin oxide (t_{ox}=1000A°). Contrasting with this agreement, in the present study we found that this theory produces significant discrepancies when applied to a structure with a more highly doped substrate ($N_A=2x10^{16}cm^{-3}$) and thicker oxide (t_{ox}=2000A°).

Through comparisons with our two-dimensional computer solution for this structure, we observed that this theory overpredicted the magnitude of source-drain current by approximately 50 percent in the saturation region of the volt-ampere characteristics. Further, this theory also overpredicted the magnitude of applied drain voltage required to produce electric current saturation in this structure. Through detailed comparisons between this theory and our computer solution, we determined that the foregoing discrepancy arose from an error in calculating E_o, the magnitude of source-drain electric field near the source end of this device. Specifically, from Eq. 4.18, E_o is exceedingly dependent on the parameter A and, therefore, an error in calculating this parameter can yield a serious error in this volt-ampere calculation.

From discussions in Section 3.0, the separation parameter, $\lambda(V_s)$ always attains a magnitude of q/kT in the limit of weak inversion operation. In this earlier MOSFET theory, the parameter A was evaluated by assuming $\lambda(V_s) \rightarrow q/kT$ at the point of channel termination (where $V_s=V_G$). This assumption implies that we have a transition between strong inversion operation (at the source end of this structure) and weak inversion operation (at the drain end)

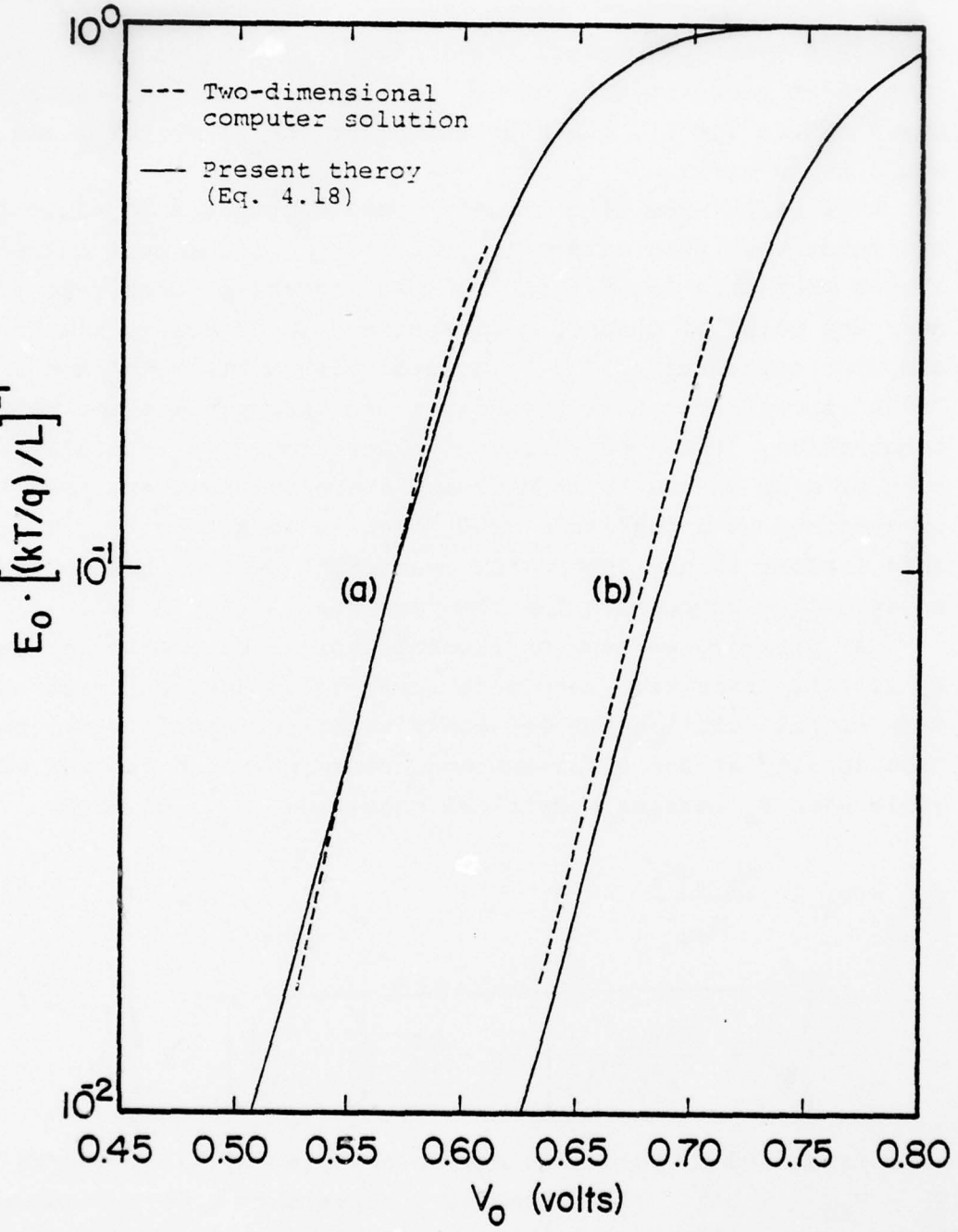


Fig. 1.28. Calculated magnitude of E_o versus V_o in a MOSFET. L = 10 m, t_{ox} = 1000A°, and a) $N_A = 2\times10^{15}cm^{-3}$, b) $N_A = 2\times10^{16}cm^{-3}$.

near this point of channel termination. If instead this transition takes place at some other location between the source and drain (where $V_s \neq V_G$), the expression for the parameter A (Eq. 3.11) would be in error.

For structures with lightly doped substrates ($N_A = 2 \times 10^{15} cm^{-3}$) and relatively thin oxides ($t_{ox} = 1000A°$), our computer calculations showed that this transition into weak inversion does take place near the point of channel termination. In contrast, similar computer calculations for structures with $N_A > 10^{16} cm^{-3}$ and $t_{ox} = 2000A°$ showed that this transition can take place prior to channel termination. From the foregoing discussion, $\lambda(V_s)$ attains a magnitude of q/kT in the limit of weak inversion; however, this limit is attained at a magnitude of V_s that is less than V_G. Therefore, this earlier theory for MOSFET operation (Section 3.0) contains an erroneous expression for the parameter A (Eq. 3.11).

At present, we have no rigorous analytical method to calculate V_s at this transition into weak inversion. However, from our one-dimensional solution for Poisson's equation (Appendix C), the electron density at the oxide-semiconductor interface becomes negligible when V_s attains a critical magnitude, V_{sc}, given by

$$V_{sc} = V_G - \frac{\kappa_s \varepsilon_o q N_A}{C_{ox}^2} \cdot \left\{ \sqrt{1 + \frac{2C_{ox}^2}{\kappa_s \varepsilon_o q N_A} \left[V_G - \frac{kT}{q} \left(1 - \frac{n_i}{N_A} \right) \right]} - 1 \right\} . \qquad (4.21)$$

As a first order approximation, we now presume $\lambda(V_s) \rightarrow q/kT$ when $V_s \rightarrow V_{sc} + 5kT/q$. (The term 5kT/q represents a correction to this one-dimensional theory, which we will justify in the following discussion.) Therefore, from Eq. 3.10

$$A = \frac{1 - (kT/q)\lambda_o}{V_{sc} - V_o + 5kT/q} . \qquad (4.22)$$

Figure 1.29 shows calculated values of V_{sc} and V_o as a function of gate voltage (V_G). From Fig. 1.29, the magnitude of V_{sc} approaches V_o in the limit of weak inversion operation. From the foregoing discussion, we have electric current saturation in a MOSFET when V_s near the drain end of this structure (i.e., $V_s=V_D+V_o$) attains a magnitude of $V_{sc}+5kT/q$. Therefore, the magnitude of drain voltage required to produce this saturation is $V_D=V_{sc}-V_o+5kT/q$ or, in weak inversion, $V_D \approx 5kT/q$. Thus, by introducing the correction term, $5kT/q$, our theory predicts full saturation of the source-drain electric current in weak inversion at a value of drain voltage consistent with published experimental data for this semiconductor device [30,31].

It should be noted that this modification of our model (Eq. 4.22) does not alter the qualitative form of the expression for E_o (Eq. 4.18) nor that of the volt-ampere relation (Eq. 3.23). However, for the saturation region of these volt-ampere characteristics, we must replace Eqs. 3.25-3.28 with suitably revised expressions. Borrowing from the theory of abrupt asymmetrical pn junctions [8], we assume a quadratic potential distribution within the drain space charge layer ($L_c(V_D)$ x L):

$$V_s(x) = (V_D+V_{bi}) \left[1 - \frac{L-x}{W_D} \right]^2 , \qquad (4.23)$$

where

$$V_{bi} = \frac{kT}{q} \log_e \left(\frac{N_A N_D}{n_i^2} \right) \qquad (4.24)$$

$$W_D = \sqrt{\frac{2\kappa_s \varepsilon_o}{qN_A} (V_D+V_{bi})} . \qquad (4.25)$$

(Source-drain channel length, L, appears in Eq. 4.23 due to our choice of origin for the coordinate x -- see Fig. 1.19.) Setting $x=L_c(V_D)$ when $V_s=V_{sc}+5kT/q$, Eq. 4.23 yields

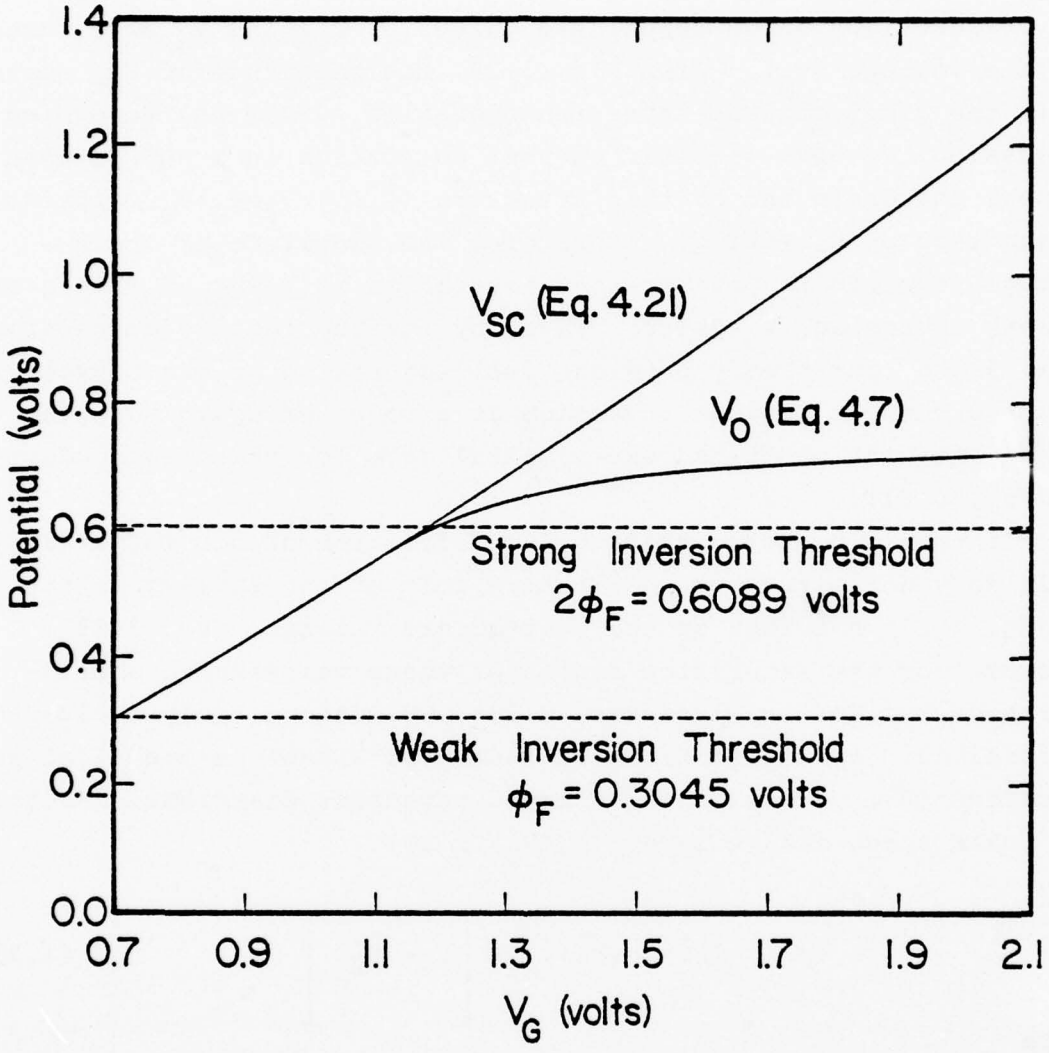


Fig. 1.29. Calculated magnitudes of V_{sc} and V_o versus gate voltage in a MOSFET. $N_A = 2 \times 10^{15} cm^{-3}$, $t_{ox} = 1000A°$.

$$L_c(V_D) = L - W_D \left[1 - \sqrt{(V_{sc}+5kT/q)/(V_D+V_{bi})}\right] \quad . \quad (4.26)$$

Equations 4.21, 4.22, and 4.24-4.26 provide a basis for deriving the volt-ampere relation of a MOSFET for the saturation region of these characteristics. This derivation will now be shown.

4.5 The Volt-Ampere Characteristics of a MOSFET

Foregoing sections of the present section outlined revisions to an earlier theory of MOSFET operation, namely the continuity theory reviewed in Section 2.0. Herein we will utilize these revisions to obtain an improved calculation of the volt-ampere characteristics for this semiconductor device.

From this earlier theory (Eq. 3.20), source-drain electric current in an MOS transistor has the qualitative form

$$I_D = W\mu E_o \left\{ Q_{io} - \frac{kT}{q}\left(\frac{dQ_i}{dV_s}\right)_{V_o} \right\} \quad . \quad (4.27)$$

In the present MOSFET theory, we obtained revised expressions for E_o (Eq. 4.18) and Q_{io} (Eq. 4.6), the magnitudes of the source-drain electric field and inversion charge, respectively, near the source end of this structure (where $V_s=V_o$).

The remaining quantity in Eq. 4.27, $(dQ_i/dV_s)_{V_o}$, represents the slope of this inversion charge distribution at $V_s=V_o$. An earlier theory of MOSFET operation (see Section 3.0) calculated this slope from an approximate form for $Q_i(V_s)$, Eq. 3.17. Through comparisons with our two-dimensional computer calculations, in the present study we found that Eq. 3.17 produces a serious error when applied to the weak inversion mode of operation.

It should be noted that the term $(dQ_i/dV_s)_{V_o}$ is evaluated at the source end of this semiconductor structure. Therefore, we can calculate this term using our revised one-dimensional approximations for dQ_i/dV_s, Eq. 4.11. At $V_s=V_o$, this equation yields

$$\left(\frac{dQ_i}{dV_s}\right)_{V_o} = C_{ox} + \frac{1}{\sqrt{2}}\left(\frac{\kappa_s \varepsilon_o}{L_D}\right)\left(\frac{kT}{q}\right)^{1/2}\left[V_o - \frac{kT}{q}\right]^{-1/2} \quad . \tag{4.28}$$

Foregoing discussions in the present section have shown that Eq. 4.11 (and, therefore, Eq. 4.28) yields satisfactory results throughout both the weak and strong inversion modes of MOSFET operation.

Substituting our revised expression for E_o, Eq. 4.18, into Eq. 4.27 yields

$$I_D = \frac{W\mu}{(\lambda_o+A)L}\left(\frac{n_{so}}{n_{so}+2N_A}\right)\left\{Q_{io} - \frac{kT}{q}\left(\frac{dQ_i}{dV_s}\right)_{V_o}\right\} \cdot \left\{1 - (1-AV_D)^{\frac{\lambda_o+A}{A}}\right\} \quad . \tag{4.29}$$

In Eq. 4.29: Q_{io} and $(dQ_i/dV_s)_{V_o}$ are given by Eqs. 4.6 and 4.28, respectively; the surface density of inversion layer electrons near the source, n_{so}, is given by Eq. 4.19; and the parameters λ_o and A are given by Eqs. 4.14 and 4.22, respectively.

For investigating the weak inversion limit of our volt-ampere relation, a substantial algebraic simplification results by first rewriting Eq. 4.27 as follows:

$$I_D = -W\mu\left(\frac{kT}{q}\right)E_o\left(\frac{dQ_i}{dV_s}\right)_{V_o} \cdot \left\{1 - Q_{io}\frac{kT}{q}\left(\frac{dQ_i}{dV_s}\right)_{V_o}^{-1}\right\} \quad . \tag{4.30}$$

The ratio of the drift component to the diffusion component of I_D is given by

$$\frac{I_{drift}}{I_{diff}} = \frac{W\mu Q_i (dV_s/dx)}{-W\mu (kT/q)(dQ_i/dx)} = \frac{-Q_i}{(kT/q)(dQ_i/dV_s)} \tag{4.31}$$

and, therefore,

$$\left(\frac{I_{drift}}{I_{diff}}\right)_{V_o} = -Q_{io} \frac{kT}{q} \left(\frac{dQ_i}{dV_s}\right)^{-1}_{V_o} . \tag{4.32}$$

Substituting Eq. 4.32 into Eq. 4.30 yields

$$I_D = -W\mu \left(\frac{kT}{q}\right) E_o \left(\frac{dQ_i}{dV_s}\right)_{V_o} \left[1 + \left(\frac{I_{drift}}{I_{diff}}\right)_{V_o}\right] . \tag{4.33}$$

From foregoing discussions, in weak inversion nearly all of the source-drain electric current is attributable to diffusion of inversion layer carriers. For this reason, the drift-to-diffusion current ratio (Eq. 4.32) attains a magnitude less than unity near the surface threshold ($V_o=2\phi_F$) and decreases rapidly with a decrease in V_o (see Fig. 1.30). Therefore, in weak inversion as $\lambda_o \to q/kT$, Eq. 4.33 has the qualitative form

$$\underset{\lambda_o \to q/kT}{\text{Lim}}(I_D) \approx -W\mu \left(\frac{kT}{q}\right)\left(\frac{dQ_i}{dV_s}\right)_{V_o} \underset{\lambda_o \to q/kT}{\text{Lim}} (E_o) . \tag{4.34}$$

Substituting Eqs. 4.20 and 4.28 into Eq. 4.34 yields a quantitative expression for this limit:

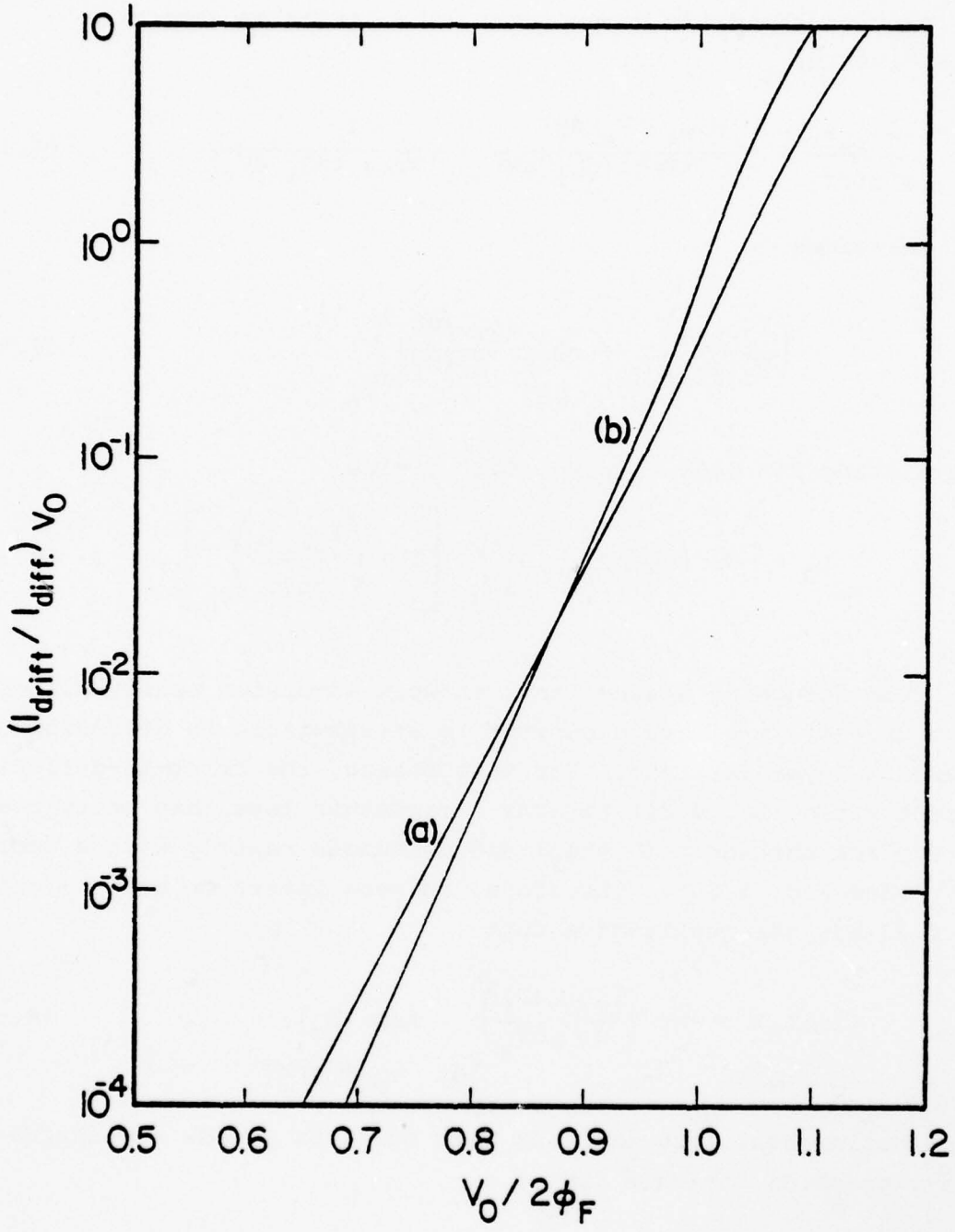


Fig. 1.30. Calculated ratio of drift-component to diffusion-component of source-drain electric current, near the source-end of a MOSFET. t_{ox} = 1000A° and a) $N_A = 2 \times 10^{15} cm^{-3}$, b) $N_A = 2 \times 10^{16} cm^{-3}$.

$$\text{Lim}(I_D) = \frac{-W\mu}{2L}\left(\frac{n_i}{N_A}\right)^2\left(\frac{kT}{q}\right)^2 C_{ox} f(V_o)\ [1-\exp(-qV_D/kT)]\ , \qquad (4.35)$$
$$\lambda_o \to q/kT$$

where

$$f(V_o) = \left\{1 + \frac{(kT/q)^{1/2}}{\sqrt{2}}\left[\frac{(C_d/C_{ox})}{\sqrt{V_o-(kT/q)}}\right]\right\} e^{-\beta V_o} \qquad (4.36)$$

$$C_d = \frac{\kappa_s \varepsilon_o}{L_D} \quad . \qquad (4.37)$$

From Eq. 4.35, in weak inversion I_D exhibits an exponential transition into electric current saturation, with full saturation at an applied drain voltage $V_D \approx 5kT/q$. Further, from Eq. 4.36, $f(V_o)$ is essentially an exponential function of V_o and, therefore, Eq. 4.35 predicts that I_D increases exponentially with an increase in V_o. For this weak inversion mode of operation, it has been shown that V_o is a near-linear function of applied gate voltage [30,31] and, therefore, Eq. 4.35 implicitly predicts an exponential increase in I_D with an increase in this gate voltage -- in qualitative agreement with experiment [63].

The preceeding discussion shows that our revised volt-ampere relation for a MOSFET (Eq. 4.29) contains all the qualitative requirements for the weak inversion mode of operation: first, Eq. 4.29 predicts that all source-drain electric current arises from diffusion; second, this expression shows that I_D exhibits an exponential increase with an increase in gate voltage; and third, this relation shows that electric current saturation exhibits an exponential form.

---

[63] R. A. Stuart and W. Eccleston, *Electronics Letters*, 8, 225 (1972).

Equation 4.29 and its weak inversion limit, Eq. 4.35, are applicable only when channel length L is constant (i.e., for the triode mode of MOSFET operation). In the saturation mode, the active source-drain inversion layer maintains a constant total voltage of $V_{sc}-V_o+5kT/q$ across its length; however, the electrical length of this inversion layer, $L_c(V_D)$, decreases with an increase in applied drain voltage. Thus, in electric current saturation, Eq. 4.29 has the form

$$I_D = \frac{W\mu}{(\lambda_o+A)L_c(V_D)} \left(\frac{n_{so}}{n_{so}+2N_A}\right) \left\{ Q_{io} - \frac{kT}{q} \left(\frac{dQ_i}{dV_s}\right)_{V_o} \right\}$$

$$\cdot \left\{ 1 - [1-A(V_{sc}-V_o+5kT/q)]^{\frac{\lambda_o+A}{A}} \right\} , \qquad (4.38)$$

where V_{sc} and $L_c(V_D)$ are given by Eqs. 4.21 and 4.26, respectively.

5.0 Evaluation of the Revised MOSFET Theory

In the preceding section, we identified and corrected inadequacies in an earlier theory for MOSFET operation. The aim of these corrections was to develop a revised MOSFET theory applicable to a wide range of device structures, throughout both the weak and strong inversion modes of operation. To assess the adequacy of this revised theory, in the present section we will show detailed comparisons between MOSFET volt-ampere characteristics calculated from this theory and our rigorous two-dimensional computer calculations of these characteristics.

For all calculations in the present section, we assumed the following values of physical and geometrical parameters.

Quantity	Symbol	Magnitude
Inversion carrier mobility	μ	2.70. cm^2/volt-sec
Source/drain junction depth	y_j	0.5μm
Source/drain doping	N_D	$1.0x10^{19}cm^{-3}$
Intrinsic carrier concentration (Si)	n_i	$1.5x10^{10}cm^{-3}$
Dielectric constant (Si)	κ_s	12
Dielectric constant (SiO_2)	κ_i	3.9
Permittivity of free space	ε_o	$8.855x10^{-14}$f/cm
Electronic charge	q	$1.6x10^{-19}$coul
Thermal voltage	kT/q	0.0258 volt

It is recognized that the mobility of inversion layer carriers in a MOSFET can be a function of both the gate and drain applied voltages. However, the present study aimed at modeling the basic conduction mechanism in this semiconductor device. For this reason, in the following calculations we assumed a constant inversion carrier mobility. Moreover, substantial justification exists for taking this approach. First, at present there is no rigorous theory for quantitatively evaluating the mobility of these inversion carriers [64]. Second, it has been shown that previous attempts to fit MOSFET theory to experiment by arbitrarily varying inversion carrier mobility have failed to account for all observed discrepancies [44]. Third, our approach of assuming constant carrier mobility permitted us to isolate and identify individual physical mechanisms responsible for "short-channel effects" in MOS transistors (see section 2.0). Without this degree of simplification, it is doubtful that the characteristics and significance of these mechanisms could be accurately assessed.

---

[64] D. P. Kennedy, *Computer Aided Engineering of Semiconductor Integrated Circuits*, Contract No. DAAB07-75-C-1344, Interim Report No. ECOM-75-1344-1 (1976).

5.1 MOSFET Volt-Ampere Calculations

To illustrate the relative merits of our revised MOSFET theory, first we will show comparisons between two other widely used MOSFET theories and our two-dimensional computer calculations. For strong inversion operation, we selected a treatment of traditional MOSFET theory given by Sze [9]. From this treatment, the volt-ampere relation of a MOSFET has the form

$$I_D = \frac{W}{L}\,\mu C_{ox}\left\{\left(V_G - 2\phi_F - \frac{V_D}{2}\right)V_D - \frac{3}{2}\frac{\sqrt{2\kappa_s\varepsilon_o qN_A}}{C_{ox}}\left[\left(V_D + 2\phi_F\right)^{3/2} - \left(2\phi_F\right)^{3/2}\right]\right\} . \quad (5.1)$$

This expression is applicable to the triode mode of operation, i.e., for applied drain voltages less than that required for electric current saturation. This drain saturation voltage (or "pinch-off" voltage) is given by

$$V_{Dsat} = V_G - 2\phi_F + \frac{\kappa_s\varepsilon_o qN_A}{C_{ox}^2}\left\{1 - \sqrt{1 + \frac{2C_{ox}^2 V_G}{\kappa_s\varepsilon_o qN_A}}\right\}, \quad (5.2)$$

where, for simplicity, we have assumed a zero flatband voltage [8].

For the saturation mode of operation, this volt-ampere relation has the form

$$I_D = I_{Dsat}\cdot\left(\frac{L}{L-\Delta L}\right) , \quad (5.3)$$

where

$$I_{Dsat} \equiv I_D(V_D=V_{Dsat}) \quad (5.4)$$

$$\Delta L \equiv \sqrt{\frac{2\kappa_s\varepsilon_o}{qN_A}\left(V_D - V_{Dsat}\right)} . \quad (5.5)$$

In Eq. 5.5, ΔL represents a reduction in electrical channel length presumed to arise from channel pinch-off -- a concept generally referred to as "channel-length modulation" [11].

This traditional MOSFET theory is based upon an assumption that all source-drain electric current, I_D, arises from drift. This assumption is reasonable for strong inversion operation of a MOSFET. Contrasting with this situation, in the weak inversion mode of operation I_D arises dominantly from thermal diffusion and, therefore, traditional MOSFET theory is inapplicable to this mode.

A general one-dimensional theory for this semiconductor device was proposed by Pao and Sah [21]. This general theory includes both drift and diffusion mechanisms of carrier transport, but requires a numerical solution to calculate the source-drain current, I_D. To achieve computational simplicity, several authors have obtained regional closed-form approximations for I_D from this Pao-Sah formulation [31,37,46,47]. From these approximate solutions, we selected the weak inversion theory proposed by Troutman and Chakravarti [37].

In this weak inversion theory, the volt-ampere relation of a MOSFET has the form

$$I_D = WL_D qD \frac{\Omega n_i}{L} [1-\exp(-qV_D/kT)] \quad , \tag{5.6}$$

where

$$\Omega \equiv \frac{\exp[q(b-1)\phi_F/kT]}{\{2[(qb\phi_F/kT)-1]\}^{1/2}} - \frac{1}{\{2[(q\phi_F/kT)-1]\}^{1/2}} \tag{5.7}$$

and L_D is the extrinsic Debye length in this semiconductor material:

$$L_D \equiv \left(\frac{\kappa_s \varepsilon_o kT}{q^2 N_A}\right)^{1/2} \quad . \tag{5.8}$$

In Eq. 5.7, the parameter b represents the band bending near the source end of this structure. This parameter can be calculated from a one-dimensional numerical solution of Poisson's equation (Eq. 4.7), for which these author's [37] proposed the following approximate form:

$$V_G = b\phi_F + \frac{\sqrt{2\kappa_s \varepsilon_o q N_A (b\phi_F)}}{C_{ox}} \quad . \tag{5.9}$$

For a wide range of device structures, Figs. 1.31-1.36 show MOSFET transfer characteristics calculated from Eq. 5.3 and Eq. 5.6 in comparison with our two-dimensional computer calculations of these characteristics. For qualitative purposes, in these figures we have shown the theoretical gate threshold voltage for each structure calculated from elementary MOSFET theory [8].

In Fig. 1.31, we have the calculated transfer characteristics of a lightly-doped long-channel MOSFET ($N_A = 2x10^{15}cm^{-3}$, $L = 10\mu m$). This figure shows that, in combination, these previous MOSFET theories yield an adequate prediction of the volt-ampere characteristics of this semiconductor structure. However, Fig. 1.31 also makes clear a difficulty in this approach to calculating MOSFET characteristics: neither of these previous theories is applicable over this entire range of assumed gate voltages and, as a result, we have a matching problem in the vicinity of the gate threshold voltage (V_T).

Figures 1.32-1.36 show similar comparisons between these previous MOSFET theories and our two-dimensional computer solution. For this assumed range of gate voltages, as source-drain channel length shrinks these figures show that Eq. 5.3 attains somewhat improved agreement with our rigorous calculations for the strong inversion mode; in contrast, for weak inversion operation, agreement between Eq. 5.6 and these computer calculations worsens. In addition, for the higher range of substrate doping ($N_A = 2x10^{16}cm^{-3}$), we encounter increased difficulty in matching these two previous theories near $V_G=V_T$ (see Figs. 1.34-1.36).

For the previously assumed device structures, Figs. 1.37-1.42 show MOSFET transfer characteristics calculated from the present theory (Eq. 4.38) in comparison with our two-dimensional computer calculations of these characteristics. These figures show that our revised MOSFET theory yields satisfactory agreement with these

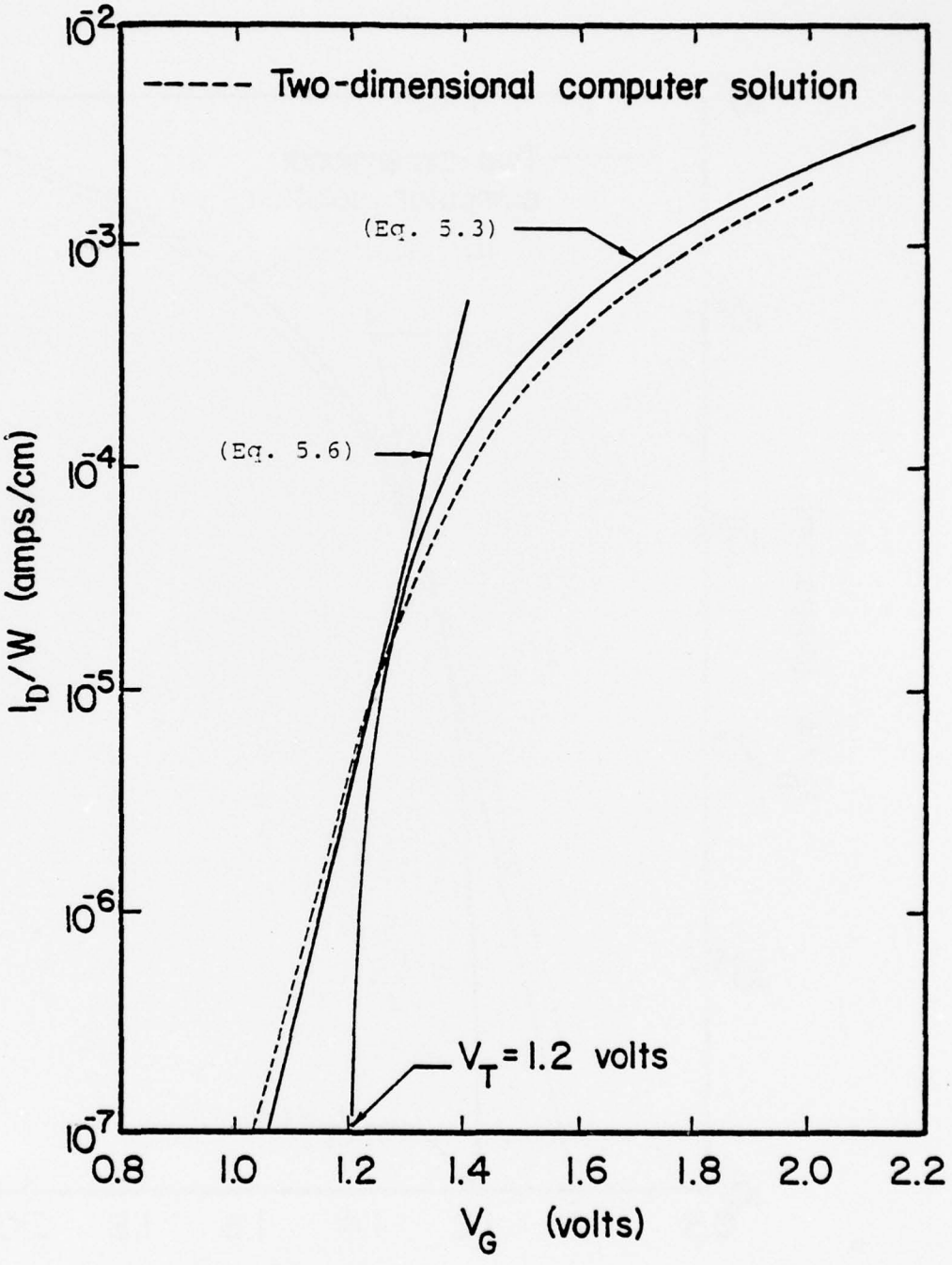


Fig. 1.31. Calculated transfer characteristic of a long-channel MOSFET. $N_A = 2 \times 10^{15} cm^{-3}$, t_{ox} = 1000A°, L = 10um. V_D = 2 volts.

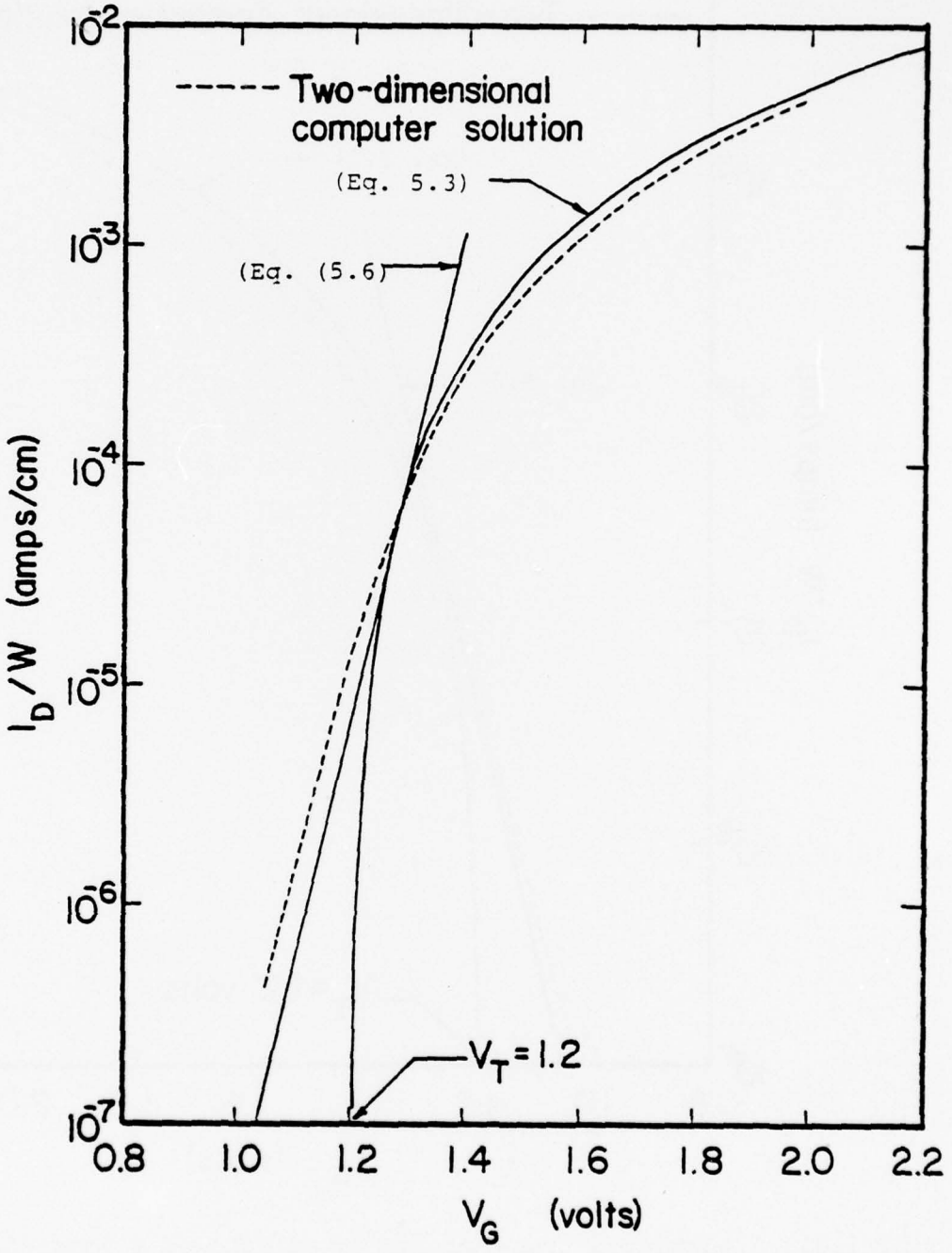


Fig. 1.32. Calculated transfer characteristic of a MOSFET. $N_A = 2 \times 10^{15} cm^{-3}$, $t_{ox} = 1000A°$, $L = 5um$. $V_D = 2$ volts.

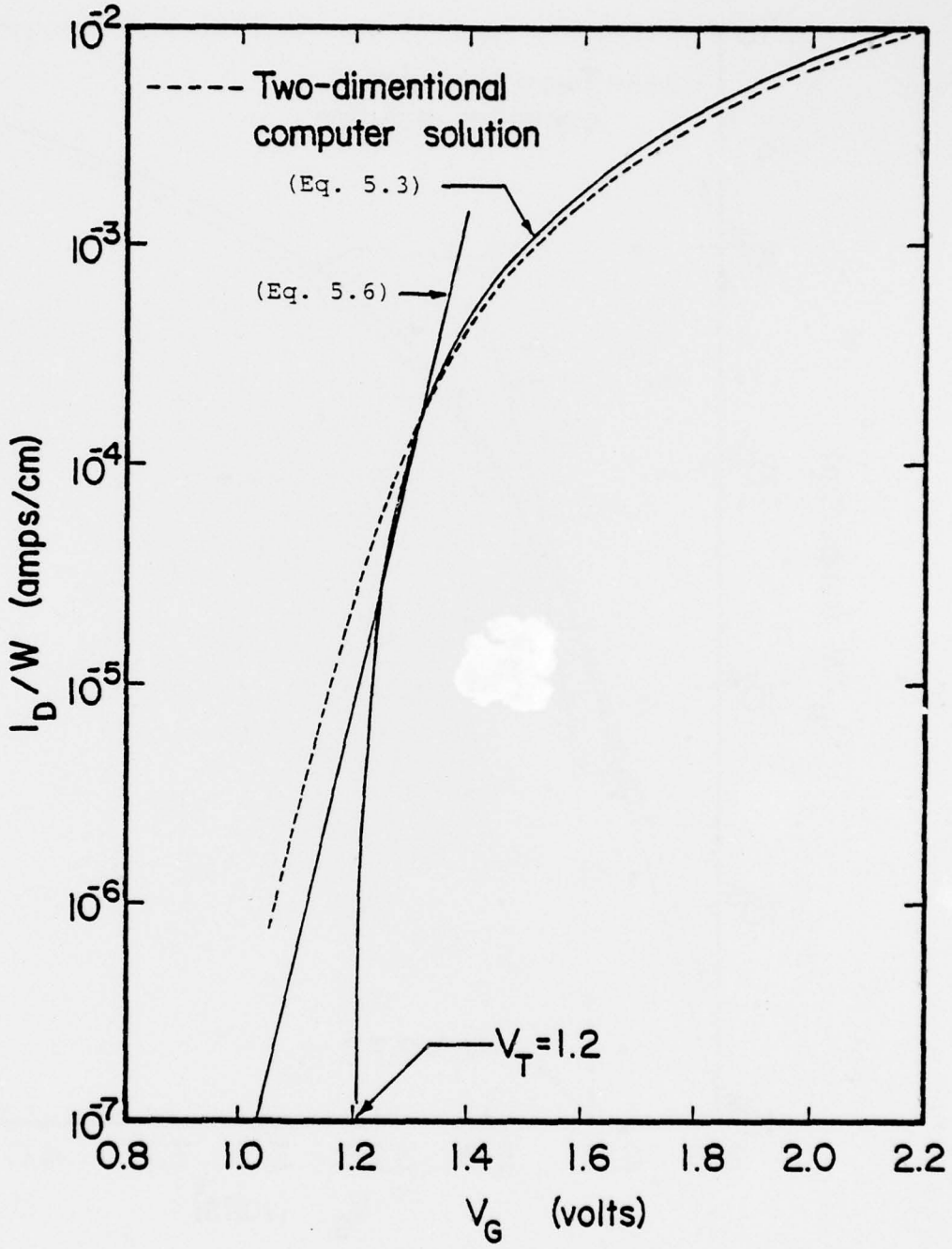


Fig. 1.33. Calculated transfer characteristic of a short-channel MOSFET. $N_A = 2 \times 10^{15} cm^{-3}$, t_{ox} = 1000A°, L = 4μm. V_D = 2 volts.

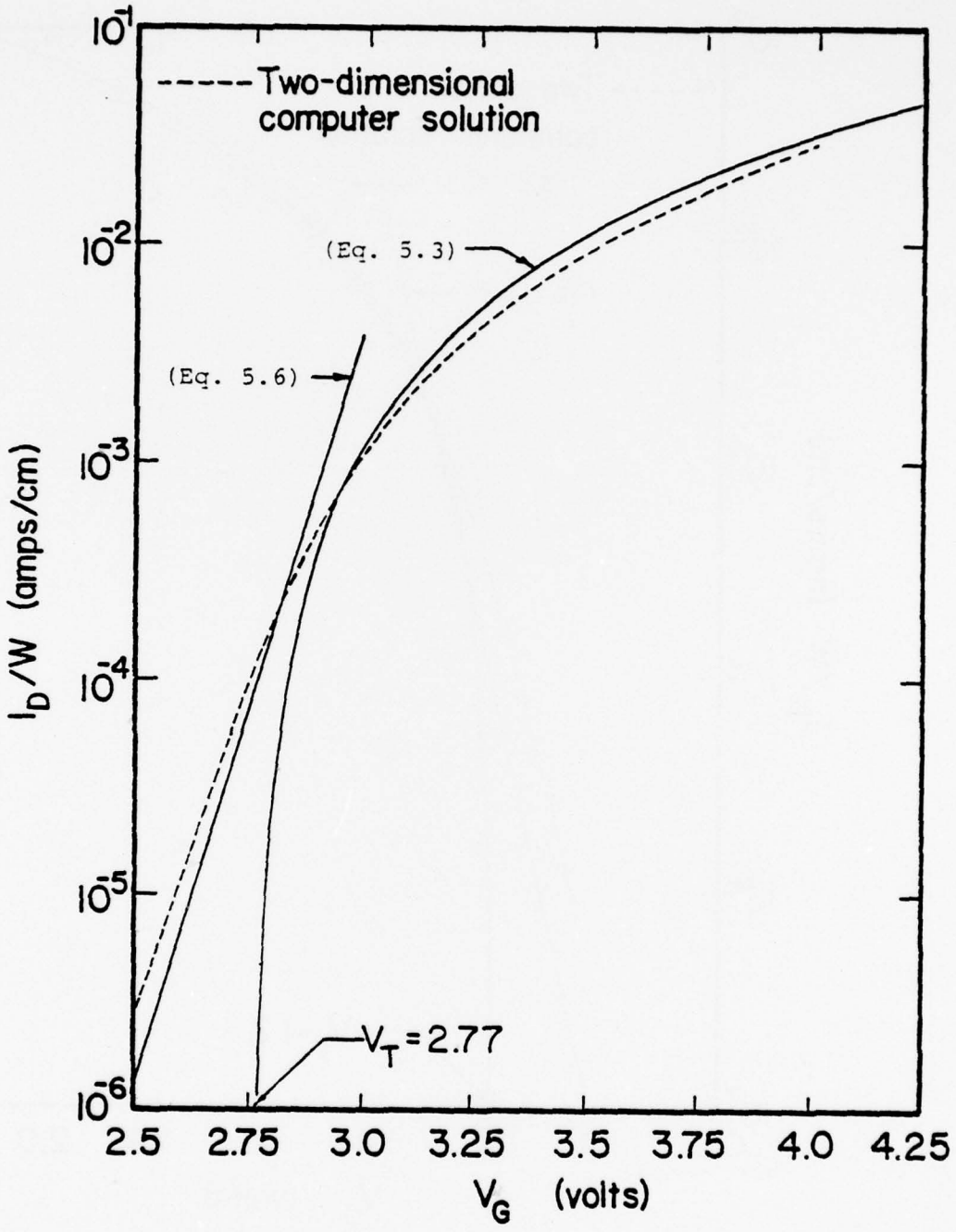


Fig. 1.34. Calculated transfer characteristic of a long-channel MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, $L = 10\mu m$. $V_D = 4$ volts.

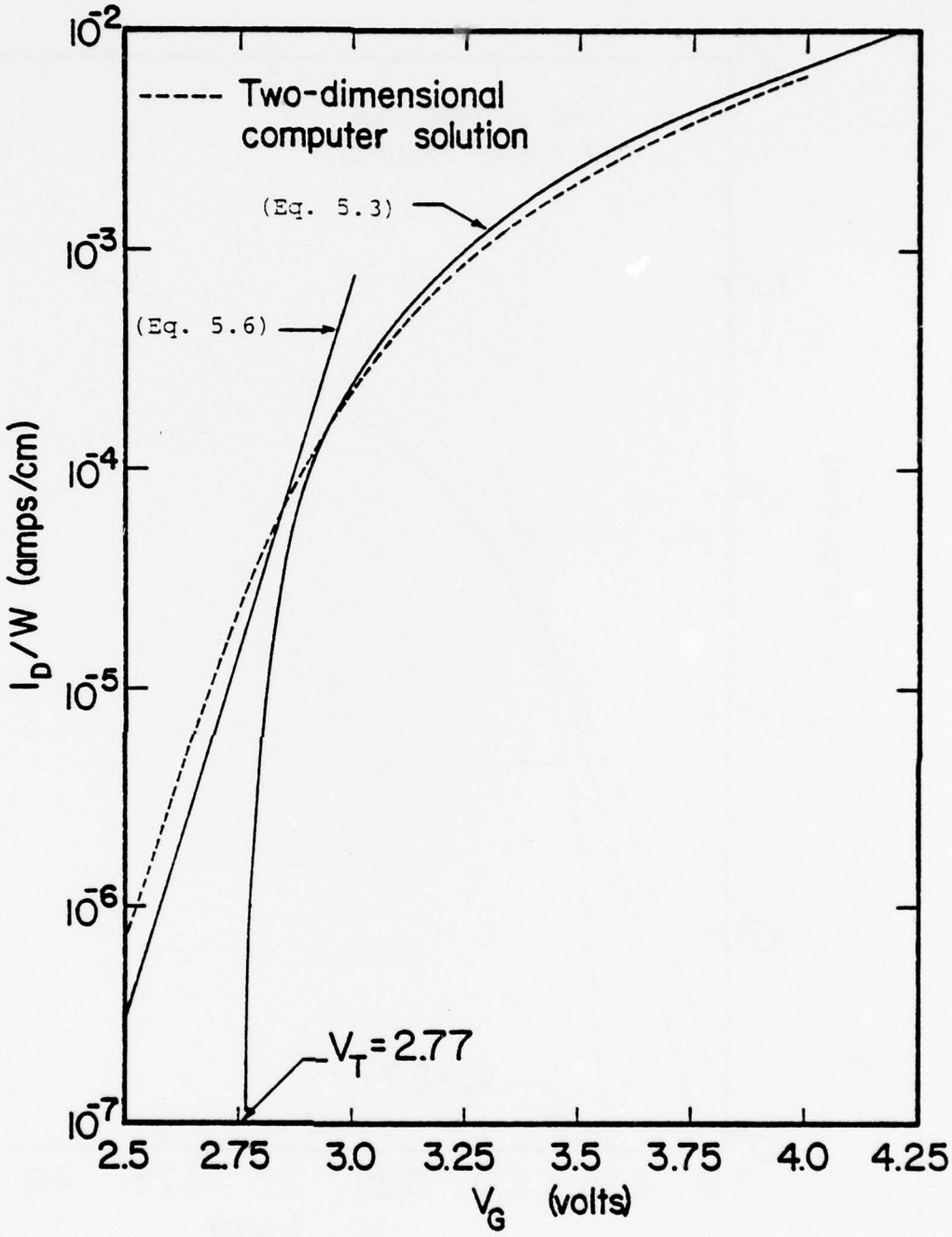


Fig. 1.35. Calculated transfer characteristic of a MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, $L = 5\mu m$. $V_D = 4$ volts.

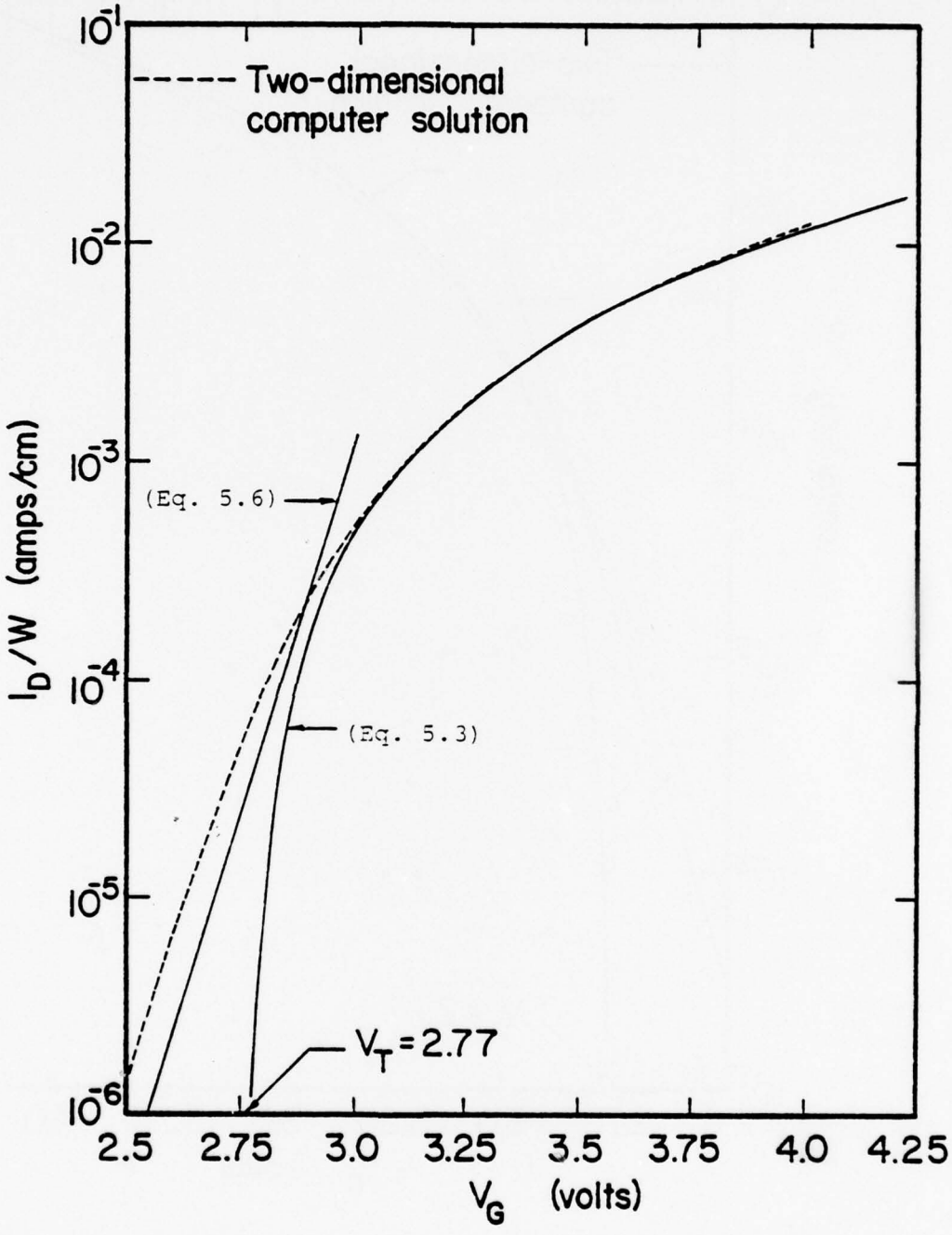


Fig. 1.36. Calculated transfer characteristic of a short-channel MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, $L = 3\mu m$. $V_D = 4$ volts.

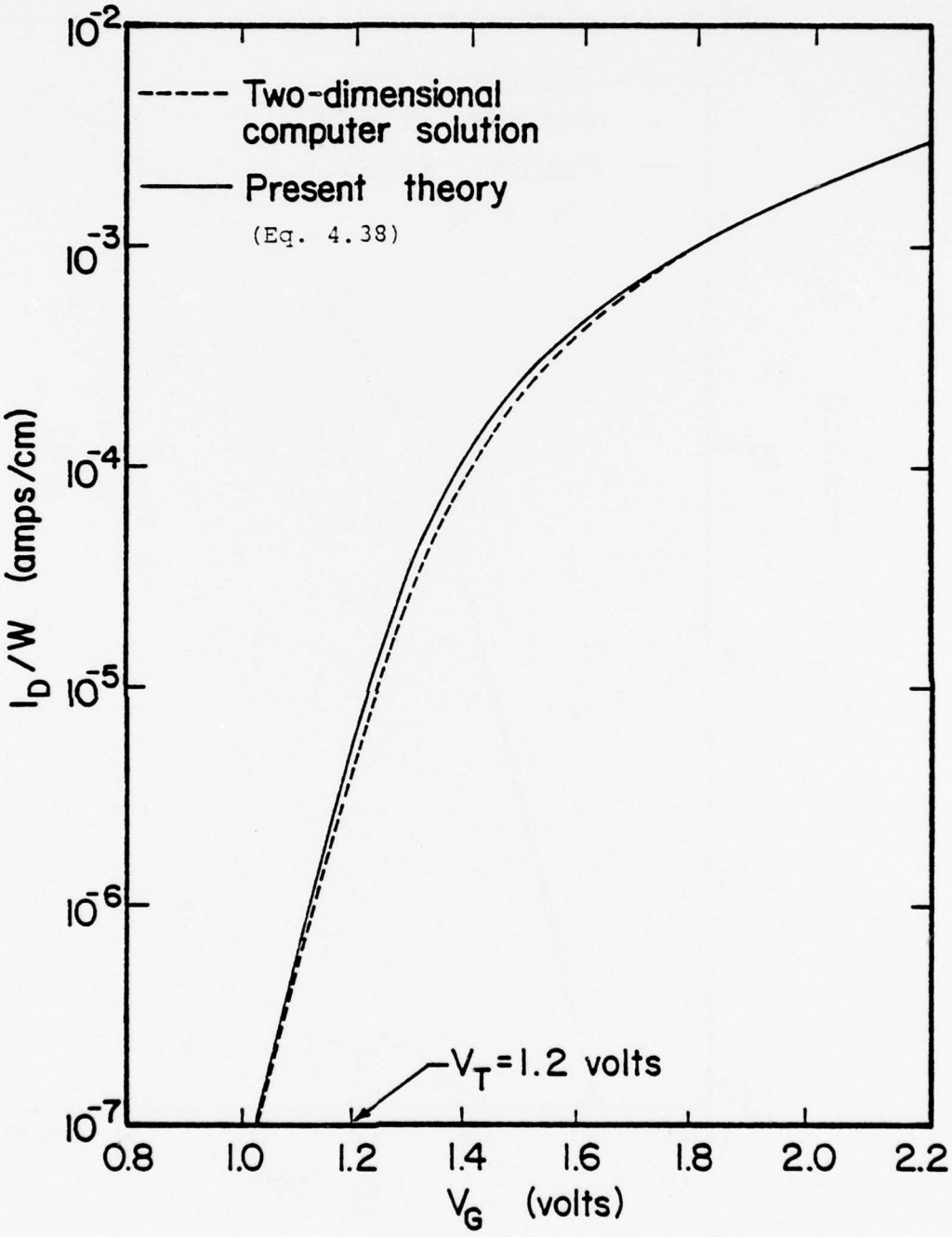


Fig. 1.37. Calculated transfer characteristic of a long-channel MOSFET. $N_A = 2 \times 10^{15} cm^{-3}$, t_{ox} = 1000A°, L = 10um. V_D = 2 volts.

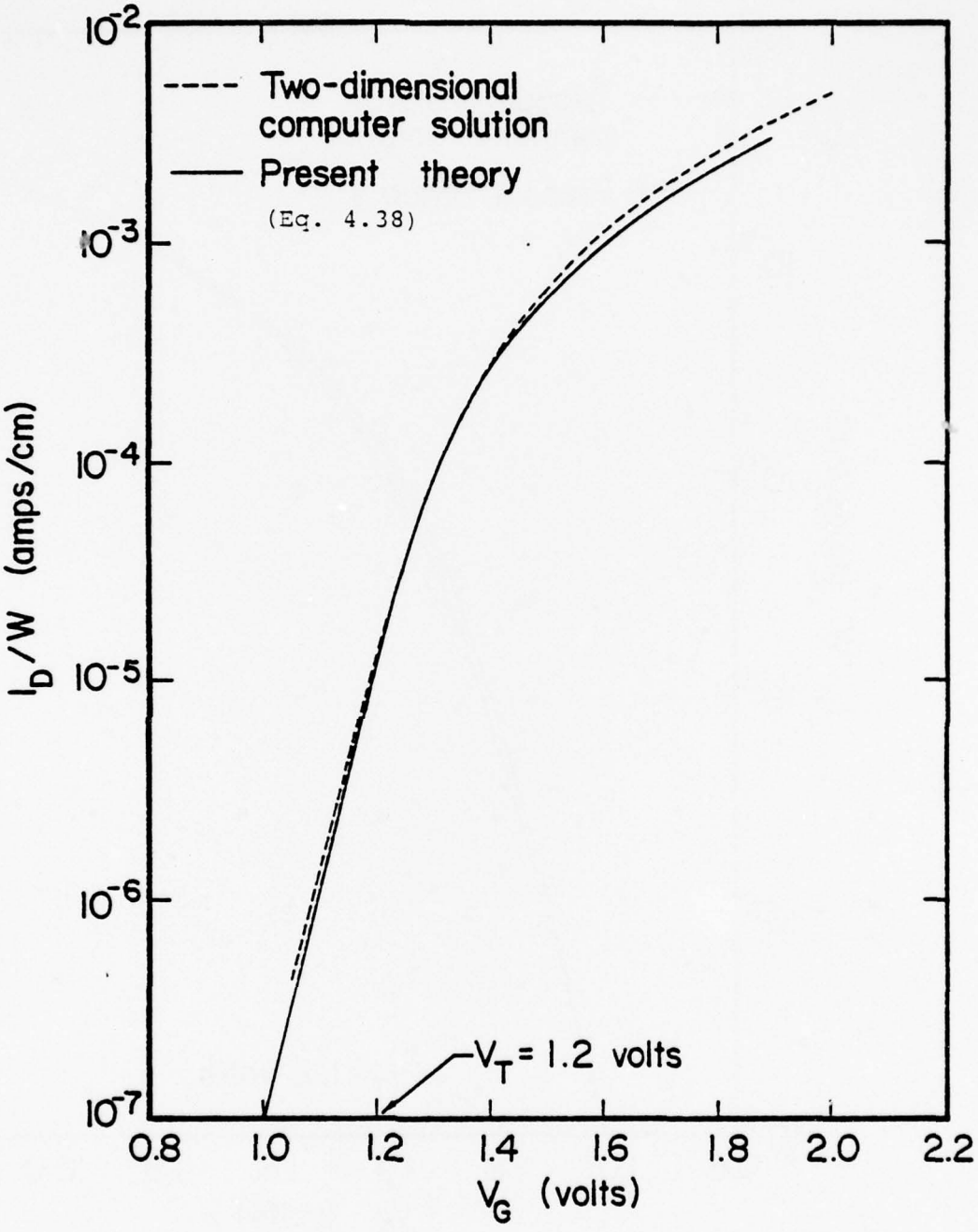


Fig. 1.38. Calculated transfer characteristics of a MOSFET. $N_A = 2 \times 10^{15} cm^{-3}$, $t_{ox} = 1000 A^{\circ}$, $L = 5 \mu m$. $V_D = 2$ volts.

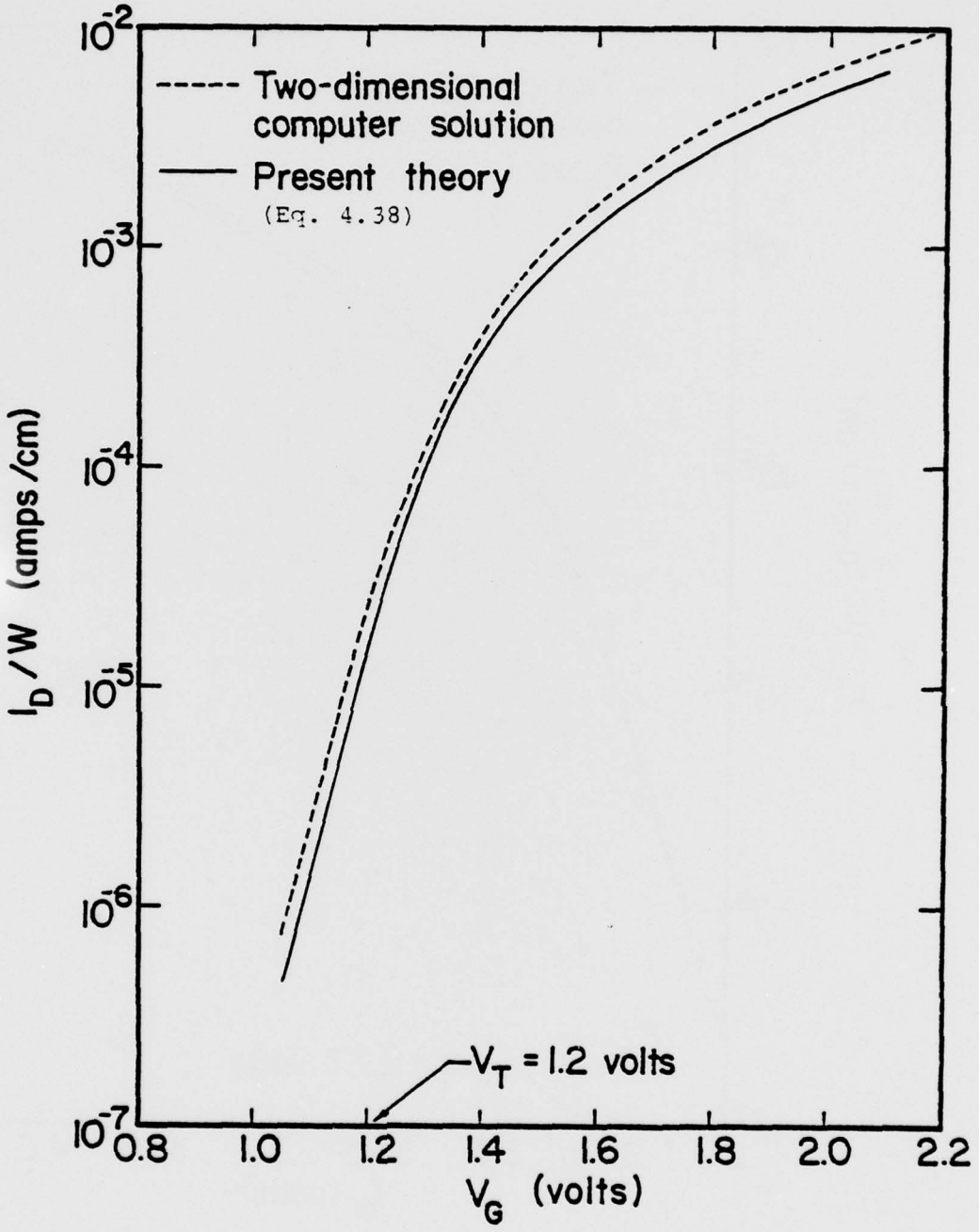


Fig. 1.39. Calculated transfer characteristic of a short-channel MOSFET. $N_A = 2 \times 10^{15} cm^{-3}$, t_{ox} = 1000A°, L = 4μm. V_D = 2 volts.

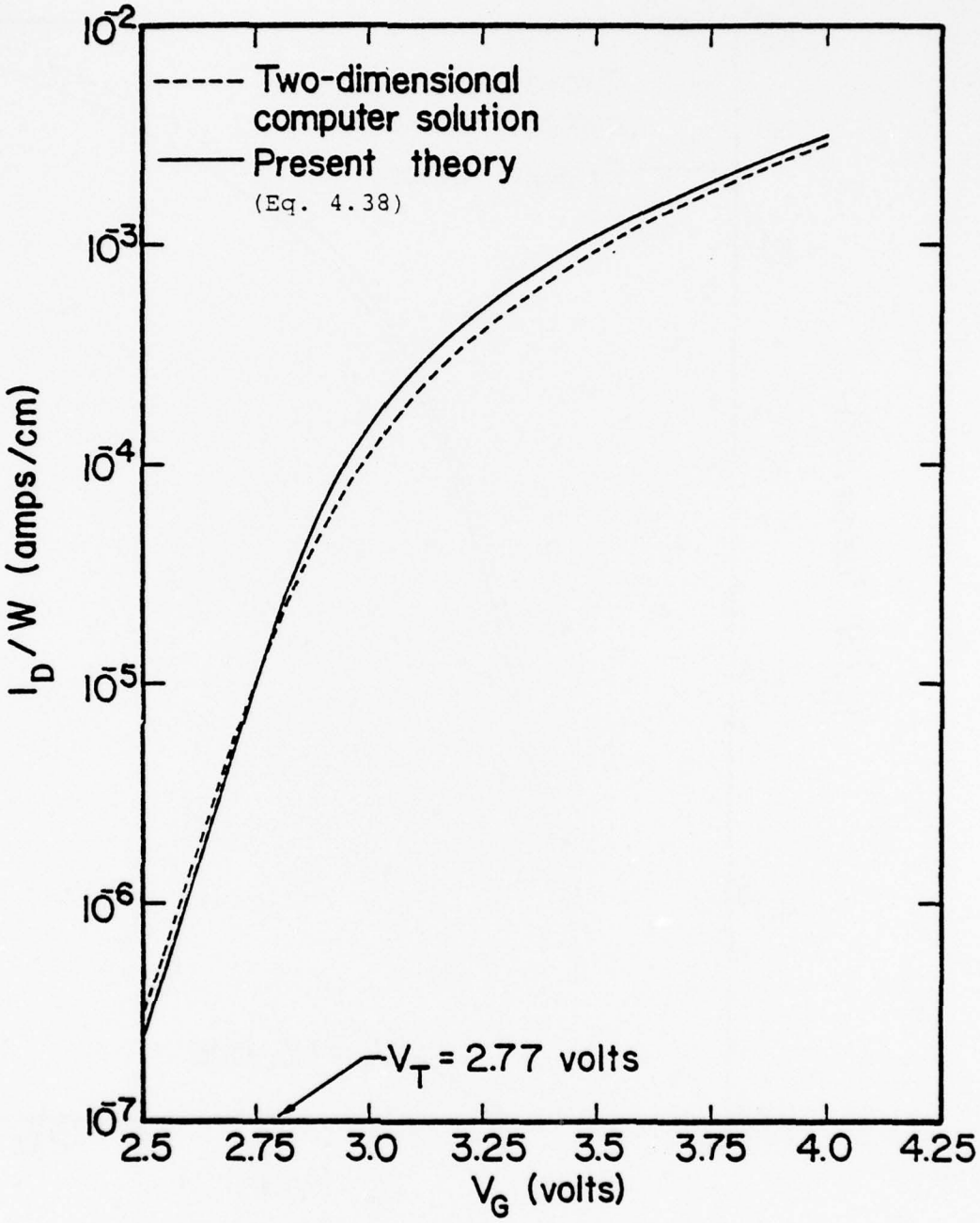


Fig. 1.40. Calculated transfer characteristic of a long-channel MOSFET. $N_A = 2 \times 10^{15} cm^{-3}$, t_{ox} = 1000A°, L = 10μm. V_D = 4 volts.

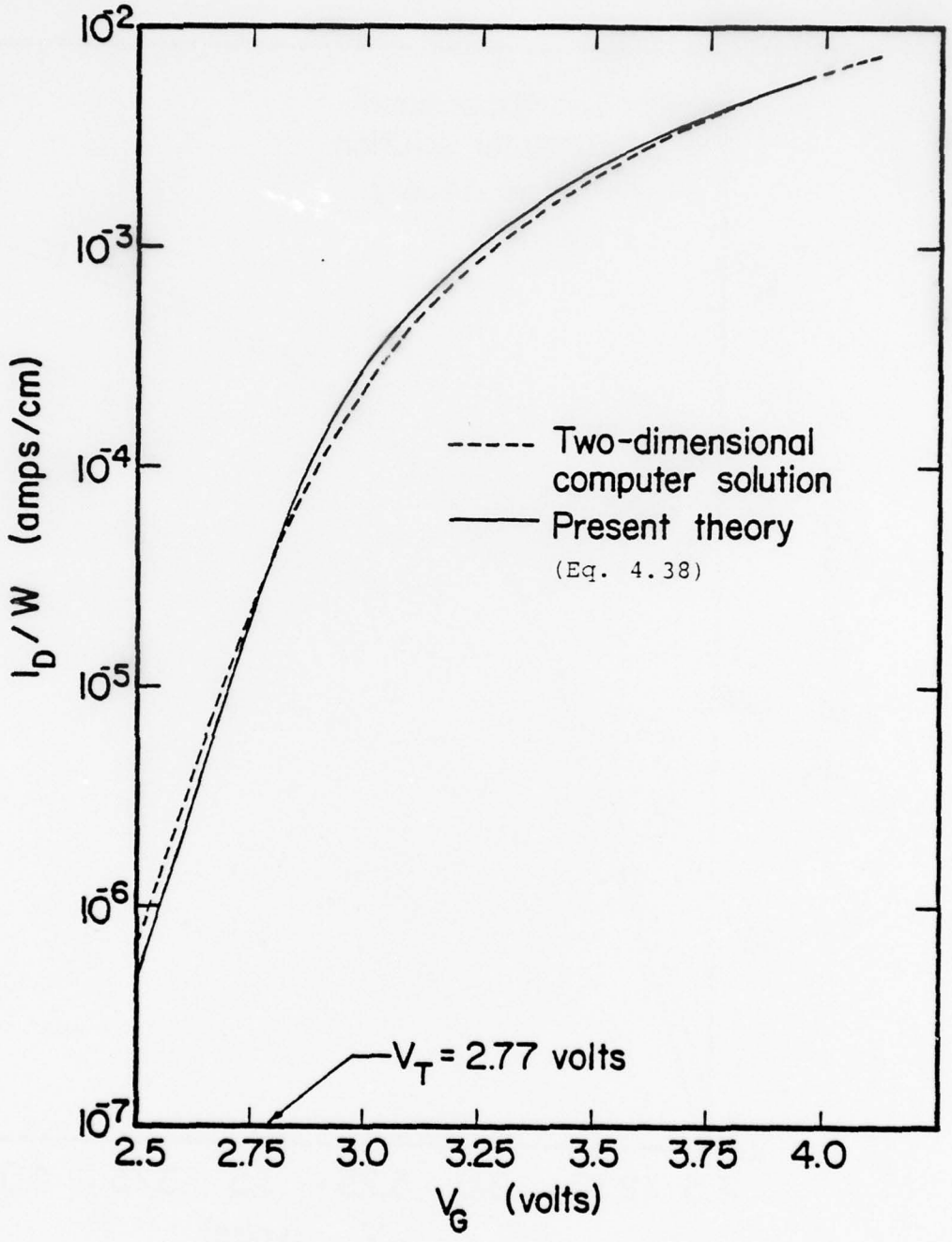


Fig. 1.41. Calculated transfer characteristic of a MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, $L = 5\mu m$. $V_D = 4$ volts.

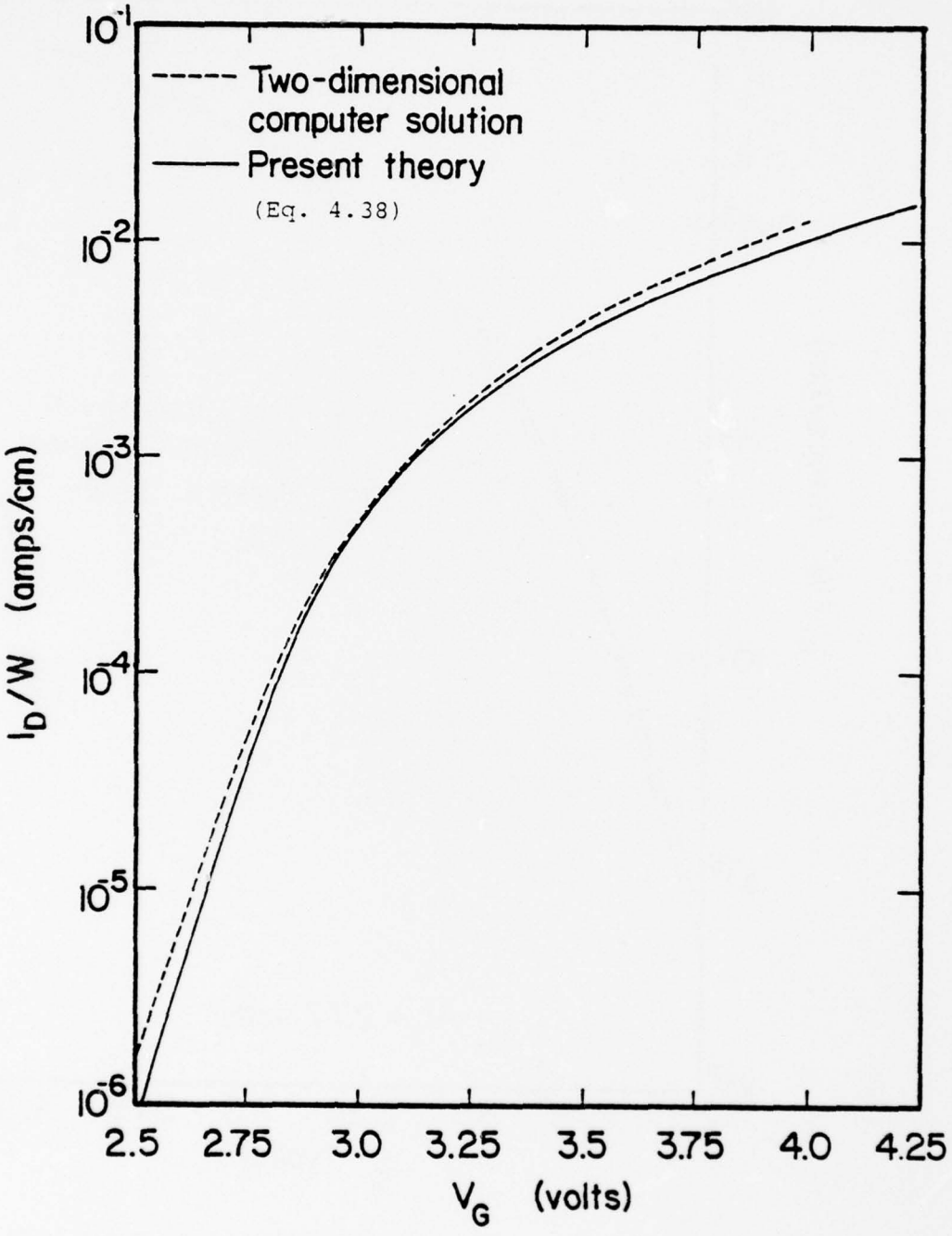


Fig. 1.42. Calculated transfer characteristic of a short-channel MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, $L = 3\mu m$. $V_D = 4$ volts.

computer calculations over many decades of source-drain current. Moreover, throughout both the weak and strong inversion modes of operation, this agreement was obtained using a single volt-ampere relation (Eq. 4.38); thus, we eliminate the matching problem (where $V_G=V_T$) encountered with previous regional models for MOSFET operation.

To further demonstrate the applicability of our revised MOSFET theory, sample drain-characteristic calculations are shown in Fig. 1.43 (for strong inversion operation) and in Fig. 1.44 (for weak inversion operation). The agreement shown in these figures is considered adequate for engineering purposes. In addition, Fig. 1.44 graphically illustrates that our model correctly predicts an exponential type of current saturation for weak inversion operation of this semiconductor device.

5.2 Short-Channel Effects

When applied to lightly doped structures ($N_A = 2x10^{15}cm^{-3}$), our revised MOSFET theory produces significant error for channel lengths $L \lesssim 4\mu m$ (see Fig. 1.39). For devices in the higher range of substrate doping ($N_A = 2x10^{16}cm^{-3}$), we observe a similar limitation for channel lengths $L \lesssim 3\mu m$ (Fig. 1.42). The foregoing observations suggest that the major discrepancy between our MOSFET theory and these computer calculations arises from two-dimensional mechanisms encountered in short-channel structures.

In our MOSFET theory (Eq. 4.27), this discrepancy could arise from our calculation of Q_{io}, $(dQ_i/dV_s)_{V_o}$, and/or E_o. To check these possible sources of error, for a short-channel device ($L=3\mu m$) we will compare these one-dimensional calculations with our two-dimensional computer solution.

Figure 1.45 shows calculated drain characteristics for this short-channel device ($L=3\mu m$). From Fig. 1.45, we note that our theory yields satisfactory results for this structure when applied to the triode mode of operation ($V_D = 0.5$ volts). In contrast,

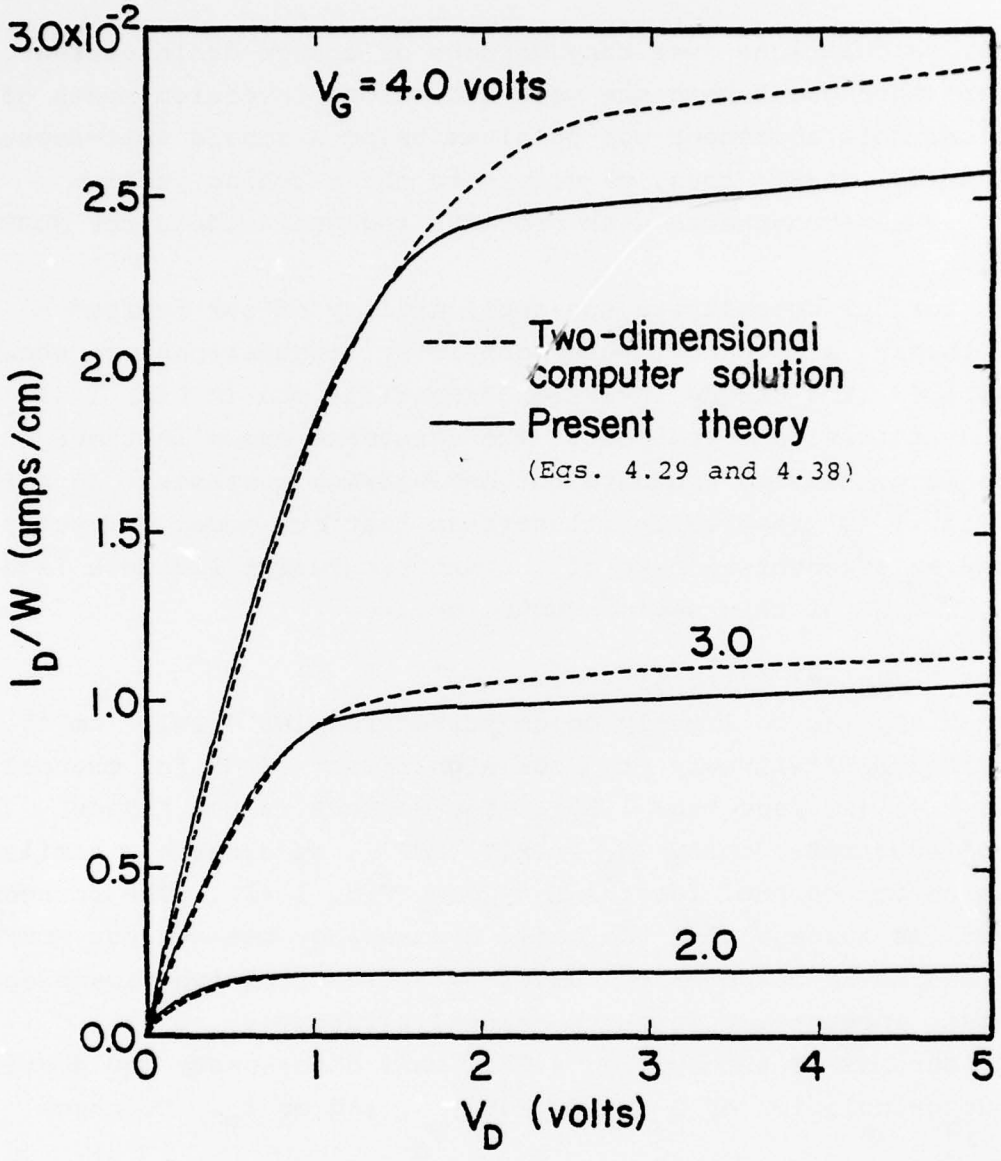


Fig. 1.43. Calculated drain characteristics of a long-channel MOSFET operating in strong inversion. $N_A = 2 \times 10^{15} cm^{-3}$, $t_{ox} = 1000A°$, L = 10μm.

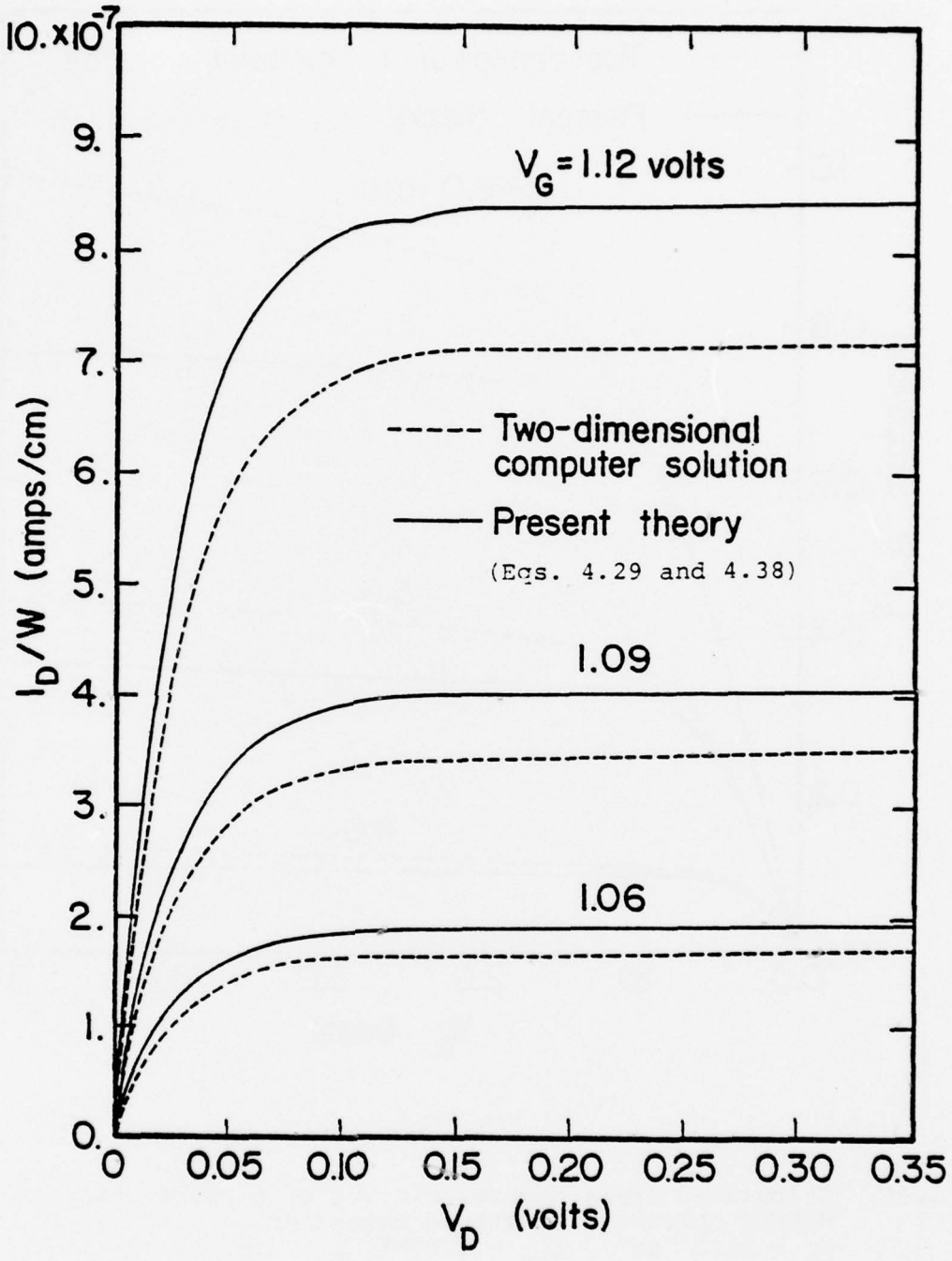


Fig. 1.44. Calculated drain characteristics of a long-channel MOSFET operating in weak inversion. $N_A = 2 \times 10^{15} cm^{-3}$, $t_{ox} = 1000A°$, $L = 10\mu m$.

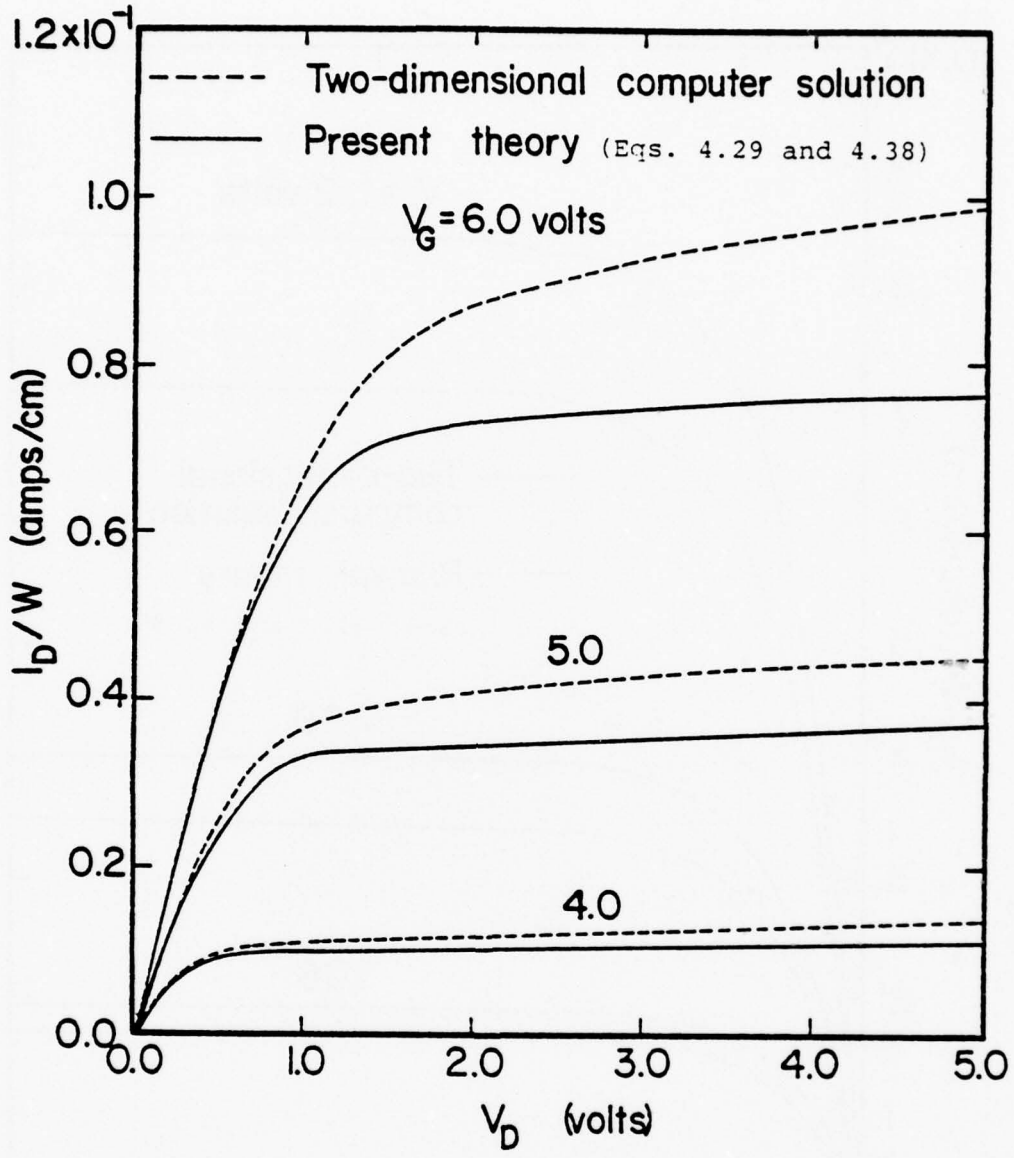


Fig. 1.45. Calculated drain characteristics of a short-channel MOSFET operating in strong inversion. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, $L = 3\mu m$.

when applied to the saturation mode (V_D = 2.0 volts), Fig. 1.45 shows that this theory produces a significant error. Corresponding to these bias conditions, Fig. 1.46 shows calculated distributions of inversion charge, $Q_i(V_s)$, from our approximate relation, Eq. 4.5, and from our two-dimensional computer solution. At values of V_s near the source and drain metallurgical junctions, our two-dimensional calculations show a rapid increase in Q_i arising from electrostatic interactions between these junctions and the source-drain inversion layer. As previously discussed in section 4.0, these interaction regions constitute infinite sinks for the mobile carrier flux and, therefore, they are not included in our one-dimensional theory. Within the active source-drain channel Fig. 1.46 shows substantial agreement between our approximation for $Q_i(V_s)$, Eq. 4.5, and this rigorous computer solution. Our one-dimensional relations for Q_{io} and $(dQ_i/dV_s)_{V_o}$ (Eqs. 4.6 and 4.11, respectively) were derived from Eq. 4.5 and, therefore, we would not expect these relations to contribute significantly to this error in I_D.

In contrast with the foregoing situation, Fig. 1.47 shows that (for V_D = 2.0 volts) we have a significant error in our one-dimensional calculation of $V_s(x)$. From Fig. 1.47, our theory obtains excellent agreement with a rigorous computer calculation of $V_s(x)$ for an applied drain-source bias voltage V_D = 0.5 volts, but produces a substantial error for V_D = 2.0 volts. Fig. 1.48 shows these calculations of $V_s(x)$ on a magnified scale near the source end of this semiconductor structure. From Fig. 1.48, for V_D = 2.0 volts we have an error in the slope dV_s/dx predicted by our theory and, hence, in our calculation of E_o.

This error in E_o could arise from two related two-dimensional mechanisms: 1) reach-through of the drain-induced electric field and 2) gate-drain ion-sharing. Figure 1.49 shows two-dimensional calculations of the potential barrier near the source junction in this MOSFET structure. From Fig. 1.49, reach-through of the source-

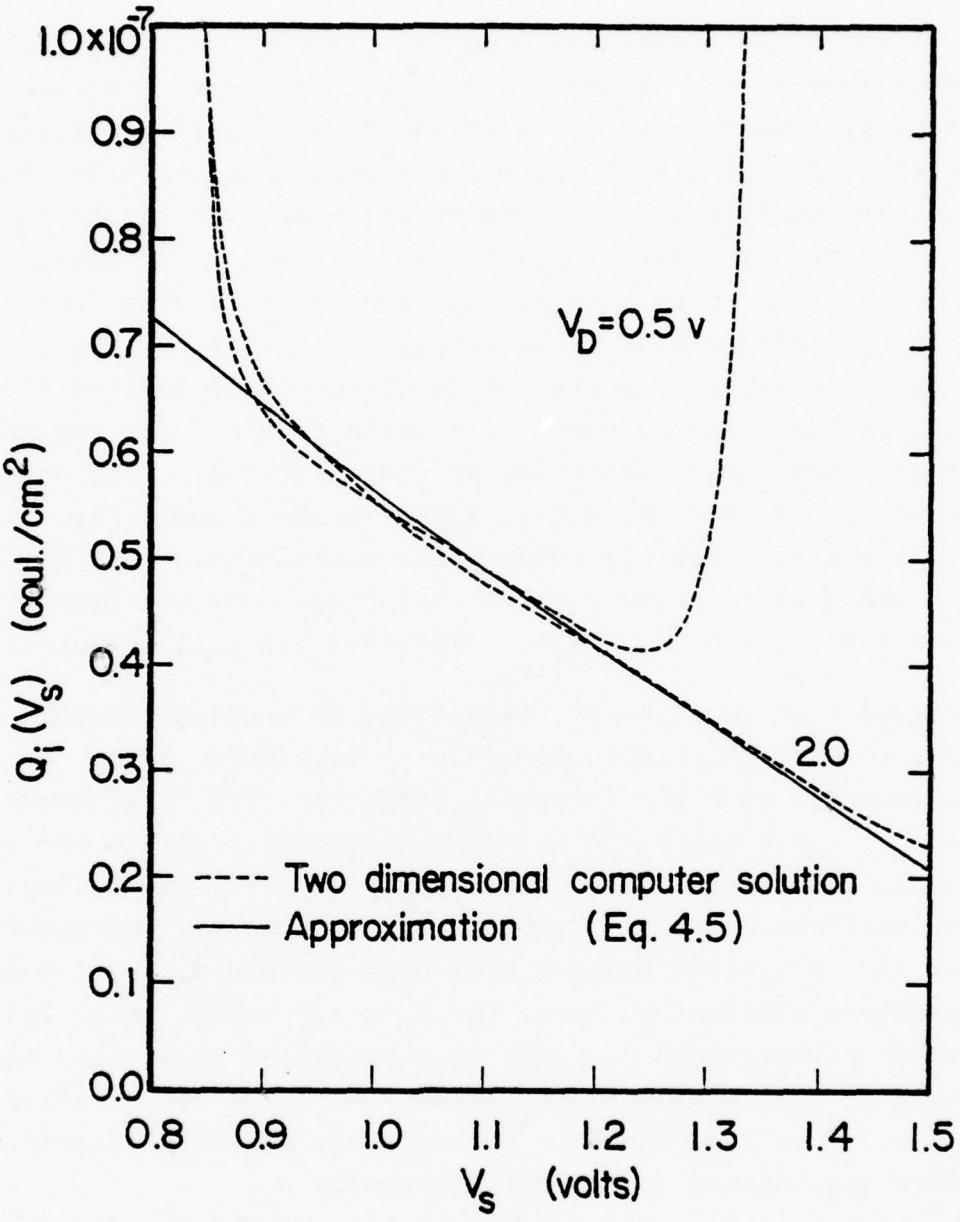


Fig. 1.46. Calculated inversion charge distribution, $Q_i(V_s)$, in a short-channel MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, $L = 3\mu m$. $V_G = 5$ volts.

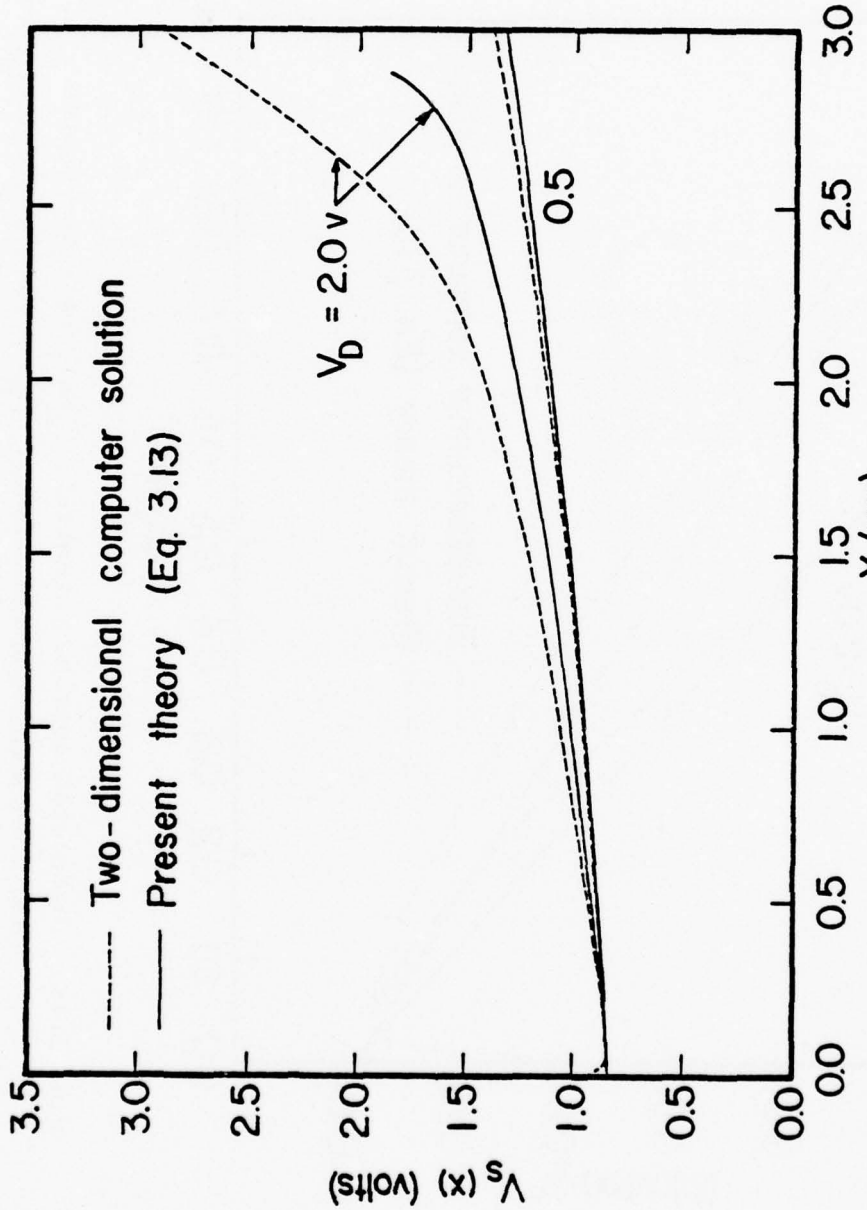


Fig. 1.47. Calculated surface potential, $V_s(x)$, in a short-channel MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, L = 3μm. V_G = 5 volts.

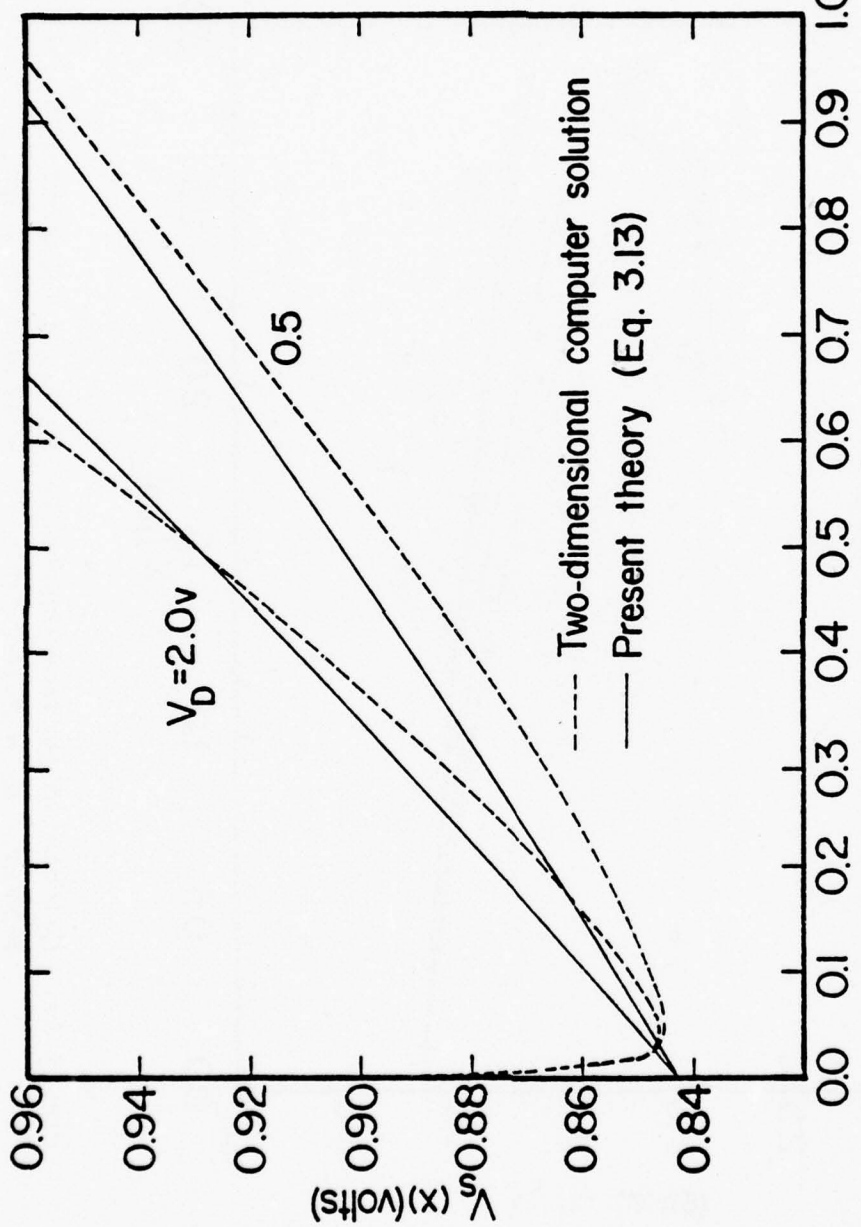


Fig. 1.48. Calculated surface potential, $V_s(x)$, near the source end of a short-channel MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, $V_G = 5$ volts.

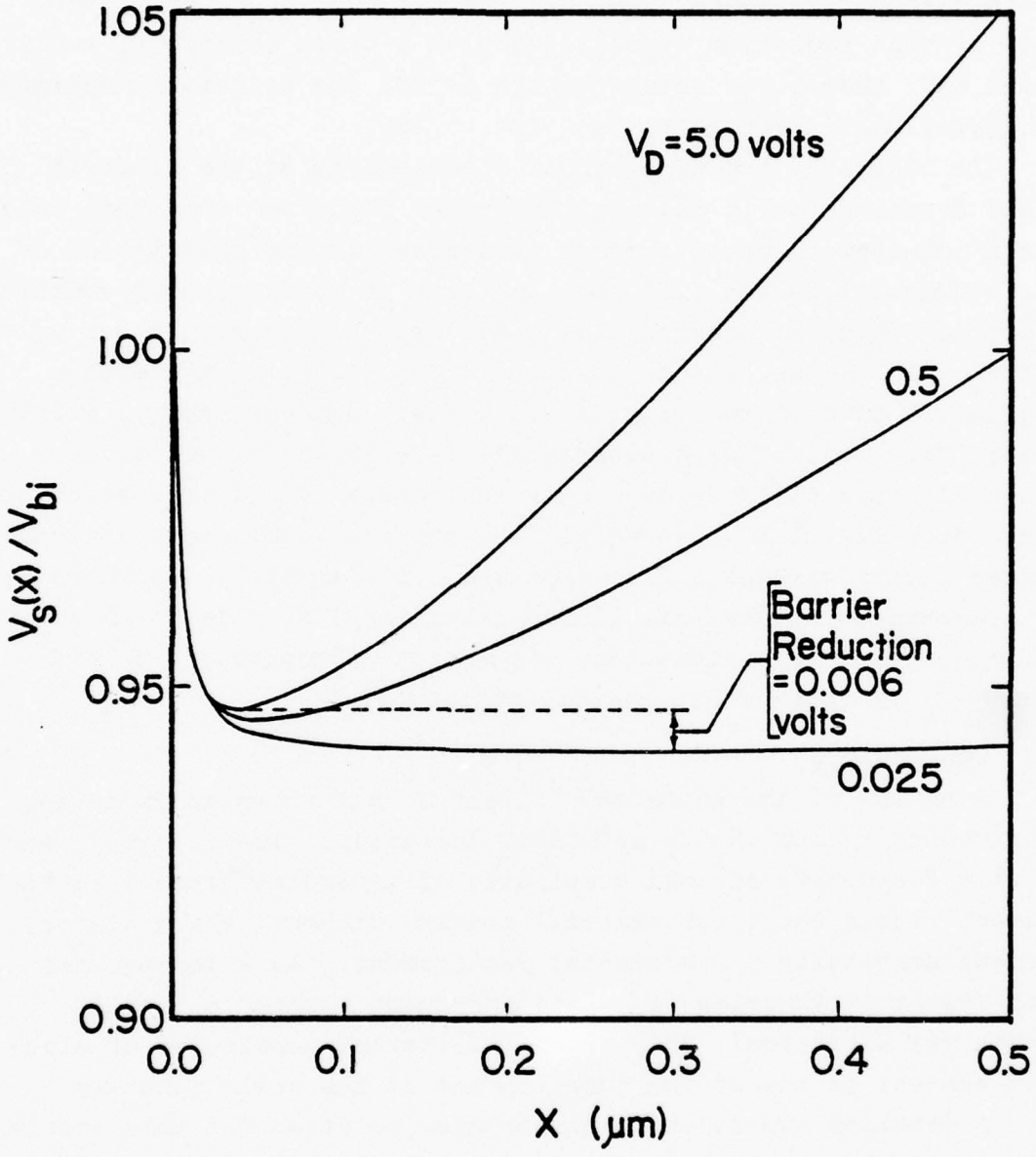


Fig. 1.49. Source-junction potential-barrier reduction (arising from reach-through) in a short-channel MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, t_{ox} = 1000A°, L = 3μm. V_G = 5 volts.

drain electric field reduces this barrier by approximately 6mV. Most of this reduction is realized with a drain voltage V_D = 0.5 volts and, therefore, reach-through is not the principle mechanism producing this error in E_o (see Fig. 1.48).

In contrast, two-dimensional calculations of the electric field distribution in this semiconductor structure show that gate-drain ion-sharing could produce this error in our calculation of E_o. Figures 1.50 and 1.51 show contours of constant source-drain electric field, E_x, corresponding to drain voltages V_D = 0.5 volts and V_D = 2.0 volts, respectively. In Fig. 1.50, E_x exhibits a negligible divergence for V_D = 0.5 volts. However, for V_D = 2.0 volts, Fig. 1.51 shows a substantial divergence of this source-drain electric field -- even near the source end of this structure. Thus, from Fig. 1.51, gate-drain ion-sharing within this semiconductor device creates a situation where E_x cannot be described by an elementary one-dimensional approximation (Eq. 4.18). As a consequence of this situation, our theory underpredicts E_o and, hence, I_D in these short-channel structures.

6.0 Conclusions

A review of the technical literature has shown an important shortcoming in our theory of MOSFET operation. Specifically, workers in this field have assumed continuity of inversion layer electric current within their mathematical models, without making electric current continuity a fundamental requirement. As a consequence, most available theories of MOSFET operation predict a totally unphysical situation: they predict different magnitudes of electric current at the source junction and at the drain junction.

A detailed two-dimensional computer solution for this problem shows that electric current continuity does, indeed, exist in the inversion layer of a MOSFET -- as it must, from a physical point of view. Further, this computer solution shows that both drift and diffusion of inversion layer carriers are important components of the associated transport process. The relative importance of drift and

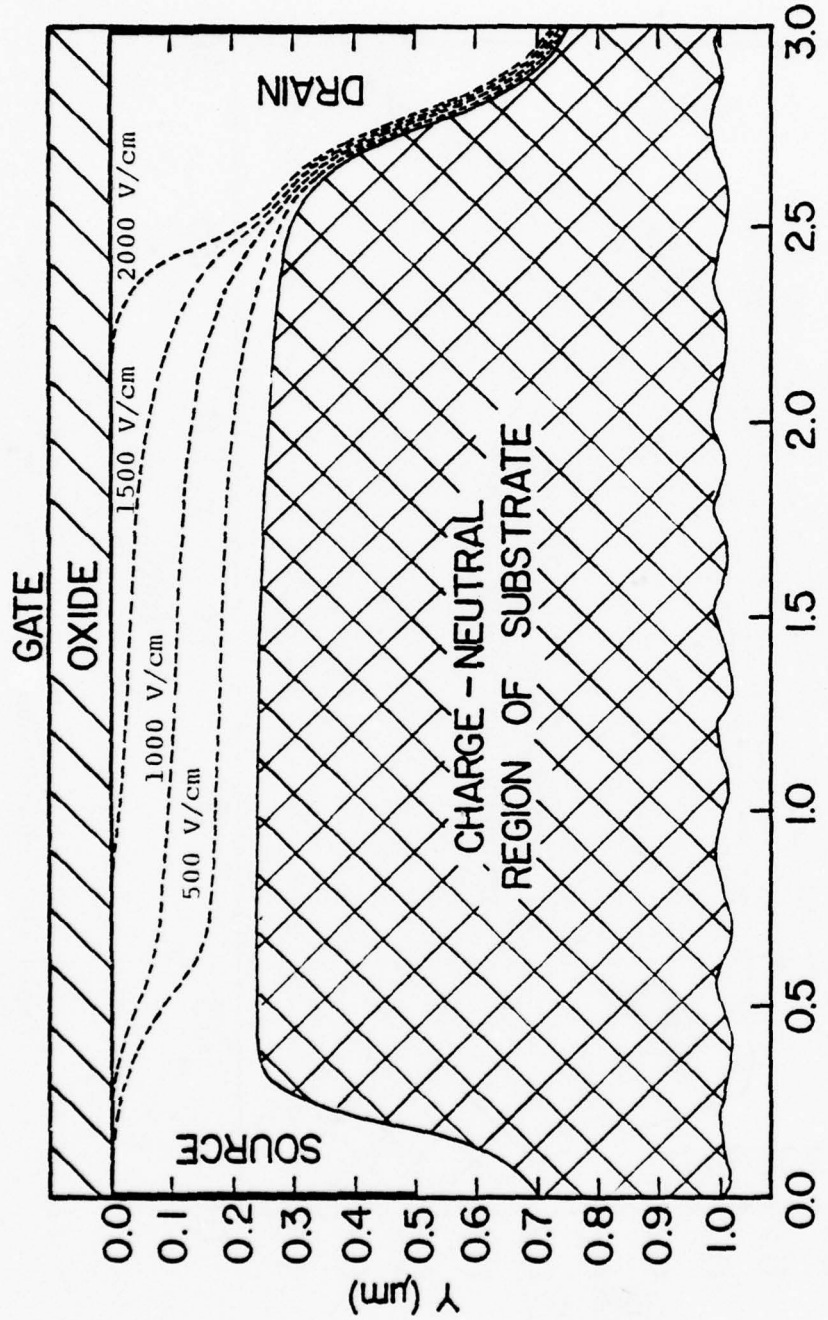


Fig. 1.50. Calculated contours of constant source-drain electric field in a short-channel MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, $L = 3\mu m$. $V_G = 5$ volts, $V_D = 0.5$ volts.

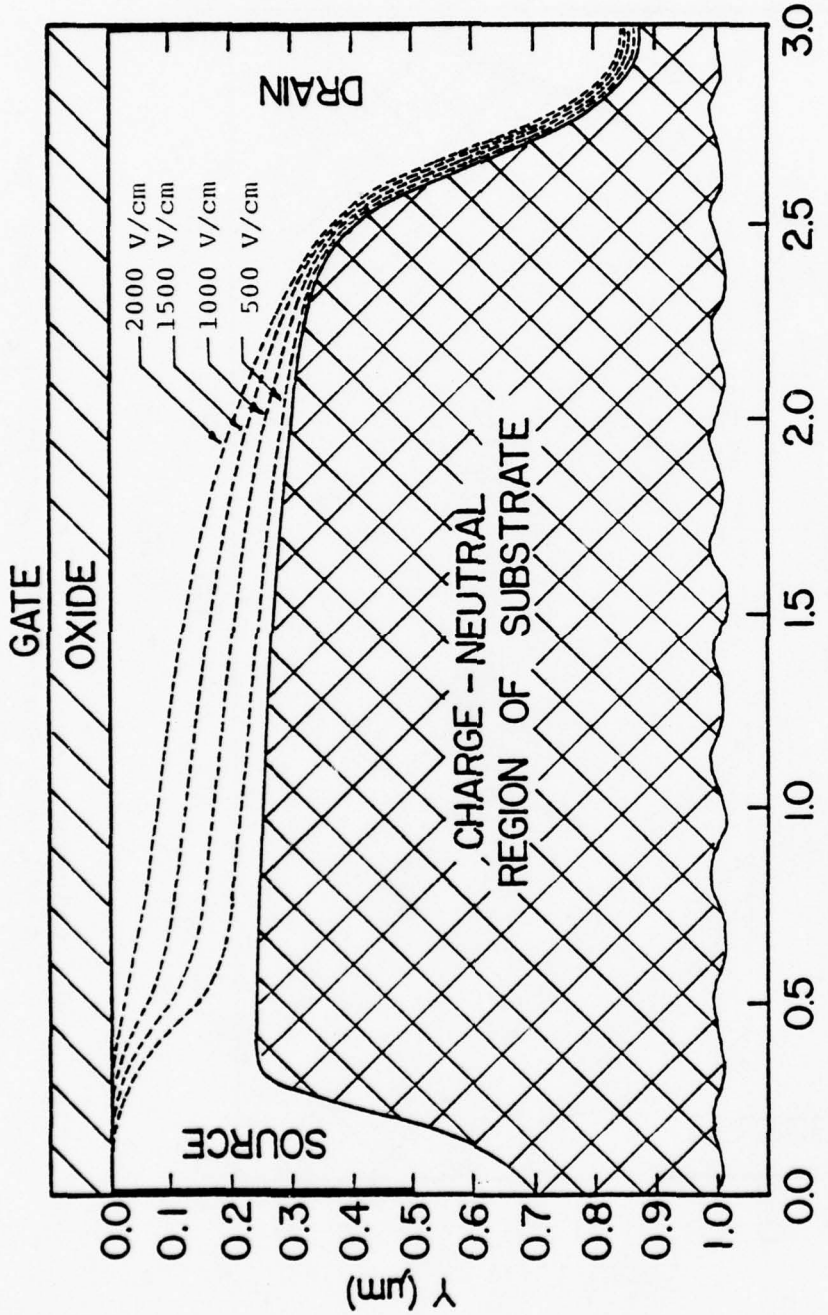


Fig. 1.51. Calculated contours of constant source-drain electric field in a short-channel MOSFET. $N_A = 2 \times 10^{16} cm^{-3}$, $t_{ox} = 1000A°$, $L = 3\mu m$. $V_G = 5$ volts, $V_D = 2$ volts.

diffusion differs between the limits of strong and weak inversion modes of operation.

In the strong inversion mode of operation, near the source junction, most (but not all) inversion layer electric current arises from the mechanism of drift. The ratio of drift current and diffusion current undergoes a continuous change, with an increase of distance from the source junction, although their sum remains constant. Thereby, electric current continuity is maintained in this semiconductor device.

In the weak inversion mode of operation, mobile carrier transport from the source to the drain is almost entirely a consequence of diffusion. In this fashion, MOSFET operation exhibits many similarities to the mechanisms encountered in a bipolar transistor. The source junction injects carriers into the inversion layer, similar to the role of an emitter junction. These carriers diffuse along the oxide-semiconductor interface and, eventually, reach the reverse-biased drain junction. Hence, the drain is like the reverse-biased collector of a bipolar transistor.

Implied by this situation is a necessity to include both drift and diffusion in any rigorous theory for MOSFET operation. This necessity shows the inherent two-dimensional nature of the mechanisms encountered in this device. On a one-dimensional basis, most authors utilize Gauss' law (or Poisson's equation) to establish the total inversion layer charge throughout the entire source-drain inversion layer. In a previous study, it was shown that this inversion charge distribution cannot satisfy a requirement of electric current continuity and, in addition, a one-dimensional form of Gauss' law -- there are insufficient degrees of freedom. On a one-dimensional basis this inversion charge distribution is determined either by Gauss' law or by a requirement for electric current continuity; if one requirement is satisfied, the other is not.

A study of two-dimensional computer solutions has established that everywhere within a MOSFET the inversion charge distribution

satisfies both the requirement of electric current continuity and Gauss' law. A fundamental difference found in this computer solution is that the inversion layer charge satisfies Gauss' law on a two-dimensional basis; thereby, the structure can exhibit an additional degree of freedom.

It is thus recognized that two-dimensional electrostatic interactions between the gate and drain always take place, in order to maintain electric current continuity within an inversion layer. As a consequence, one can mathematically analyze this structure on a one-dimensional basis by imposing a requirement of electric current continuity.

To correct this deficiency in the theory of MOSFET operation, in a previous study a new theory was proposed that assures electric current continuity in the source-drain channel. It was shown that this approach to the MOSFET problem yields results that are physically reasonable and, in addition, are in agreement with experiment. First, this approach to the problem shows that true channel pinch-off does not take place in a MOSFET. Second, for weak inversion operation this approach yields an exponential type of electric current saturation, with an increase of drain voltage. Additional agreement is obtained with the two-dimensional computer solution: in weak inversion, essentially all source-drain electric current is attributable to the mechanism of diffusion within the inversion layer.

Although this new theory implicitly contained the correct physics, it failed to yield satisfactory results when applied to short-channel structures ($L \lesssim 7\mu m$). Moreover, when applied to the weak inversion mode, this theory produced erroneous results regardless of channel length.

In the present study, inadequacies responsible for these difficulties were identified, and a revised MOSFET theory was developed. Through detailed comparisons between this revised theory and our two-dimensional computer solution, it was shown that this revised

theory yields satisfactory results for a wide range of device structures ($L \gtrsim 3\mu m$) throughout both the weak and strong inversion modes of operation. Moreover, unlike most previous analytical theories of MOSFET operation, this agreement was obtained using a single volt-ampere relation.

7.0 List of Symbols

A parameter in the present theory of MOSFET operation ($volts^{-1}$)

C_{ox} capacitance of gate oxide ($farads/m^2$)

D electron diffusivity in MOSFET inversion layer (m^2/sec)

E_o magnitude of E_x ($y=0$) at source-end of a MOSFET channel ($volts/m$)

E_x source-drain electric field ($volts/m$)

E_y gate-induced electric field ($volts/m$)

E_{ys} E_y at oxide-semiconductor interface ($volts/m$)

g_d output conductance of a MOSFET ($mhos$)

I_D source-drain electric current in a MOSFET ($amps$)

I_{diff} diffusion component of I_D ($amps$)

I_{drift} drift component of I_D ($amps$)

I_{Dsat} magnitude of I_D at the onset of channel pinch-off ($amps$)

k Boltzmann's constant ($8.62 \times 10^{-5} eV/°K$)

kT/q thermal voltage ($.0258$ volts at $T = 300°K$)

L source-drain distance in a MOSFET (m)

L_D extrinsic Debye length in semiconductor substrate of a MOSFET: $L_D = \sqrt{\kappa_s \varepsilon_o kT/q^2 N_A}$ (m)

L_c voltage-dependent channel length of a MOSFET (m)

L_{eff} effective source-drain distance in a MOSFET (m)

n electron density in semiconductor material (m^{-3})

n_i intrinsic carrier concentration in semiconductor material (m^{-3})

n_s magnitude of n along oxide-semiconductor interface (m^{-3})

n_{so} magnitude of n_s near the source-end of channel (m^{-3})

n_q nonequilibrium electron density in charge neutral semiconductor substrate (m^{-3})

N_A acceptor impurity ion density in semiconductor substrate (m^{-3})

N_D donor impurity ion density in source/drain islands (m^{-3})

N_x equivalent density of substrate ions electrostatically associated with drain (m^{-3})

N_y equivalent density of substrate ions electrostatically associated with gate (m^{-3})

p hole density in semiconductor material (m^{-3})

p_q nonequilibrium hole density in charge neutral semiconductor substrate (m^{-3})

q electronic charge (coulombs)

Q_D depletion charge in semiconductor substrate (coul /m^2)

Q_i inversion charge in semiconductor substrate (coul/m^2)

Q_{io} magnitude of Q_i at the source-end of a MOSFET channel (coul /m^2)

Q_T total electrostatic charge within the semiconductor substrate (coul /m^2)

T temperature (°K)

t_{ox} gate oxide thickness (m)

V electrostatic potential (volts)

V_{bi} built-in potential of a semiconductor p-n junction (volts)

V_D applied drain-source biasing voltage (volts)

V_G applied gate-source biasing voltage (volts)

V_i magnitude of V at which $n = n_i$ (volts)

V_o magnitude of V_s at source-end of a MOSFET channel (volts)

V_s surface potential in a MOSFET (electrostatic potential at oxide-semiconductor interface) -- (volts)

V_{sc} critical magnitude of V_s near which $\lambda(V_s) \rightarrow q/kT$ (volts)

V_{Dsat} magnitude of V_D at the onset of channel pinch-off (volts)

V_T gate threshold voltage (volts)

W width of a MOSFET (m)

W_D drain depletion layer width (m)

x distance from source towards drain, parallel to oxide-semiconductor interface (m)

x_o location near the source-end of a MOSFET channel where electrostatic interactions between the gate electrode and the drain and source junctions are at a minimum (m)

y distance from semiconductor surface into substrate, perpendicular to oxide-semiconductor interface (m)

y_i magnitude of y at which $n = n_i$ (m)

y_j source/drain island junction depth (m)

β $(kT/q)^{-1}$ $(volts^{-1})$

ΔL channel shortening $(L-L_c)$ associated with "pinch-off" or with channel termination (m)

ε_o permittivity of free space $(8.854 \times 10^{-12}$ farads/m)

κ_i relative dielectric constant of gate oxide (dimensionless)

κ_s relative dielectric constant of semiconductor substrate (dimensionless)

λ separation parameter in the present MOSFET theory $(volts^{-1})$

λ_o magnitude of λ at source-end of a MOSFET channel $(volts^{-1})$

μ electron drift mobility in a MOSFET inversion layer $(m^2/volt\text{-}sec)$

ρ electrostatic charge density within semiconductor substrate $(coul./m^3)$

ϕ_F equilibrium Fermi potential in semiconductor substrate (volts)

ϕ_n electron quasi-Fermi potential in semiconductor substrate (volts)

ϕ_p hole quasi-Fermi potential in semiconductor substrate (volts)

Chapter II

Inversion Layer Carrier Mobility in an MOSFET

C. T. Hsing and D. P. Kennedy

		Page
1.0	Introduction	119
2.0	Mathematical Analysis	121
3.0	Inversion Layer Carrier Distribution	126
4.0	Effective Carrier Mobility in the Inversion Layer of an MOSFET	131
	4.1 Surface Scattering Mechanisms	131
	4.2 Available Calculations of Inversion Layer Mobility	133
	4.3 A Modified Theory of Inversion Layer Mobility	134
	4.4 Calculations of Inversion Layer Mobility	140
5.0	Conclusions	141

Chapter II
Inversion Layer Carrier Mobility in an MOSFET

C. T. Hsing and D. P. Kennedy

1.0 Introduction

The overall goal of this research effort is to obtain an accurate evaluation of the effective carrier mobility within the MOSFET inversion layer so that it can be employed in the MOSFET modeling and design.

Traditional theory of MOSFET operation [1] is based upon an assumed constant carrier mobility within the source-drain inversion layer. This approximation describes MOSFET operation only through a limited range of gate voltages. It has been experimentally established that the inversion layer carrier mobility exhibits a large change, with a change of gate voltage, throughout the normal range of device operation. The reduction in carrier mobility in the inversion layer is believed to result from the surface scattering between carriers and the Si-SiO_2 interface, in addition to the bulk scattering. It is evident that this surface scattering process is intimately related to the average distance between these inversion layer carriers and the interface: a decrease of this distance should produce a decrease of carrier mobility. For this reason, a first step toward attaining a theory for the mobility of these inversion layer carriers is to accurately establish their distribution relative to the oxide-semiconductor interface.

In MOS structures the distribution of inversion layer carriers is traditionally assumed to be governed by solutions of Poisson's

---

[1] (a) S. M. Sze, *Physics of Semiconductor Devices*, Wiley, NY (1969); (b) A. S. Grove, *Physics and Tech. of Semicond. Devices*, Wiley, NY (1967); (c) S. K. Ghandi, *The Theory and Practice of Microelectronics*, Wiley, NY (1968); (d) R.S.C. Cobbold, *Theory and Applications of Field-Effect Transistors*, Wiley, NY (1970).

equation (Gauss' law), in conjunction with Boltzmann statistics [1]. But is has been shown that the results of such calculations are unreliable from both theoretical and experimental points of view [2,3]. Inversion layer carriers in MOS structure reside within a potential well which is created by the interface potential barrier on one side and a large electrostatic field on the other side. Since the potential well is very narrow (in the order of electron De Broglie wavelength), energy quantization in the direction normal to the oxide-semiconductor interface is expected [2]. This quantization effect has been observed in the De Haas-van Alphen oscillation experiment [3]. Therefore, an accurate solution of the inversion layer carrier distribution must satisfy quantum mechanics (expressed by the Schroedinger's wave equation) and Gauss' law (expressed by the Poisson's equation) simultaneously.

Previous solutions [4,5] of this problem have limited applicability to practical MOS devices: they are restricted to MOS capacitor operation where thermal equilibrium is retained. Furthermore, these solutions assume no tunneling at the oxide-semiconductor interface; this assumption is inconsistent with experiment [6]. Therefore, detailed studies of MOS operation require extension of the previous work to physical situation applicable to the MOSFET.

Here we present solutions of Schroedinger's equation that are consistent with both the quantum mechanical aspects and electrostatic constraints imposed by Poisson's equation for non-equilibrium MOS structures. Included in these solutions is the tunneling of inversion layer carriers from the silicon into the bounding SiO_2

---

[2] R. H. Kingston, *Semiconductor Surface Physics*, p. 55, Univ. of Penn. Press (1957).

[3] A. B. Fowler, F. F. Fang, W. E. Howard and P. J. Stiles, *Phys. Rev. Letters*, 16, 901 (1966).

[4] A. P. Gnadinger and H.E. Tally, *Solid-State Electronics*, 13, 1301 (1970).

[5] F. Stern, *Phys. Rev. B.*, 5, 4891 (1972).

[6] V.A.K. Temple and J. Shewchum, *Solid-State Electronics*, 17, 417 (1974).

layer. Sections 2.0, 3.0, and 4.0 contain the mathematical analysis, calculated results, and summary of the quantum mechanical solution of the inversion layer carrier distribution within MOS structures. Section 5.0 will discuss the approach we have used for calculating the effective carrier mobility.

2.0 Mathematical Analysis

Figure 2.1 illustrates the mathematical model which represents a one-dimensional view of energy bands in an n-channel MOSFET. The gate voltage is assumed to be sufficient to produce an inversion layer at the $Si-SiO_2$ interface. Band bending in Fig. 2.1 is a consequence of uncompensated acceptor ions and mobile electrons in the inversion layer. The interface potential barrier results from an energy difference between the work function in the oxide and the electron affinity in the semiconductor. In Eq. (2.1), ΔE is the energy difference between quasi-Fermi levels for holes, E_{Fp}, and electrons, E_{Fn}. Thus, implied by this model is a deviation from thermal equilibrium, as encountered between the source and drain of an MOSFET.

The inherent one-dimensional nature of this model is evident in Fig. 2.1 by assuming constant quasi-Fermi levels from the $Si-SiO_2$ interface to the substrate. This assumption implies the gradient of E_{Fn} between source and drain will extend throughout the entire substrate material. In two-dimensional MOSFET analysis [7], this unphysical situation is alleviated by inversion layer carrier flow in a direction normal to the $Si-SiO_2$ interface: a mechanism not considered in the present model. Since it is shown in this two-dimensional analysis that the quasi-Fermi level E_{Fn} is essentially constant across the inversion layer, the constant quasi-Fermi level assumption is adequate fo the purpose of calculation of inversion layer carrier distribution.

---

[7] R. F. Motta and D. P. Kennedy, *Steady-State Theory for the Metal-Oxide-Semiconductor Field-Effect Transistor,* Ph.D. Thesis, Univ. of Florida (1976).

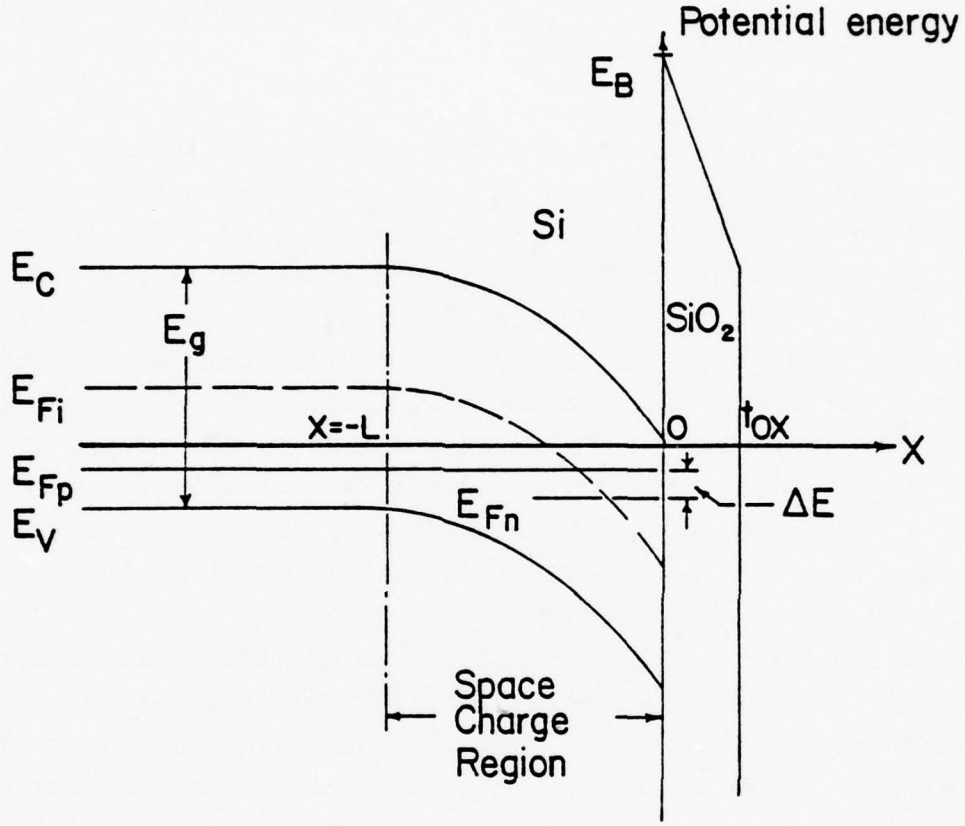


Figure 2.1. Mathematical model of an MOS structure.

Additional simplifications and approximations are used in this model, although they have little influence upon the problem under consideration. For example, we neglect charged surface states that may reside at the Si-SiO_2 interface and impurity ions within the SiO_2 layer.

The electron distribution normal to the interface (x-axis) is governed by the Schroedinger's equation which can be written as:

$$H\psi_i(x) = -\frac{h^2}{2m_l}\frac{d^2\psi_i(x)}{dx^2} + E(x)\psi_i(x) = E_i\psi_i(x) \quad , \quad (2.1)$$

where (i) H is the differential operator defined by

$$H \equiv -\frac{h^2}{2m_l}\frac{d^2}{dx^2} + E(x) \; ;$$

(ii) $\psi_i(x)$ is the eigenfunction of Eq. (2.1) corresponding to energy eigenvalue (quantized energy level) E_i;

(iii) m_l is electron conduction effective mass along the x-axis;

(iv) E(x) is the electrostatic potential energy distribution arising from uncompensated acceptor ions and mobile electrons.

The boundary conditions for $\psi_i(x)$ are $\psi_i(x) = 0$ at $x = \pm\infty$, where the origin of the coordinate system (x=0) is at the Si-SiO_2 interface.

Any solution of Eq. (2.1) requires the electrostatic potential energy distribution E(x) within the semiconductor material. Because $E(x) = -qV(x)$, this energy distribution is readily obtained from the solution of Poisson's equation:

$$\frac{d^2V(x)}{dx^2} = \frac{q}{\kappa\varepsilon_o}[N_A + n(x) - p(x)] \quad , \quad (2.2)$$

where n(x) and p(x) represent the electron and hole distributions, respectively. For simplicity, the acceptor ion density N_A is assumed constant throughout the region under consideration.

The carrier distribution n(x) is obtained through a simultaneous solution of Eq. (2.1) and Eq. (2.2). This solution is obtained by using iterative techniques. A first guess is made for the potential energy distribution E(x) in Eq. (2.1); from this guess, solutions are obtained for both the energy eigenvalues E_i and the associated eigenfunctions $\psi_i(x)$. From these $\psi_i(x)$ we next calculate the electron distribution n(x) and the hole distribution p(x). From these mobile carrier distributions, Eq. (2.2) yields an improved electrostatic potential energy distribution E(x). The improved E(x) is then used in solving Eq. (2.1), which yields improved solutions for E_i and $\psi_i(x)$. The above process is repeated until self-consistent solutions for n(x) and E(x) are obtained. This technique for solving Eq. (2.1) and Eq. (2.2) represents a Picard iteration of the type frequently used in applied mathematics [8].

It has been shown [9] that the electron distribution in an MOS structure is given by

$$n(x) = \frac{m_{11} n_v kT}{\pi h^2} \sum_{i=1}^{\infty} \{\log_e [1+\exp(E_{Fn}-E_i)/kT]\} |\psi_i(x)|^2 \quad , \qquad (2.3)$$

where m_{11} is the density-of-states effective mass of electrons in a direction parallel to the interface, and n_v is the valley degeneracy factor (depends upon the substrate crystal orientation). Because of the normalization condition for $\psi_i(x)$:

$$\int_{-\infty}^{\infty} |\psi_i(x)|^2 dx = 1 \quad , \qquad (2.4)$$

the integrated electron inversion charge density in x direction can be obtained from Eq. (2.3):

$$N_I = \frac{m_{11} n_v kT}{\pi h^2} \sum_{i=1}^{\infty} \log_e [1+\exp(E_{Fn}-E_i)/kT] \quad . \qquad (2.5)$$

---

[8] S. H. Gould, *Variational Methods for Eigenvalue Problems*, 2nd Ed., Univ. of Toronto Press (1957).

[9] C. T. Hsing and D. P. Kennedy, Contract Report HDL-CR-75-193-1, Oct. (1975).

The energy levels E_i and associated wavefunctions $\psi_i(x)$ used in the expression for $n(x)$ [Eq. (2.3)] are solutions of Eq. (2.1). We solve this equation by use of the Rayleigh-Ritz method [8]. The first step in this method is to express each $\psi_i(x)$ by the following series:

$$\psi_i(x) = \sum_{j=1}^{n} C_{ij} f_j(x) \quad , \tag{2.6}$$

where $f_j(x)$'s are linearly independent functions. The upper bound of the energy eigenvalue E_i corresponding to this $\psi_i(x)$ can then be obtained by the Rayleigh-Ritz method which is outlined below. As the number of terms (n) in Eq. (2.6) increases, this upper bound decreases and approaches a limit which is the desired energy eigenvalues. In practice, it is not necessary to use a large number of terms in Eq. (2.6), if the approximating functions $f_j(x)$ are selected such that they have close resemblance to the desired wavefunction $\psi_i(x)$; in the present analysis typically 10 to 20 terms are required.

For each eigenfunction $\psi_i(x)$ approximated by Eq. (2.6), we can write

$$(H\psi_i, \psi_i) = \int_{-\infty}^{\infty} \psi_i(x) H\psi_i(x) dx = \sum_{k,l} a_{kl} c_{ik} c_{il} \quad , \tag{2.7}$$

where

$$a_{kl} \equiv \int_{-\infty}^{\infty} f_k(x) H f_l(x) dx \quad . \tag{2.8}$$

In addition, we can also write

$$(\psi_i, \psi_i) = \int_{-\infty}^{\infty} \psi_i^2(x) dx = \sum_{k,l} b_{kl} c_{ik} c_{il} \quad , \tag{2.9}$$

where

$$b_{kl} \equiv \int_{-\infty}^{\infty} f_k(x) f_l(x) dx \quad . \tag{2.10}$$

From these relations, it can be shown [8] that

$$E_i = \min \frac{(H\psi_i, \psi_i)}{(\psi_i, \psi_i)} \quad . \tag{2.11}$$

If Eq. (2.11) is written for each eigenfunction ψ_i in a set of n eigenfunctions, we obtain a set of n equations of the form

$$\sum_{j=1}^{n} c_{ij}(a_{ji} - E_i b_{ji}) = 0 \; ; \; i = 1,2,.....,n, \tag{2.12}$$

which can be solved for the desired energy eigenvalues E_i and the expansion coefficients c_{ij} required in Eq. (2.6) can be readily determined from Eq. (2.12), once E_i is known.

3.0. Inversion Layer Carrier Distribution

The potential well in which inversion layer carriers are constrained produce a quantization of their energy levels in the direction normal to the Si-SiO_2 interface. Figure 2.2 presents a graphical illustration of this quantization for two distinctly different situations encountered in an MOS capacitor: first, for strong inversion (normal surface electric field $\xi_s = 9.0 \times 10^4$ v/cm) and second, for weak inversion ($\xi_s = 2.72 \times 10^4$ v/cm). The substrate doping is 4×10^{15} carriers/cm^3.

From Fig. 2.1, strong inversion produces a decrease in width of the potential well and, as a consequence, an increase in the degree of energy quantization. This conclusion is drawn from the fact that in strong inversion there are 8 energy levels into which more than 99% of all inversion layer carriers reside. In weak inversion, we have a substantially wider potential well containing 14 energy levels. From this observation, it is expected that substantial differences should exist between the carrier distribution resulting from strong inversion and weak inversion. A reduced degree of quantization in weak inversion could imply better agreement with classical solution for mobile carrier distribution in an MOS capacitor.

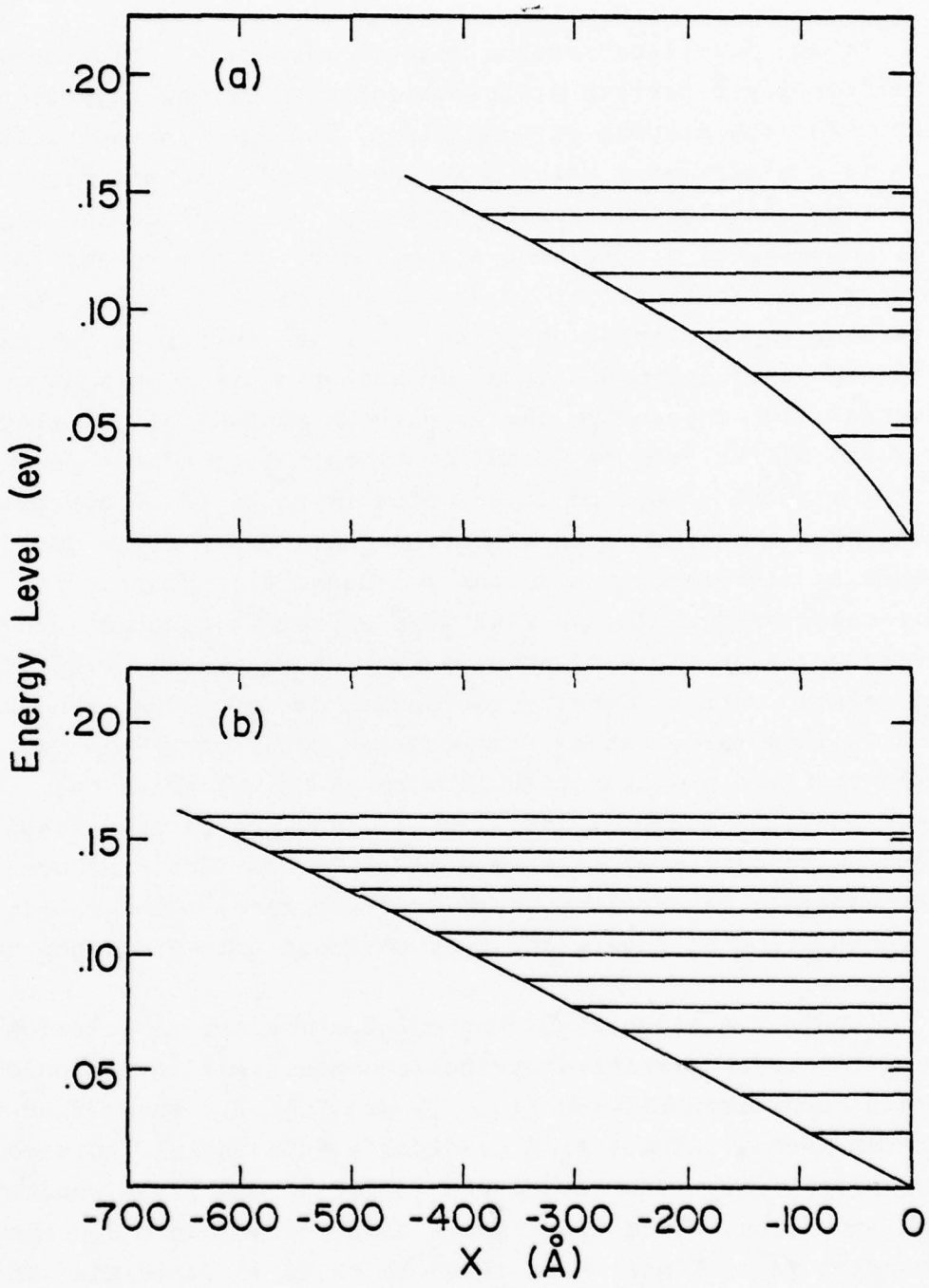


Figure 2.2. Calculated electron energy levels in the surface potential well.
(a) $\xi_s = 9.00 \times 10^4$ v/cm
(b) $\xi_s = 2.72 \times 10^4$ v/cm

Figure 2.3 illustrates a quantum mechanical solution for the inversion layer carrier distribution under strong inversion condition. For the purpose of comparison, included in this illustration is a traditional calculation based upon the solution of Poisson's equation, assuming Boltzmann statistics are applicable. From Fig. 2.3, substantial differences are observed in the inversion layer carrier distribution, yet it should be noted that the integrated inversion layer carrier densities (N_I) are identical for these two different calculations. Thus, we answer a previously unverified proposal [10] suggesting that a quantum mechanical calculation of N_I would differ from the traditional calculation of this parameter.

The total inversion layer width is shown to be approximately 300Å, when measured to the location where the carrier density equals n_i (intrinsic mobile carrier density). Contrasting with this calculated width, we find that Poisson's equation predicts an inversion layer width of about 1600Å. Furthermore, Fig. 2.3 shows the maximum carrier density is located at about 30Å from the $Si-SiO_2$ interface, rather than at this interface, which is in agreement with previous calculations of this problem [4]. Unlike these previous calculations, Fig. 2.3 shows a carrier density of about $5x10^{15}$carriers/cm^3 at the $Si-SiO_2$ interface, rather than total absence of carriers there. In addition, shown in Fig. 2.3 is penetration of inversion layer carriers into SiO_2, due to tunneling.

Figure 2.4 illustrates the quantum mechanical solution for the inversion layer carrier distribution under weak inversion condition. Direct comparison between Fig. 2.3 and Fig. 2.4 shows that the quantum mechanical solution predicts a substantial increase in inversion layer width (from 300Å to 500 Å) under the condition of weak inversion, while traditional theory predicts a decrease in inversion layer width (from 1600Å to 1250Å). It is also found that the change from strong inversion to weak inversion moves the point of maximum carrier density about 10Å further into the substrate, locating this point at about 40Å from the $Si-SiO_2$ interface.

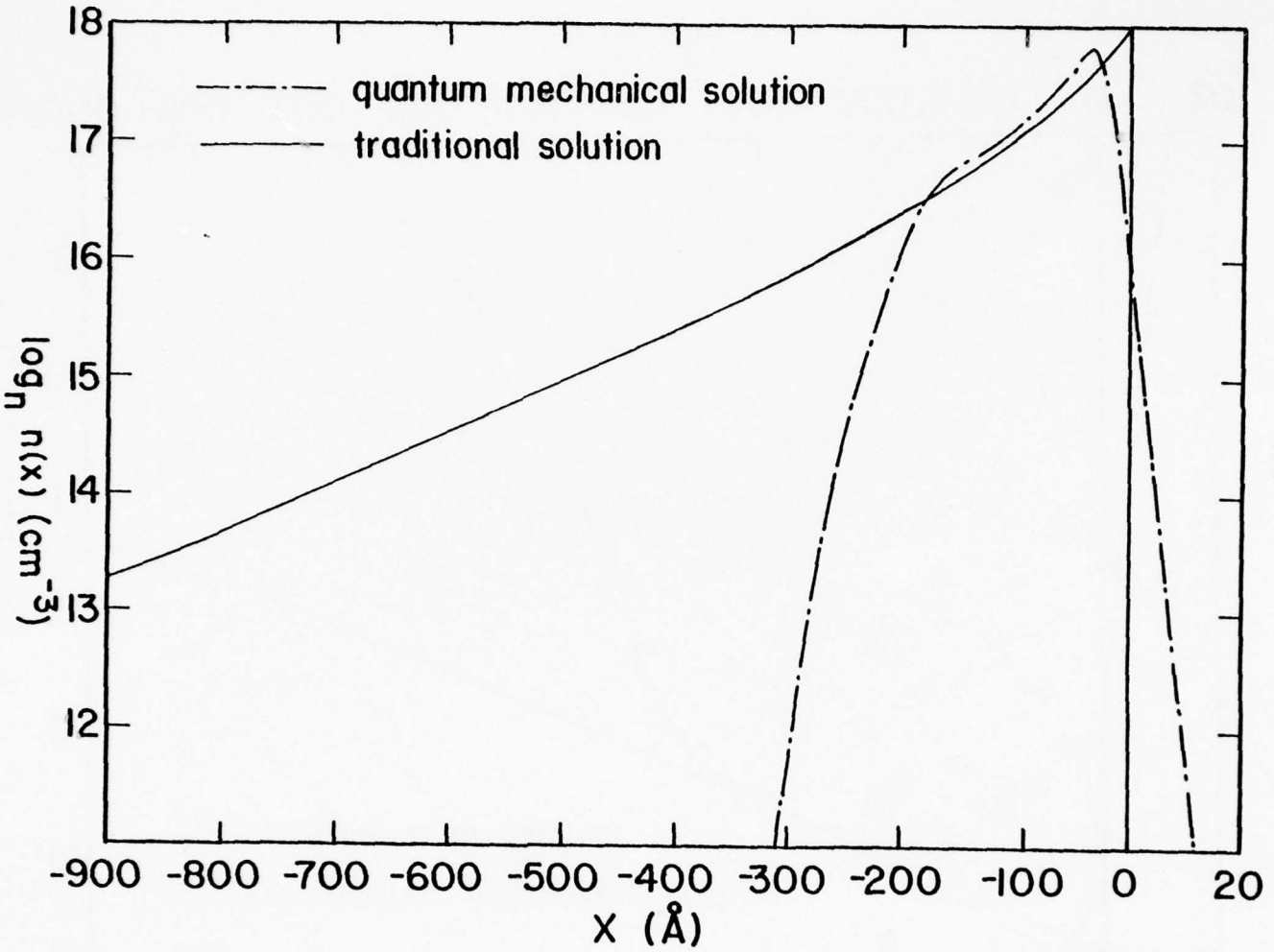


Figure 2.3. Comparison between quantum mechanical solution and traditional solution of the inversion layer carrier distribution. $\mathcal{E}_s = 9.00 \times 10^4$ v/cm; $N_I = 4.00 \times 10^{11}$ cm^{-2}.

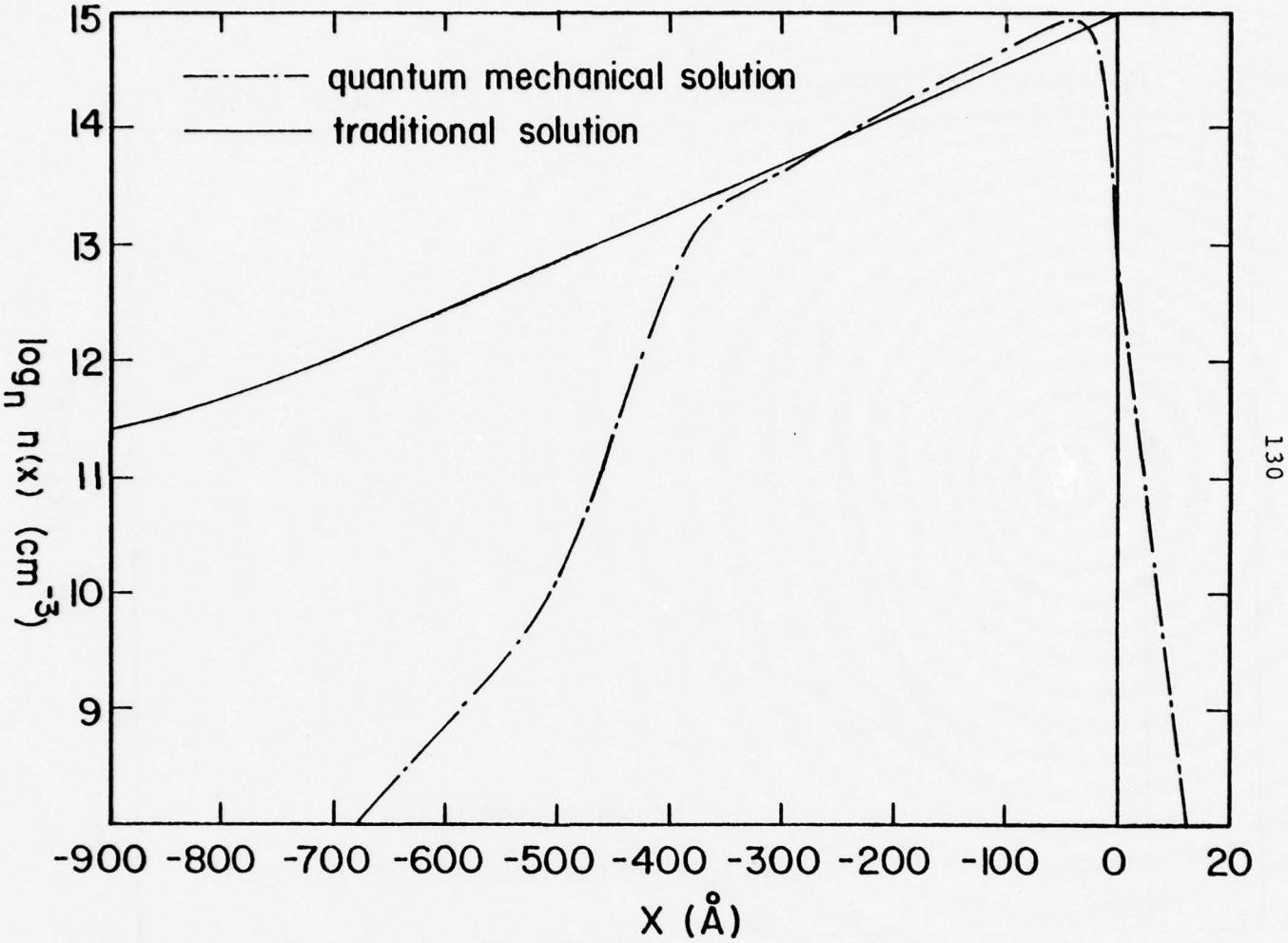


Figure 2.4. Comparison between quantum mechanical solution and traditional solution of the inversion layer carrier distribution. $\mathcal{E}_s = 9.00 \times 10^4$ v/cm; $N_I = 4.00 \times 10^{11}$ cm^{-2}.

In addition to the preceding equilibrium calculations for MOS capacitor, quasi-two-dimensional solutions are obtained for the inversion layer carrier distribution in an MOSFET as shown in Fig. 2.5. These calculations are based upon the selection of three locations between the source and drain of an MOSFET. The specific model used for this calculation is an MOSFET with channel length of 10 μm and substrate impurity density $N_A = 2x10^{15}cm^{-3}$. The gate and drain voltages are 3 volts and 4 volts, respectively. A rigorous traditional two-dimensional computer solution of this boundary value problem [7] provides the continuous surface potential distribution and normal electric field at all locations between source and drain. Figure 2.5 shows a comparison between the calculated carrier distribution at these three locations: 1.0μm, 5.0μm and 7.0μm from the source junction. From this sequence of calculations, the number of inversion layer carriers is shown to decrease from source to drain, but the layer width increases in this direction.

4.0 Effective Carrier Mobility in the Inversion Layer of an MOSFET

Next we discuss the mobility of mobile carriers in an MOSFET inversion layer, when these carriers are bounded by a potential well that is capable of producing energy quantization. We shall review the scattering mechanisms at the $Si-SiO_2$ interface and, in addition, review a solution of this mobility problem as published by J. R. Schrieffer [11]. Finally, we present a solution for this problem which is believed an improvement over presently available solutions.

4.1 Surface Scattering Mechanisms

Two types of scattering mechanisms are believed present at the $Si-SiO_2$ interface: specular (ideal) scattering and diffuse (random) scattering. Specular scattering implies that the normal

---

[10] Y. A. El-Mansy, Ph.D. Thesis, Carleton Univ., Canada (1974).
[11] J. R. Schrieffer, *Phys. Rev.*, 97, 641 (1955).

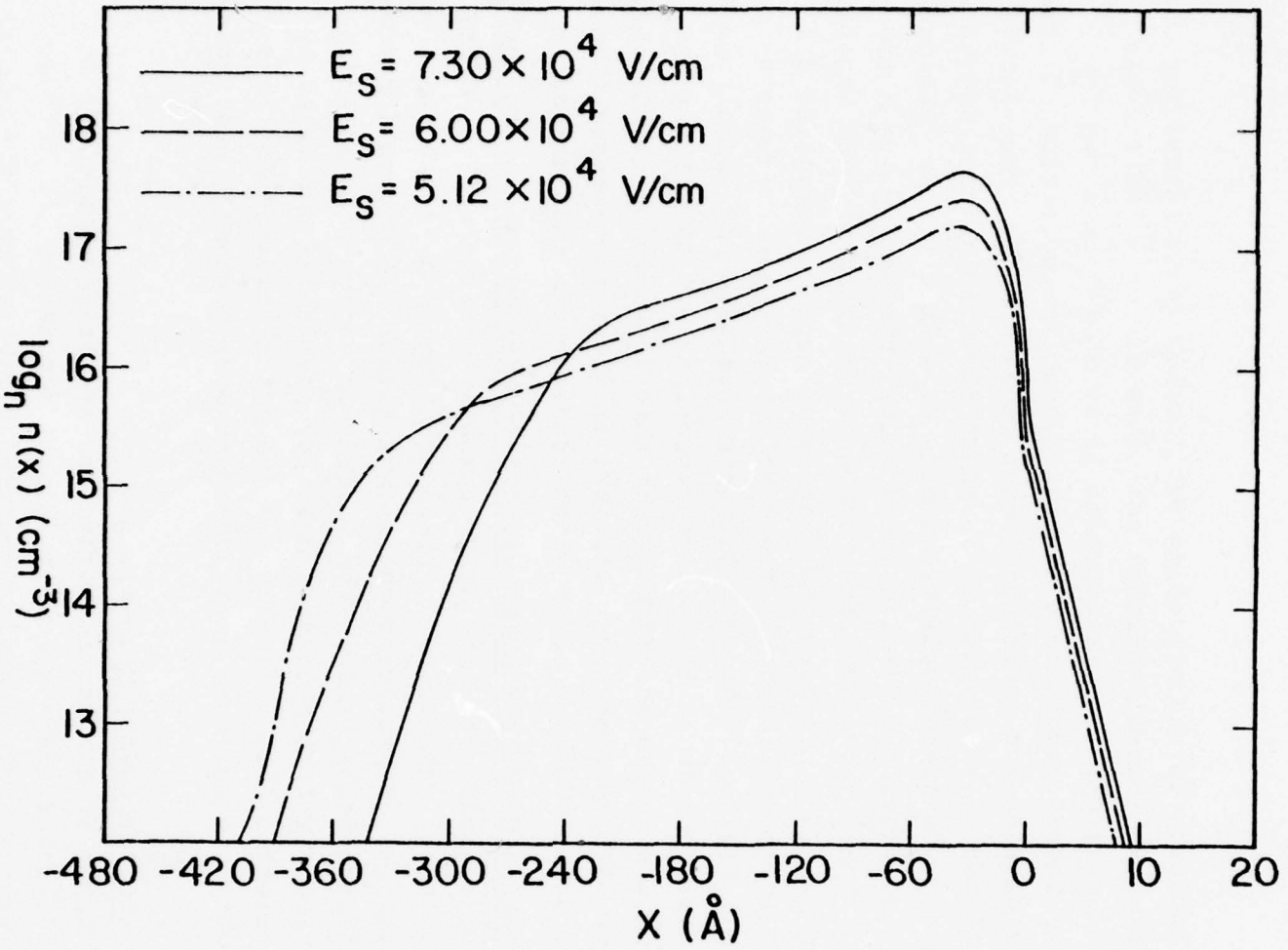


Figure 2.5. Inversion layer carrier distributions for three locations between the source and drain of an MOSFET.

component of carrier momentum will reverse upon collision with the silicon surface, while the parallel component of momentum remains constant. Since specular scattering produces no change of carrier velocity parallel to the $Si-SiO_2$ interface, this mechanism produces no change of inversion layer carrier mobility. Contrasting with specular scattering, diffuse (or random) scattering changes the parallel component of mobile carrier momentum and, as a consequence, is believed to reduce the inversion layer carrier mobility in an MOSFET. Diffuse scattering is often suggested to result from mechanical roughness at the $Si-SiO_2$ interface, in conjunction with chemical contamination at this location. Thus, it is diffuse scattering at this $Si-SiO_2$ interface that would produce a reduction of inversion layer carrier mobility, relative to the bulk mobility in semiconductor material.

4.2 Available Calculations of Inversion Layer Mobility

The first published solution of this inversion layer mobility problem [11] was based upon the Boltzmann transport equation, with an assumed diffuse scattering mechanism at the $Si-SiO_2$ interface. An important assumption in this solution is implied by the use of Boltzmann statistics -- which is inapplicable to situations where energy quantization is present. Despite this assumption, this solution did, indeed, yield results showing that the carrier mobility in an MOSFET can be substantially reduced from its value in bulk semiconductor material. It is proposed that this solution is applicable only for weak inversion operation for the following reasons:

A. The Boltzmann transport equation is rigorously applicable only for situations where the carrier energy is continuous. In an MOSFET, energy quantization exists in a direction normal to the $Si-SiO_2$ interface and, therefore, it is not obvious that the Boltzmann transport equation accurately describes all governing physical mechanisms. From Section 3.0 of this chapter, in

weak inversion the calculated quantized energy levels are separated by no more than about kT; hence, the consequences of quantization should be negligible. It was also shown in Section 3.0 that strong inversion produces an energy level separation far in excess of kT. As a consequence of this situation it is proposed that under conditions of weak inversion, solutions of the Boltzmann transport equation are applicable to this mobility problem. It is also proposed that this type of solution may not be applicable to situations where strong inversion exists in a MOSFET.

B. In general, Boltzmann statistics are applicable only for non-degenerate situations -- when the conduction band edge lies a minimum of about 2kT above the Fermi level. In most MOSFET structures, strong inversion operation produces a conduction band to Fermi level separation of less than 2kT. As a consequence, Fermi-Dirac statistics must be used in the analysis of such structures.

4.3 A Modified Theory of Inversion Layer Mobility

The Boltzmann transport equation is rigorously inapplicable to a Fermi gas of electrons. As in the solution of Schrieffer we utilize this equation with a correction for the degeneracy of the electron gas. It is expected that this approach will yield results that are in better agreement with experiment than presently available.

Throughout this calculation we assume the electric field parallel to the SiO_2 interface is very small when compared to the electric field normal to this interface. As a consequence, this parallel field component can be assumed a small perturbation in

the Boltzmann transport equation, and we solve this equation for the steady-state carrier distribution. Thereafter, the electric current along this inversion layer (and the carrier mobility) is determined from the resulting carrier distribution. In solving this equation we make the following assumptions:

(i) Fermi-Dirac statistics apply in thermal equilibrium situations;

(ii) surface scattering is totally random; and

(iii) bulk relaxation time is constant.

For a location between the source and drain of an MOSFET, we consider an electric field ξ_y parallel to the surface and field ξ_x normal to the surface. The origin of x-axis is taken at the surface. Under steady state conditions the Boltzmann transport equation for the electron distribution function can be written as

$$\underline{v}\cdot\nabla_f + \underline{a}\cdot\nabla_{\underline{v}} f = -(f-f_o)/\tau \quad , \tag{2.13}$$

where (i) $\underline{v}$ and $\underline{a}$ are the velocity and acceleration of the carrier, respectively;

(ii) τ is the relaxation time;

(iii) $f = f_o + f_1$, f_o being the distribution function for thermal equilibrium situation and f_1 being a perturbing term.

In Eq. (2.13) f_o is given by the Fermi-Dirac distribution function:

$$f_o(\underline{v},x) = \frac{C}{1 + \exp\{-[mv^2/2 - qV(x) - E_{Fn}]\}} \tag{2.14a}$$

where V(x) is the potential associated with electric field ξ_x and m is the effective mass of mobile carriers (isotropic effective mass is assumed). In Eq. (2.14), C is a normalization constant such that after integrating Eq. (2.14) we obtain the number of mobile carriers per unit area in the x-direction, N_I:

$$N_I = \int d\underline{v}dx f_o(\underline{v},x) \tag{2.14b}$$

By substituting Eq. (2.14) into Eq. (2.13) and retaining only the first order perturbation term, f_1 is found to satisfy the equation:

$$\frac{-q\xi_x}{m}\frac{\partial f_1}{\partial v_x} + v_x\frac{\partial f_1}{\partial x} + \frac{f_1}{\tau} = -\frac{qv_y\xi_y}{kT} f_o(1-f_o) \tag{2.15}$$

Equation (2.15) can be reduced to a simpler differential equation by the following substitution [11]:

$$\varepsilon_x \equiv \frac{1}{2} mv_x^2 - q[V(x)-V_s] \quad \text{and} \tag{2.16}$$

$$F(\varepsilon_x, v_x, v_y, v_z) \equiv f_1(x, v_x, v_y, v_z) \quad , \tag{2.17}$$

where V_s is the surface potential. Equation (2.16) implicitly contains the boundary condition for x, i.e., $V(x) = V_s$ at $x = 0$.

After substituting Eq. (2.16) and Eq. (2.17) into Eq. (2.15), Eq. (2.15) becomes

$$\frac{-q\xi_x}{m}\frac{\partial F}{\partial v_x} + \frac{F}{\tau} = -f_o(1-f_o)av_y\xi_y/kT \quad . \tag{2.18}$$

By assuming random scattering at the surface, the boundary condition for Eq. (2.18) is

$$F = 0 \text{ for } v_x = v_{xs} \quad , \tag{2.19}$$

where $v_{xs} = \sqrt{2\varepsilon_x/m}$ is the positive x component of carrier velocity at the surface. The solution of Eq. (2.19) can be shown to be

$$F = \frac{f_o(1-f_o)mv_y\xi_y}{kT} e^{K(v_x,\varepsilon_x)} \int_{V_{xs}}^{V_x} dv_x' \frac{e^{-K(v_x',\varepsilon_x)}}{\xi_x(v_x',\varepsilon_x)} \tag{2.20}$$

where

$$K(v_x,\varepsilon_x) \equiv \frac{m}{q\tau} \int_o^{V_x} \frac{dv_x'}{\xi_x(v_x',\varepsilon_x)} \quad . \tag{2.21}$$

The current density per unit channel width in the y-direction is

$$I_y = -q \int_{-\infty}^{\infty} dv_x \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} dv_z \int_{0}^{\infty} dx F(v_x, v_y, v_z, \varepsilon_x)$$

$$= -\frac{qm\xi_y}{kT} \int_{-\infty}^{\infty} dv_x \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} dv_z \int_{0}^{\infty} dx v_y^2 f_o(1-f_o) e^K \int_{v_{xs}}^{v_x} \frac{e^{-K'}}{\xi_x} dv_x' \tag{2.22}$$

Equation (2.22) is integrated over v_y and v_z first by the following substitutions:

$$\alpha = \frac{m}{2kT} (v_y^2 + v_z^2) \quad , \tag{2.23}$$

$$\beta = -\frac{1}{kT} [\tfrac{1}{2}mv_x^2 - qV(x) - E_{Fn}] = -\frac{1}{kT} (\varepsilon_x - qV_s - E_{Fn}) \quad . \tag{2.24}$$

Now we can write

$$f_o = \frac{C}{1 + e^{\alpha-\beta}} \quad , \tag{2.25a}$$

$$f_o(1-f_o) = \frac{C}{(1+e^{\alpha-\beta})(1+e^{-\alpha+\beta})} = -\frac{\partial f_o}{\partial \alpha} \quad . \tag{2.25b}$$

Also, let

$$v_y^2 + v_z^2 = \frac{2kT\alpha}{m} = r^2 \quad ;$$

$$v_y = r\cos\theta \quad ; \tag{2.26}$$

$$v_z = r\sin\theta \quad ;$$

then $$\int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} dv_z v_y^2 f_o(1-f_o)$$

$$= \int_o^{\infty} rdr \int_o^{2\pi} d\theta \ \frac{2kT\alpha}{m} \cos^2\theta \left(-\frac{\partial f_o}{\partial \alpha}\right)$$

$$= 2\pi \left(\frac{kT}{m}\right)^2 \int_o^{\infty} \left(-\frac{\partial f_o}{\partial \alpha}\right) \alpha d\alpha$$

$$= 2\pi \left(\frac{kT}{m}\right)^2 \log_n(1+e^{\beta})$$

$$= 2\pi \left(\frac{kT}{m}\right)^2 \log_n\{1+\exp[-(\varepsilon_x - qV_s - E_{Fn})/kT]\} \tag{2.27}$$

Now, Eq. (2.22) can be written as

$$I_y = -\frac{2\pi qC\xi_y kT}{m} \int_{-\infty}^{\infty} dv_x \int_o^{\infty} dx \log_n\{1+\exp[-(\varepsilon_x - qV_s - E_{Fn})/kT]\}$$

$$e^K \int_{v_{xs}}^{v_x} \frac{e^{-k'}}{\xi_x'} dv_x' \quad . \tag{2.28}$$

By changing the integrating variables v_x, x to v_x, ε_x and notice that K is an odd function of v_x, Eq. (2.28) reduces to

$$I_y = -\frac{2\pi q^2 C\xi_y kT\tau^2}{m^3}\int_0^\infty d\varepsilon_x \log_n\{1+\exp[-(\varepsilon_x-qV_s-E_{Fn})/kT]\}[1-2K(\varepsilon_x)-e^{-2K(\varepsilon_x)}], \tag{2.29}$$

where

$$K(\varepsilon_x) = \frac{m}{q\tau}\int_0^{\sqrt{2\varepsilon_x/m}} \frac{dv_x'}{\xi_x(v_x',\varepsilon_x)} \tag{2.30}$$

The reduction of carrier mobility in the inversion layer of an MOSFET due to random surface scattering can be expressed by the following:

$$I_y = -qN_I\mu_{eff}\xi_y \quad , \tag{2.31}$$

where N_I is the number of inversion layer carriers per unit area given by Eq. (2.14b) and μ_{eff} is the effective mobility. Equation (2.14b) can be integrated to give

$$N = \frac{4\pi CkT\tau}{m^2}\int_0^\infty d\varepsilon_x \log_n\{1+\exp[-(\varepsilon_x-qV_s-E_{Fn})/kT]\}K(\varepsilon_x) \tag{2.32}$$

By combining Eq. (2.29) to (2.32) and writing the bulk mobility $\mu_b = q\tau/m$, we obtain

$$\frac{\mu_{eff}}{\mu_b} = \frac{\int_0^\infty d\varepsilon_x \log_n\{1+\exp[-(\varepsilon_x-qV_s-E_{Fn})/kT]\}[1-2K(\varepsilon_x)-e^{-2K(\varepsilon_x)}]}{\int_0^\infty d\varepsilon_x \log_n\{1+\exp[-(\varepsilon_x-qV_s-E_{Fn})/kT]\}2K(\varepsilon_x)} \cdot \tag{2.33}$$

It is interesting to notice in Eq. (2.33) that if the surface potential is taken as reference ($V_s=0$) and Fermi energy E_{Fn} lies several kT below zero, i.e., if

$$\varepsilon_x - E_{Fn} \gtrsim 3kT. \tag{2.34a}$$

then

$$\log_n\{1+\exp[-(\varepsilon_x-qV_s-E_{Fn})/kT]\}\approx\exp[-(\varepsilon_x-qV_s-E_{Fn})/kT\ , \tag{2.34b}$$

and Eq. (2.33) reduces to

$$\frac{\mu_{eff}}{\mu_b} = \frac{\int_0^\infty d\varepsilon_x \exp[-(\varepsilon_x-qV_s-E_{Fn})/kT]\,[1-2K(\varepsilon_x)-e^{-2K(\varepsilon_x)}]}{\int_0^\infty d\varepsilon_x \exp[-(\varepsilon_x-qV_s-E_{Fn})/kT]\,2K(\varepsilon_x)} \ . \tag{2.35}$$

Equation (2.35) is identical to that expression given in Schrieffer's paper which assumed Boltzmann statistics for the carrier distribution function in thermal equilibrium. The condition, Eq. (2.34a), for reducing Eq. (2.33) to Eq. (2.35) is consistent with the ordinary criterion for determining if the electron gas is degenerate. Thus we have extended the Schrieffer's theory from non-degenerate gas to more general degenerate gas. We next use Eq. (2.33) to calculate the effective mobility for strong inversion operation, and Eq. (2.35) to calculate the effective mobility for weak inversion operation.

4.4 Calculations of inversion Layer Mobility

Numerical calculations based upon the modified Schrieffer's theory outlined above have been made: first for strong inversion ($\xi_s=9.0\times10^4$v/cm) and, second, for weak inversion ($\xi_s=2.72\times10^4$v/cm). The assumed substrate doping was 4×10^{15}cm^{-3}. Under strong inversion condition, the ratio of surface mobility to bulk mobility was found to be 0.26. Under conditions of weak inversion, this ratio becomes 0.40. These results agree reasonably well (within 20%)

with experimental data [13,14]. Since both this theory and the measuring techniques for surface mobility need further refinement, this agreement appears satisfactory.

5.0 Conclusion

There are two accomplishments on MOS devices presented in this report: a quantum mechanical solution of the inversion layer carrier distribution in MOS structure and an extension to Schrieffer's theory on MOS inversion layer carrier mobility.

The quantum mechanical solution for the inversion layer carrier distribution was obtained by solving the Schroedinger's wave equation and Poisson's equation conjunctively. The results deviate from the traditional solution (by solving Poisson's equation alone) in the following ways:

(a) the inversion layer carrier density exhibits a maximum at about 35Å from the $Si-SiO_2$ interface;

(b) there is a finite number of carriers residing at the interface, due to tunneling through the gate oxide; and

(c) the inversion layer is much narrower than predicted by traditional concepts, and this width increases from source to drain (as opposed to decreasing in that direction, as predicted by traditional solution).

A quantum mechanical solution for the inversion layer carrier distribution represents the basis upon which the carrier mobility must be calculated. Rigorously, this carrier mobility will be obtained from a solution of the von Neumann equation [15] containing a perturbation term for the electric field normal to the $Si-SiO_2$ interface. Thus far, such a solution has not been undertaken. Instead, we have solved this problem using the approach

---

[13] R. F. Pierret and C. T. Sah, *Solid State Electronics*, 11, 279 (1968).

[14] N. St. J. Murphy, F. Berg, and I. Flinn, *Solid State Electronics*, 12, 775 (1969).

[15] A. H. Kahn and H. P. R. Frederikse, *Solid State Physics*, 9, 257 (1959).

previously taken by Schrieffer. His calculation of mobility is based upon the solution of the Boltzmann transport equation. Unlike the Schrieffer solution, our calculations are corrected for strong inversion -- where the inversion layer carriers become a degenerate electron gas. This technique has yielded calculated values of inversion layer carrier mobility that are in adequate agreement with experiment for the design and development of MOSFET structures.

Chapter III

An Integral Equation-Relaxation Procedure for the Determination of Equilibrium Potentials in Semiconductor Devices

U. H. Kurzweg, M. Zahn, A. Gregalot and R. E. Wyatt

		Page
1.0	Introduction	144
2.0	Integral Equation Method	145
3.0	Numerical Results for Constant Gradient Junction	148
4.0	Concluding Remarks	161
5.0	List of Symbols	162

Chapter III

An Integral Equation-Relaxation Procedure for the Determination of Equilibrium Potentials in Semiconductor Devices

by

U. H. Kurzweg, M. Zahn, A. Gregalot and R. E. Wyatt

1.0 Introduction

One of the more important but as yet incompletely solved problems in microelectronics is the determination of equilibrium potentials and carrier densities in semiconductor devices of specified geometry, doping distribution and applied external voltage. In principal these properties can be determined by a solution of the governing non-linear, coupled, partial differential equations, however in practice, this is often economically unfeasible because of the large computation times required in an application of the standard finite difference method. Accordingly it is of interest to explore other numerical solution techniques for these equations with the hope of finding approaches requiring shorter CPU times. It has been our group's overall objective to examine such alternate solution techniques and in particular to explore the possiblity of using an equivalent integral equation approach, which is known in other areas such as mechanics to often simplify the solution procedure for boundary value problems. Our ultimate objective would be to apply such an integral technique to more complicated semiconductor geometries such as encountered in MOSFET and BIPOLAR devices. In order to simplify our problem somewhat we have concentrated in this initial phase of our study on the case of a one-dimensional situation with no net electron and hole current, no time dependence, and no recombination. In this limit the governing Maxwell and continuity equations uncouple and one is left with solving essentially a single non-linear Poisson equation with a specified doping distribution and a given set of boundary conditions.

Our discussion below will begin with a description of the equivalent non-linear Fredholm integral equation for the problem. This will be followed by a presentation of the Picard iteration technique used to solve the integral equation and the introduction of a relaxation procedure used to insure convergence. Finally, some numerical results for the special case of a linear graded junction are presented employing both constant and spatially dependent relaxation parameters. A comparison of CPU times for our integral solution approach with existing finite difference solutions is also made.

2.0 Integral Solution Method

Within the framework of the restrictions mentioned, the equation governing the normalized potential U(x) in a one-dimensional semiconductor is the non-linear Poisson equation [1]

$$\frac{d^2U}{dx^2} = - \frac{q^2}{\varepsilon_o kT} [p(U)-n(U)+N_D(x)-N_A(x)] = F(x,U) \qquad (3-1)$$

where p(U) and n(U) are hole and electron number densities which are functions of the potential via the standard Boltzmann relations. Here $N_D(x)-N_A(x)$ is the net doping distribution and the remaining constants have their standard meaning. The corresponding boundary conditions for the case of a doping distribution of odd symmetry assume the form

$$U(0) = 0, \; U(W) = U_T \qquad (3-2)$$

where U_T is a potential determined by depletion layer arguments and one which depends on the particular doping distribution $N_D(x)-N_A(x)$ and the applied external bias V_A. W represents the approximate depletion layer half-width and is typically of the order of 1000Å.

The above nonlinear differential equation together with the given boundary conditions can be converted to an equivalent

---

[1] S. M. Sze, *Physics of Semiconductor Devices*, John Wiley and Sons, Inc., N.Y. (1969).

Fredholm integral equation using an appropriate Greens function. This equation has been derived in our previous report [2] and reads

$$U(x) = \frac{U_T x}{W} - \int_0^x \frac{t(W-x)}{W} F(t,U)dt - \int_x^W \frac{x(W-t)}{W} F(t,U)dt \quad . \qquad (3-3)$$

This integral equation can be solved numerically by Picard iteration used in conjunction with an under-relaxation procedure. Specifically, one needs to evaluate the iterative form

$$U_{n+1}(x) = \frac{U_T x}{W} - \int_0^x \frac{t(W-x)}{W} F(t,U_n)dt - \int_x^W \frac{x(W-t)}{W} F(t,U_n)dt \qquad (3-4)$$

with U_1, U_2, U_3..... forming a convergent sequence provided that the error ε satisfies

$$\left| \frac{U_{n+1}(x) - U_n(x)}{U_n(x)} \right| < \varepsilon \text{ for } n > N \quad . \qquad (3-5)$$

In the present study we consider the sequence to have converged to the sought after solution $U(x)$ whenever $\varepsilon < 10^{-3}$ for all x in $[0,W]$ and use the linear form $U_T x/W$ for the zeroth approximation $U_0(x)$.

The Picard iteration procedure Eq. (3-4) will generally converge only for small F and W [2]. As this is often not the case for physically realizable conditions, there is a need to introduce a relaxation procedure such as discussed by Forsythe and Wasow [3]. The simplest such relaxation procedure which will slow down the rate of change between successive iterations and hence avoid numerical instabilities has the form

$$U^*_{n+1}(x) = U^*_n(x) + \frac{1}{K}\left[U_{n+1}(x) - U^*_n(x)\right] \qquad (3-6)$$

---

[2] D. P. Kennedy, *Computer Aided Engineering of Semiconductor Integrated Circuits*, EDRC Tech. Report 75-1344-1, pg. 129, Jan. 1976.

[3] G. E. Forsythe and W. R. Wasow , *Finite Difference Methods for Partial Differential Equations*, p. 246, John Wiley and Sons, Inc., N.Y. (1960).

where K is a constant under-relaxation parameter greater than unity and the starred functions represent the relaxed iterates. It should be pointed out that when this relaxation procedure is used, the U_n in the term F of expression Eq. (3-4) is replaced by U_n^* and also that the unrelaxed solution is recovered as K goes to one.

The relaxation procedure Eq. (3-6) will always lead to convergence of the Picard iteration Eq. (3-4) provided K is large enough. This fact can be shown by the following argument. Consider the iterative form of a slightly generalized Fredholm equation

$$U_{n+1} = f(x) + \int_0^W H[x,t,U_n^*(t)]dt \qquad (3-7)$$

and let the maximum value of H for x and t in [0,W] be M. Now introduce the infinite series

$$U(x) = U_0 + (U_1^*-U_0) + (U_2^*-U_1^*) + \ldots. (U_{n+1}^*-U_n^*) + \ldots \qquad (3-8)$$

which should converge if the iteration is numerically stable. Starting with the approximation $U_0^*=U_0=f(x)$ we find

$$(U_{n+1}^*-U_n^*) < \frac{MW}{K^n}(PW+K-1)^{n-1} \qquad (3-9)$$

where

$$P = \underset{\text{in } x,t}{\text{Max}} \{[H(U_{n+1}^*)-H(U_n^*)]/[U_{n+1}^*-U_n^*]\} . \qquad (3-10)$$

Upon substituting this inequality into series Eq. (3-8) we obtain a geometric series which converges provided K is large enough. Although expression Eq. (3-9) does not yield an explicit value for the value of K which will lead to the most rapid convergence, it suggests that this might occur when $K \approx 1-PW$.

3.0 Numerical Results for the Constant Gradient Junction

Specific numerical results have been obtained using the above described integral equation-relaxation procedure for the special case of a constant gradient p-n junction with an impurity distribution $N_D(x)-N_A(x)=Ax$, where A is the grade constant taken as 10^{22} atoms/cm^4. Our choice of this relatively simple doping distribution was that it represent a first approximation for the more complicated coerror function and Gaussian profiles and that solutions to this problem using the standard finite difference approach are known (cf. Kennedy and O'Brien [4]). For this case, as already shown earlier by Morgan and Smitz [5], Eq. (3-1) assumes the form

$$\frac{d^2U}{dx^2} = \frac{2n_i q^2}{\kappa\varepsilon_o kT}\left[\sinh U(x) - \frac{Ax}{n_i}\exp - \frac{qV_A}{2kT}\right] e^{\frac{qV_A}{kT}} \qquad (3-11)$$

and is still subjected to the boundary conditions Eq. (3-2) which now assume the explicit form

$$U(0) = 0;\ U(X) = \ln\left(\frac{AW}{n_i}\right) - \frac{qV_A}{kT} = U_T \qquad (3-12)$$

where A is again the grade constant, n_i the intrinsic number density, and V_A the applied external bias. The corresponding integral equation recast into its Picard iterative form is

$$U(x)_{n+1} = \frac{U_T x}{W} - \gamma\int_o^x \frac{t(W-x)}{W}\left[\sinh U_n^*(t) - \frac{At}{2n_i}\exp - \frac{qV_A}{2kT}\right] dt$$
$$- \gamma\int_o^W \frac{x(W-t)}{W}\left[\sinh U_n^*(t) - \frac{At}{2n_i}\exp - \frac{qV_A}{2kT}\right] dt \qquad (3-13)$$

---

[4] D. P. Kennedy and R. R. O'Brien, *IBM Jour. Res. Dev.*, 11, May 1967.

[5] S. P. Morgan and F. M. Smitz, *Bell Syst. Tel. Jour.*, 29, 1573 (1960).

where $\gamma = [2n_i q^2 \exp qV_A/2kT]/\kappa\varepsilon_o kT$. The integration interval [0,W] here extends into the n type region and the resultant solution U(x) will have odd symmetry about x=0.

Equation (3-13) together with the relaxation procedure Eq. (3-6) were programmed for solution using an IBM 370/65 computer with interactive APLSV*PLUS routines and a graphics terminal. Also the solution development was examined using a graphics equipped PDP-11/40 minicomputer. The first phase of this numerical analysis was concerned with finding the value of the constant relaxation parameter K in expression Eq. (3-6) which requires the minimum number of iteratives for convergence. The results of this study for $V_A=0$ are summarized in Fig. 3.1. We see here that for an error $\varepsilon<10^{-3}$ the best value of the relaxation parameter increases with increasing width W such that at W=1200Å the value is K=2 while at W=4000Å this value has risen to K=100. The best value of constant K in the interation-relaxation procedure is seen to increase as the third power of the width showing that the expected CPU time will increase rapidly for large integration ranges [0,W]. In Fig. 3.2 we show the relation between the number of iterations n required for convergence in the Picard procedure and the size of the under-relaxation parameter K used. Note that for fixed W, stable solutions generally do not exist until K exceeds a certain value and that the best value of K producing the must rapid convergence is only slightly larger than this. Convergence for very large K is always possible, as expected, but note that at the same time the number of iterations required becomes very large. The corresponding equilibrium potential distributions obtained by using the optimum value of K given in Fig. 3.1 are shown in Figs. 3.3-3.6 and represent the results for the half-widths W=1000Å, 2000Å, 3000Å and 4000Å, respectively. We have also recorded the values of the relaxed iterates $U_n^*(x)$ in these figures to show how the final stable solution is approached. Note that in all four cases the most rapid variation occurs in the range 0<x<1000Å. This is what

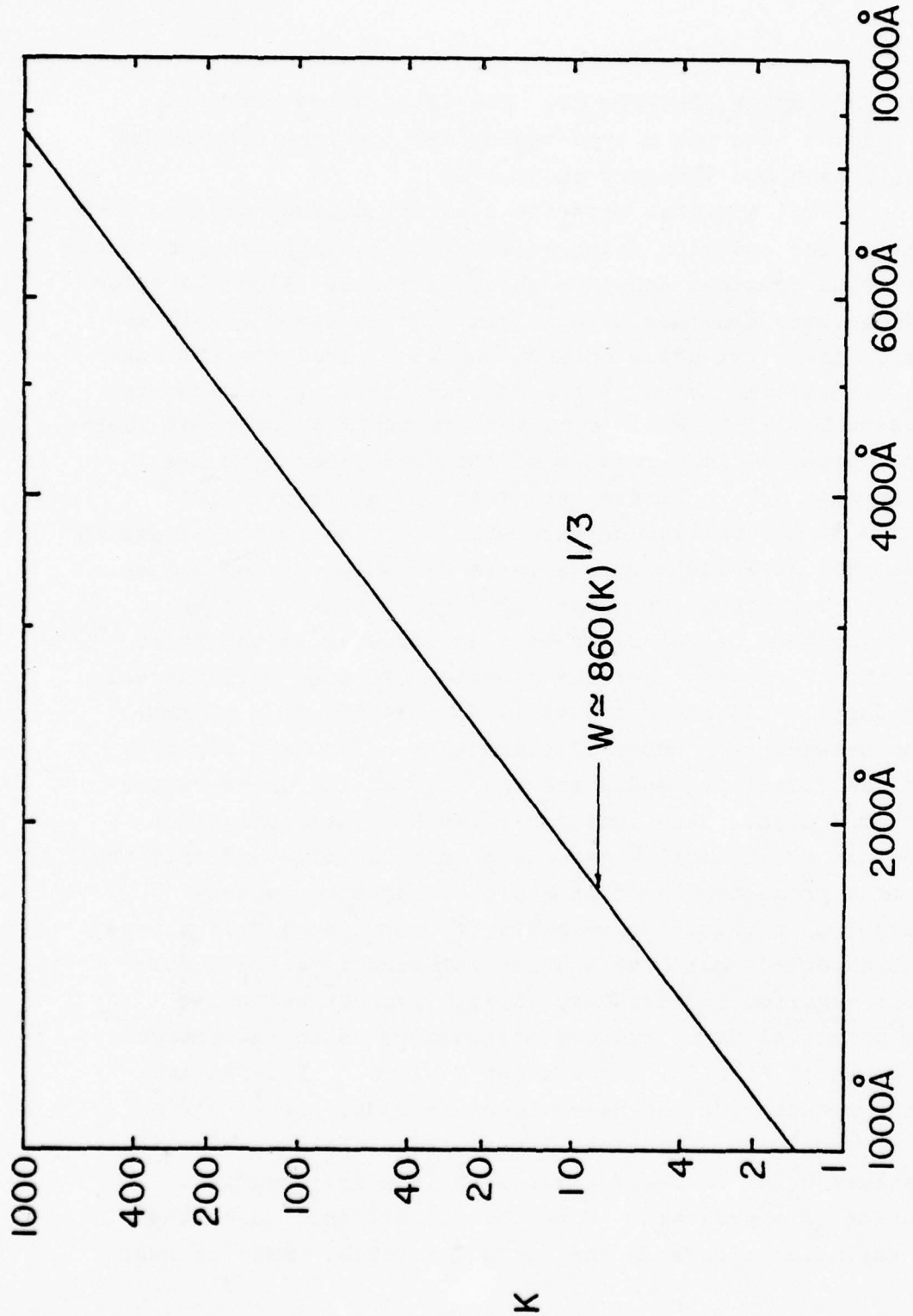


Fig. 3.1. Optimum relaxation parameter K as a function of width W for a constant gradient junction and zero applied voltage V_A.

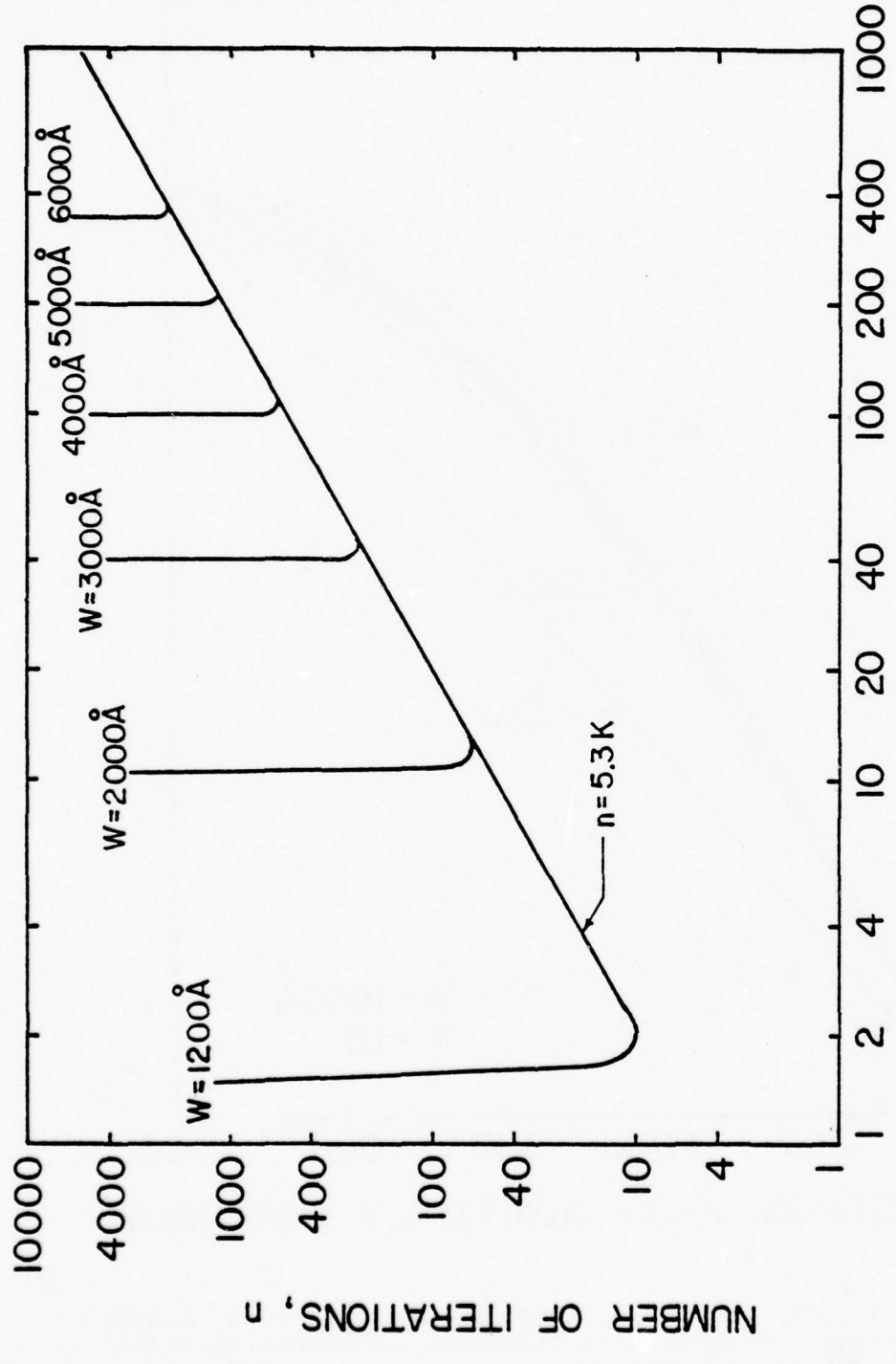


Fig. 3.2. Number of iterations required for $\varepsilon<10^{-3}$ as a function of the constant relaxation parameter K.

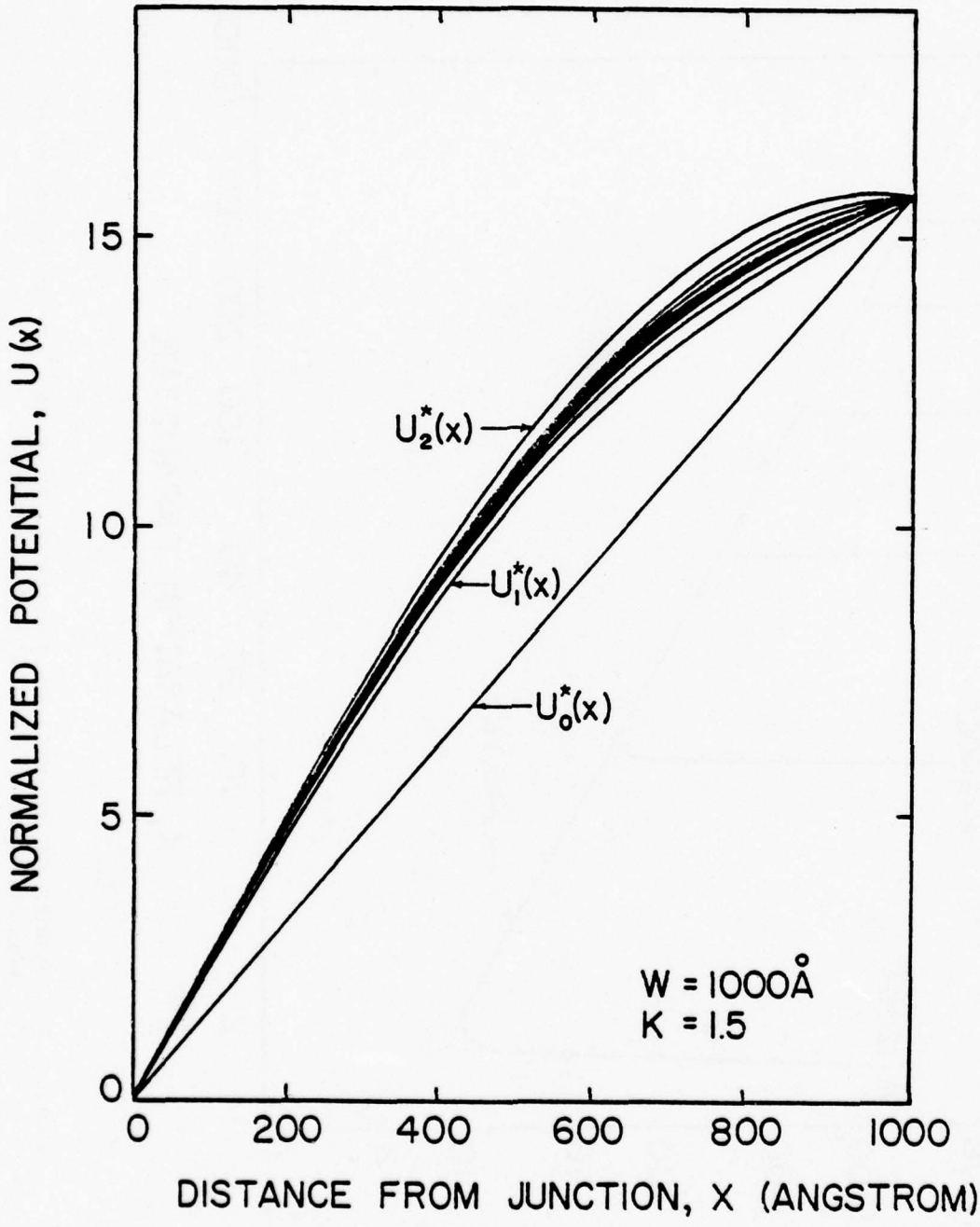


Fig. 3.3. Calculated potential distribution for a zero biased, constant gradient, P-N junction. Integration width of 1000Å.

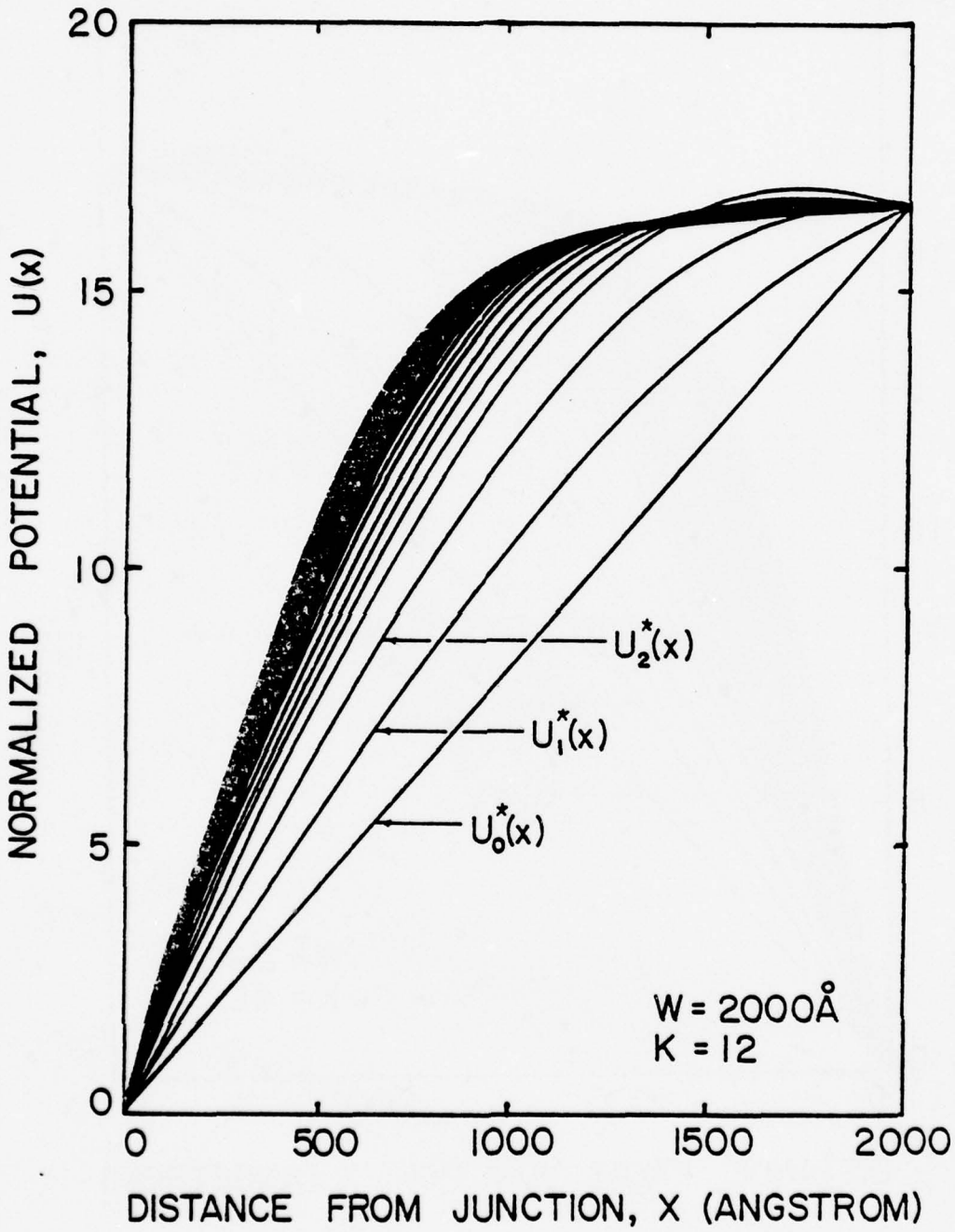


Fig. 3.4. Calculated potential distribution for a zero biased, constant gradient, P-N junction. Integration width of 2000Å.

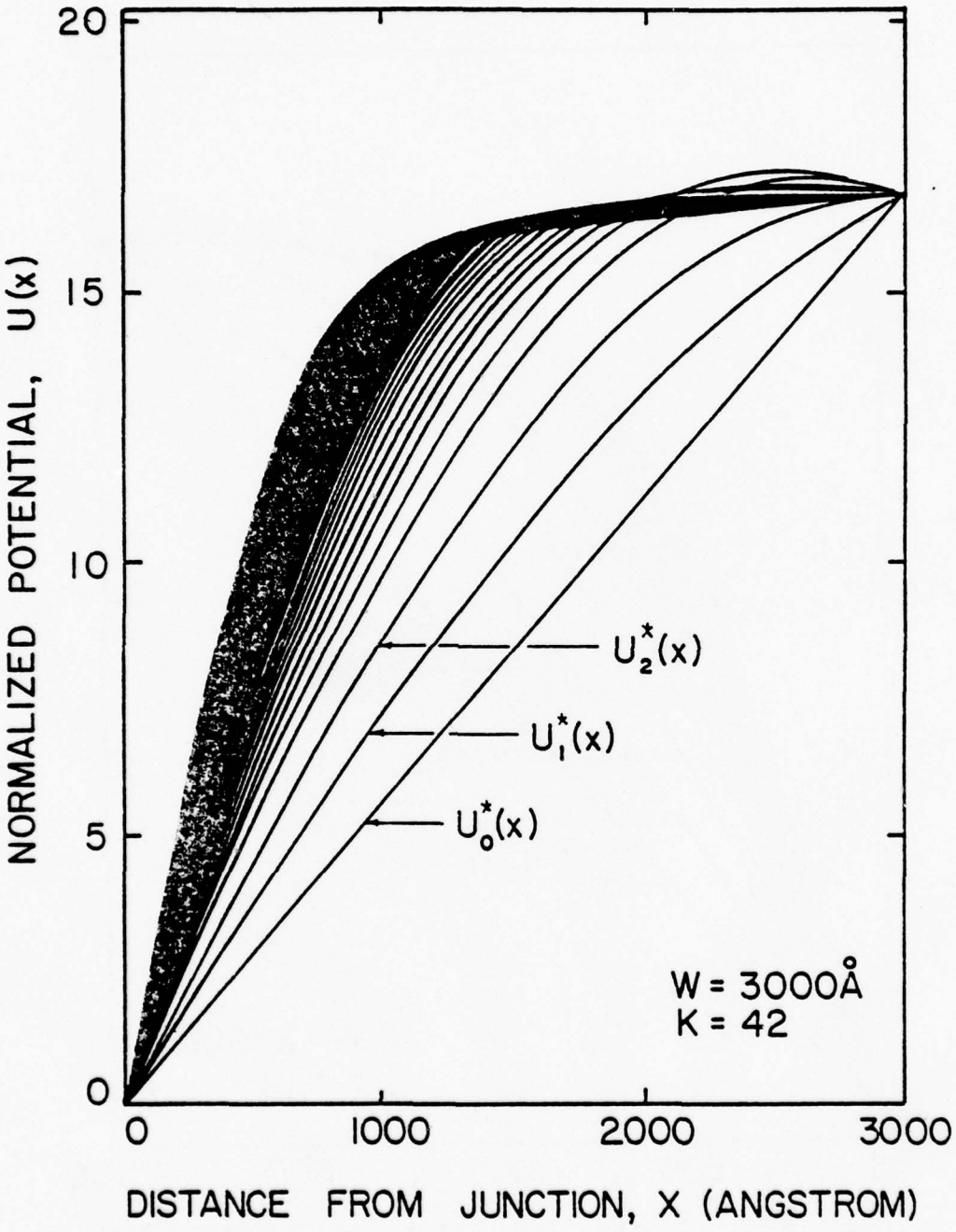


Fig. 3.5. Calculated potential distribution for a zero biased, constant gradient, P-N junction. Integration width of 3000Å.

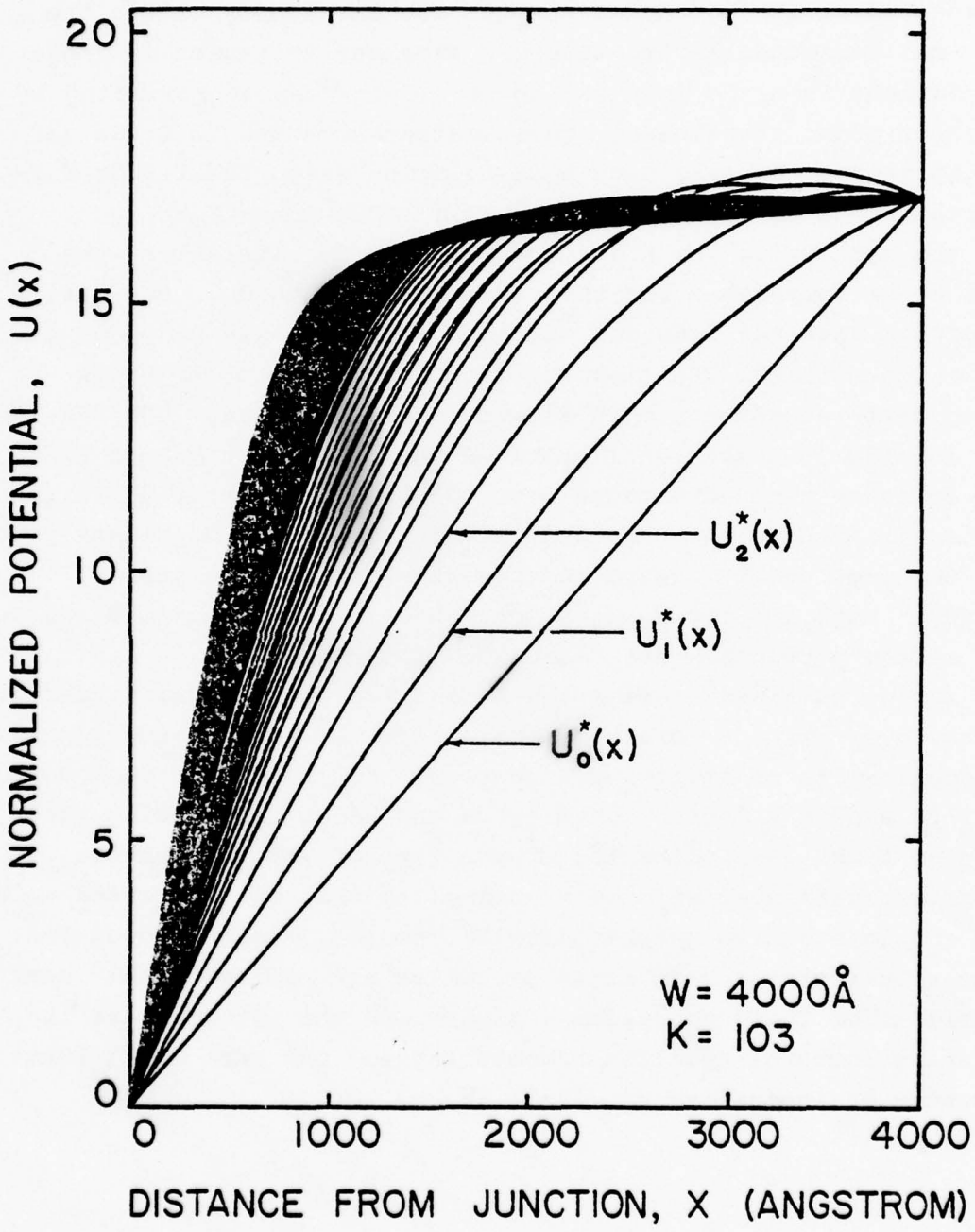


Fig. 3.6. Calculated potential distribution for a zero biased, constant gradient P-N junction. Integration width of 4000Å.

is expected as the potential changes are large only within the classical depletion layer. The CPU time per iteration in these calculations when the integral in Eq. (3-13) was approximated by a fifty-one point trapezoidal approximation amounted to 0.014 sec per iteration. We also made a calculation using the finite difference method on Eq. (3-11) to obtain a CPU comparison. Both the best value for K and the CPU time per iteration were found to be comparable for the two solution methods. Our earlier contention that the integral approach is faster was based on a comparison with the CPU time required for this problem using finite differences on a much slower machine. It may, however, be that for higher dimensional problems the integral approach will have definite time advantages over the standard finite difference approach. In Fig. 3.7 we indicate the spatially dependent potential obtained under forward and reverse bias for the case of $W=1200\mathring{A}$. Note the increase and decrease in potential relative to $V_A=0$ as the potentials are reversed.

Having obtained equilibrium potentials for several different W's using a constant relaxation parameter K, we next investigated the possibility of employing a spatially and iteration dependent $K_n(x)$ to aid in solution convergence and hence in reducing the computer time. An indication of the type of variable under-relaxation parameter which can accomplish this was suggested to us from our interactive graphic display studies. These showed that large values of the relaxation parameter are necessary only near the middle of the integration region where the integrand in the governing integral equation becomes large. One such $K_n(x)$ function suggested by inequality Eq. (3-9) is

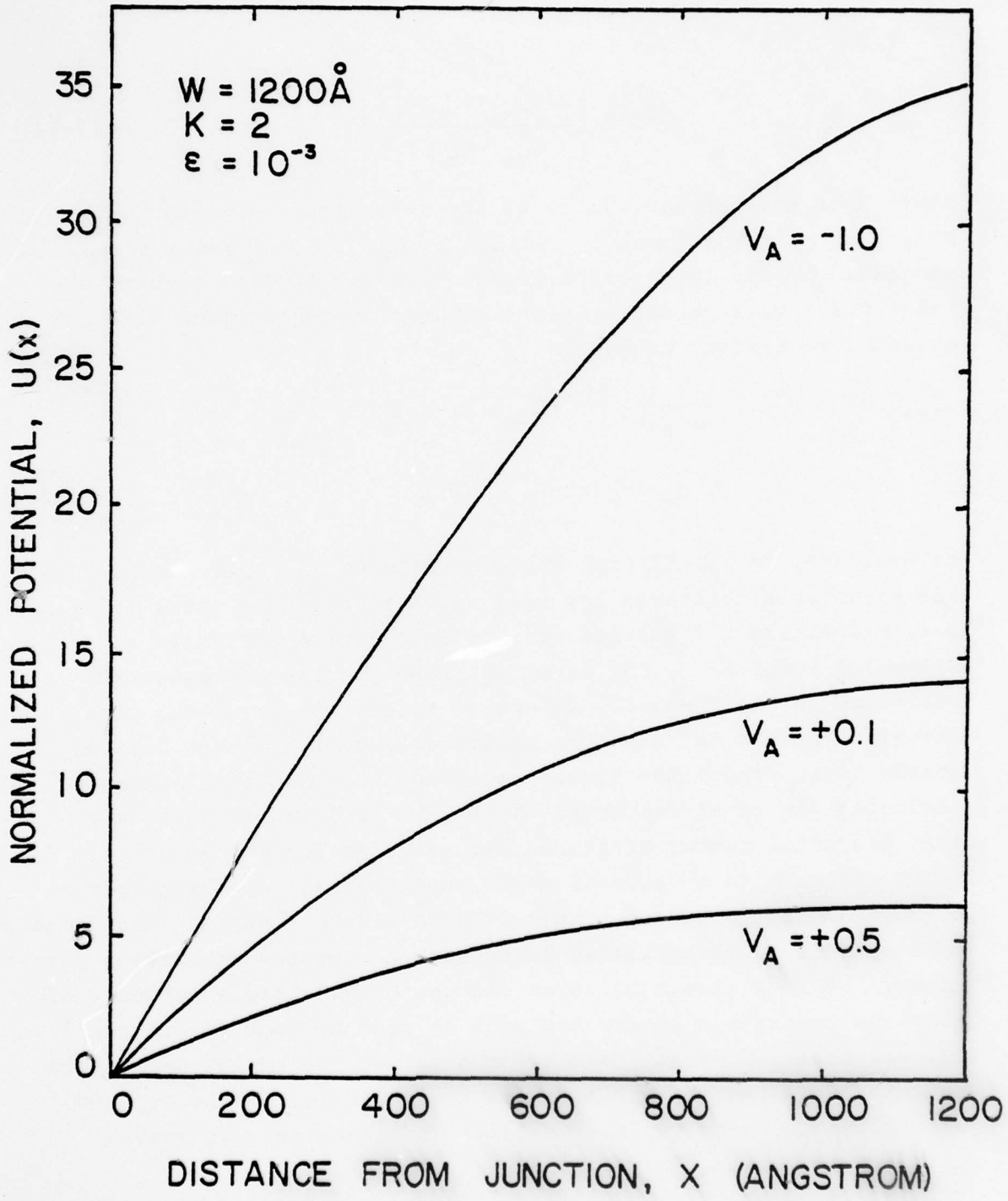


Fig. 3.7. Potential distribution in a forward and reverse biased, constant gradient, P-N junction for W = 1200Å.

$$K_n(x) = 1 + \frac{(K-1)}{P}\left[\frac{H(U^*_{n+1})-H(U^*_n)}{U^*_{n+1}-U^*_n}\right]^{\alpha} \qquad (3-14)$$

where K is the constant value of the relaxation parameter shown in Fig. 3.1, P is the quantity defined by Eq. (3-10), and α a constant suggested by our interactive graphic display studies to have the value 0.2. This variable $K_n(x)$ was used in conjunction with the extended relaxation procedure

$$U^*_{n+1}(x) = U^*_n(x) + \frac{1}{K_n(x)}\{c_1[U_{n+1}(x)-U^*_n(x)] + c_2[U_n(x)-U^*_{n-1}(x)]\} \qquad (3-15)$$

to evaluate Eq. (3-13) for the case W=3000Å and V_A=0. Here c_1 and c_2 are adjustable constants and were taken to have the value 1 and 0.4, respectively. Results of this calculation are shown in Figs. 3.8 and 3.9. The first of these figures indicates the variation in the spatially dependent relaxation parameter with iteration number and how this parameter rapidly reaches a final steady form. The other figure represents the potential iterates including the final converged value. It is found in this case that the total number of iterations required for an error $\varepsilon<10^{-3}$ is 60 compared to a value of n=240 when the best constant relaxation parameter is used for this W=3000Å width. Thus, a considerable time savings is accomplished using such a variable relaxation parameter. We have also considered the use of a multiple sequence of constant relaxation parameters such as used by Mock [6] in our calculations. However, this has not as yet led to any definitive numerical results and may not have the advantage of the type of spatially-dependent variable relaxation parameter used above.

---

[6] M. S. Mock, *Solid State Electronics*, 16, 601 (1973).

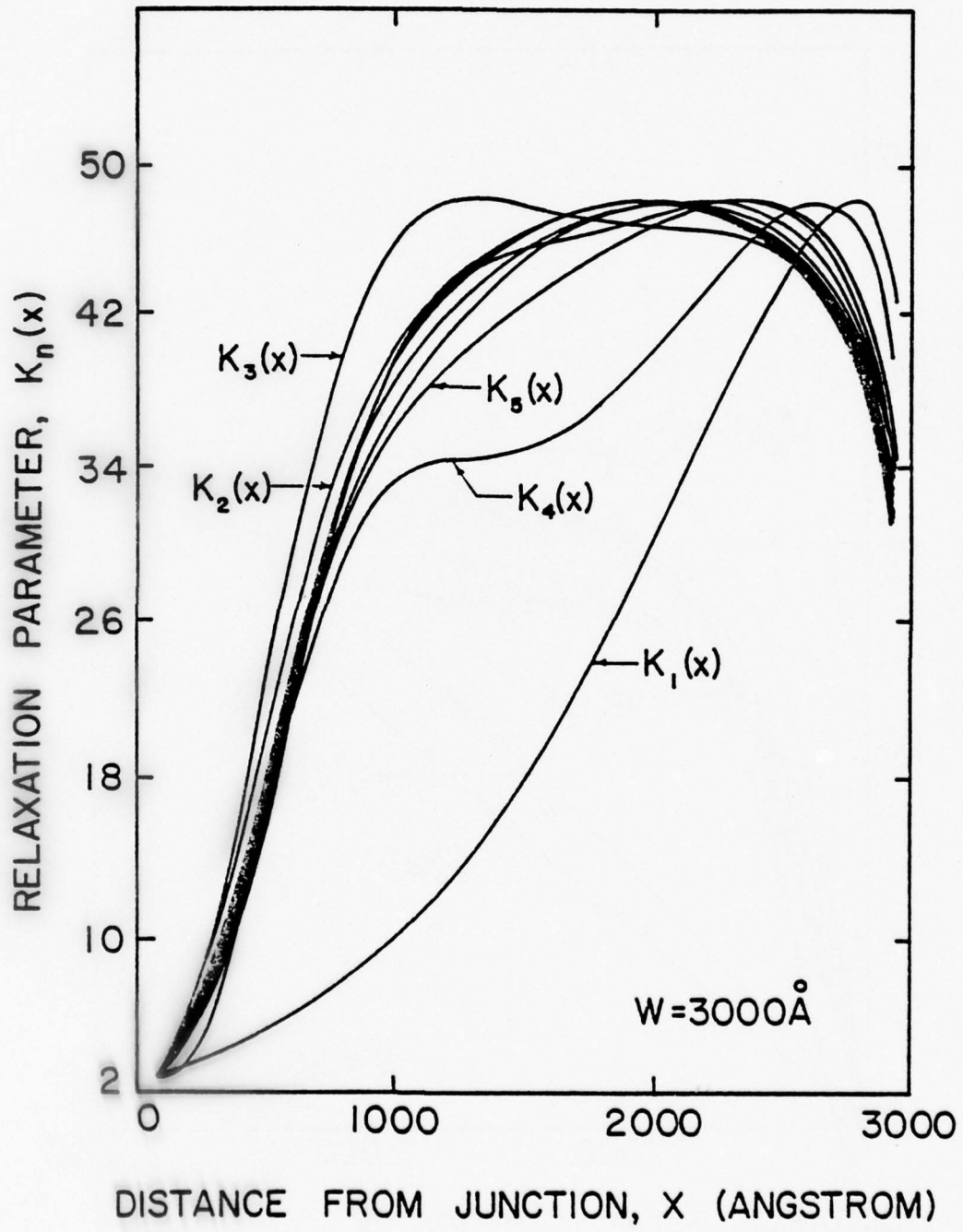


Fig. 3.8. Generated variable relaxation parameter $K_n(x)$ for a constant gradient junction with W = 3000Å.

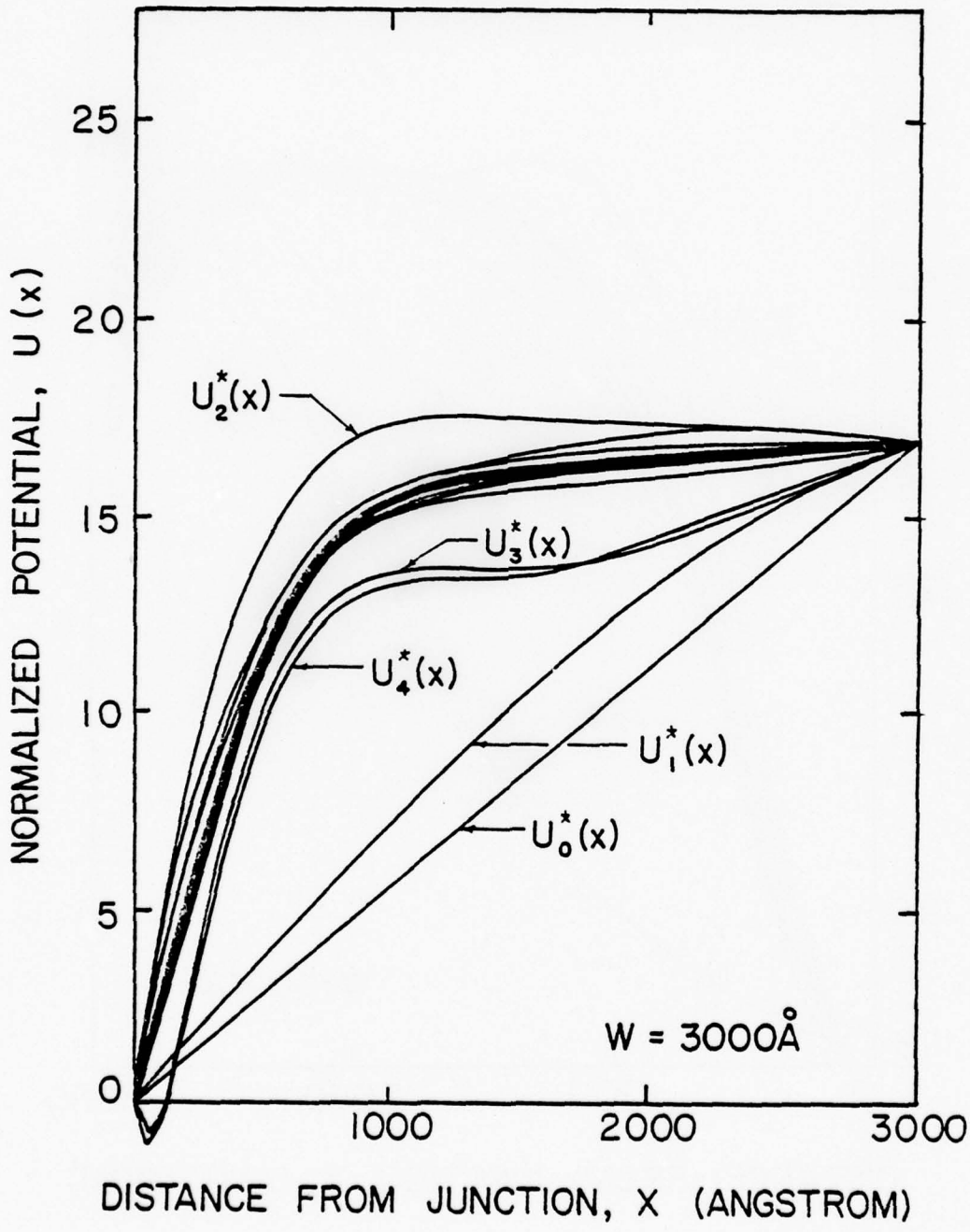


Fig. 3.9. Equilibrium potential for a constant gradient P-N junction using a variable relaxation parameter. Here W = 3000Å and $V_A = 0$.

4.0 Concluding Remarks

We have shown that the non-linear Poisson equation governing the potential in a one-dimensional semiconductor for specified doping distribution and under conditions where this equation decouples from the remaining equaticns can be converted to an equivalent integral equation. This integral equation has been solved using a Picard iteration procedure in conjunction with two different relaxation procedures. Explicit values for both the optimum relaxation parameter and the corresponding equilibrium potential are obtained for the case of a linear graded junction mainly for zero applied external voltage. The method is found to converge for all depletion half-widths W considered with the number of iterations required for convergence increasing with increasing W. CPU times for the integral solution method used here are comparable to those required by the finite difference method for the same number of interval subdivisions, but the method becomes somewhat faster when a spatially dependent relaxation parameter is used. No additional difficulties are expected in carrying out similar calculations as the above for other doping profiles including the Gaussian and error function diffusion profiles. Extension of the integral approach to higher dimensions is also a possibility and it is in this later area where the real advantage of the integral technique can be expected to be demonstrated. Some preliminary work has been done along these lines by generating Greens functions for some simple two-dimensional geometries but numerical results have not yet been obtained for this case. It seems likely in view of the above results obtained during the first year of the project that perhaps the most profitable approach for reducing CPU time in numerical solution of the equation governing semiconductor characteristics will be the use of improved spatially dependent relaxation parameters and not so much whether the finite difference or integral approach is employed.

5.0 List of Symbols

κ	Relative dielectric constant
ε_o	Permittivity of free space (8.854×10^{-14} Farad/cm)
k	Boltzmann constant
T	Temperature in degrees Kelvin
p	Hole concentration (cm^{-1})
n	Electron concentration (cm^{-1})
q	Charge on electron (1.6×10^{-19})
n_i	Intrinsic carrier concentration
N_D	Donor concentration (cm^{-3})
N_A	Acceptor concentration (cm^{-3})
U	Normalized potential ($U=qV/kT$)
V_A	Applied bias voltage in volts
W	Width of integration interval in Angstroms $\mathring{A}$
K	Constant relaxation parameter
$K_n(x)$	Variable relaxation parameter
A	Grade constant (cm^{-4})

Chapter IV

Test Pattern Model Using Monte Carlo Methods of Analysis

S. C. Taylor and D. P. Kennedy

		Page
1.0	Introduction	164
2.0	The Mathematical Model	165
3.0	The Number Generator	166
4.0	Analysis	175
5.0	Conclusion	183

CHAPTER IV

Test Pattern Model Using Monte Carlo Methods of Analysis

by

S. C. Taylor and D. P. Kennedy

1.0 Introduction

Semiconductor device fabrication techniques frequently exhibit reproducibility problems, due to inadvertent variations in each processing step. This problem is readily observed in the electrical characteristics of devices fabricated by methods involving thermal diffusion of impurity atoms into silicon [1,2]. For example, in large volume manufacturing a given diffusion furnace can be shown to introduce processing variations across any given slice of silicon, and from one slice to the next in a given diffusion run. Similarly, in any given manufacturing facility it can be shown that no two diffusion furnaces produce identical results. It is for this reason that sheet resistance test patterns are frequently used to monitor the properties of a given thermal diffusion process; any change of measured sheet resistance implies a change (either inadvertent or intentional) in the given diffusion process.

As a consequence of this situation, for process evaluation a large number of sheet resistance measurements are routinely taken during semiconductor device manufacturing. Frequently these data are evaluated on a statistical basis, establishing the mean sheet resistance and a standard deviation arising from inadvertent processing changes. To date, little information is available on how to relate observed statistical variations of sheet resistance to changes occurring in the numerous processing parameters associated with thermal diffusion. Presented here is a technique whereby such information can be inferred through Monte Carlo techniques of analysis, using a computer.

---

[1] D. P. Kennedy, *IBM Jour.*, 5, 331 (1961).

[2] D. P. Kennedy, P. C. Murley, and R. R. O'Brien, *IBM Jour.*, 8, 482 (1964).

This proposed computational technique involves a one-dimensional mathematical model for the sheet resistance test pattern, in conjunction with a pseudo random number generator. Such a computational tool offers a means whereby the test site fabrication process can be simulated on a computer and, hence, experiments can be performed that would be exceedingly difficult to accomplish on a production line. It will be shown that this method of analysis yields quantitative information concerning the relative sensitivity of sheet resistance measurements to variations in each important process parameter. It will also be shown that this method of analysis can be used to infer for a two-step diffusion process, a statistical frequency distribution for the predeposition impurity atom surface concentration, and similar information about the resulting oxide thickness.

2.0 The Mathematical Model

The mathematical model used in this analysis is one-dimensional in nature, and it approximates the impurity distribution produced by a two-step diffusion process [3]. After the predeposition step, this impurity atom distribution is approximated by the relation

$$C_1(x,t) = C_o \operatorname{erfc}\left[\frac{x}{2\sqrt{D_1 t_1}}\right] . \tag{4.1}$$

Similarly, after the oxidation step we use the approximation

$$C_2(x_1 t_1, t_2) = 2\left[\pi D_2 (t_2 - t_1)\right]^{-1/2} \int \left\{ \exp\left[- \frac{(x-x')^2}{4D_2(t_2-t_1)} \right] + \exp\left[- \frac{(x+x')^2}{4D_2(t_2-t_1)} \right] \right\} C_1(x', t_1)\, dx' , \tag{4.2}$$

where t_1 and t_2, respectively, represent the predeposition and oxidation diffusion times.

---

[3] D. P. Kennedy and P. C. Murley, *Proc. IEEE*, 52, 620 (1964).

It has been shown that oxide growth on silicon produces a redistribution of impurity atoms, due to differences of solubility between silicon and SiO_2. Thus, after this oxidation step Eq. 4.2 is modified by the following relation [4],

$$C_3(y,t_1,t_2) = C_2(y,t_1,t_2) - (0.475)C_2(0,t_1,t_2\left\{\frac{\text{erfc}\left(\frac{4}{2\sqrt{D_2t_2}}\right)}{\text{erfc}\left(\frac{0.44W_{ox}}{2\sqrt{D_2t_2}}\right)}\right\} \quad (4.3)$$

where y is the distance from the SiO_2 and silicon interface, and W_{ox} is the oxide thickness.

Thus, two basically different equations are used to model the sheet resistance arising during a two-step diffusion process. The first expression represents this resistance immediately after the predeposition of impurity atoms,

$$\rho_s^{-1} = q\int_0^{x_j} \mu(x)\Big\{C_1(x,t_1) - C_B\Big\}dx, \quad (4.4)$$

and

$$\rho_s^{-1} = q\int_0^{y_j} \mu(x)\Big[C_3(y,t_1,t_2) - C_B\Big]dx \quad (4.5)$$

represents this resistance after the oxidation step. In both Eqs. 4.4 and 4.5 the mobility is dependent upon the total impurity atom concentration [5].

3.0 The Number Generator

As previously stated, a two-step diffusion cannot be considered an invariant process. Experimental variations exist within any given diffusion run, and from one diffusion run to the next. Thus, in a computer simulation of this diffusion

---

[4] A. S. Grove, O. Leistiko, and C. T. Sah, *Jour. of Appl. Phys.*, 9, 2695 (1964).

[5] J. C. Irvin, *Bell Syst. Tech. Jour.*, 41, 387 (1962).

process the following variables must be assigned a mean value and an experimental tolerance:

C_o impurity source concentration during predeposition,

C_B uniform background doping of semiconductor material,

T_1 temperature of predeposition,

T_2 temperature of drive-in,

W_{ox} resulting thickness of oxide layer.

The experimental variations of these process parameters are approximated by a random number generator that simultaneously and independently modifies each parameter throughout its specified tolerance limits. Two different statistical filters are used to modify the generator's output, one to produce a normal distribution with specified mean and variance, and another to produce a logarithmic normal [6] distribution with specified mean, variance and degree of skewness.

The random number generator used in this computer simulation is from the Harwell Subroutine Library (Program FA01AS). This generator uses the linear congruential method to generate a sequence of uniformly distributed random numbers on the interval (0,1). These numbers $\langle U_n \rangle$ are obtained by applying the recursion relation

$$U_{n+1} = (aU_n + c) \bmod m \qquad n \geq 0,$$

where

$$a = 3^{15}$$
$$c = 0$$
$$m = 2^{32}$$
$$U_o = 21845.$$

Since the validity of the statistical results obtained in this analysis depends on the effectiveness of this random number generator, a study was undertaken to determine the randomness of

---

[6] H. Crámer, *The Elements of Probability Theory*, Wiley, p. 118, 1955.

any sequence of numbers produced by this generator. The field of statistics provides many quantitative testing procedures for determining the randomness of a particular sequence. The fact that a particular sequence passes an arbitrary number of these tests is no guarantee that it will pass all such tests. However, in practice several tests are carried out, and with each successful test our confidence in the generator increases. After several such tests have been applied with good results, the generator is assumed to produce a suitably random sequence of numbers.

The following statistical evaluation methods were selected (from Knuth [7]) to test the Harwell generator sequence:

a. The Equidistribution (Frequency) Test
b. The Gap Test
c. The Permutation Test
d. The Maximum of t Test
e. The Serial Correlation Test

Two general procedures are employed by these evaluation methods: (1) The Chi-Square Test, and (2) The Kolmogorov-Smirnov Test.

The Chi-Square Test

Assume that every observation from this generator falls into one of k categories. A fairly large number of independent observations is made. Let p_s be the probability that each observation falls into category s, and let Y_s be the number of observations which actually do fall into category s. The following statistic is formed from n observations:

$$V = \sum_{s=1}^{k} \frac{(Y_s - np_s)^2}{np_s} . \qquad (4.6)$$

Assuming k categories in Eq. (4-6), it can be shown that the parameter V has κ degrees of freedom, where $\kappa = k-1$. Thus, from Chi-Square tables we can establish the probability that V

---

[7] D. E. Knuth, *The Art of Computer Programming, "Semi-numerical Algorithms,"* Addison-Wesley (1971).

will assume a particular value. Suitably applied, this evaluation method is used to test a random number generator in several of the above listed categories.

The Kolmogorov-Smirnov (KS) Test

The distribution of a random quantity U can be specified in terms of a distribution function, F(x), where

$$F(x) = \text{probability that } U < x.$$

For a uniformly distributed random number within the interval (0,1), this distribution function assumes the ideal form:

$$F(x) = x \quad 0 \leq x \leq 1$$

$$F(x) = 1 \quad 1 \leq x \leq \infty$$

This KS test involves the generation of such a distribution from the number generator under investigation, and comparing this distribution with the ideal.

This test is accomplished by acquiring a large sequence of numbers from the generator (U_1, U_2, U_3, ..., U_n), and developing the distribution function $F_n(x)$ where

$$F_n(x) = \frac{\text{Quantity in sequence} \leq x}{n}. \tag{4.7}$$

Thereafter, $F_n(x) - F(x)$ establishes the deviation of this sequence from an ideal type of random distribution. Specifically, a maximum value is established for both positive and negative deviations of $F_n(x)$ from the ideal:

$$K_n^+ = \sqrt{n}\ \text{Max}\{F_n(x)-F(x)\} \text{ where } -\infty<x<\infty$$

$$K_n^- = \sqrt{n}\ \text{Max}\{F(x)-F_n(x)\} \text{ where } -\infty<x<\infty \tag{4.8}$$

As in the Chi-Square test, statistical tables have been developed to establish the level of significance for K_n^+ and K_n^-.

The five previously mentioned tests are now described. Each test is applied to a sequence:

$$U_n = U_1, U_2, \ldots, U_n$$

of real numbers, supposedly uniformly distributed, where $0 \leq U_n \leq 1$.

a. The Equidistribution (Frequency) Test

This test determines if the numbers of a sequence are uniformly distributed on the interval of definition. There are two ways to perform this test, one employing the Chi-Square test, the other the KS test.

For the Chi-Square test we generate from the sequence $\langle U_n \rangle$ a new sequence $\langle Y_n \rangle$ where

$$\langle Y_n \rangle = Y_1, Y_2, Y_3, \ldots, Y_n \tag{4.9}$$

and $Y_n = |dU_n|$, where d, a constant, is selected from other considerations. Thereafter, a frequency distribution is generated from the sequence $\langle Y_n \rangle$ for an interval division r where $0<r<d$. Next, the Chi-Square test is applied to this distribution assuming $K = d$ and the probability $p_s = 1/d$ for each category.

This testing technique was applied to two different sequences of numbers from the Harwell Generator:

Number Sequence	Value of (4-6)
1-1000	V = 75.88
3821-4821	V = 68.93

From the Chi-Square tables, the probability of exceeding V = 75.58 is about 25% whereas the probability of exceeding V = 68.93 is about 35%.

A similar evaluation was conducted using the Kolmogorov-Smirnov (KS) test. This is accomplished by generating a distribution for Eq. (4.7) and thereafter determining the magnitudes of K_n^+ and K_n^- in Eq. (4.8). This test was made for three sequences of numbers from the Harwell generator:

Number Sequence	Max	Probability of exceeding
0-1000	$K_n^+ = 0.6261$	29%
	$K_n^- = 0.2566$	78%
1001-2000	$K_n^+ = 0.6893$	39%
	$K_n^- = 0.356$	78%
1191-2911	$K_n^+ = 0.8197$	25%
	$K_n^- = 1.074$	12%

Although evaluating a random sequence is subjective, at best, it is suggested that the Harwell generator properties are adequate to assure that its frequency distribution is satisfactorily uniform [7].

b. Gap Test

This test is used to examine the length of "gaps" between occurrences of specific numbers. We want to establish the lengths of consecutive sequences of numbers from the output of this generator, U_j, U_{j+1}, U_{j+2}, ..., U_{j+r}, in which U_{j+r} lies within the range $\alpha<U_{j+r}<\beta$ but the others do not. This subsequence U_{j+r} represents a gap of length r.

This test is performed by setting up an output "bucket" for the random number generator of size $\beta-\alpha$, and evaluating the lengths of gaps in the output distribution. Thereby, we can tabulate the number of gaps n of lengths 0, 1, 2, ..., t-1, and the number of gaps $\geq t$. Thereafter, the Chi-Square test is applied to this distribution of gap lengths, with suitable probabilities:

$$p_0 = p,\ p_1 = p(1-p),\ p_2 = p(1-p)^2 \ \ldots$$

$$p_{t-1} = p(1-p)^{t-1},\ p_t = (1-p)^t$$

where $p = \beta-\alpha$, the probability that a U will fall within the subinterval (α,β).

This gap test was applied to the Harwell generator for different sequences of numbers. For all sequences, evaluations were made for gaps (β-α) "above the mean," (.5,1), and "below the mean," (0,.5):

Number Sequence	(β-α)	Value of (4-6)	Probability of Exceeding
1-1901	Above mean	3.068	69%
	Below mean	5.468	38%
1911-3820	Above mean	7.86	19%
	Below mean	5.104	42%
1-3600	Above mean	3.432	75%
	Below mean	4.168	63%

It is suggested [7] that these Chi-Square probabilities are indicative of a satisfactory generator.

c. Permutation Test

In this test, a sequence of numbers from the Harwell generator <U> is tested to assure that it does not contain ordered subsequences. For example, if we divide the sequence into n subsequences of 3 numbers, there are 3! possible combinations for these numbers. Because each combination has the probability 1/3! we can apply the Chi-Square test to determine whether the different orderings are uniformly distributed.

This permutation test was applied to the Harwell generator with the following results:

Number Sequence	Value of (4.6)	Probability of Exceeding
1-1500	7.384	22%
1501-3000	6.016	32%
4501-6000	1.648	88%
1-3600	20.907	59%

Again, it has been suggested [7] that this number generator is adequate for our needs.

d. Maximum of t Test

In this test we generate sequences of numbers from the generator

$$V_j = \max(U_{tj}, U_{t(j+1)}, U_{t(j+2)} \cdots U_{t(j+t-1)})$$

and apply the Kolmogorov-Smirnov test to the resulting sequence $V_o, V_1, V_2 \ldots V_{n-1}$ with the distribution function $F(x) = x^t$. In this test, we must show that the distribution function for V_j is $F(x) = x^t$. This test is based upon the fact that the probability of $\text{Max}(U_1, U_2, U_3 \ldots U_t) < x$ is the probability that $U_1 < x$ and $U_2 < x \ldots$, and this is the product of each individual probability -- x.x.x --- = x^t.

This test was performed for several different sequences of numbers from the Harwell generator:

Number Sequence	Max	Probability of Exceeding
1-1500	$K_n^+ = 1.103$	11%
	$K_n^- = 0.278$	84%
1911-3410	$K_n^+ = 0.9110$	21%
	$K_n^- = 0.3853$	74%
1-3600	$K_n^+ = 0.8154$	26%
	$K_n^- = 0.3399$	79%

As before, it is suggested [7] the resulting probability indicate satisfactory randomness in the number generator.

e. Serial Correlation Test

The intent of this test is to evaluate the degree of correlation at the output of this generator between two sequential numbers, (U_j, U_{j+1}). This correlation test is performed by calculating the following correlation coefficients:

$$c = \frac{n(U_oU_1 + U_1U_2 \ldots + U_{n-2}U_{n-1} + U_{n-1}U_o) - (U_o+U_1+\ldots+U_{n-1})^2}{n(U_o^2 + U_1^2 + \ldots + U_{n-1}^2) - (U_o + U_1 + \ldots + U_{n-1})^2} \tag{4.10}$$

This coefficient always lies between -1 and +1. When c is zero (or very small) it indicates U_j, U_{j+1} are relatively independent of each other. When the correlation is near ±1 it represents a complete linear dependence.

A good value of c is conjectured to be between $\mu n - 2\sigma$ and $\mu n + 2\sigma$ [8] where

$$\mu n = -\frac{1}{(n-1)} \qquad \sigma n = \frac{1}{n-1}\frac{n(n-3)}{n+1} \tag{4.11}$$

It has been found that when a normal distribution is assumed for Eq. (4.10), these are the values for Eq. (4.11).

Applying this test to the Harwell number generator, we obtain the following results:

Number Sequence	Values of (4.11)
1-1000	c = -0.0573
1911-3820	c = -0.0120

For adequate lack of serial correlation, the first run (1-1000) should lie between (c = -0.0642) and (c = 0.0622), and the second run (1911-3820) should lie between (c = -0.0463) and (c = 0.0452). Clearly, this test indicates little serial correlation at the output of this number generator.

None of the above tests disqualify the Harwell random number generator, but in no way does this imply that the numbers from this generator are truly random. Instead it is inferred from these results that the generator produces a sufficiently random sequence of numbers which are adequate for the purposes of this study.

4.0 Analysis

Figure 4.1 shows the manner in which this mathematical model is used for resistance calculations. A mean value is selected for each individual process parameter and, in addition, a variability is assigned to these parameters. Thereafter, the number generator produces sets of process parameters that statistically approximate a practical situation involving a large number of individual diffusion runs. Each set of diffusion parameters is introduced into the appropriate model for sheet resistance [Eq. 4.4 or Eq. 4.5] and, thereby, we obtain calculated values of sheet resistance that approximate those obtained from large volume manufacturing.

It is emphasized that one cannot adequately characterize the impurity atom distribution by assigning only process parameters to the variables in Eqs. 4.4 and 4.5. Inpurity atom diffusion into silicon is seldom accurately described by the classical theory of thermal diffusion. For this reason, published values are used for the impurity atom diffusion coefficient, and an iterative procedure is used to approximate the total diffusion time required to attain a given mean junction depth x_j. Thereafter, the influence of small process changes upon this junction location are determined by the relation

$$D = D_o \exp(-\Delta H/RT). \tag{4.12}$$

Throughout this study, the following parameters were used to approximate a fabrication process for which experimental data is available,

C_B	3.5×10^{15} (atoms/cm^3)	±15%
C_o	5.43×10^{19} (atoms/cm^3)	≈±7%
W_{ox}	3630Å	≈±7%
T_1	1100°C	±1.5°C
T_2	1100°C	±1.5°C

A sequence of calculations was first undertaken to determine the relative influence of these parameters upon the calculated sheet resistance.

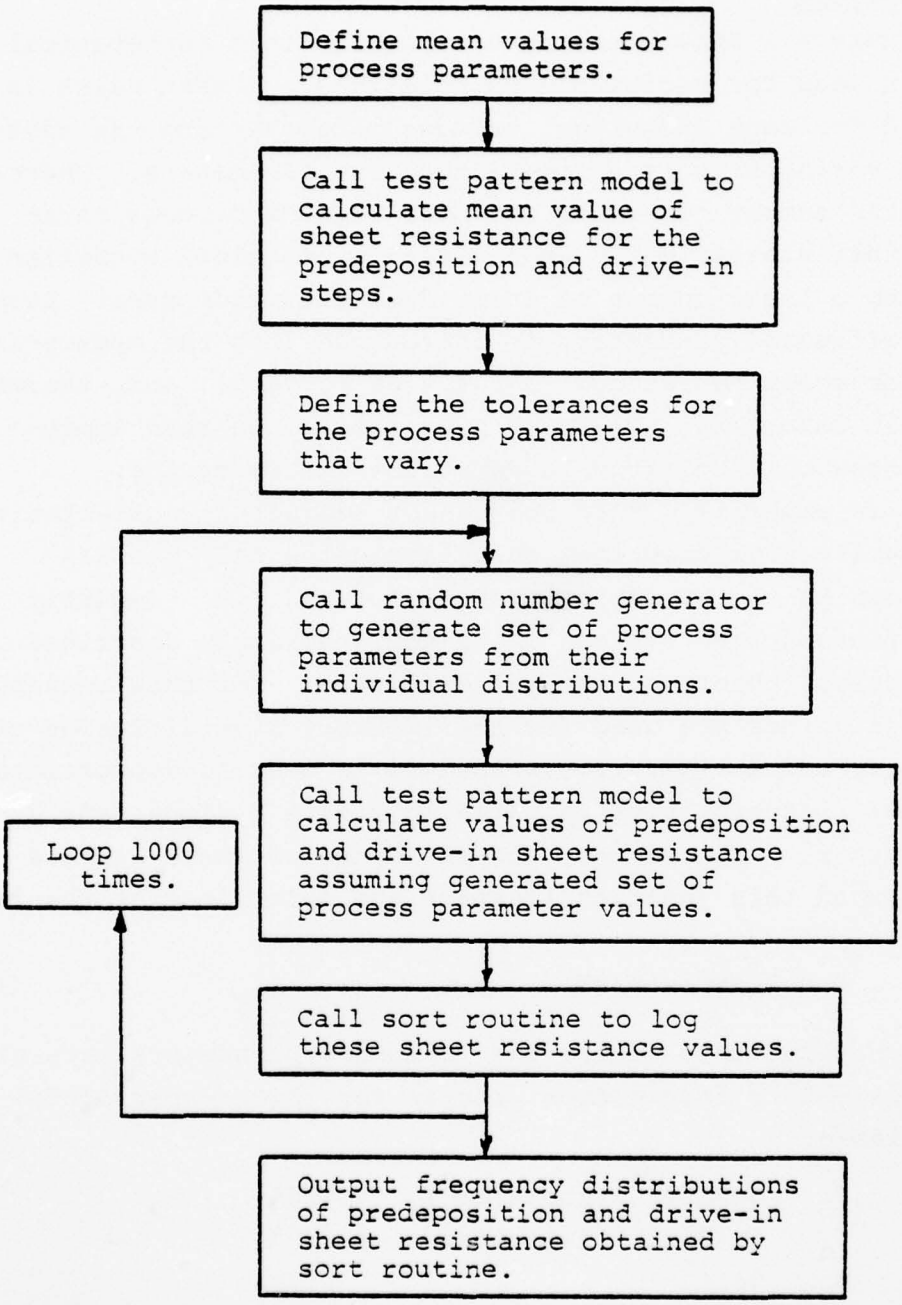


Fig. 4.1 Flowchart of computer program for sheet resistance calculations.

For the predeposition step, the following represents our calculated results:

Variable	Calculated Sheet Resistance mean ($\Omega/\square$)	Std. Deviation
C_o, C_B, T_1	89.31	3.03
C_B, T_1	89.4	0.82
C_o	89.3	2.90

Here we find a sheet resistance standard deviation of 3.03 assuming all process parameters change simultaneously. If, instead, only the background doping (C_B) and temperature (T_1) are permitted to vary this standard deviation is reduced to 0.82, which is exceedingly small. Implied by this series of calculations is the fact that in a typical fabrication process, variations of impurity atom surface concentration (C_o) produce most of the variation observed in measured sheet resistance.

Similar calculations were performed for the oxidation phase of this two-step diffusion process. First, we assume all process parameters associated with the predeposition step are subject to variation, while no process variation is attributable to the oxidation step; these assumptions yield a mean sheet resistance of 140.3 ($\Omega/\square$), with a standard deviation of 4.01. If, instead, we assume the same variabilities in the predeposition step and, in addition, W_{ox} varies by the specified magnitude, this standard deviation increases to 4.3.

A sequence of such calculations readily establishes that there are two process variables producing most of the measured sheet resistance variations observed after oxidation: the predeposition surface concentration (C_o) and the oxide thickness (W_{ox}). This particular situation provides a means whereby we can draw from measured sheet resistance calculations the frequency distributions of both C_o and W_{ox}.

To illustrate this procedure, Fig. 4.2 shows a comparison between the measured and calculated sheet resistance after predeposition. This measured frequency distribution is based upon several thousand measurements, whereas the calculated distribution is based upon a statistical sample of one thousand.

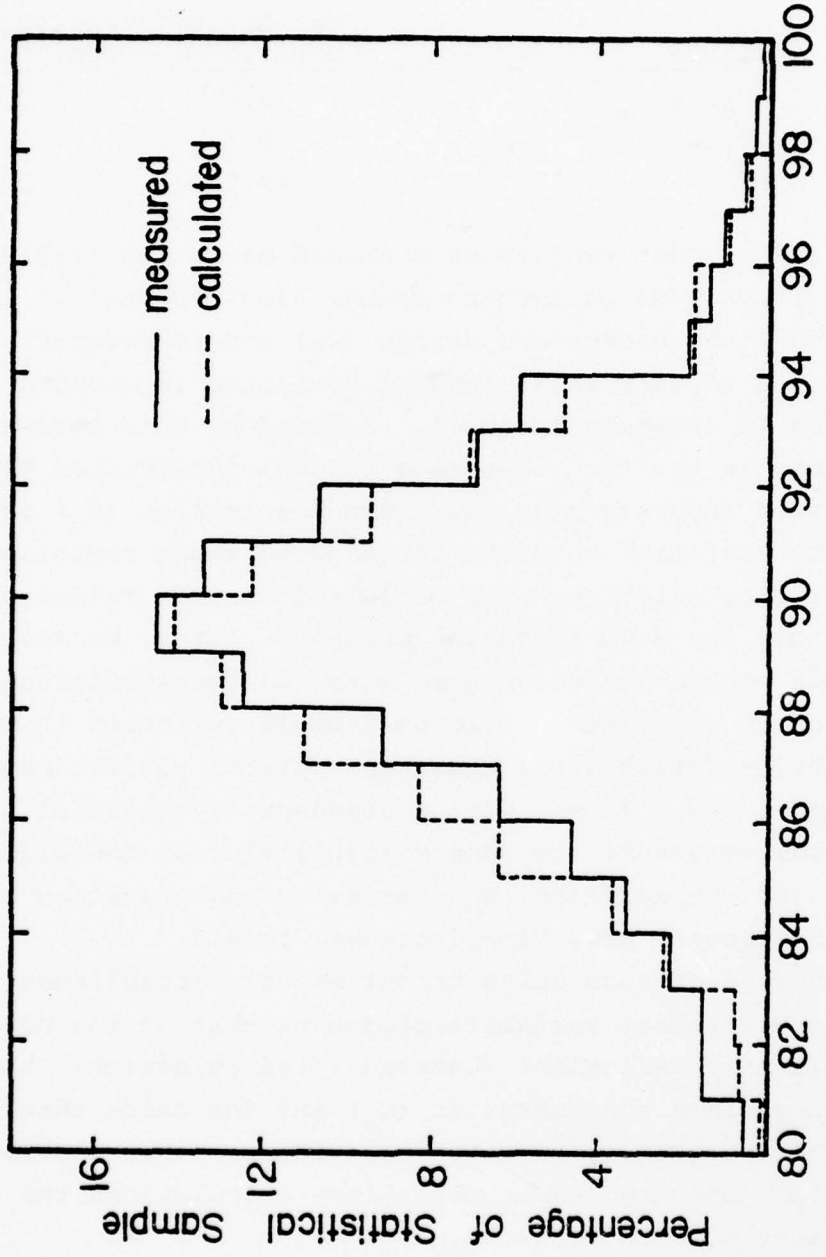


Fig.4.2 Predeposition sheet resistance frequency distribution.

As previously stated, our calculations indicate that in this particular diffusion process the variation of measured sheet resistance is attributable primarily to variations of impurity atom surface concentration. For this reason, the available experimental sheet resistance measurements, in conjunction with our calculated results, can be used to infer the experimental distribution of surface concentration. This is accomplished by adjusting in this mathematical model the C_o frequency distribution to yield calculated sheet resistances that are in substantial agreement with experiment. Using this procedure, Fig. 4.2 presents a direct comparison between the calculated and measured sheet resistance frequency distributions after such an adjustment. Further, Fig. 4.3 presents the C_o frequency distribution implied by this calculation to exist within the particular diffusion process under consideration.

It has been shown that in the absence of oxidation, the impurity atom surface concentration after "drive-in" is directly related to this concentration after predeposition in [3]. Thus, C_o variabilities associated with predeposition will influence the sheet resistance produced by all subsequent processing steps. It is presumed that, for this reason, our calculations show the C_o variations, associated with impurity atom predeposition contribute significantly to the standard deviation of sheet resistance after oxidation.

Computational experiments also indicate that, in a two-step diffusion, oxide thickness has an important influence upon the final distribution of sheet resistance. Thus, with knowledge of the C_o distribution (see Fig. 4.3) this model can be used to infer the distribution of oxide thickness. Figure 4.4 shows the frequency distribution of measured sheet resistance from this diffusion process and, in addition, the calculated distribution. This particular calculation was accomplished by assuming in our model (Eq. 4.3) the C_o distribution inferred from previous calculations (Fig. 4.3) and also adjusting the W_{ox} distribution to attain satisfactory agreement with experiment. Thereby, from this final calculation we infer the frequency distribution of oxide thickness produced by this particular diffusion process, Fig. 4.5.

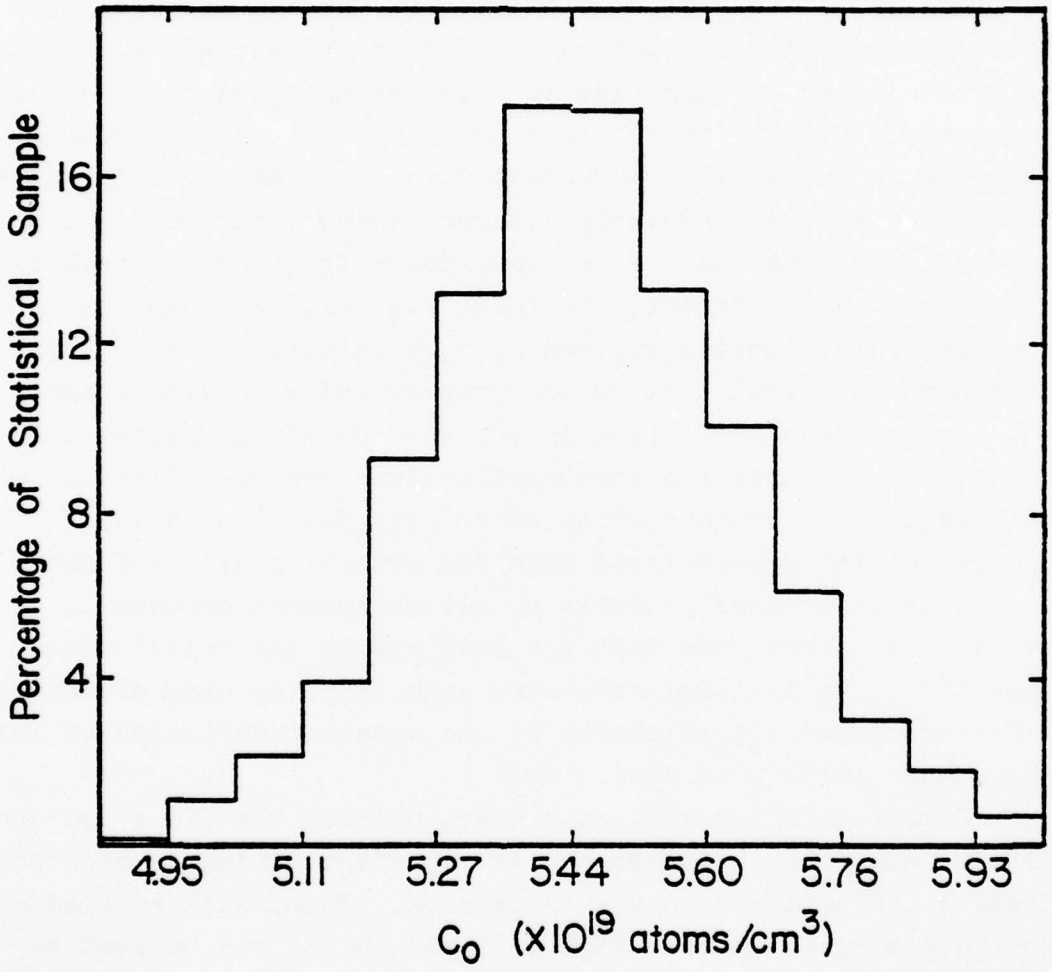


Fig. 4.3 C_0 frequency distribution.

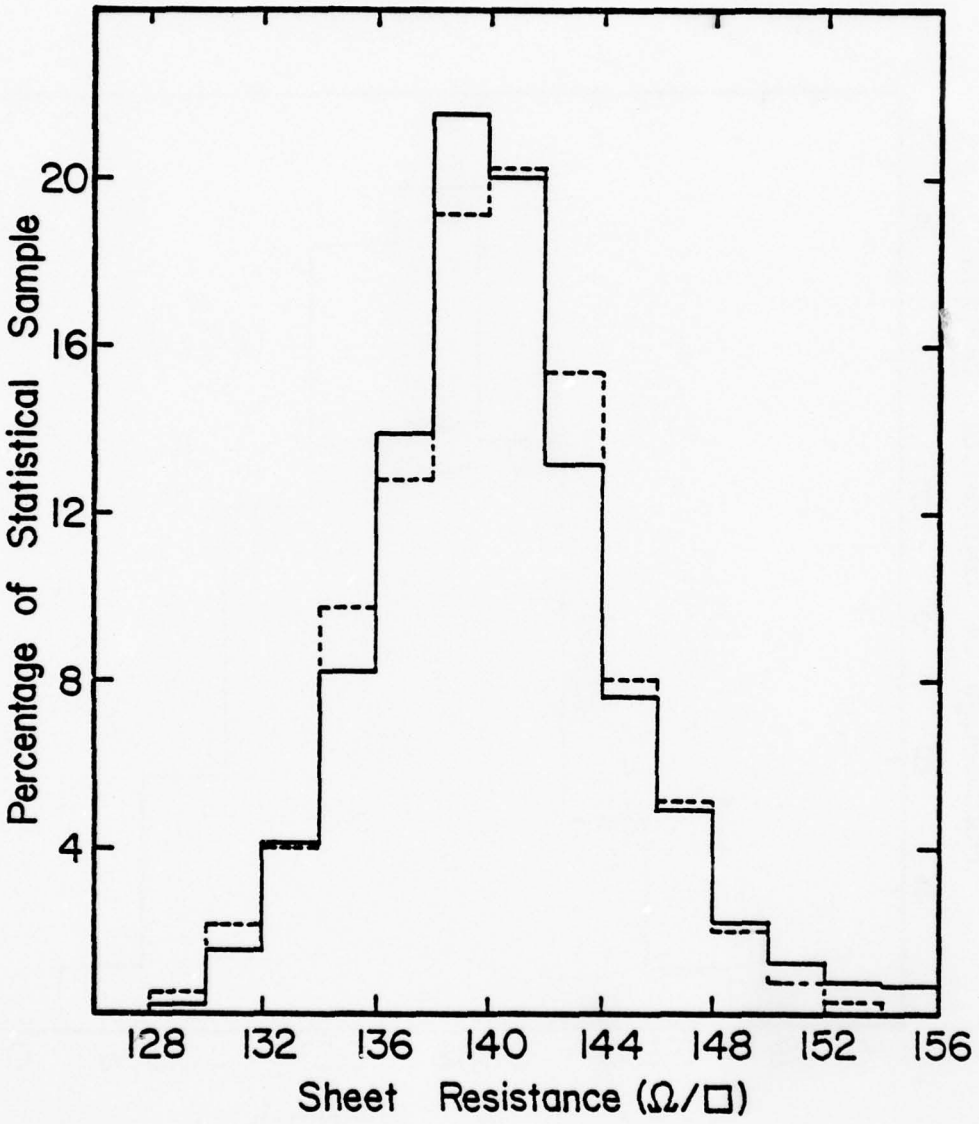


Fig. 4.4 Drive-in sheet resistance frequency distribution.

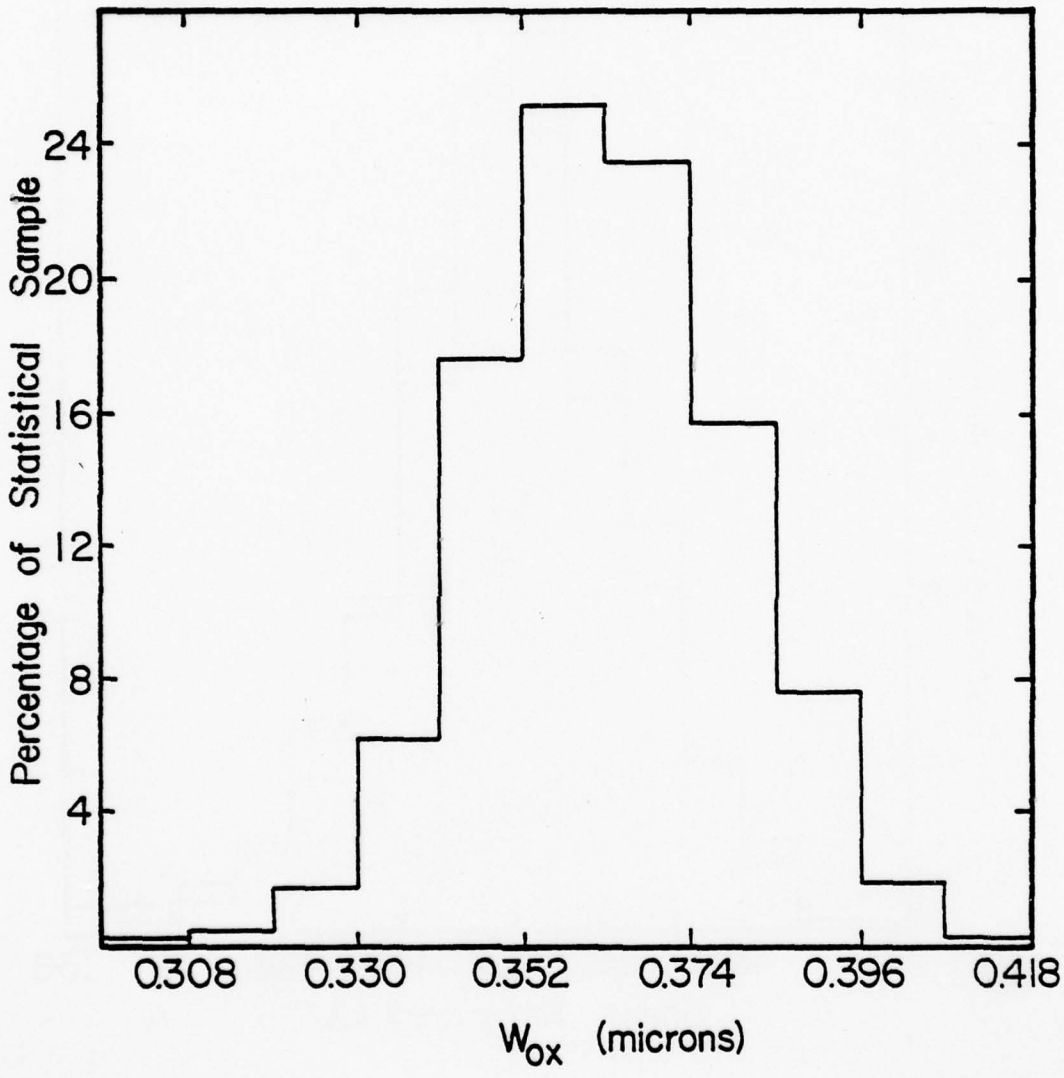


Fig. 4.5 W_{ox} frequency distribution.

5.0 Summary and Conclusions

A computer program has been developed simulating a two step diffusion of impurity atoms into silicon. Random process variations are introduced by a random number generator which has been tested and found to be statistically suitable. Results of a sensitivity analysis suggest that two parameters, C_o and W_{ox}, dominate the diffusion process. Finally, a simple method is used to infer the variability in C_o and W_{ox} from experimentally measured sheet resistance distributions.

Chapter V
Equivalent Circuit Studies

Professor F. A. Lindholm;
Graduate Student J. I. Arreola

		Page
I.	A Nonlinear Indefinite Admittance Matrix for Modeling Electronic Devices	186
	I.1 Indefinite Admittance Matrix	188
	I.2 Extension for Nonlinear Electronic Devices	189
	I.3 Conclusions	193
II.	Special Considerations for the MOSFET	195
	II.1 Transport Current	198
	II.2 Charging Currents	198
	II.3 Network Representation of the Intrinsic MOSFET	200
III.	Functional Dependence of Charge, Current, and Model Circuit Elements	200
	III.1 New Expressions for the Inversion and Depletion Charges	203
	III.2 Merit of the Expressions	208
	III.3 Drain Current	208
	III.4 Saturation	209
	III.5 Charge Apportionment	209
	III.6 Illustrative Circuit Example	210

Chapter V
Equivalent Circuit Studies

Professor F. A. Lindholm;
Graduate Student, J. I. Arreola

The transient computer simulation of large-signal MOSFET circuits can disagree sharply with observed behavior. Hence the equivalent-circuit models used to represent MOS transistors in network form deserve scrutiny for possible inadequacies. In the present effort, we are examining three defects of present-day models in the representation of the internal device physics, seeking to set down a basis for removing these defects.

We identify the three defects in present-day models as follows:

(a) inadequate representation of the MOSFET as a four-terminal device;

(b) inadequate inclusion of the effects of the two-dimensional (and three-dimensional) configuration of flux lines present, particularly in short-channel structures; and

(c) inadequacy of the quasi-static approximation, which underlies all models in common use for computer circuit simulation.

In relation to the defect labeled (c) above, we have developed a test for the self-consistent validity of the quasi-static approximation that is easily implemented in circuit-analysis programs. A description of this work appeared in Chapter III of the Interim Report for the period -- 1 July 1975 to 31 December 1975.

In relation to defects (a) and (b), we have developed a methodology yielding network representations that include elements to account for four-terminal behavior and for multi-dimensional effects during transients. Because of basic asymmetries in the device structure, "capacitive" currents associated with the accumulation of mobile carriers within the MOSFET require in the net-

work representation circuit elements in addition to capacitors. By applying the methodology to a particular physical model of the MOSFET, we have shown that appreciable errors can potentially result from use of an all-capacitor model. This work is described in this final report.

The presentation that follows begins in Section I with the methodology. The approach described in Section I yields equivalent-circuit models for the large-signal transient response of all electronic devices described by charge control. Its relation to the indefinite admittance matrix of circuit theory offers advantages in the modeling of devices having more than three terminals. In Sections II and III these advantages are demonstrated for the four-terminal MOSFET.

I. A NONLINEAR INDEFINITE ADMITTANCE MATRIX FOR MODELING ELECTRONIC DEVICES

This section describes a new approach for developing equivalent-circuit models of electronic devices. The models developed by this approach represent the large-signal, hence nonlinear, response to transient excitation. The approach applies to all devices whose operation is described by the principles of charge control [1-3], including therefore field-effect transistors of various kinds, bipolar transistors, and certain electron tubes.

The models yielded by the approach are compact, composed of few circuit elements. As a result of their compactness, the models are meant to be useful in the computer-aided analysis of electronic

---

[1] E. O. Johnson and A. Rose, "Simple General Analysis of Amplifier Devices with Emitter, Control, and Collector Functions," *Proc. IRE,* Vol. 47, pp. 407-418, March 1959.

[2] R. D. Middlebrook, "A Modern Approach to Semiconductor and Vacuum Device Theory," *IEE Proc.*, Vol. 106B, suppl. 17, pp. 887-902, March 1960.

[3] P. E. Gray, D. DeWitt, A. R. Boothroyd, and J. F. Gibbons, *Physical Electronics and Circuit Models of Transistors,* Wiley, New York, 1964.

circuits. This intended use contrasts with that intended for equivalent-circuit models [4] containing many circuit elements, which pertain chiefly to detailed studies of the physics underlying electronic device behavior.

The approach to be described applies independently of the number of device terminals. Indeed, the greater is that number, the more the power of the approach is disclosed. The approach applies also independently of multi-dimensional spatial dependence that may be present in the boundary-value problem describing the device. This generality is needed, for example, in modeling the MOS field-effect transistor (MOSFET), because the substrate terminal constitutes a fourth terminal through which sizable transient currents flow in some circuit applications, and because short-channel devices give rise to multi-dimensional effects.

Models of four-terminal devices [5,6] and models that include multi-dimensional effects [7] have been proposed earlier. But this previous work has not focussed on laying down systematic procedures for developing models, which is the aim of the present section.

Systematic procedures exist for modeling the linear response of multi-terminal circuits subjected to small-signal excitation. These procedures are linked to the indefinite admittance matrix (i.a.m.), which we first shall review and then exploit to model the nonlinear response of multi-terminal electronic devices to large-signal excitation.

---

[4] C. T. Sah, "Equivalent Circuit Models in Semiconductor Transport for Thermal, Optical, Auger-Impact, and Tunneling Recombination-Generation-Trapping Processes," *Phys. Stat. Sol. (a)*, Vol. 7, pp. 541-559, 1971.

[5] F. A. Lindholm and P. R. Gray, "Large-Signal and Small-Signal Models for Arbitrarily-Doped Four-Terminal Field-Effect Transistors," *IEEE Trans. Elec. Dev.*, Vol. ED-13, pp. 819-829, December 1966.

[6] R.S.C. Cobbold, *Theory and Applications of Field-Effect Transistors*, Chapter 8, Wiley-Interscience, New York, 1970.

[7] H. C. Poon, L. D. Yau, R. L. Johnson, and P. Beecham, "D.C. Model for Short-Channel IGFET's," *1973 IEDM Digest*, p. 156, 1973; L. D. Yau, "A Simple Theory to Predict the Threshold Voltage of Short Channel IGFET's," *Solid-State Electron.*, Vol. 17, p. 1059, 1974; Y. A. El-Mansy and A. R. Boothroyd, "A Simple Two-Dimensional Saturation Model for Short Channel IGFETs for CAD Applications," *1974 IEDM Digest*, p. 35, 1974.

I.1 Indefinite Admittance Matrix

Consider a lumped electrical network which has n terminals. Let an additional external node be the common reference. From the standpoint of its behavior at the terminals, the network, if linear, may be described by a set of equations as follows:

$$\underline{I} = \underline{y}\,\underline{V} \quad . \tag{5.1}$$

The required linearity is assured for any network operating under small-signal conditions. The matrix elements of $\underline{y}$ are

$$y_{jk} = \frac{I_j}{V_k}\Big|_{V_i = 0,\ i \neq j} \tag{5.2}$$

where $\underline{I}$ and $\underline{V}$ correspond to the current and voltage at the terminals.

The matrix $\underline{y}$ defined in (6.1) and (6.2) is called the indefinite admittance matrix [8,9], and its elements satisfy the following property imposed by Kirchhoff's laws:

$$\sum_j y_{jk} = \sum_k y_{jk} = 0 \quad , \tag{5.3}$$

that is, the elements in any row or any column sum to zero.

As will be seen, our development of large-signal models for electronic devices will make use of two special cases of the i.a.m. In the first case, the matrix $\underline{y}$ is symmetric and has one of the following forms:

$$\underline{y} = \underline{a};\ \underline{y} = \underline{b}\,\frac{d}{dt};\ \underline{y} = \underline{c}\int dt \quad . \tag{5.4}$$

Here $\underline{a}$, $\underline{b}$, and $\underline{c}$ are real symmetric matrices, and each matrix element corresponds to a single lumped resistor or capacitor or inductor connected between each pair of the n terminals. In the second case, the matrix $\underline{y}$ is nonsymmetric but is the sum of two indefinite admittance matrices: a symmetric matrix, like (5.4), and a residual nonsymmetric matrix, each element of which corresponds to a controlled current source placed between each pair of

---

[8] J. Shekel, "Matrix Analysis of Multi-Terminal Transducers," *Proc. IRE*, Vol. 42, pp. 840-847, May 1954.

[9] L. A. Zadeh, "Multipole Analysis of Active Networks," *IRE Trans. on Circuit Theory*, Vol. CT-4, pp. 97-105, September 1957.

terminals. In this second case, then, the circuit representation of the i.a.m. results from connecting the network corresponding to the symmetric matrix in parallel with that corresponding to the nonsymmetric matrix. In general, summing of indefinite admittance matrices corresponds to connecting their circuit representations in parallel.

I.2 Extension for Nonlinear Electronic Devices

Consider an electronic device having n terminals. The modeling begins by specifying the physical mechanisms relevant to the operation of the device. For many devices, only three such mechanisms, at most, are relevant: the transport of charged carriers between terminals; the net recombination of charged carriers within the device; the accumulation of these carriers within the device. Thus, the current i_J flowing at any terminal J is the sum of three components: a transport current $(i_J)_T$, a recombination current $(i_J)_R$, and a charging current $(i_J)_C$. That is,

$$i_J = (i_J)_T + (i_J)_R + (i_J)_C \quad . \tag{5.5}$$

We now characterize these components.

The transport mechanism consists of the injection of a charged carrier in one terminal, followed by its transport across the device until it reaches any of the other terminals, where it recombines at a surface with a carrier of opposite charge. The recombination mechanism differs from the transport mechanism only in that the carriers recombine within the bulk of the device instead of at the terminals. Therefore, both mechanisms can be characterized by the same form:

$$(i_J)_{T,R} = \sum_{K \neq J} (i_{JK})_{T,R} \quad . \tag{5.6}$$

Here i_{JK} represents the current due to the charged carriers injected from terminal J, which recombine, at a surface or in the bulk, with opposite-charged carriers injected from terminal K. From this characterization it follows that $(i_{JK})_{T,R}$ satisfies the following properties:

$$i_{JK} = -i_{KJ}; \quad i_{JJ} = 0 \tag{5.7}$$

These properties allow transport and net recombination to be represented by controlled current sources connected between pairs of terminals. The value of the current source between terminals J and K is i_{JK}.

The last mechanism to be considered is the accumulation of mobile carriers within the device, which requires the charging current,

$$(i_J)_C = \frac{dq_J}{dt} \quad . \tag{5.8}$$

As Fig. 5.1 illustrates, dq_J is the part of the total charge accumulated within the device in time dt that is supplied from terminal J. The charge accumulation expressed in (5.8) is a mechanism basic to any electronic device that operates by charge control [1-3].

Now, using (5.6) and (5.8), we may rewrite (5.5) as

$$i_J = \sum_{K \neq J} (i_{JK})_{T,R} + \frac{dq_J}{dt} \quad . \tag{5.9}$$

Although (5.9) is valid, it does not correspond to a convenient network. To get a convenient network representation, we apply one additional constraint which costs small loss in generality in that it holds for all charge-control devices [1-3]. We apply the constraint that the overall device under study is charge neutral. Or, more exactly and less demanding, we assume the device accumulates no net overall charge as time passes. This constraint of overall charge neutrality requires a communication of the flux lines amongst the terminals to occur that maintains charge neutrality by coulomb forces and by drift and diffusion currents. The requisite <u>overall</u> neutrality may result either from neutrality occurring at each macroscopic point, as in a transistor base, or from a balancing of charges that are separated, as on the gate and in the channel of a MOSFET.

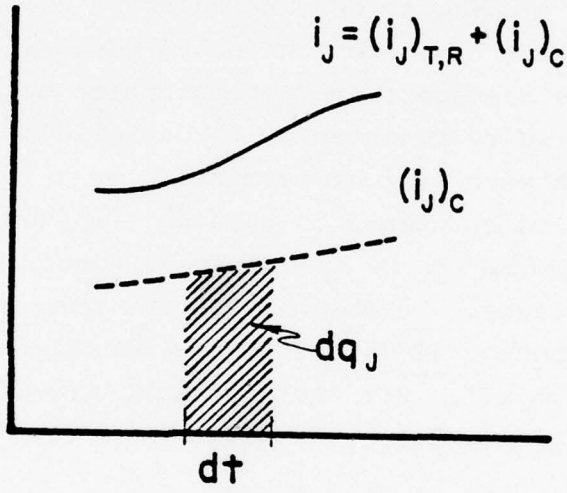


Figure 5.1. The charging current $(i_J)_C$ at terminal J produces the accumulated charge dq_J.

As a result of overall neutrality, the current at any terminal J becomes the sum of the currents flowing out of all of the other terminals:

$$i_J = - \sum_{K \neq J} i_K \quad . \tag{5.10}$$

This global counterpart of the Kirchhoff current-node law implies for the charging currents of (5.8) that

$$(i_J)_C = - \sum_{K \neq J} (i_K)_C \quad , \tag{5.11}$$

which means that a charging current entering one terminal flows, in its entirity, out of all of the other terminals. Hence, as is true also for the transport and recombination mechanisms, charge accumulation can be represented by a controlled current source connected between each pair of terminals.

For a model to be useful in circuit analysis requires that the elements of the model all be specified as functions of the terminal currents and voltages. To do this, we now make use of the principles of charge control [1-3] and of the closely allied quasi-static approximation [5,6,10]. For the transport and recombination mechanisms, charge control gives directly

$$(i_{JK})_{T,R} = q_{JK}/t_{JK} \quad . \tag{5.12}$$

Here q_{JK} is the charge of the carriers that contribute to the current flowing between terminals J and K. The recombination time t_{JK} is the time constant associated with that current: a transit time if the mechanism being described is transport, a lifetime if it is recombination. Then, to produce the desired functional dependence, a quasi-static approximation [5,6,10] is used that specifies each $(i_{JK})_{T,R}$ as a function of the instantaneous voltages at the device terminals.

---

[10] F. A. Lindholm, "Unified Modeling of Field-Effect Devices," *IEEE Journal of Solid-State Circuits*, Vol. SC-6, August 1971.

This characterization of $(i_{JK})_{T,R}$, combined with the properties expressed in (5.7) can be manipulated to describe transport and recombination by a matrix like $\underline{a}$ in (5.4). The matrix is symmetric, and, as one can verify, it satisfies the key properties, given in (5.3), that belong to the indefinite admittance matrix.

Similar procedures are applied to model charge accumulation. To the charging current defined in (5.8) a quasi-static approximation is applied [5,6,10], specifying the functional dependence of q_J on the terminal voltages and enabling thereby the employment of the chain rule of differentiation. The resulting characterization of $(i_J)_C$ describes charge accumulation by a matrix that has the form of $\underline{b}$ in (5.4), a matrix whose elements are

$$b_{JK} = \frac{\partial q_J}{\partial v_K}\Big|_{dv_I = 0,\ I \neq K} \quad . \tag{5.13}$$

The matrix $\underline{b}$ also satisfies the key properties of the indefinite admittance matrix that are given in (5.3). For a general n-terminal electronic device, this matrix describing charge accumulation is nonsymmetric, and is therefore the sum of a symmetric and a residual nonsymmetric part. The symmetric part corresponds to an all-capacitor network; the network representation of the residual nonsymmetric matrix consists of controlled current sources.

I.3 Conclusions

From the properties of the i.a.m. it follows that the three-branch circuit of Fig. 5.2 serves as a building block for model generation. Connecting a circuit of this form between each terminal pair yields the general network representation for an n-terminal electronic device. For any particular device of interest, certain of the circuit elements may vanish. In a MOSFET, for instance, no transport or recombination currents flow to the gate, and the corresponding circuit elements will be absent.

Any equivalent-circuit model generated by this approach can be regarded in two ways: either as a product of the building block

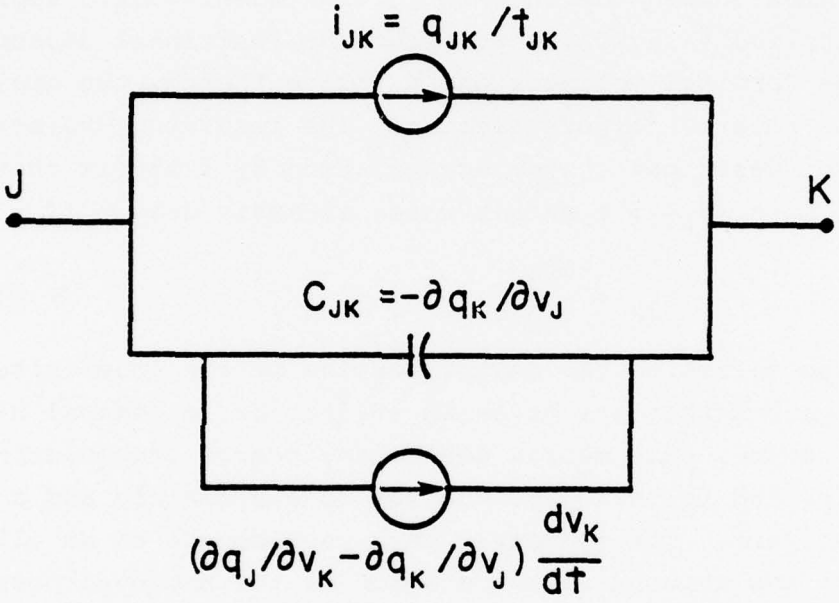


Figure 5.2. General equivalent circuit between each pair of terminals at an n-terminal electronic device.

of Fig. 5.2 or as a circuit described by a matrix which obeys the key properties of the i.a.m. Description by the i.a.m. treats all terminals equally in that none is singled out as the reference node; the advantages of this show up plainly in the modeling of a four-terminal device, such as the MOSFET [11].

From Fig. 5.2 notice that the mobile charge accumulation within a general n-terminal electronic device is not represented by the flow of displacement currents in an all-capacitor model. The residual nonsymmetric matrix, and the corresponding transcapacitance current source of Fig. 5.2 provides the needed correction. This correction has practical engineering consequences in certain MOSFET circuits [11], though a discussion of that lies beyond the intent of the present section.

To use the approach set forth here in modeling any particular device requires that the static dependence on the terminal voltages be specified for the currents and charges defined in (5.12) and (5.13). This requires that a physical model for the device be chosen to describe the dc steady state. For the MOSFET this has been done. The corresponding equivalent-circuit model is derived in the following sections.

II. Special Considerations for the MOSFET

The previous section gives a lumped network representation of all electronic devices obeying the principles of charge control whose large-signal transient behavior depends on three physical mechanisms: mobile-charge transport, net recombination within the device, and mobile-charge accumulation. The result is the equivalent circuit of Fig. 5.2, which applies between any pair of terminals and is the basic building block from which an equivalent circuit is constructed for the overall n-terminal device. The currents representing transport and net recombination flow in the current

---

[11] W. Shockley, "A Unipolar 'Field-Effect'transistor," *Proc. IRE*, Vol. 40, pp. 1365-1376, November 1952.

source i_{JK}. The charging current representing mobile-charge accumulation flow through the capacitor

$$C_{JK} = - \frac{\partial q_K}{\partial v_J} \quad , \tag{5.14}$$

and through the controlled current source characterized by

$$d_{JK} = \frac{\partial q_J}{\partial v_K} - \frac{\partial q_K}{\partial v_J} \quad . \tag{5.15}$$

To develop an equivalent-circuit model for any charge-control device, therefore, one needs only to describe the circuit elements of (5.14) and (5.15) by compact analytical expressions consistent with the physics underlying the device behavior. In Section III we do this for a particular model of the four-terminal MOSFET. To provide the background for this modeling, however, it is instructive to consider first in general terms the physical mechanisms underlying the transport and the charging currents in a MOSFET.

Consider the <u>intrinsic</u> portion of the enhancement-mode MOSFET shown in Figure 5.3. Let terminals 1 and 2 represent the intrinsic source and drain contacts. Let terminals 3 and 4 represent the contacts of the gate and the substrate. If the gate and substrate voltages are controlled to induce a conduction channel between the source and drain regions, it is common to adopt the following approximation: neglect net recombination/generation in both the inversion and depletion regions, and neglect leakage currents in the insulator. These approximations, together with others in common use that will be specified in Section III, will lead to the functional dependencies of the circuit elements in the model.

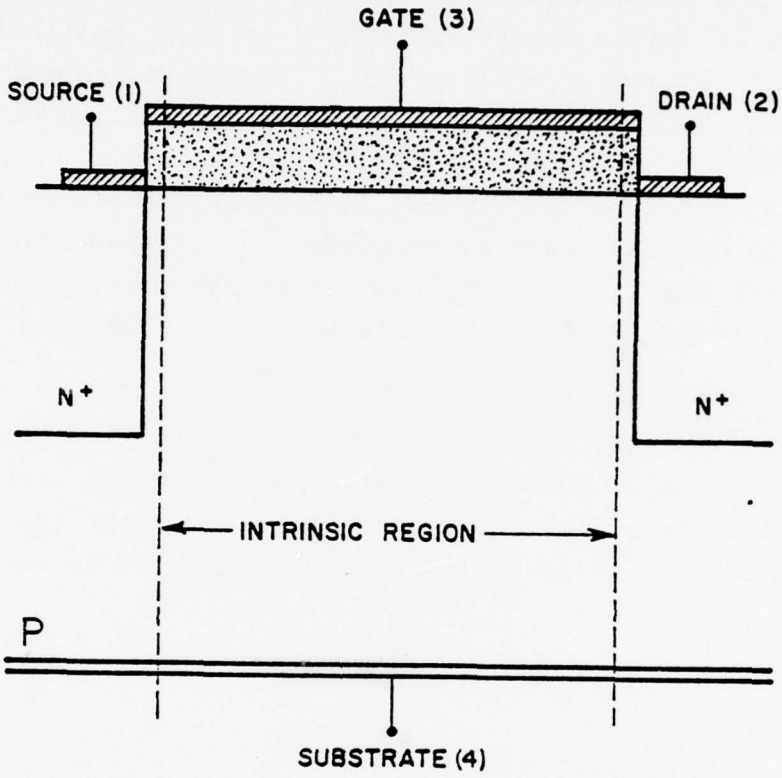


Figure 5.3. The MOSFET divided into intrinsic and extrinsic parts.

II.1 Transport Current

Because net recombination within the device is neglected and the transport of mobile charge is assumed to occur only between the source and drain terminals, the only transport element existing in the network representation is i_{12}. This current source represents the total current flowing in the conduction channel. The main approximation needed to specify its functional dependence on the device make-up and the terminal voltages is a quasi-static approximation [5,6,10] that is commonly used in charge control analysis. The functional dependence of i_{12} will be given in Section III.

II.2 Charging Currents

Consider first the charging currents flowing into the gate and substrate. Because we assume no transport mechanisms influence these currents, their only function is charging, that is, they cause the accumulation of mobile charge within the device in accordance with,

$$i_3 = (i_3)_c = i_G = \frac{dq_3}{dt} = \frac{dq_G}{dt} \qquad (5.16)$$

$$i_4 = (i_4)_c = i_B = \frac{dq_4}{dt} = \frac{dq_B}{dt} \quad . \qquad (5.17)$$

Thus dq_3 is the incremental change in the total charge accumulated at the metal of the gate, and dq_4 is the incremental change in the total charge accumulated in the substrate region.

Now consider the induced conduction channel, which, for concreteness, we assume is an electron channel. Here, a charge q_N is induced by the charge accumulated in the gate and substrate regions. Overall charge neutrality requires: $q_N = -(q_3 + q_4)$. This induced charge has to be supplied by the charging currents $(i_1)_c$ and $(i_2)_c$ flowing at the source and drain terminals. Hence,

$$(i_1)_c + (i_2)_c = \frac{dq_1}{dt} + \frac{dq_2}{dt} = \frac{dq_N}{dt} \quad . \qquad (5.18)$$

Here the charge dq_1 is the part of the total induced charge in the conduction that flows in at the source; dq_2 is the part that flows in at the drain.

For the network representation, we require to know what part of dq_N is supplied by dq_1 and what part by dq_2. In other words, we need to account for the apportionment of the charge between the source and drain terminals. This important requirement has apparently been overlooked by previous workers.

The correct apportionment of the charge for the MOSFET channel must take account that the apportionment depends on the terminal voltages. Consider two limiting cases. First consider the device operating with an inversion channel but with no current flow ($v_{DS}=0$). Then the distribution at the inversion charge is entirely symmetric; it is asymmetric when current flows because of the potential drop in the channel. For $v_{DS}=0$, the resulting symmetric distribution requires that both charging currents contribute equally to dq_N. As a second limiting case, consider the device operating in saturation [$v_{DS}>V_{DSS}$ as defined in Eq. (5.40)]. Then practically no displacement current flows between the channel termination point and the drain, and thus dq_N is mainly supplied by dq_1, the time integral of the charging current flowing at the source.

From the two limiting conditions just considered, the drain and source charging currents can be expressed, as a first approximation, in the form:

$$(i_1)_c = \frac{dq_1}{dt} = \lambda \frac{dq_N}{dt} \qquad (5.19)$$

$$(i_2)_c = \frac{dq_2}{dt} = (1-\lambda)\ \frac{dq_N}{dt} \qquad (5.20)$$

Here λ is a function that has the following properties:

$$\lambda = \begin{cases} \frac{1}{2} \text{ if } v_{DS} = 0 \\ 1 \text{ in saturation } [V_{DS}>V_{DSS}; \\ \quad \text{see Eq. (5.40)}] \end{cases} \qquad (5.21)$$

A suitable approximation for the voltage dependence of λ for intermediate values of v_{DS} is considered in Section III. Previous works [5,6] have assumed implicitly that $dq_2 = 0$, which is unjustified for non-saturated operation (see Eq. 5.20).

II.3 Network Representation of the Intrinsic MCSFET

In this section we have described two important considerations that underlie the modeling of the four-terminal MOSFET:

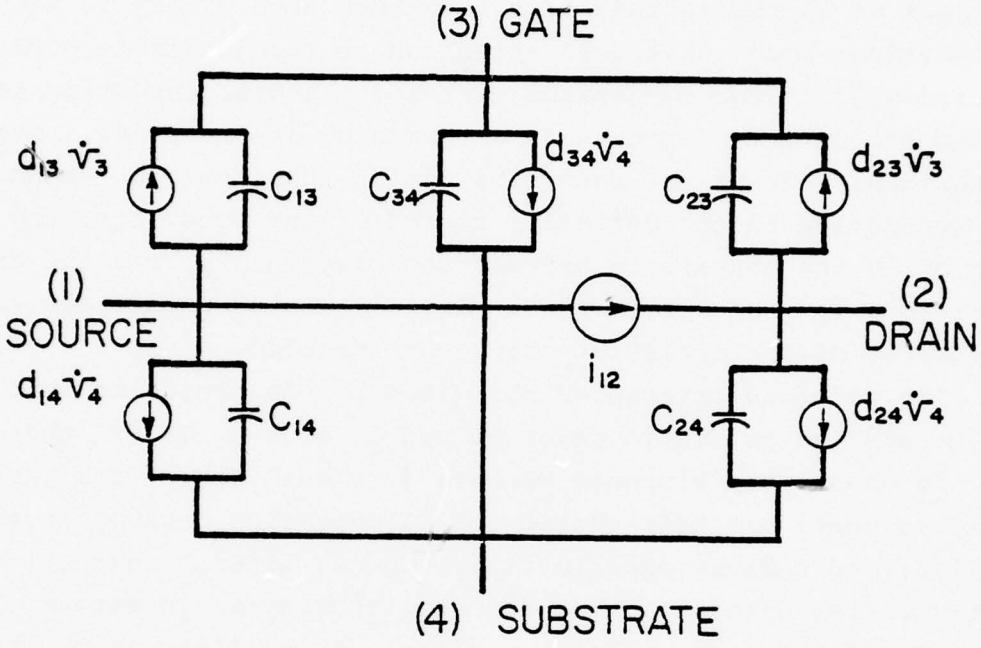
(a) that transport current occurs only between the drain and the source (terminals 1 and 2); and

(b) that the charging of the inversion channel with mobile carriers for non-saturated operation is apportioned between the source and the drain - that is, part comes by an influx of electrons from the source and part by an influx from the drain.

These considerations, joined with the building block of Fig. 5.4, lead directly to the form of the equivalent-circuit model for the intrinsic MOSFET, which is shown in Fig. 5.4.

III. FUNCTIONAL DEPENDENCE OF CHARGE, CURRENT, AND MODEL CIRCUIT ELEMENTS

The modeling procedures and considerations of Section I apply, in principle, to many kinds of electronic devices. Their application in Section II to the intrinsic four-terminal MOSFET lead to the form of the equivalent circuit model shown in Fig. 5.4, for the transient response of the MOSFET. It remains now to determine the functional dependencies of each capacitor, transcapacitor, and controlled source in that model. Upon the complexity of these functional dependencies depends whether the circuit model will be simple enough for use in circuit-analyses programs. The purpose of this section is to develop these functional dependencies for a specific physical model of static MOSFET behavior.



$$C_{13} = -\frac{\partial q_G}{\partial v_1} \qquad C_{23} = -\frac{\partial q_G}{\partial v_2} \qquad C_{34} = -\frac{\partial q_B}{\partial v_3}$$

$$C_{14} = -\frac{\partial q_G}{\partial v_1} \qquad C_{24} = -\frac{\partial q_G}{\partial v_2}$$

$$d_{13} = -d_{14} = \lambda \frac{\partial q_N}{\partial v_3} - \frac{\partial q_G}{\partial v_1}$$

$$d_{23} = -d_{24} = (1-\lambda) \frac{\partial q_N}{\partial v_3} - \frac{\partial q_G}{\partial v_2}$$

$$d_{34} = d_{13} + d_{23} = \frac{\partial q_G}{\partial v_4} - \frac{\partial q_B}{\partial v_3}$$

$$\lambda = f(V_{DS})$$

Figure 5.4. Equivalent-circuit model of the intrinsic MOSFET.

The physical model we choose to illustrate the modeling procedures and to examine the suitability of the resulting circuit model is based on approximations that have been used widely in the past. The physical model restricts attention to the intrinsic MOSFET. It assumes that two mechanisms, transport and accumulation of charge within the device, dominate in determining its behavior. Other mechanisms, such as net recombination in the inversion channel and net generation in the depletion region of the substrate, are neglected. Current in the channel is assumed to arise mainly from the drift of carriers. The Shockley gradual case [11] is assumed. The voltage dependence of the depletion charge is included.

The modeling principles described in the preceding sections can be applied to static physical models of the MOSFET other than the one whose main elements we have just outlined. This particular physical model has been chosen for illustration because it is widely familiar and because experimental evidence attests that it provides a first-order description of the static behavior in strong inversion of all but short-channel devices. By application of the modeling procedures developed here to this physical model, the static characterization will be converted into an equivalent-circuit description of the transient response.

Previous equivalent circuits for the MOSFET transient response suffer two major shortcomings. First, they failed to account for the non-reciprocity in circuit representation required in properly characterizing the charging currents that flow during transients in a four-terminal MOSFET. This non-reciprocity manifests itself in the circuit of Fig. 5.4 by the presence of the transcapacitors d_{JK}. Second, the complexity of the most complete of the previous circuit models for the four-terminal behavior [5,6] diminishes its utility in computer-aided circuit analysis. The complexity results from the characterizations used for the inversion and depletion charges and the channel current under static (dc) conditions. As we have noted before, the development of suitable descriptions for the static charges and current is a key to the modeling of the

transient response. A main element underlying all compact equivalent circuits for the transient response is the quasi-static approximation, in which the dynamic response is viewed as a succession of steady-state responses.

Thus, in what follows, we consider first the development of a better static characterization than has previously been available.

III.1 New Expressions for the Inversion and Depletion Charges

The main elements of the physical model we will use have been noted above. It is a four-terminal model that includes the voltage dependence of the depletion charge Q_B in the bulk. This dependence was first included by Sah and Pao [12] and then extended by Cobbold [13] to the four-terminal MOSFET.

In Cobbold's derivation, for strong inversion in the channel, the principal charge components per unit area in the intrinsic portion are written in terms of the physical make-up of the device and the voltages at its terminals as follows: for the gate charge Q_g (coulombs/cm^2),

$$Q_g = C_o(V_{GS} - \phi_{MS} - 2\phi_F - V(y)) \quad ; \qquad (5.22)$$

for the substrate charge Q_b (coulombs/cm^2),

$$Q_b = -kC_o(V(y) - V_{BS} + 2\phi_F)^{1/2} \quad . \qquad (5.23)$$

The expression for the induced charge in the channel follows from the condition of overall charge neutrality joined with the one-dimensional approximation imbedded in the Shockley gradual case [11]:

$$Q_n = -(Q_g + Q_b + Q_{ss}) \quad . \qquad (5.24)$$

In these expressions C_o is the oxide capacitance per unit area, ϕ_{MS} is the metal to semiconductor work function difference, Q_{ss} is the oxide charge expressed as an effective interface charge,

---

[12] C. T. Sah and H. C. Pao, "The Effects of Fixed Bulk Charge on the Characteristics of Metal-Oxide Semiconductor Transistors," *IEEE Trans. Electron Devices*, Vol. ED-13, pp. 393-409, April 1966.

[13] R.S.C. Cobbold, "M.O.S. Transistor as a 4-Terminal Device," *IEE Electronic Letters*, Vol. 2, No. 6, June 1966.

and $k = (2qN_A \kappa_s \varepsilon_o / C_o^2)^{1/2}$ is a constant that depends in the ionized doping concentration N_A, the dielectric permittivities of the oxide and the semiconductor, and the oxide thickness. The potential V(y) is the voltage at the surface of the channel measured at a distance y from the intrinsic source terminal.

The handling of the expressions in the derivation that follows is simplified by defining the following potentials:

$$U_{GS} = V_{GS} - \phi_{MS} - 2\phi_F + \frac{Q_{ss}}{C_o} \qquad (5.25)$$

$$U_{BS} = V_{BS} - 2\phi_F \quad . \qquad (5.26)$$

All quantities appearing in the right-hand sides of these definitions, except for V_{GS} and V_{BS}, are assumed to be constant independent of the voltages.

The use of the previous definitions simplifies the expressions for the charges:

$$\frac{Q_g}{C_o} + \frac{Q_{ss}}{C_o} = U_{GS} - V(y)$$

$$\frac{Q_b}{C_o} = -k(V(y) - U_{BS})^{1/2} \qquad (5.27)$$

$$\frac{Q_n}{C_o} = -\left(\frac{Q_g}{C_o} + \frac{Q_{ss}}{C_o} + \frac{Q_b}{C_o}\right)$$

The total charges in the channel and the bulk depletion regions are found by integration. As is done by Sah and Pao [12] and by Cobbold [6], the integration is accomplished by transforming the variable of integration into y the potential V(y). This involves using the relation between the drain current I_D and the channel charge Q_n. Since only drift flow is assumed, this relation can be written as

$$\frac{I_D}{L} = -K \frac{Q_n}{C_o} \frac{dV}{dy} \qquad (5.28)$$

with

$$K = \frac{\mu Z C_o}{L} \quad ,$$

where μ is an effective (average) mobility and L is the effective channel length. Both μ and L depend on the applied bias conditions. Various dependencies have been proposed by several authors [14], however, for the purposes of this derivation we will assume μ and L to be constant.

The procedure just indicated yields for the total charges:

$$\frac{Q_N}{C_o^*} = -\frac{K}{I_D}\int_o^{V_{DS}} \left(\frac{Q_n}{C_o}\right)^2 dV$$

$$\frac{Q_G}{C_o^*} = -\frac{K}{I_D}\int_o^{V_{DS}} \frac{Q_n}{C_o}\,\frac{Q_g}{C_o}\, dV \qquad (5.29)$$

$$\frac{Q_B}{C_o^*} = -\frac{K}{I_D}\int_o^{V_{DS}} \frac{Q_n}{C_o}\,\frac{Q_b}{C_o}\, dV$$

$$\frac{Q_{SS}}{C_o^*} = \frac{Q_{ss}}{C_o}$$

Here $C_o^* = LZC_o$ is the total oxide capacitance.

The condition of overall charge neutrality allows rewriting (5.29) as

$$\frac{Q_N}{C_o^*} = -\frac{K}{I_D}(A + B + 2D)$$

$$\frac{Q_G}{C_o^*} + \frac{Q_{SS}}{C_o^*} = \frac{K}{I_D}(A + D) \qquad (5.30)$$

$$\frac{Q_B}{C_o^*} = \frac{K}{I_D}(B + D)$$

---

[14] R. H. Crawford, *MOSFET in Circuit Design*, McGraw-Hill, New York, 1967; W. E. Armstrong, "Effective Surface Mobility Theory, M.S. Thesis, Arizona State Univ., Tempe, June 1971.

where we are using the following definitions:

$$A = \int_o^{V_{DS}} \left(\frac{Q_g}{C_o} + \frac{Q_{ss}}{C_o} \right)^2 dV$$

$$B = \int_o^{V_{DS}} \left(\frac{Q_b}{C_o} \right)^2 dV \qquad (5.31)$$

$$D = \int_o^{V_{DS}} \left(\frac{Q_g}{C_o} + \frac{Q_{ss}}{C_o} \right) \frac{Q_b}{C_o} \, dV$$

Substituting expressions (5.27) into (5.31) and evaluating the integrals, we obtain:

$$A = \frac{1}{3} \left[U_{GS}^3 - (U_{GS} - V_{DS})^3 \right]$$

$$B = k^2 \frac{V_{DS}}{2} (V_{DS} - 2U_{BS}) \qquad (5.32)$$

$$D = -\frac{2}{15} k \left[(V_{DS}-U_{BS})^{3/2} (5U_{GS} - 2U_{BS} - 3V_{DS}) \right.$$

$$\left. - (-U_{BS})^{3/2} (5U_{GS} - 2U_{BS}) \right]$$

Expressions for the partial derivatives of the total charge components are evaluated by applying chain-rule differentiation to (5.30). This yields,

$$\frac{1}{C_o^*} \frac{\partial Q_N}{\partial V_J} = -\frac{K}{I_D}(A_J+B_J+2D_J) - \frac{1}{I_D} \frac{Q_N}{C_o^*} \frac{\partial I_D}{\partial V_J}$$

$$\frac{1}{C_o^*} \frac{\partial Q_G}{\partial V_J} = \frac{K}{I_D}(A_J+D_J) - \frac{1}{I_D} \left(\frac{Q_G}{C_o^*} + \frac{Q_{SS}}{C_o^*} \right) \frac{\partial I_D}{\partial V_J} \qquad (5.33)$$

$$\frac{1}{C_o^*} \frac{\partial Q_B}{\partial V_J} = \frac{K}{I_D}(B_J+D_J) - \frac{1}{I_D} \frac{Q_B}{C_o^*} \frac{\partial I_D}{\partial V_J}$$

where

$$A_J = \frac{\partial A}{\partial V_J} \; ; \; B_J = \frac{\partial B}{\partial V_J} \; ; \; D_J = \frac{\partial D}{\partial V_J} \qquad (5.34)$$

In writing (5.33) and (5.34), we have used a condensed notation in which V_J with J = S,D,G,B represents successively the source, drain, gate, and substrate voltages. The partial derivatives in (5.33) give expressions pertaining to the capacitors and transcapacitors of the circuit model of Fig. 5.4.

Substituting (5.32) into (5.33), and evaluating the partial derivatives we obtain

$$\begin{aligned}
A_S &= -U_{GS}^2 \\
A_D &= (U_{GS} - V_{DS})^2 \\
A_G &= V_{DS}(2U_{GS} - V_{DS}) \\
A_B &= 0 \\
(A_S+A_D+A_G+A_B &= 0) \qquad (5.35) \\
B_S &= k^2 U_{BS} \\
B_D &= k^2 (V_{DS} - U_{BS}) \\
B_G &= 0 \\
B_B &= -k^2 V_{DS} \\
(B_S+B_D+B_G+B_B &= 0) \\
D_S &= k\, U_{GS}(-U_{BS})^{1/2} \\
D_D &= -k[(V_{DS}-U_{BS})^{1/2}(U_{GS}-V_{DS})] \\
D_G &= -\tfrac{2}{3} k\, [(V_{DS}-U_{BS})^{3/2} - (-U_{BS})^{3/2}] \\
D_B &= \tfrac{1}{3} k\, [(V_{DS}-U_{BS})^{1/2}(3U_{GS}-2U_{BS}-V_{DS}) \\
&\qquad -(-U_{BS})^{1/2}(3U_{GS}-2U_{BS})] \\
(D_S+D_D+D_G+D_B &= 0)
\end{aligned}$$

III.2 Merit of the Expressions

The quantities A, B, D and A_J, B_J, D_J, are continuously repeated in the expression. Hence the equations for the twelve storage elements indicated in (5.33) are reduced to simple multiplications and additions of these quantities. This is convenient for use in computer-aided circuit analysis. The result is a circuit representation of the charging current during transients which is both simpler and more accurate than that of representations proposed previously.

III.3 Drain Current

To complete the model we need expressions for the current flowing in the inversion channel. The drain current is found by integrating both sides in equation (5.28) and the result is

$$I_D = -K \int_0^{V_{DS}} \frac{Q_n}{C_o} \, dV \quad . \tag{5.36}$$

Evaluation of the integral yields

$$I_D = K \left\{ U_{GS}V_{DS} - \frac{V_{DS}^2}{2} - \frac{2}{3} K[(V_{DS}-U_{BS})^{3/2}-(-U_{BS})^{3/2}] \right\} \tag{5.37}$$

By the quasi-static approximation, this expression gives the controlled source i_{12} of the model of Fig. 5.4. In terms of the quantities defined in (5.35), Eq. (5.37) can be simplified as follows:

$$I_D = K \left(\frac{A_G}{2} + C_G\right) \quad . \tag{5.38}$$

For use in (5.33), we note that the evaluation of the partial derivatives of the drain current is straightforward, and the result is

$$\frac{\partial I_D}{\partial V_S} = -K[U_{GS} - k(-U_{BS})^{1/2}]$$

$$\frac{\partial I_D}{\partial V_D} = K[U_{GS} - V_{DS} - k(V_{DS} - U_{BS})^{1/2}] \qquad (5.39)$$

$$\frac{\partial I_D}{\partial V_G} = KV_{DS}$$

$$\frac{\partial I_D}{\partial V_B} = K\,k[(V_{DS}-U_{BS})^{1/2} - (-U_{BS})^{1/2}]$$

These partial derivatives also have meaning in a circuit model for small-signal excitation.

III.4 Saturation

All the expressions just derived apply under non-saturation conditions. To extend these expressions so that they hold for saturated conditions requires only the substituting for V_{DS} of its value V_{DSS} at saturation. This is the voltage at which I_D reaches a maximum:

$$V_{DSS} = U_{GS} + \frac{k^2}{2}\left\{1 \pm [1 + \frac{4}{k^2}(U_{GS}-U_{BS})]^{1/2}\right\} \qquad (5.40)$$

Since V_{DS} is no longer an independent variable, note that all the partial derivatives with respect to V_{DS} are equal to zero.

III.5 Charge Apportionment

As was discussed in Section II, the evaluation of the equivalent-circuit elements indicated in Fig. 5.4 requires not only knowledge of the current, the charges and the charge derivatives. It requires also knowledge of how the charging currents are apportioned among the four terminals to change the charges in the inversion and bulk-depletion regions. The function λ, a charge-apportionment function, characterizes this apportioning. To complete the modeling, a suitable approximation must be found for the functional dependence of λ on the voltages at the terminals. Here we

use an approximation which, although first-order, offers an improvement over the approximation implicitly used in the past. We will assume that λ is a linearly increasing function of V_{DS} in the non-saturated region. When $V_{DS} = 0$, $\lambda = 1/2$, which represents equal charging of the inversion charge Q_N by the source and the drain. In saturation, $\lambda=1$, which represents that all charging of Q_N comes from current at the drain. Thus:

$$\lambda = \begin{cases} \frac{1}{2} \text{ when } V_{DS} = 0 \\ \frac{1}{2} \frac{V_{DS}+V_{DSS}}{V_{DSS}} \text{ in non-saturation} \\ 1 \text{ in saturation } (V_{DS}>V_{DSS}) \end{cases} \quad (5.41)$$

Past works [5,6] have implicitly assumed $\lambda=1$. Plots of the circuit elements calculated from combining this approximation with the expressions shown in Fig. 5.4 are given in Fig. 5.5 and each is defined in Table 5.1.

III.6 Illustrative Circuit Example

To illustrate the practical importance of the new model developed here, consider the MOS circuit of Fig. 5.6, which employs a substrate bias of V_{BB} volts. In this figure are shown qualitative sketches of the waveforms of the output and input voltages and of certain other voltages of interest. The new model developed here makes several contributions, as follows, to the analysis of this circuit that appear to have practical significance.

For MOS transistor T_2, the new model provides an expression for the capacitance C_{24} between the drain and the substrate which apparently was not available before in a form convenient for computer-aided circuit analysis. Thus the new model provides for an accurate calculation of the transient current flowing between the drain and the substrate of T_2. This current flows through the substrate power supply and in a large circuit could influence the current-handling capability of this power supply.

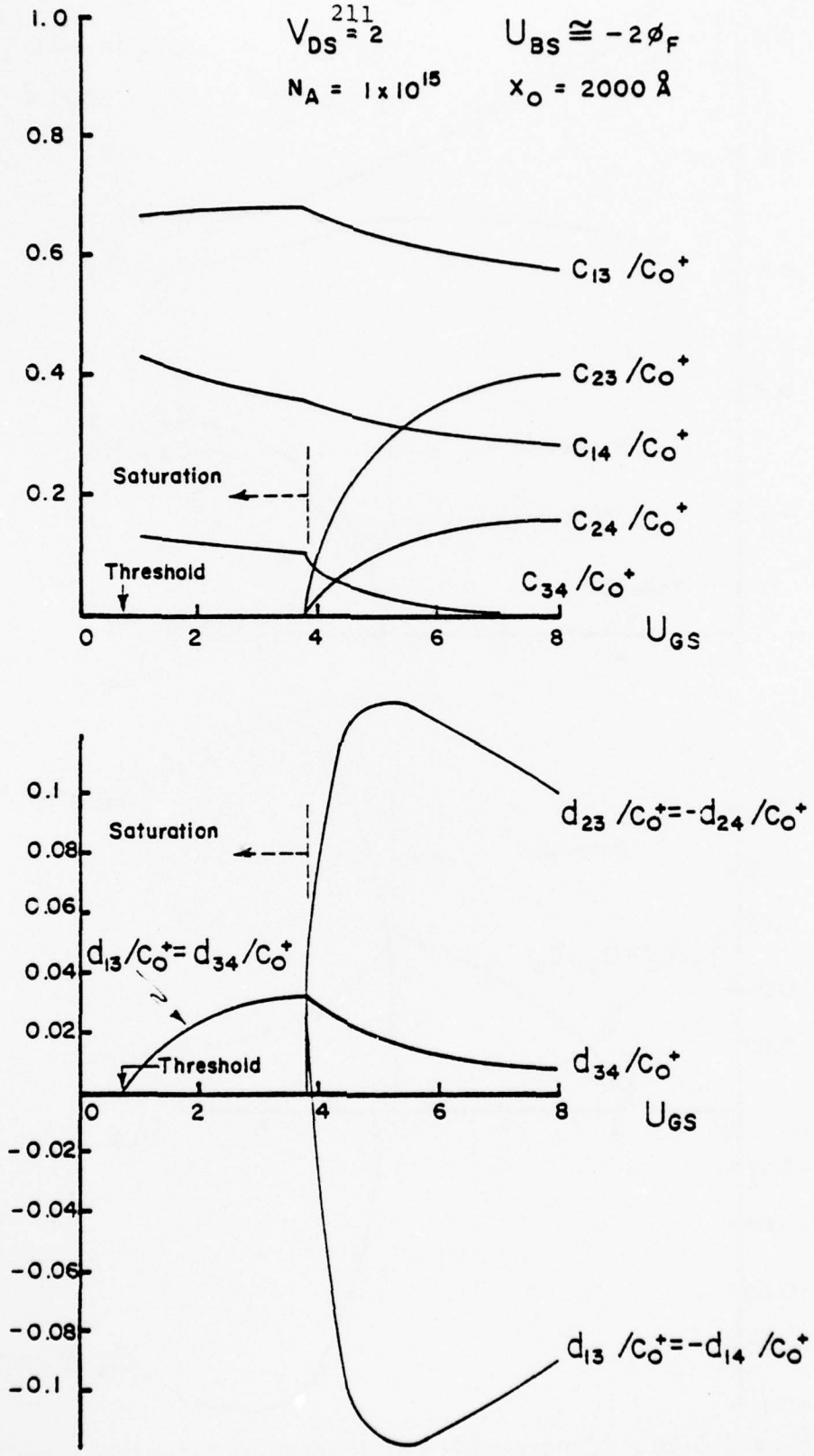


Figure 5.5(a). Functional dependence of capacitors and transcapacitors.

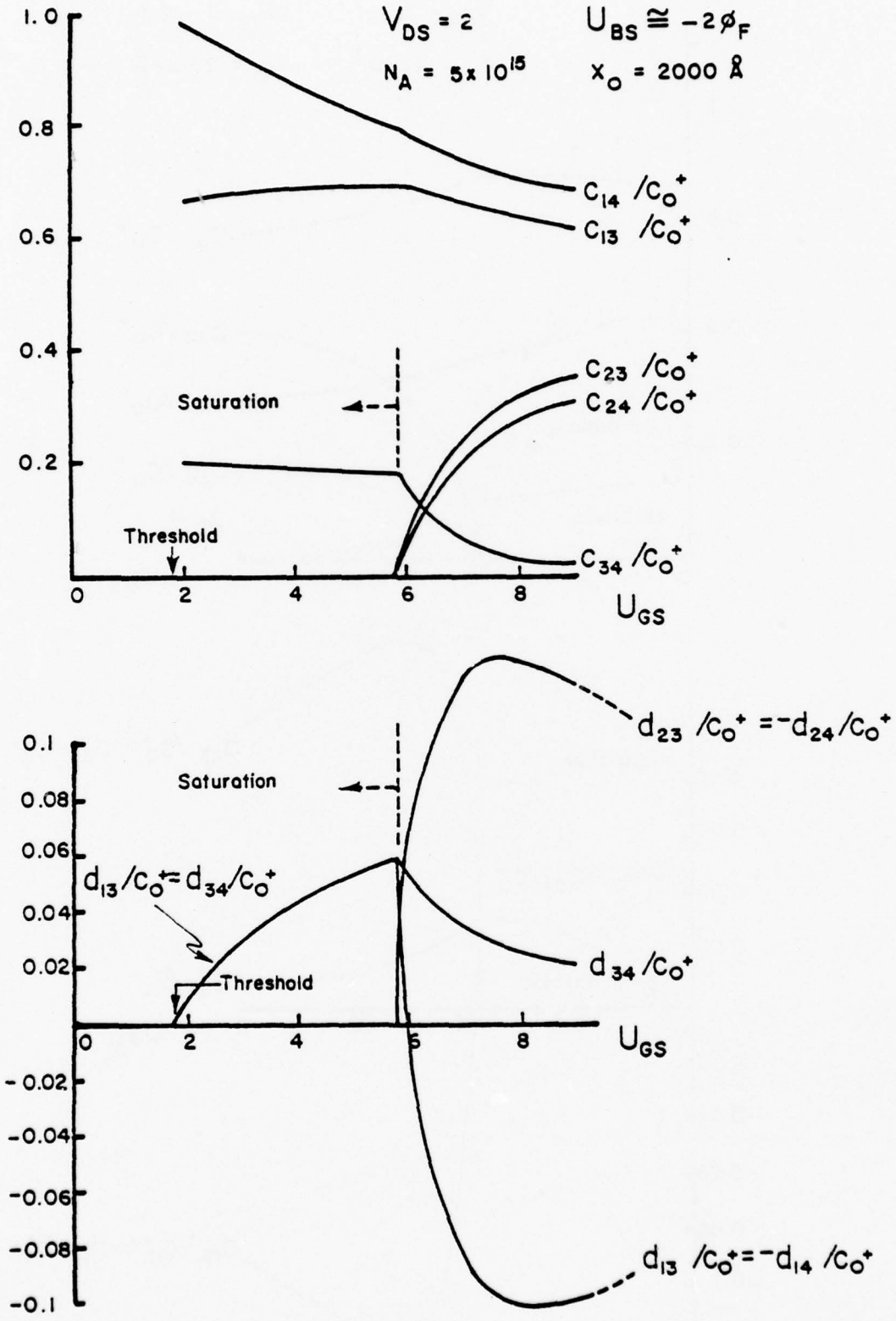


Figure 5.5(b). Functional dependence of capacitors and transcapacitors.

TRANSPORT CURRENT

$$\frac{i_{12}}{K} = -\left(\frac{A_G}{2} + D_G\right)$$

CAPACITORS

$$\frac{C_{13}}{C_o^*} = -\frac{K}{I_D}\left[A_S+D_S - \frac{K}{I_D}(A+D)\,\frac{1}{K}\,\frac{\partial I_D}{\partial V_S}\right]$$

$$\frac{C_{23}}{C_o^*} = -\frac{K}{I_D}\left[A_D+D_D - \frac{K}{I_D}(A+D)\,\frac{1}{K}\,\frac{\partial I_D}{\partial V_D}\right]$$

$$\frac{C_{14}}{C_o^*} = -\frac{K}{I_D}\left[B_S+D_S - \frac{K}{I_D}(B+D)\,\frac{1}{K}\,\frac{\partial I_D}{\partial V_S}\right]$$

$$\frac{C_{24}}{C_o^*} = -\frac{K}{I_D}\left[B_D+D_D - \frac{K}{I_D}(B+D)\,\frac{1}{K}\,\frac{\partial I_D}{\partial V_D}\right]$$

$$\frac{C_{34}}{C_o^*} = -\frac{K}{I_D}\left[B_G+D_G - \frac{K}{I_D}(B+D)\,\frac{1}{K}\,\frac{\partial I_D}{\partial V_G}\right]$$

TRANSCAPACITORS

$$\frac{d_{13}}{C_o^*} = -\frac{d_{14}}{C_o^*} = -\lambda\,\frac{K}{I_D}\left[A_G+B_G+2D_G + \frac{K}{I_D}(A+B+2D)\,\frac{1}{K}\,\frac{\partial I_D}{\partial V_G}\right] + \frac{C_{13}}{C_o^*}$$

$$\frac{d_{23}}{C_o^*} = -\frac{d_{24}}{C_o^*} = -(1-\lambda)\,\frac{K}{I_D}\left[A_G+B_G+2D_G + \frac{K}{I_D}(A+B+2D)\,\frac{1}{K}\,\frac{\partial I_D}{\partial V_G}\right] + \frac{C_{23}}{C_o^*}$$

$$\frac{d_{34}}{C_o^*} = \frac{d_{13}}{C_o^*} + \frac{d_{23}}{C_o^*}$$

The terms

$$\frac{I_D}{K},\ \frac{1}{K}\frac{\partial I_D}{\partial V_S},\ \frac{1}{K}\frac{\partial I_D}{\partial V_D},\ \frac{1}{K}\frac{\partial I_D}{\partial V_G},\ A_S, B_S, D_S;\ A_D,\ B_D,\ D_D;\ A_G,\ B_G,\ D_G;$$

A, B, C are defined in the text.

Table 5.1. Functional dependence of each circuit element in the model for the intrinsic MOSFET.

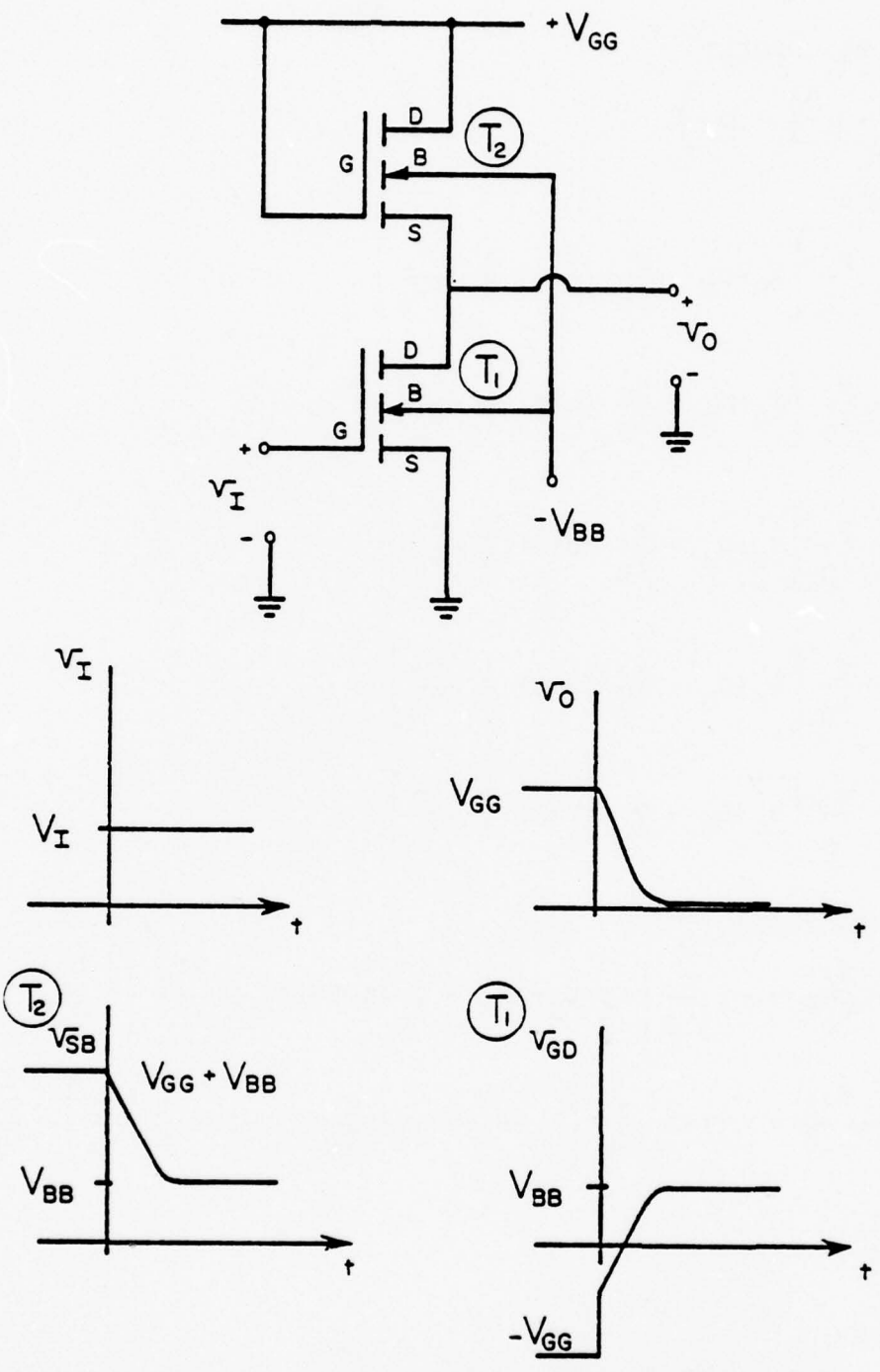


Figure 5.6. Example of a circuit illustrating significance of the model.

In a similar way, for MOS transistor T_1 the model provides new and convenient expressions for C_{14} and C_{24} that enable an accurate calculation of components of the current through the substrate power supply. These currents have the engineering importance described directly above. For calculating the rise times and other related behavior while the gate voltage is changing rapidly ($\dot{v}_{GS}$ large), the transcapacitors d_{13} and d_{23} need to be included. An examination of the dependencies of the capacitors and transcapacitors in Fig. 5.5 shows that neglect of the transcapacitors can introduce errors of about 25% in the calculation of the charging currents to the gate. This emphasizes for this particular example the inaccuracies in calculation that could result from neglect in the MOSFET network representation of the circuit elements required by the non-reciprocity of the device.

Chapter VI

A Two-Dimensional Model for MOSFET Operation

A. D. Sutherland

		Page
1.0	Introduction	217
2.0	Description of the Model	218
2.01	Basic Geometry	218
2.02	Dimensionless Variables	221
2.03	Basic Equations	221
2.04	Boundary Conditions	224
2.04.1	Boundary Conditions on ψ and n	224
2.04.2	Boundary Conditions on θ	228
2.05	Determination of the Current Scale Factor J_o	228
3.0	Finite-Difference Algorithms	229
3.01	Finite-Difference Algorithms for ψ	231
3.01.1	Implementation of the Boundary Conditions on ψ	233
3.01.2	Numerical Algorithm for the Charge Term Q_p	238
3.02	Finite-Difference Algorithm for the Stream Function	242
3.02.1	Implementation of the Boundary Conditions on θ	247
3.03	Finite-Difference Algorithm for the Current Scale Factor J_o	248
3.04	Finite-Difference Algorithm for the Electron Number Density n	249
4.0	The Iterative Procedure	251
4.01	Picard Iteration	253
4.02	Gummel's Algorithm	255
5.0	Concluding Remarks	256

CHAPTER VI

A Two-Dimensional Model for MOSFET Operation

A. D. Sutherland

1.0 Introduction

The one-dimensional model of the MOSFET described in Chapter I could not have been developed without a detailed understanding of the two-dimensional physical processes occurring in this semiconductor device. Those mechanisms, discussed in detail in that chapter, were studied through extensive use of a two-dimensional computer model of the MOSFET developed for IBM by M. S. Mock [1-3]. This useful analytic tool, which has produced results in good agreement with experiment [4], was provided to us by IBM with the understanding that it would be regarded as their proprietary property.

In order to make such a two-dimensional analytic tool available to government laboratories and the semiconductor industry, we have undertaken the development of such a computer model under this contract. It is our intention to disseminate this computer program to those desiring it, together with a complete USER'S MANUAL which will be published as a separate report.

It should be emphasized that the proprietary nature of the IBM program has been scrupulously observed during this development. Some of the algorithms implemented in our model are due to Mock, but only those which he has described in the literature [1-3]. No attempt has been made to copy the IBM program. We have, of course, made liberal use of that program as a means of providing numerical checks of the performance of our model. The availability of that program for this purpose is gratefully acknowledged.

---

[1] M. S. Mock, Solid State Electronics, 16, 601 (1973).

[2] D. P. Kennedy, Mathematical Simulation of the Effects of Ionizing Radiation on Semiconductors, Scientific Report No. 1, AFCRL-71-0272 (1971).

[3] D. P. Kennedy, Mathematical Simulation of the Effects of Ionizing Radiation on Semiconductors, Final Report, AFCRL-72-0257 (1972).

[4] D. P. Kennedy, Private Communication.

2.0 Description of the Model

The computer model treats an n-channel enhancement mode MOSFET in which the conducting channel is formed by a gate-induced inversion layer just below the interface between a uniformly doped p-type substrate and the SiO_2 layer (hereinafter named the oxide) upon which the gate metallization is deposited. The inherent thinness of the inversion layer justifies the assumptions made that the net electron-hole recombination rate and the hole current are negligible in the steady state operation of this device. Details are given in the following subsections.

2.01 Basic Geometry

Figure 6.1(a) depicts such an n-channel device in which the source and drain consist of n^+ islands diffused into the p-substrate. The thinness of the gate-induced inversion layer of mobile electrons formed just below the oxide-silicon interface makes possible the adoption of the simplified rectangular geometry of Figure 6.1(b), wherein the vertical metallurgical p-n junctions (assumed to be abrupt transitions between p and n^+ regions) are located at the left and right borders of the rectangular region treated. Large gradients in electric potential and, consequently, in the electron and hole densities are expected in the immediate vicinity of those junctions, requiring a rather dense array of the geometric points at which those variables are calculated with the finite-difference equations employed. The same is true in the region just below the oxide-silicon interface where the inversion layer forms. On the other hand, the density of such points in regions remote from those surfaces can be relatively sparse. By placing the source and drain junctions at the borders of the rectangular region treated, such a graded lattice of points is readily defined. Figure 6.2 illustrates the graded lattice employed by the model with a 41 (horizontal) by 25 (vertical) array overlayed on the rectangle. The prescription adopted for generating this graded lattice is precisely that suggested by Mock [1].

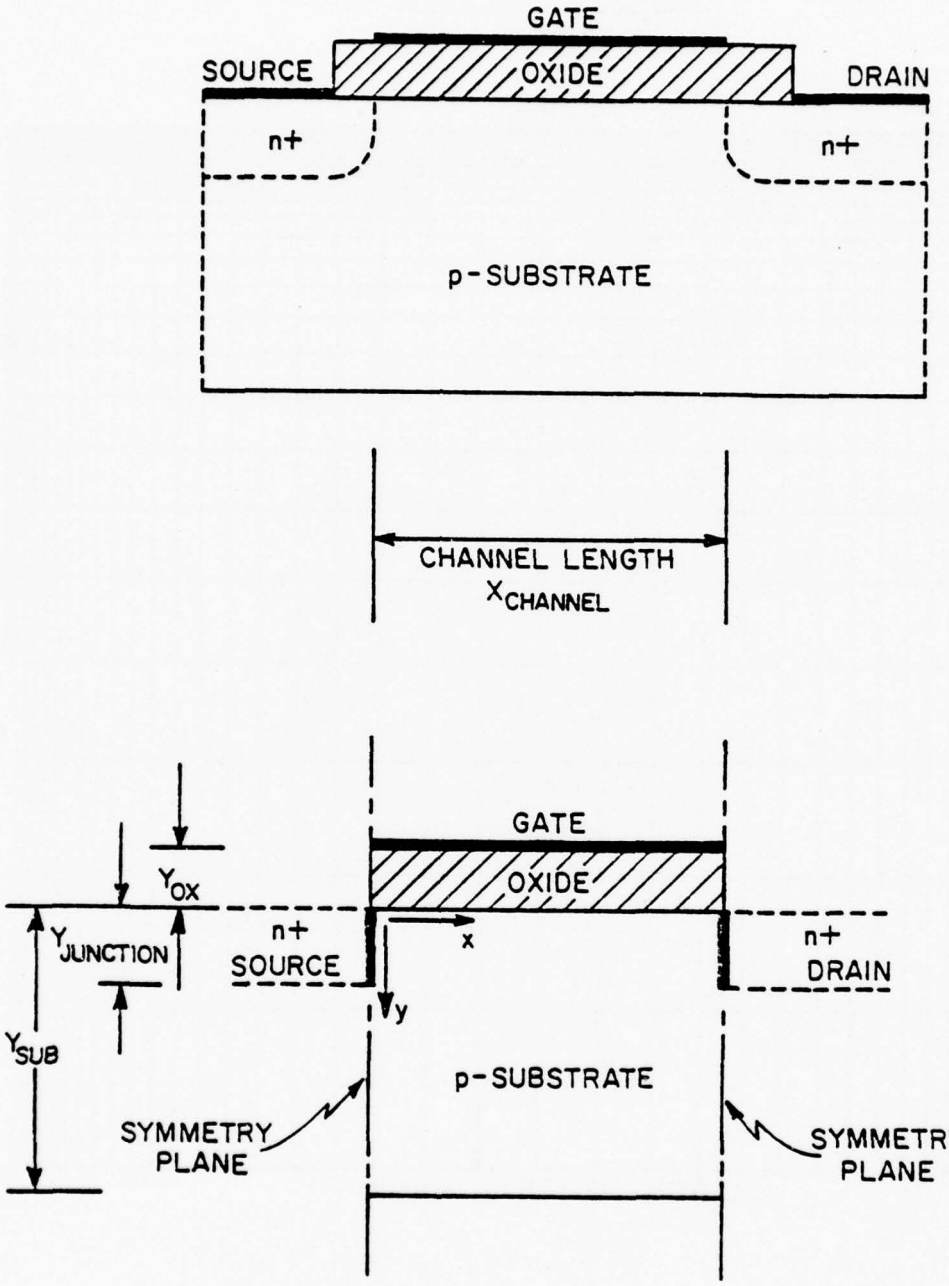


Figure 6.1 (a) Cross-section of an n-channel MOSFET.

(b) Rectangular region modeled. The abrupt metallurgical junctions between source-substrate and drain-substrate are regarded as planar and at the left and right borders of the region modeled.

GRADED LATTICE

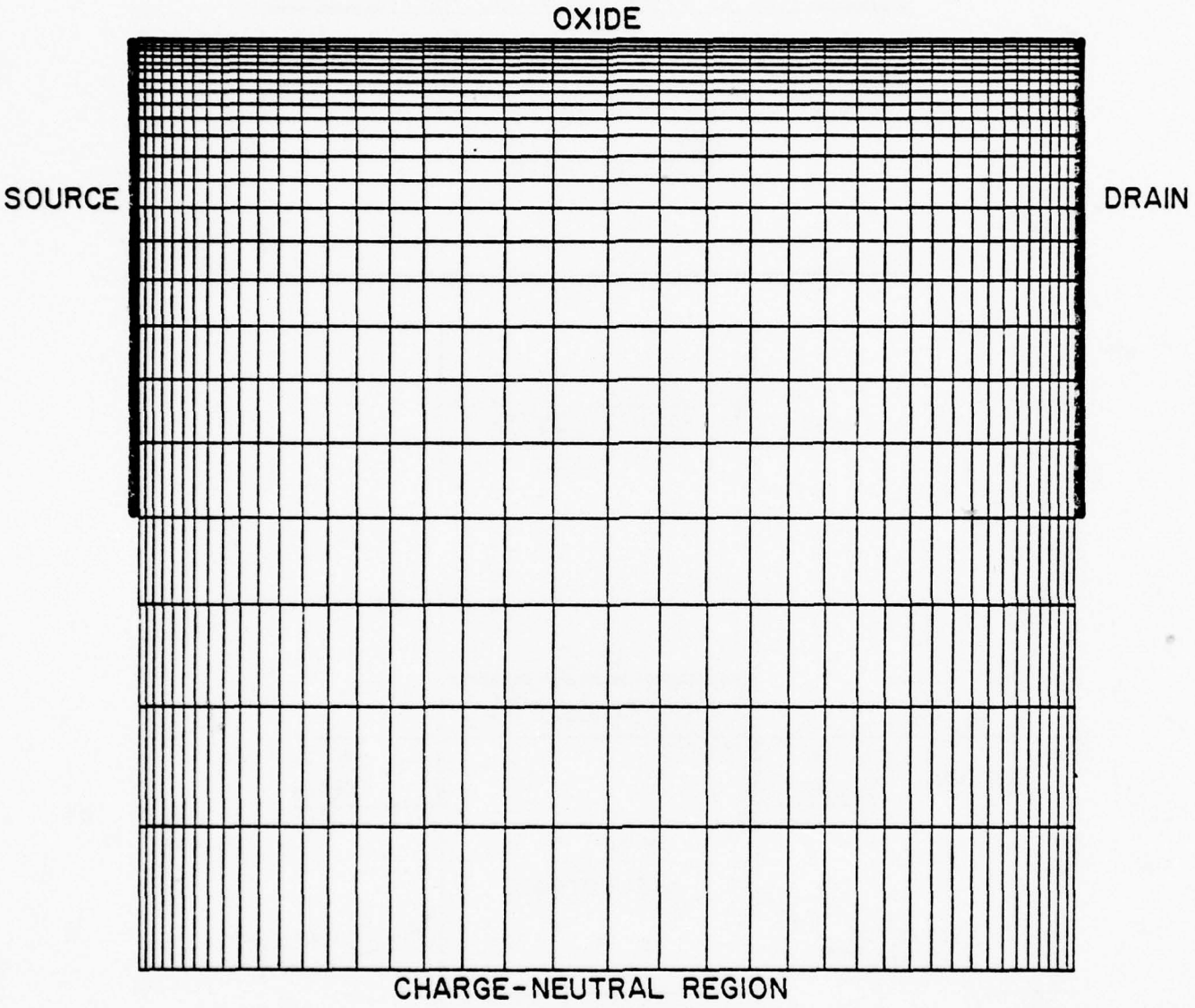


Figure 6.2 Graded lattice of points used in the finite-difference equations dealt with in the computer model.

2.02 Dimensionless Variables

For computational purposes, the adoption of dimensionless (normalized) variables at the outset results in two benefits: (1) unnecessary multiplications by redundant constants such as ε_o, q, kT/q, n_i, etc. can be precluded, and (2) numerical scaling causing most variables to lie within reasonable numerical ranges can be achieved.

Understanding that such dimensionless variables will be utilized in all of which follows (with few exceptions which will be clearly flagged), we shall designate physical variables (i.e., those which possess dimensions) in this subsection by means of primed symbols, using corresponding unprimed symbols for their dimensionless counterparts. This convention provides the advantage of utilizing conventional symbols such as $\bar{J}$ for current density, n and p for electron and hole densities, ψ for electric potential, etc. even though those variables are normalized. In this vein, the dimensionless variables employed in our model are defined in Table I.

2.03 Basic Equations

Having adopted the assumptions of negligible hole current and electron-hole recombination, the dimensionless equations describing the physical mechanisms within this semiconductor device are:

$$\nabla^2\psi = N_A + n - p \ , \tag{6.01}$$

$$n = \exp(\psi-\phi_n) \ , \tag{6.02}$$

$$p = \exp(\phi_p-\psi) \ , \tag{6.03}$$

$$\bar{J} = -\mu n\nabla\phi_n \ , \tag{6.04}$$

$$0 = \nabla\phi_p \ , \tag{6.05}$$

$$\nabla\cdot\bar{J} = 0 \ . \tag{6.06}$$

Equation (6.01) is Poisson's equation for the electric potential ψ, (6.02) and (6.03) express the use of Boltzmann statistics

TABLE I--DEFINITION OF DIMENSIONLESS VARIABLES
(κ_s is the relative permittivity of silicon)

Define:

$$L_D' = \sqrt{\frac{\kappa_s \varepsilon_o' (k'T'/q')}{q' n_i'}} \quad , \qquad \text{Intrinsic Debye length of silicon}$$

$$\mu_n' \quad , \qquad \text{Low field electron mobility}$$

Then Let:

$$\left.\begin{aligned} \psi' &= (k'T'/q')\psi \\ \phi_n' &= (k'T'/q')\phi_n \\ \phi_p' &= (k'T'/q')\phi_p \end{aligned}\right\} \quad , \qquad \text{Electric potential, and electron and hole quasi-Fermi potentials}$$

$$\left.\begin{aligned} N_D' &= n_i' N_D \\ N_A' &= n_i' N_A \\ n' &= n_i' n \\ p' &= n_i' p \end{aligned}\right\} \quad , \qquad \text{Donor, acceptor, electron, and hole number densities}$$

$$\left.\begin{aligned} x' &= L_D' x \\ y' &= L_D' y \end{aligned}\right\} \quad , \qquad \text{Cartesian coordinates}$$

$$\bar{J}' = \left[\frac{(k'T'/q')\mu_n' n_i' q'}{L_D'} \right] \bar{J} \quad , \qquad \text{Electron current density}$$

$$Q_{ss}' = \left[\frac{(k'T'/q')\kappa_s \varepsilon_o'}{L_D'} \right] Q_{ss} \quad , \qquad \text{Surface charge density}$$

$$\mu' = \mu_n' \mu \quad , \qquad \text{Electron mobility}$$

for the electron and hole number densities n and p (i.e., non-degenerate doping of the p-substrate is implied), (6.04) relates the electron current density vector $\bar{J}$ to the electron quasi-Fermi potential ϕ_n and number density n in the usual manner [5], (6.05) enforces zero hole current density, while (6.06) assures conservation of charge under the assumed condition of zero net recombination of electron-hole pairs.

The above equations apply only in the rectangular p-substrate region. The conditions in the oxide are described by:

$$\nabla^2\psi = 0 \ , \tag{6.07}$$

and $n = p = \bar{J} \equiv 0$, with appropriate boundary conditions to match the tangential and normal electric fields at the oxide-silicon interface.

The divergence-free flow of electron current expressed by (6.06) can be enforced by introducing a stream function $\theta(x,y)$, as did Mock, such that the x and y components of $\bar{J}$ are given by:

$$J_x = J_o \frac{\partial\theta}{\partial y} \ , \tag{6.08}$$

$$J_y = -J_o \frac{\partial\theta}{\partial x} \ , \tag{6.09}$$

thereby assuring that $\nabla\cdot\bar{J} = 0$. Here, J_o is a scalar constant whose determination will be discussed later. We next seek the differential equation which must be satisfied by this stream function.

Solving (6.02) for the electron quasi-Fermi potential ϕ_n, and substituting the result into (6.04), we obtain:

$$\bar{J} = \mu e^{\psi}\nabla(ne^{-\psi}) \ . \tag{6.10}$$

Identifying the x and y components of (6.10) with the correspoding expressions in (6.08) and (6.09), we have:

$$J_o \frac{\partial\theta}{\partial y} = \mu e^{\psi} \frac{\partial}{\partial x}(ne^{-\psi}) \ , \tag{6.11}$$

---

[5] S. M. Sze, Physics of Semiconductor Devices, p. 96, Wiley, N.Y. (1969).

$$J_o \frac{\partial \theta}{\partial x} = -\mu e^{\psi} \frac{\partial}{\partial y} (ne^{-\psi}) \quad . \tag{6.12}$$

Divide both sides of these intermediate equations by μe^{ψ}, then differentiate the first partially with respect to y, the second partially with respect to x, and add the two, obtaining:

$$\frac{\partial}{\partial x} \left[\mu^{-1} e^{-\psi} \frac{\partial \theta}{\partial x} \right] + \frac{\partial}{\partial y} \left[\mu^{-1} e^{-\psi} \frac{\partial \theta}{\partial y} \right] = 0 \quad . \tag{6.13}$$

Eq. (6.13) is a second order differential equation in θ, given $\mu(x,y)$ and $\psi(x,y)$. Given its solution, the electron current density vector $\bar{J}$ is then known, from (6.08) and (6.09).

The manner in which the current density scale factor J_o is determined is best deferred until after a discussion of the boundary conditions which must be satisfied by the solutions of these basic equations.

2.04 Boundary Conditions

We regard the left and right borders of the rectangular region dealt with by the model as symmetry planes such that ψ, θ, n, p, and J_y are even functions of x, while J_x is an odd function. Image theory then requires that these functions all be periodic in x, the periodicity being twice the channel length $X_{channel}$ of Figure 6.1(b). A consequence of this implied periodicity is that the source and drain islands shown dotted in Figure 6.1(a) are "squeezed out of the picture," being replaced by infinitesimally thin source and drain "contacts" which serve as sources or sinks for current and electric flux lines. The assignment of boundary conditions on ψ and n to these "contacts" bears careful scrutiny.

2.04.1 Boundary Conditions on ψ and n--We select the charge-neutral source island as the zero reference for electric potential, $\psi = 0$ (even though that region has been "squeezed out of the picture" in the model), and assume that the contact potentials of all metal-semiconductor and metal-oxide contacts are identical and, therefore, can be ignored. The p-substrate, in the absence of any external bias voltages, assumes the normalized "built-in" potential:

$$\psi_{EQ} = -\ell n(N_D N_A) \ , \tag{6.14}$$

which is also assumed by the gate. If the substrate and gate are then biased positively* with respect to the source at potentials V_{SUB} and V_G, respectively, the boundary conditions stipulated along the two horizontal borders in Figure 6.3(a) are obtained.

Consider next the source and drain "contacts" which, as stated above, are assumed to coincide with the metallurgical junction planes. Letting U be the normalized potential with zero reference redefined (for the present discussion) to be that of the charge-neutral p-region, Kennedy [6] shows rigorously that the potential U_J resulting in the plane of the metallurgical junction defining an abrupt p-n junction is:

$$U_J = U_T \left[\frac{N_D}{N_A+N_D}\right] + \left[\frac{N_A-N_D}{N_A+N_D}\right]\left[1 - \exp(-U_T)\right] \ , \tag{6.15}$$

where U_T is the total potential appearing across the p-n junction:

$$U_T = \ell n(N_D N_A) + U_{applied} \ . \tag{6.16}$$

(Note that positive $U_{applied}$ in (6.16) results in a reverse-biased junction, consistent with Kennedy's convention.)

Consider some typical numerical values in (6.15) and (6.16). With $n_i' = 1.5\text{x}10^{10}$ cm^{-3}, $N_D' = 10^{19}$ cm^{-3}, $N_A' = 10^{16}$ cm^{-3} (the primes flag unnormalized constants), $\ell n(N_D N_A)$ assumes the numerical value 33.7, and the exponential term in (6.15) is entirely negligible except for negative values of $U_{applied}$ (i.e., forward bias) approaching that numerical magnitude. Furthermore, since $N_D >> N_A$, (6.15) simplifies tremendously, becoming:

$$U_J \approx U_T - 1 \ . \tag{6.17}$$

Since the drain junction is always operated reverse-biased, with $V_D > 0$, (6.17) is valid at the drain "contact."

*Later in this development, we tacitly assure that V_{SUB} will always be a negative numeric.

[6] D. P. Kennedy, IEEE Trans. Electron Devices, ED-22, 988 (1975).

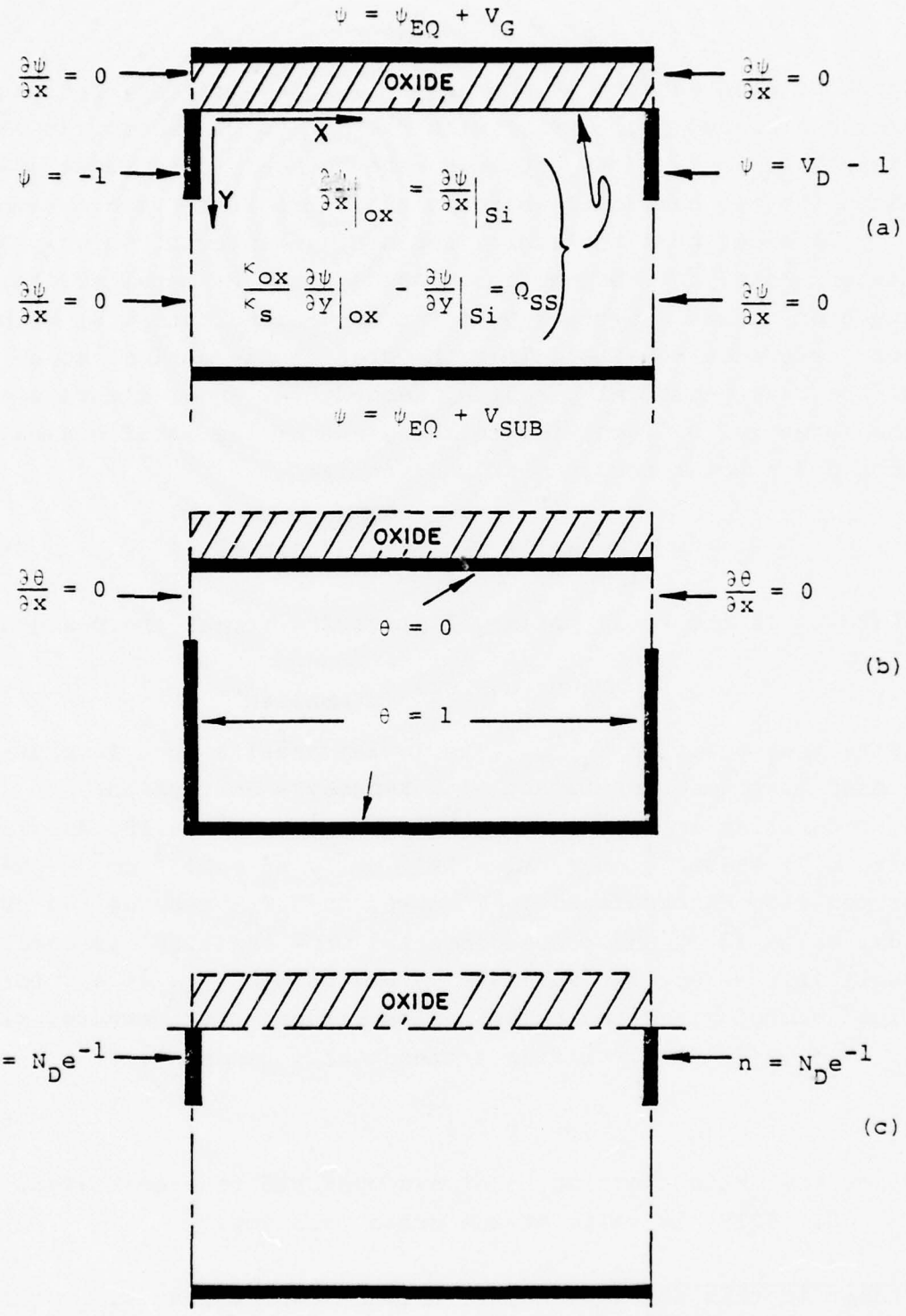


Figure 6.3 Boundary conditions for: (a) $\psi(x,y)$, (b) $\theta(x,y)$, and (c) $n(x,y)$.

Consider the source "contact," which does become forward-biased due to band bending near the oxide-silicon interface as the applied gate potential V_G is made positive with respect to the source. At the onset of strong inversion, the total normalized potential drop across the n^+p source junction right at the oxide-silicon interface will be reduced to $U_T = \ell n(N_D/N_A)$, which assumes the value $U_T = 6.9$, using the same numerical values for N_D' and N_A' assumed above. Again, the exponential term in (6.15) is negligible. We conclude that (6.17) is still a valid approximation for the source contact, at least for N_D/N_A ratios of 10^3 or higher. The approximation begins to become questionable only when $N_D/N_A < 10^2$, but at the source "contact" only.

Reverting to our convention of selecting the charge-neutral source island for zero reference potential, we conclude from (6.17) that the appropriate boundary conditions on ψ at the source and drain "contacts" are $\psi = -1$, and $\psi = V_D - 1$, respectively, as shown in Figure 6.3(a).

The remaining boundary conditions on ψ shown in that figure assure the symmetry assumed at the left and right borders of the region, and assure proper matching of the solutions of (6.01) and (6.07) at the oxide-silicon interface. Note that the condition imposed upon the normal derivatives of ψ at that interface allows for the inclusion of surface charge Q_{SS}, a feature not included by Mock. (The condition stated in the figure for the normal derivatives assumes an unfamiliar form because of the normalized variables employed.)

The numerical method later to be described for determining the electron number density $n(x,y)$ requires that n be specified along the borders of the rectangular region only at the source and drain "contacts." The fact that those "contacts" have potentials one normalized unit lower than those of the corresponding charge-neutral source and drain islands (where $n = N_D$) leads to the boundary conditions on n depicted in Figure 6.3(c).

2.04.2 Boundary Conditions on θ--Referring to (6.08) and (6.09), we note that θ is defined only in terms of its spatial derivatives. This means that its magnitude can be arbitrarily bounded. A convenient choice for those bounds is $0 \leq \theta \leq 1$, with θ specified to be zero along the oxide-silicon interface as in Figure 6.3(b). Tacitly assuming that the substrate (back-gate) will always be biased negatively with respect to the source ($V_{SUB} \leq 0$), so that no back-gate current flows, the appropriate boundary conditions on θ along the remaining borders in the figure are those shown. The stipulation $\frac{\partial \theta}{\partial x} = 0$ along the source and drain "contacts" assures zero tangential current at those "contacts," and is consistent with the implied symmetry conditions discussed earlier.

Bounding the stream function between zero and unity has a useful consequence. The numerical value of $\theta(x,y)$ at any arbitrary point (x,y) within the p-substrate gives, by inspection, the decimal fraction of the total drain current I_D crossing a vertical plane erected between the point (x,y) and the point $(x,0)$ at the oxide-silicon interface. This interpretation follows immediately from (6.08).

2.05 Determination of the Current Scale Factor J_o

The constant J_o appearing in (6.08) and (6.09), which relates the current density vector $\bar{J}$ to the stream function θ, is obtained in the following manner. Eq. (6.11) gives:

$$\frac{\partial}{\partial x}\left(ne^{-\psi}\right) = \mu^{-1}e^{-\psi}J_o \frac{\partial \theta}{\partial y} . \qquad (6.18)$$

Integrating with respect to x along a horizontal line (y = constant) which intersects the source and drain "contacts," one obtains:

$$N_D\left[e^{-V_D} - 1\right] = J_o \int_0^{X_{channel}} \mu^{-1}e^{-\psi} \frac{\partial \theta}{\partial y}\, dx . \qquad (6.19)$$

Solving for J_o:

$$J_o = \frac{-N_D[1-e^{-V_D}]}{\int_0^{X_{channel}} \mu^{-1}e^{-\psi} \frac{\partial \theta}{\partial y}\, dx} . \qquad (6.20)$$

This is the same result used by Mock for calculating J_o, except for the factor N_D appearing in (6.20), which is absent in Mock's result ([1], eq. (10)). The reason for this difference is seated in the fact that Mock chose to set the electron quasi-Fermi potential ϕ_n equal to zero in the charge-neutral source region, which makes $\psi = \ell n(N_D)$ in that region, whereas we have chosen to set $\psi = 0$ there. Note that the sign of J_o changes if one were to perform the integration from x = 0 to the "image" drain contact at $x = -X_{channel}$. Thus, J_o is an odd function of x, as it must be in order to satisfy the symmetry conditions on J_x and J_y described in Section 2.04.

3.0 Finite-Difference Algorithms

Figure 6.4 defines the indices (I,J), and their bounds, used to identify each lattice point involved in implementing finite-difference algorithms for determining ψ, θ, and n. (The figure does not attempt to convey the graded nature of the lattice actually used.) As shown, two "interlaced" lattices are employed. The first, depicted by circles, represents the set of points at which the electrostatic potential ψ and the corresponding electron and hole densities n and p are numerically determined while the second, depicted by squares, represents the set of points at which the stream function θ is evaluated. This choice, although complicating the satisfaction of the boundary conditions on θ, allows the use of simple differencing between adjacent lattice points in evaluating $\partial\theta/\partial x$ and $\partial\theta/\partial y$.* Mock also employed such an interlaced pair of lattices.

---

*In an earlier version of the model, the author chose to deal with a single lattice of points, rather than the interlaced lattices presently employed. In the development of the computer code for that choice, it was soon discovered that the nonuniform point spacings of the lattice precluded the use of simple differencing for evaluating derivatives. It was found necessary to fit parabolas through sets of three adjacent points in a row or column in order to evaluate the partial derivatives needed. While this was easily accomplished with a single subroutine for that purpose, nonetheless each such operation required 7 multiplications and 7 additions each time the subroutine was called. This imposed an unacceptable cost in computation time, and the approach was abandoned.

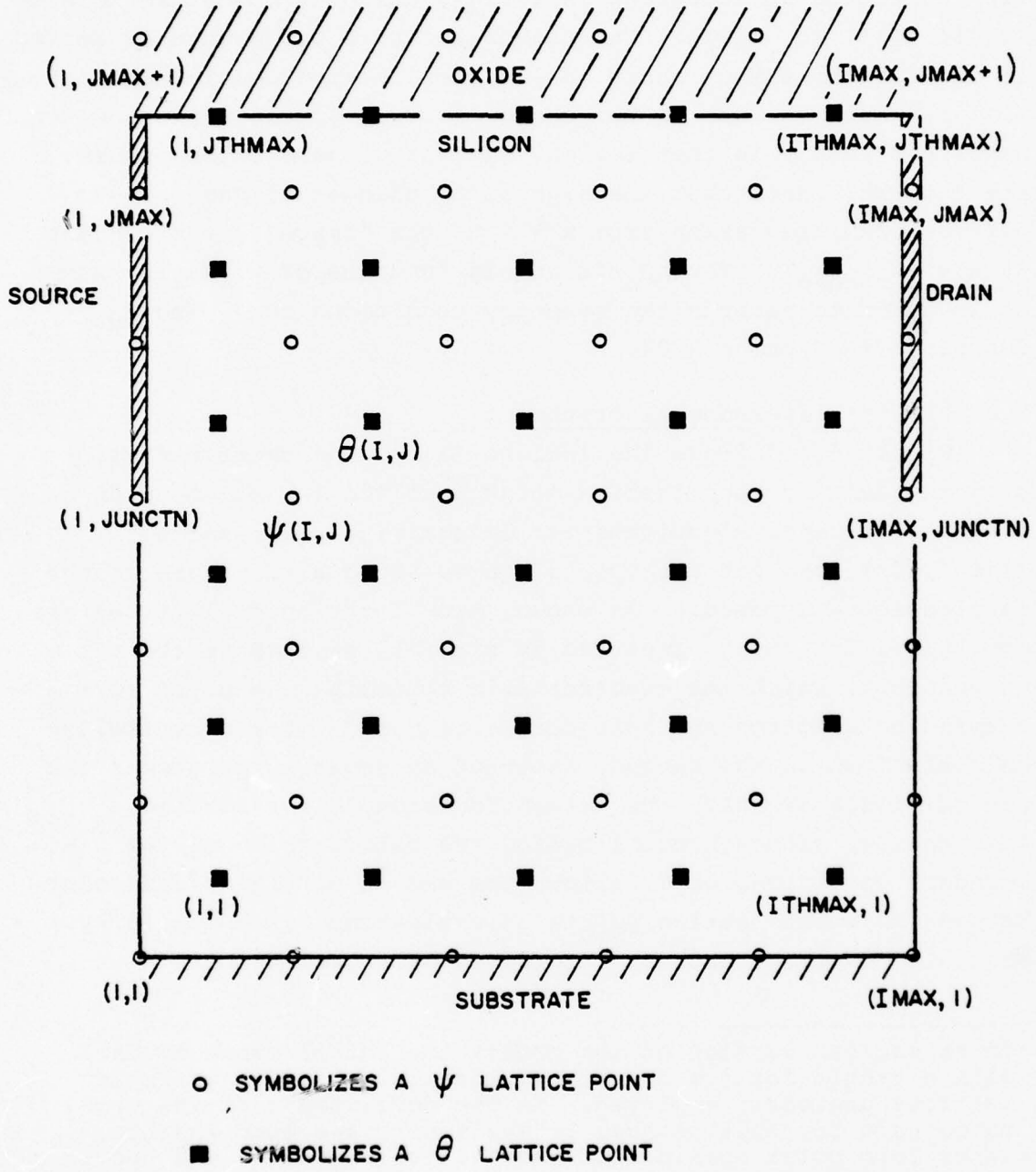


Figure 6.4 Layout of the interlaced ψ and θ lattices used in the finite-difference equations.

Note that the oxide-silicon interface coincides with the upper border, J = JTHMAX, of the θ lattice. Although not shown in Figure 6.4, the ψ lattice extends vertically an additional four rows (uniformly spaced) beyond the row J = JMAX + 1 shown, to model the electrostatic potential within the oxide region. Note, also, the convention adopted for the indices of the ψ lattice with respect to those of the θ lattice. (Given the ordered pair (I,J), the point of the θ lattice so identified is "northeast" of the corresponding point of the ψ lattice.)

3.01 Finite-Difference Algorithms for ψ

Consider the five point "star" of points in the ψ lattice shown in Figure 6.5. The dotted rectangle shown in that figure, having dimensions $\tilde{\Delta}x(I)$ by $\tilde{\Delta}y(J)$, is bounded at its four corners by the corresponding θ lattice points "NE, NW, SE and SW" of the central point P. (We shall adopt "points of the compass" notation, using subscripts in the difference equations to follow, to avoid repeated use of the index pairs (I,J), (I+1,J), etc., which are used in their FORTRAN counterparts in the computer program.) The lengths $\tilde{\Delta}x(I)$, $\tilde{\Delta}y(J)$, abbreviated $\Delta\tilde{x}_P$ and $\tilde{\Delta}y_P$ in the difference equations, are defined to be:

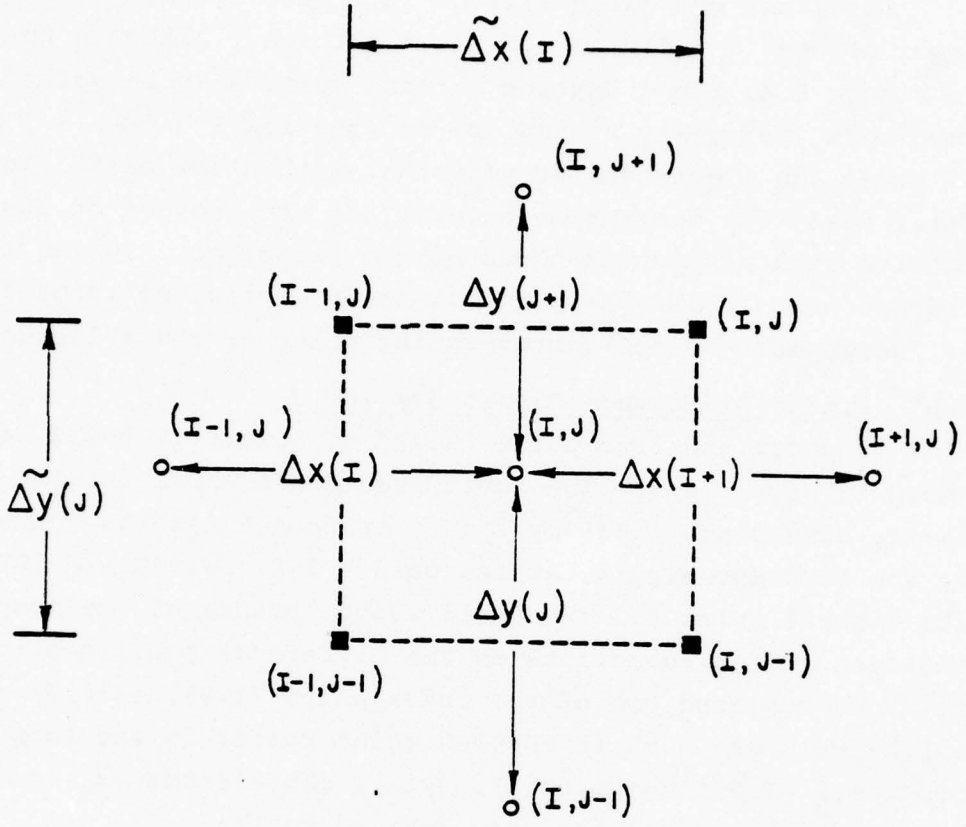
$$\left.\begin{aligned} \tilde{\Delta}x &= (\Delta x_E + \Delta x_W)/2 \ , \\ \tilde{\Delta}y &= (\Delta y_N + \Delta y_S)/2 \ . \end{aligned}\right\} \tag{6.21}$$

This locates the θ lattice points always midway between their neighboring ψ lattice points.

Applying Gauss's law to the rectangular region surrounding the point (I,J) in Figure 6.5, one obtains:

$$B_P\psi_S + D_P\psi_W + E_P\psi_P + F_P\psi_W + H_P\psi_N = Q_P \ , \tag{6.22}$$

where:



○ SYMBOLIZES A ψ LATTICE POINT

■ SYMBOLIZES A θ LATTICE POINT

Figure 6.5 Five-point "star" of ψ-lattice points used in the finite-difference equation for ψ.

$$\left.\begin{aligned} B_P &= \tilde{\Delta}x_p/\Delta y_s \ , \\ D_P &= \tilde{\Delta}y_p/\Delta x_w \ , \\ E_P &= -(B_P+D_P+F_P+H_P) \ , \\ F_P &= \tilde{\Delta}y_p/\Delta x_E \ , \\ H_P &= \tilde{\Delta}x_p/\Delta y_N \ , \\ Q_P &= \int\int_{BOX} (N_A+n-p)\,dxdy \ . \end{aligned}\right\} \qquad (6.23)$$

With a ψ lattice whose dimensions are I_{MAX} by (JMAX+5), the set of equations represented by (6.22) can be expressed in the matrix form:

$$[L] \cdot [\psi] = [Q] \ , \qquad (6.24)$$

where $[\psi]$ and [Q] are rectangular matrices having the same dimensions as those of the ψ lattice, while [L] is a square matrix of dimensions (IMAX)(JMAX+5) by (IMAX)(JMAX+5). This matrix is sparse, having only five nonzero symmetrically positioned diagonal elements. Stone's iterative matrix factorization method [7] is employed in our model to solve (6.24) for the array ψ, given the array Q.

3.01.1 Implementation of the Boundary Conditions on ψ

The boundary conditions on ψ depicted in Figure 6.3(a) are enforced by appropriate modifications of the factors B_P, D_P, E_P, F_P, H_P, and Q_P in (6.22) at those ψ lattice points coinciding with the borders of the rectangular region modeled. These are next discussed.

3.01.1(a) Border Segments where ψ is Specified--For ψ lattice points at which ψ is fixed, the coefficients B_P, D_P, F_P, and H_P are set to zero, while E_P is set to unity. Then the term Q_P in (6.22) is set equal to the known value of ψ at the point in question. In accordance with (6.22), this forces ψ_P to assume that known value at such border points.

---

[7] H. L. Stone, SIAM Jour. Numer. Anal., 5, 530 (1968).

3.01.1(b) Border Segments along Symmetry Planes--The boundary condition $\partial\psi/\partial x = 0$ is enforced at the symmetry planes in the following manner. At the left border, the coefficient D_p in (6.22) is set to zero, while F_p is doubled. The converse is applied at the right border. This assures the required symmetry of ψ at those border segments.

3.01.1(c) The Oxide-Silicon Interface--Since the ψ lattice has no points coinciding with the interface, the enforcement of the proper boundary conditions on ψ there is somewhat more involved. We develop special forms of the finite-difference equation for ψ, which apply for the two rows of the ψ lattice, viz. J = JMAX and J = (JMAX+1), which bracket that interface. Figure 6.6 shows one set of points in those rows of the ψ lattice, and defines spacings $\Delta\eta_{ox}$ and $\Delta\eta_{si}$ which locate the interface. Through an appropriate choice of the ratio $(\Delta\eta_{ox}/\Delta\eta_{si})$, we develop an expression for the potential ψ_{surf} at the interface, in terms of the known potentials ψ_{ox} and ψ_{si} of the ψ lattice points north and south of it, such that the required discontinuity in $\partial\psi/\partial y$ is automatically satisfied. Then, an application of Gauss's law to the two rectangular regions shown crosshatched in the figure leads to the required special forms of the finite-difference equation in ψ for the two rows in question.

The boundary condition on $\partial\psi/dy$ at the interface is:

$$\left(\frac{\partial\psi}{\partial y}\right)_{si} = \left(\frac{\kappa_{ox}}{\kappa_{si}}\right)\left(\frac{\partial\psi}{\partial y}\right)_{ox} - Q_{ss} \ , \qquad (6.25)$$

which comes from applying Gauss's law to an infinitesimally thin gaussian "pillbox" bracketing the interface, with the result expressed in terms of the normalized variables employed in our model. Here, κ_{ox} and κ_{si} are the relative permittivity of the oxide and silicon regions, respectively, while Q_{ss} is the surface charge density (if any) residing at the interface. In addition to this condition, we have:

$$\psi_{si} = \psi_{ox} + \left(\frac{\partial\psi}{\partial y}\right)_{ox} \Delta\eta_{ox} + \left(\frac{\partial\psi}{\partial y}\right)_{si} \Delta\eta_{si} \ , \qquad (6.26)$$

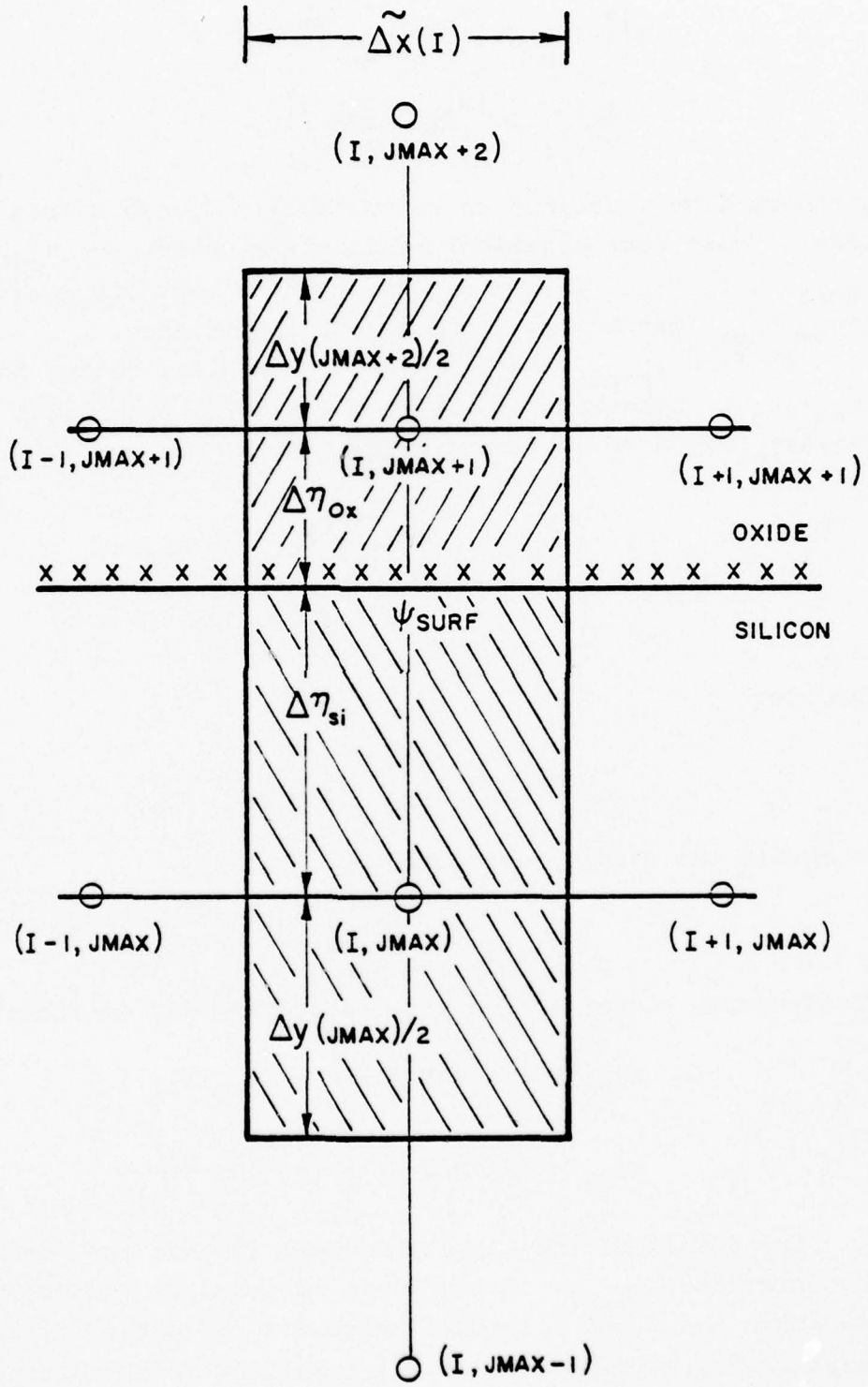


Figure 6.6 Boundary conditions on ψ at the oxide-silicon interface are enforced through an appropriate choice for the location of the interface, leading to special finite-difference equations applicable for the two ψ-lattice rows bracketing the interface.

$$\left(\frac{\partial\psi}{\partial y}\right)_{ox} = \left[\frac{\psi_{surf} - \psi_{ox}}{\Delta\eta_{ox}}\right] , \tag{6.27}$$

$$\left(\frac{\partial\psi}{\partial y}\right)_{si} = \left[\frac{\psi_{si} - \psi_{surf}}{\Delta\eta_{si}}\right] , \tag{6.28}$$

with positive y defined to be vertically downward from the interface. These four equations relate five unknowns--$\Delta\eta_{ox}$, $\Delta\eta_{si}$, ψ_{surf}, $(\partial\psi/\partial y)_{ox}$, $(\partial\psi/\partial y)_{si}$--which means that the ratio $(\Delta\eta_{ox}/\Delta\eta_{si})$ can be specified at our convenience.

To find ψ_{surf} (and discover a convenient choice for $\Delta\eta_{ox}/\Delta\eta_{si}$), substitute (6.25) for $(\partial\psi/\partial y)_{si}$ and (6.27) for $(\partial\psi/\partial y)_{ox}$ into (6.26), obtaining:

$$\psi_{surf} = \left\{ \frac{\psi_{si} + \left(\frac{\kappa_{ox}}{\kappa_{si}}\right)\left(\frac{\Delta\eta_{si}}{\Delta\eta_{ox}}\right)\psi_{ox} + Q_{ss}\Delta\eta_{si}}{1 + \left(\frac{\kappa_{ox}}{\kappa_{si}}\right)\left(\frac{\Delta\eta_{si}}{\Delta\eta_{ox}}\right)} \right\} . \tag{6.29}$$

Choosing:

$$\Delta\eta_{ox} = \left(\frac{\kappa_{ox}}{\kappa_{si}}\right)\Delta\eta_{si} , \tag{6.30}$$

we obtain the simple relationship:

$$\psi_{surf} = \left[\frac{\psi_{ox} + \psi_{si}}{2}\right] + Q_{ss}\,\frac{\Delta\eta_{si}}{2} . \tag{6.31}$$

Furthermore, since $(\Delta\eta_{ox} + \Delta\eta_{si}) = \Delta y(JMAX+1)$, we find:

$$\Delta\eta_{si} = \Delta y(JMAX+1)/[1 + (\kappa_{ox}/\kappa_{si})] , \tag{6.32}$$

$$\Delta\eta_{ox} = \Delta y(JMAX+1)\left[\frac{(\kappa_{ox}/\kappa_{si})}{1 + (\kappa_{ox}/\kappa_{si})}\right] . \tag{6.33}$$

The result of the above procedure is that the potential at the interface, ψ_{surf}, is now known in terms of the potential at its north and south ψ lattice neighbors, with the simple relationship expressed by (6.31) resulting from locating the interface as prescribed by (6.32) and (6.33). These three equations assure that the boundary condition on $(\partial\psi/\partial y)$ at the interface is automatically satisfied.

Now apply Gauss's law to the rectangular region shown in Figure 6.6 which encompasses the ψ lattice point (I,JMAX) just below the interface. Again, adopting "points of the compass" notation, the electrostatic flux entering the upper surface of that rectangle is $-\tilde{\Delta}x_P(\psi_P-\psi_{surf})/\Delta\eta_{si}$. Using (6.31) to relate ψ_{surf} to ψ_N and ψ_P, the resulting finite-difference equation still assumes the form given by (6.22). However, the terms H_P and Q_P appearing in (6.23), (and only those) are now modified, becoming:

$$\left.\begin{aligned} H_P &= \tilde{\Delta}x_P/2\Delta\eta_{si} \ , \\ Q_P &= \int\int_{BOX} (N_A+n-p)\,dxdy - Q_{ss}\tilde{\Delta}x_P/2 \ , \end{aligned}\right\} \qquad (6.34)$$

which apply for any ψ lattice point in the row J = JMAX.

An identical procedure, applied to the rectangular region encompassing the point (I,JMAX+1) in Figure 6.6 shows the terms B_P and Q_P to be:

$$\left.\begin{aligned} B_P &= \tilde{\Delta}x_P/2\Delta\eta_{ox} \ , \\ Q_P &= -\left(\frac{\kappa_{si}}{\kappa_{ox}}\right)Q_{ss}\tilde{x}_P/2 \ , \end{aligned}\right\} \qquad (6.35)$$

which apply for any Ψ lattice point in the row J = JMAX+1.

As a check of these results, it is satisfying to find that if there <u>were</u> no interface (i.e., if $\kappa_{ox} \equiv \kappa_{si}$, and $Q_{ss} \equiv 0$), then (6.32) and (6.33) result in $\Delta\eta_{si} = \Delta\eta_{ox} = \Delta y(JMAX+1)/2$, so that both (6.34) and (6.35) reduce to the corresponding expressions for B_P, H_P, and Q_P given in (6.23), which apply for all rows of the ψ lattice except the two which bracket the oxide-silicon interface.

Mock did not use the above procedure for dealing with the interface. Instead, he defined a transformation of coordinate variable $y \rightarrow z$ in the oxide region, chosen so that $(\partial\psi/\partial z)_{ox} = (\partial\psi/\partial y)_{si}$ at the interface, even though $(\partial\psi/\partial y)$ is discontinuous there. (This is equivalent to defining a modified

oxide thickness.) While this approach possesses the advantage of requiring no special forms of the finite-difference equations at the interface, it precludes the inclusion of surface charge Q_{ss} in the model. Since we wish to include such, we choose not to follow Mock's approach, and implement the above described procedure instead.

3.01.2 Numerical Algorithm for the Charge Term Q_p

The charge term Q_p in (6.23) represents the net bulk charge enclosed by the rectangular region surrounding each ψ lattice point P. A simple approximation for Q_p is obtained from (6.23) by:

$$Q_p \approx (N_A+n-p)\tilde{\Delta}x_p\tilde{\Delta}y_p \ , \qquad (6.36)$$

which assumes that n and p can be regarded as constant throughout the rectangle $\tilde{\Delta}x_p\tilde{\Delta}y_p$. This assumption is a poor one, particularly in regions within the inversion layer, and adjacent to the source and drain "contacts." Examination of the numerical results obtained with the IBM program reveals that, in spite of the high density of lattice points in those regions (see Figure 6.2), the electron number density can change by an order of magnitude between neighboring points of the ψ lattice, particularly under conditions of strong inversion.* We seek, therefore, a better numerical algorithm for evaluating Q_p in regions where (6.36) is inadequate.

Semilogarithmic plots of n(x,y) versus x and y in the regions where n is found to vary rapidly reveal that its functional variation between adjacent lattice points is well represented by fitting an exponential function to the known values of n at those points. Since exponential functions are readily integrated analytically, this proves to be a convenient choice for accounting for the variation of n in such regions, to achieve a better estimate for Q_p.

---

*Similar steep gradients in the hole density are also found, but only in regions where $p << N_A$. Thus, the use of (6.36), insofar as the holes are concerned, causes negligible errors.

Figure 6.7 shows a ψ lattice point P and its four nearest neighbors. The figure defines a cartesian coordinate system (ξ,η) whose origin coincides with P. We subdivide the rectangular region centered on P into four quadrants, in each of which $n(\xi,\eta)$ is approximated by a function of the form:

$$n(\xi,\eta) = n_P e^{\alpha\xi} e^{\beta\eta} \quad . \tag{6.37}$$

The constants α and β appearing in (6.37) are then evaluated by equating $n(\xi,\eta)$ to the known values n_E, n_W, n_N and n_S at the four neighboring points, with ξ and η assigned the appropriate Δx and Δy values for those points. One obtains:

$$\left.\begin{aligned} \alpha_E &= \frac{1}{\Delta x_E} \ln(n_E/n_P) \quad , \\ \alpha_W &= \frac{-1}{\Delta x_W} \ln(n_W/n_P) \quad , \\ \beta_N &= \frac{1}{\Delta y_N} \ln(n_N/n_P) \quad , \\ \beta_S &= \frac{-1}{\Delta y_S} \ln(n_S/n_P) \quad . \end{aligned}\right\} \tag{6.38}$$

Given (6.37), with α and β appropriately defined for each of the four quadrants, the electron density is integrated analytically to find its contribution to the total charge Q_P in the rectangle. One obtains:

$$\int\int_{BOX} n(\xi,\eta)\,d\xi d\eta = n_P\left[(F_N+F_S)(F_E+F_W)\right] \quad , \tag{6.39}$$

with the functions F being defined by:

$$\left.\begin{aligned} F_N &= \left[\frac{\sqrt{n_N/n_P} - 1}{\beta_N}\right] \quad , \\ F_S &= -\left[\frac{\sqrt{n_S/n_P} - 1}{\beta_S}\right] \quad , \\ F_E &= \left[\frac{\sqrt{n_E/n_P} - 1}{\alpha_E}\right] \quad , \end{aligned}\right\} \tag{6.40}$$

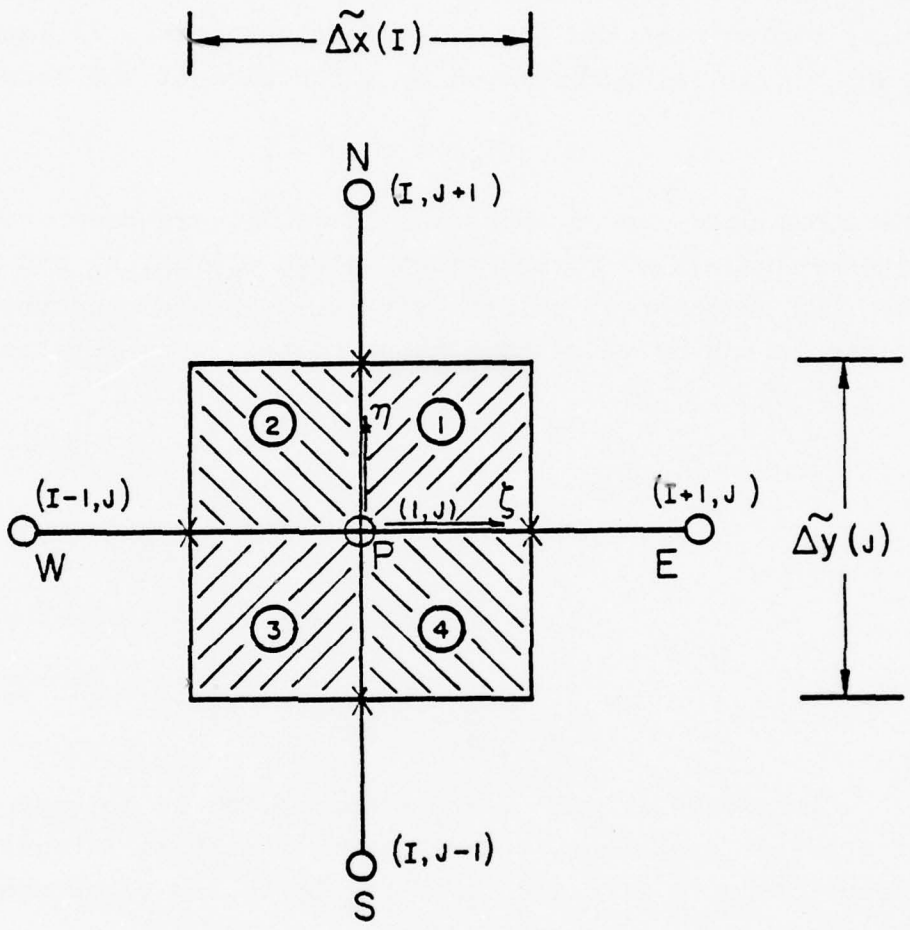


Figure 6.7 The charge term Q_P in the finite-difference equation for ψ is obtained by integrating the electron density $n(\xi,\eta)$ over the four quadrants shown.

$$F_W = -\left[\frac{\sqrt{n_W/n_P} - 1}{\alpha_W} \right] . \Bigg|$$

This procedure is followed only for the electron charge, and only in regions where the electron density assumes values in excess of 1% of the background acceptor density. In either event, (6.36) is utilized to find the contributions to Q_p due to the acceptor and hole densities.

In addition to Q_p, the Gummel algorithm to be discussed later requires the evaluation of the following integral:

$$\int\int_{BOX} n(\xi,\eta)\psi(\xi,\eta)\,d\xi d\eta \ . \tag{6.41}$$

For this purpose, we express $\psi(\xi,\eta)$ as:

$$\psi(\xi,\eta) = \psi_P + \left(\frac{\partial\psi}{\partial\xi}\right)\xi + \left(\frac{\partial\psi}{\partial\eta}\right)\eta \ , \tag{6.42}$$

where $\frac{\partial\psi}{\partial\xi}$ and $\frac{\partial\psi}{\partial\eta}$ are evaluated by simple differencing:

$$\left.\begin{aligned}
\left(\frac{\partial\psi}{\partial\eta}\right)_N &= \left[\frac{\psi_P - \psi_N}{\Delta y_N} \right] , \\
\left(\frac{\partial\psi}{\partial\eta}\right)_S &= \left[\frac{\psi_S - \psi_P}{\Delta y_S} \right] , \\
\left(\frac{\partial\psi}{\partial\xi}\right)_E &= \left[\frac{\psi_E - \psi_P}{\Delta x_E} \right] , \\
\left(\frac{\partial\psi}{\partial\xi}\right)_W &= \left[\frac{\psi_P - \psi_W}{\Delta x_W} \right] .
\end{aligned}\right\} \tag{6.43}$$

Then, using appropriate pairs of (6.43) in (6.42) for each of the four quadrants of Figure 6.7, together with (6.37), the desired integral is again obtained analytically.

The result of this integration is:

$$\iint_{BOX} n(\xi,\eta)\psi(\xi,\eta)\,d\xi d\eta = n_P \left\{ (F_E+F_W)(F_N+F_S)\psi_P \right.$$

$$+ G_E(F_N+F_S)\left[\frac{\psi_E-\psi_P}{\Delta x_E}\right] + G_W(F_N+F_S)\left[\frac{\psi_P-\psi_W}{\Delta x_W}\right]$$

$$\left. + G_N(F_E+F_W)\left[\frac{\psi_N-\psi_P}{\Delta y_N}\right] + G_S(F_E+F_W)\left[\frac{\psi_P-\psi_S}{\Delta y_S}\right] \right\}, \qquad (6.44)$$

where the functions G_N, G_S, G_E, and G_W are:

$$\left.\begin{aligned} G_N &= \left[\frac{\sqrt{n_N/n_P}\,[\ell n(n_N/n_P - 1) + 1}{\beta_N^2} \right], \\ G_S &= -\left[\frac{1 + \sqrt{n_S/n_P}\,[\ell n(n_S/n_P) - 1]}{\beta_S^2} \right], \\ G_E &= \left[\frac{\sqrt{n_E/n_P}\,[\ell n(n_E/n_P) + 1]}{\alpha_E^2} \right], \\ G_W &= -\left[\frac{1 + \sqrt{n_W/n_P}\,[\ell n(n_W/n_P) - 1]}{\alpha_W^2} \right]. \end{aligned}\right\} \qquad (6.45)$$

Again, this procedure for evaluating (6.42) is implemented only in regions where the electron density assumes values in excess of 1% of the background acceptor density.

Since this procedure for evaluating integrals of the electron density requires a significant increase in the number of numerical operations per lattice point where it is implemented, it is computationally less efficient than the simple algorithm for Q_p given by (6.36) and its counterpart for (6.39). For this reason, the computer program allows the user to circumvent the procedure, if he deems it unnecessary for his purposes, by means of an input data parameter.

3.02 Finite-Difference Algorithm for the Stream Function

The partial differential equation in θ, eq. (6.13), is repeated here for convenience of reference:

$$\frac{\partial}{\partial x}\left[\mu^{-1}e^{-\psi}\frac{\partial\theta}{\partial x}\right] + \frac{\partial}{\partial y}\left[\mu^{-1}e^{-\psi}\frac{\partial\theta}{\partial y}\right] = 0 \ . \qquad (6.13)$$

Our present version of the computer model does not incorporate provisions for field-dependent electron mobility. Therefore, μ^{-1} can be factored from (6.13), and ignored. If the field dependence of μ is to be retained, as a refinement to be later incorporated, only two subroutines in the computer program will require revision.

Figure 6.8 focuses upon a point (I,J) of the θ lattice, and its nearest neighbors. Simple differencing gives the partial derivatives of θ at the four points shown as crosses in the figure:

$$\left.\begin{aligned}
\left(\frac{\partial\theta}{\partial x}\right)_E &= \left[\frac{\theta_E - \theta_P}{\tilde{\Delta}x_E}\right] , \\
\left(\frac{\partial\theta}{\partial x}\right)_W &= \left[\frac{\theta_P - \theta_W}{\tilde{\Delta}x_W}\right] , \\
\left(\frac{\partial\theta}{\partial y}\right)_N &= \left[\frac{\theta_P - \theta_N}{\tilde{\Delta}y_N}\right] , \\
\left(\frac{\partial\theta}{\partial y}\right)_S &= \left[\frac{\theta_S - \theta_P}{\tilde{\Delta}y_S}\right] .
\end{aligned}\right\} \qquad (6.46)$$

The numerical value of $e^{-\psi}$ at each of those four points is determined in the manner next to be described.

Consider the four ψ lattice points located NE, SE, SW, and NW of the θ lattice point P. Assume a linear dependence of ψ along the borders of the rectangle defined by those four points, then evaluate an average value of $e^{-\psi}$ along each of those four borders, as did Mock [1], assigning the four averaged values to the respective points indicated by the crosses in the figure. (Mock describes this approach as a critical factor in achieving accuracy.) The result is:

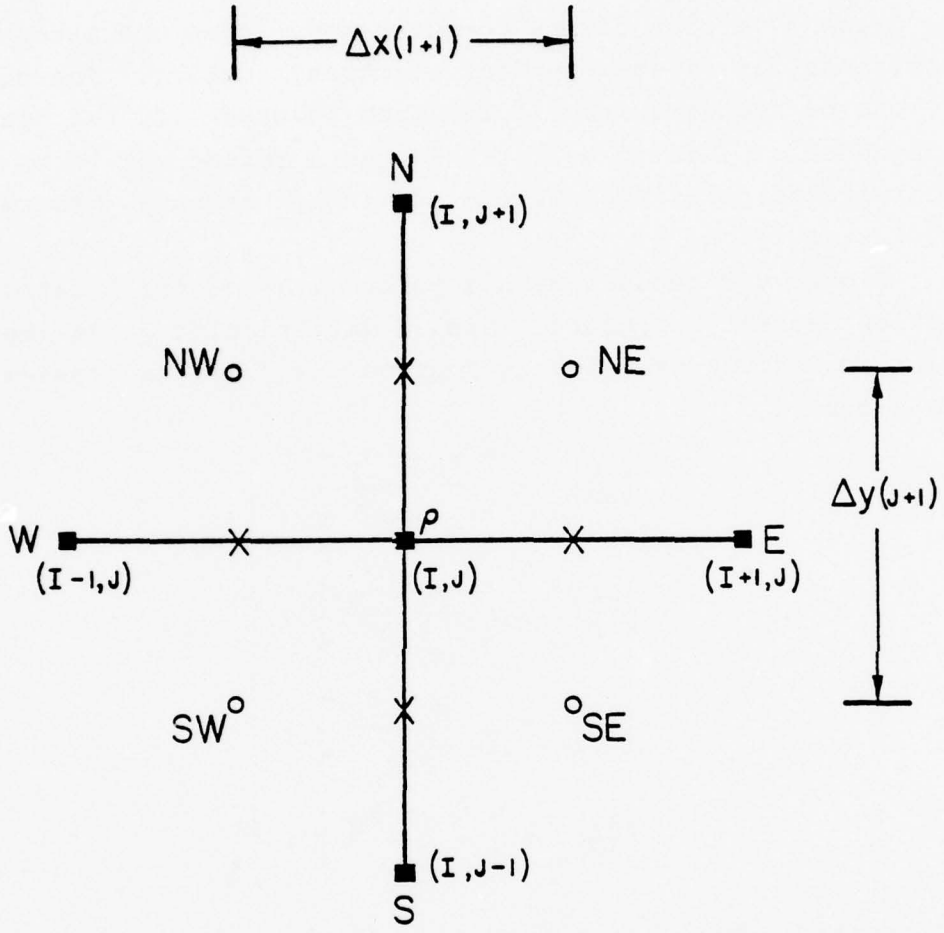


Figure 6.8 Five point "star" of θ-lattice points used in the finite-difference equation for θ.

$$\left.\begin{aligned}
\langle e^{-\psi}\rangle_E &= \left[\frac{e^{-\psi_{NE}} - e^{-\psi_{SE}}}{\psi_{SE} - \psi_{NE}}\right], \\
\langle e^{-\psi}\rangle_W &= \left[\frac{e^{-\psi_{NW}} - e^{-\psi_{SW}}}{\psi_{SW} - \psi_{NW}}\right], \\
\langle e^{-\psi}\rangle_N &= \left[\frac{e^{-\psi_{NE}} - e^{-\psi_{NW}}}{\psi_{NW} - \psi_{NE}}\right], \\
\langle e^{-\psi}\rangle_S &= \left[\frac{e^{-\psi_{SE}} - e^{-\psi_{SW}}}{\psi_{SW} - \psi_{SE}}\right].
\end{aligned}\right\} \qquad (6.47)$$

With the $e^{-\psi}$ values thus determined at the four points midway between the N, E, S, and W points of the θ lattice, the finite-difference form of (6.13) becomes (assuming constant mobility μ):

$$\Delta y_P\left[\langle e^{-\psi}\rangle_E\left(\frac{\theta_E - \theta_P}{\tilde{\Delta}x_E}\right) - \langle e^{-\psi}\rangle_W\left(\frac{\theta_P - \theta_W}{\tilde{\Delta}x_W}\right)\right]$$

$$+ \Delta x_P\left[\langle e^{-\psi}\rangle_S\left(\frac{\theta_S - \theta_P}{\tilde{\Delta}y_s}\right) - \langle e^{-\psi}\rangle_N\left(\frac{\theta_P - \theta_N}{\tilde{\Delta}y_N}\right)\right] = 0 , \qquad (6.48)$$

where Δx_p and Δy_p have been introduced as shorthand notation for $\Delta x(I+1)$, $\Delta y(J+1)$ in Figure 6.8. (See Figure 6.5 for the geometric definitions of the $\tilde{\Delta}x$ and $\tilde{\Delta}y$ factors in the above equation.)

As was the case with the finite-difference equation for ψ discussed earlier, the set of equations (6.48) can be represented in matrix form. The resulting sparse matrix contains only five nonzero symmetrically positioned diagonal elements. The matrix equation is solved for θ, given ψ, using the block-line iterative procedure described by Varga [8].

---

[8] R. S. Varga, Matrix Iterative Analysis, p. 194, Prentice-Hall, New York (1962).

This iterative method requires an initial "guess" for the stream function θ. With the aid of the IBM program discussed in the introductory section of this chapter, $\theta(x,y)$ was examined for several cases. It was found that θ ceases to vary with x and y, becoming essentially unity, at about the point where the electron density equals the intrinsic value n=1 in our normalized units. Furthermore, θ was observed to be predominantly a function of y, with x playing a minor role except in the immediate vicinity of the source and drain "contacts." Based on this, our initial "guess" for θ is quite simple. Choosing the plane midway between source and drain, we locate the particular y coordinate where $n \approx 1$. (The initial value of $n(x,y)$ depends upon the initial "guess" of the electrostatic potential ψ, to be described later.) A parabola is then fitted to match the conditions $\theta=1$, $\partial\theta/\partial y = 0$ at that point, and the boundary condition $\theta=0$ at y=0. The resultant parabolic fit at mid-channel is then applied at all other planes of constant x, resulting in a one-dimensional initial "guess."

The finite-difference equation for θ expressed by (6.48) is deceptively simple. Its implementation produces nontrivial challenges because of the exponential dependence upon ψ evident in (6.47). To illustrate, consider a modest drain voltage of 10 volts. In our normalized variables, this translates to a numerical value on the order of 400. But the IBM 370/165 used in this development produces overflow and underflow with the operation $e^{\pm|x|}$ when $|x|$ exceeds approximately 175. Not only is this limit obviously violated in regions near the drain, in the example cited, but along the entire channel as well when the gate voltage exceeds about 4.5 volts.

The modus operandi developed for dealing with this problem is the following. One <u>never</u> deals with positive x values in the operation e^{x}. One <u>always</u> alters the equation being treated so that only negative exponents are encountered. Furthermore, one endeavors <u>always</u> to reformulate that equation so that x represents <u>differences</u> in the potential between adjacent ψ

lattice points, rather than the magnitude of ψ itself. Finally, one always tests the magnitude of the negative exponent before attempting the operation of exponentiation, avoiding "underflow" messages by use of the approximation $e^{-x} \approx 0$ when x is found sufficiently large.

The details of the manner in which this procedure is implemented in the evaluation of θ will be described in the Users Manual to be published later. In brief, the average $e^{-\psi}$ functions in (6.47) are first reexpressed by factoring the exponential term whose exponent is least negative. This leaves within the brackets an exponential term whose negative exponent represents differences in ψ between adjacent lattice points. Following this, the four terms in (6.47) are again factored by the exponential term outside of the brackets whose exponent is least negative. Since the right hand side of (6.48) is zero, the latter factor can be cancelled from the equation. Again, this results in exponential terms whose negative exponents represent differences in ψ between adjacent lattice points. The subprogram which implements this procedure selects one of 16 possible cases in the process of determining which exponents are least negative.

3.02.1 Implementation of the Boundary Conditions on θ

The finite-difference equation in θ, equation (6.48), can be reexpressed to read:

$$B_P\theta_S + D_P\theta_W + E_P\theta_P + F_P\theta_E + H_P\theta_N = 0 \ , \quad (6.49)$$

which then takes on the same form as the finite-difference equation in ψ, equation (6.22). As was the case with ψ, the boundary conditions on θ shown in Figure 6.3(b) are established by means of appropriate modifications in the coefficients B_P, D_P, E_P, F_P, and H_P in (6.49).

3.02.1(a) Border Segments where θ=1--The fact that the θ lattice interlaces the ψ lattice (Figure 6.4) means that the boundary condition θ=1 cannot strictly be enforced at the borders where that condition applies (Figure 6.3). However, those borders

are remote from the region under the oxide-silicon interface where "the action" occurs, so this is not considered to be a serious drawback. What is done is to enforce θ=1 along the column (or row) of θ lattice points imagined to be present just outside of the rectangular region modeled. Thus, for example, at the left border in Figure 6.3(b) the coefficient D_P is calculated in the usual manner. Then the right hand side of (6.49) is set equal to the negative of D_P, which is equivalent to enforcing θ_W to be unity. Similar procedures are applied at the right border and along the lower border of the substrate.

3.02.1(b) Border Segments where $\partial\theta/\partial x = 0$--Again, we imagine two columns of the θ lattice just outside of the rectangular region modeled in Figure 6.4. At the left border, therefore, we wish to enforce $\theta_W = \theta_P$. This is achieved by calculating D_P and E_P in the usual manner, then summing them to obtain a revised value for E_P. The coefficient D_P in (6.49) is then set to zero. A similar procedure is applied at the right border.

3.02.1(c) The Oxide-Silicon Interface where θ=0--Here, the row J = JTHMAX of the θ lattice coincides with the interface. The condition θ=0 is enforced along that row by setting B_P, D_P, F_P, and H_P to zero, while setting E_P to unity in (6.49).

3.03 Finite-Difference Algorithm for the Current Scale Factor J_o

The row of the ψ lattice just below the oxide-silicon interface is selected for the line integral contained in (6.20), to evaluate the current scale factor J_o. The trapezoidal rule is applied to evaluate the integral. Simple differencing of θ on the interlaced θ lattice provides $\partial\theta/\partial y$ at the centroid of each interval Δx. The value of $e^{-\psi}$ at that centroid is again specified to be an average obtained by assuming a linear variation of ψ between adjacent ψ lattice points along the row:

$$\langle e^{-\psi} \rangle = \left[\frac{e^{-\psi_E} - e^{-\psi_W}}{\psi_W - \psi_E} \right] . \qquad (6.50)$$

As in the case of evaluating θ discussed earlier, special procedures are employed to avoid overflow and underflow due to machine limitations.

3.04 Finite-Difference Algorithm for the Electron Number Density n

Equation (6.10), when separated into its x and y components, gives:

$$\frac{\partial}{\partial x}\left(ne^{-\psi}\right) = \frac{J_x}{\mu} e^{-\psi} , \tag{6.51}$$

$$\frac{\partial}{\partial y}\left(ne^{-\psi}\right) = \frac{J_y}{\mu} e^{-\psi} . \tag{6.52}$$

Expressing (6.51) in finite-difference form:

$$\left[\frac{n(I+1,J)e^{-\psi(I+1,J)} - n(I,J)e^{-\psi(I,J)}}{x(I+1)}\right] = \frac{J_x}{\mu} \langle e^{-\psi(I,J)} \rangle . \tag{6.53}$$

(Here we resort to the use of the indices (I,J) identifying ψ lattice points because the "points of the compass" notation employed earlier becomes awkward.) Note that an average value for $e^{-\psi}$ is assigned to the point midway between the points I+1 and I of the row J. That is:

$$\langle e^{-\psi(I,J)} \rangle = \left[\frac{e^{-\psi(I+1,J)} - e^{-\psi(I,J)}}{\psi(I,J) - \psi(I+1,J)}\right] . \tag{6.54}$$

Equations (6.53) and (6.54), when combined, can be reexpressed so as to involve only differences in ψ:

$$n(I+1,J) = n(I,J)e^{-[\psi(I,J)-\psi(I+1,J)]} + \frac{J_x}{\mu}\left[\frac{1 - e^{-[\psi(I,J)-\psi(I+1,J)]}}{\psi(I,J) - \psi(I+1,J)}\right] \Delta x(I+1) . \tag{6.55}$$

Alternatively:

$$n(I,J) = n(I+1,J)e^{-[\psi(I+1,J)-\psi(I,J)]} - \frac{J_x}{\mu}\left[\frac{1 - e^{-[\psi(I+1,J)-\psi(I,J)]}}{\psi(I+1,J) - \psi(I,J)}\right] \Delta x(I+1) . \tag{6.56}$$

Equations (6.55) and (6.56), together with the vertical counterpart of (6.55) derived in the same manner from (6.52), form

the basis for the finite-difference algorithm for the electron density n, which proceeds as follows. Consider any row $J \geq$ JUNCTN (see Figure 6.4) which intersects the vertical source and drain "contacts" at the left and right borders where n is known. Starting at the left border, (6.55) is applied recursively to "march" along that row, toward the right border, until a potential minimum is encountered. When this occurs, then one switches to the right border and applies (6.56), recursively, to "march" to the left toward the last point at which n was previously calculated. In this manner, the electron mmeber density is filled in row by row, for all rows of the ψ lattice which intersect the source and drain "contacts." Note that the exponentials encountered with this procedure always involve negative exponents.

The next step in the procedure then utilizes the vertical counterpart of (6.55), derived from (6.52). In this case, one starts at the row J = JUNCTN, where n is now known, and "marches" vertically downward to the substrate contact, column by column. Again, only negative exponents are encountered in any exponentiation operations required.

To avoid underflow, appropriate approximations are involved in the subprogram which implements this procedure when the negative exponents encountered become sufficiently large. Similarly, appropriate approximations are applied when the denominators in (6.55), (6.56), and the vertical counterpart of (6.55) approach zero, to avoid overflow.

The "marching" algorithm described above is precisely the one suggested by Mock [1], who states that the procedure of switching directions at the point where ψ passes through a minimum suppresses the buildup of roundoff errors. In spite of this assertion, one intuitively feels uneasy about any numerical method which only makes use of information about the solution at one border without "pumping into" the solution some information about the boundary condition at the other border as well. The "marching" method fails to do so, and blithely assumes that, where the two solutions meet, continuity in n will not be violated.

The author felt sufficiently uneasy about this to try a different approach. Combining pairs of equations of the form of (6.55) by summing them, so as to couple the points I-1, I, and I+1 in a common row J, a tridiagonal matrix results from the set of equations relating n along that row. Given J_x and ψ as knowns, Gaussian elimination is then implemented to obtain n along such a row. This approach achieves the goal of "pumping into" the equations simultaneous information about the boundary conditions at both ends of the row in question. Interestingly, the two methods are found to give identical results (within the three significant figures printed out) in regions where the electron density is numerically important, i.e., in the inversion layer. Both approaches produce negative electron densities in regions where the numerical magnitude of n is negligible in comparison with the acceptor density N_A. (These nonphysical negative densities are believed due to the approximate forms of (6.55) and its vertical counterpart, and (6.56), discussed above to avoid underflow and overflow.) Insofar as these negative densities are concerned, the "marching" method is found superior. With it, they occur with values of n an order of magnitude lower than observed with the more sophisticated tridiagonal matrix approach.

4.0 The Iterative Procedure

The basic equations defined in Section 2, and their finite-difference counterparts developed in Section 3, constitute a set of coupled nonlinear equations for which a self-consistent solution is obtained by means of an iterative procedure. The one followed is basically the same as that adopted by Mock [1].

Given the geometric parameters of Figure 6.1(b), and given N_D, N_A, V_D, V_G, and V_{SUB}, the iterative procedure requires an initial "guess" of $\psi(I,J)$ and $\theta(I,J)$ at their respective lattice points. The initial "guess" employed for θ has been discussed in Section 3.02. We describe, next, that which has been developed for ψ.

The initial estimate for ψ utilizes the one-dimensional Kennedy-Motta theory described in Chapter I to provide an estimate for $\psi(x,0)$ along the oxide-silicon interface, which is then used to establish the boundary conditions at y=0 for a set of one-dimensional solutions of Poisson's equation (one per column, with y as the geometric variable). These give an estimate for $\psi(x,y)$ throughout the substrate region which fails to meet the boundary conditions along the source and drain "contacts." This defect is overcome by introducing matching space-charge sheaths along these "contacts," using abrupt junction depletion theory. The potential within the oxide region is assumed to vary linearly with y, connecting the known constant potential of the gate electrode with the estimate of the potential along the interface obtained from the 1-D Kennedy-Motta theory. In this manner, a reasonable first "guess" of $\psi(x,y)$ is achieved.

With ψ and θ thus "known," the corresponding estimates of the electron and hole densities n and p throughout the substrate region are next obtained. The evaluation of n utilizes the "marching" algorithm described in the preceding subsection. The evaluation of p utilizes the Boltzmann relation, equation (6.03), since the hole quasi-Fermi potential is known not to vary (equation 6.05)), as a consequence of the assumption of negligible hole current employed in our model.

The iterative process then proceeds as follows. Given n and p, the charge Q_p associated with each ψ lattice point P is then known (Subsection 3.01.2). Equation (6.22) is then solved, using Stone's method, to provide a new estimate for ψ. A new estimate for θ is then generated, followed by new estimates for n and p consistent with the new ψ and θ distributions. The procedure is then repeated until convergence is achieved, as determined by convergence criteria specified as input data by the user. Two differing procedures--PICARD iteration, and GUMMEL's algorithm--are employed in the implementation of this iterative sequence. These will next be discussed.

4.01 Picard Iteration

The Picard iteration procedure utilizes only "past history" to predict the next iterate for ψ, and is employed to achieve "coarse" convergence. Consider the m^{th} iteration step to have just been completed, so that $\psi^{(m)}$, $\theta^{(m)}$, $n^{(m)}$, and $p^{(m)}$ have been determined. (Values for step (m-1) are also available, as needed, except when m=0, which corresponds to the initial "guess.") Given these, (6.22) is solved, yielding a new solution for the potential distribution, which we designate $v^{(m+1)}$. With it, the improved approximation $\psi^{(m+1)}$ for the correct potential $\psi(x,y)$ is obtained from:

$$\psi^{(m+1)} = \psi^{(m)} + \alpha_m[v^{(m+1)}-\psi^{(m)}] + \beta_m[\psi^{(m)}-\psi^{(m-1)}] , \quad (6.57)$$

where α_m and β_m are underrelaxation parameters, adjusted for each iteration step m in the manner to be described next. With $\psi^{(m+1)}$ thus specified, the corresponding new estimates for $\theta^{(m+1)}$, $n^{(m+1)}$, and $p^{(m+1)}$ are then determined.

As was done by Mock [1], we adopt the Chebyscheff sequence of relaxation parameters α_m and β_m discussed by Forsythe and Wasow [9] for estimating appropriate values of those parameters at each iteration step m and, as did Mock, we reinitiate the sequence whenever the new residual $|v^{(m+1)}-\psi^{(m)}|_{max}$ turns out to be greater than the residual found in step m. (Mock interprets this as an indicator that the nonlinearity of the problem is becoming evident.) This sequence of relaxation parameters is given by, [9]:

$$\alpha_m = \frac{4\cosh(m\omega)}{(b-a)\cosh([m+1]\omega)} , \quad (6.58)$$

$$\beta_m = \frac{\cosh([m-1]\omega)}{\cosh([m+1]\omega)} , \quad (6.59)$$

where

$$\omega = \cosh^{-1}[\frac{b+a}{b-a}] . \quad (6.60)$$

---

[9] G. E. Forsythe and W. R. Wasow, Finite-Difference Equations for Partial Differential Equations, pp. 226-235. Wiley, New York (1960).

The parameters b and a in (6.60) are the upper and lower bounds, respectively, of the eigenvalues of the matrix [L] in equation (6.24). The initial values of this Chebyscheff sequence, as prescribed by Forsythe and Wasow, are determined by:

$$\alpha_o = 2/(b+a) \ , \tag{6.61}$$

$$\beta_o = 0 \ . \tag{6.62}$$

The implementation of this procedure requires an estimate for the upper and lower bounds b and a of the finite-difference Laplacian operator matrix [L]. The estimate for a is not critical, but that of b is. Too small an estimate for b leads to failure of the relaxation procedure to converge, while too large an estimate leads to extremely slow convergence. We have adopted the estimate for b suggested by Mock ([1], eq. (21))*:

$$b = \frac{Y_{sub}}{4X_{channel}} \iint (n+p)\,dxdy \ , \tag{6.63}$$

the double integral being evaluated over the entire rectangular substrate region by application of the trapezoidal rule in two dimensions. The parameter b is initially evaluated using the initial estimates for n and p described above. But, each time the Chebyscheff sequence is reinitiated due to the detection of a growing residual, the value of b is reassessed, using (6.63), on the basis that the estimates for n and p at that stage of the computation should be better than the previous estimates employed for that purpose.

Mock [1] points out that since large values of b lead to slow rates of convergence, and since its magnitude is proportional to the total number of free carriers in the substrate region, a savings in computation time can be achieved by applying the Picard iteration process, initially, with a re-

*Mock's eq. (21) contains a factor ε_s which is missing in (6.63). This is because of differences in Mock's choice of normalized variables and the choice which we have adopted.

duced donor density assigned to the source and drain, until coarse convergence criteria are satisfied. Having achieved this, he then assigns the correct donor density to those regions and switches to Gummel's iterative algorithm for the remainder of the computation, with tightened convergence criteria applied. Our model adopts the same procedure.

4.02 Gummel's Algorithm

Unlike the Picard iteration method just described, which uses only "past history," Gummel's algorithm [10] incorporates a new prediction for p and n at each iterative step. Because this prediction is based upon a linear approximation of the exponential functions appearing in the Boltzmann relations for those densities, it is efficient in terms of computation speed only when the estimates of ψ, p, and n are close to the "correct" values when it is initiated.

Let $\psi^{(m)}$, $n^{(m)}$ and $p^{(m)}$ be the values obtained at iteration step m. Let $\psi^{(m+1)} = \psi^{(m)} + \delta^{(m+1)}$, and assume that $\delta^{(m+1)} << \psi^{(m)}$. Then:

$$e^{\psi^{(m+1)}} \approx e^{\psi^{(m)}}[1+\delta^{(m+1)}] \ . \tag{6.64}$$

Applying this approximation to (6.02) and (6.03), one has

$$n^{(m+1)} \approx n^{(m)}[1 + \delta^{(m+1)}] \ , \tag{6.65}$$

$$p^{(m+1)} \approx p^{(m)}[1 - \delta^{(m+1)}] \ , \tag{6.66}$$

which assumes that the electron quasi-Fermi potential ϕ_n does not change appreciably between iteration steps (m) and (m+1). Assuming these conditions to be met, Poisson's equation, (6.01), gives for $\psi^{(m+1)}$:

$$\nabla^2\psi^{(m+1)} - [n^{(m)}+p^{(m)}]\psi^{(m+1)} = [N_A+n^{(m)}-p^{(m)}] - [n^{(m)}+p^{(m)}]\psi^{(m)} \ , \tag{6.67}$$

which is Gummel's algorithm.

---

[10] H. K. Gummel, IEEE Trans. Electron Devices ED-11, 455 (1964).

The implementation of (6.67) in finite-difference form requires the evaluation of the integrals:

$$\int\int_{BOX} [n^{(m)}+p^{(m)}]\psi^{(m+1)}\, dxdy\ , \tag{6.68}$$

$$\int\int_{BOX} [n^{(m)}+p^{(m)}]\psi^{(m)}\, dxdy\ , \tag{6.69}$$

in addition to the evaluation of the charge term Q_p appearing in equation (6.22). The evaluation of integrals of this form was discussed earlier in Section 3.01.2, leading to the derivation of equation (6.44) for the term in the integrands of (6.68) and (6.69) involving the electron density. (The term involving the hole density is approximated simply as $p^{(n)}\tilde{\Delta}x_p\tilde{\Delta}y_p$.) Insofar as the integral (6.68) is concerned, which appears on the left side of (6.67), inspection of (6.44) reveals that the effect of that term is to modify the coefficients B_p, D_p, E_p, F_p, and H_p in the finite-difference form of Laplaces equation, eq. (6.22). Insofar as the integral (6.69) is concerned, its effect is to modify the charge term Q_p in (6.22). With these modifications incorporated, the solution of the finite-difference equation for ψ using the Gummel algorithm uses the same subprogram which implements Stone's method for solving (6.22).

5.0 Status of the Model

The preceding sections have some details about our two-dimensional MOSFET model. More complete details, including flow charts, program listings, and instructions for using this computer model are contained in a supplementary report which has been devised to serve as a User's Manual.

Part II

Integrated Circuit Process Modeling

Stanford University
Stanford, California

Table of Contents

Page

Illustrations iii

Tables. vi

Introduction. 1

Chapter

1 Ion Implantation 4

1.1 Effects of Implantation Damage on Impurity Profiles in Annealed Si. 4

1.1.1 Objective 4

1.1.2 Introduction. 4

1.1.3 Discussion of Initial Conditions for the Calculations and the Experimental Data to be Predicted. 5

1.1.4 Development of Model. 8

1.1.5 Comparison of Calculated and Experimental Results 11

1.1.6 Discussion. 15

1.2 Calculation of Range Profiles and Recoil Implantation in Multi-Layered Media. 15

1.2.1 Introduction. 15

1.2.2 Experimental Measurements of Knock-On Atoms. 16

1.2.3 Application of the Boltzmann Transport Equation to the Calculation of Range Profiles and Recoil Implantation in Multi-Layered Media 23

2 Thermal Oxidation. 37

2.1 Introduction 37

2.2 General Theory 39

2.3 Silicon Crystal Orientation Effects. 41

2.4 Impurity Doping Effects. 43

2.5 Oxidation in HCL/O_2 Mixtures 47

Chapter		Page
3	Silicon Epitaxy	73
	3.1 Introduction	73
	3.2 Objective	74
	3.3 Transient Response to the Dopant Incorporation Mechanism	77
	3.4 Systems Analysis	82
	3.5 Verification of the Approach	84
	3.6 System Delay	86
	3.7 A Two-Block Model	89
Appendix A		93
Appendix B		96

Illustrations

Figure	Title	Page
1.1	As Implanted Boron Profile, calculated vacancy concentration profile and the annealed profile at 900°C for 35 min.	6
1.2	Electrical activity vs. time	8
1.3	10^{14} ions/cm^2 Dose Implant	12
1.4	10^{15} ions/cm^2 Dose Implant	13
1.5	Comparison of calculated electrical activity vs. time with experimental isothermal results by Seidel and MacRae.	14
1.6	Knock-on of oxygen during high energy implantation of heavy ions through an oxide mask	17
1.7	Curve showing integral density of oxygen in silicon wafers versus oxide thickness obtained from experiments for an implant dose of 5 x 10^{13} As/cm^{-2} at 100 KeV	18
1.8	Oxygen dose versus surface state density	20
1.9	Distribution of implanted As and recoil oxygen in SiO_2	22
1.10	Schematic showing the process by which the quantity F(E,x) evolves	26
1.11	Range profile for 100 kev As → SiO_2	30
1.12	Production of an energetic oxygen recoil	31
1.13	Oxygen recoil implantation	31
1.14	Oxygen recoil implantation	32
2.1	Oxide thickness vs. time for <111> and <100> silicon	57
2.2	Effect of redistribution on impurity doping profiles	58
2.3	Surface concentrations on diffused heavily doped samples	59
2.4	Oxide thickness vs. doping	60
2.5	Oxide thickness vs. time for various doping levels at 900°C.	61
2.6	Oxide thickness vs. time for various doping levels at 1000°C	62

Figure	Title	Page
2.7	Theoretical curves of oxide thickness vs. time for various linear rate constants.	63
2.8	Linear and parabolic rate constants vs. chemical dopant concentrations at 900°C	64
2.9	Oxide thickness vs. oxidation time for the oxidation of <111> and <100> oriented N-type silicon in various O_2/HCL mixtures at 900°C	65
2.10	Oxide thickness vs. oxidation time for the oxidation of <111> and <100> oriented N-type silicon in various O_2/HCL mixtures at 1000°C.	66
2.11	Oxide thickness vs. oxidation time for the oxidation of <111> and <100> oriented N-type silicon in various O_2/HCL mixtures at 1100°C.	67
2.12	Parabolic rate constant vs. % HCL for <111> and <100> oriented N-type silicon at 900°, 1000°, and 1100°C	68
2.13	Linear rate vs. % HCL for <111> and <100> oriented N-type silicon at 900°, 1000°, and 1100°C.	69
2.14	Arrhenius plots of the parabolic rate constants for silicon oxidized in various O_2/HCL mixtures.	70
2.15	Arrhenius plots of the linear rate constants for <111> and <100> oriented silicon oxidized in various O_2/HCL mixtures.	71
2.16	Typical output generated by moving boundary diffusion program.	72
3.1	Horizontal epitaxial reactor schematic	75
3.2	(a) Cross-sectional view of the reaction chamber (b) Block diagram of the reactor as a system (c) One block transfer function relating the dopant gas flow input f(t) to dopant concentration in the epitaxial film N(x).	76 76 76
3.3	Experimentally observed dopant concentration as a function of distance from the surface of the epitaxial film for a decreasing step change in the dopant gas flow.	80
3.4	Experimental dopant profile for an increasing step change in dopant gas flow and exponential fit to the experimental profile.	81
3.5	Response of the reactor to pulse inputs in the dopant gas flow	85

Figure	Title	Page
3.6	Response of the reactor to a ramp followed by a decreasing step input in the dopant gas flow	87
3.7	Reactor response to a sequency of two pulses, each 0.8 min. wide.	88
3.8	Concept of the two block model of the epitaxial reactor.	90
3.9	Comparison of the one block model (exponential fit) and the two block model for an increasing step change in dopant gas flow.	91

Tables

Table	Title	Page
1.1	Parameters for use in the cross section expression	28
1.2	Comparison between theory and experiment for boron range profile discontinuity.	34
2.1	Rate constants for <111> and <100> silicon	56

COMPUTER-AIDED
SEMICONDUCTOR PROCESS MODELING

1. Introduction

In military, industrial, commercial, and consumer applications, frequently there is a great necessity for "customizing" the design of an integrated circuit to fulfill the critical needs of a specific system or class of systems. A major barrier which prevents the economic production of small quantities of high performance custom integrated circuits is the cost of design. That is, the initial cost of designing optimum fabrication processes, device structures, and circuit configurations is prohibitively large because of the amount of empirical human effort which must be invested. The root cause of this problem is a glaring lack of adequate process, device, and circuit models and accompanying computer aided design techniques to ease the burden of custom design. Perhaps the most serious deficiency among these is the unavailability of suitable models for predicting accurately the physical characteristics of a monolithic structure on the basis of the control parameters for the corresponding fabrication processes. The salient objective of this program is the development of new basic models for integrated circuit processes which will permit accurate prediction of the characteristics of a monolithic structure on the basis of its proposed process parameters. These models will serve as the basis for economic computer aided design of optimum fabrication processes for custom integrated circuits.

The four key generic integrated circuit fabrication processes which are being investigated are (1) ion implantation, (2) thermal oxidation and chemical vapor desposition, (3) epitaxy, and (4) thermal diffusion. This report describes the progress which has been made in the first 15 months of this program in all four areas.

In ion implantation, studies have been done in two areas. Effects of implantation damage on impurity profiles is annealed Si have been investigated. A three-stream diffusion model for boron in silicon involving boron, vacancies and boron vacancy complexes has been proposed. The model is capable of predicting ordinary diffusion, proton enhanced diffusion and the annealing behavior of room-temperature implanted boron when appropriate restrictions on dose and annealing temperature are obeyed. In the second study a straight forward approach to calculations of range profiles of dopant ions in multilayer

media (e.g. arsenic implementation in silicon through a thin layer of silicon dioxide) has been developed based on numerical integration of the linearized Boltzmann transport equation. Using this approach, we have also calculated the range profiles of the recoiling ions which are knocked on from one layer of the substrate into the other layer by the projectiles (e.g. oxygen recoiling from SiO_2 into Si, by ions).

In the area of thermal oxidation and chemical vapor deposition the principal goal of the first year of the program has been to achieve accurate analytic prediction of oxide thickness for an arbitrary sequence of oxidations. At the present time, such prediction is possible only for lightly doped, (111) oriented silicon, with no chlorine species present during the oxidation process.

Toward this, kinetic oxidation data have been gathered under the following conditions:

(a) Data for (111) and (100) silicon over the temperature range 700°C to 1200°C.

(b) Data for dry O_2/HCL oxidation ambients containing 0-10% HCL for temperatures between 900°C and 1100°C.

(c) Data for heavily phosphorous doped (111) substrates for surface concentrations up to solid solubility and temperatures between 900°C and 1100°C.

The availability of the above data now makes it possible to predict analytically the oxide thickness under a very wide range of ambient conditions and substrate orientation and doping levels.

In epitaxy, preliminary effort has been directed toward understanding the kinetics of growth and the mechanism of dopant incorporation in the epitaxial layer.

The dopant system of a horizontal silicon epitaxial reactor has been characterized by changing the dopant gas flow during the continuous deposition of epitaxial layers from silane. A System "transfer function" has been found relating the dopant profile in the epitaxial layer (system output) to the time-varying dopant gas flow (system input). The transfer function allows the calculation of the dopant profile in the epitaxial film for any time varying dopant gas flow. The calculation of the dopant gas flow, as a function of the time required to achieve a desired dopant profile, is thus possible.

In thermal diffusion, a new mathematical model is being developed for the diffusion of impurities into silicon for both a constant source and drivein diffusion process. The model considers the influence of the internal electric field on the motion of impurity ions at elevated temperatures. The electric field is due to charge density produced by the ionization of impurities present in the material. Appendix B reviews the results of this study.

CHAPTER I

Ion Implantation

J. R. Gibbons, A. Chu, D. Smith and K. C. Saraswat

1.1 Effects of Implantation Damage on Impurity Profiles in Annealed Si

1.1.1 Objective

A major unsolved problem in the field of ion implantation is the general question of how the annealing of implantation damage in a crystal affects the diffusion of the implanted species. Briefly, the problem is that implanted impurity ions create damage in the semiconductor lattice as they come to rest. Annealing of this damage produces a rich source of vacancies and Si interstitial ions, both of which may produce orders of magnitude more diffusion of the implanted impurities during the annealing cycle than one would estimate from a simple calculation based on the impurity diffusion coefficient at the annealing temperature.

The objective, therefore, is to study the annealing of implanted profiles in common semiconductor materials. The immediate emphasis is placed on dopants such as boron, phosphorus, and arsenic in silicon substrates. During the first fifteen months of the program a model has been developed to predict the diffusive-annealing characteristics of boron in silicon. The model is capable of predicting ordinary diffusion, proton enhanced diffusion and the annealing behavior of room-temperature implanted boron when appropriate restrictions on dose and annealing temperature are obeyed. Parameters required in the model were selected so that the model would predict correctly the impurity profiles that are obtained under conventional thermal diffusion conditions. The same parameter set is then used in the annealing calculation of implanted boron.

The computed profiles and electrical activities, as a function of time, compare very well with experimental data published in the literature. When the implantation dose exceeds 10^{15} ions/cm^2 and/or the annealing temperature is below 900°C, precipitation effects may occur. We outline a systematic approach to incorporate these effects into the basic three-stream model.

1.1.2 Introduction

Recently Anderson and Gibbons have proposed a two-stream diffusion model for the ordinary diffusion of boron [1] and proton enhanced diffusion [2] of boron in silicon. This model appears to contain the key attributes for explaining the annealing behavior of ion implanted boron in silicon. For instance, activation of boron during annealing requires at least two

forms of boron, electrically active and inactive; and the damage enhanced diffusion during initial stages of annealing is qualitatively similar to the enhanced diffusion produced by protons.

It is the purpose of this work to show that by extending the two-stream diffusion model to include POSITIVELY CHARGED VACANCIES, we can calculate the annealed profiles and the activation of boron in silicon. The resulting three-stream diffusion model will predict ordinary diffusion, proton enhanced diffusion and the annealing behavior of boron that is ion implanted into silicon at room temperature and subsequently annealed.

Naturally, a full description of boron annealing is a complicated matter and the model presented here deals with only a part of the overall picture. In particular we will concentrate on ion doses in the range below 10^{15} ions/cm^2 and annealing temperatures of 900°-1000°C. The reasons for these limitations will be apparent from a consideration of Table 1.1, where we have abstracted some of the global features of the data presented by Hofker et. al. [3]. These authors have measured boron carrier concentration profiles and the total boron profiles under a number of dose and annealing conditions of interest to use. As indicated in the table, their dose range extends from 10^{14} ions/cm^2 to 10^{16} ions/cm^2 and their annealing temperature range from 800°C to 1000°C. As we will show, the three-stream diffusion model to be developed here is capable of predicting the annealing behavior in the lower left corner of Table 1.1; i.e. for doses of 10^{14} ions/cm^2 to 10^{15} ions/cm^2 and anneal temperatures of 900°C to 1000°C. The adjacent cases require extensions of the model that are straightforward but not of interest to our present purpose. For instance, in the 10^{16} ions/cm^2 dose cases, the boron peak concentration

ANNEAL TEMP °C	Dose, ions/cm^2 10^{14}	10^{15}	10^{16}
800°C	Boron Near ~25% Activity	Absorption Profile Peak	Boron <16%
900°C	3 STREAM DIFFUSION MODEL APPLIES		Precipitates when 40%
1000°C	(B^-, V^+, B^-V^+) 100% Activity		$C_B > C_{SS}$ 65%

Table 1.1 Isochronal Annealing Results.
Annealing Time = 35 min.

exceeds solid solubility. Therefore there will be precipitation of boron. In the lower dose cases, there is a fraction of inactive boron near the peak that is similar to a precipitate, when the anneal is carried out at 800°C. However, the boron concentration is below the solid solubility limit for these cases, so this form of "precipitate" is probably boron trapped by dislocation networks that do no fully anneal at 800°C in 35 minutes. For these cases we will have to include a new species and pertinent reaction kinetics.

1.1.3 Discussion of Initial Conditions for the Calculations and the Experimental Data to be Predicted

Our purpose in formulating a model is, of course, to develop a theoretical basis from which impurity profiles and electrical activities can be calculated. It is useful to begin with a brief summary of the initial conditions for the calculation and the experimental data we wish the theory to predict.

The as-implanted profile as determined by Secondary Ion Mass Spectrometry is shown in Figure 1.1. The vacancy concentration profile is calculated

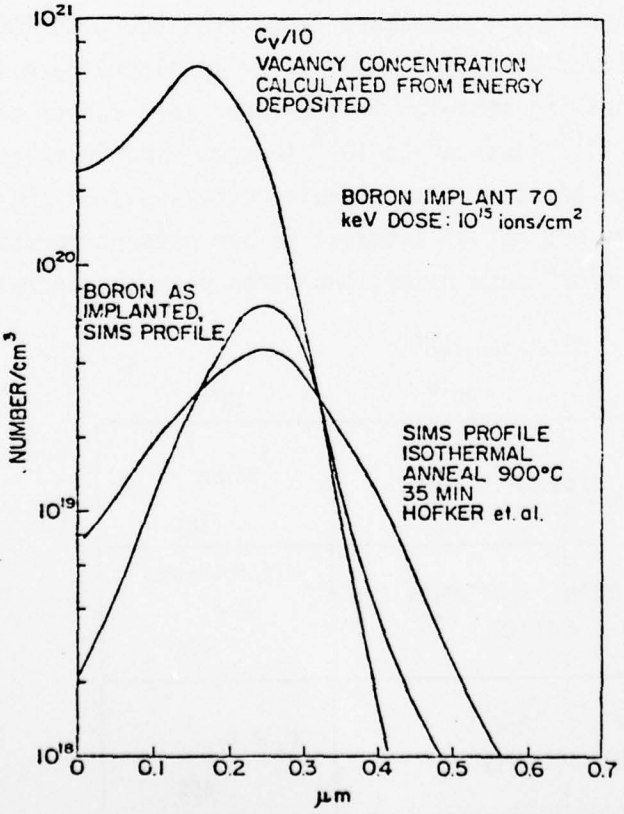


Fig.1.1 As implanted Boron Profile, calculated vacancy concentration profile and the annealed profile at 900°C for 35 min.

from the Frenkel-pairs produced by the 70 Kev incident boron ions as they deposit energy into atomic processes [4] along their trajectories. In what follows we will make the conservative assumption that, as the boron atoms come to rest, they will all combine with nearby vacancies to form boron-vacancy (BV) pairs, which are electrically inactive. Hence, at the outset of the anneal the electrical activity is zero. Although a non-zero activity is more reasonable, it will also require an initial profile for the electrically active boron. At present, until we obtain short-annealing-time electrical carrier profiles, we do not wish to introduce this unknown. We further assume that the remaining high vacancy concentration will condense into clusters and dislocation loops. Room temperature annealing will occur during the implantation, in agreement with backscattering experiments [5], and what remains will anneal very quickly at temperatures of 900°C and higher, as indicated by our unpublished T.E.M. work [6]. Therefore we will use thermal equilibrium concentration of vacancies as an initial condition in our analysis. This assumption will surely be inadequate for a liquid nitrogen implant or for the case of a heavy ion or high dose implant that produces an amorphous damage layer. For the cases to be dealt with, however, the assumption of a thermal equilibrium concentration of vacancies seems plausible.

Our model will describe the rapid diffusion and conversion of Boron Vacancy pairs into electrically active Substitutional Boron, producing at given times profiles and electrical activities that we can use to compare with experimental results. Hofker et. al. [3] have measured electrical carrier and total boron profiles after 35 minutes of annealing at various temperatures and implanted doses. Seidel and MacRae [7] have performed isothermal annealing experiments in which a $1.5\ 10^{15}$ ions/cm^2 dose, room temperature implant was annealed at three temperatures. They reduced all their electrical activation data to an anneal temperature of 855°C by shifting time with an exponential factor with an activation energy of 5 eV. We have scaled their 855°C data to 900°C using the same technique. The result is shown in Figure 1.2.

We assume that a realistic annealing model must provide not only reasonable prediction of the annealed boron impurity profile but also a reasonable prediction of the annealed boron impurity profile but also a reasonable prediction of the Seidel-MacRae annealing curve.

1.1.4 Development of the Model

The three species in this model are Substitutional (electrically active Boron, Boron-Vacancy pairs (electrically inactive), and Positively Charged Vacancies. Electrically-active boron is assumed to diffuse substitutionally by means of random encounters with neutral vacancies. The boron-vacancy (BV) pair is assumed to diffuse much more rapidly. The diffusion of boron then becomes a weighted diffusion with contributions from a slow, electrically active fraction (substitutional boron) and a fast, electrically inactive fraction (the boron-vacancy pair). Hence a large fraction of boron-vacancy pairs results in low electrical activity and fast diffusion, and conversely a large fraction of substitutional boron results in high electrical activity and slow diffusion. Annealing will proceed with the conversion of a large population of BV-pairs into substitutional boron.

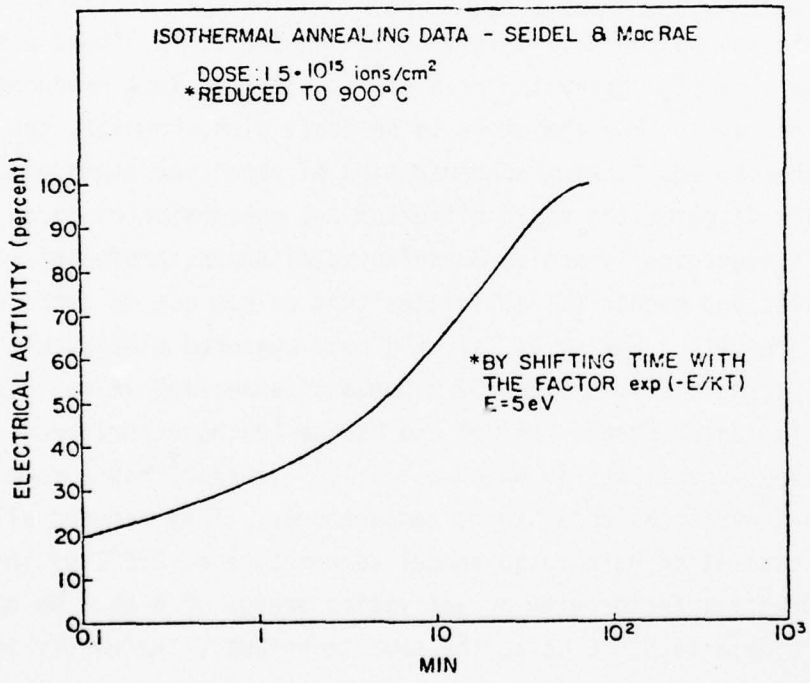


Fig.1.2 Electrical Activity vs. time. Scaledexperimental isothermal annealing data at 900°C by Seidel and MacRae.

The model consists of three diffusion equations (eq.1.7) modified by terms representing the boron-boron vacancy pair reaction; in the positive vacancy equation an additional term is introduced to represent the tendency for positive vacancies to reach the thermal equilibrium concentration in

the silicon crystal. Next we will describe the features of this model and the methods used to select the various constants required in the equations.

A) Diffusion Coefficients - Only Functions of Temperature.

We assume that substitutional boron diffuses by the same mechanism as the self diffusion of silicon. Therefore we use the self diffusion coefficient of silicon for D_B. At 900°C this value is approximately 1.6 10^{-21} cm^2/sec. For positive vacancies we use the diffusion coefficient from Seidel and MacRae [7], which at 900°C is D_v= 4.7 10^{7} cm^2/sec. Finally the diffusion coefficient for BV-pairs is estimated from measurements of enhanced diffusion by Hofker et.al. [3] and Anderson and Gibbons [2]. Hofker and coworkers measure the enhanced diffusion coefficient after 35 minutes of annealing at 900°C to be 1.4 to 1.7 10^{-14} cm^2/sec. Since this is an average value it provides only lower limit for the value of D_{BV}. As an alternative, Anderson and Gibbons estimated the value of D_{BV} to be 5.7 x 10^{-14} cm^2/sec at 750°C from proton-enhanced experiments. We use for the present work D_{BV} = 7 10^{-14} cm^2/sec.

$$\frac{\partial C_B}{\partial t} = D_B \frac{\partial^2 C_B}{\partial x^2} + \frac{C_{Bv} - ko\, C_B\, C_{v+}}{\tau} \qquad 1.1\ (a)$$

$$\frac{\partial C_{v+}}{\partial t} = D_{v+} \frac{\partial^2 C_{v+}}{\partial x^2} + \frac{C_{Bv} - ko\, C_B\, C_{v+}}{\tau} - \frac{C_{v+} - C_{v+eq}}{\tau_v} \qquad 1.1\ (b)$$

$$\frac{\partial C_{Bv}}{\partial t} = D_{BV} \frac{\partial^2 C_{Bv}}{\partial x^2} - \frac{C_{Bv} - ko\, C_B\, C_{v+}}{\tau} \qquad 1.1\ (c)$$

B) The Thermodynamic Reaction. Boron and positive vacancies react to form BV-pairs according to the chemical equation 1.2.

$$B + V^+ \quad BV \qquad 1.2$$

At thermal equilibrium the concentrations of reactants and product are related by the thermodynamic equation 1.3.

$$C_{BV} = k_o\, C_B\, C_V \qquad 1.3$$

The generation or disappearance of reactants or product that result from the reaction is modeled by first order kinetics with the departure from equilibrium as the driving force and a time constant.

The equilibrium constant for the reaction k_o is calculated from equation 1.3 as follows. Under equilibrium conditions, such as in an ordinary diffusion, the electrical activity is near 100%. Assuming 98% activity and rearranging equation 1.3 we obtain an expression for k_o

$$k_o = (C_{BV}/C_B)\ (1/C_V) \qquad 1.4$$

Estimating $C_{vo}(T)$ from Seidel and MacRae [8] to be $1.25\ 10^{11}/cm^3$ and substituting in equation 1.4 we find $k_o = 1.63\ 10^{13}\ cm^3$.

C) Equilibrium Concentration of Positively Charged Vacancies.

The tendency to reach equilibrium is again modelled by first order kinetics and a time constant. Equation 1.5 gives the dependence of the concentration of positive vacancies on the energy level E_V +, the Fermi level E_F and the concentration of Neutral vacancies $C_{vo}(T)$, which is a function of temperature only.

$$C_{V+eg} = C_{vo}(T)\ \exp\ \frac{E_{V+} - E_F}{KT} \qquad 1.5$$

The Fermi level is calculated using the Maxwell-Boltzmann approximation instead of the Fermi-integral of order 1/2. Hence doping degeneracy effects are not thus far accounted for.

To obtain values for and this set of space equations is solved by numerical analysis techniques with boundary conditions for an ordinary thermal diffusion. The numerical analysis techniques were discussed in great detail in a previous report of this program [9]. Values for and are then chosen to fit experimental ordinary diffusion profiles. The values of and are 450 sec and 10^{-4} respectively. It is worthwhile to emphasize that the apparent concentration-dependent diffusion coefficient of boron is achieved in this model through the Fermi level. A high substitutional boron concentration shifts the Fermi level closer to the valence band edge, thus increasing the concentration of positive vacancies (E_V = 0.35 ev above the valence band). This in turn increases the ratio of BV-pairs to substitutional boron, thus increasing the overall diffusion of boron.

D) The Relation Between Diffusion and Electrical Activity.

The two forms of boron in this model have dichotomous attributes. Namely substitutional boron is electrically active and slow diffusing; on the other

hand the boron-vacancy pair is electrically inactive and fast diffusing. Because of this dichotomy and because the diffusion of the total boron is a weighted diffusion of these two species, a relation between diffusion and electrical activity ought to exist. We assume near 100% electrical activity in the selection of parameters to fit thermal diffusion experiments. This ensures that when the abnormally high concentration of boron vacancy pairs that is produced by the implantation has annealed, the enhanced diffusion (associated with the initially low electrical act) will relax to ordinary diffusion which is simply the near-thermal-equilibrium solution of the set of equations. The analytical analogue of this discussion is carried out in Appendix A. The boron and BV pair equations are combined and the predominance of D_{BV} over D_B is used to arrive at a single approximated diffusion equation for total boron. In this equation we can identify the overall diffusion coefficient D_{exp} with a product of D_{BV} and the fractional concentration of BV pairs. Furthermore the fractional concentration can be expressed in terms of electrical activity. Hence we arrive at a diffusion coefficient (the only one accessible in an actual experiment, hence the subindex), which is a function of the electrical activity.

$$D_{exp} = D_{BV} \frac{C_{BV}}{C_{BV} + C_B} = D_{BV} \text{ (1 - elec. actv.)} \qquad 1.6$$

At the outset of the anneal the assumed electrical activity is zero, the value of D_{exp} is then initially equal to D_{BV}. During anneal, as the electrical activity increases D_{exp} will decrease. Finally when the electrical activity approaches 98%, D_{exp} will approach the value of $1.4\ 10^{-15}$ cm^2/sec, the ordinary diffusion coefficient of boron at 900°C measured in thermal diffusion experiments.

1.1.5 Comparison of Calculated and Experimental Results

With the choice of parameters described in the previous section, equation 1.1 can be used to calculate the diffusive redistribution of boron under ordinary conditions of thermal diffusion. These parameters, together with the initial conditions just discussed, then permit us to solve for the annealing behavior of ion implanted boron. The results of such a calculation for 10^{14} ions/cm^2 dose implant is shown in Figure 1.3. The as-implanted profile-determined experimentally by Hofker et. al. [3] is represented by the dotted line. Upon annealing, the BV-pairs in this profile diffuse and convert into substitutional boron. Hence a sequence of electrically active boron profiles will develop

with time. Figure 1.3 shows calculated profiles under 1, 3, 10 and 35 minutes of annealing with corresponding electrical activities of 10, 28, 65 and 94 percent. The total boron concentration at 35 minutes is represented by triangles; the fit to the solid line representing the experimental S.I.M.S. profile is excellent. Since the ordinary diffusion of boron at 900°C for 35 minutes would only modify the as-implanted profile very slightly, the fit between the calculated results and the deeper experimental profile is indicative of the existence of enhanced diffusion which progressively diminishes to the ordinary diffusion rate as the electrical activity approaches 100 percent. The calculated results are thus in good agreement with the experimental observations of Hofker and coworkers [3].

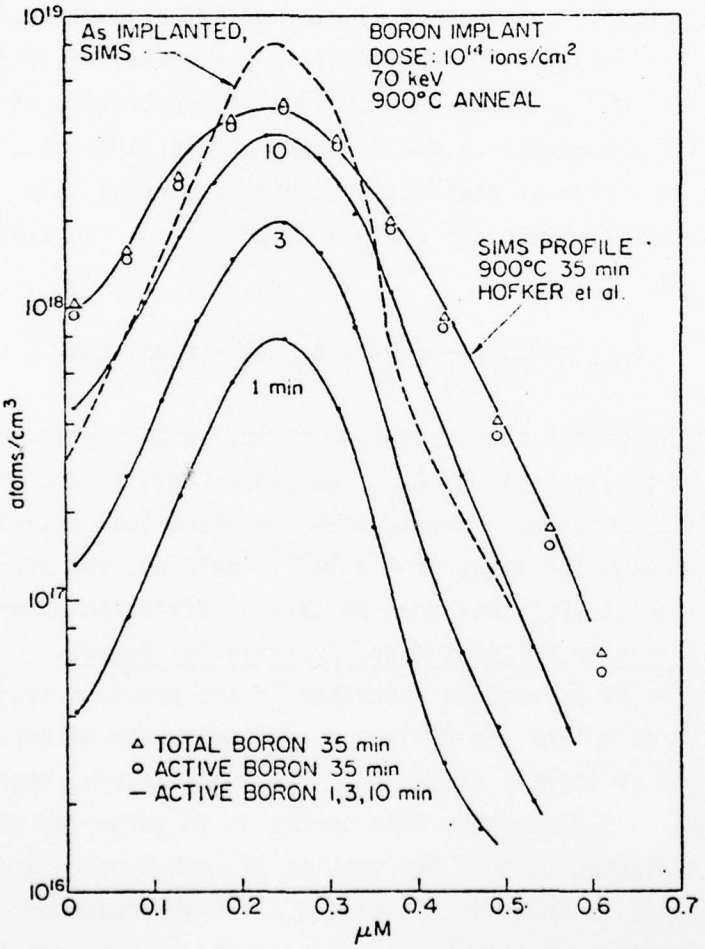


Fig. 1.3 10^{14} ions/cm^2 Dose Implant - Evolution of calculated active boron profiles at 1, 3, 10, 35 min. in a 900°C isothermal-anneal and comparison of the calculated Total Boron concentration Profile with the experimental S.I.M.S. profile by Hofker at 35 min.

Figure 1.4 shows the results of a calculation performed with the same set of parameters for a 10^{15} ions/cm^2 dose; which produces an order-of-magnitude increase in the impurity and damage concentrations. The results depicted are analogous to the previous example except for some loss in the quality of the fit; the calculated total boron concentration is again represented by triangles.

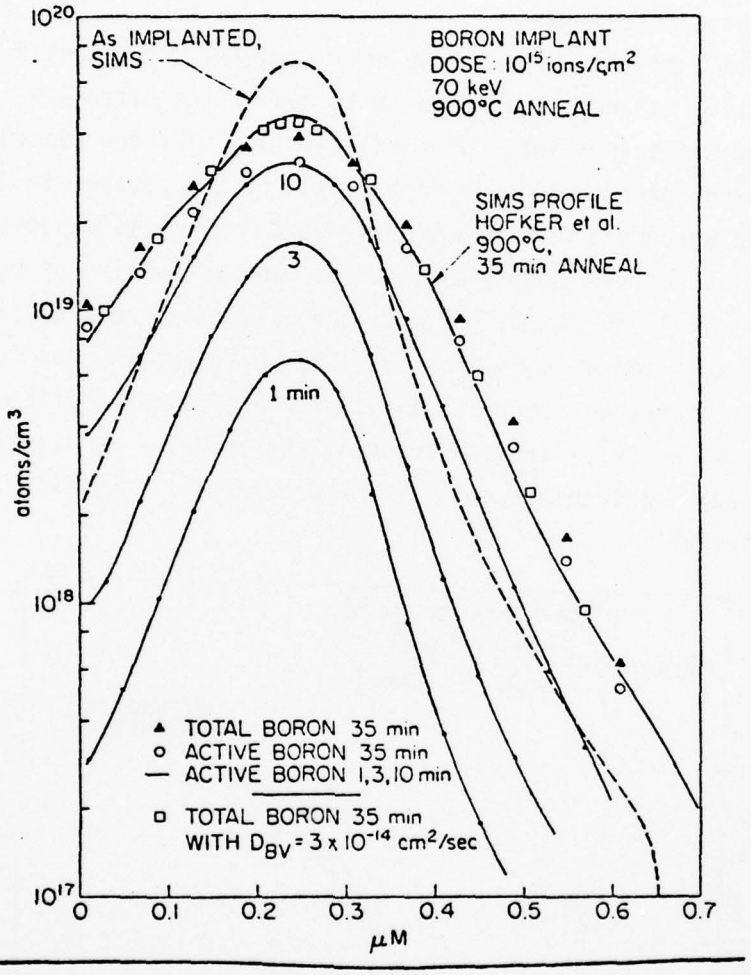


Fig. 1.4 10^{15} ions/cm^2 Dose Implant - Evolution of calculated active boron profiles at 1, 3, 10, 35 min. in a 900°C isothermal-anneal and comparison of the calculated Total Boron concentration Profile with the experimental S.I.M.S. Profile by Hofker at 35 min.

We attribute this difference to the doping degeneracy effects that were not correctly included in our Fermi-level calculation. The Fermi level is calculated using the Maxwell-Boltzmann approximation to the Fermi integral of order 1/2. Consequently the Fermi level was much too close to the valence band and the concentration of positive vacancies was too high. This in turn caused the relative concentration of BV-pairs to be high and finally the overall effect of excessive diffusion.

To verify this assumption, we have estimated the error factor in the concentration of positive vacancies and used it to reduce the diffusion coefficient of boron vacancy pairs in a subsequent calculation. The new calculated 35 minute total boron concentration is represented by the squares in Figure 1.4. As expected, the agreement with the experimental profile is improved.

The comparison of the experimental electrical activation of boron versus time by Seidel and MacRae [6] with the calculated results from the previous examples is shown in Figure 1.5. The agreement is good for long times. For short times we attribute the difference to our initial condition of zero electrical activity at the outset of the anneal. As discussed previously, we expect improvements with a non-zero initial condition for the electrical activity.

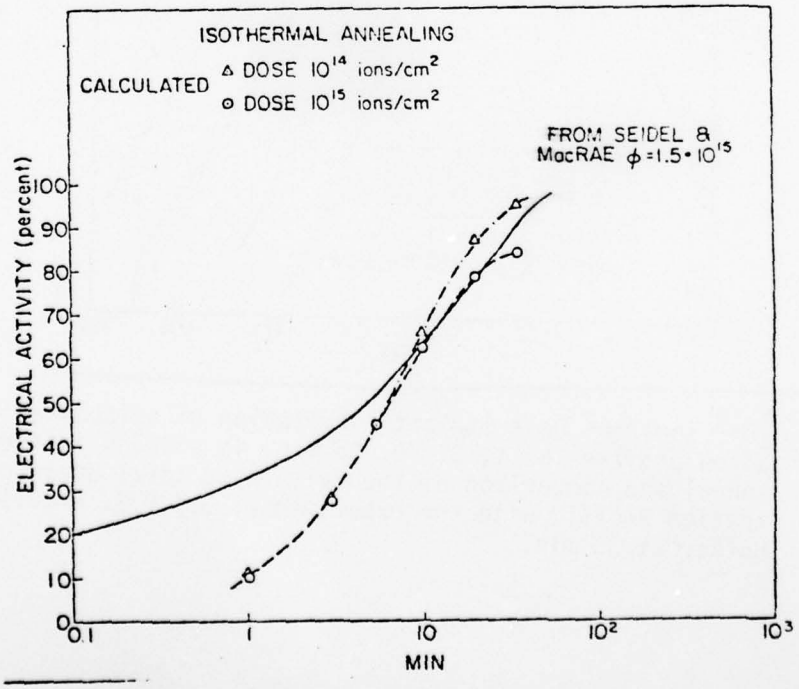


Fig. 1.5 Comparison of calculated electrical activity vs time with experimental isothermal results by Seidel and MacRae.

1.1.6 Discussion

The three stream diffusion model is capable of predicting boron impurity profiles under conditions of ordinary thermal diffusion, proton enhanced diffusion and radiation enhanced diffusion of the sort that occurs when boron is implanted at room temperature and subsequently annealed. In this latter case the model is valid when the dose and energy of the boron implant are such that

(a) the solid solubility limit for B in Si is not exceeded at the annealing temperature.

(b) an amorphous damage layer is not produced by the implantation. It is also necessary for the annealing temperature to be sufficiently high so that gross damage is removed in the first few minutes of the annealing cycle.

From a practical standpoint these restrictions do not impose a serious limit on the applicability of the theory since for B in Si they will meet for implantation energies above 70 keV, doses below 10^{15} ions/cm^2 and annealing temperatures above 900°C.

In reference to Table 1.1, for 70 keV B in Si, when the aforementioned implantation and annealing restrictions are not observed, a new species and additional kinetic terms need to be incorporated to the model. We have preliminary results for the low temperature case that are encouraging. However we are still awaiting experimental results to further support assumptions that we make. The high dose case and also the liquid nitrogen temperature implant case require modelling of the annealing kinetics of the amorphous layer produced during implantation.

In conclusion we believe the model will be successful in predicting a number of important cases of annealing. Furthermore we believe the model can be expanded as suggested in Section I to include more complex annealing situations, and that it will then satisfactorily predict profile shapes and electrical activities in these cases as well.

1.2 Calculation of Range Profiles and Recoil Implantation in Multi-Layered Media

1.2.1 Introduction

Application of ion implantation to device fabrication often involves processing in which surface coatings, e.g. SiO_2 and Si_3N_4, are present on the silicon substrates. Two questions of interest that arise in cases such as these are as follows: (1) when an implantation is done through a

thin layer of surface coating on a target, what is the range distribution of the primary projectile in the coated (or multilayer) target, and (2) during this process the highly energetic primary projectile ions impart enough energy to the target atoms so as to dislodge them from their lattice sites (Fig. 1.6). What is the range and distribution of the knock-on atoms that are recoil implanted from the surface coating into the target?

During the first six months of this program the work was primarily concerned with the experimental and verification of the knock-on phenomena in the case of arsenic implantation in Si through a mask of SiO_2. During the remaining nine months the work was directed toward calculation of the range profiles of the knock-on atoms by the application of Boltzmann transport equation.

1.2.2 Experimental Measurements of Knock-On Atoms [19]

The recoil of oxygen from As-implantation was studied by employing MOS surface state techniques. Arsenic was implanted into an SiO_2 film on Si such that only the recoil oxygen reached the silicon. By varying oxide thickness, the depth distribution could be determined. The distribution at greater depth is exponential with a characteristic length of 217 Å for an As-implant at 100 keV.

Oxygen in thermal equilibrium in silicon is very difficult to detect. Therefore, very little quantitative experimental data on this effect has been available to date. We have used an MOS method that is very sensitive even at low concentrations of oxygen. In this study, we implanted As into an oxide film on top of a silicon single crystal such that the projected range of As was well within the oxide and only recoil oxygen reached the silicon. In this manner, the two species could be separated. The oxide thickness was varied, and the total integrated oxygen concentration in the silicon was measured by determining the surface state density induced by oxygen (Figure 1.7). The correlation of surface state density and oxygen concentration was established by implanting various doses of oxygen into test samples.

Fahrner and Goetzberger [10] have shown that ions implanted into an $Si\text{-}SiO_2$ interface produce characteristic surface state peaks that can be studied by MOS capacitance techniques. Fahrner [11] has subsequently observed in more detail the behavior of oxygen implanted into the interface. He found that oxygen yields well resolvable peaks of surface state density that are

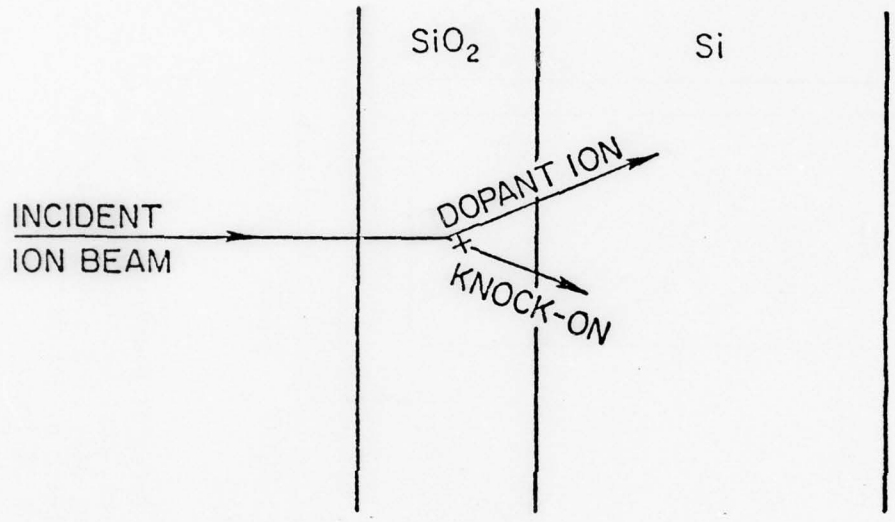


Fig. 1.6 Knock-on of oxygen during high energy implantation of heavy ions through an oxide mask.

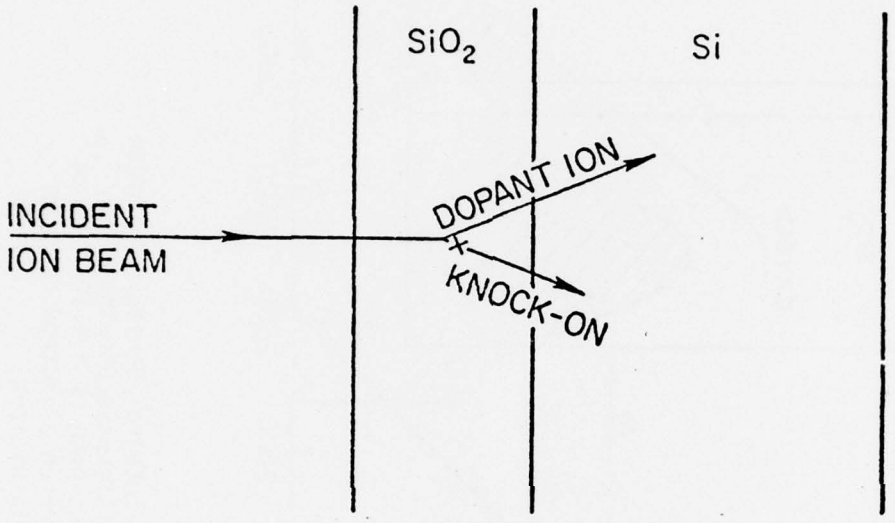


Fig. 1.6 Knock-on of oxygen during high energy implantation of heavy ions through an oxide mask.

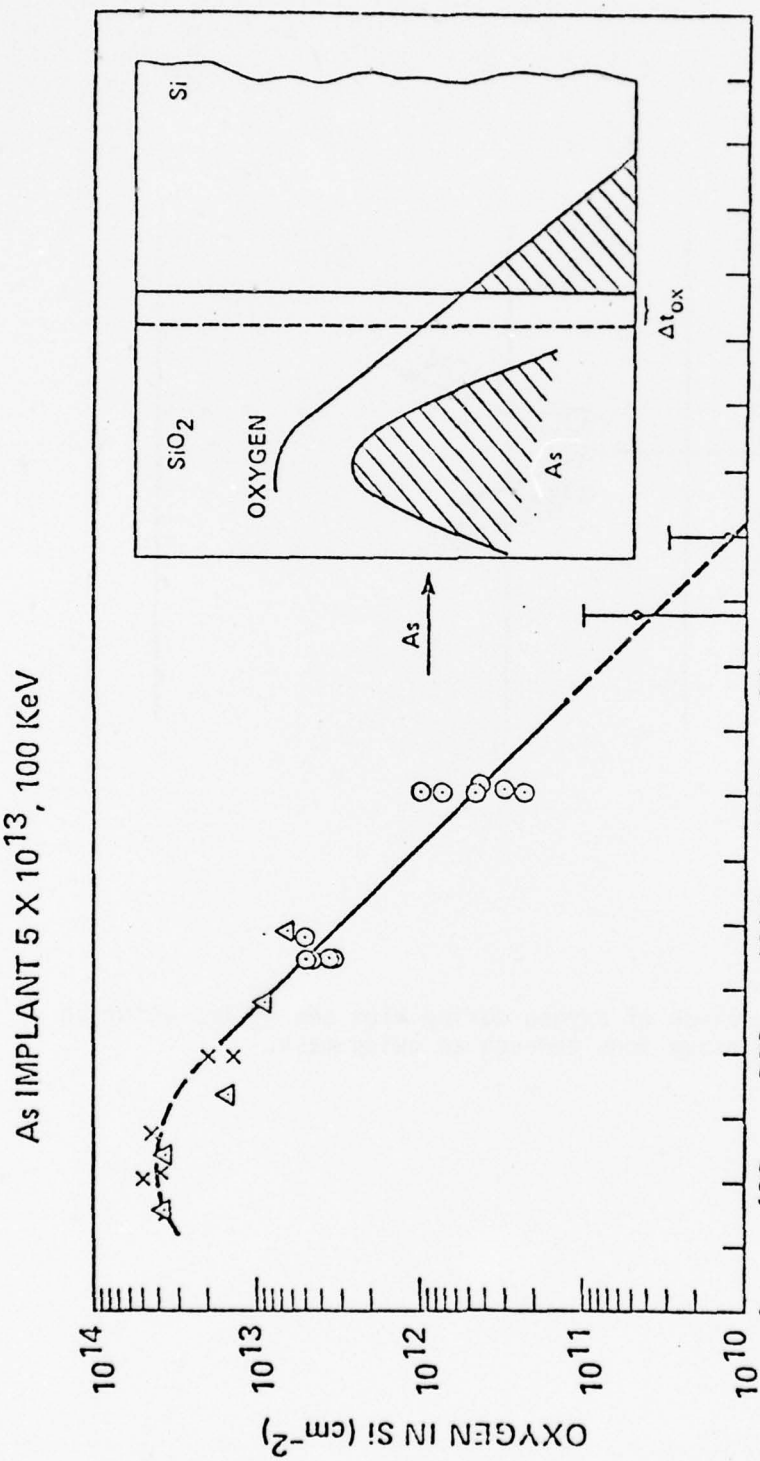


Fig. 1.7. Curve Showing Integral Density of Oxygen in Silicon versus Oxide Thickness Obtained from Experiments for an Implant Dose of 5×10^{13} As/cm^{-2} at 100 KeV. (Insert: Sketch of Method Used for Producing Recoil Oxygen in Silicon)

correlated with implantation dose and annealing conditions. We have in this investigation been able to verify most of Fahrner's observations.

The experimental conditions were as follows: silicon crystals of (100) orientation and a resistivity of 1-2 ohm-cm n-type were oxidized under clean conditions using modern MOS technology. A dry oxidation with subsequent high temperature annealing in nitrogen was employed. Aluminum metal contacts with an area of 3.6×10^{-3} cm^2 area were evaporated in an e-gun evaporation. Implantations were carried out before metalization. All samples were annealed at 450°C for 60 minutes in a forming gas atmosphere following metalization. This also constituted the annealing step for e-gun and ion implantation damage. Unimplanted control samples had surface state densities and fixed charge densities in the low 10^{10} cm^{-2} range or below.

Electrical measurements of surface state density were carried out by employing the slow ramp techniques [12,13] as in previous investigations [10,11]. An automatic data acquisition system permitted immediate plotting of surface state density versus energy in the gap for each sample.

First, we wish to discuss results on the control runs unde[illegible] to establish the correlation between surface state density and oxygen concentration. Since the peak surface state density of the oxygen peak measured in units of cm^{-2} eV^{-1} was used to measure oxygen concentration in units of cm^{-2}, a correlation had to be established first. For this purpose, samples with an oxide thickness of 764 $\pm$ 10 Å (measured by Talystep) were implanted with varying doses of oxygen with the implant energy chosen such that the peak of the distribution was at the interface. Oxygen range was adjusted using newly published tables [18]. Since the third moment ratio of oxygen in this energy range is very small, it was assumed that one-half of the implanted dose was located in the silicon and contributed to the signal. The following pertinent points were observed regarding the calibration runs: at low doses only one surface state peak located in the middle of the gap could be observed. At higher doses, a second peak in the upper half of the gap emerged. This peak grew at a faster rate and merged with the other peak at very high doses. Figure 1.8 shows the correlation of both peaks with implanted dose. Only the mid-gap peak was used for our investigations. An estimate of the total area of the surface state peak showed that only about one surface state was

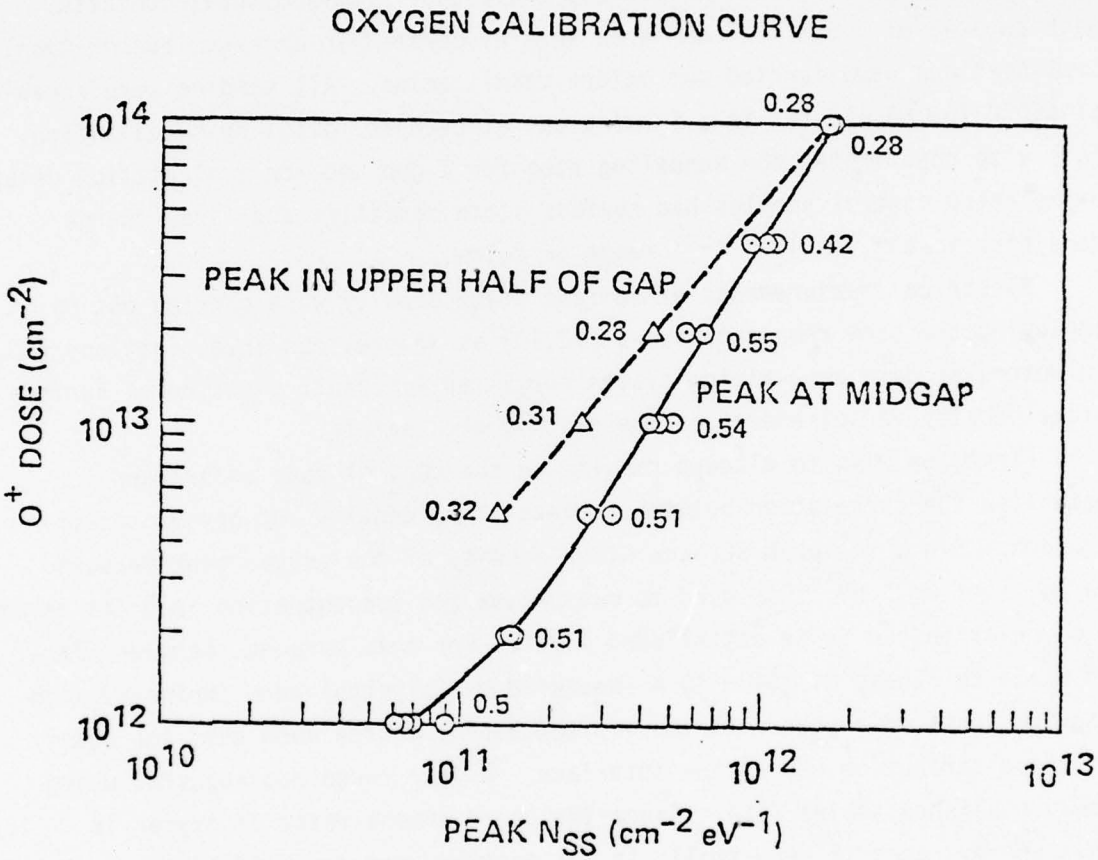


Fig. 1.8. Oxygen Dose versus Surface State Density

observed for 100 implanted oxygen ions. The location of the peaks in the gap was very sensitive to annealing temperature as well as to implant dose. The dependency of energy of the two observed peaks on dose is indicated at each of the measured points in Figure 1.8. It is likely that this dependence is caused by the superposition of two or three broad peaks, each of which depends differently on dose. The surface states are probably caused by complexes of oxygen with residual radiation damage from implantation.

Another calibration run was undertaken keeping the dose constant and varying oxide thickness. The results agreed well with those of the previous run. The results of Figure 1.8 are also in good agreement with Fahrner's sparser data. The position of the peaks did not agree with Fahrner's, perhaps because the annealing temperature was different.

For the second experiment, As was implanted at an energy of 100 keV into oxide layers of varying thickness. The projected range for this energy is 473 Å. After an annealing treatment identical to that of the calibration samples, both the C-V characteristics and surface state distribution had the same features indicating that the effect was also caused by oxygen. Potentially, As or Si could have interfered with the measurement. Because of the chosen implant energy, As could not reach the interface in large concentration. With high As doses and thin samples, however, doping effects due to As were seen. Si from the SiO_2 was not expected to interfere because its recoil efficiency is much less.

The results of this experiment are shown in Figure 1.7. Plotted are total number of oxygen atoms in the silicon versus oxide thickness for a dose of 5×10^{13} As cm^{-2} implanted at an energy at 100 keV. Recall, that although the data were obtained for oxygen in Si, Figure 1.7 represents the integral of the distribution of the recoil oxygens in SiO_2 because the oxide thickness was varied and the silicon was only used as a sensor for the oxygen atoms. It is seen that the oxygen concentration follows an exponential for larger oxide thicknesses with a characteristic length of 217 Å. This curve can be differentiated to obtain the real distribution of recoil oxygen in SiO_2. The exponential part gives $C(Ox) = 7.34 \times 10^{6} \, \phi_{As} \, (exp) - (x/217)$ for As implanted at 100 keV. $C(Ox)$ is the oxygen concentration, ϕ_{As} is the As dose cm^{-2}, and x is the depth coordinate measured in Å. The distribution obtained by differentiation is plotted in Figure 1.9, together with the distribution of

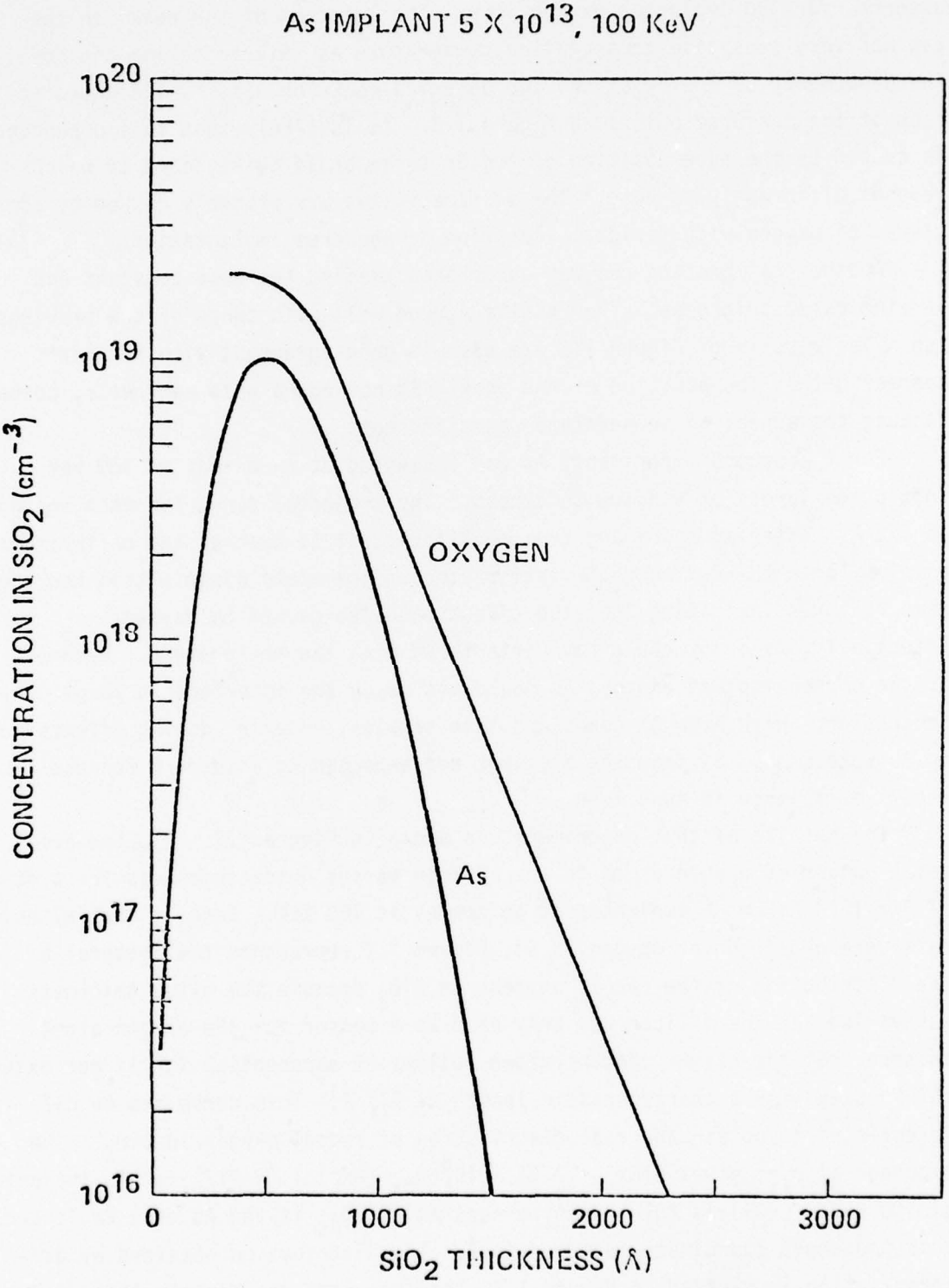


Fig. 1.9. Distribution of Implanted As and Recoil Oxygen in SiO_2.

implanted As. It is seen that the number of recoil oxygen atoms is slightly larger than that of the As implant and that their range is much larger. This distribution is easily converted to one in silicon by considering that in Si the range of oxygen is about 20% larger than in SiO_2.

1.2.3 Application of the Boltzmann Transport Equation to the Calculation of Range Profiles and Recoil Implantation in Multi-Layered Media

I. The answers to both of the questions raised in Section 1.2. can be obtained when the energy distribution of the primary projectile is known as a function of depth in the multilayer target. In particular, if an As ion is implanted into a Si_3N_4-SiO_2-Si target as illustrated in Figure 1.9 then a knowledge of the As energy distribution is sufficient to calculate both the recoil energy and recoil angle for each N and/or O recoil that is produced in a slab of width Δx_p at the position x_p. A knowledge of the As energy distribution at the Si_3N_4-SiO_2 interface is also sufficient to predict the relative concentration of As ions on the two sides of the interface.

It is of course possible to calculate energy distributions for semi-infinite targets using either the moments method of Winterbon et. al. [14] or the direct construction technique of Brice [15]. However these methods are not readily applicable in a multilayer target and it is therefore natural to resort to the Boltzmann transport equation, where the evolution of the energy distribution with distance is a quantity that may be calculated directly.

Use of the Boltzmann equation to calculate the energy distribution for the primary ion is in principle straightforward when each layer of the target is assumed to be a random stopping medium. However even for this case the analysis requires that several simplifying assumptions be made to construct numerical solutions in particular cases. Fortunately, results show good agreement with experiments where such comparison is possible, and the analysis provides a useful basis for additional calculations.

The technique is similar in concept to the Quasi-Monte-Carlo calculations of Furukawa and Ishiwara [16].

II. The Boltzmann Equation

Consider a collection of a larger number of particles with different velocities, $(\vec{v})$ located at different points in spaces $(\vec{x})$. We describe the collection by an average number,

$$dN_p = F(\vec{v},\vec{x})\ d^3\vec{v}\ d^3\vec{x} \qquad 1.7$$

for each differential element of phase space $d^3\vec{v}\, d^3\vec{x}$. We further assume that particles can undergo transitions from one region in phase space to another by means of collisions. The probability for a particle with velocity $\vec{v}$ to be scattered into the energy interval $d^3\vec{x}$ about $\vec{v}$ during a time dt is given by

$$K(\vec{v}\rightarrow\vec{v}')\, d^3\vec{v}'d\, t = N_s|\vec{v}|d\sigma\,(\vec{v}\rightarrow\vec{v}')\, d\, t \qquad 1.8$$

where $K\,(\vec{v}\rightarrow\vec{v}')$ is the transition rate, N_S is the density of scattering centers, and $d\sigma(\vec{v}\rightarrow\vec{v}\,)$ is the differential cross section for scattering from $\vec{v}$ to $d^3\vec{v}$ Note that the spatial dependence has been omitted for simplicity. A consideration of the average number of particles scattered into and out of a differential element of phase space leads to an integro-differential equation governing the density of F (v, x)

$$\frac{\partial F(\vec{v})}{\partial t} + \vec{v}\cdot\frac{\partial F(\vec{v})}{\partial \vec{x}} = N_s\int\left\{d\sigma(\vec{v}'\rightarrow\vec{v})\,|\vec{v}'\,|F(\vec{v}')-d\sigma(\vec{v}\rightarrow\vec{v}')\,|\vec{v}\,|F(\vec{v})\right\} + Q\,(\vec{v}) \qquad 1.9$$

The Quantity Q $(\vec{v})$ is a source term which describes the generation of new members of the ensemble. For example if the particles under consideration are generated in collisions (i.e. recoils), then Q $(\vec{v})$ would be calculated by consideration of these collisions. The spatial dependence of the quantities in Equation 1.9 has again been omitted for simplicity. Of course if more than one type of particle is being considered (e.g. primary ions and recoils) approximate subscripts should be added to the various quantities. Equation 1.9 is the standard form of the Boltzmann equation.

Equation 1.9 in conjunction with appropriate boundary conditions provides a basis for a complete description of the complex processes which result from energetic ion moving in an inhomogeneous media. A numerical solution in six dimensional phase space, while possible in principle, is unfortunately considerably beyond the reach of even the most energetic programmer equipped with the most sophisticated, modern computer. However, simplifying assumptions can be made which allow the extraction of a considerable amount of information in certain cases.

III. One Dimensional Penetration

For heavy ions incident on light targets, as for example As on SiO_2, angular scattering of the incident ion is relatively small and the ion

travels in nearly a straight line. The density function is then a function only of the depth of penetration x and the energy E of the ion. With this simplifying assumption equation 1.9 can be rewritten.

$$\frac{\partial f(E,x)}{\partial x} = N_s \int \left((d\sigma(E' \rightarrow E)\ f(E',x) - d\sigma(E \rightarrow E')\ f(E,x) \right) \qquad 1.10$$

where f(E,x) is the total flux of ions at depth x,

$$f(E,x) = \int dt\ |\vec{v}|\ F(E,x,t) \qquad 1.11$$

f(E,x) dAdE is then total number of particles with energy E to E + dE which cross an element of area perpendicular to the direction of incidence. Equation 1.10 has a simple physical interpretation, as in Figure 1.10. The first term on the RHS represents the rate for scattering from E' into dE, while the second term represents scattering out of the energy interval dE. Integration of equation 1.10 is carried out starting from x = 0 (the sample surface) and integrating inward, the starting condition being

$$f(E,0) = D\ \delta(E-E_0) \qquad 1.12$$

where D is the total dose in ions/cm^2, E_0 is the energy of the incident ion, and $\delta(E-E_0)$ is the Dirac delta function. It is tacitly understood that the terms on the right hand side must be summed over all collision types (i.e. elastic scattering via nuclear collisions as well as inelastic or electronic scattering).

IV. Scattering Cross Sections and Separation of Low Energy Levels

For numerical analysis we have used an analytical form for the nuclear cross section suggested by Sigmund [17].

For $T \geq T_{max}$, $d\sigma = 0$.

By a proper choice of X, q, and m this expression provides a good analytical approximation to scattering cross sections derived by various workers. A number of examples, given by Sigmund, are tabulated in Table 1.2. This expression provides a power law scattering for small t, and Rutherford

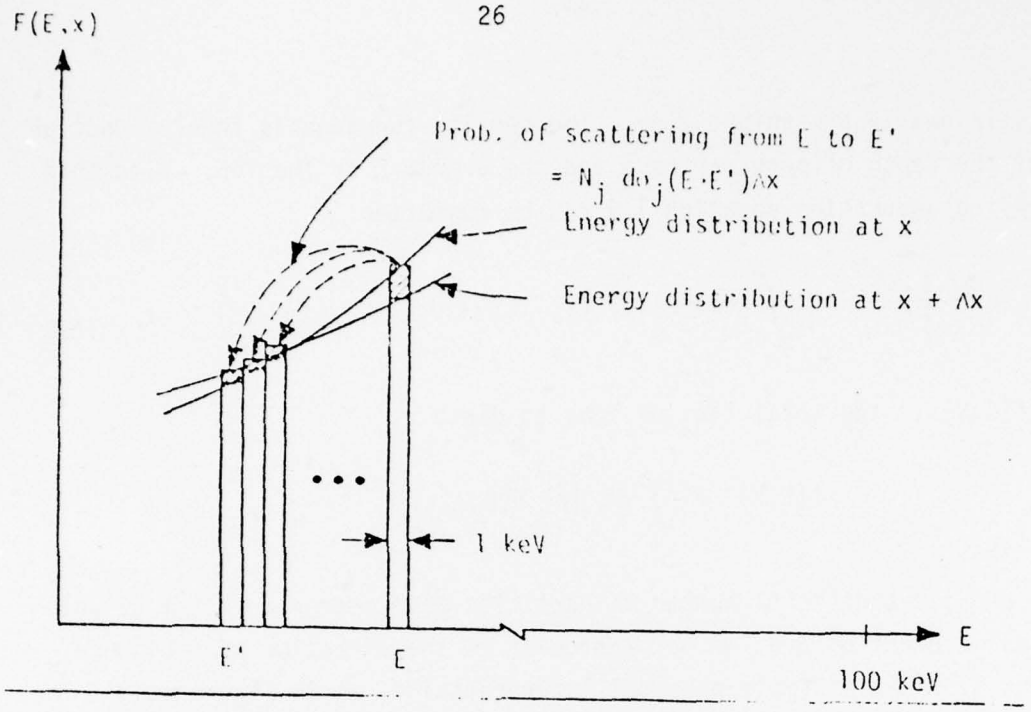


Figure 1.10 Schematic showing the process by which the quantity F(E,x) evolves.

$$d\sigma(E' \cdot E) = \left(\frac{\pi a^2}{4}\right) \left\{ \frac{2\lambda t^{-1-m} dt}{\left[1 - (2\lambda t^{1-m})^{q}\right]^{1/8}} \right\} \qquad 1.13$$

where λ, q, and m are constants and

$$a = .8853 a_o \, (Z_1^{2/3} + Z_2^{2/3})^{-1/2}$$

$$t = T E \left\{ \frac{M_2}{4M_1} \left(\frac{a}{Z_1 Z_2 e^2} \right)^2 \right\}$$

$a_o = \hbar/me^2 = .529$ Å

T = E' - E = transfered energy

M_1 = ion mass

M_2 = target atom mass

Z_1 = ion atomic number

Z_2 = target atom atomic number

e = electronic charge

$\hbar$ = plank's constant

Equation 1.13 is understood to apply only for $T \leq T_{max} = \frac{4M_1M_2}{(M_1+M_2)^2} E$

scattering at higher values of t.

As will be shown below the inelastic (electronic) scattering does not require a knowledge of the differential cross section for this process, but can be properly accounted for if the stopping power,

$$S_e(E) = \int T d\sigma_e \qquad 1.14$$

is known. For this purpose we use the expression,

$$S_e(E) = CE^p \qquad 1.15$$

where C and p are constants depending on the ion-target atom combination.

V. Separation of Low Energy and Electronic Collisions, Numerical Integration

It is convenient to rewrite the collision integral of equation 1.10 as follows:

$$I = \int \left| d\sigma(E' \to E)\, F(E') - d(E \to E')\, F(E) \right| = -\frac{\partial}{\partial E} J(E') \qquad 1.16$$

where $$J(E) = N_s \int_{E'=E}^{E_{max}} K(E',E)\, F(E')\, dE' \qquad 1.17$$

and $$K(E',E) = \int_{E''=0}^{E''=E} d\sigma(E'' \to E') = \pi p^2 \qquad 1.18$$

where p is the impact parameter for the transition $E' \to E$. For distributions F (E'), where F(E') does not vary appreciably over the energy interval where K(E',E) is non-zero (small energy or soft collisions) we can take $F(E') \approx F(E)$ in equation 1.17 and write

$$J(E) \approx N_s S(E)\, F(E) \qquad 1.19$$

Screening function	m	q	λ
Thomas-Fermi	.333	.667	1.309
Thomas-Fermi-Sommerfeld	.331	.588	1.70
Lenz-Jensen	.191	.512	2.92
Moliere	.216	.530	3.07
Bohr	.103	.570	2.37

Table I Parameters for use in the cross section expression (From Sigmund [8]).

where we have used the fact that

$$\int d E' \; K(E',E) = \int T \, d\sigma = S(E) \qquad 1.20$$

(10) and (14) are proven by partial integration.

The technique for numerical integration of equation 1.10 will now be described. Let E_o be the energy of the incident ion beam. Then the range $o \geq E \geq E_o$ is divided into energies E_1, E_2, ---, E_2, --- $E_n = E_o$. The flux f_j is defined by

$$f_j = \int_{E=E_{j-1}}^{E=E_j} f(E)\, d E \qquad 1.21$$

For electronic scattering and for nuclear events such that $T < \frac{E_j - E_{j-1}}{2}$ the collision integral is represented by terms of the type shown in equation 1.16 with J (E) given by equation 1.19. The following describes the development of the energy spectrum on passing from x to $x^+ \Delta x$.

$$x \rightarrow x + \Delta x \qquad 1.22(a)$$

$$f_j \rightarrow f_j + N_s \Delta x \sum_{j'} \left| f_j / d\sigma_{j' \rightarrow j} - f_j d\sigma_{j \rightarrow j'} \right| \qquad 1.22(b)$$

$$E_j \rightarrow E_j - N_s \Delta x \, S_n(E_j) \qquad 1.22(c)$$

Here,

$$d\sigma_{j' \rightarrow j} = \int_{E=E_{j-1}}^{E=E_j} d\sigma(E_{j'} \rightarrow E) \qquad 1.23$$

and S accounts for electronic and small energy nuclear scattering events. This scheme can be shown to be mathematically rigorous, and as Δx and ΔE approach zero, one obtains an increasingly accurate approximation to the exact solution for f(E). The details of proof for this assertion will not be given here.

VI. An Example

We have used the algorithm outlined above as the basis for a computer program, which calculates the function f(E). Results for 100 kev As incident on a SiO_2 target are shown in Figure 1.11. This calculation uses Lenz-Jensen cross sections and 20% of the electronic stopping given by the Linhard formula.

Special considerations are required at very low energies and this region is shown shaded in the figure. One can show however that the total number of particles that pass into the shaded region between x and x^+ Δx is equal to $J(E_1)\Delta x$, where E_1 is the lowest energy considered. Since these ions have very little energy they can be considered to have stopped in this interval.

Figure 1.12 shows the range profile calculated using this technique. This profile is essentially identical to the one which would be calculated from LSS Theory using a joined half-Gaussian distribution and the stopping powers selected for the Boltzmann analysis.

VII. Knock-Ons

A natural consequence of the calculation outlined above is the source density $Q(\vec{v},x)$ for oxygen knock-ons produced in energetic collisions (FIgure 1.13). Thus an As-oxygen collision involving energy transfer T,

produces a recoiling oxygen with energy T traveling in the direction.

$$\cos \Theta = \sqrt{\frac{T}{\gamma E}} \qquad 1.24$$

By following the penetration of recoiling oxygen atoms one can construct the distribution of recoil implanted oxygen atoms.

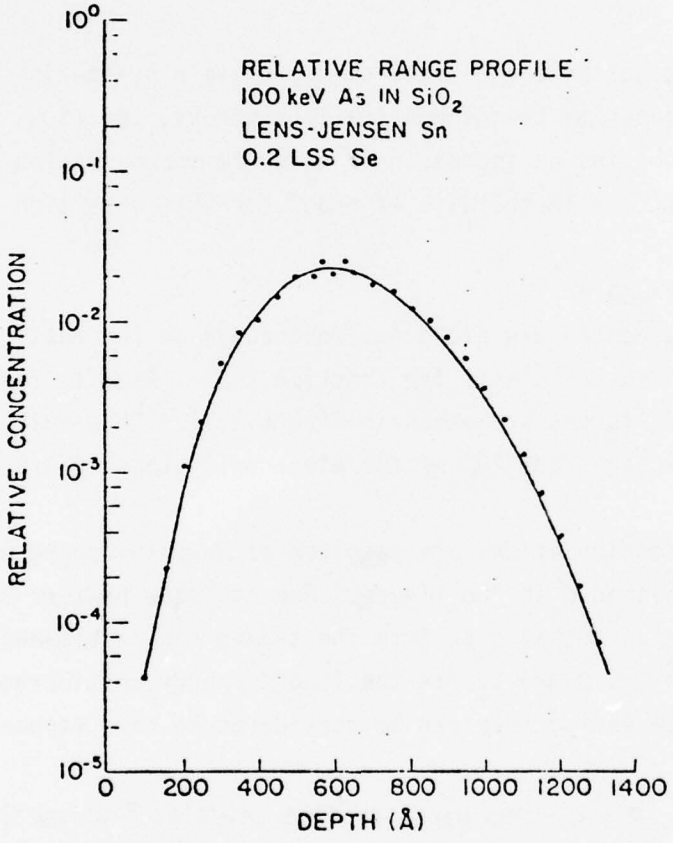
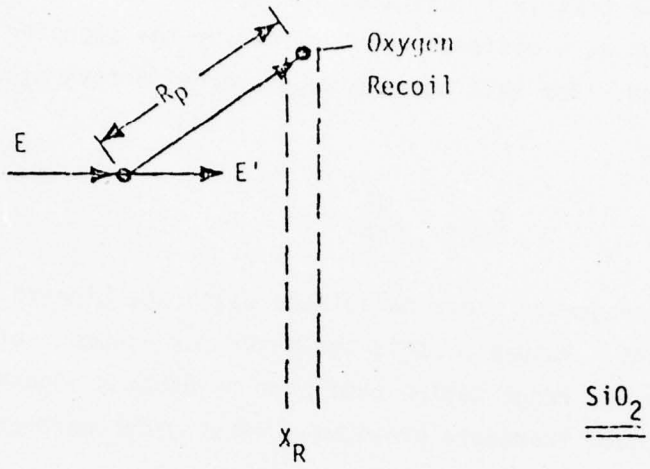


Fig. 1.11 Range profile for 100 kev As → SiO_2.



Calculation of Oxygen Recoil Distribution in SiO_2: Method 1.

Fig. 1.12 Production of an energetic oxygen recoil.

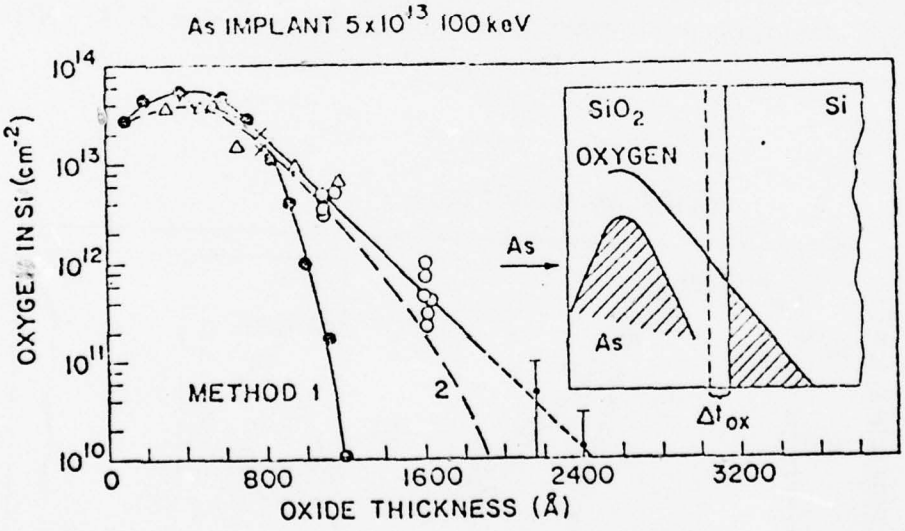


Fig. 1.13 Oxygen recoil implantation. The vertical axis shows the total dose of recoil implanted oxygen atoms as a function of oxide thickness. Experimental data is from Goetzberger et al [10]. The curves labeled method 1 and method 2 are calculated by techniques explained in the text.

A simple way to do this is illustrated in Figure 1.13. The recoiling atom is assumed to stop at a distance r determined by the stopping powers of the materials involved. For this purpose we use an effective stopping power defined by

$$S_{eff} = \frac{dE}{dRp} \qquad 1.25$$

where R_p is the projected range calculated using the Linhard approach to range calculations. Values of this parameter for a number of species can be obtained from the range tables published by Gibbons, Johnson and Mylroie [18]. This approach therefore provides a first order correction for effects due to straggling.

We have used this technique to calculate the total doses due to recoil implantation of oxygen, when 100 kev As is incident on a silicon target covered with a SiO_2 layer. The thickness of the SiO_2 layer is taken as a variable and the total oxygen dose implanted into silicon is calculated as a function of this thickness. For this purpose we consider only recoiling oxygen atoms which penetrate more than 100 A below the interface. The results of this calculation are denoted by "METHOD 1" in Figure 1.14. Also

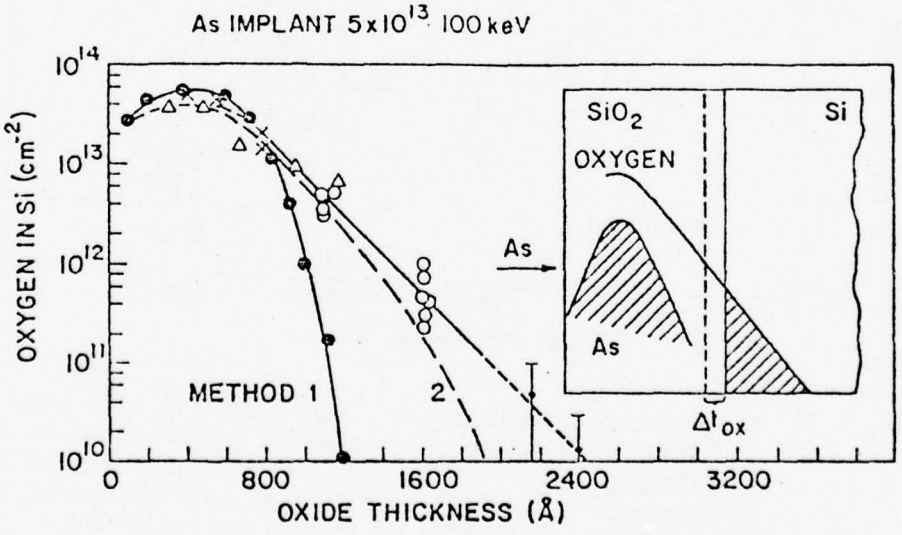


Fig. 1.14 Oxygen recoil implantation. The vertical axis shows the total dose of recoil implanted oxygen atoms as a function of oxide thickness. Experimental data is from Goetzberger et al [10]. The curves labeled method 1 and method 2 are calculated by techniques explained in the text.

shown are experimental data of Goetzberger et. al. [19]. As can be seen the agreement is quite good except in the tail of the distribution where straggling becomes important. The curve labelled "METHOD 2" was obtained using a fairly rough estimation of the effects of straggling of the recoil ions on the final recoil distribution.

VIII. Discontinuity At An Interface

Because of the differences in stopping power associated with various substances, a concentration discontinuity is expected at an interface between two layers [18]. This result appears naturally in range calculations using the technique outlined above. It is also possible to obtain an analytical expression which should be fairly accurate even for light ions, or ions crossing the interface at an angle.

The number of ions scattering into energies below any given energy E is proportional to J(E) defined in equations 1.16 - 1.18 for $E \to 0$ this then gives the number stopping at depth x. If we use the approximation of equation 1.19 for J we can write

$$\frac{n_1}{n_2} \approx \left(\frac{S_1}{S_2}\right)_{E \to 0} \qquad 1.26$$

Here C_j is the concentration and S_j is the stopping power, in substrates 1 and 2 respectively. Using equation 1.13 one can take the limit indicated in 1.26 and obtain

$$\frac{n(1)}{n(2)} = \frac{\left(\Sigma_j C_\rho^j \left(C_\rho^j\right)^{-2m}\right)_1}{\left(\Sigma_j C_\rho^j \left(C_\varepsilon^j\right)^{-2m}\right)_2} \qquad 1.27$$

where C_ρ and C_ε are given by

$$C_\rho^j = N_j \pi a_j^2 \gamma_j$$

$$C_\varepsilon^j = \left[\frac{a_j M_2^j}{Z_1 Z_2 j e^2 (M_1 + M_2^j)}\right]$$

and the superscript j refers to different target atom type. The quantities C_ρ and C_ε have been tabulated for a number of target-ion combinations in the book by Gibbons, Johnson, and Mylroie.

Values of the calculated interface discontinuity rates for three interface types are tabulated in Table 1.3. Note that the values are fairly insensitive to the choice of the parameter m. Experimental measurements of Combosson et. al. are included in column 3 of the table.

IX. Conclusions

The approach using the Boltzmann equation is seen to be a useful one for calculation of ion implantation effects in multilayered targets. In particular, the energy distributions for a heavy primary ion can be calculated as a function of depth in a multilayer target, and from this information range distributions of both the primary ion and the recoils can be constructed.

Interface	T - F (m = .333)	L - J (m=.191)	Combasson, et al (experiment)
SiO_2 on Si	1.68	1.86	0.7 - 2.2
Si_3N_4 on Si	2.18	2.42	2.1
Si_3N_4 on SiO_2	1.3	1.3	1.1

Tabel 1.2 Comparison Between Theory and Experiment for Boron Range Profile Discontinuity. Tabulated Quantities are C_1/C_2 as given by eq. 21. The last columns are experimental results of Combasson, et al [11].

REFERENCES

1. J. R. Anderson and J. F. Gibbons, "New Boron Diffusion in Silicon," Appl. Phys. Lett. 28, 184 (1976).

2. J. R. Anderson and J. F. Gibbons, "Measurements of Diffusion Parameters in Silicon Using Proton Enhanced Diffusion," Appl. Phys. Lett., in press.

3. W. K. Hofker, H. W. Werner, D. P. Oosthoek, H. A. M. de Grefte, "Influence of Annealing on the Concentration Prof. of Boron Implantations in Si," Appl. Phys. 2, 265-278, Springer-Verlag, 1973.

3[1]. W. K. Hofker, H. W. Werner, D. P. Oosthoek, N. J. Koenan, "Boron Implantations in Silicon: A Comparison of Charge and Boron Conc. Profiles," Appl. Phys. 4, 125-133, Springer-Verlag, 1974.

4. D. K. Brice, "Spatial Distribution of Energy Deposited Into Atomic Processes in Ion Implanted Silicon," Proc. of the First Int. Conf. on Ion Imp. in Semiconductors, Thousand Oaks, CA, Gordon and Breach Pub., New York, 1970.

5. J. E. Westmoreland, J. W. Meyer, F. H. Eisen, B. Welch, "Production and Annealing of Lattice Disorder by 200 Kev Boron Ions," Appl. Phys. Lett. 15, no. 9. Nov. 1969.

6. Tom Magee, personal communication.

7. T. E. Seidel and A. U. MacRae, "The Isothermal Annealing of Boron Implanted Silicon," Proc. of the First Int. Conf. on Ion Imp. in Semiconductors, Thousand Oaks, CA, Gordon and Breach Pub., New York, 1970.

8 T. E. Seidel and A. U. MacRae, "Some Properties of Boron Implanted Silicon," Trans. of the Metallurgical Society of AIME, vol. 245, 497, March 1969.

9. "First Semi-Annual Report on Computer-Aided Semiconductor Process Modeling," Stanford Electronics Laboratories Technical Report No. TR-4969-73-1, December 1975.

10. W. Fahrner and A. Goetzberger, Appl. Phys. Lett. 21, 329, (1972).

11. W. Fahrner, J. Electrochem. Soc. 121, 784 (1974).

12. R. Castagne, Compt. Rend, Acad. Sci., Paris 267, Series B, 866 (1968).

13 M. Kuhn, Solid State Electronics 13, 873 (1970).

14. K. B. Winterbon, P. Sigmund and J. B. Sanders, Mat. Fys. Medd. Dan. Vid. Selsk. 37, no. 14 (1970).

15. D. K. Brice, Proc. First International Conf. on Ion Implantation (Thousand Oaks, CA), pp 101-111, F. Eisen and L. Chadderton, Eds., Gordon and Breach, London, 1971.

16. S. Furukawa and H. Ishiwara, J. Appl. Phys. 43, 1268 (1972).

17. Peter Sigmund, Rev. Roumanian Physics, vol. 17, pp 283 (1972).

18. J. F. Gibbons, W. S. Johnson, S. Mylroie, "Projected Range Statistics," (Dowden, Hutchinson and Ross, Inc., Stroudsburg, Penn., 1975.)

19. A. Goetzberger, D. J. Bartelink, J. P. McVittie, J. F. Gibbons, Appl. Phys. Lett., Nov. 1976 (in press).

20. J. L. Combasson, J. Bernard, G. Guernet, N. Hilleret, M. Bruel, Proc. International Conf. on Ion Implantation in Semiconductors and Other Materials, edited by B. L. Crowder (Plenum Press, New York, 1973) p. 285.

CHAPTER 2

THERMAL OXIDATION*

James D. Plummer, Bruce E. Deal
Charles P. Ho, De is W. Hess, Robert W. Dutton

2.1 Introduction

The principal goal of the first year of the program has been to achieve accurate analytic prediction of oxide thickness for an arbitrary sequence of oxidations. At the present time, such prediction is possible only for lightly doped, (111) oriented silicon, with no chlorine species present during the oxidation process.

Toward this, kinetic oxidation data have been gathered under the following conditions:

a) Data for (111) and (100) silicon over the temperature range 700°C to 1200°C.

b) Data for dry O_2/HCl oxidation ambients containing 0-10% HCl for temperatures between 900°C and 1100°C.

c) Data for heavily phosphorus doped (111) substrates for surface concentrations up to solid solubility and temperatures between 900°C and 1100°C.

The availability of the above data now makes it possible to predict analytically the oxide thickness under a very wide range of ambient conditions and substrate orientation and doping levels.

The interpretation of the data which have been obtained have, to a large extent, been based on the general oxidation relationship developed in 1965 by Deal and Grove [1]. This relationship (as explained in the next section), uses two principal rate constants - B, the parabolic constant and B/A, the linear constant - to predict oxide thickness. The data generated in each of the three categories, above, have been largely interpreted in terms of modified linear and parabolic rate constants.

This technique allows analytic prediction of oxide thickness given that B and B/A may be expressed as functions of orientation, doping level and HCl concentration. Since B and B/A are based on a model in which oxidant diffusion through an existing oxide layer, and Si-SiO_2 interface reaction kinetics

*This work represents a joint effort by the Stanford University Integrated Circuits Laboratory and Fairchild Camera and Instrument Corp. Research and Development Laboratory.

respectively determine the overall oxidation rate, some understanding of the underlying physical mechanisms involved when HCl or heavy doping are present, may also be obtained from this model. This kind of interpretation of experimental results will be presented later in this report.

An even more basic physical understanding of the mechanisms involved in thermal oxidation will be needed, we feel, to fully understand oxidation kinetics and in particular to analytically predict oxide charge densities. What is really needed here, and what we are now beginning to pursue is a more microscopic model of the oxidation process.

Many of the important oxidation kinetic processes and most of the charges associated with SiO_2 are determined by the 20-40 Å transition region between Si and SiO_2. A clear understanding on an atomic or molecular level of this region appears to be necessary for a physical understanding of the kinetic and charge density experimental data obtained here and elsewhere. Toward this end, we are now beginning to work in the following three areas.

a) Profiles through thermal SiO_2-Si structures of Si, O, and dopants such as phosphorus are presently being obtained using the ion microprobe facility of the Materials Characterization Laboratory of Texas Instruments Inc.

b) Interaction with Spicer and his associates at Stanford and the sputtering Auger analysis equipment they have access to, is continuing.

c) A supplemental proposal to this grant is being prepared by Tiller and Pound of the Materials Science Department at Stanford. The essence of this proposal is a proposed calculation of SiO_2 film formation on an atomic or molecular level, including transport of ions in the film due to the presence of both donors and acceptors in the silicon and the presence of certain important species (such as Cl) in the gas phase. This analysis will be carried out with different interface orientations and will include the effect of strain, interface roughness and non-equilibriums interface effects.

The following sections of this report will present the experimental results obtained during the past year with HCl, (100) orientation and heavy doping as the new parameters in the oxidation process. The interpretation of these results will be based largely on the macroscopic diffusion - interface reaction model of Deal and Grove.

2.2 General Theory

The basic thermal oxidation model of Deal and Grove, based on the diffusion of the oxidant from the ambient through the existing oxide to react at the Si-SiO_2 interface, has been presented as:

$$X^2 + AX_o + B(t + \tau) \tag{2-1}$$

where

$A = 2D_{eff}(1/k + 1/h)$

$B = 2D_{eff}C^*/N_1$

$\tau = (X_i^2 + AX_i)/B$

$X_i = X_o\ (t = 0)$

C^* = equilibrium concentration of oxidant in oxide

h = gas-phase transport coefficient (from ambient to outer oxide surface)

D_{eff} = effective diffusion coefficient of oxidant in oxide

N_o = number of oxidant molecules incorporated per unit volume of oxide

k = interfacial reaction rate constant

The relation may be presented in two alternatives, providing more workable forms:

$$X_o = \frac{A}{2}\left[\left(1 + \frac{t + \tau}{A^2/4B}\right)^{1/2} - 1\right]$$

time as input parameter
thickness as output (2-2)

$$t = \frac{X_o^2 - X_i^2}{B} + \frac{X_o - X_i}{B/A}$$

desired thickness as input parameter
necessary time as output (2-3)

For oxidation time t, long, relative to the characteristic time $A^2/4B$, the oxidation model reduces to a parabolic behavior $X_o^2 \cong Bt$ with B, therefore, termed the parabolic rate constant.

For oxidation t, short, relative to the characteristic time, a linear behavior $X_o \cong B/A\ (t + \tau)$ results, with $B/A = (kh)(k+h)^{-1}(C^*/N_1)$ therefore labeled the linear rate constant.

Examination of the limiting forms indicates that the oxidation process in the parabolic domain is diffusion-limited, and in the linear region is surface-reaction limited (generally $h >> k$). Both regions should be directly dependent on the equilibrium concentration of oxidant in the oxide. Factors affecting the diffusion process should therefore be most influential for long oxidation times, etc.

Experiments have indicated that with an initially clean bare Si surface, $X_i = 0$ for wet O_2 oxidation, while an effective $X_i \cong 200$ Å is found for dry O_2 oxidation (due to an initial phase of rapid oxidation by a different mechanism).

At the present time, B, B/A and τ are known only for (111) oriented, lightly doped conditions. Under these restrictions these parameters may be expressed in the following form.

$$B = C_1 e^{-E_1/kT} \tag{2-4}$$

$$B/A = C_2 e^{-E_2/jT} \tag{2-5}$$

$$= (X_i^2 + AX_i)/B \tag{2-6}$$

where

Dry O_2: $C_1 = 7.72 \times 10^2$ micron^2/hr

$C_2 = 6.23 \times 10^6$ micron/hr

$E_1 = 28.5$ kcal/mole $= 1.23$ ev/molecule

$E_2 = 46.0$ kcal/mole $= 2.0$ ev/molecule

Wet O_2: $C_1 = 2.14 \times 10^2$ micron^2/hr

$C_2 = 8.95 \times 10^7$ micron/hr

$E_1 = 16.3$ kcal/mole $= 0.71$ ev/molecule

$E_2 = 45.3$ kcal/mole $= 1.97$ ev/molecule

Orientation, the addition of a chlorine species and heavy substrate doping have all been shown to affect either B, B/A or both rate constants. The following sections treat these individually.

2.3 Silicon Crystal Orientation Effects

A log-log plot of oxide thickness versus oxidation time for the thermal oxidation of silicon in dry oxygen at temperatures ranging from 700°-1200°C is shown in Figure 2.1. As expected, the difference in oxidation rate between (100) and (111) oriented silicon decreases gradually from 700°C, where the differences are on the order of 40%, to 1200°C, where the differences are less than 2% for all points, with the (111) oxidizing faster than the (100). It should be noted that in the region below 100 Å at 700°C, the oxide thicknesses obtained for (100) and (111) silicon are essentially the same. Only after the oxide thickness exceeds 100 Å, do the normal orientation differences appear. This observation suggests that in the initial oxide growth regime at 700°C, the silicon surface plays a minimal role. Such conclusions are consistent with previous information [2] obtained by Auger analysis of thin oxides grown between 200° and 800°C, which show that the hydration of the silicon surface plays a crucial role in silicon oxidation. The results can be of interest in the fabrication of MNOS devices.

The oxidation rate data were evaluated using the general relationship for the thermal oxidation of silicon as developed by Deal and Grove [1]:

$$x_o^2 + Ax_o = B\,(t + \tau), \qquad (2\text{-}1)$$

Since the parabolic rate constant B depends upon the partial pressure of oxidant in the oxidizing ambient, the solubility of the oxidizing species in SiO_2, and the effective diffusion coefficient of the oxidizing species in SiO_2 [1], there should be no variation of B with silicon orientation. On the other hand, the linear rate constant B/A involves the reaction rate constants at the SiO_2 surface and at the silicon surface [1], so that any orientation dependence of the oxidation rate should appear in this term via the constant A.

Since the method of analysis utilized depends upon the correction factor τ, which is determined by extrapolation [1], and thus prone to some error, oxidation times at least three times greater than the τ value for any set of oxidation conditions were generally used for the determination of B and B/A. In this way, any error in τ should have little effect on the calculated values of B and B/A.

The results of least squares analyses of oxidation rate data for (100) and (111) oriented silicon wafers using Equation (2-1) are given in Table 2-1. It should be mentioned that the oxide thickness (x_i) at zero oxidation time as determined by extrapolation of the x_o vs t curve to zero oxidation time, was found to be 160 $\pm$ 40 Å for all oxidation conditions. Table 2-1 demonstrates that, according to theory, there is essentially no difference in B for the two orientations, at least from 1200° to 900°C. Below 900°C, very long oxidation times (>100 hrs) become necessary to eliminate the effects of τ, and the scatter observed was considerable. At 700°C, a value for B could not be obtained, since a plot of oxide thickness versus time yielded a straight line, indicating that the parabolic contribution to the overall oxidation rate was negligible. The linear rate constant B/A, shows the expected orientation dependence, with B/A for (111) always being equal to or greater than that for (100). It should be noted that the kinetic parameters for (111) oriented silicon given in Table I are in very close agreement with those obtained by Deal and Grove [1]. Such agreement is remarkable when one considers the different oxidation conditions (humidity, gas quality, flow rates, furnaces, etc.) prevailing in two laboratories carrying out similar experiments eleven years apart.

Activation energies were obtained by fitting the data in Table 2-1 to an Arrhenius equation of the form

$$k = k_o e^{-E_a/kT} \qquad (2\text{-}7)$$

where E_a = activation energy, T = temperature in °K, and k = gas constant. Plots of B and B/A vs temperature are shown in Figures 2-14 and 2-15 in connection with the HCl data described later. This analysis gave values of 1.3 eV (30 kcal/mole) and 1.2 eV (28 kcal/mole) for the activation energy of B for (100) and (111) oriented silicon, respectively, which agrees well with previously determined values for the dry thermal oxidation of (111) oriented silicon [2], [3].

A similar analysis for the linear rate constant for the (111) orientation yielded a value of 2 eV (47 kcal/mole), which is in close agreement with the energy required to break a Si-Si bond [4]. However, determination of the activation energy for B/A for (100) oriented silicon affords two choices. If the data from 900° to 1200°C are analyzed, and the low temperature points

neglected due to large scatter as stated previously, a value of 2.5 eV (57 kcal/mole) is obtained, i.e., an increase of ~25% over that for (100). This increase was predicted by Ligenza from consideration of steric hindrance in the (100) and (111) planes of silicon during thermal oxidation [5].

On the other hand, if the data from 700° to 1200°C are analyzed, the 900°, 1100°, and 1200°C points could be considered as scatter from a line drawn parallel to the one for (111) orientation, passing through the 700°, 800°, and 1000°C points, thereby resulting in the same activation energy as (111) oriented silicon. Indeed, some faith could be put in such an assumption, since the 700°C data give B/A directly from an X_o vs t plot as indicated above. Thus, one could argue that the higher temperature analyses were in error due to the inaccuracy in determining A from a plot of x_o vs $t+\tau/x_o$. The slopes of these lines (B) are quite large, therefore a small change in the slope would shift the intercept (A), and thus B/A, considerably. At present, it is not known which of the above alternatives is the correct one. In any case, the preceeding discussion indicates why there has been little or no work reported since 1961 concerning the differences between oxidation rates and corresponding activation energies of (100) and (111) oriented silicon in dry oxygen.

2.4 Impurity Doping Effects

The effects of impurity doping levels on thermal oxidation rates are intimately connected with a widely encountered phenomenon in semiconductor processing - namely, impurity redistribution. Figure 2-2 illustrates how redistribution and thermal oxidation interact.

As a thermal oxide is grown over a doped silicon substrate, redistribution of the impurity results. In the case of phosphorus, arsenic, and antimony, the dopant atoms tend to pile up at the surface resulting in a higher surface concentration than background concentration ($C_s > C_B$). In the case of boron, the opposite effect takes place resulting in surface depletion ($C_s > C_B$).

In the case of very heavily doped substrates (i.e., C_B typically > 10^{19}), it has been observed [6] for both phosphorus and boron, that the oxidation rates can be substantially different (generally faster) than those observed on lightly or moderately doped substrates. With respect to Figure 2-2, the two parameters that have been correlated [6] with this increased oxidation are C_s, the dopant concentration in the silicon at the surface, and $\overline{C_{ox}}$, the average impurity concentration in the oxide.

Intuitively, the effect of C_s is to reduce the amount of energy required to break Si bonds and thus to affect the surface reaction. C_s would thus be expected to influence B/A, the linear rate constant and not B.

The effect of $\overline{C_{ox}}$ intuitively is to change the diffusion constant for the O_2 or H_2O oxidizing species in the SiO_2. $\overline{C_{ox}}$ would thus be expected to affect B, the parabolic rate constant.

In order to study these enhanced oxidation rates in more detail, a large number of heavily doped wafers were prepared as illustrated in Figure 2-3. Because of the difficulty of purchasing wafers with doping levels $>10^{19}/cm^3$, the samples were prepared by diffusion of phosphorus into standard lightly doped, (111) substrates. The predeposition and drive-in schedules were chosen so that the diffused profiles were flat (within 10%) over the first 2 μ into the silicon. The drive-ins were done in N_2 to prevent oxidation and hence redistribution during the formation of the heavily doped layers.

Five types of samples were prepared with surface doping levels between $5 \times 10^{19}/cm^3$ and solid solubility ($\approx 3 \times 10^{20}/cm^3$), as indicated in Figure 2-3. The resulting diffused profiles were measured using spreading resistance and anodic sectioning techniques. Agreement between the two measurement techniques was excellent as illustrated in the figure. It is important to note that the doping concentrations tabulated in Figure 2-3 represent <u>electrically active concentrations</u> not chemical concentrations. The implications of this will be described later.

Following preparation of the samples, a series of initial experiments were conducted to evaluate the magnitude of the enhanced oxidation effect over N^+ regions. A typical result is shown in Figure 2-4. The horizontal scale is again electrically active surface concentration. The data show approximately a three to one range in oxide thickness for these wafers which were oxidized simultaneously.

The choice of a low oxidizing temperature tends to maximize the enhanced oxidation rate for the common N type impurities because, as explained above, they tend to pile up at the silicon surface. As a result the dominant effect is on the reaction occuring at the $Si\text{-}SiO_2$ interface and hence the linear rate constant B/A which dominates the overall reaction at low temperatures. It should be noted that low temperature oxidation is becoming increasingly important in the semi-

conductor industry because of the trend towards larger diameter silicon wafers.

Wet oxidation also tends to increase the effect observed in Figure 2-4 because there is more redistribution for wet oxides when N type impurities are involved (i.e. Cs/Cb is larger for wet than dry oxides) [6], [7].

Following these initial experiments, a more carefully designed series of oxidations were made to evaluate the enhanced oxidation over N^+ regions under a wide variety of temperatures and ambient conditions. Typical of the results which were obtained are the data shown in Figures 2-5 and 2-6. The bottom curve (A) in each case is lightly doped (1×10^{15}) material and the results for this material agree with previously published data [1]. The upper curve in each case (F) corresponds to a substrate doped to approximately solid solubility.

Several observations can be made about these data even in the form shown in these figures.

(1) The enhanced oxidation effect is more pronounced at lower temperatures (900°C) than at higher temperatures (1000°C).

(2) These effects are also more pronounced for shorter times and thinner oxides than they are for longer times and thicker oxides.

Both of these observations are in agreement with our previously stated expectations. That is, N type dopants which pile up at the silicon surface and segregate into the Si, should affect the reaction kinetics at the Si-SiO_2 interface far more than they affect the diffusion of the oxidizing species through the SiO_2. Thus we would expect B/A to be affected and not B, in agreement with the observations stated above.

Based upon the general oxidation relationship of Deal and Grove [1], we can consider from a theoretical point of view, what the curves of Figures 2-5 and 2-6 should look like if the only effect of N^+ doping were on the surface reaction kinetics. Figure 2-7 shows a series of theoretical curves generated from this oxidation relationship. The various curves were generated for values of the linear rate constant B/A between the normal or lightly doped substrate value (2.8×10^{-4} μ/min at 900°C) and

500 times this value (corresponding to a much enhanced surface reaction rate). The qualitative agreement between these theoretical curves and the experimental results in Figure 2-5 is apparent.

We can carry the analyses of these data in Figures 2-5 and 2-6 one step further by extracting linear and parabolic rate constants according to standard techniques [1]. The results for the 900°C oxidation are shown in Figure 2-8. Two conclusions may be drawn from these data.

(1) The parabolic rate constant is essentially unchanged by the N^+ substrate. This is in accordance with our expectations since most of the phosphorus segregates into the silicon.

(2) The linear rate constant is virtually unchanged for doping levels $< 5 \times 10^{19}/cm^3$. Above this concentration, B/A increases approximately one order of magnitude as the doping level is increased to the solid solubility limit.

It should be noted that the horizontal scale in this figure has been corrected to chemical phosphorus concentration in the substrate, rather than the electrically active concentration previously used. The chemical phosphorus concentration and hence the percentage of the species electrically active in our diffused samples was determined by sputtering Auger techniques by Johannessen et al at Stanford [8]. There is still some question about the absolute accuracy of the horizontal scale in Figure 2-8 because of sensitivity limitations of the Auger technique. Independent measurements on these samples are currently being performed using SIMS techniques at the Texas Instruments Materials Characterization Laboratory. These should resolve the question of chemical vs electrically active phosphorus concentrations in these samples.

The 1000°C extracted rate constants show a form very similar to Figure 2-8. Samples have also been oxidized at 1100°C and at 800°C. The resulting oxide thicknesses are currently being measured. Completion of these data at these four temperatures will allow us to plot both B and B/A as a function of temperature and thus to extract activation energies for these rate constants as a function of substrate doping level.

The reported activation energy for lightly doped wafers for B/A, the linear rate constant, is approximately 46 kcal/mole [1]. This compares quite well with the energy required to break a Si-Si bond, which is 42.2 kcal/mole [4]. Our 900°C and 1000°C data indicate that B/A becomes

extremely large for the samples doped close to solid solubility. The activation energy for B/A in this case therefore approaches zero, indicating that the reaction at the Si-SiO_2 interface is no longer a rate-limiting step at these doping levels. That is, it appears that the incorporation of dopant atoms in concentrations close to solid solubility, so strains the lattice structure that bond breaking becomes extremely easy. Whether or not this simplified model is correct, can of course, only be verified by a better model of the molecular and atomic structure near the Si-SiO_2 interface. Such a model is clearly necessary also for prediction of charge densities and is therefore a goal of our research program. It is also something we are presently beginning to pursue in the manner indicated at the beginning of this oxidation report.

2.5 Oxidation in HCl/O_2 Mixtures

One of the most significant developments in the passivation of thermally grown silicon dioxide films over the past several years has been the addition of a chlorine species during silicon oxidation. It has been demonstrated [9], [10],[11] that such additions result in improved threshold stability and increased dielectric strength. In addition, it has been observed that chlorine additions increase the rate of silicon oxidation [12], [13], [14] but extensive data for widely varying oxidation conditions have not been reported. Therefore, in this investigation the thermal oxidation kinetics of silicon in O_2/HCl mixtures are characterized as a function of oxidation temperature, HCl concentration and silicon orientation, and the results are compared to those obtained from silicon oxidations performed in a dry oxygen atmosphere.

A. Rate Constants

Log-log plots of oxide thickness versus oxidation time for the thermal oxidation of (100) and (111) oriented silicon in O_2/HCl mixtures at temperatures of 900°, 1000°, and 1100°C, are shown in Figures 2-9, 2-10, and 2-11, respectively. It is clear from the figures that the (111) silicon always oxidizes faster than does the (100). Also, it can be seen that the effect of HCl addition is to increase the oxidation rate relative to oxidation in dry oxygen. In particular, a relatively large increase in the oxidation rate occurs with the initial 1% HCl addition. Subsequent additions have a somewhat smaller effect, but the overall rate is systematically increased.

In addition, it can be observed in these figures that the overall increase in oxide thickness for a particular oxidation time upon increasing the HCl concentration from 0 to 10%, is larger for (100) than for (111) oriented silicon. Although we can presently offer no explanation for this observation, it may be related to the silicon etching phenomenon mentioned below. It should be noted that very few data points are plotted for oxidations carried out at 1100°C with HCl concentrations greater than 1%. This is due to silicon etching and oxide bubble formation which occurs at the $Si-SiO_2$ interface during silicon oxidation with O_2/HCl mixtures at 1100°C [13], [14], [15], [16].

In order to separate the effects of HCl on the parabolic B and linear B/A rate constants, and thus obtain some indication of the role HCl plays in silicon oxidation, Equation 2-1 was utilized. It should be mentioned that during the determination of τ, x_i was assumed to be essentially constant at 160 $\pm$ 40 Å for all oxidation conditions investigated. Although x_i should decrease with HCl addition due to water vapor generation as per the reaction of O_2 and HCl,

$$4\ HCl + O_2 \rightarrow 2\ H_2O + 2\ Cl_2 \qquad (2\text{-}8)$$

the extrapolations performed to determine x_i and thus τ could not detect small changes in x_i. Nevertheless, since τ depends upon x_i, B, and B/A, a decrease in τ was observed upon HCl addition, probably in part as a result of water generation, since x_i and therefore τ are zero for oxidation in a steam ambient [1].

The results of least squares analyses on the data shown in Figures 2-9, 2-10 and 2-11 are summarized in Figures 2-12 and 2-13 which are semi-log plots of the effect of HCl concentration on the parabolic and linear rate constants, respectively. It should be noted that the rate constants generated via Equation 2-1 and plotted in Figures 2-12 and 2-13 are <u>effective</u> rate constants, in that they represent the combined effects of oxygen, water, and chlorine species on B and B/A.

Figure 2-12 shows that essentially no orientation effect is observed in the parabolic rate constant B, in accordance with theory. Also, increasing the HCl concentration above 1% results in a linear increase in B (the plots are put on a semi-log scale for convenience). The increase in B going from 0 to 1% HCl in dry oxygen is not as consistent among the three temperatures, however. For 1000° and 1100°C, a large increase in B due to 1% HCl addition occurs, whereas this trend is not observed at

900°C. These observations are in agreement with previous work which showed a linear increase of B with 0-9% HCl addition at 900°C [14] and a large increase in B due to 1% HCl addition at 1100°C [13]. Presumably, the large increase in B, as well as the linear increase with subsequent HCl additions are at least in part due to water generation as described in Equation (2-8), since it is well known that small amounts of water increase the thermal oxidation rate of silicon in dry oxygen [17]. However, as pointed out by van der Meulen and Cahill [14], gas phase thermodynamic equilibrium calculations indicate that the amount of water generated via Equation (2-8) cannot account for the observed increase in oxidation rate. This is consistent with the published results that Cl_2 addition to a dry oxygen atmosphere yields a considerably higher silicon oxidation rate than an equivalent HCl addition at 1150°C [17], suggesting that the chlorine species, rather than the water generated, may be primarily responsible for the increase in oxidation rate at high temperatures. However, since it is difficult to obtain Cl_2 which contains less than 3 ppm water, some effect, although small, would be anticipated from the water present.

In regard to the above argument concerning the amount of water generated via Equation (2-8), it should be pointed out that thermodynamic calculations such as those in Ref. [14] give only the idealized gas phase equilibrium conditions. The presence of a surface--especially one such as SiO_2 which is useful for catalytic purposes--could shift the equilibrium, and thus alter the amount of water generated and incorporated into the oxide surface.

Some insight can be gained into the silicon oxidation regime affected by water or chlorine (as generated via Equation (2-8)) from consideration of the definitions of the parabolic (B) and the linear (B/A) rate constants. The definitions are as follows [1]:

$$B \equiv 2\, D_{eff}\, C^*/N_1, \qquad (2\text{-}9)$$

where D_{eff} = effective diffusion coefficient, C^* = maximum concentration of oxidant in oxide, and N_1 = number of oxidant molecules incorporated into a unit volume of oxide;

$$B/A = \frac{C^*}{N_1(\frac{1}{k} + \frac{1}{h})} \quad , \tag{2-10}$$

where k = surface reaction rate constant, and h = gas phase transport coefficient.

Inspection of these definitions along with comparison of Figures 2-12 and 2-13 suggests that the monatonic increase in B with HCl concentration is related to the effective diffusion coefficient D_{eff}, since this term does not appear in the expression for B/A. Likewise, it appears that the rapid initial increase in B and B/A might be related to the solubility of the oxidizing species in the oxide film. Although the latter conclusion could be explained by water addition, this explanation would not account for the increase in D_{eff}, since this parameter is known to decrease with water addition [1].

In light of the above arguments, the gradual increase of the parabolic rate constant with increasing HCl concentration could be partially related to a chlorine species, since it has been established that a higher concentration of chlorine is incorporated into the silicon dioxide film as the HCl concentration and/or the silicon oxidation temperature are increased [18]. Consequently, it might be anticipated that increasing the chlorine concentration would cause the SiO_2 lattice to be strained (especially near the interface), thus allowing diffusion of oxidant to occur more easily, and thereby increase the oxidation rate as observed in the oxidation of heavily boron-doped silicon [19]. ince the ionic radius of boron when substituted for silicon in a silicon lattice (which should be close to the radius of boron substituted for silicon in SiO_2) is 0.88 Å [20], while the ionic radius of chlorine is 1.8 Å [21], it would be expected that the SiO_2 lattice expansion would be even greater in the case of chlorine as compared to boron. This speculation is also consistent with the previously mentioned observation that Cl_2 additions result in larger silicon oxidation rates than do HCl additions at high temperatures, since more chlorine is apparently incorporated into the SiO_2 with O_2/Cl_2 than with O_2/HCl mixtures [18].

Inspection of Figure 2-13 reveals a strong orientation effect on the linear rate constant B/A, which apparently decreases with increasing temperature. This observation was discussed in the previous section on

orientation effects. Also, as generally found for B, an initial rapid increase in B/A is observed upon increasing HCl concentration in the oxygen atmosphere from 0 to 1%. However, further increase of the HCl concentration resulted in essentially no change in B/A. Although an explanation for these observations is not apparent, it is conceivable that the breaking of Si-Si bonds could be promoted by etching of the silicon substrate which is known to occur with the use of HCl [13], [14], [15], [16], so that an increase in the linear rate constant might result. However, increased HCl concentrations should then further increase B/A, unless an equilibrium between oxidation and etching is established.

Clearly, additional investigation of this phenomenon is needed if we are to understand the mechanism involved. It should be mentioned that preliminary results obtained in the Fairchild laboratory with O_2/Cl_2 oxidation atmospheres substantiate the above interpretation, in that relatively large increases in B/A due to Cl_2 addition at 1100°C were observed, with considerably smaller increases noted for B.

B. Activation Energies

An Arrhenius plot of the parabolic rate constant B for the thermal oxidation of silicon in O_2 and several O_2/HCl mixtures is shown in Figure 2-14. Since essentially no orientation effect was observed for B, only the (111) orientation is plotted. A least squares analysis of the data for dry oxygen (0% HCl) gave a value of 1.2 eV (28 kcal/mole), which agrees well with previously determined values for the thermal oxidation of silicon in dry oxygen [1], [22], as well as for the diffusivity of oxygen through fixed silica [23]. It can be seen from Figure 2-14 that HCl addition to the dry oxygen atmospheres resulted in a convex curvature (relative to dry oxygen oxidation) of the Arrhenius plot.

An Arrhenius plot of the linear rate constant B/A for the thermal oxidation of silicon in O_2 and several O_2/HCl mixtures is shown in Figure 2-15. A least squares analysis of the data for dry oxygen (0% HCl) oxidation of (111) oriented silicon yielded a value of 2 eV (47 kcal/mole), which is in close agreement with the energy required to break a Si-Si bond [4]. However, determination of the activation energy for (100) oriented silicon affords two choices as explained in the previous section on orientation effects.

As in the case of the parabolic rate constant, the Arrhenius plot of the linear rate constant for HCl/O_2 mixtures (Figure 2-15) is curved. In this instance, the curvature is concave relative to dry oxygen oxidation. Upon consideration of Equation (2-8), it is not surprising that Arrhenius plots of the rate constants for the oxides grown in HCl atmospheres should result in curved lines. Certainly more than one activation process is occurring, e.g., the oxidation due to oxygen and water vapor, along with several possible effects due to chlorine, as mentioned above. As the oxidation temperature is changed, the relative importance of these activated processes may also change. Thus it becomes necessary to separate out the effects of water and chlorine on the silicon oxidation rate. Toward this end, experiments are currently being performed to investigate the oxidation rate of silicon in O_2/H_2O and O_2/Cl_2 mixtures.

Moving Boundary Oxidation Problem

Brief mention will be made here of a future goal of this oxidation study. In work by Rodoni and Dutton [24] at Stanford, funded under a different grant, computer techniques for solving the moving boundary oxidation problem have been developed. The goal of this work is to allow prediction of diffused and implanted profiles following a diffusion accomplished in an oxidizing ambient.

The growth of oxide during the diffusion process gives rise to redistribution which most often cannot be described by the usual gaussian or error function approximations to the diffusion process. Significant deviations occur near the SiO_2-Si boundary which, except for the special case of initial uniform impurity concentration and quadriatic oxide growth (25), require numerical methods for the solution of impurity profiles. [25], umerical techniques developed by Rodoni and Dutton are based upon the direct application of the conservation law in integral form [24]. This is the same conservation law from which the usual partial differential equations describing diffusion are derived under suitable conditions. By working directly from the integral form, numerical methods have been developed which are of a more general type than those based solely upon partial differential equations. As one example, the computer program is capable of generating concentration profiles throughout the oxide and silicon for general boundary conditions at both the gas/solid and oxide/silicon interfaces.

This computer program has recently been implemented on the same HP2100 computer that is being used in these oxidation studies. Figure 2-16 shows one example of the type of output which can be obtained. In this particular example, an initial boron implant (dashed line) has been subjected to a thermal oxidation of 90 minutes in dry O_2 at 1050°C. This oxidation step grows about 1080 Å of SiO_2, and might, for example, be a typical gate oxidation step in an MOS process. The resulting boron profile in both the silicon and the SiO_2 generated by the computer are shown in the figure.

At the present time, this computer program makes a number of simplifying assumptions including constant diffusion coefficients independent of concentration. It also neglects orientation effects, heavy doping effects and the presence of HCl in calculating the oxide growth. Our intention in the near future is to incorporate the results of this oxidation study into this computer program. We will thus be able to correct the computer model of oxidation kinetics to take into account orientation, heavy doping and the presence of HCl during the oxidation.

Finally, we expect this computer analysis technique to be extremely useful, in conjunction with the experimentally attained SIMS dopant profiles described at the beginning of this section, in contributing to our understanding of the physical mechanisms involved in the kinetcs of thermal oxidation.

REFERENCES

1 B. E. Deal and A. S. Grove, J. Appl. Phys., 36, Dec. 1965, pp. 3770-3778.

2 J. Ruzyllo, I. Shiota, N. Miyamoto, and J. Nishizawa, J. Electrochemical Soc., 123, 26 (1976).

3 A. G. Revesz and R. J. Evans, J. Phys. Chem. Solids, 30, 551 (1969).

4 L. Pauling, "The Nature of the Chemical Bond," Cornell University Press, Ithaca, New York, 3rd ed., 1960.

5 J. R. Ligenza, J. Phys. Chem., 65, 2011 (1961).

6 B. E. Deal and M. Skalar, J. Elec. Soc., 112, April 1965, pp. 430-435.

7 B. E. Deal, A. S. Grove, E. H. Snow and C. T. Sah, "Observation of Impurity Redistribution During Thermal Oxidation of Silicon Using the MOS Structure," J. Elec. Soc., 112, March 1965, pp. 308-314.

8 J. S. Johannessen, W. E. Spicer, J. F. Gibbons, J. D. Plummer and N. J. Tayler, "Observations of Phosphorus Pileup at the SiO_2-Si Interface, to be published.

9 R. J. Kriegler, Denki Kagaku 46, 466 (1973).

10 R. S. Ronen and P. H. Robinson, J. Electrochem. Soc., 119, 747 (1972).

11 C. M. Osburn, ibid 121, 809 (1974).

12 R. J. Kriegler, ibid 119, 388 (1972).

13 K. Hirabayashi and J. Iwamura, ibid 120, 1595 (1973).

14 Y. J. van der Meulen and J. G. Cahill, J. Electron. Materials, 3, 371 (1974).

15 R. J. Kriegler, "Semiconductor Silicon," H. R. Huff and R. R. Burgess, Eds., The Electrochemical Society, Inc., (1973), p. 363.

16 E. A. Irene, J. Electrochem. Soc., 121, 1613 (1974).

17 R. J. Kriegler, Y. G. Cheng and D. R. Colton, ibid 119, 388 (1972).

18 Y. J. van der Meulen, C. M. Osburn, and J. F. Ziegler, J. Electrochem. Soc., 122, 284 (1975).

19 B. E. Deal and M. Sklar, J. Electrochem. Soc., 112, 430 (1965).

20 Helmut F. Wolf, "Semiconductors," John Wiley and Sons, Inc., New Yor, 1971, p. 199.

21 A. F. Wells, "Structural Inorganic Chemistry," Oxford University Press, 1962, p. 68.

22 A. G. Renesz and R. J. Evans, J. Phys. Chem Solids, 30, 551 (1969).

23 F. J. Norton, Nature, 171, 701 (1961).

24 M. Rodoni and R. W. Dutton, "A Study of Numerical Methods for Solving the Moving Boundary Diffusion Problem," T. R. 5012-1, Stanford Electronics Labs, June 1975.

25 A. S. Grove, O. Leistiko, and C. T. Sah, "Redistribution of Acceptor and Donor Impurities During Thermal Oxidation of Silicon," J. App. Phys., vol 35, Sept. 1964, pp. 2695-2702.

Table 2-1

Rate Constants for <111> and <100> Silicon

Temperature (°C)	Orientation	τ(hr)	A(μm)	B(μm^2/hr)	B/A(μm/hr)
1200	(100)	0.03	0.0399	0.0453	1.14
	(111)	0.03	0.0404	0.0458	1.13
1100	(100)	0.09	0.101	0.0247	0.246
	(111)	0.09	0.0845	0.0244	0.289
1000	(100)	0.35	0.195	0.00913	0.0467
	(111)	0.35	0.120	0.00956	0.0797
900	(100)	3.2	0.429	0.00332	0.00775
	(111)	1.2	0.214	0.00381	0.0178
800	(100)	10.0	0.441	0.000755	0.00171
	(111)	5.0	0.354	0.00119	0.00335
700	(100)	--	--	--	0.000222
	(111)	--	--	--	0.000348

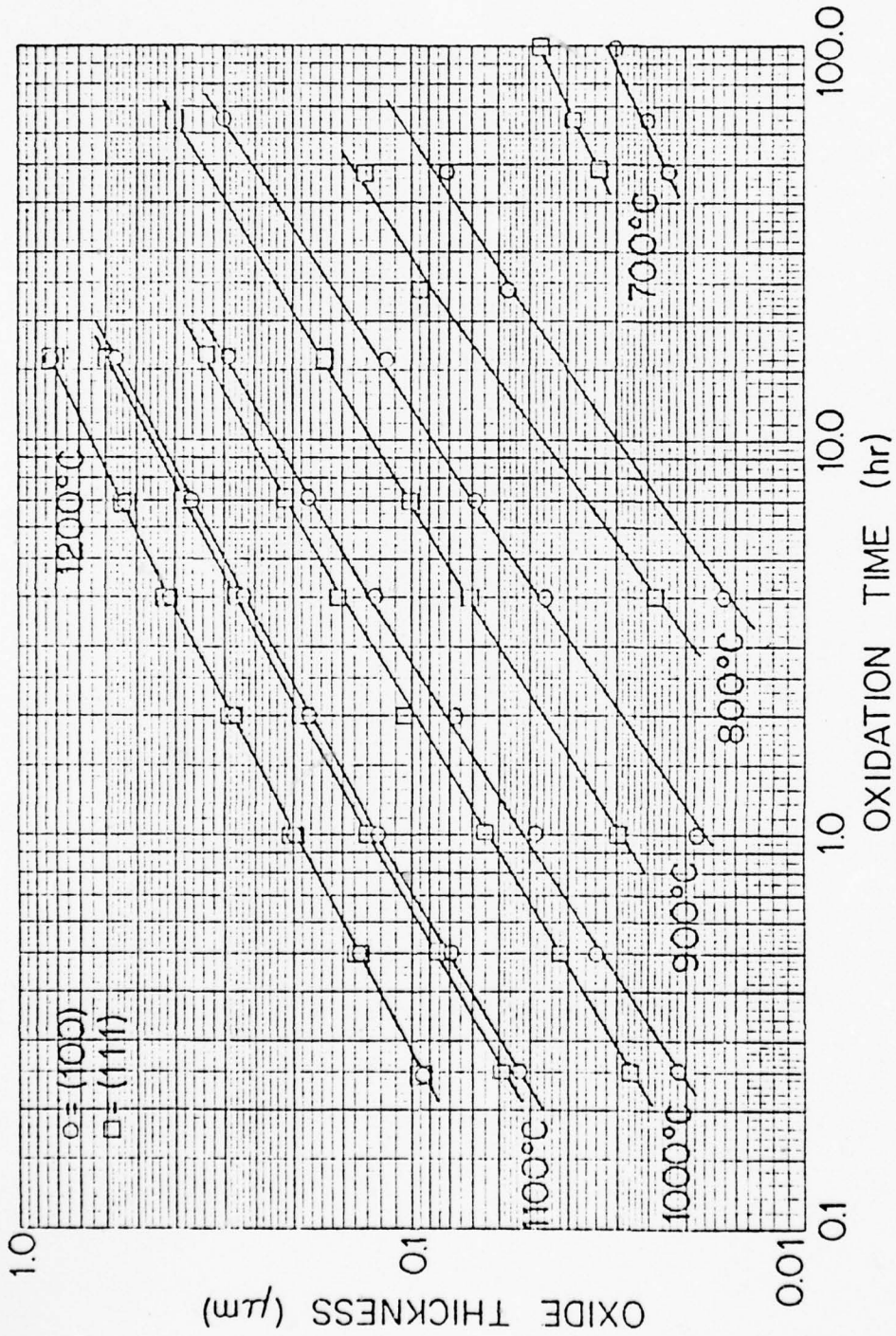


Fig. 2.1. Oxide Thickness vs. Time for <111> and <100> Silicon.

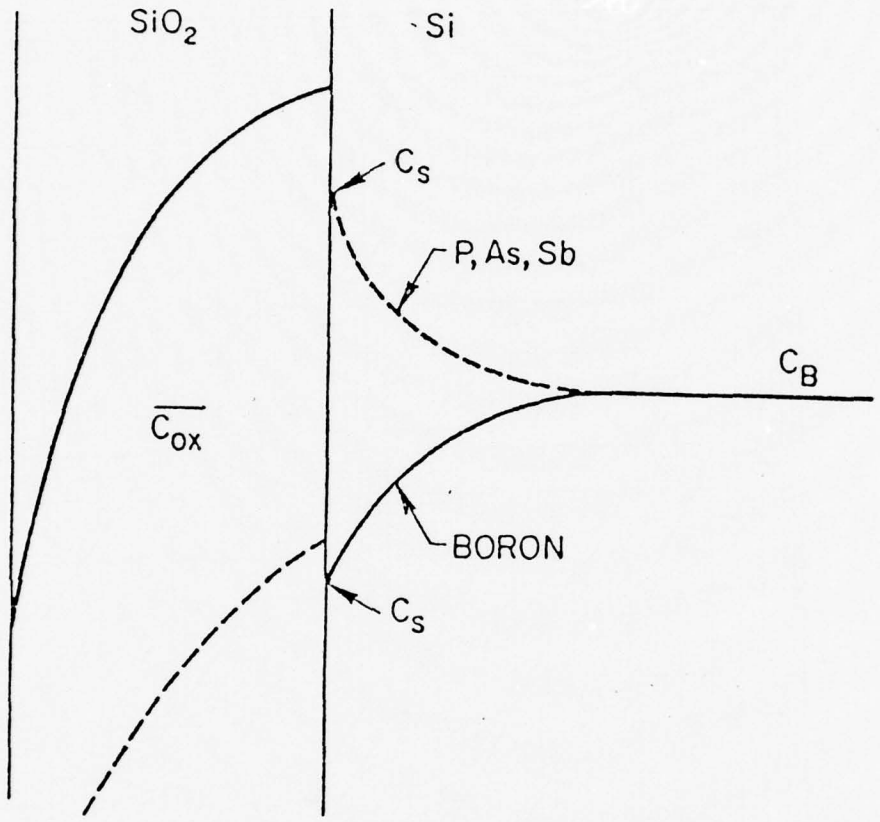
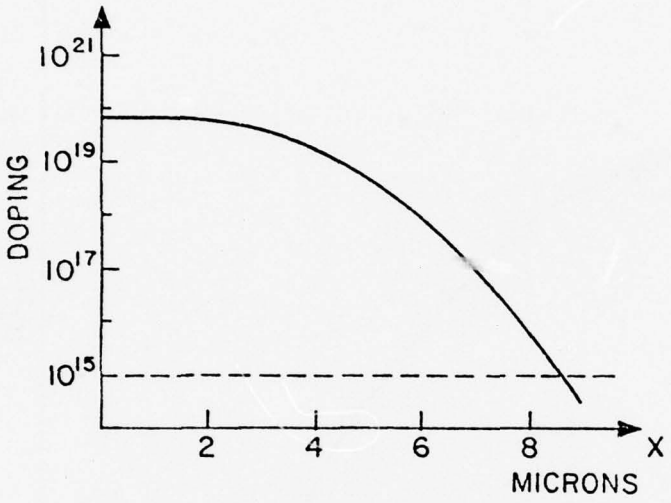


Fig. 2-2. Effect of Redistribution on Impurity Doping Profiles.



SAMPLE	C_s	
	SPREADING *P*	ANODIC SECTIONING
A	5×10^{19}	5.2×10^{19}
B	7.2×10^{19}	
C	1.8×10^{20}	
D	2.8×10^{20}	2.8×10^{20}
E	3.1×10^{20}	3.3×10^{20}

Fig. 2-3. Surface Concentrations of Diffused Heavily Doped Samples.

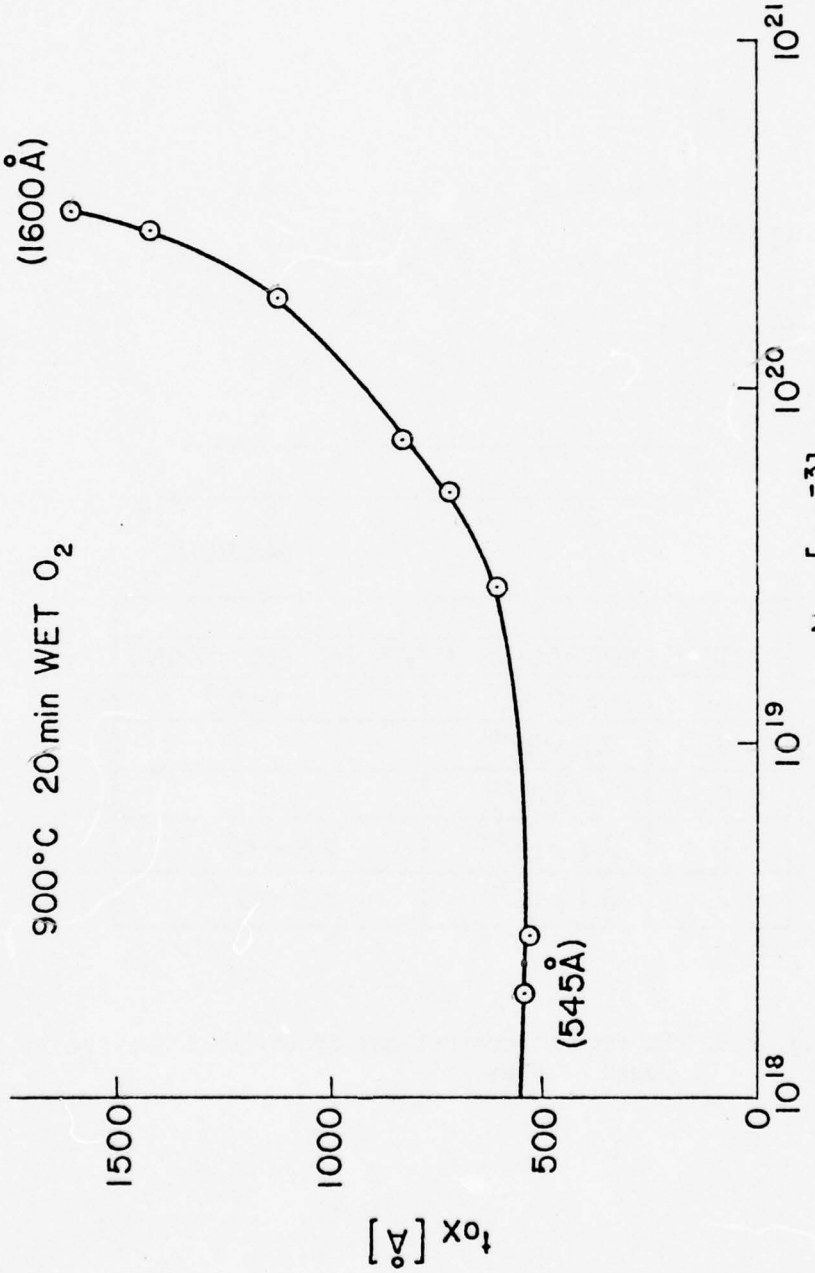


Fig. 2-4. Oxide Thickness vs. Doping - Samples Oxidized Simultaneously.

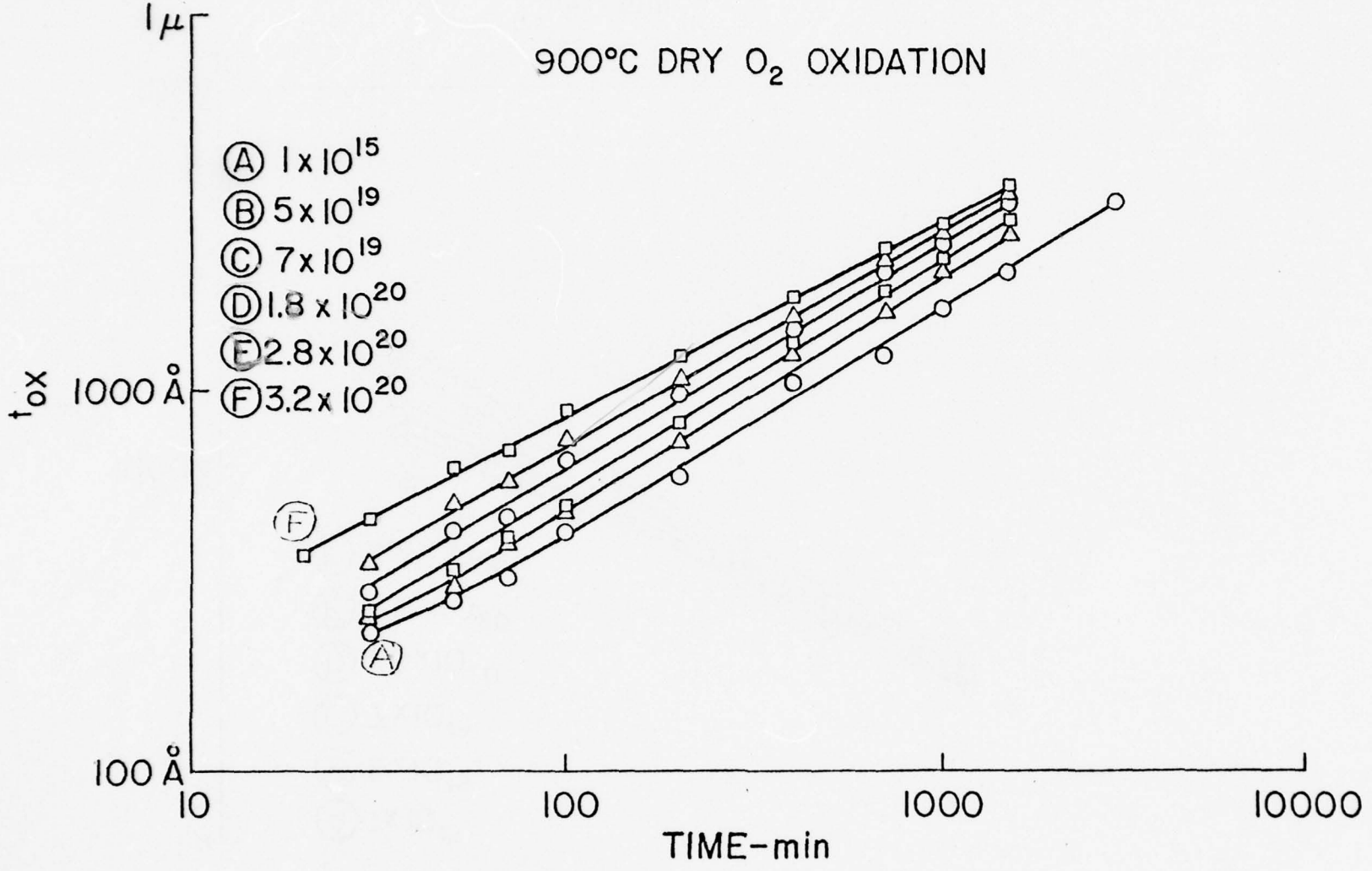


Fig. 2-5. Oxide Thickness vs. Time for Various Doping Levels at 900°C.

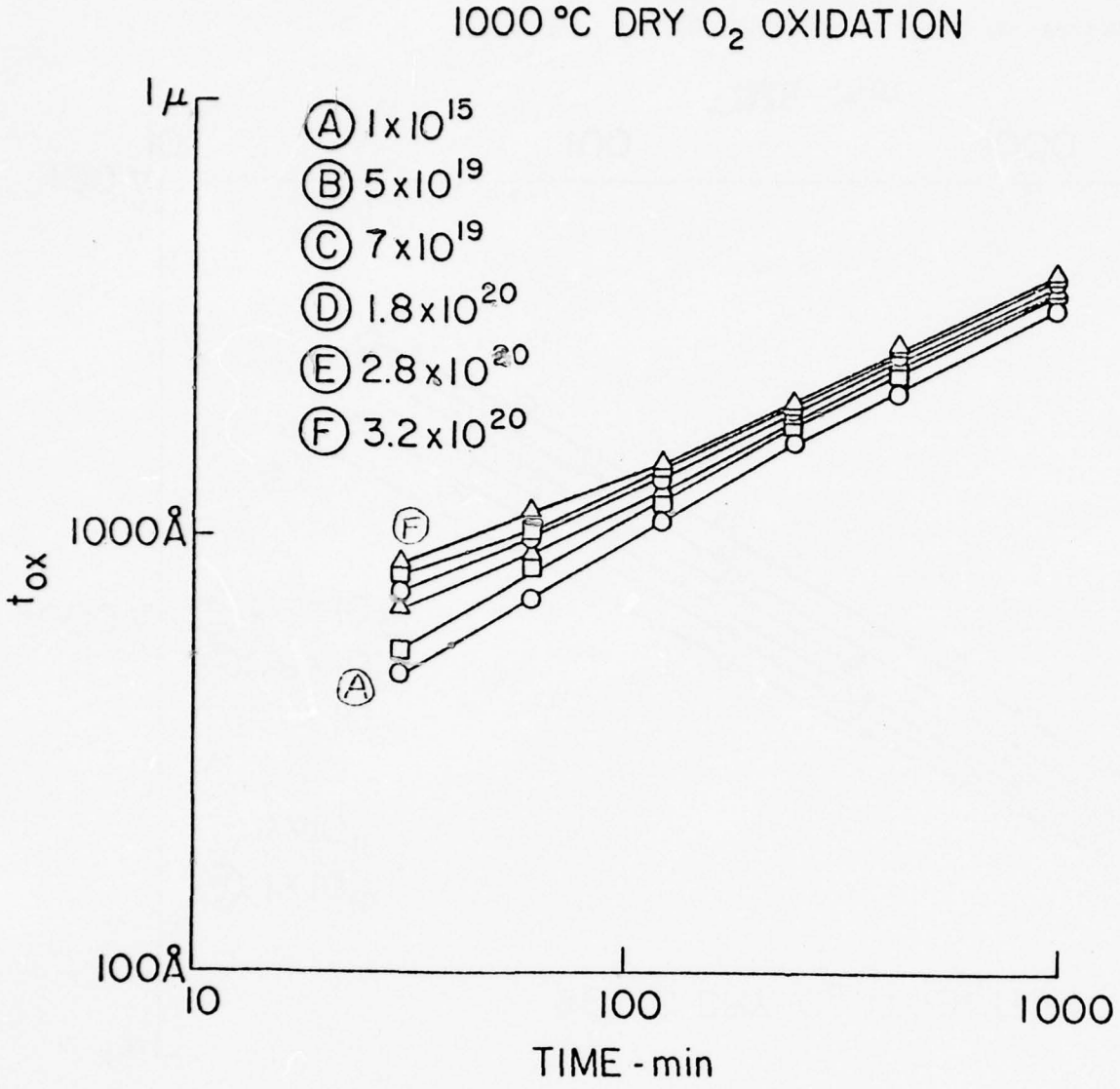


Fig. 2-6. Oxide Thickness vs. Time for Various Doping Levels at 1000°C.

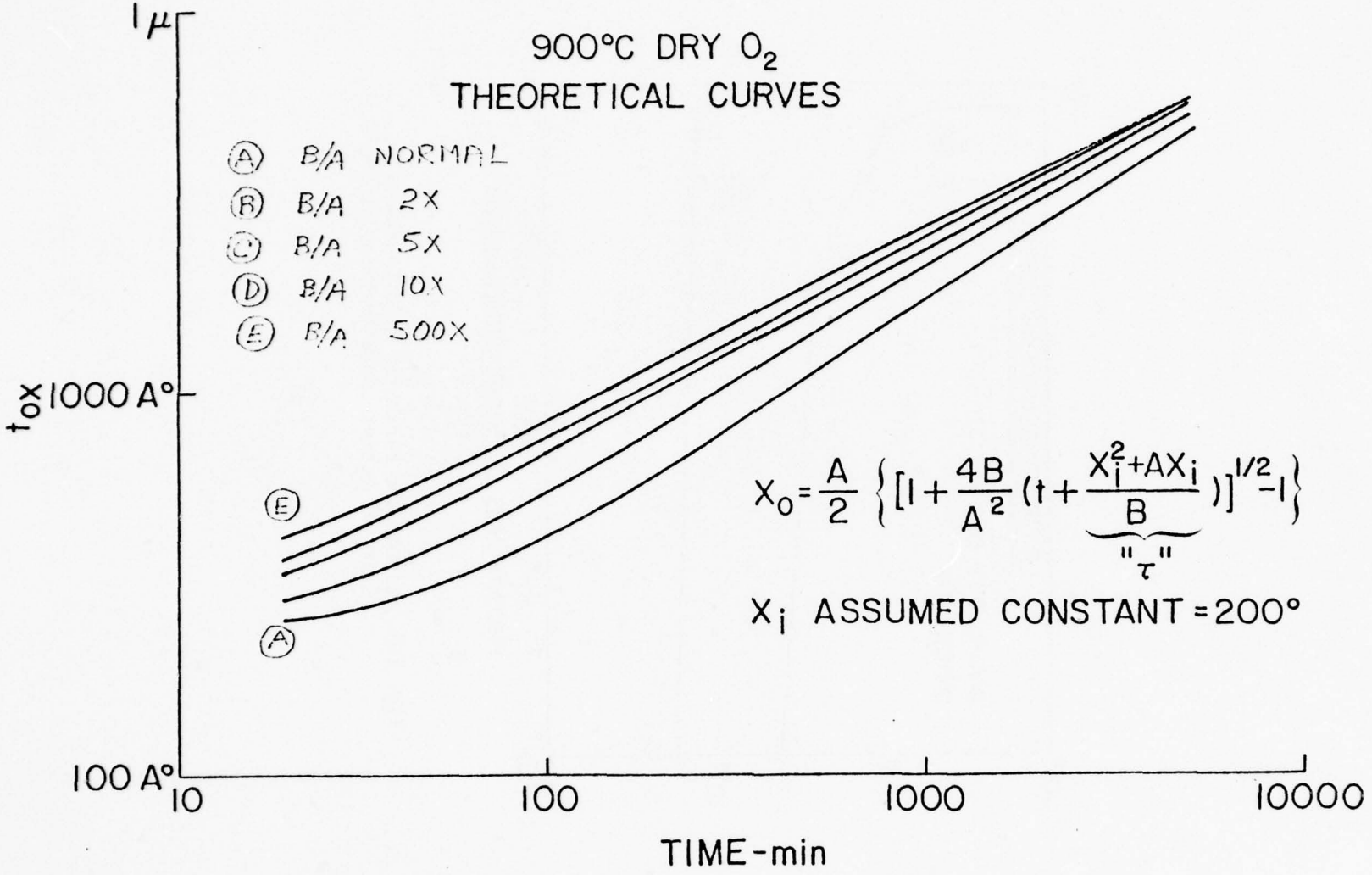


Fig. 2-7. Theoretical Curves of Oxide Thickness vs. Time for Various Linear Rate Constants.

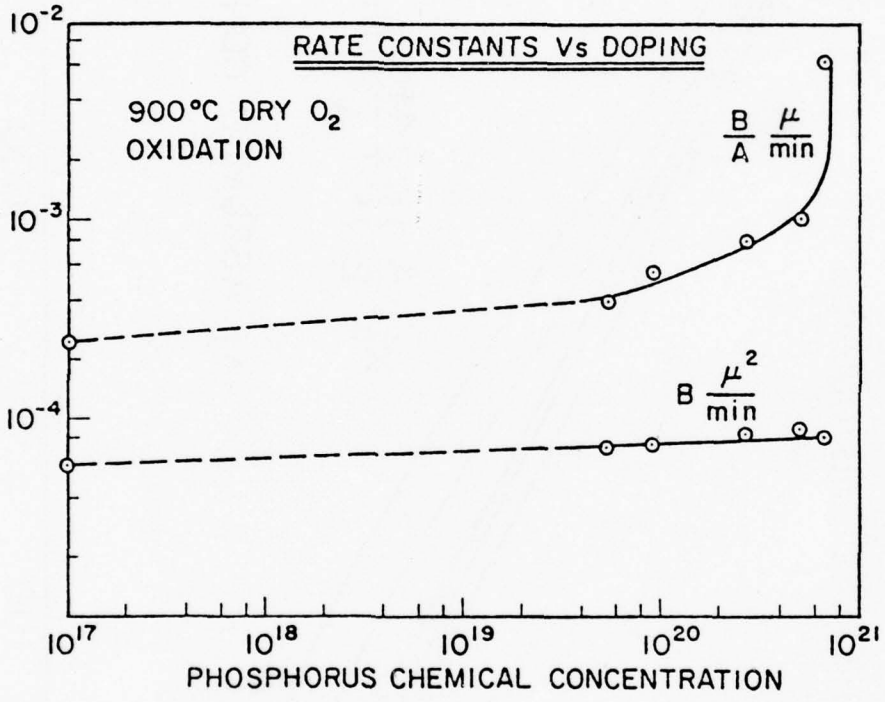


Fig. 2-8. Linear and Parabolic Rate Constants vs. Chemical Dopant Concentrations at 900°C.

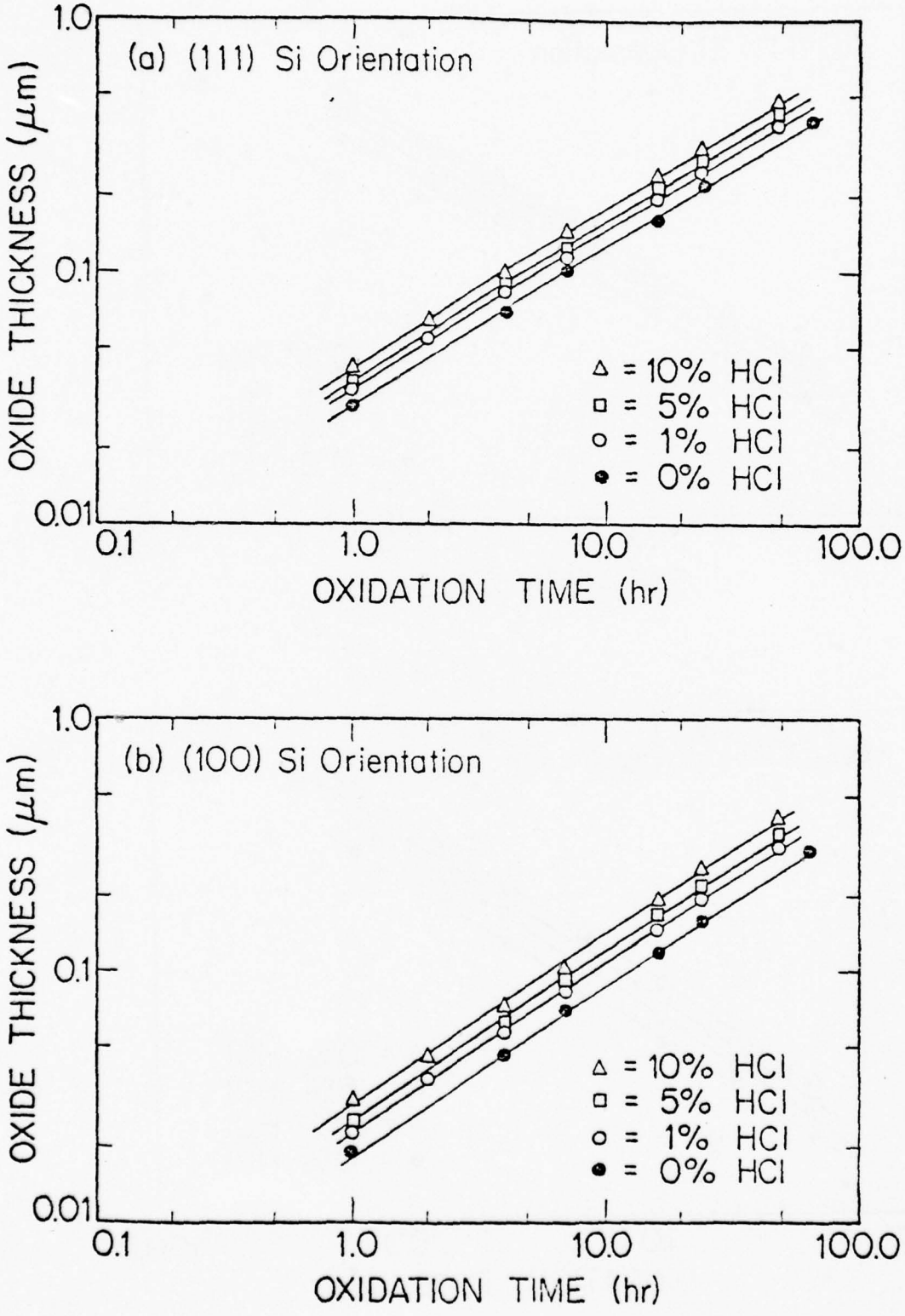
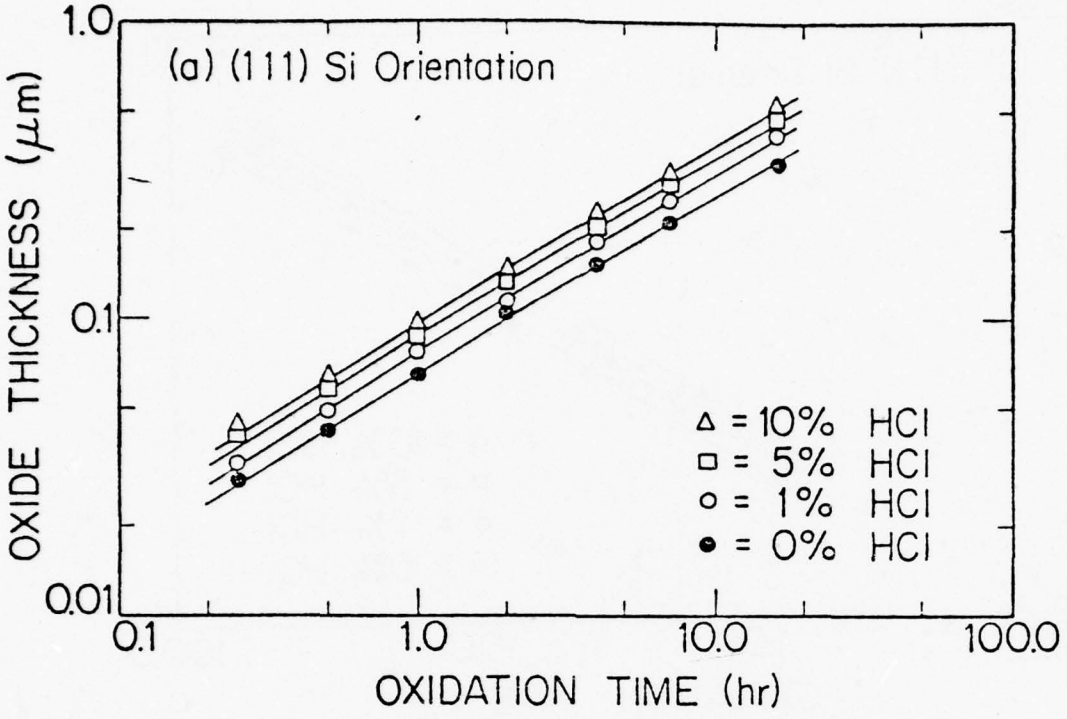


Fig. 2-9. Oxide Thickness vs. Oxidation Time for the Oxidation of <111> and <100> Oriented N-Type Silicon in Various O_2/HCl Mixtures at 900°C.



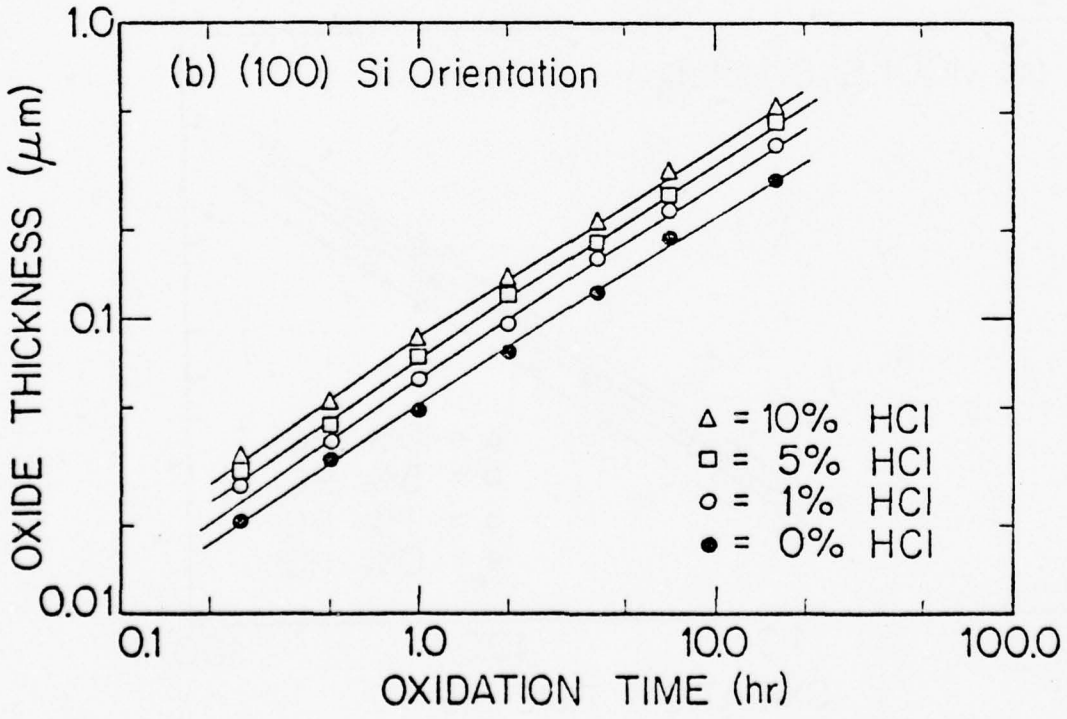
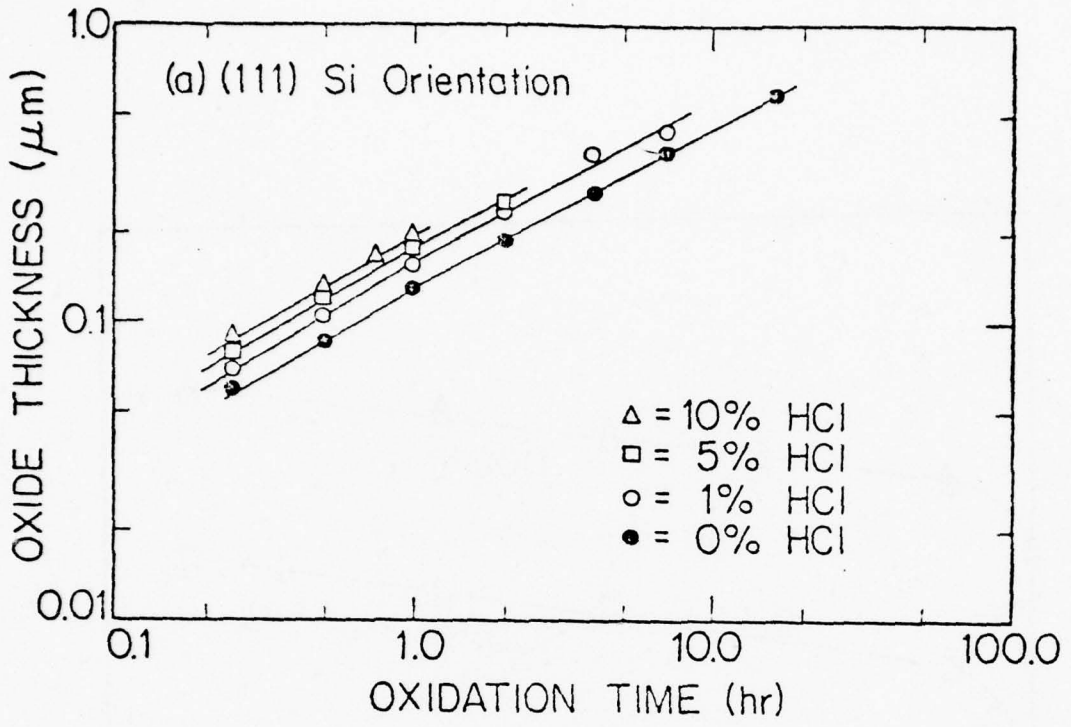


Fig. 2-10. Oxide Thickness vs. Oxidation Time for the Oxidation of <111> and <100> Oriented N-Type Silicon in Various O_2/HCl Mixtures at 1000°C.



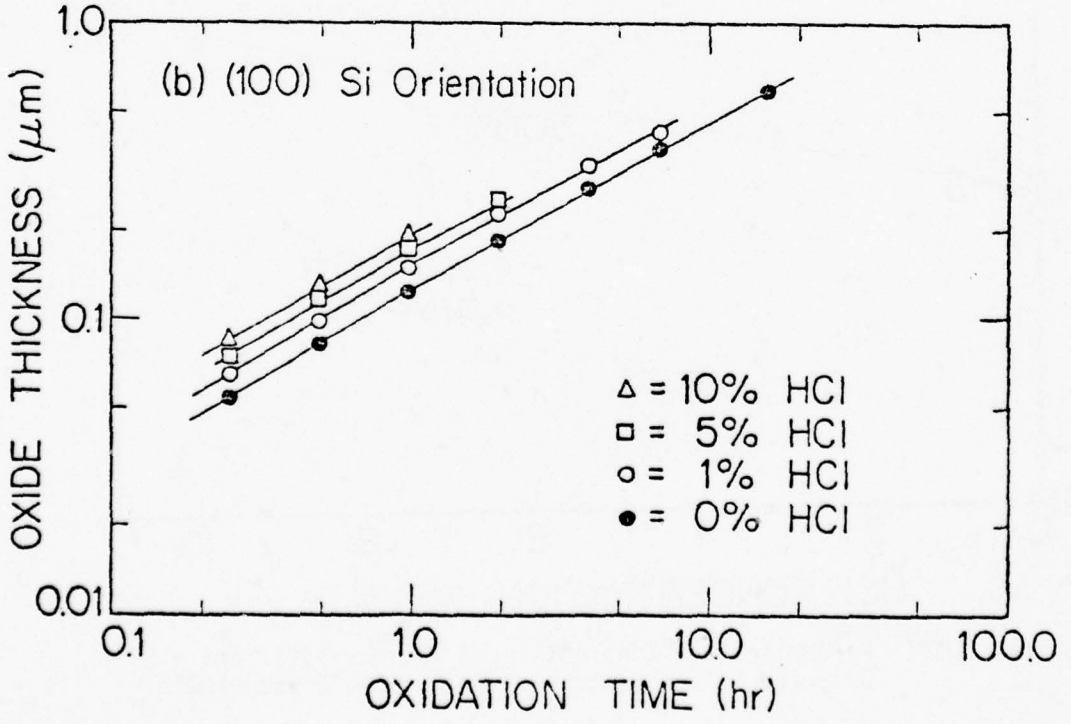


Fig. 2-11. Oxide Thickness vs. Oxidation Time for the Oxidation of <111> and <100> Oriented N-Type Silicon in Various O_2/HCl Mixtures at 1100°C.

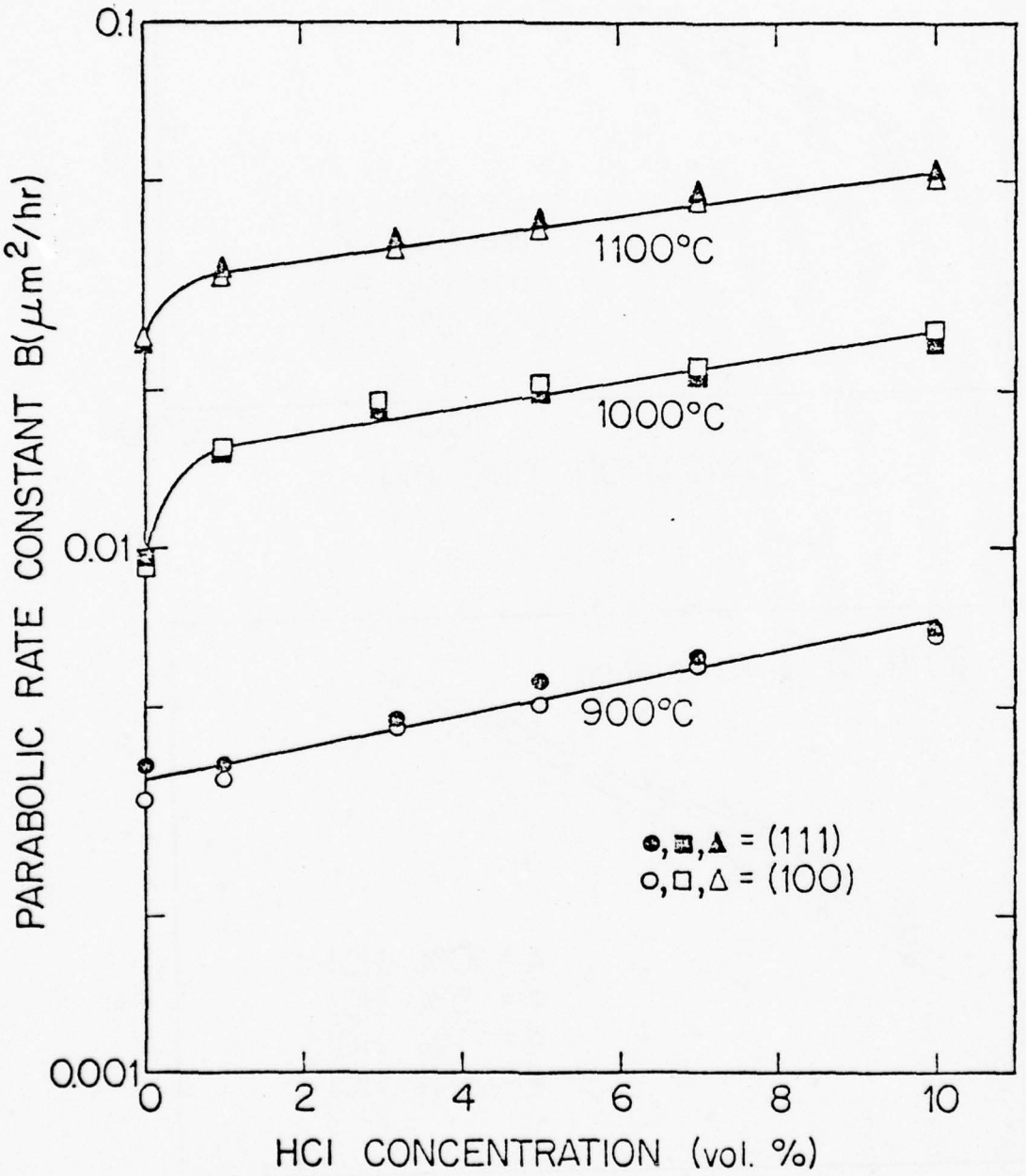


Fig. 2-12. Parabolic Rate Constant vs. % HCl for <111> and <100> Oriented N-Type Silicon at 900°, 1000°, and 1100°C.

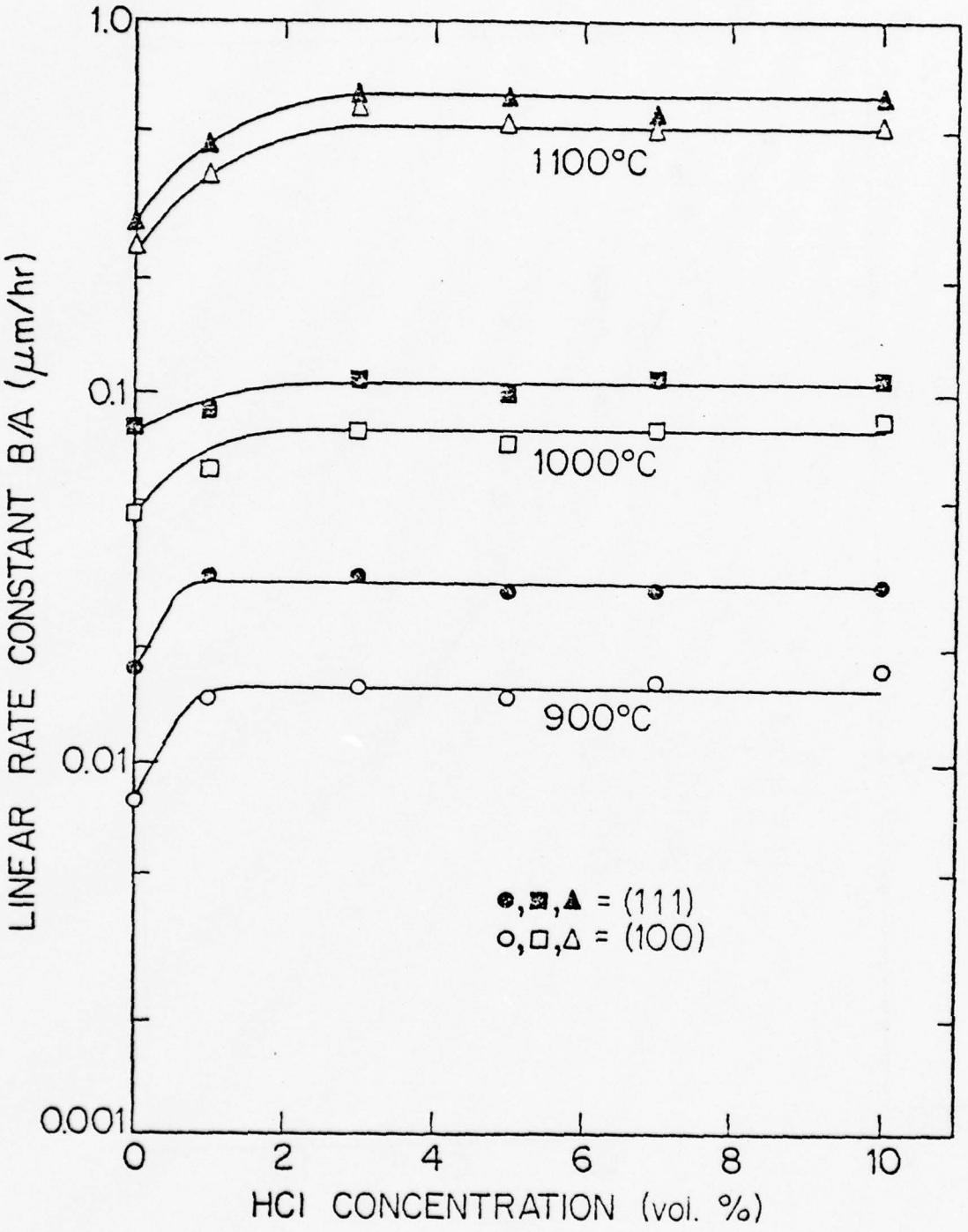


Fig. 2-13. Linear Rate Constant vs. % HCl for <111> and <100> Oriented N-Type Silicon at 900°, 1000°, and 1100°C.

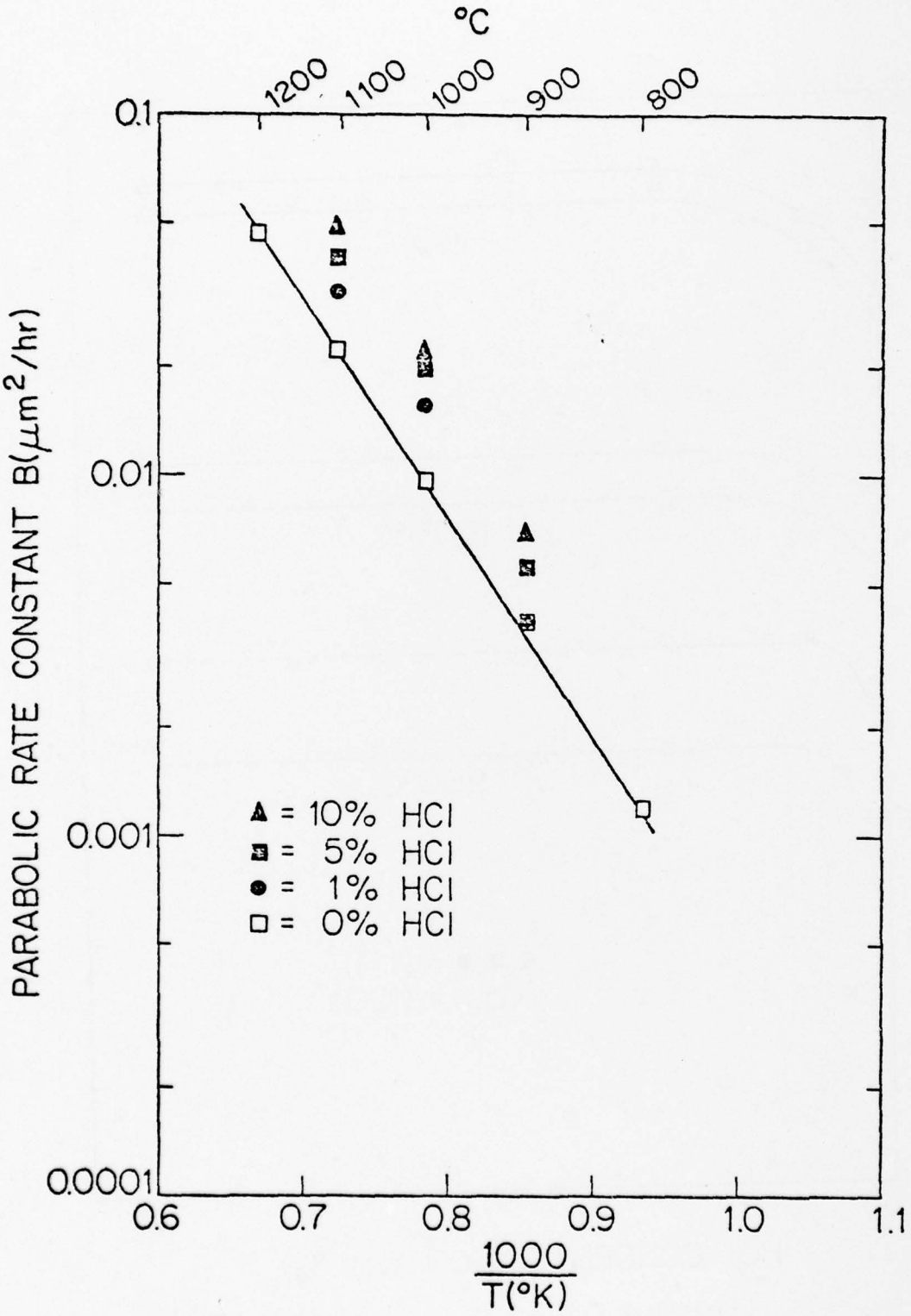


Fig. 2-14. Arrhenius Plots of the Parabolic Rate Constants for Silicon Oxidized in Various O_2/HCl Mixtures. Silicon Orientation is <111>.

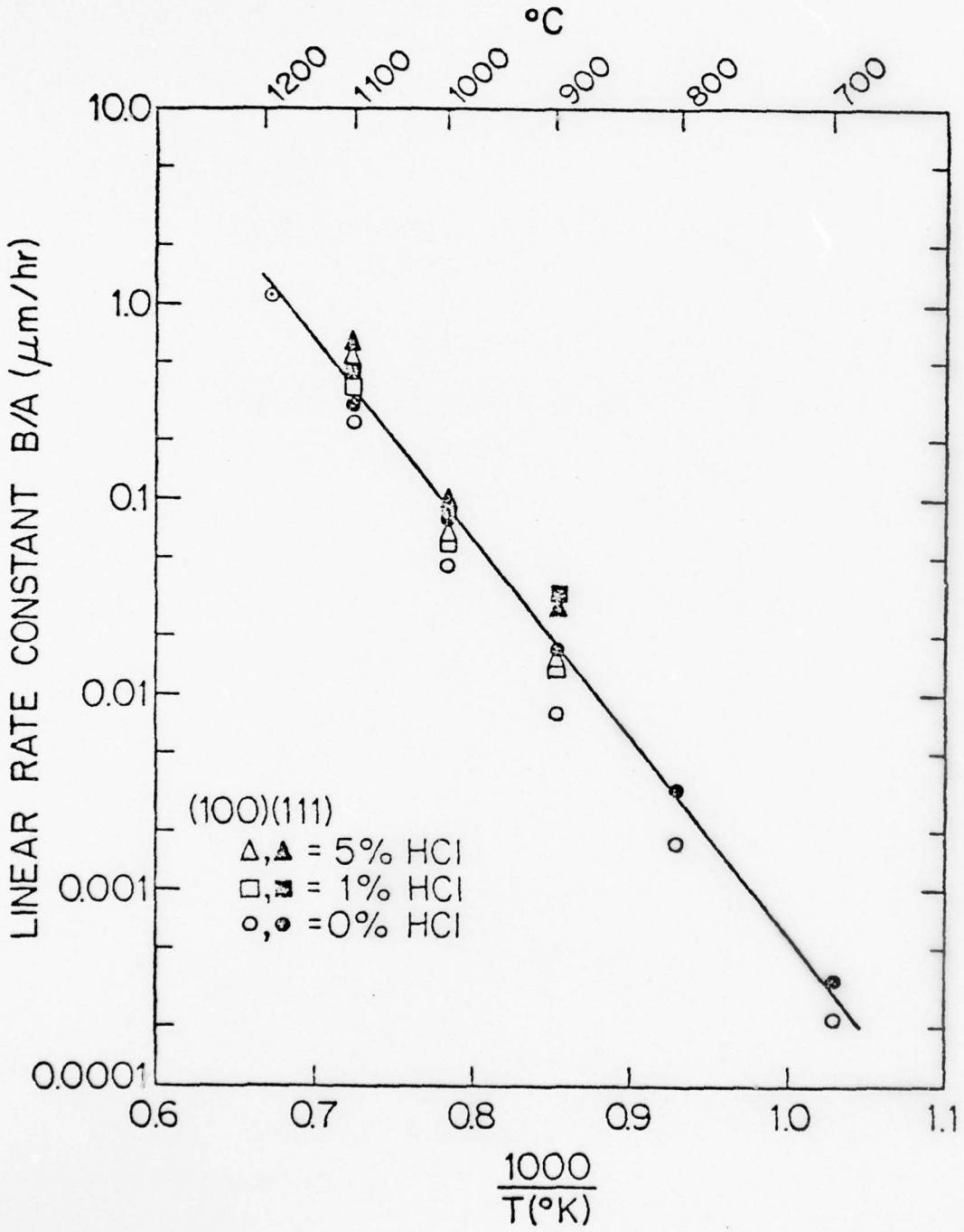


Fig. 2-15. Arrhenius Plots of the Linear Rate Constants for <111> and <100> Oriented Silicon Oxidized in Various O_2/HCl Mixtures.

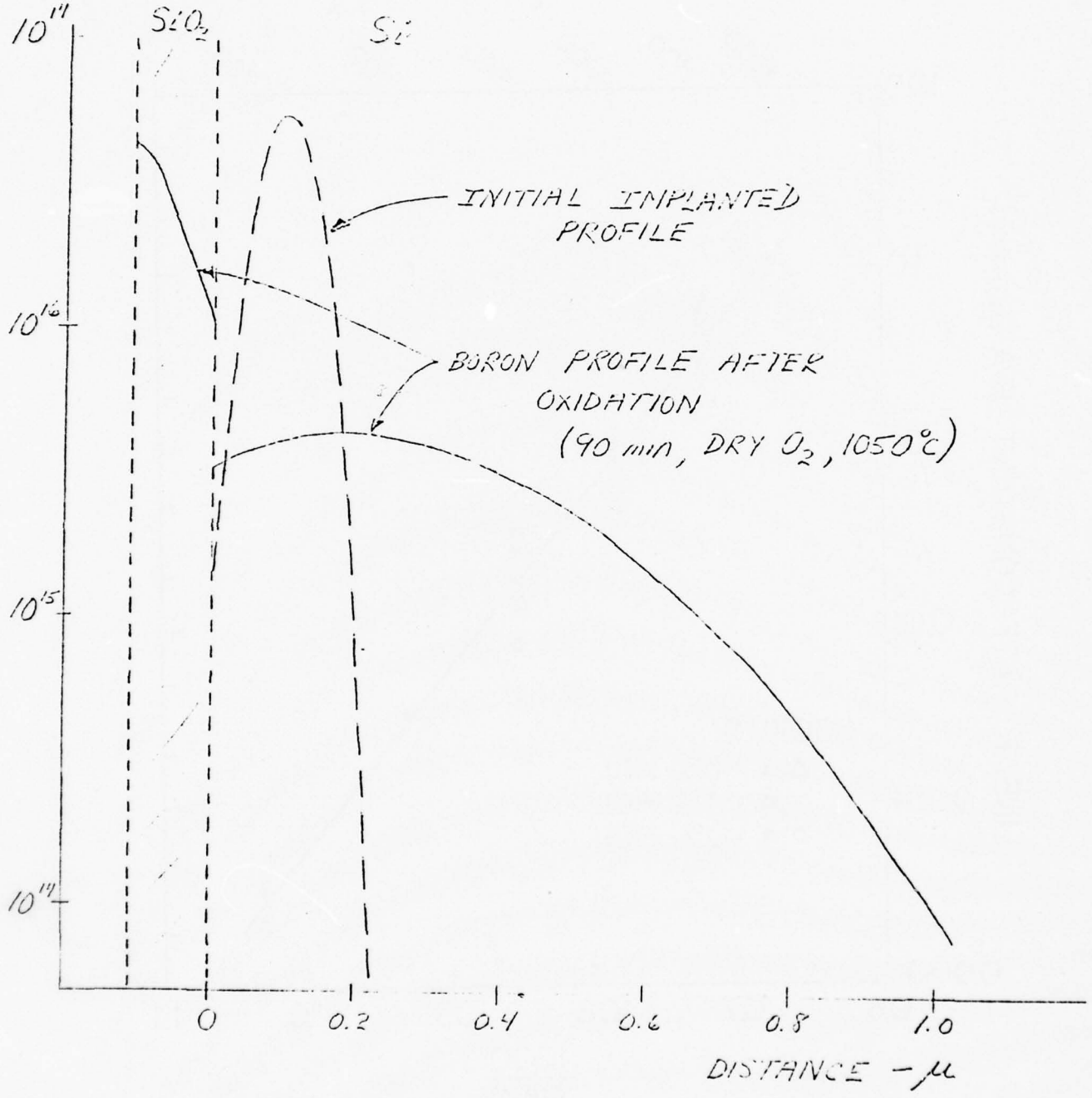


Fig. 2-16. Typical Output Generated by Moving Boundary Diffusion Program.

CHAPTER 3

Silicon Epitaxy

T. I. Kamins, K. C. Saraswat and R. Reif

3.1 Introduction

The third part of this program deals with modeling of epitaxial growth of silicon on a silicon substrate. Initially the work was divided into three areas:

1) Dopant profile measurement techniques
2) Kinetics of epitaxial growth
3) Kinetics of dopant inclusion in the epitaxial layer.

During the first six months of the program the attention was primarily focused on the first two areas (see the Semi-annual Report [1] for details). Preliminary work was begun in the third area during the remaining nine months of the program.

The dopant system of a horizontal silicon epitaxial reactor has been characterized by changing the dopant gas flow during the continuous deposition of epitaxial layers from silane. The different response for increasing and decreasing dopant gas flow indicates the importance of different physical mechanism in the two cases. A System "transfer function" has been found relating the dopant profile in the epitaxial layer (system output) to the time-varying dopant gas flow (system input). The transfer function allows the calculation of the dopant profile in the epitaxial film for any time varying dopant gas flow. The calculation of the dopant gas flow as a function of time required to achieve a desired dopant profile are thus possible.

3.2 Objective

All work presented here was done in the commercially available horizontal reactor illustrated in Figure 3.1. The major features are:

1) Water-cooled r.f. induction coils
2) Silicon-carbide-coated graphite susceptor (1/4" x 2-1/2" x 10")
3) Quartz susceptor cradle to tilt the susceptor to 2°
4) Quartz reactor tube (2-1/2" x 3-1/2" x 36")
5) The main gas manifold in which the mixing of the reactant gases occurs
6) Double-dilution doping system.

A horizontal reactor cross section is shown in Figure 3.2(a). It has been shown both theoretically and empirically that during epitaxial growth in a horizontal reactor, the gas stream above the susceptor divides into two parts, a flow region and a boundary layer [2]. The boundary layer of about 1/2 cm in thickness is next to the susceptor. There is no displacement gas flow in the boundary layer either vertically or horizontally across the wafer. The flow region has horizontal flow corresponding to the main flow displacement and also vertical flow due to thermal convection tumbling. The flow region is considered to be well mixed both chemically and thermally, and at a relatively low temperature. Consequently, the major portion of the temperature differential is across the boundary layer. Epitaxial growth proceeds by the following steps [3]:

1) Mass transfer of the reactant molecules (e.g., SiH_4) by diffusion from the flow region reservior across the boundary layer to the silicon surface.
2) Adsorption of reactant atoms on the surface.
3) One or more chemical reactions at the surface.

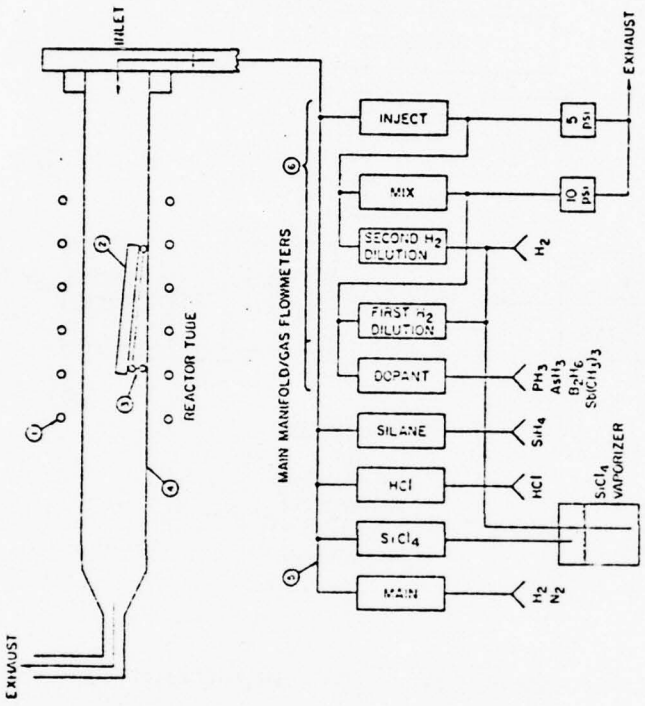


Fig. 3.1. Horizontal epitaxial reactor schematic.

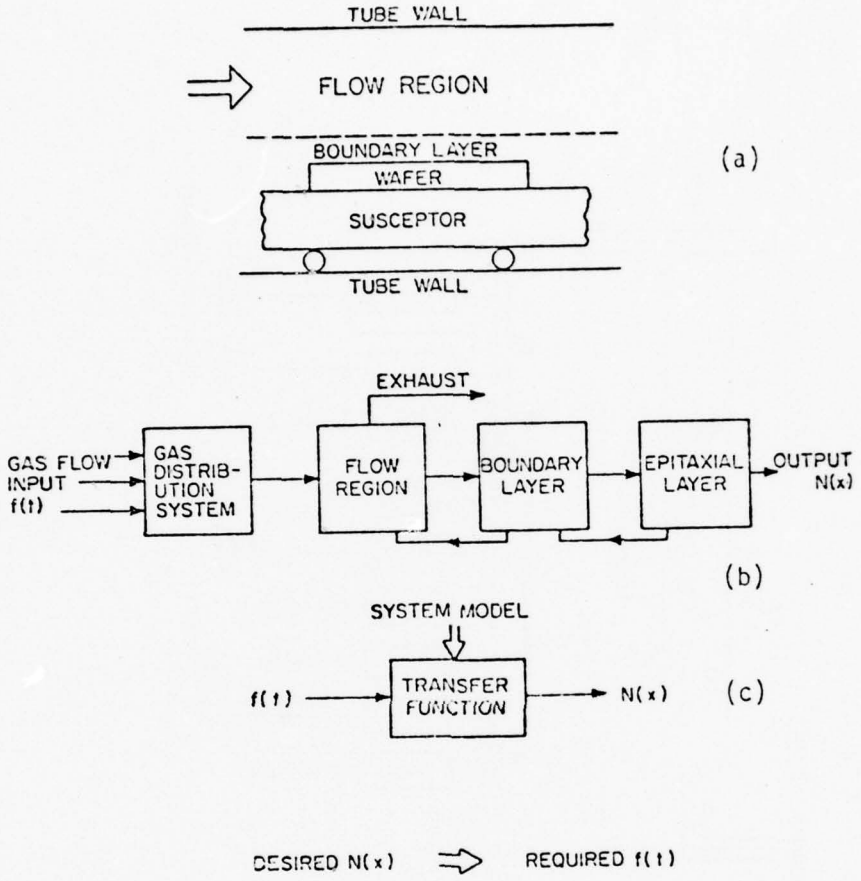


Fig. 3.2. (a) Cross sectional view of the reaction chamber
(b) Block diagram of the reactor as a system
(c) One block transfer function relating the dopant gas flow input f(t) to dopant concentration in the epitaxial film N(x).

4) Desorption of product molecules (e.g., H_2).

5) Mass transfer of the product molecules by diffusion through the boundary layer, back to the flow region.

6) Lattice arrangement of the adsorbed silicon atoms. This step may occur as part of (3).

Figure 3.2(b) shows a block diagram representation of the epitaxial reactor as a system. The input to the system being the gas flow as a function of time, f(t), and the output being the doping density profile in the epitaxial layer N(x). All the four blocks interact, as shown in Figure 3.2(b). As far as the doping mechanism is concerned, despite such complexity within limitations the reactor as a whole can be considered analogous to a linear system. (This will become evident later in this report). The implication of this is, that it is possible to obtain a transfer function of the reactor, relating the dopant gas flow imput f(t) to the doping profile output N(x) as shown in Figure 3.2(c). Once the transfer function has been obtained it is possible to calculate the dopant gas flow as a function of time required to achieve a desired dopant profile in an epitaxial film.

3.3 Transient Response of the Dopant Incorporation Mechanism

It is a well known fact that a linear system can be characterized by obtaining its transient response. Once the transient response is known it is a relatively easy task to determine the transfer function.

While such characterization of an epitaxial reactor has not been attempted in the past, a brief literature survey indicates that the characterization of transient doping effects is promising. For example, diffusion-rate processes in catalytic reactors have been characterized by "admittance functions" analogous to ac electrical parameters [4]. In this work, the time-varying output-gas concentrations were related to the time

variations of the input gases. Further work has considered diffusion through a boundary layer, adsorption on a surface, and diffusion into a solid [5]. Several of the same kinetic processes are present in our study of dopant in epitaxial layers. In our study, however, additional chemical processes are involved since the output is not a time-varying gas concentration, but a variation of the solid dopant concentration as a function of position in the deposited epitaxial layer. The earlier work does, however, indicate that the proposed approach is promising.

In addition to providing a method of fabricating a desired dopant profile, this portion of the study will produce basic information concerning the various mechanisms involved in the dopant inclusion processes. As indicated by Kobayashi and Kobayashi [6], transient studies may reveal the importance of various mechanisms, which may not be apparent in steady-state studies. In particular, consideration of limiting cases may allow separation of the several mechanisms involved; e.g., dopant-gas flow without film growth would involve only a selected number of the mechanisms involved in the total epitaxial doping process.

Thus, the study will allow the twin goals of the ARPA program to be addressed: The initial results will allow the calculation of dopant flow necessary for a desired dopant profile, which can be used in specialized devices. The further study of the mechanisms entering into the dopant-incorporation process will allow more detailed modeling and understanding and may suggest better methods of obtaining the desired dopant profile by revealing the limiting mechanisms.

Using the results of the work described in the Semi-annual Report [1], the reactor was first optimized for a nominal deposition of approximately

0.6 microns/min and uniform doping during the entire deposition process. Silane was used as the source of Si and deposition was done at 1070°C. Arsine was used as the dopant gas and flow settings were found for the typical partial pressure of approximately 2×10^{-6} and 6×10^{-6} atm., corresponding to dopant concentrations of roughly 1×10^{15} and 3×10^{15} cm^{-3} in the epitaxial layer. The substrates were (100) oriented silicon wafers with phosphorous dopant concentrations in the 10^{15} cm^{-3} range. The epitaxial layers were grown with a step-function change in the dopant gas flow during the continuous deposition, i.e. the dopant gas flow was changed from one of the above described to the other one during the deposition to simulate the step input. Both the increasing and decreasing steps were used in these experiments.

The dopant profiles in the epitaxial layers were obtained by capacitance-voltage measurements on deep-depletion MOS structures, and planar p-n junctions and mesa p-n junction. The thickness and dopant concentrations used in the samples were chosen to be compatible with the C-V technique. Spreading-resistance measurements and stacking-fault thickness determinations were used to confirm the capacitance-voltage measurements. Figures 3.3 and 3.4 shows the resulting dopant profiles in the two cases. In order to ensure that the experimental profiles were not limited by the resolution of the capacitance-voltage data-reduction technique, a theoretical capacitance-voltage curve was generated by solving Poisson's equation [7]. The data-reduction program was then applied to this curve. The results showed that the experimental profiles varied slowly compared to the resolution possible, indicating that the major features observed were not artifacts of the analysis technique.

From Figures 3.3 and 3.4, we see that the transition from one dopant concentration to the other occurred in approximately 1.2 microns, corresponding

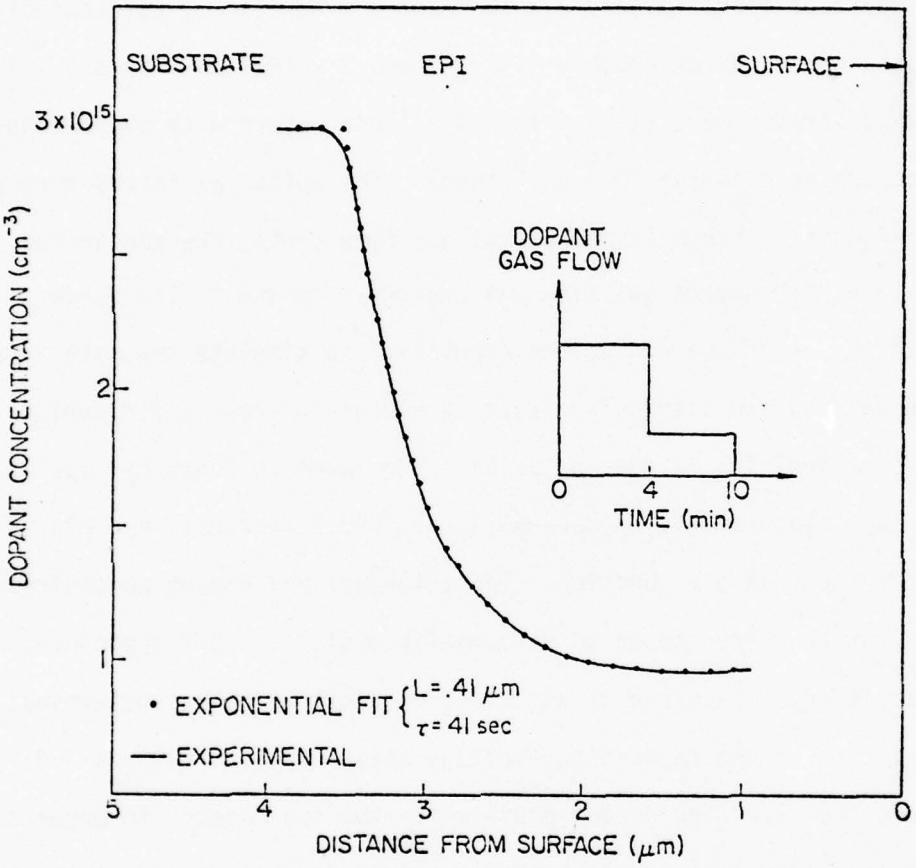


Fig. 3.3. Experimentally observed dopant concentration as a function of distance from the surface of the epitaxial film for a decreasing step change in the dopant gas flow, as indicated in the inset. Also shown is an exponential theoretical fit to the experimental curve.

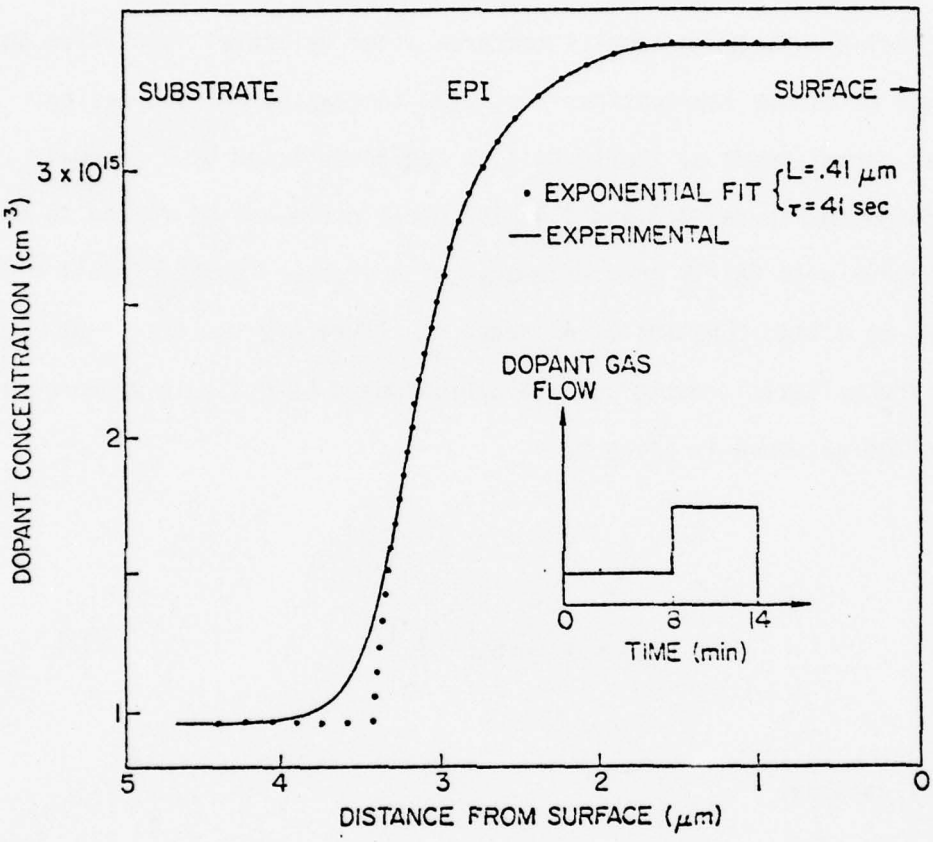


Fig. 3.4. Experimental dopant profile for an increasing step change in dopant gas flow and exponential fit to the experimental profile.

to a time of roughly two minutes. The heat cycling during the fabrication of the MOS and p-n junction C-V samples produced a square-root-of-Dt of only 0.07 microns because of the low diffusivity of the arsenic dopant in the epitaxial layer. Spreading resistance data confirm that there is no significant deviation between samples measured after epitaxial deposition and after complete CV sample fabrication. Both the increasing and decreasing steps show an abrupt onset of the transition region followed by a gradual decay. As shown in Figures 3.3 and 3.4, the decay curve can be fitted to an exponential curve with fairly good accuracy, with a decay length of 0.41 microns, corresponding to a time constant of 41 seconds. Therefore the transient response of the epitaxial reactor can be approximated by a single exponential function and the response is given by

$$h(t) = 1 - \exp\left(-\frac{t}{T_1}\right) \qquad 3.1$$

where T_1 = 41 seconds.

3.4 Systems Analysis

Once the transient response of a system has been characterized, it is possible to do further analysis to predict output for a certain input. There are several numerical techniques for doing this. One of the simplest methods is to use the convolution integral [8].

For a linear system the output g(t) to an input f(t) can be calculated by convolving the input with the impulse response I(t)

$$g(t) = \int_{-\infty}^{\infty} I(\tau)\, f(t-\tau)\, d\tau \qquad 3.2$$

The impulse response can be determined by taking the time derivative of the step response h(t), $I(t) = (\partial h(t))/(\partial t)$. Since h(t) is zero for negative values of t, $I(\tau)$ is also zero for negative values of τ. Hence the lower limit of integration can be simplified

$$g(t) = \int_0^t I(\tau)\ f(t-\tau) \qquad 3.3$$

The limits of integration can be further simplified depending upon the form of the input. For example if f(t) = 0 for t 0, then f(t-) is zero for t and thus the upper limit of the integration becomes t

$$g(t) = \int_0^t I(\tau)\ f(t-\tau)\ d\tau \qquad 3.4$$

Equation 3.3 or 3.4 can be easily evaluated either analytically or numerically by using any of the popular numerical integration methods [9].

Similar calculations can be done in the frequency domain [8] using either the Laplace Transform method or Fourier Transform method. The former has the advantage that it is possible to draw analogs to electrical circuits, whereas the latter has the advantage that numerical methods are very efficient, e.g. the fast Fourier transform. In both of these techniques repeated transforms have to be taken to convert the information from time domain to frequency domain and back to time domain.

In the case of the epitaxial reactor, by differentiating Equation 3.1 we get

$$I(t) = \frac{1}{\tau_1}\ \exp\left(-\frac{t}{\tau_1}\right) \qquad 3.5$$

Since the growth rate, G, is constant in time the doping density, N(x), is

given by

$$N(x) = Gg(t) \qquad 3.6$$

combining equation 3.4, 3.5 and 3.6 we get

$$N(x) = \frac{G}{T_1} \int_0^{\infty} \exp\left(-\frac{\tau}{T_1}\right) f(t-\tau)\, d\tau \qquad 3.7$$

As explained earlier the limits of integration can be simplified depending upon the type of f(t).

3.5 Verification of the Approach

In a series of experiments the dopant gas flow consisted of an increasing step followed by a decreasing step, approximating a pulse in the input. Four different pulse widths were used - 4.3, 3, 1.8, and 0.8 minutes. The resulting doping profiles in the substrates were measured using the C-V techniques. Figure 3.5 shows the results of measurements. For pulse widths of 4.3 minutes and 3 minutes the resulting profiles seem to be a simple superposition of profiles obtained in Figures 3.3 and 3.4, because the pulse widths are longer compared to the system time constant (41 seconds). However, when using the shorter pulses of 1.8 minutes and 0.8 minutes, the higher limit of doping was never reached, indicating the limits of the deposition system in responding to arbitrary changes in the rate of dopant gas flow.

In an effort to verify the validity of the transfer function approach, doping profiles were calculated using Equation 3.7 for the four pulse inputs. The results are plotted in Figure 3.5(a) - 3.5(d). There seems to be an excellent agreement between the theory and the experiments except at the onset of the high-to-low or low-to-high region. Even for the shortest pulse, where the pulse width of 48 seconds is comparable to the system time constant

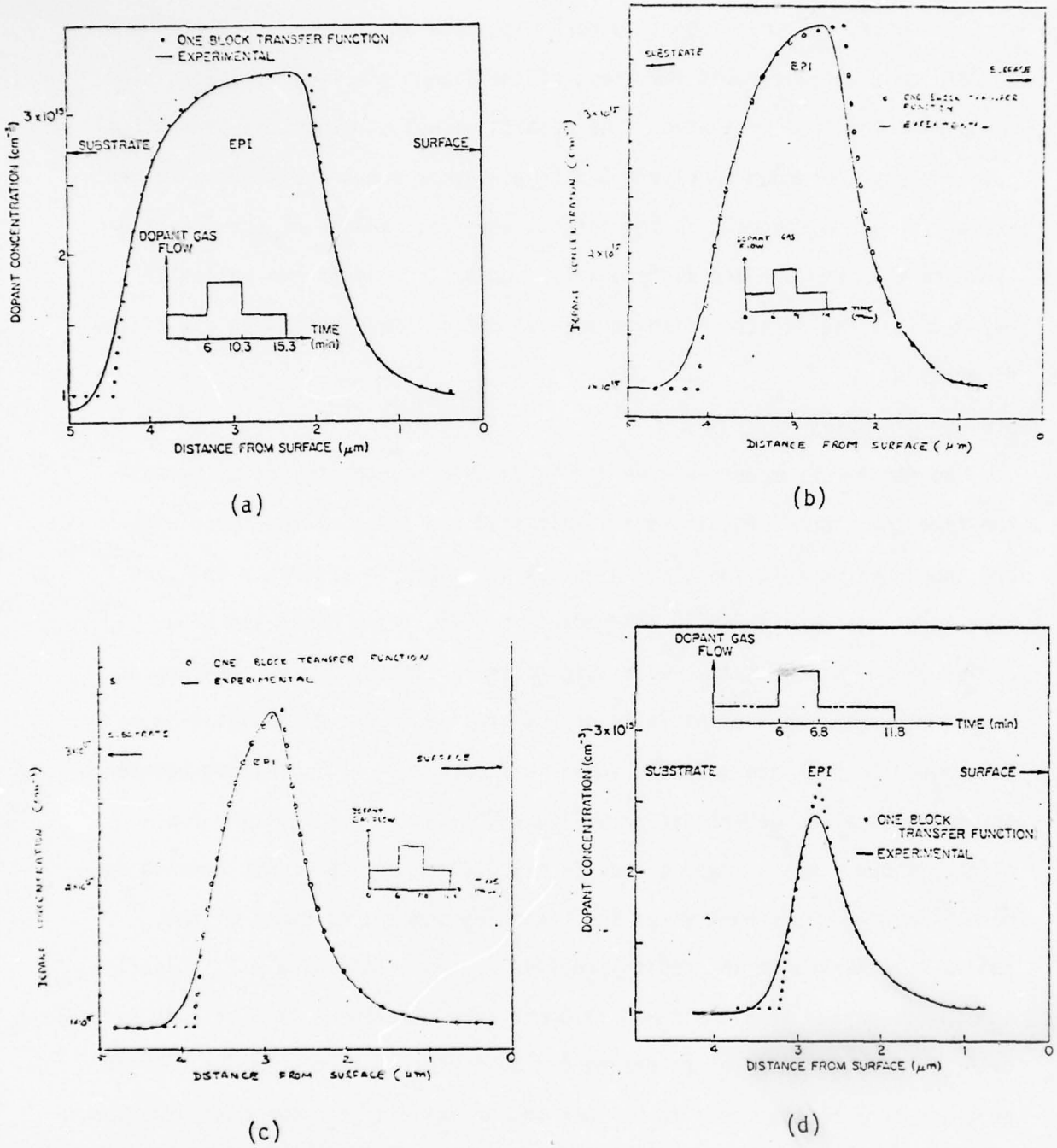


Fig. 3.5. Response of the reactor to pulse inputs in the dopant gas flow. The pulse width in time is (a) 4.3 min (b) 3 min (c) 1.8 min and (d) 0.8 min.

of 41 seconds, the model seems to work with good accuracy.

In another experiment the shape of the dopant gas flow input function was a ramp followed by a step. The results of measurements and theoretical calculations are shown in Figure 3.6, which again shows excellent agreement. Except for one discrepancy - the initial and final values of the dopant gas flow are the same but are different for doping profile in the epitaxial layer. This can be attributed to some mechanical malfunction in one of the flowmeters.

3.6 System Delay

So far in our model we have not taken into account any delay between the time the flowmeter changes are initiated and the time the impurities are incorporated into the growing epitaxial film. To ascertain this, an experiment was done in which the dopant gas flow input consisted of a series of two 0.8 min wide pulses separated in time by 5 min. The dopant gas flow input, and the resulting dopant profile in the epitaxial film as determined by C-V technique are shown in Figure 3.7. The distance between the two pulses was determined from Figure 3.7 and, since the time was already known, this allowed a precise calculation of the growth rate to be made. Knowing the growth rate, the beginning and ending times of the pulses were marked on the doping profiles as shown in Figure 3.7. Clearly the points marked as 6.8 min and 11.8 min show that there is little appreciable delay in the system. The points marked by 6 min and 11 min do not confirm anything because they seem to be lost in the majority carrier concentration profile. Another possible explanation of this portion of the profile will be attempted in the next section.

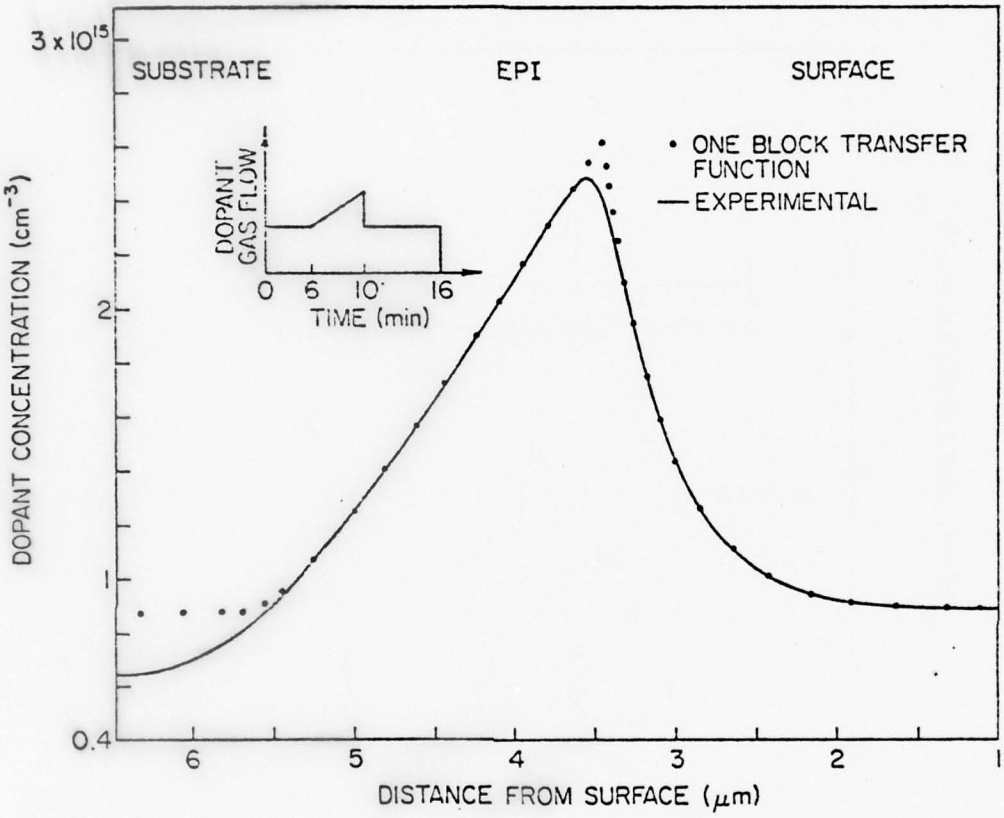


Fig. 3.6. Response of the reactor to a ramp followed by a decreasing step input in the dopant gas flow.

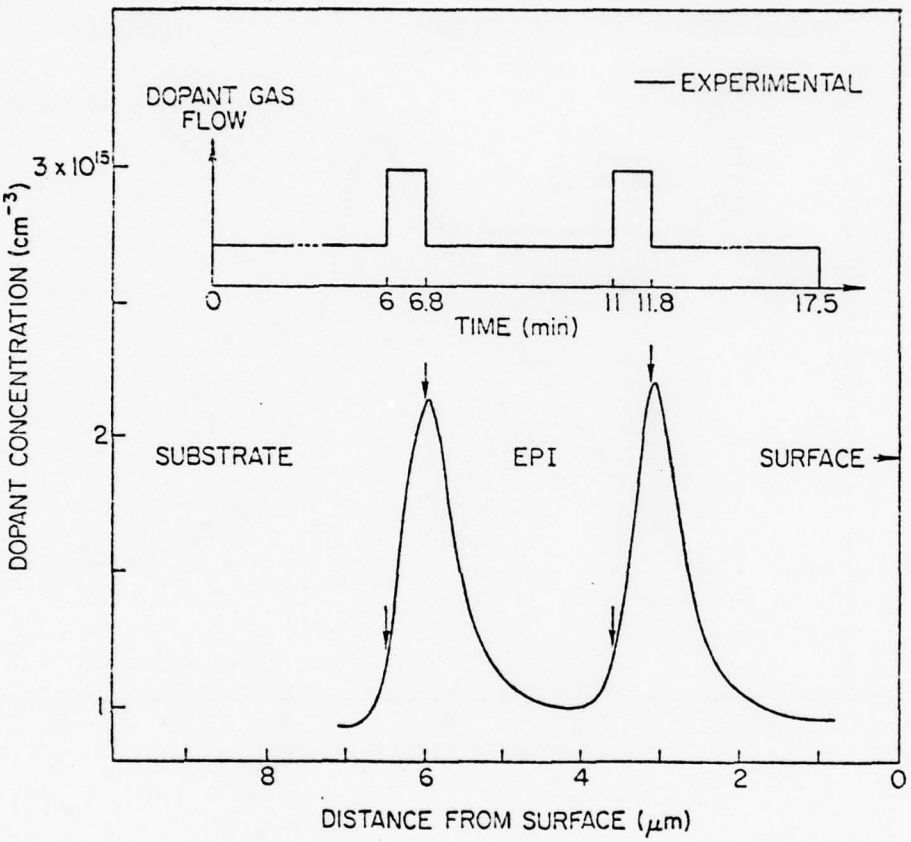


Fig. 3.7. Reactor response to a sequence of two pulses, each 0.8 min wide.

3.7 A Two-Block Model

As we have discussed earlier the one block, exponential transfer function, although adequate, does not explain some of the finer features of the profiles discussed earlier. To cover the finer features of Figures 3.3 through 3.7, a two-block model was attempted, the concept of which is shown in Figure 3.8, where the reactor is modeled by a combination of two transfer functions. The impulse response of the two blocks are I_1 (t) and $I_2(2)$. Therefore the output N(x) for an input f(t) is given by

$$N(x) = G[f(t) * I_1(t) * I_2(t)] \qquad 3.8$$

(the sign * represents convolution)

$I_1(t)$ is still given by Equation 3.5, however, $I_2(t)$ was chosen to be an error function.

$$I_2(t) = \left(\frac{N_2 - N_1}{2}\right)\left[1 + \text{erf}\left(\frac{t - t_0}{T_2}\right)\right] \qquad 3.9$$

where constants N_1 and N_2 represent the upper and lower limits of the dopant concentration in the epitaxial film, T_2 is a time constant and t_o is the time when the step change is initiated in the gas flow input.

Using Equation 3.8, an effort was made to calculate the step response of the reactor to the increasing step, with $N_1 = 10^{15}$ cm^{-3}, $N_2 = 3.5 \times 10^{15}$ cm^{-3}, t_o = 8 min. Out of several values tried T_2 = 30 seconds gave the best results, Figure 3.9 shows the results of these calculations and the C-V measurements (as shown earlier in Figure 3.4). The agreement seems to be excellent. Thus, the rounding of the dopant profile curve when an abrupt change is initiated can be modeled by incorporating an added block in the model where the overall

impulse response is given by

$$I(t) = I_1(t) * I_2(t) \qquad 3.10$$

The validity of this model has yet to be confirmed on other types of profiles, however it appears to be promising.

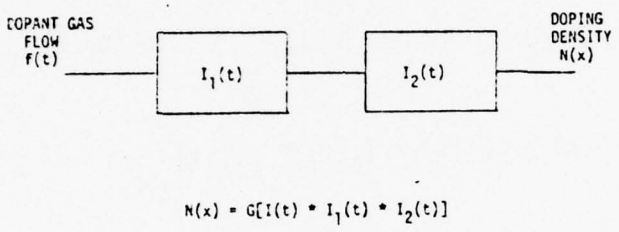


Fig. 3.8. Concept of the two block model of the epitaxial reactor.

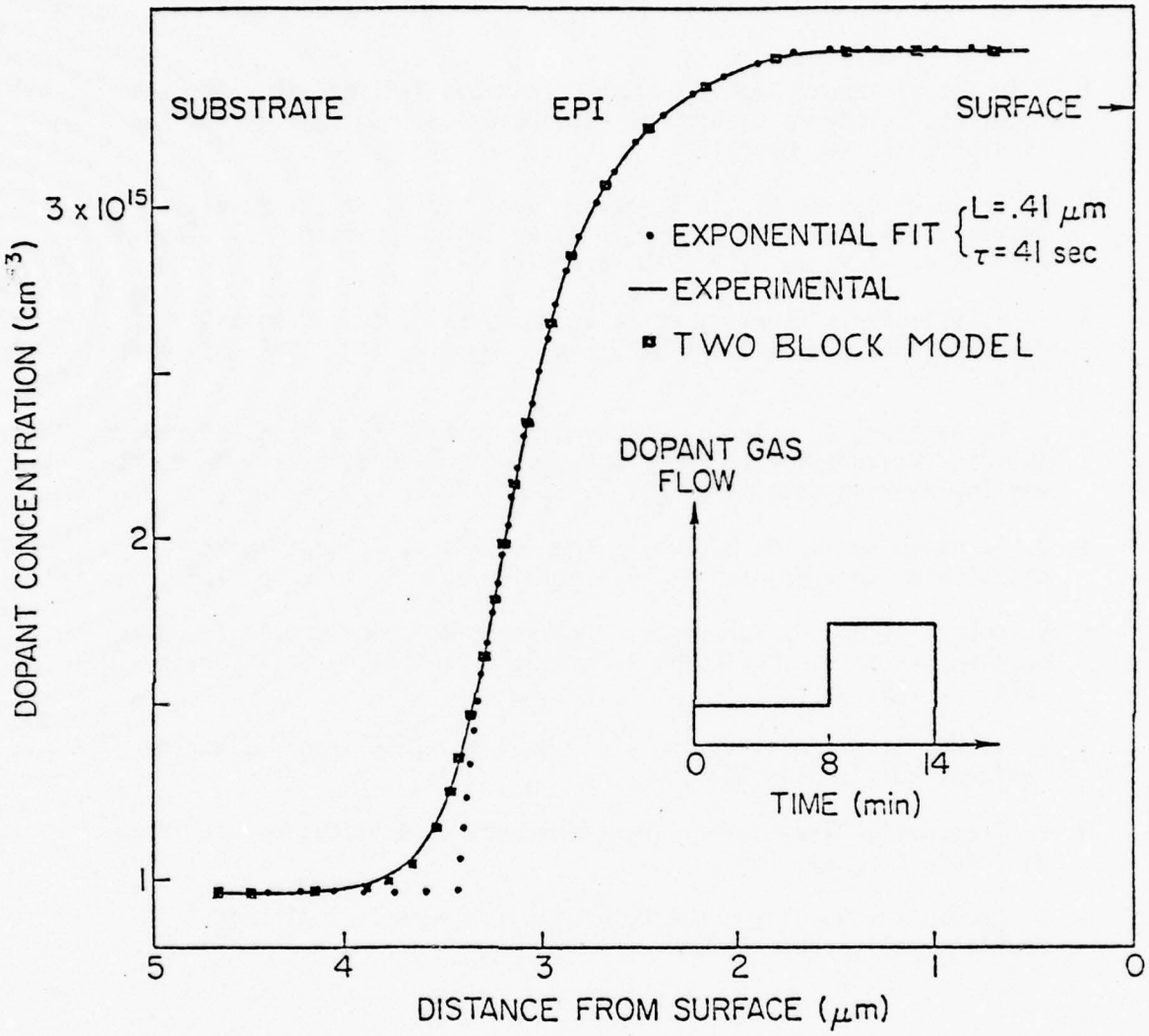


Fig. 3.9. Comparison of the one block model (exponential fit) and the two block model for an increasing step change in dopant gas flow.

References

1 "First Semi Annual Report on Computer-Aided Semiconductor Process Modeling," Stanford Electronics Laboratories Technical Report No. TR-4969-73-1, December 1975.

2 F. C. Eversteyn et al, "A Stagnant Layer Model for the Epitaxial Growth of Silicon from Silane in a Horizontal Reactor," J. Electrochem. Soc., 117, 7, July 1970, pp. 925-931.

3 E. G. Bylander, "Kinetics of Silicon Crystal Growth from SiCl Decomposition," J. Electrochem. Soc., 109, 12, Dec. 1962, pp. 1171-1175.

4 P. F. Deisler, Jr., and R. H. Wilhelm, "Diffusion in Beds of Porous Solids: Measurement by Frequency Response Techniques," Industrial and Engineering Chemistry, 45, 1953, p. 1219.

5 J. B. Rosen and W. E. Winsche, "The Admittance Concept in the Kinetics of Chromatography," J. Chem. Phys., 18, 1950, p. 1587.

6 H. Kobayashi and M. Kobayashi, "Transient Response Method in Heterogeneous Catalysis," Catalysis Reviews--Science and Engineering, 10, 1974, p. 139.

7 C. T. Wu, E. C. Douglas, and C. W. Mueller, IEEE Trans. Electron Devices, ED-22, 319 (1975).

8 R. Bracewell, "The Fourier Transform and its Applications," McGraw-Hill Book Company, 1965.

9 G. Dahlquist, A. Bjorck and N. Anderson, "Numerical Methods," Prentice Hall, Inc., 1974.

APPENDIX A - CALCULATION OF AN EFFECTIVE DIFFUSION COEFFICIENT

Equations A1 and A2 are equations 1.1(a) and 1.1(c) in Section III rewritten in terms of fluxes, J_B and J_{BV}.

A1
$$\frac{dC_B}{dt} = \frac{dJ_B}{dx} + \left(\frac{C_{BV} - k_o C_B C_{V+}}{\tau}\right)$$

A2
$$\frac{dC_{BV}}{dt} = \frac{dJ_{BV}}{dx} - \left(\frac{C_{BV} - k_o C_B C_{V+}}{\tau}\right)$$

Addition of equation A1 and A2 yields:

A3
$$\frac{d}{dt}\ (C_B + C_{BV}) = \frac{d}{dx}\ (J_B + J_{BV})$$

Expressing the fluxes in terms of concentration gradients and diffusion coefficients we obtain equation A4.

A4
$$J_B + J_{BV} = D_B\ \frac{\partial C_B}{\partial x} + D_{BV}\ \frac{\partial C_{BV}}{\partial x}$$

For this equation we can show that the last term dominates as follows: D_{BV} is much greater than D_B, at 900°C; their values are 7×10^{-14} cm^2/sec and 1.6×10^{-21} cm^2/sec respectively. At the outset of the anneal, $C_{BV} > C_B$ and $\frac{\partial C_{BV}}{\partial x} > \frac{\partial C_B}{\partial x}$, hence we can neglect the first term. As the annealing proceeds the diffusion of Boron will approach ordinary diffusion conditions, namely,

$$\left(\frac{\partial C_{BV}}{\partial x} \Big/ \frac{\partial C_B}{\partial x}\right) \sim \frac{C_{BV}}{C_B} \sim .02$$

The inequality between gradients has reversed; however because D_{BV} is much greater than D_B, the last term still overwhelms the first one. We can then

approximate the total flux as in equation 1.5.

A5 $$J_B + J_{BV} \simeq D_{BV} \frac{\partial C_{BV}}{\partial x}$$

We now let ξ be the fractional concentration of BV-pairs, which is a function of the electrical activity:

A6 $$\xi = \frac{C_{BV}}{C_{BV} + C_B} = 1 - \alpha$$

then:

A7 $$C_{BV} = \xi (C_{BV} + C_B)$$

and the gradient becomes:

A8 $$\frac{\partial C_{BV}}{\partial x} = \frac{\partial \xi}{\partial x} (C_{BV} + C_B) + \xi \frac{\partial (C_{BV} + C_B)}{\partial x}$$

We can now identify several situations in which the first term in equation A8 is negligible in comparison with the last one. For instance, under equilibrium conditions we can rewrite equation 1.3 in Section III as:

$$\frac{C_{BV}}{C_B} = k_o C_{V+}$$

Then substitution of the above equation in the expression of $1/\xi$ yields

$$\frac{1}{\xi} = 1 + \frac{C_B}{C_{BV}} = \frac{1}{1 + k_o + C_{V+}}$$

where:

$$C_{V+} = C_V o(T) \exp \frac{E_{V+} - E_F}{kT}$$

In general the Fermi level, as a function of the acceptor concentration, will be a function of distance; however for annealing temperatures near 900°C and the boron concentration below $\sim 10^{17}$ atoms/cm^3, this dependence is very weak, hence $C_{V^+} \simeq C_{V^\circ}$, $\partial \xi / \partial x \simeq 0$ and the first term in equation A8 is negligible. An analogous case arises when the annealing temperature is high, above 1100°C. In this case the Fermi level is fixed in the middle of the bandgap.

Under non-equilibrium conditions, the comparison of the terms in equation A8 can be performed numerically. For the particular annealing case in Section IV, the results of the calculation show that the approximation if equation A8 is correct under non-equilibrium and equilibrium conditions, failing only under <u>equilibrium conditions</u> when the boron concentration is near and above 10^{18} atoms/cm^3. In other words, when the boron concentration is high and the Fermi level is a function of distance, $\partial \xi / \partial x$ may no longer be small.

For cases in which $d\xi/dx$ may be neglected, we may then simplify equation (8) to

$$\frac{\partial C_{BV}}{\partial x} \simeq \xi \frac{\partial (_{BV} + C_B)}{\partial x} \qquad \text{A9}$$

Substitution of equation A9 in equation A5 and equation A5 in equation A3 then yields

$$\frac{d(C_B + C_{BV})}{dt} = (D_{BV} \cdot \xi) \frac{d^2(C_{BV} + C_B)}{\partial x^2} \qquad \text{A10}$$

where $(D_{BV} \cdot \xi)$ can be identified with the overall diffusion coefficient D_{exp}, hence:

$$D_{exp} = D_{BV} \xi = D_{BV}(1 - \alpha) \qquad \text{A11}$$

DAAB07-75-C-1344

APPENDIX B

MATHEMATICAL MODEL OF IMPURITY DIFFUSION

Alan H. Marshak
Electrical Engineering Department
Louisiana State University
Baton Rouge LA 70803

September 1976

Final Report for Period 1 July 1975--30 September 1976

Prepared for

ADVANCED RESEARCH PROJECTS AGENCY
1400 Wilson Boulevard
Arlington VA 22209

Stanford Electronics Laboratory
Department of Electrical Engineering
Stanford University
Stanford CA 94305

MATHEMATICAL MODEL OF IMPURITY DIFFUSION

The term diffusion, when applied in semiconductor device fabrication, is used loosely to describe the motion of impurity atoms in semiconductors at elevated temperatures. The diffusion of impurities has been the subject of considerable work. However, it has been known for many years that diffusion in silicon at high concentrations produce impurity profiles that differ significantly from those predicted by simple theory. Thus there is still a need to be able to accurately determine impurity profiles in modern semiconductor devices and integrated circuits.

The object of this research is to develop a mathematical model for the diffusion of impurities into silicon for both a constant source and drive-in diffusion process. The model considers the influence of the internal electric field on the motion of impurity ions at elevated temperatures. The electric field is due to charge density produced by the ionization of impurities present in the material.

1. Introduction

Diffusion in semiconductor material is an ambipolar process involving mobile holes, electrons and ionized impurity atoms [1,2]. In the fabrication of modern devices and integrated circuits, plane parallel structures are of primary importance. We thus restrict the analysis to diffusion flow in one dimension. Furthermore, in the interest of simplicity, we assume all impurities to be ionized. The impurities can be described by a flux equation of the form

$$f(x,t) = -D_c \frac{\partial c}{\partial x} + Z\mu_c cE \tag{1}$$

where $c(x,t)$ represents the concentration of an arbitrary impurity species, D_c and μ_c denote the diffusion constant and mobility, respectively, and $Z = 1$ if c represents a donor impurity and $Z = -1$ if c represents an acceptor impurity. The particles must also satisfy a continuity equation of the form

$$\frac{\partial c}{\partial t} + \frac{\partial f}{\partial x} = 0 \tag{2}$$

Substitution of (1) into (2) yields the transport equation which determines the ionized impurity distribution in the wafer; thus

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(D_c \frac{\partial c}{\partial x} - Z\mu_c cE\right) \tag{3}$$

Holes and electrons also satisfy similar equations and it follows that

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}\left(D_p \frac{\partial p}{\partial x} - \mu_p pE\right) - R \tag{4}$$

$$\frac{\partial n}{\partial t} = \frac{\partial}{\partial x}\left(D_n \frac{\partial n}{\partial x} + \mu_n nE\right) - R \tag{5}$$

Gauss' law is used to relate the field to the charge density

$$\frac{\partial E}{\partial x} = \frac{e}{\varepsilon}\left(p - n + Zc\right) \tag{6}$$

The net recombination term is represented by

$$R = \frac{pn - n_i^2}{\tau_n(p + n_i) + \tau_p(n + n_i)} \tag{7}$$

where τ_n and τ_p are the electron and hole lifetimes. These equations, together with appropriate boundary conditions, constitutes the complete general problem. Although the transport of carriers is well defined, no analytic solution appears possible.

2. Technical Approach

The body of this work is concerned with the solution of the general problem under one assumption. The time derivative of holes and electrons is sufficiently small compared to other associated terms that they can be neglected. This is a valid assumption because holes and electrons have a much larger mobility than impurity ions. Thus as the impurity ions change during a diffusion, the holes and electrons readjust almost instantaneously, staying in a steady-state determined by the impurity ion distribution. Under this approximation the problem can be simplified to a form where numerical analysis is both accurate and efficient.

2.1 Classical Case. When the material remains nondegenerate so that the Einstein relation $D/\mu = \lambda$ remains constant, it is easy to show that under the above assumption $pn = n_i^2$ and thus $R = 0$. It is convenient to make the substitution

$$E = -\frac{\partial \phi}{\partial x} \tag{8}$$

where ϕ is the electrostatic potential. A given value of E will determine ϕ apart from an integration constant, which, without loss of generality, will be chosen zero for intrinsic material. Integrating (4) under these conditions yields

$$p = n_i e^{-\phi/V_T} \tag{9}$$

where $V_T = D_p/\mu_p = KT/e$. In a similar manner, (5) yields

$$n = n_i e^{\phi/V_T} \tag{10}$$

The hole and electron density can now be eliminated from the problem statement by substituting (9) and (10) into (6). Thus the quasi-static problem can be specified by

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(D_c \frac{\partial c}{\partial x} + Z\mu_c c \frac{\partial \phi}{\partial x} \right) \tag{11}$$

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{e}{\varepsilon}\left(2n_i \sinh \phi/V_T - Zc\right) \tag{12}$$

Since the wafer is very thick compared to the depth of diffusion, we treat the wafer as being semi-infinite with $x \geq 0$. The boundary and initial conditions for a constant source diffusion process are given by,

$$c(0,t) = C_o \tag{13a}$$

$$\frac{\partial \phi(0,t)}{\partial x} = 0 \tag{13b}$$

$$c(\infty,t) = 0 \tag{13c}$$

$$\phi(\infty,t) = 0 \tag{13d}$$

$$c(x,0) = 0,\ x > 0 \tag{13e}$$

and for a drive-in process, (13a) and (13e) are replaced by

$$\frac{\partial c(0,t)}{\partial x} = 0 \tag{13f}$$

$$c(x,0) = N(x),\ x > 0 \tag{13g}$$

The problem now is to solve the set of coupled equations given by (11) and (12), subject to the conditions given by (13).

2.2 Degenerate Case. When the doping concentration is sufficiently high the Einstein relation for the majority carrier no longer remains constant and the equations of the previous section do not apply. We assume a donor type impurity diffusion so that electrons represent the majority carrier. Integrating (5) under quasi-static conditions yields

$$\lambda_n \frac{\partial n}{\partial x} - n \frac{\partial \phi}{\partial x} = 0 \tag{14}$$

In order to obtain a relation between the carrier density and the electrostatic potential, we integrate again; thus

$$\phi = \int_{n_i}^{n} \lambda_n \frac{dn}{n} \tag{15}$$

Now the Einstein relation can be expressed by [3]

$$\lambda_n = \frac{KT}{e} \frac{F_{1/2}(\eta)}{F_{-1/2}(\eta)} \tag{16}$$

where $\eta = (E_{fn} - E_c)/KT$ and $F_{1/2}(\eta)$ represents the Fermi integral of order

1/2. Since

$$n = N_c F_{1/2}(\eta) \tag{17}$$

the Einstein relation can be rewritten as

$$\lambda_n = \frac{KT}{e} n \frac{d\eta}{dn} \tag{18}$$

By using (18) in (15) we obtain

$$\phi = \frac{KT}{e} (\eta - \eta_i) \tag{19}$$

Thus the electron density can be written as a function of the electrostatic potential. By using (9) and (17) in (6), we have

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{e}{\epsilon} [N_c F_{1/2}(\eta) - n_i e^{-\phi/V_T} - c] \tag{20}$$

where $\eta = e\phi/KT + \eta_i$. In this case, the quasi-static problem is specified by (20) coupled with

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x} (D_c \frac{\partial c}{\partial x} + \mu_c c \frac{\partial \phi}{\partial x}) \tag{21}$$

For an acceptor type impurity, a similar set of equations result.

3. Analysis

The quasi-static problem is specified by two non-linear partial differential equations in terms of impurity density and electrostatic potential. A solution is computed employing numerical techniques. Discretization of the time coordinate yields sets of ordinary differential equations. These are solved using a quasi-linearization technique. This process is iterated until sufficient accuracy is obtained.

3.1 Numerical Method. The problem can be further simplified by normalization of the variables. If $\hat{x}$, $\hat{t}$, $\hat{c}$ and $\hat{\phi}$ represent the normalized variables, then an appropriate normalization scheme is given by

$$c = n_i \hat{c} \tag{22a}$$

$$\phi = V_T \hat{\phi}/Z \tag{22b}$$

$$x = \sqrt{\frac{\varepsilon V_T}{e n_i}}\ \hat{x} \tag{22c}$$

$$t = \frac{\varepsilon V_T}{e D_c n_i}\ \hat{t} \tag{22d}$$

However, in order to simplify notation for this section, the normalized variables will be represented by c, ϕ, x and t. For concreteness, the equations describing diffusion for nondegenerate conditions will be used as the vehicle to explain the numerical method.

Substituting the normalized variables into (11) and (12) yields

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(\frac{\partial c}{\partial x} + c\,\frac{\partial \phi}{\partial x}\right) \tag{23}$$

$$\frac{\partial^2 \phi}{\partial x^2} = 2\sinh\phi - c \tag{24}$$

To facilitate numerical techniques, the independent variables were discretized, thereby giving sets of difference equations. These variables will be denoted now as t_i and x_j, defined by

$$t_i = (i-1)\Delta t + t_o,\quad i = 1, \ldots, n \tag{25a}$$

$$x_j = (j-1)\Delta x,\qquad j = 1, \ldots, m \tag{25b}$$

where Δt and Δx were chosen to be fixed for simplicity.

A non-zero starting time t_o was chosen due to the inability to numerically handle the abruptness of the initial condition at the surface. The particular choice of diffusion profile used at t_o varied and will be discussed later.

The time derivative was approximated by a two-point implicit scheme,

$$\frac{\partial f(t_i)}{\partial t} \simeq \frac{f(t_i) - f(t_{i-1})}{\Delta t} \tag{26}$$

Evaluating (23) and (24) at $t = t_i$ yields

$$\frac{d^2\phi}{dx^2} - 2 \sinh \phi + c = 0 \tag{27a}$$

$$\frac{d}{dx}\left[\frac{dc}{dx} + c \frac{d\phi}{dx} \right] - \frac{c - N}{\Delta t} = 0 \tag{27b}$$

where the notation

$$c = c(x,t_i) \tag{28a}$$

$$N = c(x,t_{i-1}) \tag{28b}$$

has been used. Notice that this constitutes a set of ordinary differential equations with a driving function N and boundary conditions

$$c(0) = \beta = C_o/n_i \tag{29a}$$

$$\frac{d\phi(0)}{dx} = 0 \tag{29b}$$

$$c(\infty) = 0 \tag{29c}$$

$$\phi(\infty) = 0 \tag{29d}$$

Thus the set of PDE is reduced to a set of ODE by using a two-point formula for the time derivative. At a given time t, the set of ODE in matrix form $f(\overline{y},\overline{y}',\overline{y}'',x) = 0$ is solved, where the vector $\overline{y}$ represents the dependent variables c and ϕ, primes indicate derivatives and x is the independent variable. This set of equations is reduced to successive approximate sets of linear equations which can be more easily solved. The linearization technique is based on the Newton approximating technique for

finding roots of an arbitrary function. After linearization, the set of ODE is of the form

$$A\Delta\bar{y}'' + B\Delta\bar{y}' + C\Delta\bar{y} = D \tag{30}$$

where A,B,C,D matrices are known for an initial guess of $\bar{y}$. We now discretize x and solve (30) satisfying the proper boundary conditions. An improved guess is calculated using $\bar{y}^1 = \bar{y} + \Delta\bar{y}$. Now the process is repeated with new A,B,C,D based on the new $\bar{y}$. A convergence criterion is used to stop the above process.

The procedure is straightforward. Once the distribution at some time t_{i-1} is known, (27) and (29) can be solved to find the distribution at t_i. By beginning with i=1, the step is performed for each time increment until the desired final time is reached. This procedure is shown pictorially in Figure 1.

Notice that the only parameter other than the dependent and independent variables is the normalized surface concentration, β. Thus, a series of solutions over a range of β will give a general solution to the quasi-static problem.

3.2 Evaluation of Physical Parameters. The basic parameters needed are n_i, N_c and η_i. Care must be taken since these parameters are evaluated at the diffusion temperature, say in the range between 900 and 1200°C. While it is possible to extrapolate the results obtained at lower temperature to higher temperature [4], in general our knowledge of m_e^* and m_h^* is poor. Thus the evaluation of $N_c(m_e^*,T)$ at diffusion temperature is suspect. For this reason, we use (17) evaluated for intrinsic material,

$$N_c = \frac{n_i}{F_{1/2}(\eta_i)} \tag{31}$$

The intrinic density $n_i(T)$ is evaluated using the data of Morin and Maita [5] and the Fermi integral is calculated using the polynomial approximation suggested by Battocletti [6].

The reduced Fermi level is evaluated using

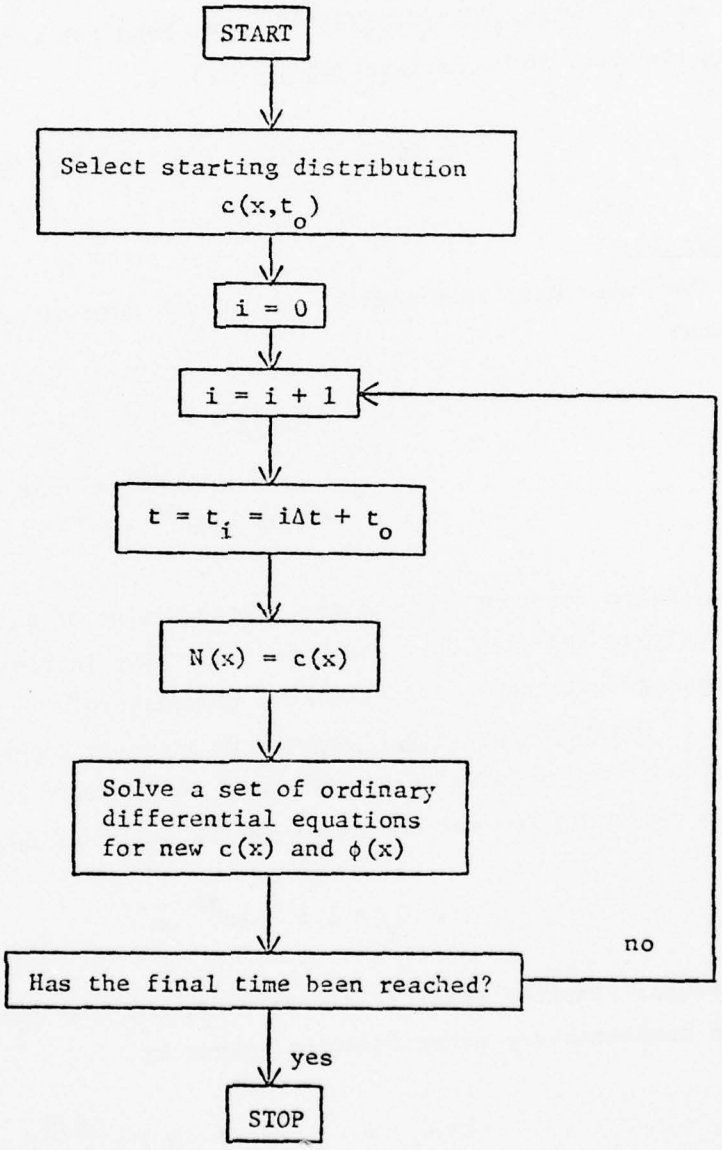


Figure 1. Flow Diagram of the Numerical Method

$$n_i(T) = -\frac{E_g}{2KT} + \frac{3}{4} \ln \frac{m_h^*}{m_e^*} \tag{32}$$

with $m_h^*/m_e^* = (0.67818/1.1925)^{2/3}$. The band gap in silicon is evaluated using the data of Macfarlane, et al, [7]

$$E_g(T) = 1.205 - 2.8 \times 10^{-4}\ T \quad {}^{o}K \tag{33}$$

4. Results

The input data used corresponds to the case of arsenic diffusion in silicon:

$$T = 1050^{o}C$$
$$D_c = 6.44 \times 10^{-15}\ cm^2/sec$$
$$n_i = 1.17 \times 10^{19}\ cm^{-3}$$

The diffusion constant used is the average value of data given by Masters and Fairfield [8] with Chiu and Chosh [9]. The intrinsic density represents the value calculated at the diffusion temperature.

4.1 Constant Source Diffusion. In order to determine the maximum influence of the internal electric field on the diffusion process, the surface concentration was set equal to the solid solubility value,

$$C_o = 1.6 \times 10^{21}\ cm^{-3}$$

Several starting conditions were used, the most prominent of which was the complementary error function, given by

$$c(x,t_o) = C_o\ \mathrm{erfc}\ [x/(4D_c t_o)^{1/2}] \tag{34}$$

Although a starting condition for ϕ was not needed, it was necessary to pick an initial guess for the first time step. This was chosen by assuming charge neutrality, giving

$$\phi(x,t_o) = V_T \sinh^{-1} [c(x,t_o)/2n_i] \qquad (35)$$

Typically t_o was chosen as 2 minutes with $\Delta x = 0.001\mu$ and $\Delta t = 0.5$ sec.

Figure 2 shows the numerical results for arsenic in silicon for a constant source diffusion. Also shown is the complementary error function which is the exact solution for negligible electric field. Inspection of the curves indicate significant deviation in the impurity profiles from those predicted by simple theory. Figures 3 and 4 show the electric field and charge density for the constant source diffusion process shown in Figure 2. These curves are reminiscent of those for a high-low junction. The field has a maximum value of approximately 10^5 volts/cm which decreases as the field moves into the material with increasing diffusion time. These results are typical for diffusion with high surface concentration.

At 900°C the Fermi level is just at the conduction band edge for a concentration of 2.4×10^{20} cm^{-3}; at 1200°C, for 3.8×10^{20} cm^{-3}. Thus for most constant source diffusions in practice, the material is degenerate. Figure 5 shows the numerical results for arsenic in silicon using Fermi-Dirac statistics for electrons compared to the profile obtained assuming non-degenerate material. The same input and control data was used in each case. Note that treating the material as if it were nondegenerate always produces an underestimate in the true value of the impurity profile. Typically the error produced ranges from 4% near the surface to 50% deep in the material.

4.2 Drive-in Diffusion. Here redistribution of impurities occurs from an initial distribution, which may be due to a prior diffusion or from an ion implantation. As an example, we assume a Gaussian initial distribution given by

$$N(x) = C_m e^{-(x-x_m)^2/2\sigma^2} \qquad (36)$$

Figure 6 shows the numerical results for $C_m = 10^{20}$ cm^{-3}, $x_m = 0.1\mu$ and $\sigma = 0.02\mu$. In order to determine the affect of the electric field on the motion of the impurities for this case, the profiles are compared to the zero field solution given by [10]

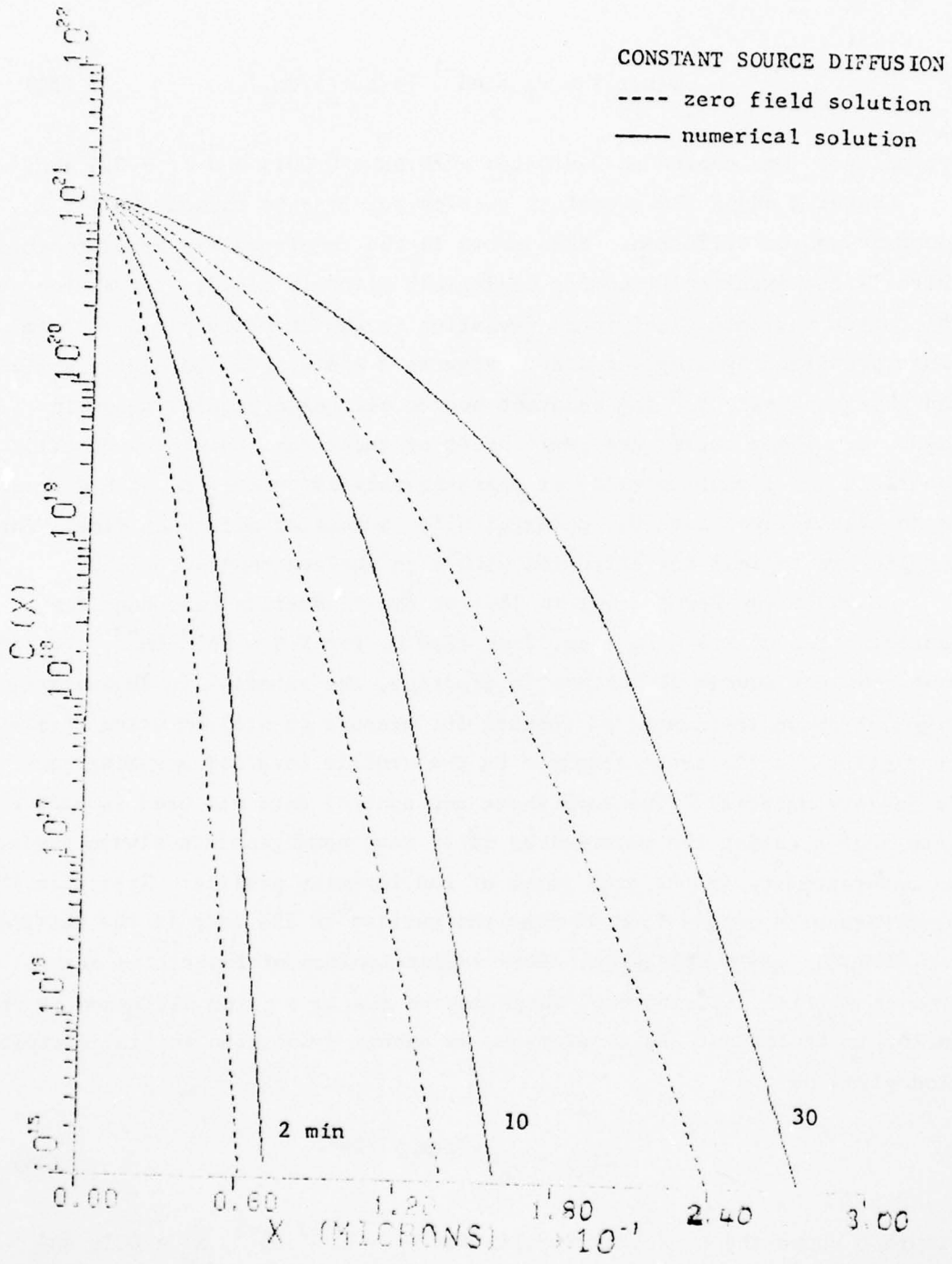


Figure 2. Impurity profiles of arsenic in silicon @ 1050 °C with $C_o = 1.6 \times 10^{21}$ cm^{-3}

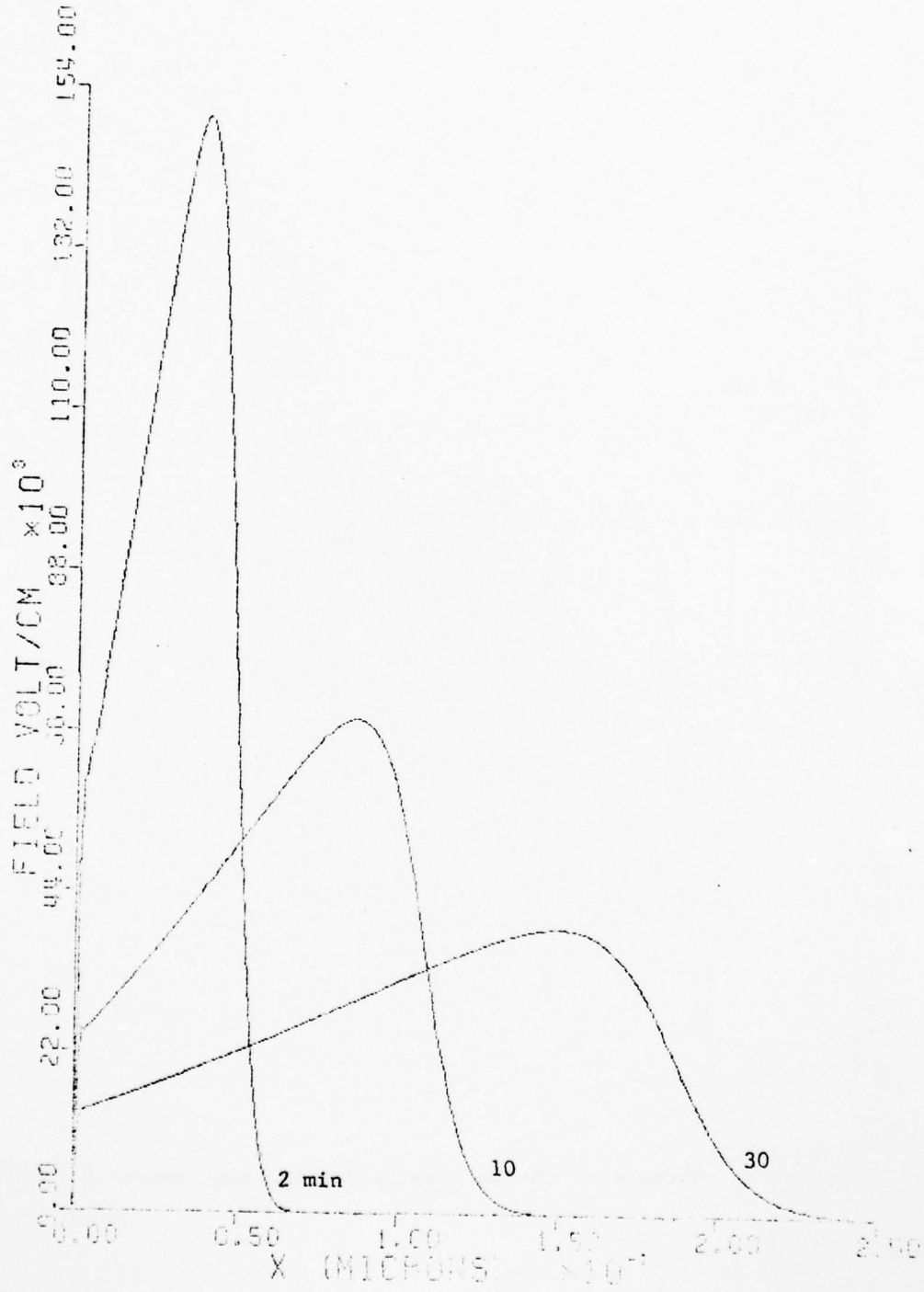


Figure 3. Electric field for constant source diffusion

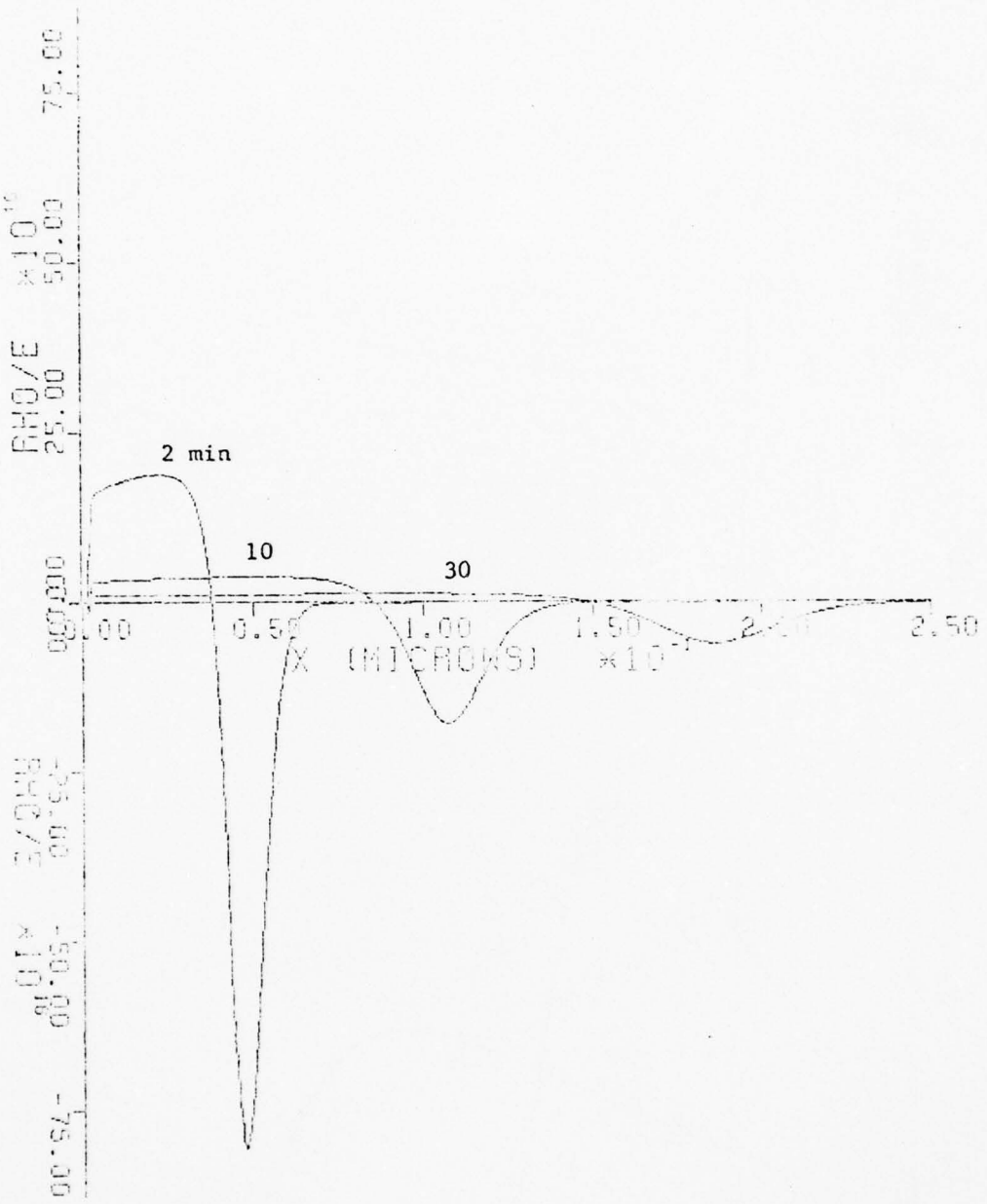


Figure 4. Normalized charge density for constant source diffusion

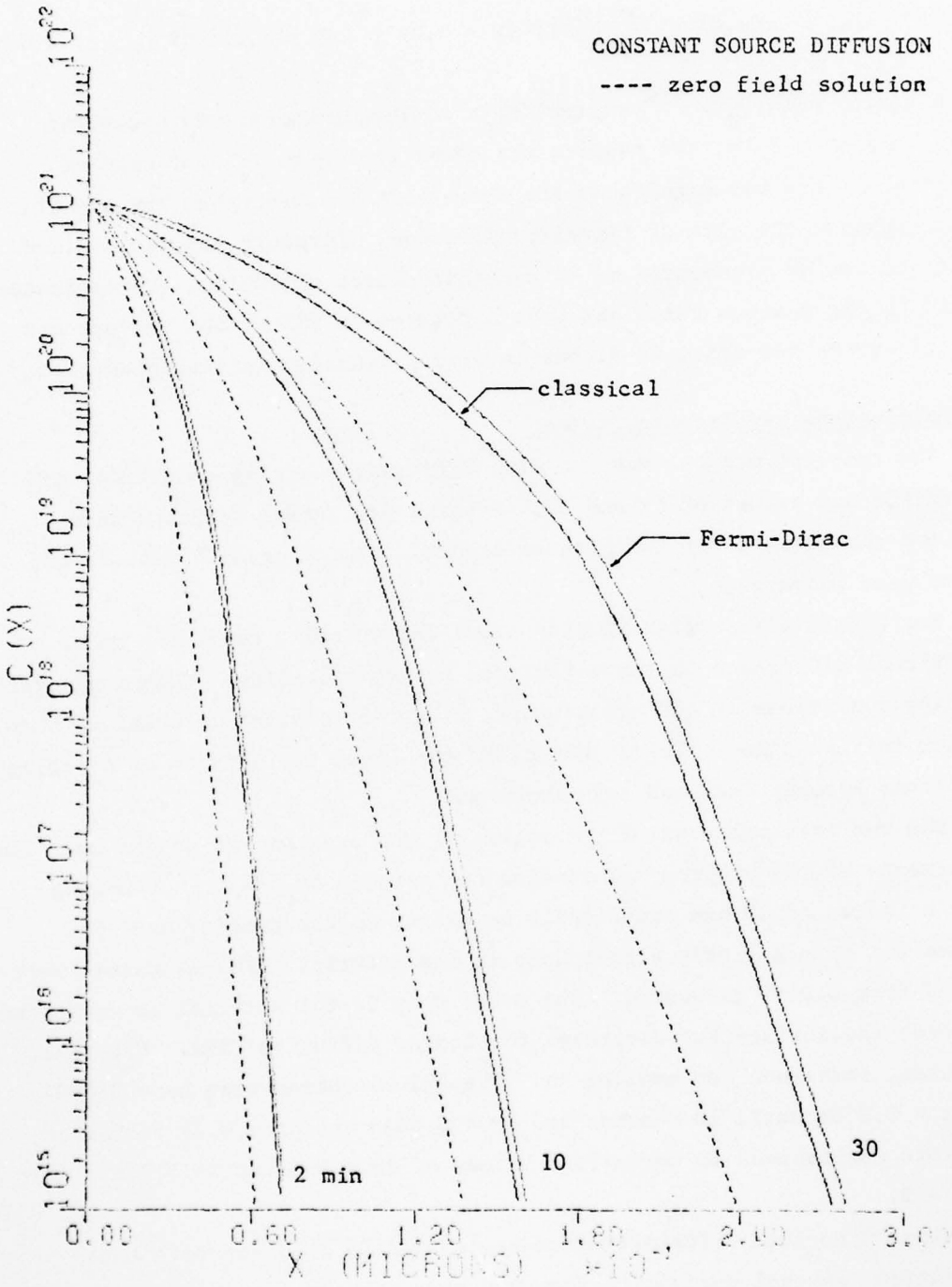


Figure 5. Comparison of the results for arsenic in silicon using classical and Fermi-Dirac statistics

$$c(x,t) = \int_0^\infty N(\alpha)[f(x-\alpha,t) + f(x+\alpha,t)]d\alpha \qquad (37)$$

where $f(x,t) = (4\pi D_c t)^{-1/2} \exp(-x^2/4D_c t)$. This integral was evaluated using Simpson's rule; the results are shown in Figure 7. For ease of comparison, only two magnitudes are shown. It is clear that the electric field enhances the rate of impurity diffusion. Although the field-aided effect is not as pronounced as in constant source diffusion, overestimates of 25% in the maximum value and underestimates of 50% at the surface are typical errors for drive-in diffusion from a Gaussian initial condition.

5. Conclusions and Recommendations

The computer program was run with different starting conditions and with different values of Δx and Δt. Several derivative formulas were used and the behavior of the program studied. The program is stable and offers good convergence.

For relatively short diffusion times the starting condition makes a significant difference in the calculated impurity density. Comparing erfc starting conditions at different times, a better solution is obtained with smaller initial times. For longer diffusion times the effects of starting conditions becomes less and less important.

The two critical control parameters of the program are Δx and Δt. The convergence of the program was studied for various Δx and Δt. Changing Δt by a factor of 10 had practically no effect on the density near the surface and only a little effect deep in the material. The Δx values were changed from 0.0005 to 0.005μ. The error deep in the material is more than that near the surface but decreases for longer diffusion time. CPU time, of course, increases for smaller Δx. The values chosen were $\Delta x = 0.0001\mu$ and $\Delta t = 0.005$s until 10 seconds and $\Delta x = 0.001\mu$ and $\Delta t = 0.5$s from then on. This corresponds to normalized values of $\Delta x \approx 0.4$ and $\Delta t \approx 0.05$; thus $\Delta x/\Delta t \approx 8$.

Since numerical differentiation is inherently less accurate than integration, several derivative formulas were studied with a view of noting the effect on the calculated values of impurity density, electric field and charge density. To make the differentiation more accurate, you must

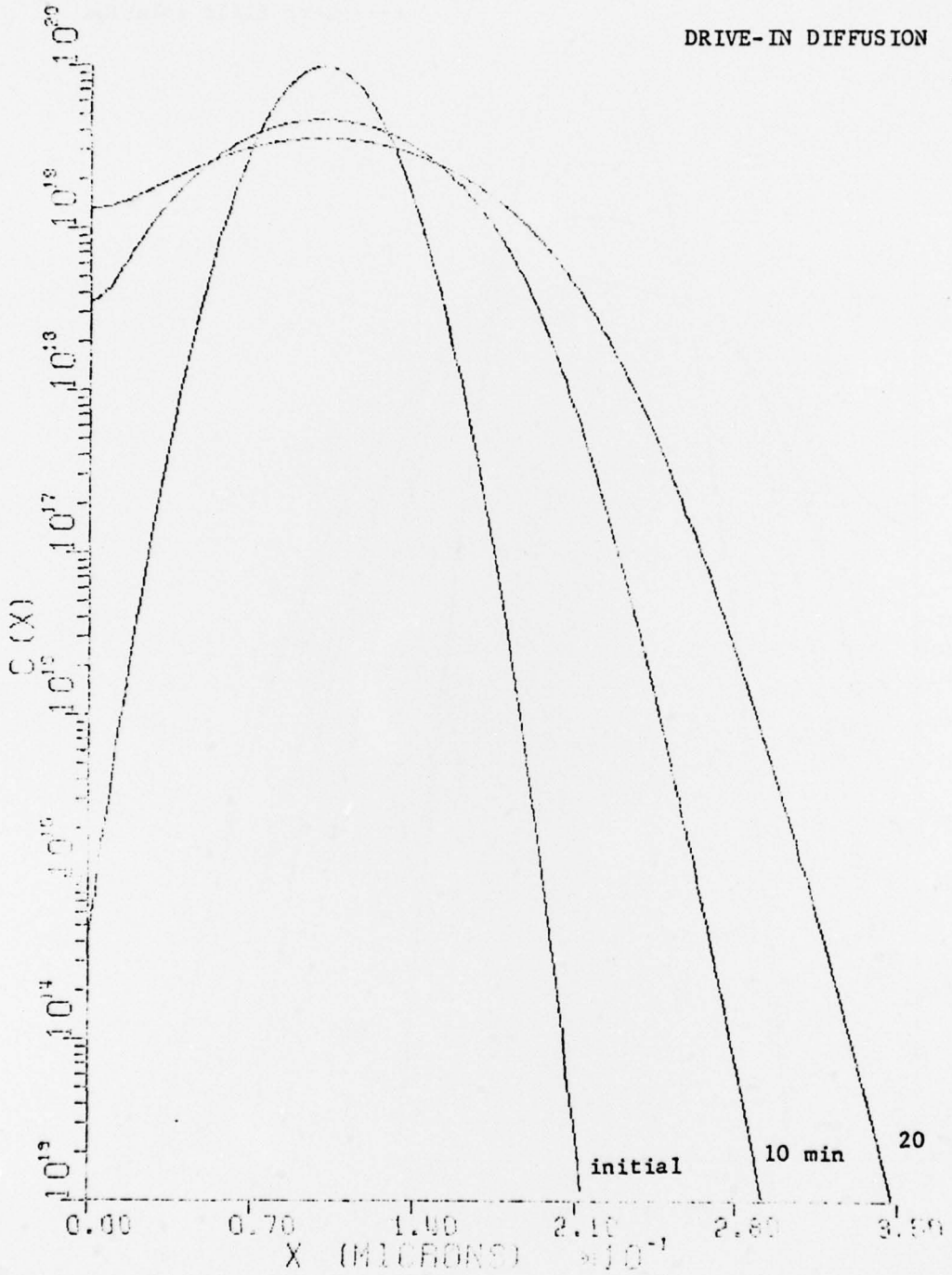


Figure 6. Impurity profiles of arsenic in silicon @ 1050 °C with Gaussian initial condition

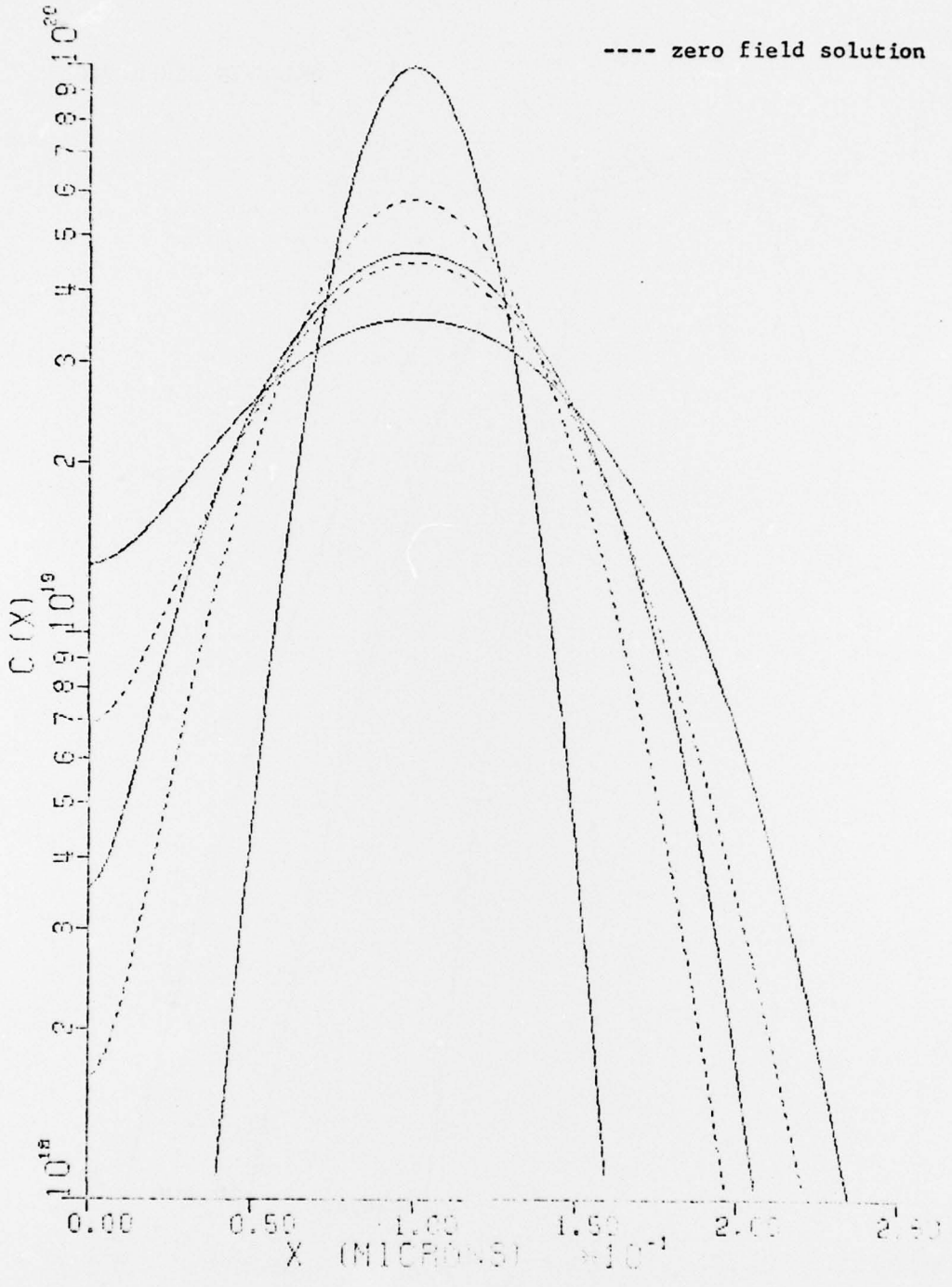


Figure 7. Effect of the electric field for drive-in diffusion

decrease Δx or increase the order of the formula used. Small values of Δx tend to magnify the round off errors; on the other hand, a higher order formula results in a larger CPU time. A 4 to 6 order formula seems to be adequate; a five point formula was used because of its symmetry about the central point.

The computer programs developed for calculating impurity profiles for constant source and drive-in diffusion allow a rigorous analysis of field-aided diffusion using a quasi-static approach. Since these programs require extensive CPU time, they are too costly to be used as engineering aids in process design. The purpose here is to use these programs to gain basic understanding of the diffusion process in semiconductor material and to check the accuracy of models that are suitable for process design and control.

The calculated impurity profiles for arsenic disagree with experimental results [11]. This discrepancy between theory and experiment must be due to using a diffusion model which includes only the field aided effect.

While the internal electric field significantly influences the motion of impurity ions, other effects must also be considered. Specific recommendation for extending this research are:

(1) investigate the diffusion process assuming charge neutrality,
(2) the numerical technique must be generalized to include a variable diffusion coefficient so that excess vacancy generation [12] and plastic deformation [13] effects can be incorporated into the diffusion model,
(3) since vacancies are not conserved [14], a new recombination term must be added to the diffusion model,
(4) the effects of a moving boundary due to oxide growth and redistribution during epitaxial growth should be studied.

References

1. T. V. [illegible]kin, V. A. Uskov and M. Ya. Shirobokov, "Effect of the intern[illegible]lectric field on the diffusion of impurities in semiconductors," Soviet [illegible]sics - *Solid State*, pp. 2703-2707, May 1966.

2. [illegible]. Q. [illegible]ewitt and A. H. Marshak, "A numerical study of field-aided diffusion," *Solid-State Electronics*, pp. 257-265, March 1974.

3. [illegible]. S. [illegible] and F. A. Lindholm, "Alternative formulation of generalized Einstein relation for degeneration material," *Proc. IEEE*, pp. 1256-1257, [illegible] 1968.

4. [illegible]. K. [illegible] and R. J. Van Overstreten, "Theoretical calculations of the Fermi level and of other parameters in phosphorus doped silicon at diffusion temperatures," *IEEE Trans. ED*, pp. 155-165, February 1974.

5. [illegible]. J. [illegible]n and J. P. Maita, "Electrical properties of silicon containing arsenic and boron," *Phys. Rev.*, pp. 28-35, October 1954.

6. [illegible]. E. [illegible]nocletti, "Polynomial approximation of the Fermi integral," Proc. [illegible], pp. 2162-2163, December 1965.

7. [illegible]. G. [illegible]rlane, T. P. McLean, J. E. Quarrington and V. Roberts, "Fine structure in the absorption edge spectrum of Si," *Phys. Rev.*, pp. 1245-1254, [illegible]ember 1958.

8. [illegible]. J. [illegible]ers and J. M. Fairfield, "Arsenic isoconcentration diffusion studies in silicon," *J. Applied Phys*, pp. 2390-2394, May 1969.

9. [illegible]. L. [illegible] and H. N. Ghosh, "A diffusion model for arsenic in silicon," IBM J. [illegible] *Develop.*, pp. 472-476, November 1971.

10. A. H. [illegible]rshak and J. E. Taylor, "Synthesis of general impurity profiles using a two-step diffusion process," *IEEE Trans. ED*, pp. 1037-1043, September 1972.

11. D. P. [illegible]nedy and P. C. Murley, "Concentration dependent diffusion of arsenic in silicon," *Proc. IEEE*, pp. 135-136, February 1971.

12. S. M. [illegible]. "General theory of impurity diffusion in semiconductors via the vacancy mechanism," *Physical Review*, pp. 773-784, 1969.

13. N. D. [illegible], "Anomalous diffusion in semiconductors-a quantitative analysis," *Solid-State Electronics*, pp. 165-172, 1970.

14. P. G. [illegible]mon, *Diffusion in Solids*. New York: McGraw-Hill, 1963.

	# of Copies
Defense Documentation Center ATTN: DDC-TCA Cameron Station (Bldg 5) Alexandria, VA 22314	2
Office of Naval Research Code 427 Arlington, VA 22217	1
Naval Ship Engineering Center ATTN: CODE 6157D Prince Georges Center Hyattsville, MD 20782	1
Commander Naval Electronics Lab Center ATTN: Library San Diego, CA 92152	1
Commander US Naval Ordance Lab ATTN: Technical Library White Oak Silver Springs, MD 20910	1
Rome Air Development Center ATTN: Documents Library (TILD) Griffiss AFB, NY 13441	1
HQ ESD (DRI) L.G. Hanscom Field Bedford, MA 01731	1
Deputy for Science & Technology Office, Assist Sec Army (R&D) Washington, DC 20310	1
Commander US Army Missile Command Redstone Scientific Info Ctr ATTN: Chief, Documents Section Redstone Arsenal, AL 35809	2
Commander Harry Diamond Laboratories ATTN: Library 2800 Powder Mill Road Adelphi, MD 20783	1

	# of Copies
Commander US Army Materiel Development and Readiness Command ATTN: DRCDE-R 5001 Eisenhower Ave Alexandria, VA 22333	1
NASA Scientific & Tech Info Facility ATTN: Acquisition Br (S-AK/DL) PO Box 33 College Park, MD 20740	2
Advisory Group on Electron Devices 201 Varick Street, 9th Floor New York, NY 10014	2
Director Naval Research Laboratory ATTN: Code 2627 Washington, DC 20375	1
USA Security Agency ATTN: IARDA Arlington Hall Station Arlington, VA 22212	1
Director Defense Communications Agency Technical Library Center Code 25 Washington, DC 20305	1
Director Naval Research Laboratory ATTN: Mr. Eliot Cohen Code 5211 Washington, DC 20375	1
Commander Harry Diamond Laboratories ATTN: Mr. Horst W. A. Gerlach 2800 Powder Mill Road Adelphi, MD 20783	1
**GIDEP Engineering & Support Dept - TE Section P. O. Box 398 Norco, CA 91760	1

**Do not submit if report is not releasable to public

	# of Copies
Institute of Defense Analysis Arlington, VA 22209	1
Mr. Jack Kilby 5924 Royal Lane Suite 150 Dallas, TX 75230	1
Commander Harry Diamond Laboratories ATTN: Mr. A. J. Baba 2800 Powder Mill Road Adelphi, MD 20783	1
Naval Research Laboratory ATTN: Dr. David F. Barbe (Code 5260) 4555 Overlook Ave., S.W. Washington, DC 20375	1
Commander Naval Electronics Laboratory Ctr ATTN: Mr. C. E. Holland Jr. 271 Catalina Blvd San Diego, CA 92152	1
Air Force Avionics Lab ATTN: AFAL/DHE (Mr. Stanley Wagner) Wright-Patterson AFB, OH 45433	1
Commander Rome Air Development Center ATTN: Mr. Joseph E. Brauer (RBRM) Griffiss AFB, NY 13441	1
Dr. Gerald B. Herzog Solid-State Technology Center RCA David Sarnoff Research Center Princeton, NJ 08540	1
Mr. Harold D. Toombs Texas Instruments, Inc. P.O. Box 5474; M/S 72 Dallas, TX 75222	1
Dr. George E. Smith Bell Telephone Labs, Inc. Room 2A-323 Murray Hill, NJ 07974	1
Dr. Barry Dunbridge TRW Systems Group One Space Park Redondo Beach, CA 90278	1
Dr. Paul E. Greene Hewlett-Packard Co. 1501 Page Mill Road Palo Alto, CA 94304	1
Naval Research Laboratory ATTN: L. J. Palkutti Code 5216 Washington, DC 20375	1
Bell-Northern Research IC Design Aids ATTN: D. M. Cauchey P. O. Box 3511, Station C Ottowa, Ontario K1Y Canada	1
Dr. Gordon E. Moore Intel Corporation 3065 Bowers Road Santa Clara, CA 95951	1
Naval Electronics Systems Cmd. ATTN: Mr. L. W. Sumney (Code 3042) Washington, DC 20360	1
Mr. Fred Sharf Actron, Dept 515 700 Royal Oaks Drive Monrovia, California 91016	1
Mr. Jim Mader Harris Electronic Systems Div P. O. Box 37 MS: 15/966 Melbourne, Florida 32901	1
Mr. Wesley Jones Bendix LSI Design Center East Joppa Road Baltimore, Maryland 21204	1
Mr. Charles Gwyn Sandia Laboratories Division 2142 Albuquerque, New Mexico 87115	1

	# of Copies
Mr. David Ford Dept 631, Magnavox Government & Industrial Elec. Co. 4624 Executive Blvd Fort Wayne, Indiana 46808	1
Mr. Dick Rath Hughes Aircraft Co. P. O. Box 92919 Airport Station Los Angeles, CA 90009	1
Mr. Don Gibson Rockwell International Collins Radio Group MOS/LSI Marketing Development Newport Beach, CA 92663	1
Mr. Jeff Steinwedel Solid State Scientific Montgomeryville Industrial Ctr Montgomeryville, PA 18936	1
Mr. John Travalent Sperry Univac M.S. UZU25 Univac Park, P. O. Box 3525 St. Paul, Minn. 55165	1
Mr. Ralph Alspaugh ITT Semiconductors 3301 Electronics Way West Palm Beach, FLA. 33407	1
Commanding Officer Naval Avionics Facility ATTN: Mr. Bob Murphy Code 908, 6000 East 21st St. Indianapolis, Indiana 46218	1
Commander Naval Electronics Lab Dr. I. Lagnado, Code 4800 271 Catalina Blvd. San Diego, California 92152	1
RAC RADC ATTN: Mr. I. L. Krulac Griffiss AFB, NY 13441	1

	# of Copies
Commander	
ATTN: DRSEL-MS-TI	2
ATTN: DRSEL-TL-DT	1
ATTN: DRSEL-TL-E	1
ATTN: DRSEL-TL-M	1
ATTN: DRSEL-TL-P	1
ATTN: DRSEL-PL-ST	1
ATTN: DRSEL-TL-B	1
ATTN: DRSEL-TL-I	1
ATTN: DRSEL-TL-I (CPC File)	2
ATTN: DRSEL-TL-IT	1
ATTN: DRSEL-TL-IC	1
ATTN: DRSEL-TL-ID	1
ATTN: DRSEL-TL-IR	1
ATTN: DRSEL-TL-IG	5
ATTN: DRSEL-TL-IS	1
ATTN: DRSEL-TL-IJ	1
ATTN: DRSEL-TL-DD	1
Fort Monmouth, NJ 07703	