



12
B.S.

Research and Development Technical Report

ECOM - 75-1344-1

AD A 024459

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

April 1976

Interim Report for Period 1 July 1975-31 December 1975

Distribution Statement

Approved for Public Release, Distribution Unlimited

Prepared for

Advanced Research Agency
1400 Wilson Boulevard
Arlington, Va. 22209

ECOM

US ARMY ELECTRONICS COMMAND FORT MONMOUTH, NEW JERSEY 07703

EDRC-76-1

**BEST
AVAILABLE COPY**

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.

ACCESSION for	
NTIS	☒
DTIC	☐
UNCLASSIFIED	☐
JUSTIFICATION	
BY	
DESCRIPTION	
A	

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER ECOM-75-1344-1	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER (1 Jul-31 Dec 75)	
4. TITLE (and Subtitle) Computer Aided Engineering of Semiconductor Integrated Circuits.		5. TYPE OF REPORT & PERIOD COVERED Interim Report. July 1, 1975-Dec. 31, 1975	
7. AUTHOR(s) David P. Kennedy Principal Investigator		8. PERFORMING ORG. REPORT NUMBER EDRC-76-1	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Electron Device Research Center Univ. of Fla., Dept. of Elec. Eng. Gainesville, Florida 32611		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61101E 5D10	
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency (ARPA) 1400 Wilson Blvd. Arlington, Virginia 22209		12. REPORT DATE April 1976	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) U.S. Army Electronics Command Attn: DRSEL-TL-IG Fort Monmouth, New Jersey 07703		13. NUMBER OF PAGES 271	
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		15. SECURITY CLASS. (of this report) UNCLASSIFIED	
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		16a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
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) COMPUTER Aided Engineering INTEGRATED Circuits SEMICONDUCTORS, Computer Aided Engineering of IC. CAD, Semiconductor ICs.			
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 and to reduce the cost of these circuits by devising improved computer-aided engineering techniques. During this period emphasis has been placed on integrated circuit device and process modeling. For device modeling the			

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 68 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

409438-43

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract

results presented include the: (1) development of two-dimensional model for MOS and bipolar transistors; (2) development of a one-dimensional mathematical model for MOSFET operation; (3) studies on equivalent circuit modeling of transistor operation; (4) theoretical investigations of carrier mobility in the inversion layer of an MOSFET; (5) mathematical investigations for high-speed analysis of devices; (6) sheet resistivity test pattern model development, using Monte-Carlo techniques; and (7) hot carrier studies for inhomogeneous semiconductor material. For process modeling the results presented include (1) ion implantation; (2) thermal oxidation and chemical vapor deposition; (3) silicon epitaxy; and (4) thermal diffusion.

19. MODELING, Semiconductors and Devices
SEMICONDUCTOR Processing
PROCESS Modeling of Semiconductors

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SUMMARY

First Semi-Annual 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 process; (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 first six 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 first six months of the program.

Table of Contents

	<u>Page</u>
Summary.....	1
Illustrations.....	5
Tables.....	9

Part I

Semiconductor Device Modeling (Univ. of Florida)

Introduction to Part I.....	13
Chapter I Two-Dimensional Models of MOS and Bipolar... Transistor Operation	17
Chapter II One-Dimensional Mathematical Model of..... MOSFET Operation	47
Chapter III Equivalent Circuit Studies.....	73
Chapter IV Carrier Mobility in an MOSFET.....	99
Chapter V Integral Methods for the Analysis of..... Semiconductor Devices	123
Chapter VI Test Pattern Model Using Monte Carlo..... Methods	143
Chapter VII Hot Carrier Mechanisms in Semiconductor..... Devices	159

Part II

Semiconductor Process Modeling (Stanford Univ.)

Introduction to Part II.....	169
Chapter VIII Ion Implantation.....	173
Chapter IX Thermal Oxidation.....	197
Chapter X Silicon Epitaxy.....	211
Chapter XI Mathematical Modeling of Impurity..... Diffusion	239

Appendices and References

References.....	253
Appendix A.....	261

Illustrations

<u>Figure</u>	<u>Title</u>	<u>Page</u>
I-1	Cross-section of an n-channel MOSFET.....	22
I-2	Graded lattice points used in the..... computer model.	23
I-3	Boundary conditions used in the two-..... dimensional analysis of an MOSFET.	31
I-4	Five-point star of mesh points.....	37
I-5	Basic mathematical model of a bipolar..... transistor.	46
II-1	Illustrative model of an MOSFET.....	52
II-2	Calculated inversion charge in an..... MOSFET.	65
II-3	Calculated first derivative of inver-..... sion charge for strong inversion operation.	66
II-4	Calculated inversion charge in an..... MOSFET for weak inversion operation.	67
II-5	Calculated first derivative of inversion... charge for weak inversion operation.	68
III-1	Illustrating the division into intrinsic... and extrinsic regions for an MOSFET and a bipolar transistor.	78
III-2	General charge-control device.....	80
III-3	Illustrating the development of the..... general network representation of the extended ideal charge-control model.	83
III-4	One possible network representation of..... the extrinsic part of the MOSFET transistor.	92
IV-1	Simplified energy diagram at the inter-..... face of Si-SiO ₂ .	104
IV-2	Calculated electron energy levels in the.... surface potential well.	112
IV-3	Calculated electron energy levels in the.... surface potential well.	113

IV-4	Comparison between quantum mechanical..... solution and classical solution for the inversion layer carriers.	115
IV-5	Quantum mechanical solutions of inversion.. carrier distributions.	116
IV-6	Potential distribution at the interface of. Si-SiO ₂ .	122
V-1	Computational instability in the absence of. under-relaxation.	134
V-2	Number of Picard iterations vs. under-..... relaxation parameter.	135
V-3	Solution of Poisson's equation using the.... Fredholm integral method.	136
V-4	Number of Picard iterations vs. under-..... relaxation parameter.	138
V-5	Calculated potential distribution in a..... reverse biased, constant gradient, p-n junction.	139
VI-1	Sheet resistance frequency distribution..... from a statistical sample of about 5000 test pattern measurements.	146
VI-2	Calculated frequency distribution for the... boron surface concentration during a pre-dep. diffusion.	147
VIII-1	Block diagram for study of implant annealing. Profiles in common semiconductor materials.	177
VIII-2	Isochronal annealing of boron in silicon.....	178
VIII-3	Curve showing integral density of oxygen in.. silicon versus oxide thickness obtained from experiments for an implant dose of 5×10^{13} as cm ⁻³ at 100 KeV. (Insert: sketch of method used for producing recoil oxygen in silicon.)	188
VIII-4	Oxygen dose versus surface state density.....	190
VIII-5	Distribution of implanted as and recoil..... oxygen in SiO ₂ .	193
VIII-6	(a) Division of substrate into thin layers... to calculate energy distribution of the ions. (b) Total incident ion flux per unit energy as a function of distance traveled.	195

IX-1	Oxide thickness versus time for the oxidation of (100) and (111) oriented N-type silicon in dry oxygen at 1000°C.	203
IX-2	Determination of rate constants for the dry oxidation of (100) and (111) oriented N-type silicon at 1000°C. The slopes are equal to B and the intercepts to -A as per eq. (IX-6).	205
IX-3	Effect of redistribution of Si and SiO ₂ impurity profiles.	207
X-1	Horizontal epitaxial reactor schematic.	214
X-2	Horizontal reactor cross section with thermal and chemical profiles during growth.	215
X-3	Arrhenius plot of SiH ₄ growth rate.	219
X-4	SiH ₄ growth rate versus SiH ₄ concentration.	219
X-5	(a) Dopant concentrations near an epitaxial interface. (b) Band bending near a growing silicon surface. (c) Dopant and free electron concentrations near a growing silicon surface.	221
X-6	Solid-state doping versus gas phase dopant concentration for various dopant sources, silicon sources, and temperatures.	223
X-7	Doping profile of the step epitaxial from C-V measurements.	227
X-8	The DDMOS capacitor structure.	232
X-9	Typical energy band diagram for the substrate under the gate oxide.	233
X-10	Conceptual diagram of autodoping and out-diffusion profiles.	237
XI-1	Flow diagram of the numerical method.	247
XI-2	Diffusion profiles of arsenic in silicon at 1050°C with $C_0 = 1.6 \times 10^{21}$.	249

Tables

<u>Number</u>	<u>Title</u>	<u>Page</u>
I-1	DEFINITION OF DIMENSIONLESS VARIABLES....	25
VIII-1	DEFINITION OF THE VECTOR $\vec{y} = (y_1 \dots y_{99}) \dots$	181
VIII-2	NONZERO ELEMENTS OF THE JACOBIAN.....	185
IX-1	DRY O ₂ OXIDATION PARAMETERS.....	204
IX-2	HCl OXIDATION PARAMETERS.....	209

Part I

Semiconductor Device Modeling

University of Florida
Gainesville, Florida

Introduction - Part I
Semiconductor Device Modeling

At the University of Florida, efforts are directed toward the development of mathematical models that accurately predict the electrical characteristics of semiconductor devices. Involved here are 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 been also initiated on mathematical models for test patterns to be used during IC fabrication process evaluation.

Included in this device model development program is the research needed to further enhance our modeling capabilities. For example, two-dimensional transient solutions for device operation cannot presently be attained, from a practical point of view-the required computer time is prohibitive. Therefore, research is being directed toward implementing mathematical techniques that are not presently used for device analysis; methods that could significantly reduce the required computer time. Research is also underway on physical mechanisms of MOS and Bipolar transistor operation that are known to be inadequately, or inaccurately, described by available modeling techniques.

Initial studies for these two-dimensional models have been directed toward the MOSFET. A computer program for MOSFET analysis was used as a test vehicle for modifications, new algorithm, development, etc. All decisions have now been made, and a new MOSFET analysis program is being written. It is expected that this new program will be completed, in final form, in approximately six weeks. In addition, an available computer program for the two-dimensional analysis of bipolar transistors is now operational at the University of Florida; this task involved modifying a computer program designed for operation on a CDC machine, so that it can be run on an IBM machine.

An engineering model for MOSFET operation has been brought to near completion during the past six months. Present efforts are directed toward testing this model. A few difficulties arose through some unwarranted assumptions, and these difficulties are being corrected. It is expected that this model will be completed within the next few weeks.

The quasi-static approximation underlies all device models commonly used in the transient computer simulation of large-signal bipolar or MOSFET circuits. We propose a test for the self-consistent validity of this approximation. The test applies to each device model in a circuit, and is easily implemented in circuit-analysis programs. It applies to any kind of device, such as the bipolar transistor and the MOSFET, whose intrinsic structure operates by charge control. Emphasis is given to testing the extrinsic as well as the intrinsic device, and to the non-idealities present in actual devices. Prominent among these non-idealities is the four-terminal nature of the MOSFET. Repeated application of the test can help determine the degree of complexity needed, for each device model in a circuit, to assure self-consistency with the quasi-static approximation.

In MOSFET operation, the mobility of inversion layer carriers undergo important changes, with a change of applied gate voltage. This difficulty is presumed to result from scattering at the oxide-semiconductor interface, although there is no adequate model for predicting its influence on carrier mobility. We discuss here a program of research directed toward a solution of this important problem, and the advances made within the first six months of this contract period.

Initial studies are underway on mathematical methods that are suitable for the two-dimensional transient analysis of semiconductor devices. Extensive 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. In addition, preliminary studies have also been undertaken of other computational methods: finite elements,

weighted residual, and collocation. It is our intent to find a method whereby these transient calculations can be undertaken without requiring an unreasonable large amount of computer time.

A statistical model is being developed for test patterns used to evaluate sheet resistance. Unlike other available models, the present development is aimed toward approximating the statistical distribution of sheet resistance actually observed during IC manufacturing. A suitable psuedo random number is needed for this task. Such a number generator has been acquired, it was tested for randomness, and it has been found satisfactory.

In device modeling, some regions of large electric field are known to produce hot-electron mechanisms. Presently, the consequences of these mechanisms are accounted for by introducing a field dependent carrier mobility. It can be shown that this modeling technique is only applicable to homogeneously doped semiconductor material, and that substantial error arises through its application to regions containing a large impurity atom gradient. Research has been initiated on this topic, with an aim toward accurately modeling the consequences of hot electron mechanisms in semiconductor devices.

University of Florida

D. P. Kennedy

U. Kurzweg

F. A. Lindholm

A. D. Sutherland

M. Zahn

Chapter I

Two-Dimensional Mathematical Models for MOS and Bipolar Transistor Operation

A. D. Sutherland

	<u>Page</u>
1.0 Introduction.....	19
2.0 Two-Dimensional MOSFET Model.....	20
2.01 The Basic Geometry.....	21
2.02 Dimensionless Variables.....	24
2.03 Basic Equations.....	24 ¹
2.04 Iteration Method.....	26
2.05 The Stream Function.....	28
2.06 The Determination of $n(x,y)$	29
2.07 Boundary Conditions.....	30
2.08 Determination of the Current Constant J_0 ...	35
2.09 Finite-Difference Equations.....	36
2.10 Method Used for Solving the Finite-..... Difference Equations	42
3.0 Two-Dimensional Bipolar Transistor Model.....	44

Chapter I
Two-Dimensional Mathematical Models for MOS and
Bipolar Transistor Operation
A. D. Sutherland

1.0 Introduction

Two distinct categories of computer models for devices are required in order to achieve one ultimate objective of this research program-the ability to model accurately the effects of statistical variations in process parameters, characteristic of a given IC production line, upon the performance of a given IC configuration. The first category of device models, which will be embedded in computer programs using Monte-Carlo techniques to simulate statistical process parameter variations, must be ultrafast; they will be called upon thousands of times during the course of a single computer run to simulate device characteristics. Thus, this class of models, which we distinguish as "statistical models" because of their intended use, will generally be one-dimensional in nature and will make use of both analytic and empirical approximations of the device physics in order to achieve both accuracy and high speed. The second category of device models, which we distinguish as "engineering models," serves an entirely different purpose-accurate modeling of the physical processes within a given device so as to (1) provide physical insight aiding in the development of good high speed "statistical models" and (2) serve as an engineering tool for the design of devices per se. This section of this report deals with the latter category of models.

The development of engineering models under this program involves more than the achievement of a working computer model alone. It is our intention to disseminate these models to government laboratories and industry, and to encourage their use and critical evaluation of them. We seek to achieve feedback which will lead to improved 2nd, and possibly 3rd generation models as the program progresses. To this end, it is vital that

adequate documentation of the models be provided, in order that others will find them easy to use and understand. Thus, we are developing rather complete USER'S MANUALS which will include detailed instructions for using the computer program, detailed flow charts, a description of how the program functions, including the various numerical algorithms used, a glossary of FORTRAN terms, sample INPUT and OUTPUT formats, and a listing of the computer program itself.

Presently under development, in the above context, are two 2-dimensional steady-state computer models, one devised to simulate MOS field effect transistors, the second to simulate bipolar transistors. Each of these will be described in what follows. Transient models for each of these devices are planned for the future.

2.0 Two-Dimensional MOSFET Model

The starting point for the development of a 2-D MOSFET model has been the prior work of Mock [1], who provides a somewhat concise description of his basic computational approach in the reference cited. Good agreement with experiment has been reported with Mock's model [2].

We have a working version of Mock's computer program, provided by IBM, which was developed for them under the principal investigator's direction while he was with IBM. We have documented this program to an extent sufficient to enable us to run computations with it, and have used it as a tool to assist in the development of a one-dimensional "statistical model" for MOSFETs described elsewhere in this report. But there are substantial details of that program's inner workings which are not fully understood in terms of the detailed manner in which Mock's algorithms are implemented. Rather than attempt to develop full documentation for this program in the detail described above, then seek IBM's agreement to its dissemination, the author's philosophy

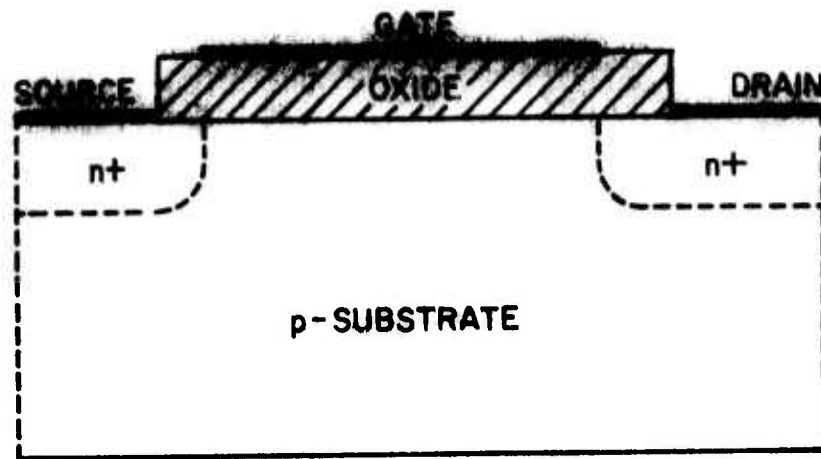
[1] M. S. Mock, *Solid State Electronics*, 16, 601 (1973).

[2] D. P. Kennedy, private communication.

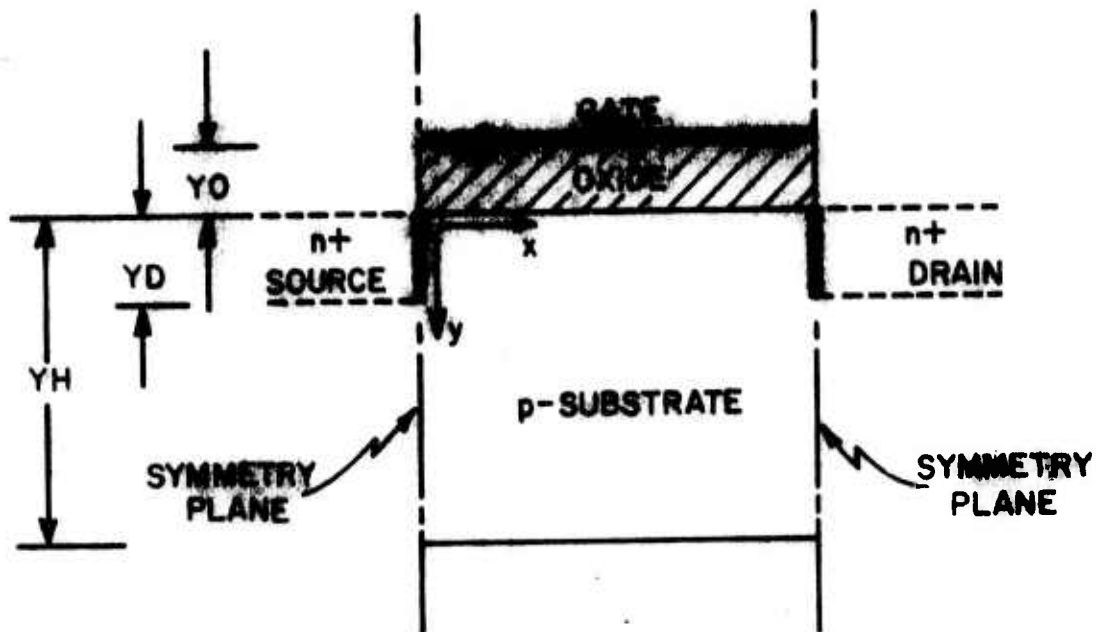
has been that he can more effectively provide full documentation for a computer program whose detailed inner workings he fully understands, i.e., one written by us. Thus, what is described below is our own version of Mock's basic model, based more on the methods described in reference [1] than upon his computer program itself, with some modifications in approach which are our own.

2.01 The Basic Geometry.-Figure I-1 depicts the configuration along the channel of a typical n-channel MOSFET structure resulting from the diffusion of n+ source and drain regions into a p-type substrate. Since, in such a device, the "action" takes place due to the formation of a thin inversion layer of mobile electrons immediately below the silicon-oxide interface, it is reasonable to simplify the geometry dealt with by the model, as shown in Figure I-1(b). This is indeed what Mock did in his computer model although, in reference [1], he implies the treatment of the more general geometry of Figure I-1(a).

The principal reason for selecting the simplified geometry of Figure I-1(b) as the basis for the model is that it places the metallurgical junctions between p-substrate and n+ source and drain at the left and right borders of the rectangular region under study. Large gradients in electric potential and in electron and hole number density are to be expected in the near vicinity of those junctions, as well as in the near vicinity of the oxide-silicon interface. Since finite-difference methods are employed to solve the partial differential equations relating these functions, a rather fine spacing between the lattice points at which these functions are numerically determined is required in those regions to achieve adequate resolution, whereas the spacing between lattice points remote from those regions can be relatively coarse. By placing the source and drain metallurgical junctions at the borders of the region, such a graded lattice of points is easily defined. Figure I-2 illustrates the graded lattice used for a 41 (horizontal) by 25 (vertical) array. The prescription



(a)



(b)

Figure I-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.

GRADED LATTICE

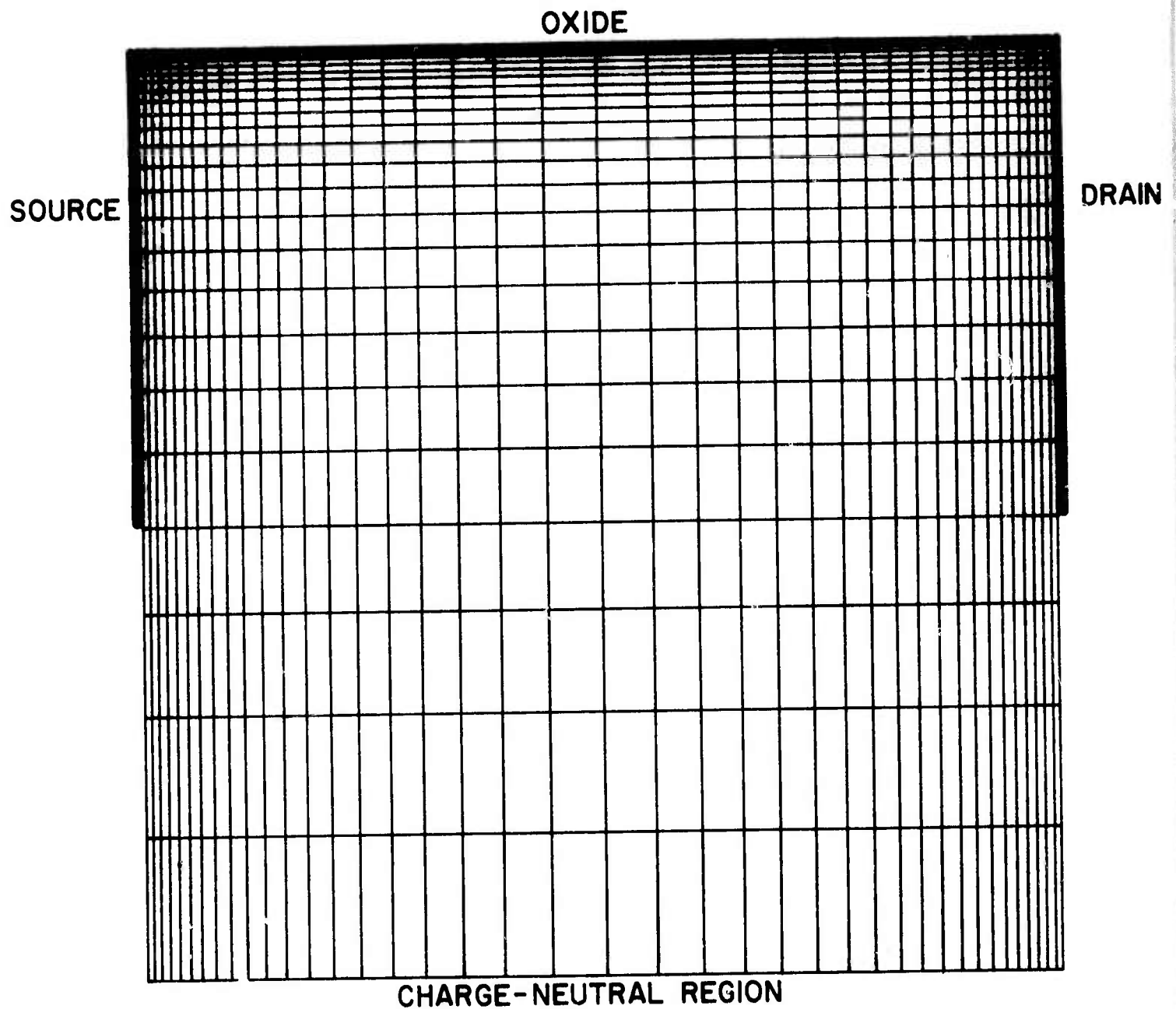


Figure I-2. Graded lattice of points used in the finite-difference equations dealt with in the computer model.

followed for generating such a graded lattice is precisely that suggested by Mock (reference [1], eq. (24)).

2.02 Dimensionless Variables.-To avoid unnecessary repeated multiplications by redundant constants, such as kT/q , q , etc., as well as to provide scaling which causes variables to lie within reasonable numerical ranges, we introduce normalized variables. Since variables having physical dimensions will be referred to only in this paragraph, it is convenient to designate them with primed symbols, enabling the use of the same symbols, but unprimed, to represent their dimensionless counterparts. Thus, we can utilize familiar symbols such as $\bar{J}$ for current density, n and p for electron and hole density, ψ for potential, μ for mobility, etc., even though these all represent dimensionless variables. In this vein, the dimensionless variables employed are defined in TABLE I-1.

2.03 Basic Equations.-As did Mock, we assume negligible electron-hole generation/recombination, and ignore the flow of hole current. These approximations are reasonable for an n-channel MOSFET under normal operating conditions. Their effect is to cause the electron current density vector to be divergenceless-a condition which is readily enforced by deriving that current from a stream function.

In terms of the dimensionless variables just defined, the equations describing the physical behavior of electrons and holes in the p-substrate region of Figure I-1(b) are:

$$\nabla^2 \psi = N_A + n - p \quad (\text{I-01})$$

$$n = \exp(\psi - \phi_n) \quad (\text{I-02})$$

$$p = \exp(\phi_p - \psi) \quad (\text{I-03})$$

$$\bar{J} = -\mu n \nabla \phi_n \quad (\text{I-04})$$

$$0 = \nabla \phi_p \quad (\text{I-05})$$

$$\nabla \cdot \bar{J} = 0 \quad (\text{I-06})$$

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

TABLE I-1.-DEFINITION OF DIMENSIONLESS VARIABLES. (κ_s IS THE RELATIVE PERMITTIVITY OF SILICON.)

Define:

$$L_D' = \left[\frac{\kappa_s \epsilon_0' (k'T'/q')}{q' n_i'} \right]^{1/2}, \quad \text{Intrinsic Debye length of silicon}$$

$$\mu_n', \quad \text{Low Field electron mobility}$$

Then Let:

$$\left. \begin{aligned} \psi' &= (k'T'/q')\psi \\ \phi_n' &= (k'T'/q')\phi_n \\ \phi_p' &= (k'T'/q')\phi_p \end{aligned} \right\} \quad \text{Electric potential, and electron \& hole quasi-Fermi potentials}$$

$$\left. \begin{aligned} N_D' &= n_i' N_D \\ N_A' &= n_i' N_A \\ n' &= n_i' n \\ p' &= n_i' p \end{aligned} \right\} \quad \text{Donor, acceptor, electron, \& hole number density}$$

$$\left. \begin{aligned} x' &= L_D' x \\ y' &= L_D' y \end{aligned} \right\} \quad \text{cartesian coordinates}$$

$$\bar{J}' = \left[\frac{(k'T'/q') \mu_n' n_i' q'}{L_D'} \right] \bar{J} \quad \text{electron current density}$$

$$\mu' = \mu_n' \mu, \quad \text{electron mobility}$$

Equation (I-01) is Poisson's equation for the electric potential ψ , (I-02) and (I-03) express the use of Boltzmann statistics (i.e., nondegenerate doping is assumed), (I-04) relates the electron current density vector to the electron number density and to the electron quasi-Fermi potential in the usual manner [3], (I-05) enforces zero hole current, while (I-06) assures conservation of charge under the assumed condition of zero generation/recombination of electron-hole pairs.

The above equations apply in the p-substrate region of Figure I-1(b). The conditions in the oxide region are described by:

$$\nabla^2 \psi = 0, \quad (\text{I-07})$$

and $p=n=\bar{J}=0$, with appropriate boundary conditions applied to tangential and normal electric fields to match the solutions of (I-01) and (I-07) at the oxide-silicon interface.

2.04 Iteration Method.—We adopt Mock's iterative procedure for solving the above set of coupled nonlinear partial differential equations. At each step m in the iteration, the unrelaxed potential function $v(x,y)^{(m)}$ is computed, satisfying (I-01) and (I-07) in the substrate and oxide regions, respectively, subject to the boundary conditions to be discussed in a later subsection. The electron and hole densities $n^{(m)}$ and $p^{(m)}$ used in computing $v^{(m)}$ are those determined using the potential $\psi^{(m)}$ obtained in the preceding iteration step, in a manner also to be described later. Then given $v^{(m)}$, the improved approximation $\psi^{(m+1)}$ to the correct potential $\psi(x,y)$ is obtained from:

$$\begin{aligned} \psi(x,y)^{(m+1)} = & \psi(x,y)^{(m)} + \alpha_m [v(x,y)^{(m)} - \psi(x,y)^{(m)}] \\ & + \beta_m [\psi(x,y)^{(m)} - \psi(x,y)^{(m-1)}] \end{aligned} \quad (\text{I-08})$$

The iteration is terminated when the maximum residual, $\text{Max}|v^{(m)} - \psi^{(m)}|$, falls below a specified value.

- [3] S. M. Sze, *Physics of Semiconductor Devices*, p. 96, Wiley, N.Y. (1969).

The relaxation parameters α_m and β_m appearing in (I-08) are assigned the Chebyshev sequence of values discussed by Mock, and are reinitialized in the manner described by him whenever the maximum residual obtained at a given step of the iterative process fails to be less than that obtained at the previous step. Since the rate of convergence of this iterative scheme depends upon the total number of free carriers in the system, we follow Mock's idea for speeding convergence by dealing with modest source and drain doping levels, e.g. 10^{17} cm^{-3} , until the iteration has terminated, following which the doping levels in those regions are set to their desired values, with further iterative operations then using a linearized version of (I-01), obtained by invoking Gummel's algorithm [4]. (See reference [1], Section 4.)

The initial approximation $\psi(x,y)^{(0)}$ required to initiate the above iterative scheme is determined in the manner now to be described. Given fixed applied gate and substrate voltages V_G and V_{SUB} , the use of the model usually involves the sequential selection of a number of ascending values of the applied drain voltage V_D , with the above iterative scheme applied to achieve a self consistent solution of (I-01) through (I-07) at each such V_D . (This effectively traces out a curve of I_D versus V_D , with V_G fixed, as one does experimentally with a curve tracer.) Thus, once this sequence is started, $\psi(x,y)^{(0)}$, and the corresponding values for $n(x,y)^{(0)}$ and $p(x,y)^{(0)}$ are merely chosen to be those values just previously found for a lower value of V_D . Thus, we need only have a means for approximating $\psi(x,y)^{(0)}$ at the initiation of the sequence of drain voltage values.

By initiating such a sequence with V_D set to zero, a reasonable approximation for $\psi(x,y)^{(0)}$ can be achieved by solving a one-dimensional version of (I-01) and (I-07), with the direction along which ψ varies being perpendicular to the gate electrode. This we do, using the same iterative scheme described above, but

[4] H. K. Gummel, *IEEE Trans. Electron Devices*, ED-11, 455 (1964).

starting with the one-dimensional solution obtained using the depletion approximation to initiate the one-dimensional iteration, the latter being analytically expressible. Again, this basically follows the procedure adopted by Mock.

2.05 The Stream Function.—In view of (I-06), we introduce, as does Mock, a stream function $\theta(x,y)$ such that:

$$\begin{aligned} J_x &= J_0 \frac{\partial \theta}{\partial y} (x,y), \\ J_y &= -J_0 \frac{\partial \theta}{\partial x} (x,y), \end{aligned} \quad (I-09)$$

which assures the satisfaction of (I-06). Here, J_0 is a scalar constant. We seek next the differential equation which must be satisfied by this stream function.

Solving (I-02) for the electron quasi-Fermi potential ϕ_n , and substituting the result in (I-04), we obtain:

$$\bar{J} = \mu e^\psi \nabla (n e^{-\psi}). \quad (I-10)$$

Identifying the x and y components of (I-10) with the corresponding expressions in (I-09), we have:

$$\begin{aligned} J_0 \frac{\partial \theta}{\partial y} &= \mu e^\psi \frac{\partial}{\partial x} (n e^{-\psi}) \\ J_0 \frac{\partial \theta}{\partial x} &= -\mu e^\psi \frac{\partial}{\partial y} (n e^{-\psi}). \end{aligned}$$

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, obtaining:

$$\frac{\partial}{\partial x} \left(\frac{e^{-\psi}}{\mu} \frac{\partial \theta}{\partial x} \right) + \frac{\partial}{\partial y} \left(\frac{e^{-\psi}}{\mu} \frac{\partial \theta}{\partial y} \right) = 0. \quad (I-11)$$

In the iterative procedure described above, (I-11) may be regarded as a second order differential equation whose solution gives $\theta(x,y)^{(m)}$, knowing $\psi(x,y)^{(m)}$ at iteration step m . Then, using (I-09), the current density vector $\bar{J}(x,y)^{(m)}$ is known, once the constant $J_0^{(m)}$ has been evaluated. The method of determining that constant will be described after we establish the boundary conditions imposed upon ψ , θ , and n . Before doing so, however, we show how knowledge of ψ and $\bar{J}$ enables the determination of the electron number density n by means of a similar partial differential equation derivable from (I-10).

2.06 The Determination of $n(x,y)$. - Starting with (I-10), divide both sides by μe^ψ , then take the divergence of both sides, obtaining:

$$\nabla^2 (ne^{-\psi}) = \nabla \cdot \left(\frac{e^{-\psi}}{\mu} \bar{J} \right).$$

Then, using the vector identity $\nabla \cdot (g\bar{F}) = g\nabla \cdot \bar{F} + \bar{F} \cdot \nabla g$, where g is any scalar function, while $\bar{F}$ is any vector function, this becomes:

$$\nabla^2 (ne^{-\psi}) = \frac{e^{-\psi}}{\mu} \nabla \cdot \bar{J} + \bar{J} \cdot \nabla \left(\frac{e^{-\psi}}{\mu} \right).$$

In view of (I-06) the leading term on the right side of this intermediate result vanishes. Thus, we obtain:

$$\nabla^2 (ne^{-\psi}) = \bar{J} \cdot \nabla \left(\frac{e^{-\psi}}{\mu} \right). \quad (I-12)$$

As above, in the case of the stream function $\theta(x,y)$, (I-12) may be regarded as a differential equation in the electron density $n(x,y)$ in the iterative procedure. That is to say, at iteration step m , with $\psi^{(m)}$ and $\theta^{(m)}$ (hence $\bar{J}^{(m)}$) being known, (I-12) will give $n(x,y)^{(m)}$.

It is of interest to note that, since $ne^{-\psi} = e^{-\phi_n}$, from (I-02) then (I-12) is really a differential equation for the electron quasi-Fermi potential ϕ_n . Recognition of this assists in establishing boundary conditions for the solutions of (I-12).

2.07 Boundary Conditions.—We regard the left and right borders of the rectangular region of Figure I-1(b) as symmetry planes, such that ψ , θ , n , p , and J_y are even functions of x , while J_x is an odd function of x . Image theory then requires that these functions are all periodic in x , with a periodicity of twice the channel length XL of Figure I-1(b). A consequence of this implied periodicity is that the source and drain regions shown dotted in the geometry of Figure I-1(a) are essentially "squeezed out of the picture." They are replaced by infinitesimally thick source and drain "contacts" which serve as "sources or sinks" for electric and current flux lines. This being the case, the electric potential ψ and number density n assigned to those "contacts" requires careful scrutiny.

We adopt the common convention of specifying the potential of the gate, drain, and substrate with respect to that of the source. Thus, we specify $\psi = 0$ in the charge-neutral source region, even though that region has been "squeezed out of the picture" in our model. In the absence of any externally applied bias voltages, therefore, the substrate will assume its thermal equilibrium potential ψ_{EQ} , given by:

$$\psi_{EQ} = -\ln(N_D N_A) , \quad (I-13)$$

which will also be the potential assumed by the gate electrode, assuming zero difference in the work function of the gate metal and the p-substrate. If the substrate is then externally biased at a voltage V_{SUB} , and the gate at a voltage V_G (both measured with respect to the charge-neutral source) the potentials assumed by those electrodes will have the values shown in Figure I-3(a). With regard to that figure, we assume that the lower border of the region modeled extends sufficiently deep into the charge-neutral substrate region that it can be regarded as an equipotential surface to which we ascribe the properties of a substrate contact electrode.

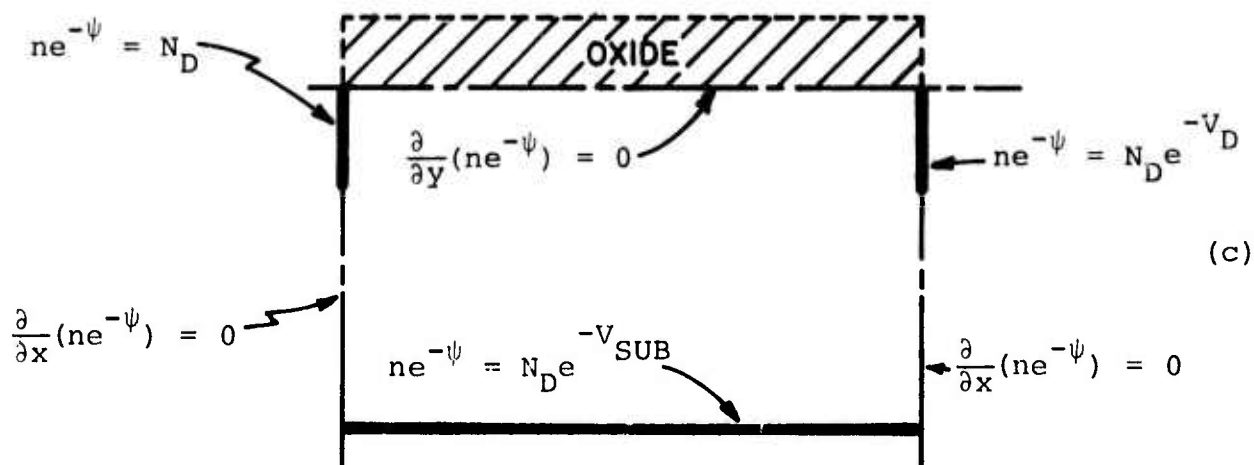
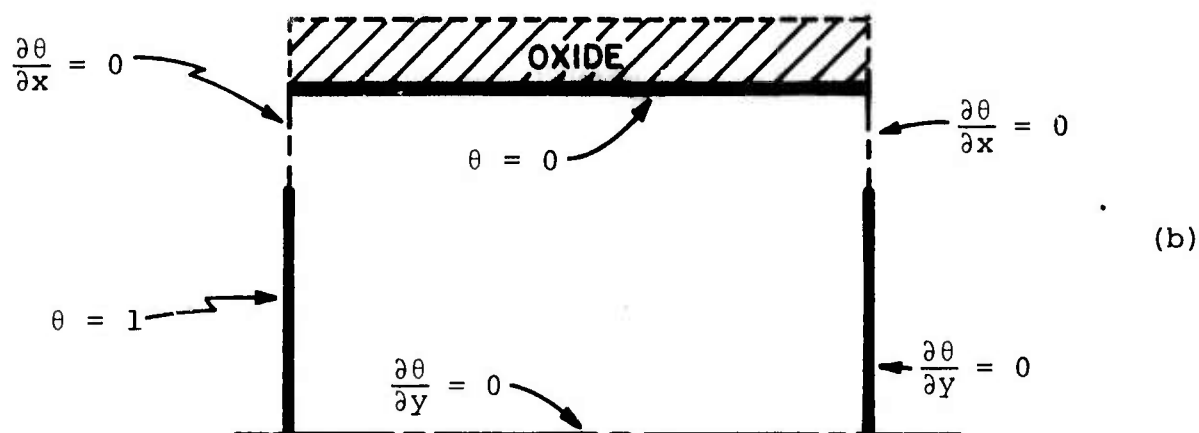
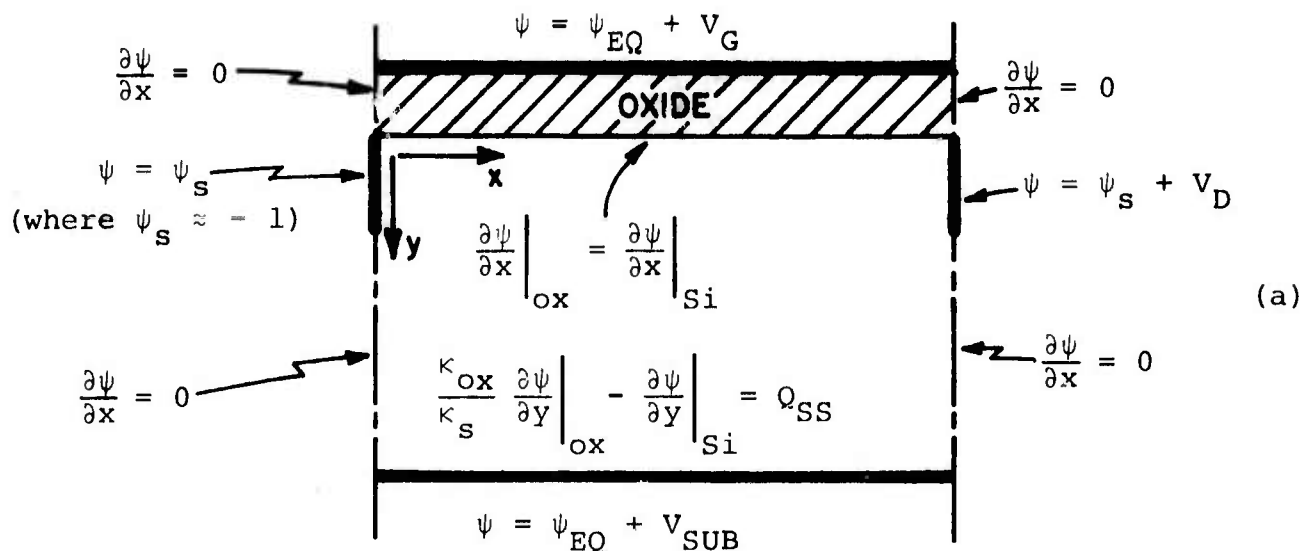


Figure I-3. Boundary conditions for: (a) $\psi(x, y)$, (b) $\theta(x, y)$, and (c) $ne^{-\psi}(x, y)$.

What potential should be assigned to the source and drain "contacts" in Figure I-3(a)? Assuming identical doping of the drain and source, the source and drain "contacts" will assume identical "built-in" potentials ψ_s , in the absence of any external biasing. With a voltage V_D applied to the drain, the drain "contact" will assume the value shown in that figure.

There remains to be specified the potential ψ_s itself, which is the potential assumed at the plane of the metallurgical junction of the n^+ source and the p -substrate. Kennedy [5] shows rigorously that, for an abrupt junction, that potential is given by*:

$$\psi_s = \psi_T \left[\frac{N_D}{N_A + N_D} - 1 \right] + \left[\frac{N_A - N_D}{N_A + N_D} \right] \left[1 - \exp(-\psi_T) \right] \quad (I-14)$$

where ψ_T is the magnitude of the total potential drop appearing across the p - n junction, i.e. the sum of the "built-in" potential plus any applied potential difference:

$$\psi_T = \ln(N_D N_A) + V_{\text{appl.}} \quad (I-15)$$

(Positive $V_{\text{appl.}}$ in (I-15) causes the junction to be reverse biased.)

In spite of the fact that (I-14) shows ψ_s to be, in general, a function of the total voltage drop appearing across the junction, an assessment of typical numerical values shows that ψ_s can be assumed constant, independent of junction voltage, except in cases where it is heavily forward-biased. For example, with $n_i' = 1.5 \times 10^{10} \text{ cm}^{-3}$, $N_D' = 10^{19} \text{ cm}^{-3}$, $N_A' = 10^{16} \text{ cm}^{-3}$ (unnormalized units), $\ln(N_D N_A)$ assumes the value 33.7, and the exponential term in (I-14) is entirely negligible except for negative values of V_{appl} approaching that magnitude. Since $N_D \gg N_A$, (I-14) simplifies

[5] D. P. Kennedy, *IEEE Trans. Electron Devices*, ED-22, 988 (1975).

* The factor of unity appearing in the leading bracket is not present in Kennedy's result. It accounts for the choice of zero potential as the n^+ region, whereas Kennedy chose the p -region.

tremendously under such conditions, becoming:

$$\psi_s \approx -1.0 \quad (\text{I-16})$$

Since the drain junction is always reverse biased, with $V_D > 0$, (I-16) is always valid at the drain "contact."

Is the above approximation valid at the source junction, which does become forward biased as V_G is made positive, due to band bending at the p-substrate surface? At the onset of strong inversion, the total potential drop across the n^+ -p source junction right at that surface becomes $\psi_T = \ln(N_D/N_A)$. Thus, using the same values for N_D and N_A as in the above numerical example, $\psi_T = 6.9$, and neglecting the exponential term in (I-14) is still a valid approximation for estimating ψ_s . We conclude that (I-16) can be used to specify ψ_s at the source contact, with good accuracy, for N_D/N_A ratios of 10^3 or higher. The approximation begins to become questionable when $N_D/N_A \approx 10^2$ or less, but at the source contact only.

The remaining boundary conditions shown in Figure I-3(a) assure the symmetry assumed at the left and right borders of the region modeled, and assure correct matching of the solutions of (I-01) and (I-07) at the silicon-oxide interface. Note that the condition imposed on the normal derivatives of ψ at that interface allow for the inclusion of surface charge Q_{ss} at that interface, a feature not included by Mock. (The equation stated for the normal derivatives takes on a slightly unfamiliar form because of the normalized variables used.)

Figure I-3(b) states the boundary conditions imposed upon the stream function θ . The choice of $\theta = 1$ and $\theta = 0$ along the surfaces shown assures no normal current flow at those surfaces, as does the specification $\partial\theta/\partial y = 0$ at the right edge, as shown. This latter condition allows that surface (which is obviously a contour of constant θ) to "float" to assume a value of θ consistent with Kirchhoff's current law in the event that the substrate is biased in such a manner as to draw current.

The choice of unity and zero as the bounds on θ is arbitrary, but a convenient one. (Any other choice would result only in a different numerical value of the constant J_0 appearing in (I-09). The manner in which J_0 is determined remains to be described below.) The particular choice used allows rapid assessment of where current is flowing in the region. Since contours of constant θ are current streamlines, then the value of θ at any x,y gives directly the fraction of the total source current crossing the vertical plane through (x,y) between the surface and that point.

Finally, the specifications on the normal derivatives of θ shown at the left and right borders are consistent with the treatment of those borders as symmetry planes, while that specified at the lower border is consistent with our treatment of it as a substrate contact electrode. Note that these conditions imply, through (I-09), that current flux lines enter or leave those surfaces perpendicular to them.

Figure I-3(c) states the boundary conditions on $ne^{-\psi}(x,y)$. At those surfaces where the normal derivatives of $ne^{-\psi}$ are set to zero, no current crosses those surfaces, in accordance with (I-10). At the substrate contact, the value specified assures that $n = 1/N_A$, $p = N_A$. To see this, recall (I-02) and (I-03):

$$n = e^{\psi} e^{-\phi_n} \quad (I-02)$$

$$p = e^{\phi_p} e^{-\psi} \quad (I-03)$$

Setting $\psi = -\ln(N_D N_A) + V_{SUB}$ in (I-03) and requiring $p = N_A$, one finds for ϕ_p :

$$\phi_p = -\ln N_D + V_{SUB}.$$

Since $\phi_n = \phi_p$ in the charge-neutral substrate region at the substrate contact, substitution of this result into (I-02) yields $n = 1/N_A$, as it should be. Similarly, at the source and drain contacts, n must take on the value $N_D e^{-1}$, since the potential

of these contacts is one normalized unit below the potential of the charge-neutral source and drain regions, respectively. The boundary conditions stated there assure that this is the case.

2.08 Determination of the Current Constant J_0 .—The constant J_0 appearing in (I-9), which relates the current density vector $\vec{J}$ to the stream function θ , is obtained in the following manner. Writing the x-component of (I-10):

$$\mu e^{\psi} \frac{\partial}{\partial x} (n e^{-\psi}) = J_0 \frac{\partial \theta}{\partial y} ,$$

divide by μe^{ψ} , then integrate with respect to x along a horizontal line ($y = \text{constant}$) which intersects both the source and the drain "contacts." (Any value of $y < Y_D$ in Figure I-1(b) will qualify.) One obtains:

$$n e^{-\psi} \Big|_{x=XL} - n e^{-\psi} \Big|_{x=0} = J_0 \int_0^{XL} \frac{e^{-\psi}}{\mu} \frac{\partial \theta}{\partial y} dx .$$

Substituting the boundary conditions of Figure I-3(c) in the left side of this intermediate result:

$$N_D \left[e^{-V_D} - 1 \right] = J_0 \int_0^{XL} \frac{e^{-\psi}}{\mu} \frac{\partial \theta}{\partial y} dx .$$

Finally, solving for J_0 , one obtains:

$$J_0 = \frac{-N_D \left[1 - e^{-V_D} \right]}{\int_0^{XL} \frac{e^{-\psi}}{\mu} \frac{\partial \theta}{\partial y} dx} . \quad (I-17)$$

This is the same result used by Mock for calculating J_0 (reference [1], eq. (I-10)), except for the factor N_D appearing in (I-17) which

is absent in Mock's result. The reason for this difference lies in Mock's choice of setting the electron quasi-Fermi potential $\phi_n = 0$ in the charge-neutral source region, which makes $\psi = \ln(N_D)$ in that region, rather than the $\psi = 0$ reference value that we have chosen to use. Note that the sign of J_0 changes if one performs the integration from $x = 0$ to the "image" drain contact at $x = -XL$. Thus, J_0 is an odd function of x , as it should be.

2.09 Finite-Difference Equations.—Figure I-4 singles out a mesh point of the graded lattice shown in Figure 2, together with its four nearest neighbors. We distinguish such mesh points by means of indices i, j , with i and j assuming the values 1,1 at the lower left corner of the lattice shown in Figure 2. The index i signifies x position, j signifies y position. The five-point "star" of Figure I-4 forms the basis for forming finite-difference approximations of (I-01) and (I-07) for ψ , of (I-11) for θ , and of (I-12) for $ne^{-\psi}$. Note the definitions of $\Delta x(i)$, $\Delta y(j)$, their average values $\tilde{\Delta x}(i)$, $\tilde{\Delta y}(j)$, and the indexing convention adopted for them in Figure I-4.

Except at the silicon-oxide interface, the finite-difference equation for electric potential ψ is:

$$\begin{aligned} \tilde{\Delta x}(i) & \left\{ \left(\frac{\psi(i,j) - \psi(i,j+1)}{\Delta y(j+1)} \right) + \left(\frac{\psi(i,j) - \psi(i,j-1)}{\Delta y(j)} \right) \right\} \\ + \tilde{\Delta y}(j) & \left\{ \left(\frac{\psi(i,j) - \psi(i+1,j)}{\Delta x(i+1)} \right) + \left(\frac{\psi(i,j) - \psi(i-1,j)}{\Delta x(i)} \right) \right\} \end{aligned} \quad (I-18)$$

$$\begin{aligned} &= \tilde{\Delta x}(i) \tilde{\Delta y}(j) [p(i,j) - n(i,j) - N_A(i,j)], \text{ in silicon} \\ &= 0, \text{ in oxide.} \end{aligned}$$

For mesh points (i,j) coinciding with the silicon-oxide interface, a special form of (I-18) results by applying Gauss' law to equate the electric flux leaving the dotted rectangle shown in Figure I-4 to the charge enclosed within that rectangle. One obtains:

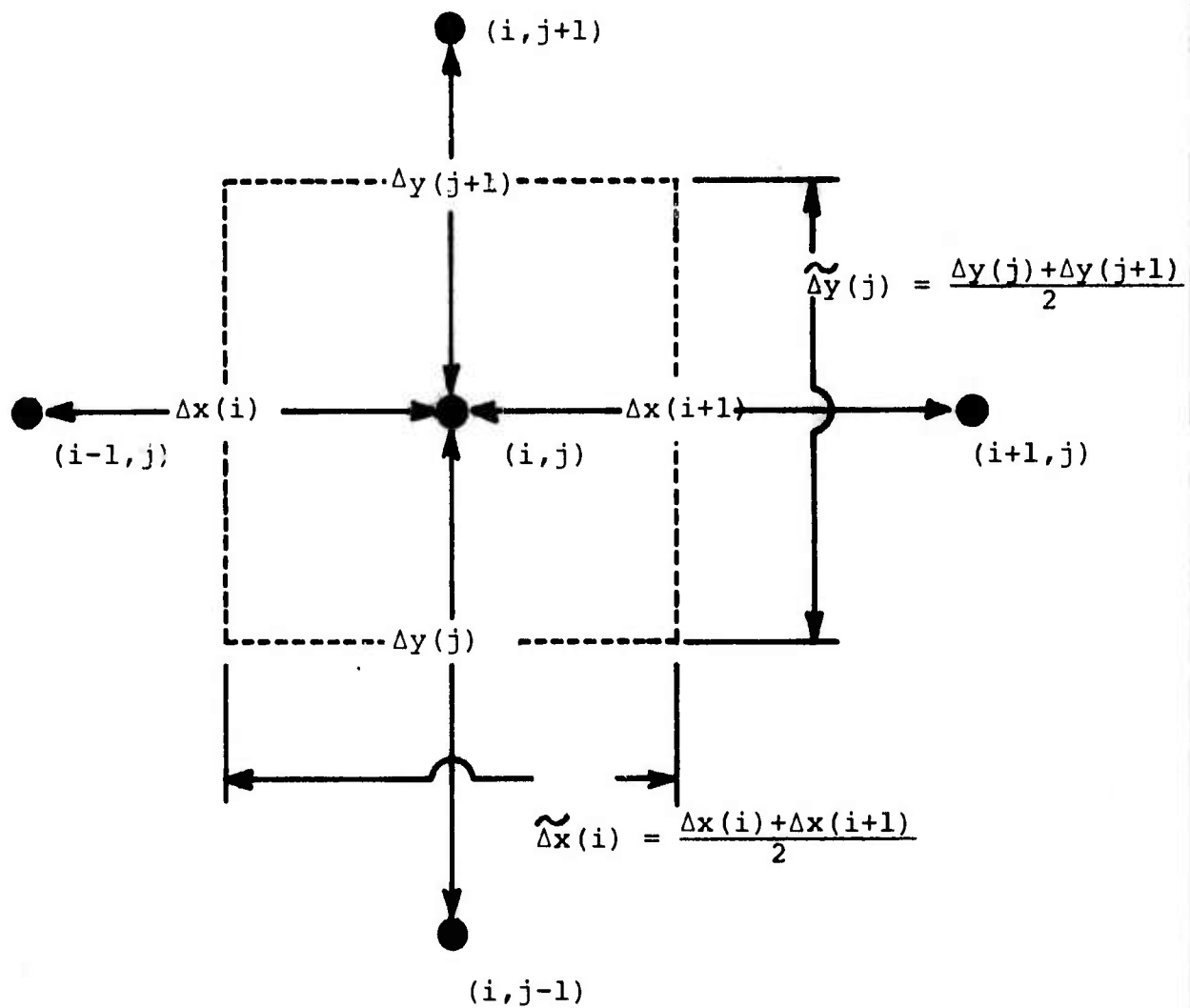


Figure I-4. Five-point star of mesh points used in forming finite-difference equations.

$$\begin{aligned}
& \tilde{\Delta x}(i) \left\{ \left(\frac{\kappa_{ox}}{\kappa_s} \right) \left(\frac{\psi(i,j) - \psi(i,j+1)}{\Delta y(j+1)} \right) + \left(\frac{\psi(i,j) - \psi(i,j-1)}{\Delta y(j)} \right) \right\} \\
& + \left[\frac{\left(\frac{\kappa_{ox}}{\kappa_s} \right) \Delta y(j+1) + \Delta y(j)}{2} \right] \left\{ \left(\frac{\psi(i,j) - \psi(i+1,j)}{\Delta x(i+1)} \right) + \left(\frac{\psi(i,j) - \psi(i-1,j)}{\Delta x(i)} \right) \right\} \\
& = Q_{ss} \tilde{\Delta x}(i) + \frac{\tilde{\Delta x}(i) \tilde{\Delta y}(j)}{2} [p(i,j) - n(i,j) - N_A(i,j)] .
\end{aligned} \tag{I-19}$$

Here, the first term on the right represents surface charge at the interface, while the second term represents bulk charge in the portion of the dotted rectangle lying below the interface. Equation (I-19) assures the satisfaction of the boundary conditions on ψ at the silicon-oxide interface stated in Figure I-3(a). Our method of treating this boundary condition differs from Mock's.

The finite difference equation for the stream function θ , using the same 5-point star of mesh points, becomes:

$$\begin{aligned}
& \tilde{\Delta x}(i) \left\{ a(i,j-1/2) \left(\frac{\theta(i,j) - \theta(i,j-1)}{\Delta y(j)} \right) + a(i,j+1/2) \left(\frac{\theta(i,j) - \theta(i,j+1)}{\Delta y(j+1)} \right) \right\} \\
& + \tilde{\Delta y}(j) \left\{ a(i-1/2,j) \left(\frac{\theta(i,j) - \theta(i-1,j)}{\Delta x(i)} \right) + a(i+1/2,j) \left(\frac{\theta(i,j) - \theta(i+1,j)}{\Delta x(i+1)} \right) \right\} \\
& = 0 .
\end{aligned} \tag{I-20}$$

where

$$\left. \begin{aligned}
a(i,j-1/2) &= \frac{e^{-\psi(i,j-1)} - e^{-\psi(i,j)}}{\psi(i,j) - \psi(i,j-1)} \\
a(i,j+1/2) &= \frac{e^{-\psi(i,j+1)} - e^{-\psi(i,j)}}{\psi(i,j) - \psi(i,j+1)} \\
a(i-1/2,j) &= \frac{e^{-\psi(i-1,j)} - e^{-\psi(i,j)}}{\psi(i,j) - \psi(i-1,j)} \\
a(i+1/2,j) &= \frac{e^{-\psi(i+1,j)} - e^{-\psi(i,j)}}{\psi(i,j) - \psi(i+1,j)}
\end{aligned} \right\} \tag{I-21}$$

Equations (I-21) are approximations for the function $e^{-\psi}$ at x,y positions midway between the points of the star (i.e., where the dotted rectangle of Figure I-4 intersects the "rays" of the star). These approximations are obtained by assuming linear variations of ψ between neighboring mesh points, then averaging $e^{-\psi}$ along the line interconnecting them. This method of approximating $e^{-\psi}$ midway between mesh points is a critical factor discussed in detail by Mock. (See reference [1], Section 6.)

The finite-difference equation (I-20) for θ differs from that used by Mock for a fundamental reason. Mock chooses the lattice of points at which $\theta(i,j)$ is defined not to be the same set of points at which $\psi(x,y)$ and $n(x,y)$ are defined (i.e., the star of points in Figure I-4), but rather to correspond to the location of the corners of the dotted rectangle shown in that figure. Thus, his lattice of points for θ "interlaces" the lattice of points used for ψ and n . While this convention leads to some computational convenience in his method for treating conditions at the silicon-oxide interface (he places that interface so as to coincide with his θ lattice, not his ψ,n lattice) and in his method for determining n (which differs from ours) it leads to severe problems in meeting the boundary conditions on θ at the left border of Figure I-3(b), where θ must be specified to be constant (i.e., that portion of the border where $\theta = 1$ in Figure I-3(b)), because the lattice of points at which θ is defined cannot coincide with that border, using Mock's convention. Thus, how can one specify $\theta = 1$ there?

Again, using the same five point star of lattice points in Figure I-4, the finite-difference approximation of (I-12) determining the function $ne^{-\psi}(x,y)$ becomes:

$$\begin{aligned}
& J_x(i) \left\{ \left(\frac{ne^{-\psi}(i,j) - ne^{-\psi}(i,j-1)}{\Delta y(j)} \right) + \left(\frac{ne^{-\psi}(i,j) - ne^{-\psi}(i,j+1)}{\Delta y(j+1)} \right) \right\} \\
& + \Delta y(j) \left\{ \left(\frac{ne^{-\psi}(i,j) - ne^{-\psi}(i-1,j)}{\Delta x(i)} \right) + \left(\frac{ne^{-\psi}(i,j) - ne^{-\psi}(i+1,j)}{\Delta x(i+1)} \right) \right\} \\
& = -\Delta x(i) \Delta y(j) \left\{ J_x(i,j) \frac{\partial}{\partial x} \frac{e^{-\psi}}{\mu}(i,j) + J_y(i,j) \frac{\partial}{\partial y} \frac{e^{-\psi}}{\mu}(i,j) \right\} \quad (I-22)
\end{aligned}$$

where, in accordance with (I-09),

$$\begin{aligned}
J_x(i,j) &= J_0 \frac{\partial \theta}{\partial y}(i,j) \\
J_y(i,j) &= -J_0 \frac{\partial \theta}{\partial x}(i,j). \quad (I-23)
\end{aligned}$$

Note that the determination of the numerical values of the right-side of (I-22) requires the numerical evaluation of several partial derivatives, namely $\partial\psi/\partial x$ and $\partial\psi/\partial y$ in order to determine the field-dependent mobility μ , then the determination of $\partial/\partial x(e^{-\psi}/\mu)$ and $\partial/\partial y(e^{-\psi}/\mu)$, and finally the determination of $\partial\theta/\partial x$ and $\partial\theta/\partial y$, in order to find J_x and J_y . If the spacing between mesh points were uniform, the determination of these partial derivatives is readily achieved by simple differencing. For example, $\partial\theta/\partial x(i,j) = [\theta(i+1,j) - \theta(i-1,j)]/2\Delta x(i)$. But with a graded lattice of points, a more accurate algorithm for determining these partials is required.

We base our numerical algorithm for so doing on the following approach. Let $g(x,y)$ be the function whose partial derivatives are required, with $g(x,y)$ defined numerically at each of the lattice points (i,j) . Consider determining $\partial/\partial x(g(x,y))$ at the mesh point (i,j) . Knowing the mesh spacings $\Delta x(i)$ and $\Delta x(i+1)$, fit a parabola through the points $g(i-1,j)$, $g(i,j)$, and $g(i+1,j)$, obtaining an analytic expression of the form:

$$\left. g(x,y) \right|_{y=\text{const}} = ax^2 + bx + c$$

where the constants a, b , and c are known. Then $\partial g / \partial x(x, y) = 2ax + b$, which can be evaluated at the value of x corresponding to the point (i, j) . A similar approach yields $\partial g / \partial y$.

Implementation of the above approach leads to the following finite-difference expressions for $\partial g / \partial x(i, j)$ and $\partial g / \partial y(i, j)$:

$$\frac{\partial g}{\partial x}(i, j) = \frac{\Delta x^2(i)g(i+1, j) - \Delta x^2(i+1)g(i-1, j) + [\Delta x^2(i+1) - \Delta x^2(i)]g(i, j)}{\Delta x(i)\Delta x(i+1)[\Delta x(i) + \Delta x(i+1)]} \quad (I-24)$$

$$\frac{\partial g}{\partial y}(i, j) = \frac{\Delta y^2(j+1)g(i, j-1) - \Delta y^2(j)g(i, j+1) + [\Delta y^2(j) - \Delta y^2(j+1)]g(i, j)}{\Delta y(j)\Delta y(j+1)[\Delta y(j) + \Delta y(j+1)]} .$$

These expressions for $\partial g / \partial x$ and $\partial g / \partial y$ reduce to the simple differencing scheme cited above, in the event of a uniform lattice point spacing.

It should be noted that Mock does not use the finite difference equation (I-22) to determine the electron number density n . Instead, he develops a method which basically involves the numerical evaluation of a line integral, starting from the source or drain contact where n is known, and terminating at the point (i, j) where n is to be determined. Not only does this method appear to the author to invite the accumulation of systematic numerical errors inherent in the discretization scheme used to evaluate the line integral, but also Mock notes that the method must be applied with care to avoid the development of numerical instabilities.

There remains to be described the numerical determination of the hole density $p(i, j)$, and of the current constant J_0 . Consider first $p(i, j)$. In Section 2.07, while discussing boundary conditions, it was shown that the hole quasi-Fermi potential ϕ_p at the substrate contact is given by:

$$\phi_p = -\ln N_D + V_{SUB} .$$

Then, in view of (I-05), which enforces the assumed condition of zero hole current, this is the value of ϕ_p everywhere throughout the substrate region. Equation (I-03) then gives for the hole density:

$$p(i,j) = \frac{e^{V_{SUB}}}{N_D} e^{-\psi(i,j)} . \quad (I-25)$$

Note that, since $\psi = -\ln(N_D N_A) + V_{SUB}$ at the substrate electrode, $p = N_A$ there, as it should.

Finally, consider the numerical implementation of (I-17) to determine the current constant J_0 . Here, as did Mock, we implement the integration indicated along a horizontal mesh line, $j = \text{constant}$, intersecting the source and drain "contacts," using the trapezoidal rule.

2.10 Method Used for Solving the Finite-Difference Equations.-

The finite difference equations (I-18), (I-20), and (I-22) for ψ , θ , and $ne^{-\psi}$, respectively, each may be cast in the form:

$$\begin{aligned} B(i,j)V(i,j-1) + D(i,j)V(i-1,j) + E(i,j)V(i,j) \\ + F(i,j)V(i+1,j) + H(i,j)V(i,j+1) = Q(i,j) . \end{aligned} \quad (I-26)$$

where the variable V stands for either ψ , θ , or $ne^{-\psi}$, and the coefficients B , D , E , F , H , and the function Q , are to be identified with the corresponding terms in (I-18), (I-20), or (I-22). If we define column vectors $[V]$ and $[Q]$:

$$[V] = \begin{bmatrix} V_{11} \\ V_{21} \\ \vdots \\ V_{n1} \\ V_{12} \\ V_{22} \\ \vdots \\ V_{n2} \\ \vdots \end{bmatrix} \quad [Q] = \begin{bmatrix} Q_{11} \\ Q_{21} \\ \vdots \\ Q_{n1} \\ Q_{12} \\ Q_{22} \\ \vdots \\ Q_{n2} \\ \vdots \end{bmatrix}$$

then the set of equations (I-26) may be expressed as the matrix equation:

$$[M] \cdot [V] = [Q] \quad (I-27)$$

where $[M]$ is a sparse square matrix having only five non-zero diagonals:

$$[M] = \begin{bmatrix} E_{11} & F_{11} & 0 & 0 & & & H_{11} & 0 \\ D_{21} & E_{21} & F_{21} & 0 & 0 & & H_{21} & \\ 0 & D_{31} & E_{31} & F_{31} & 0 & 0 & H_{31} & \\ 0 & 0 & D_{41} & E_{41} & F_{41} & 0 & 0 & H_{41} \\ & & & & & & & \\ 0 & B_{ij} & & 0 & D_{ij} & E_{ij} & F_{ij} & 0 \\ & & & & & & & H_{ij} \end{bmatrix}$$

J+1
ELEMENTS

J+1
ELEMENTS

If the dimensions of the lattice of points used to model the rectangular region, as in Figure I-2, are J by K , then the matrix M has JK rows and JK columns.

We adopt the very fast iterative matrix factorization method developed by Stone [6] for dealing with the problem of finding the elements of the column vector $[V]$, given $[Q]$ and $[M]$, when $[M]$ has the five-diagonal structure of that shown above. In this manner, the solutions of (I-18), (I-20), and (I-22) for ψ , θ , and $ne^{-\psi}$ are obtained.

[6] H. L. Stone, *SIAM Jour. Numer. Anal.*, 5, 530 (1968).

2.11 Status of the Model. -Our computer model consists of an executive main program and ten subroutines upon which it calls:

MAIN	Executive program
RELAX	Implements the iterative procedure
STONE	Implements Stone's method
POISSN	Calculates ψ
STREAM	Calculates θ
FERMI	Calculates $ne^{-\psi}$
GRID	Defines the graded lattice of points
BORDER	Assigns boundary conditions
CURRNT	Calculates the current constant J_0
PSINIT	Calculates initial 1-D potential
OUTPUT	Provides output data

Flow charts for these have been developed and their reduction to FORTRAN code is now in progress. Debugging will commence within the next two weeks.

3.0. TWO-DIMENSIONAL BIPOLAR TRANSISTOR MODEL

We have a computer program which was developed to provide a two-dimensional model of bipolar transistors. This computer program was developed for us, under subcontract, by Professor David Navon, at the University of Massachusetts, under a contract between the University of Florida and IBM, in 1974, which was concerned with investigations of several factors concerning power transistors. Under the terms of that contract, all results from these studies are non-proprietary, and we are free to use this computer model as we choose.

This computer model deals with the same basic equations listed in the preceding section for the MOSFET model (equations (I-01) through (I-06)) except that the assumptions of zero electron-hole recombination and of negligible hole current invoked there are removed. The recombination rate modeled assumes the well known Shockley-Reed-Hall steady state law corresponding to

uniformly distributed recombination centers with a single energy level at the center of the bandgap. Figure I-5 shows the rectangular region with which this model deals.

Navon's program was written for use with a CDC CYBER 74 computer whose word length precludes the need for double precision arithmetic operations. During this reporting interval, we have adapted this program for use with an IBM 370/65. This required the introduction of double precision arithmetic, and numerous other program modifications. The program is now operational, and provides a starting point for the development of a revised version devised to handle bipolar transistor structures more typical of IC geometries, e.g., collector contact at the upper surface of the structure, the inclusion of a buried n^+ layer, etc. Work on this revised computer model will be deferred until the 2-D MOSFET model work has been completed.

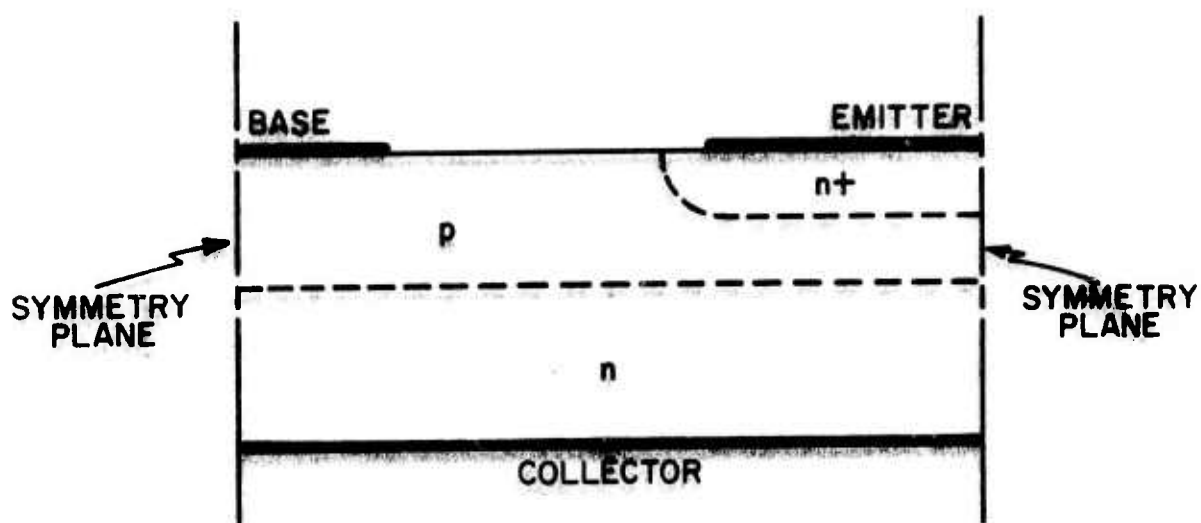


Figure I-5. Basic bipolar transistor structure modeled by the transistor analysis computer program.

Chapter II

A One-Dimensional Mathematical Model for
MOSFET Operation

R. F. Motta and D. P. Kennedy

	<u>Page</u>
1.0 Introduction.....	49
2.0 An Earlier Theory for MOSFET Operation.....	50
2.1 General Mathematical Development.....	51
2.2 Interpretation of the Separation.....	53
Parameter $\lambda(V_s)$	
2.3 The Volt-Ampere Characteristics of a MOSFET..	54
2.4 The Saturation Mode of Operation.....	57
2.5 Calculation of Q_{io}	59
2.6 Calculation of λ_o	60
3.0 Model Evaluation and Refinements.....	62
3.1 Electric Current Saturation.....	63
3.2 Weak Inversion Operation.....	70
4.0 Conclusions.....	71
5.0 List of Symbols.....	71

Chapter II

A One-Dimensional Mathematical Model for MOSFET Operation

R. F. Motta and D. P. Kennedy

1.0 Introduction

Elementary MOSFET theory [1-5] contains simplifications and approximations that render it inapplicable in many situations. However, this theory offered great mathematical simplicity and, in the early years of MOS technology, it produced satisfactory agreement with experiment. In time, the rapid growth of semiconductor integrated circuit (IC) technology created applications for which this theory was found to be inadequate. Specifically, when applied to the weak inversion mode of operation and/or to short-channel structures, this theory yielded results in poor agreement with experiment.

This situation was recognized by many workers and, subsequently, extensive research efforts were undertaken to alleviate this problem. From these efforts, a multitude of new theories emerged as modifications of elementary MOSFET theory. In a previous study [6], we carefully reviewed these new theories and found them to be extensions of elementary MOSFET theory without consideration of inherent weaknesses in this theory. In particular, this theory contains no constraint whereby electric current

-
- [1] H.K.J. Ihantola and J. L. Moll, *Solid State Electronics*, 7, 423 (1964).
 - [2] C. T. Sah and H. C. Pao, *IEEE Trans. Electron Devices*, ED-13, 393 (1966).
 - [3] R. H. Crawford, *MOSFET in Circuit Design*, McGraw-Hill Company, N.Y. (1967).
 - [4] A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley and Sons, Inc., N.Y. (1967).
 - [5] S. M. Sze, *Physics of Semiconductor Devices*, Wiley Interscience, N.Y. (1969).
 - [6] D. P. Kennedy and F. A. Lindholm, *Physical Models of MOSFET Devices*, Final Report, HDL-CR-75-193-1 (1975).

continuity can be attained in the source-drain channel of this semiconductor device. Moreover, many of these new theories are heuristic in nature and require adjustable parameters to obtain agreement with experiment. For any particular device structure, these adjustable parameters must usually be determined by laboratory experiments, thus increasing development costs of new integrated circuits.

In a previous study [7], a new theory for MOSFET operation was proposed that assures electric current continuity in this semiconductor device. Furthermore, this new theory for MOSFET operation was reduced to a simplified one-dimensional system of equations that are suitable for engineering purposes. The adequacy of this theory was demonstrated for long-channel structures operating in the strong inversion mode. However, although implicitly containing the current physics, this new theory also failed to yield satisfactory results for the weak inversion mode and/or for short-channel structures. Therefore, the main direction of the present research has been to alleviate these difficulties.

At the inception of this research program we had completed an important revision of this theory. We therefore initiated an extensive evaluation of our revised theory, for a wide range of device structures and bias conditions. This evaluation revealed additional errors in our model which have necessitated other revisions. In this report, we present these revisions and discuss progress we have made in the development of this model.

2.0 An Earlier Theory for MOSFET Operation

In this section, it is our purpose to present an overview of an earlier theory for MOSFET operation. This presentation gives the basic structure of our theory, prior to a number of refinements we made during the present research program. Therefore, in addition to acquainting the reader with this earlier work, the present section constitutes a necessary preliminary to the discussion of these refinements which follows in §3.0.

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

2.1 General Mathematical Development

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

$$I_D = -W\mu \left[Q_i \frac{dv_s}{dx} - \frac{kT}{q} \frac{dQ_i}{dx} \right]. \quad (\text{II-1})$$

In this expression, I_D represents the total electric current parallel to the oxide-semiconductor interface (see Fig. II-1).

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 (II-1) we have

$$0 = \frac{dI_D}{dx} = W\mu \left[Q_i \frac{d^2v_s}{dx^2} + \frac{dQ_i}{dx} \frac{dv_s}{dx} - \frac{kT}{q} \frac{d^2Q_i}{dx^2} \right]. \quad (\text{II-2})$$

From the chain rule of differentiation,

$$\frac{dQ_i}{dx} = \frac{dQ_i}{dv_s} \frac{dv_s}{dx} \quad (\text{II-3a})$$

$$\frac{d^2Q_i}{dx^2} = \frac{dQ_i}{dv_s} \frac{d^2v_s}{dx^2} + \frac{d^2Q_i}{dv_s^2} \left(\frac{dv_s}{dx} \right)^2 \quad (\text{II-3b})$$

and, by making these substitutions in (II-2), we obtain

$$\frac{d^2v_s}{dx^2} \left(\frac{dv_s}{dx} \right)^{-2} = \left[\frac{kT}{q} \frac{d^2Q_i}{dv_s^2} - \frac{dQ_i}{dv_s} \right] \left[Q_i - \frac{kT}{q} \frac{dQ_i}{dv_s} \right]^{-1}. \quad (\text{II-4})$$

A physically meaningful modification of (II-4) is realized by introducing into this expression a separation parameter $\lambda(V_s)$, yielding

$$\lambda(V_s) = \frac{d^2v_s}{dx^2} \left(\frac{dv_s}{dx} \right)^{-2} \quad (\text{II-5a})$$

$$\lambda(V_s) = \left[\frac{kT}{q} \frac{d^2Q_i}{dv_s^2} - \frac{dQ_i}{dv_s} \right] \left[Q_i - \frac{kT}{q} \frac{dQ_i}{dv_s} \right]^{-1}. \quad (\text{II-5b})$$

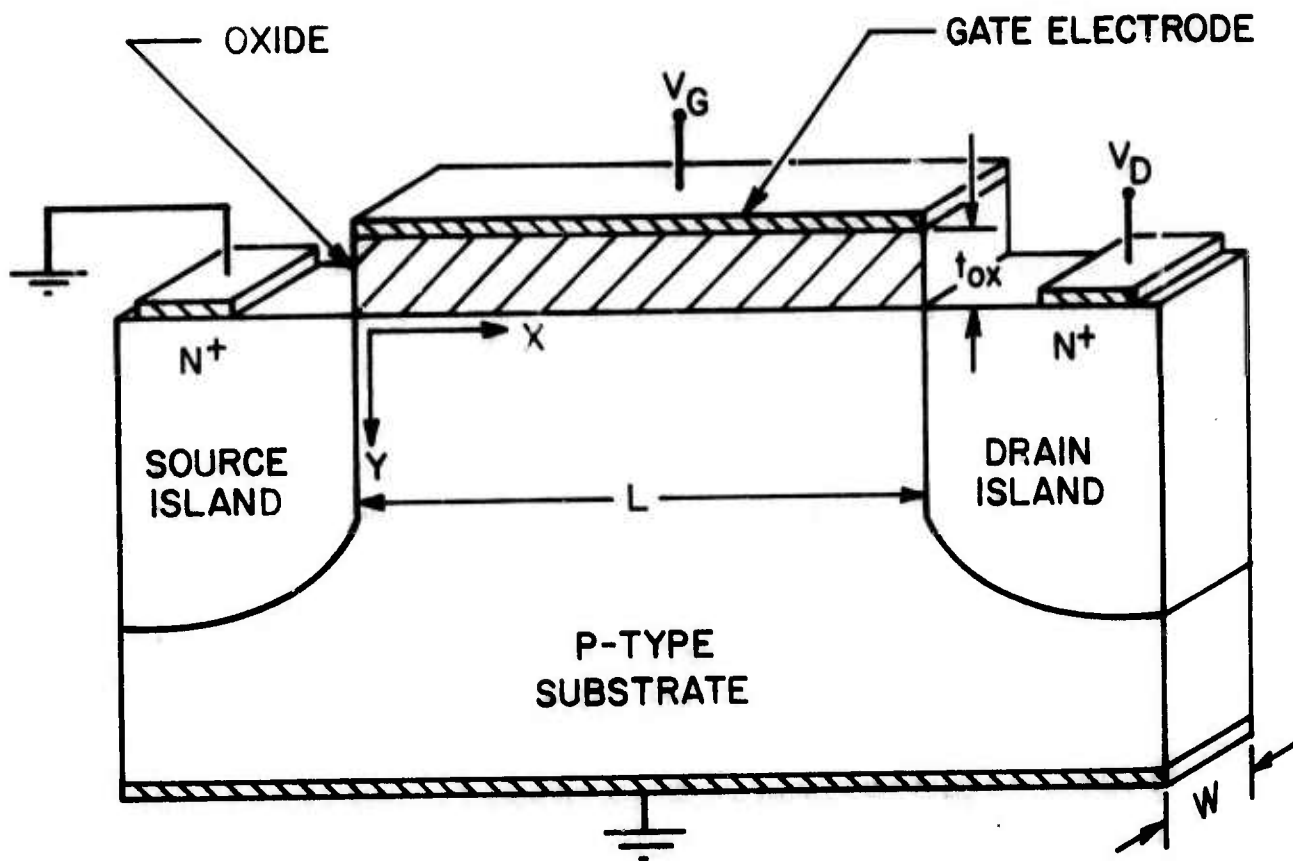


Figure II-1. Illustrative Model of a MOSFET.

Therefore,

$$\frac{d^2 V_s}{dx^2} - \lambda(V_s) \left(\frac{dV_s}{dx} \right)^2 = 0 \quad (\text{II-6a})$$

$$\frac{d^2 Q_i}{dV_s^2} + \left[\lambda(V_s) - \frac{q}{kT} \right] \frac{dQ_i}{dV_s} - \left(\frac{q}{kT} \right) \lambda(V_s) Q_i = 0. \quad (\text{II-6b})$$

Equations (II-6a) and (II-6b) have the solution equations [6]

$$x = x_o - \frac{1}{E_o} \int_{V_o}^{V_s} \exp[\Omega(\xi)] d\xi \quad (\text{II-7a})$$

$$Q_i(V_s) = Q_{io} e^{qV_s/kT} \left\{ e^{-qV_o/kT} - \left[\frac{q}{kT} - \frac{1}{Q_{io}} \left(\frac{dQ_i}{dV_s} \right)_{V_o} \right] \int_{V_o}^{V_s} \exp \left[-\frac{q\xi}{kT} + \Omega(\xi) \right] d\xi \right\}, \quad (\text{II-7b})$$

respectively, where

$$\Omega(\xi) = - \int_{V_o}^{\xi} \lambda(\eta) d\eta. \quad (\text{II-8})$$

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

2.2 Interpretation of the Separation Parameter $\lambda(V_s)$

In order to obtain explicit solutions from (II-7a) and (II-7b), it is necessary to establish the separation parameter $\lambda(V_s)$. From (II-5b), the magnitude of this parameter is given by

$$\lambda(V_s) = - \left(\frac{1}{Q_i} \frac{dQ_i}{dV_s} \right) \left[1 - \frac{kT}{q} \frac{d^2 Q_i}{dV_s^2} \left(\frac{dQ_i}{dV_s} \right)^{-1} \right] \left[1 - \frac{kT}{q} \frac{1}{Q_i} \frac{dQ_i}{dV_s} \right]^{-1}. \quad (\text{II-9})$$

A manipulation of (II-9), in combination with (II-1), yields

$$\lambda(V_s) = \frac{I_{\text{diff}}}{I_D} \left[\frac{q}{kT} - \frac{d}{dV_s} \log_e \left(\frac{dQ_i/dx}{dV_s/dx} \right) \right], \quad (\text{II-10})$$

where I_{diff} is the diffusion component of the total current I_D .

From (II-10), 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 our 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. From (II-10), 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 the first two terms of a Taylor series expansion of (II-10) about a location near the source junction:

$$\lambda(V_s) = \frac{\lambda_o}{1 - A(V_s - V_o)}, \quad (\text{II-11})$$

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

$$A = \frac{1 - (kT/q)\lambda_o}{(V_G - V_o)}. \quad (\text{II-12})$$

2.3 The Volt-Ampere Characteristics of a MOSFET

From the implicit solution equations for $V_s(x)$ and $Q_i(V_s)$ [(II-7a) and (II-7b), respectively] in conjunction with the approximate relation for $\lambda(V_s)$, (II-11), we can readily obtain an expression for the volt-ampere characteristics of a MOSFET. From (II-7a) we obtain

$$V_s = V_o + \frac{1}{A} \left\{ 1 - \left[1 - (\lambda_o + A)E_o x \right]^{\frac{A}{\lambda_o + A}} \right\}, \quad (\text{II-13})$$

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

$$E_o = - \left(\frac{dV_s}{dx} \right)_{x=0} . \quad (II-14)$$

In a similar manner, from (II-7b) we obtain

$$Q_i(V_s) = Q_{io} e^{q(V_s - V_o)/kT} \left\{ 1 - \left[1 - \frac{kT}{q} \frac{1}{Q_{io}} \left(\frac{dQ_i}{dV_s} \right)_{V_o} \right] \cdot z_o^{\lambda_o/A} \exp(z_o) \left[\Gamma\left(1 + \frac{\lambda_o}{A}, z_o\right) - \Gamma\left(1 + \frac{\lambda_o}{A}, z\right) \right] \right\} , \quad (II-15)$$

where

$$z_o = - \frac{q}{kT} \cdot \frac{1}{A} \quad (II-16a)$$

$$z = z_o \left[1 - A(V_s - V_o) \right] . \quad (II-16b)$$

For the range of variables encountered in this boundary value problem, it was proposed that (II-15) could be adequately approximated by the relation

$$Q_i(V_s) \approx Q_{io} \left[1 - A(V_s - V_o) \right]^{\lambda_o/A} . \quad (II-17)$$

This system of equations has been constrained to yield a divergence free electric current. Therefore, the magnitude of this current [from (I-1)],

$$I_D = -W\mu Q_i \frac{dV_s}{dx} \left[1 - \frac{kT}{q} \frac{1}{Q_i} \frac{dQ_i}{dV_s} \right] , \quad (II-18)$$

can be calculated at any location along the source-drain channel. Selecting for this calculation the source end of this structure, (II-18) has the form

$$I_D = -W\mu Q_{io} \left(\frac{dV_s}{dx} \right)_{x=0} \left[1 - \frac{kT}{q} \frac{1}{Q_{io}} \left(\frac{dQ_i}{dV_s} \right)_{V_o} \right] , \quad (II-19)$$

where $V_o = V_s(0)$. Substituting (II-14) into (II-19) yields

$$I_D = W\mu Q_{io} E_o \left[1 - \frac{kT}{q} \frac{1}{Q_{io}} \left(\frac{dQ_i}{dV_s} \right)_{V_o} \right]. \quad (II-20)$$

From (II-17) we obtain the relation

$$\left(\frac{dQ_i}{dV_s} \right)_{V_o} = -\lambda_o Q_{io}. \quad (II-21)$$

Further, from (II-13) we have, upon recognizing that $V_s = V_D$ when $x=L$,

$$E_o = \frac{1}{(\lambda_o + A)L} \left\{ 1 - \left[1 - A(V_D - V_o) \right]^{\frac{\lambda_o + A}{A}} \right\}. \quad (II-22)$$

Thus, after substituting (II-21) and (II-22) into (II-20), we have an expression for the volt-ampere characteristics of a MOSFET

$$I_D = \frac{W\mu Q_{io}}{(\lambda_o + A)L} \left[1 + \frac{kT}{q} \lambda_o \right] \left\{ 1 - \left[1 - A(V_D - V_o) \right]^{\frac{\lambda_o + A}{A}} \right\}. \quad (II-23)$$

It should be noted that (II-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 (II-23) to the weak inversion mode of operation. From (II-12), as $\lambda_o \rightarrow q/kT$ in weak inversion the parameter A becomes small and, therefore, (II-23) has the limit

$$\lim_{\lambda_o \rightarrow q/kT} (I_D) \sim \frac{2WDQ_{io}}{L} \left[1 - e^{-q(V_D - V_o)/kT} \right]. \quad (II-24)$$

Clearly, (II-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 current saturation in weak inversion that is in substantial agreement with both experimental observation and other recent theoretical studies [8,9].

Further insight is gained from (II-24) if we consider the drain junction as a minority carrier (electron) sink-like the collector junction of a bipolar transistor. Assuming Boltzmann statistics, the term $\{1 - \exp[-q(V_D - V_O)/kT]\}$ in (II-24) is identical in form to the minority carrier sink offered by a reverse biased collector junction. In addition, when $V_D \gg V_O$, we have from (II-24)

$$I_D = \frac{2WDQ_{i0}}{L} . \quad (\text{II-25})$$

In this situation I_D might be assumed a consequence of minority carrier diffusion from a source of magnitude Q_{i0} to an ideal sink that is located a distance L from this source.

Thus, (II-23) contains all the qualitative requirements for the weak inversion mode of operation. First, this expression shows that electric current saturation exhibits an exponential form in weak inversion. Second, in weak inversion (II-23) predicts that all source-drain electric current arises from diffusion.

2.4 The Saturation Mode of Operation

Equation (II-23) and its weak inversion limit, (II-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

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

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

(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 . Thus, in electric current saturation (II-23) has the form

$$I_D = \frac{W\mu Q_{i0}}{(\lambda_O + A)L_C(V_D)} \left[1 + \frac{kT}{q} \lambda_O \right] \left\{ 1 - \left[1 - A(V_G - V_O) \right]^{\frac{\lambda_O + A}{A}} \right\}, \quad (\text{II-26})$$

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. From this theory, the drain junction space-charge layer extension into the substrate is given by

$$W_D^2 = \frac{2\kappa_s \epsilon_O}{qN_A} V_D, \quad (\text{II-27})$$

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

$$\Delta L = W_D - \left[W_D^2 - \frac{2\kappa_s \epsilon_O}{qN_A} (V_D - V_G) \right]^{1/2}. \quad (\text{II-28})$$

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

$$L_C(V_D) = L - \left\{ W_D - \left[W_D^2 - \frac{2\kappa_s \epsilon_O}{qN_A} (V_D - V_G) \right]^{1/2} \right\}. \quad (\text{II-29})$$

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, (II-26) in conjunction with (II-29) establishes the drain current when $V_D > V_G$.

2.5 Calculation of Q_{i0}

Before (II-7b) and, hence, (II-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_{i0} - the value of Q_i at the source end of this semiconductor structure (where $V_s = V_0$). From our rigorous computer calculations of MOSFET operation, we know that electrostatic interactions between the gate and drain are at a minimum near the source junction. For this reason, it is presumed that a one-dimensional calculation of inversion charge is applicable in this region.

From elementary MOSFET theory, this inversion charge has the form [1]

$$Q_i(V_s) = Q_T - Q_D = -C_{ox}(V_G - V_s) + \sqrt{2\kappa_s \epsilon_o q N_A V_s}, \quad (\text{II-30})$$

where Q_T and Q_D represent total electrostatic charge and depletion charge, respectively, and where

$$C_{ox} \equiv \frac{\kappa_i \epsilon_o}{t_{ox}} \quad (\text{II-31})$$

represents the static capacitance of a gate oxide of thickness t_{ox} . In a previous study [6], we showed that (II-30) neglects holes in establishing the depletion charge (Q_D) and, as a result, (II-30) is inapplicable for the weak inversion mode of transistor operation. Further, from this study [6], we have a revised expression for $Q_i(V_s)$ applicable near the source junction for any degree of inversion:

$$Q_i(V_s) = -C_{ox}(V_G - V_s) + \sqrt{2} \left(\frac{\kappa_s \epsilon_o}{L_D} \right) \left(\frac{kT}{q} \right) \left[e^{-\beta V_s} + \beta V_s - 1 \right]^{1/2}. \quad (\text{II-32})$$

In this expression, L_D is the extrinsic Debye length,

$$L_D = \left[\frac{\kappa_s \epsilon_o kT}{q^2 N_A} \right]^{1/2}. \quad (\text{II-33})$$

It should be noted that (II-32) yields the magnitude of $Q_{i0} = Q_i(V_0)$ explicitly in terms of V_0 and implicitly in terms of V_G . From a one-dimensional solution of Poisson's equation at the source end of this structure [6], we have (II-34)

$$V_G = V_0 + \sqrt{2} \left(\frac{\kappa_s}{\kappa_i} \right) \left(\frac{t_{ox}}{L_D} \right) \left(\frac{kT}{q} \right) \left\{ e^{-\beta V_0} + \beta V_0 - 1 + \left(\frac{n_i}{N_A} \right)^2 \left[e^{\beta V_0} - 1 \right] \right\}^{1/2}.$$

Clearly, this problem represents only a minor mathematical complication.

2.6 Calculation of λ_0

In the present MOSFET theory, source-drain electric current is given by (II-23). This equation contains two unknown parameters (Q_{i0} and λ_0) which must be evaluated in order to calculate this current. In §2.5, we gave an expression, (II-32), which can be used to evaluate Q_{i0} . Herein we will derive an expression for λ_0 , the remaining parameter needed for calculating the source-drain electric current.

To obtain an expression for λ_0 , we will utilize a general relation for $\lambda(V_s)$ given by (II-5b). In addition, since $\lambda_0 \equiv \lambda(V_0)$ represents the magnitude of $\lambda(V_s)$ at the source end of this semiconductor structure, we will use (II-32) to approximate $Q_i(V_s)$ in this vicinity. A substantial degree of mathematical simplification results if we neglect the term $e^{-\beta V_s}$ in this equation; thereby

$$Q_i(V_s) \approx -C_{ox}(V_G - V_s) + \sqrt{2} \left(\frac{\kappa_s \epsilon_0}{L_D} \right) \left(\frac{kT}{q} \right) \left[\beta V_s - 1 \right]^{1/2}. \quad (II-35)$$

We can neglect this term when

$$V_s \gg (kT/q) [1 - \exp(-qV_s/kT)], \quad (II-36)$$

which is clearly the situation in all cases of practical interest.

From (II-35), by differentiating with respect to V_s , we have

$$\frac{dQ_i}{dV_s} = C_{ox} + \frac{1}{\sqrt{2}} \left(\frac{\kappa_s \epsilon_o}{L_D} \right) \left(\frac{kT}{q} \right)^{1/2} \left[V_s - \frac{kT}{q} \right]^{-1/2} \quad (II-37)$$

$$\frac{d^2 Q_i}{dV_s^2} = - \frac{1}{2\sqrt{2}} \left(\frac{\kappa_s \epsilon_o}{L_D} \right) \left(\frac{kT}{q} \right)^{1/2} \left[V_s - \frac{kT}{q} \right]^{-3/2} \quad (II-38)$$

Substituting (II-35), (II-37), and (II-38) into (II-5b) and, thereafter, evaluating the result for $V_s = V_o$, we obtain

$$\lambda_o = \left\{ \frac{C_{ox} + \frac{1}{2} \left[2\kappa_s \epsilon_o q N_A / (V_o - \frac{kT}{q}) \right]^{1/2} \left[1 + \frac{1}{2} \left(\frac{kT}{q} \right) (V_o - \frac{kT}{q})^{-1} \right]}{C_{ox} \left[V_G - V_o + \frac{kT}{q} \right] - \left[V_o - \frac{3}{2} \left(\frac{kT}{q} \right) \right] \left[2\kappa_s \epsilon_o q N_A / (V_o - \frac{kT}{q}) \right]^{1/2}} \right\} \quad (II-39)$$

For $V_o \gg \frac{3}{2} (kT/q)$, (II-39) has the approximate form

$$\lambda_o = \frac{C_{ox} + \frac{1}{2} \left[2\kappa_s \epsilon_o q N_A / (V_o - \frac{kT}{q}) \right]^{1/2}}{C_{ox} \left[V_G - V_o + \frac{kT}{q} \right] - \left[V_o - \frac{3}{2} \left(\frac{kT}{q} \right) \right] \left[2\kappa_s \epsilon_o q N_A / (V_o - \frac{kT}{q}) \right]^{1/2}}, \quad (II-40)$$

where V_o is given implicitly by (II-34).

For the weak inversion mode, (II-34) has the approximate form

$$V_G \approx V_o + \frac{\sqrt{2\kappa_s \epsilon_o q N_A (V_o - \frac{kT}{q})}}{C_{ox}} \quad (II-41)$$

Upon substituting (II-41) into (II-40) and, thereafter, simplifying the result through straightforward algebraic manipulations, we obtain $\lambda_o = q/kT$. Thus, our approximate expression for λ_o yields the correct magnitude for this separation parameter in the limit of weak inversion. In §3.1, we demonstrate quantitatively that Eq. (II-40) provides a satisfactory approximation of λ_o for any degree of inversion.

3.0 Model Evaluation and Refinements

The original version of our model for MOSFET operation contained inadequate approximations for Q_{i0} and λ_0 which rendered this model inapplicable to the weak inversion mode. In a previous study [6], we alleviated this difficulty by obtaining improved approximations and, thereby, we revised our model so that it would apply for any degree of inversion. Therefore, at the start of the present research program, we initiated an extensive evaluation of our revised model for a variety of device structures and a wide range of applied bias voltages.

We compared calculated volt-ampere characteristics from our model with rigorous two-dimensional computer calculations of these characteristics. In this comparison, we obtained satisfactory agreement between our model and these computer calculations for a long channel structure ($L = 10\mu\text{m}$) with lightly doped substrate ($N_A = 2 \times 10^{15} \text{cm}^{-3}$) and moderately thin oxide ($t_{\text{ox}} = 1000 \text{\AA}$). In contrast, we observed significant discrepancies between our model and these computer calculations for a structure with a more highly doped substrate ($N_A = 2 \times 10^{15} \text{cm}^{-3}$) and thicker oxide ($t_{\text{ox}} = 2000 \text{\AA}$). For this structure, our model over predicted the magnitude of source-drain current by approximately 50%, in the saturation region of these volt-ampere characteristics. Further, this model also over predicted the magnitude of applied drain voltage required to produce electric current saturation in this structure. During the present research we determined an important source of this error and, thereafter, revised our model to alleviate this difficulty. In §3.1, we discuss the details of this revision.

In other calculations for weak inversion operation, we also observed disagreement between our model and our rigorous computer solution for MOSFET operation. From these calculations, our model under predicted the source-drain current in this weak inversion mode. A detailed study of this problem revealed that our model contained an inadequate approximation for the distribution of inversion charge, $Q_i(V_s)$, within the source-drain channel.

Further, we determined that the foregoing discrepancy in our volt-ampere expression resulted from an erroneous approximation for the slope of this inversion charge distribution, dQ_i/dV_s , at the source end of this structure. This aspect of our model has been revised and we now obtain substantially improved agreement between our model and two-dimensional computer calculations of these volt-ampere characteristics. Details of this revision are discussed in §3.2.

3.1 Electric Current Saturation

For a long channel MOSFET ($L = 10\mu\text{m}$) with $2 \times 10^{16} \text{ cm}^{-3}$ substrate doping and an oxide thickness of $2000 \text{ \AA}$, we observed a substantial discrepancy between our model and rigorous computer calculations for the strong inversion mode. Briefly, our model over predicted both the drain voltage required to produce electric current saturation and the magnitude of source-drain current in this saturation mode. From (II-20), this discrepancy could arise from our calculation of Q_{i0} , E_0 , or $(dQ_i/dV_s)_{V_0}$. By checking these calculations against our computer solution, we determined that the foregoing discrepancy was produced by an error in our calculation of E_0 .

From (II-22), E_0 is exceedingly dependent upon the parameters λ_0 and A . Therefore, it was assumed the error in our calculation of E_0 could be attributed to either an error in A or λ_0 . Thus, we first undertook a detailed assessment of our approximate expression for λ_0 .

In our theory for MOSFET operation, λ_0 represents the magnitude of our separation parameter, $\lambda(V_s)$, at the source end of this semiconductor structure. Thus, from (II-5b), λ_0 has the qualitative form

$$\lambda_0 = \left\{ \frac{\frac{kT}{q} \left(\frac{d^2 Q_i}{dV_s^2} \right)_{V_0} - \left(\frac{dQ_i}{dV_s} \right)_{V_0}}{Q_{i0} - \frac{kT}{q} \left(\frac{dQ_i}{dV_s} \right)_{V_0}} \right\}, \quad (\text{II-42})$$

where $Q_{i0} = Q_i(V_0)$. In §2.6, we derived a quantitative relation for λ_0 based upon our revised one-dimensional approximations for $Q_i(V_s)$ and its first two derivatives [(II-35) through (II-37), respectively]. To assess the adequacy of using these approximations to calculate λ_0 , we made comparisons between (II-35)-(II-37) and a rigorous nonequilibrium solution of Poisson's equation in one spatial dimension (see Appendix A).

These comparisons are illustrated in Figs. II-2 and II-3, for strong inversion operation, and in Figs. II-4 and II-5, for weak inversion operation. From Figs. II-2 through II-5 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. II-3 and II-5, we have a qualitative illustration that good agreement is attained between our approximation for d^2Q_i/dV_s^2 and this rigorous solution. Thus, from the foregoing discussion, we would not expect our approximate expression for λ_0 [(II-40)] to produce a significant error in our volt-ampere calculation.

We next sought to check the adequacy of our expression for the parameter A. From §2.2, the separation parameter, $\lambda(V_s)$, always attains a magnitude of q/kT in the limit of weak inversion operation. In the Kennedy 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) near this point of channel termination. If, instead, this transition takes place at some other location between the source and drain (where $V_s \neq V_G$), our expression for the parameter A [(II-12)] would be in error.

For structures with lightly doped substrates ($N_A = 2 \times 10^{15} \text{ cm}^{-3}$) and relatively thin oxides ($t_{ox} = 1000 \text{ Å}$), 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 $N_A > 10^{16} \text{ cm}^{-3}$ and

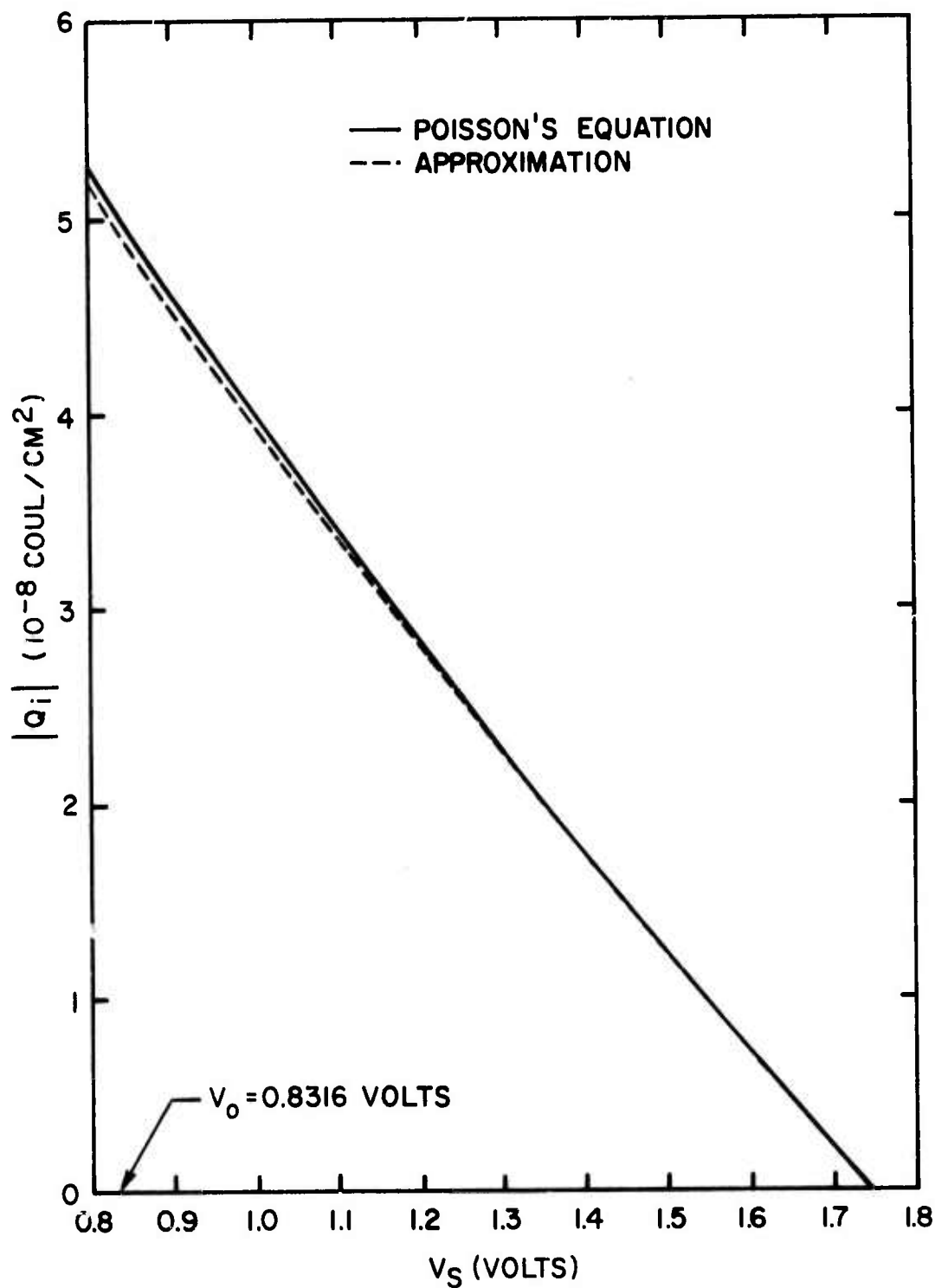


Figure II-2. Calculated inversion charge distribution in a MOSFET for strong inversion operation.

$N_A = 2 \times 10^{16} \text{ cm}^{-3}$, $t_{\text{ox}} = 2000 \text{ \AA}$, $V_G = 8.0$ volts.

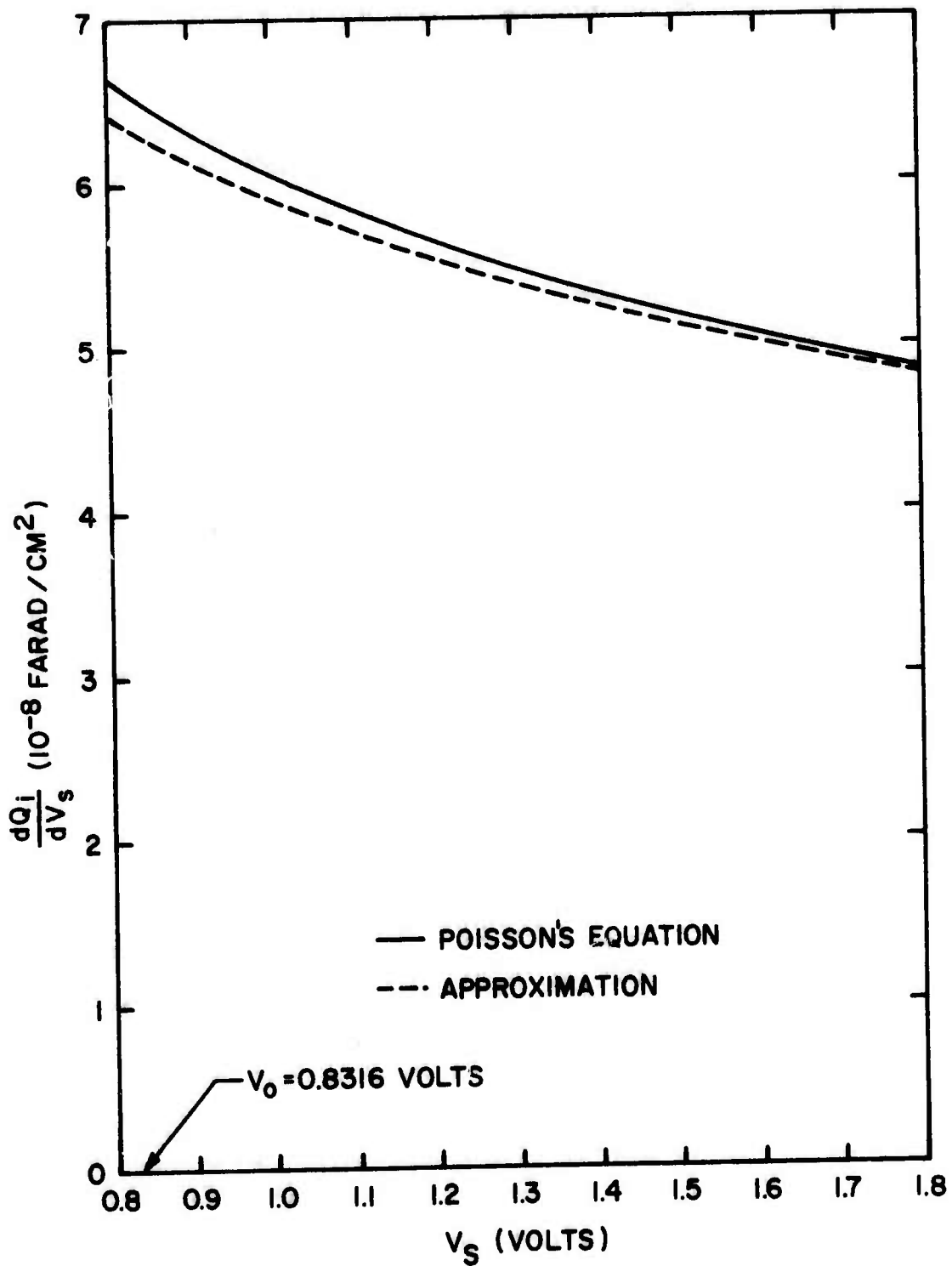


Figure II-3. Calculated dQ_i/dV_s in a MOSFET for strong inversion operation.

$N_A = 2 \times 10^{16} \text{ cm}^{-3}$, $t_{ox} = 2000 \text{ \AA}$, $V_G = 8.0$ volts.

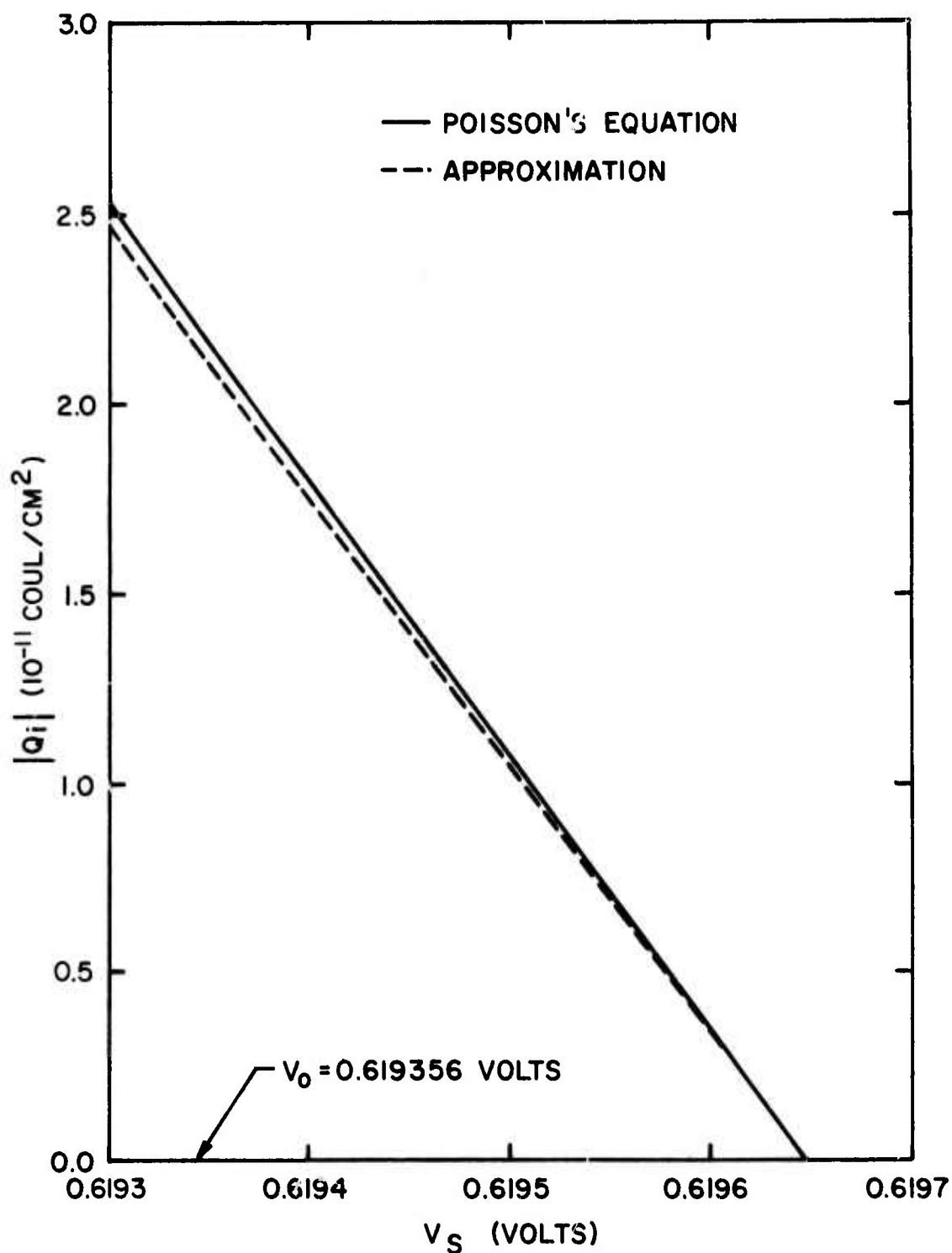


Figure II-4. Calculated inversion charge distribution in a MOSFET for weak inversion operation.

$N_A = 2 \times 10^{16} \text{ cm}^{-3}$, $t_{\text{ox}} = 2000 \text{ \AA}$, $V_G = 4.3$ volts.

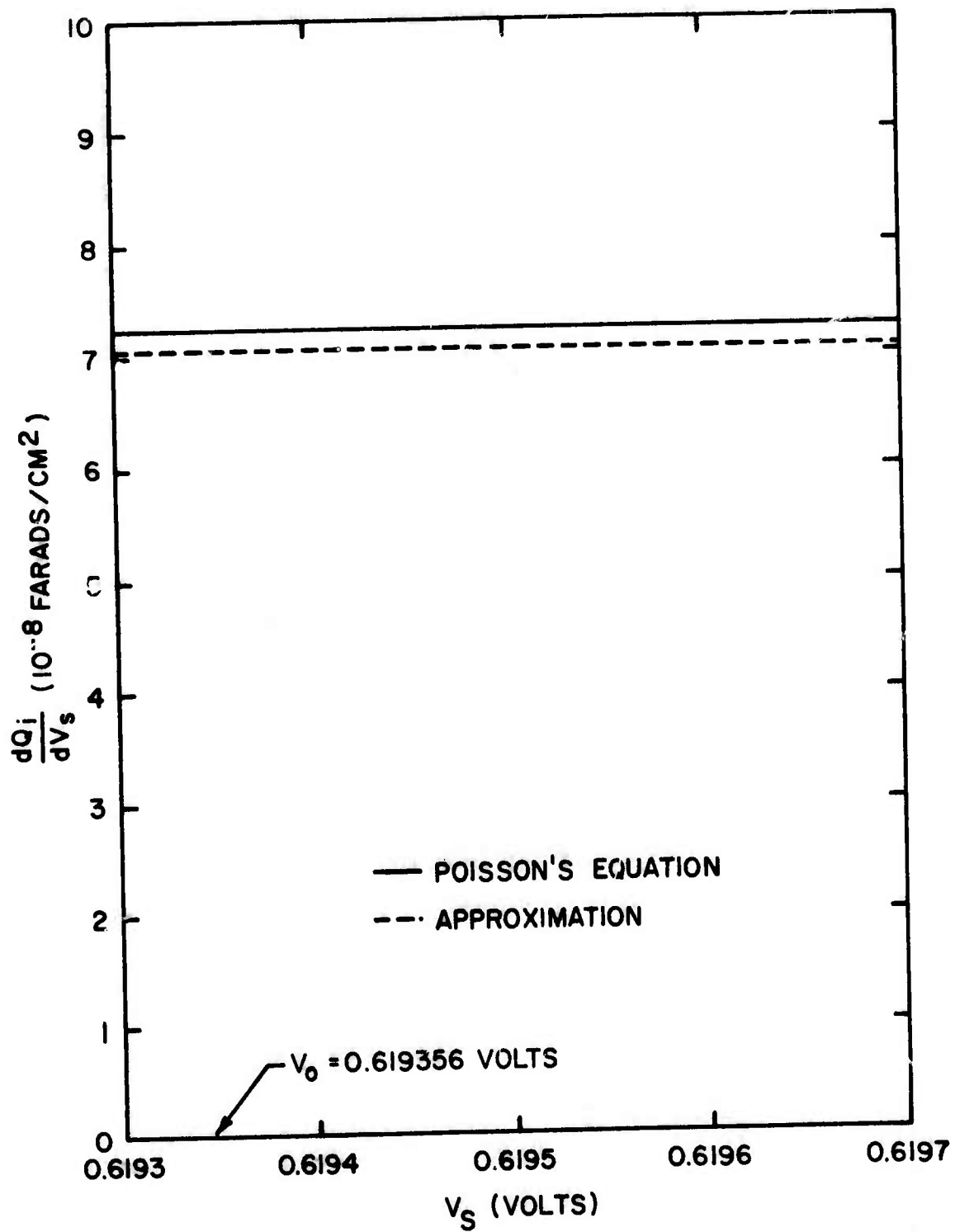


Figure II-5. Calculated dQ_i/dV_s in a MOSFET for weak inversion operation.
 $N_A = 2 \times 10^{16} \text{ cm}^{-3}$, $t_{ox} = 2000 \text{ \AA}$, $V_G = 4.3 \text{ volts}$.

$t_{ox} = 2000 \text{ \AA}$ 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 . Thus, we have an error in our expression for the parameter A (II-12).

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

$$V_{sc} = V_G - \frac{\kappa_s \epsilon_o q N_A}{C_{ox}^2} \left\{ \sqrt{1 + \frac{2C_{ox}^2}{\kappa_s \epsilon_o q N_A} \left[V_G - \frac{kT}{q} \left(1 - \frac{n_i}{N_A} \right) \right]} - 1 \right\} \quad (\text{II-43})$$

As a first order approximation, we presume $\lambda(V_s) \rightarrow q/kT$ when $V_s \rightarrow V_{sc}$. Therefore, from (II-11)

$$A = \frac{1 - (kT/q)\lambda_o}{V_{sc} - V_o} \quad (\text{II-44})$$

It should be noted that this modification of our model (II-43) does not alter the qualitative form of our expression for E_o (II-22) nor that of our volt-ampere relation (II-23). However, for the saturation region of these volt-ampere characteristics, we replace (II-26)-(II-29) with the following equations:

$$I_D = \frac{W\mu Q_{io}}{(\lambda_o + A)L_c(V_D)} \left[1 + \frac{kT}{q} \lambda_o \right] \left\{ 1 - \left[1 - A(V_{sc} - V_o) \right]^{\frac{\lambda_o + A}{A}} \right\} \quad (\text{II-45})$$

$$W_D^2 = \frac{2\kappa_s \epsilon_o}{qN_A} V_D \quad (\text{II-46})$$

$$\Delta L = W_D - \left[W_D^2 - \frac{2\kappa_s \epsilon_o}{qN_A} (V_D - V_{sc}) \right]^{1/2} \quad (\text{II-47})$$

$$L_c(V_D) = L - \left\{ W_D - \left[W_D^2 - \frac{2\kappa_s \epsilon_0}{q N_A} (V_D - V_{sc}) \right]^{1/2} \right\}, \quad (\text{II-48})$$

where V_{sc} and A are given by (II-43) and (II-44), respectively.

With these revisions in our model [(II-44)-(II-48)], we obtained improved agreement with our computer solution, for the device under consideration. Specifically, the discrepancy in our volt-ampere calculation was reduced from 50% to approximately 5%.

3.2 Weak Inversion Operation

In weak-inversion mode calculations, we found an additional discrepancy in our model for MOSFET operation: our approximate expression for $Q_i(V_s)$, (II-17) was found in error, when compared with rigorous computer calculations of this inversion charge distribution. Moreover, because we used this incorrect expression to calculate the term $(dQ_i/dV_s)V_o$ in our volt-ampere relation [see §2.3], we also have a discrepancy in our calculation of the source-drain current in weak inversion.

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 one-dimensional approximation, (II-37). Substituting (II-22) into (II-20), our volt-ampere relation now has the form

$$I_D = \frac{W \mu Q_{io}}{(\lambda_o + A)L} \left[1 - \frac{kT}{q} \frac{1}{Q_{io}} \left(\frac{dQ_i}{dV_s} \right)_{V_o} \right] \left\{ 1 - \left[1 - A(V_D - V_o) \right]^{\frac{\lambda_o + A}{A}} \right\}, \quad (\text{II-49})$$

where $(dQ_i/dV_s)_{V_o}$ is given by (II-37).

Using (II-49) we again calculated these volt-ampere characteristics for weak inversion operation and compared these characteristics with those calculated from our two-dimensional computer program. We obtained substantially improved agreement between these two calculations as a result of the foregoing revision.

4.0 Conclusions

An evaluation of our previous theory for MOSFET operation revealed discrepancies in both the strong and weak inversion modes of operation. We have identified the factors producing these errors and have initiated appropriate revisions in our model. During the next quarter, we plan to complete these revisions and resume our evaluation of this model. In addition, we plan to make a comparative study between this model and elementary MOSFET theory. From this comparison, we aim to demonstrate the superiority of our model for application to the design and development of MOSFET structures.

5.0 List of Symbols

C_{ox}	capacitance of gate oxide
D	electron diffusivity in a MOSFET inversion layer
E_o	magnitude of source-drain electric field at source end of a MOSFET channel
E_y	gate-induced electric field
E_{ys}	E_y at oxide-semiconductor interface
I_D	source-drain electric current in a MOSFET
I_{diff}	diffusion component of I_D
kT/q	thermal voltage (.0259 volts at $T = 300^\circ K$)
L	source-drain distance in a MOSFET
L_c	voltage-dependent channel length in a MOSFET
L_D	extrinsic Debye length in semiconductor substrate of a MOSFET
n	electron density in semiconductor material
n_i	intrinsic carrier concentration in semiconductor material
n_s	magnitude of n along oxide-semiconductor interface
n_q	nonequilibrium electron density in charge neutral semiconductor substrate.
N_A	acceptor impurity ion density in semiconductor substrate
p	hole density in semiconductor material
p_q	nonequilibrium hole density in charge neutral semiconductor substrate

q	electronic charge
Q_i	inversion charge in semiconductor substrate
Q_{io}	magnitude of Q_i at source end of a MOSFET channel
t_{ox}	gate oxide thickness
V	electrostatic potential (referenced to charge-neutral regions of semiconductor substrate)
V_D	applied drain-source biasing voltage
V_G	applied gate-source biasing voltage
V_i	magnitude of V at which $n = n_i$
V_o	magnitude of V_s at source end of a MOSFET channel
V_s	surface potential in a MOSFET (electrostatic potential at oxide-semiconductor interface)
W	width of a MOSFET
W_D	drain depletion layer width
x	distance from source towards drain, parallel to oxide-semiconductor interface
y	distance from semiconductor surface into substrate, perpendicular to oxide-semiconductor interface
y_i	magnitude of y at which $n = n_i$
β	$(kT/q)^{-1}$
ϵ_o	permittivity of free space
κ_i	relative dielectric constant of gate oxide
κ_s	relative dielectric constant of semiconductor substrate
λ	Separation parameter in our model for MOSFET operation
λ_o	magnitude of λ at source-end of a MOSFET channel
μ	electron drift mobility in a MOSFET inversion layer
ρ	electrostatic charge density within semiconductor substrate
ϕ_n	electron quasi-Fermi potential in semiconductor substrate
ϕ_p	hole quasi-Fermi potential in semiconductor substrate

Chapter III
Equivalent Circuit Studies

F. A. Lindholm

	<u>Page</u>
General.....	75
1.0 Introduction.....	76
2.0 Self Consistency.....	77
2.1 Self Consistency from Extended Ideal.....	77
Charge-Control Model for Intrinsic	
Device	
2.2 Model for the Extrinsic Device.....	90
2.3 Self-Consistency Test.....	93
3.0 Commentary.....	94
3.1 Model Updating.....	95

Chapter III

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 (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 network 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 now being refined and extended, and is meant for publication in the final report.

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 detailed reporting of this work now follows.

A SELF-CONSISTENCY TEST FOR DEVICE MODELS IN
TRANSIENT COMPUTER SIMULATION OF LARGE-SIGNAL
CIRCUITS

(F. A. Lindholm)

1.0 Introduction

From a fundamental standpoint, the analysis of such semiconductor devices as MOS and bipolar transistors is based on a set of differential equations [1] which express relationships among currents, mobile-carrier concentrations, and potential. In the full generality needed to describe large-signal, dynamic response, analytic solution of these equations, without approximations, has proven to be intractable. Even by computer, solutions subject to the appropriate boundary conditions are done only with great difficulty. Hence, to simplify treatment of these basic equations, various methods of approximation have evolved.

Among these is the *quasi-static approximation*. In this method, one first omits all time-dependent terms appearing in the basic equations. Analytic solution then becomes possible if various additional approximations are employed. This solution yields expressions relating the currents and charge components within the device to the boundary conditions, for example, to the terminal voltages. In a strict sense these expressions apply only if the boundary conditions are time-invariant. To obtain an approximate solution if the boundary conditions vary with time, one makes the quasi-static approximation that the expressions hold in general, despite departures from the dc steady state.

[1] W. Shockley, *Bell System Tech. J.*, 28, 435 (1949).

Because of their origin in conventional device analysis, all device models commonly used in the computer-aided analysis of bipolar and MOS circuits depend for their validity on the quasi-static approximation. Based on this premise, we develop here a test for the *self-consistent validity* of device models that is easily implemented in circuit-analysis programs.

The self-consistency test to be developed applies both to MOS and bipolar transistors, and, indeed, to other devices as well. Use of the concept of the ideal charge-control model, extended appropriately, enables a development in this generality, and with this subject the rest of the paper begins. Using the extended ideal charge-control model, we then develop the test for self-consistency and unify alternative views of its meaning. Emphasis is given to the extrinsic as well as to the intrinsic device, and to various nonidealities that distinguish the behavior of actual devices from that of the ideal charge-control model.

Application of the test developed here to a practical MOS NAND gate demonstrates marked violations of self-consistency over appreciable intervals of time [2]. These violations motivate the comments with which we conclude the paper, about various methods for improving models to remedy self-inconsistency.

2.0 Self Consistency

In conventional device analysis, one of the additional approximations referred to earlier involves dividing the device into *extrinsic* and *intrinsic* parts, as Fig. III-1 illustrates. The self-consistency test to be developed will apply to both of these parts. We begin with intrinsic device.

2.1 Self Consistency from Extended Ideal Charge-Control Model for Intrinsic Device

The charge-control viewpoint was first applied to restricted types of devices: to photoconductive devices [3] and to junction

[2] D. L. Fraser, Jr., and F. A. Lindholm, "Violations of the Quasi-Static Approximation in Large-Signal MOSFET Models," to be published.

[3] A. Rose, "La Photoconductivité," *L'Onde Electrique*, 34, 645-651 (1954).

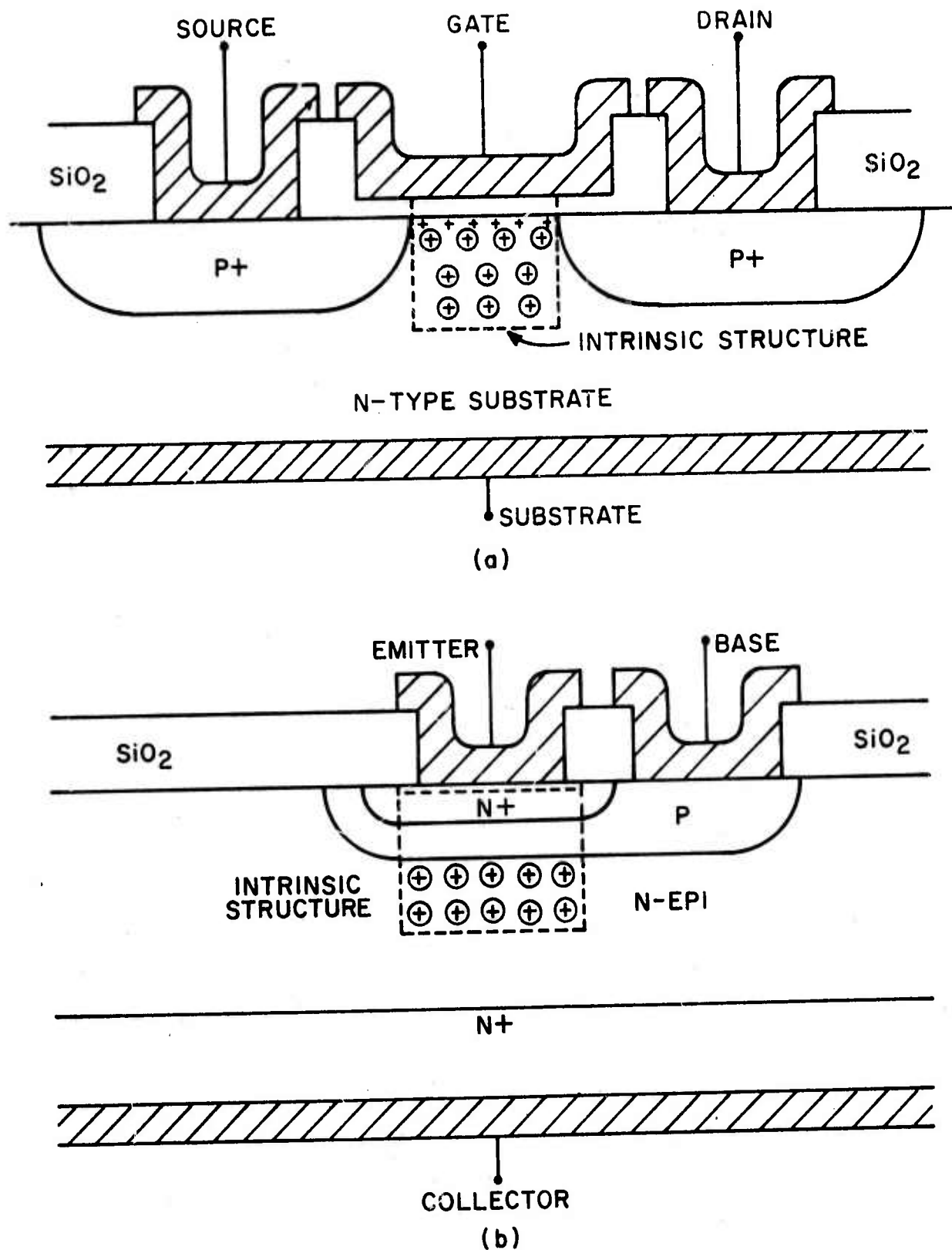


Fig. III-1. Illustrating the division into intrinsic and extrinsic devices for: (a) MOSFET, and (b) bipolar transistor. The dotted boxes enclose the intrinsic structures. (Not to scale.)

transistors [4]. In 1959, however, Johnson and Rose [5] and Middlebrook [6] introduced charge control as a means for unifying the characterization of many different kinds of electronic devices. It is this unifying attribute that we exploit here, together with the property that charge control is rooted in the quasi-static approximation.

Earlier work will guide our treatment. Previous presentations and applications of charge control do not satisfy our needs for assessing self-consistency, however, partly because most earlier workers concentrated on small-signal, amplifying applications, partly because they restricted consideration only to the intrinsic device, and partly for other reasons that will become plain as our treatment progresses.

Following others [5,6], we begin by considering the general three-terminal, electronic device pictured in Figure III-2. This is a charge-control device, which means that the basis of its operation involves the current i_{21} flowing between terminals 1 and 2 being controlled by the charge Q_3 placed at terminal 3. For this device, we isolate the essence of operation through introduction of an *ideal charge-control model* defined by the following properties:

- (a) The current i_{21} flowing between terminal 1 (source or emitter, for example) and terminal 2 (drain or collector, for example) is the ratio of a controlled charge Q_{21} of mobile carriers within the device to the mean transit time t_{21} needed for these carriers to move between these two terminals. Hence,

$$i_{21} = Q_{21}/t_{21} \quad . \quad \text{(III-1)}$$

-
- [4] J. J. Sparkes and R. Beaufoy, *Proc. IRE*, 45, 1740 (1957); *ATE J.*, 13, 310 (1957).
 - [5] E. O. Johnson and A. Rose, *Proc. IRE*, 47, 407 (1959).
 - [6] R. D. Middlebrook, *Proc. IRE*, 106, Part B, Suppl. No. 17, 887 (1959).

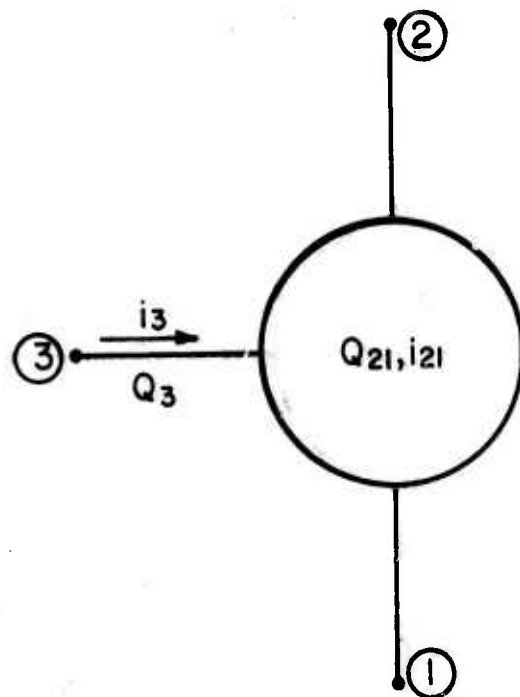


Fig. III-2.-General charge-control device.

- (b) The controlling charge Q_3 associated with terminal 3 (gate or base, for example) exercises total control over the controlled charge according to

$$dQ_{21} = -dQ_3 . \quad (\text{III-2})$$

- (c) The current i_3 flows only to change the charge Q_3 placed there. Thus,

$$i_3 = \dot{Q}_3 , \quad (\text{III-3})$$

where the dot notation designates time differentiation.

In all devices, property (a) is an idealization, valid in a strict sense only in the dc steady state, but extrapolated to apply at each instant during transients by use of the *quasi-static approximation*. Property (b) is an idealization for most devices in that Q_3 only partially controls Q_{21} . For example, in an MOS transistor, some of the flux lines starting from gate charge may end on depletion charge rather than on the charge of mobile carriers in the channel. As another example, in a short-channel MOS transistor, the two-dimensional (or three-dimensional) configuration of flux lines in the insulator and in the semiconductor may invalidate the full control expressed in (III-2). Similar violations of (III-2) occur in bipolar transistors and other devices. Property (c) is an excellent approximation for some devices; an example is the MOS transistor. For the bipolar transistor, however, it is an idealization because the base fails to retain all of the charge Q_3 placed at its terminal. Net recombination, for instance, occurring within the device or at its surfaces or contacts, requires a component of i_3 in addition to that expressed in property (c).

For the purposes of this paper, however, these nonidealities are inconsequential. As we shall see for the bipolar and the MOS transistor, formulation of a test for self-consistency of the quasi-static approximation can be based on straightforward extensions of the ideal charge control model defined above.

To specify the network representation of the ideal charge control model, consider first the simple case in which terminal 2 acts as the sole collector of the transport current i_{21} . This

corresponds to a first-order approximation of forward-active operation for the bipolar transistor and of pinch-off operation for the MOSFET. For this case, the properties defined by (III-1) through (III-3) imply the network representation of Fig. III-3(a), in which chain-rule differentiation applied to (III-3) gives for the capacitance:

$$C_{31} = \left. \frac{\partial Q_3}{\partial V_{31}} \right|_{V_{32}} = \left. \frac{\partial Q_{21}}{\partial V_{31}} \right|_{V_{32}} \quad (\text{III-4})$$

Fig. III-3(a) represents the large-signal transient response, in contrast to the small-signal representations of Johnson and Rose [5] and Middlebrook [6].

Some circuit applications reverse the roles of terminals 1 and 2, with terminal 1 becoming the sole collector of the current i_{21} , as in the inverse-active operation of the bipolar transistor. This corresponds to the network representation of Fig. III-3(b), with

$$C_{32} = \left. \frac{\partial Q_3}{\partial V_{32}} \right|_{V_{31}} = \left. \frac{\partial Q_{21}}{\partial V_{32}} \right|_{V_{31}} \quad (\text{III-5})$$

In the most general case, terminals 1 and 2 both collect and emit the transport current i_{21} , as in the saturated mode of the bipolar transistor and the non-pinch-off mode of the MOSFET. Thus the general network representation of the ideal charge-control model, extended to apply for large-signal applications and to account for the dual roles of terminals 1 and 2, derives from a parallel combination of Figs. III-3(a) and III-3(b). The general network representation appears in Fig. III-3(c).*

* Rather than the plausibility argument given, one can derive the network representation of Figure III-3(c) by more strict reasoning. This representation follows from chain-rule differentiation of (III-3) joined with the implication of (III-2) that all flux lines of Coulomb force associated with the control of Q_{21} by Q_3 must contact terminal 3 (none link terminals 1 and 2).

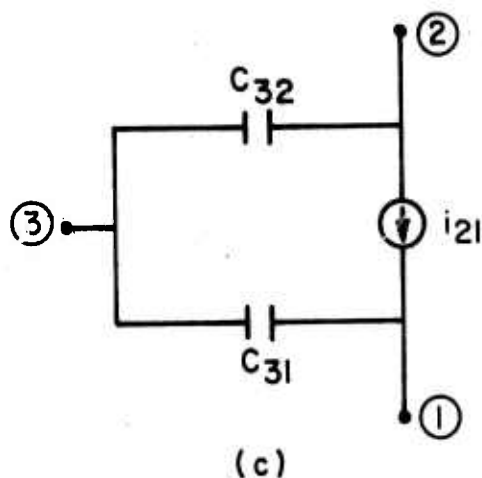
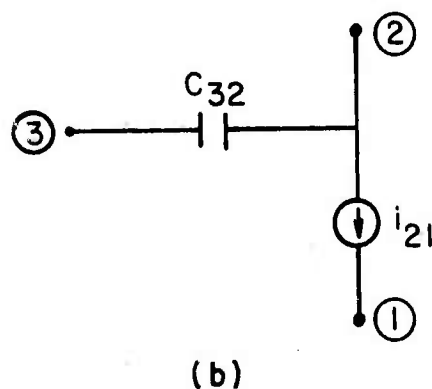
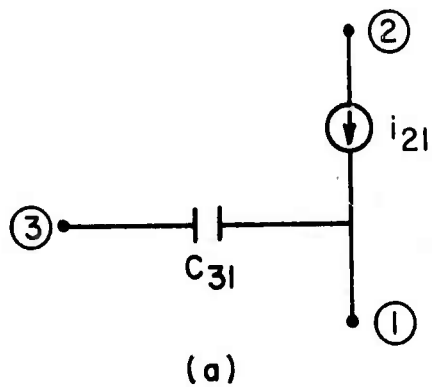


Fig. III-3. Illustrating the development of the general network representation, shown in (c), of the extended ideal charge-control model. It approximates the behavior of the *intrinsic* part of MOSFETs, bipolar transistors, and other charge-control devices.

In this network, the circuit elements all are functions of the device make-up and the terminal voltages and currents. The particular functional dependencies that apply result from the details of the physical modeling used to characterize the charge Q_{21} and the transit time t_{21} . For example, the Ebers-Moll model [7] implies

$$Q_{21} \approx \frac{qn_i^2 W_B}{2N_B} \exp(qV_{BE}/kT) \text{ and } t_{21} \approx W_B^2/2D_B \quad (\text{III-6})$$

for the bipolar transistor in the forward-active mode; and the simple square-law characterization [8,9] for light substrate doping implies

$$Q_{21} \approx (2/3)C_O ZLV_{GS} \text{ and } t_{21} \approx (4/3)L^2/\mu V_{GS} \quad (\text{III-7})$$

for the MOSFET in pinch-off operation. Here q denotes the electron charge; n_i , the intrinsic carrier density; W_B , the quasi-neutral base thickness; D_B , the diffusivity of minority carriers in the base; kT/q , the thermal voltage; V_{BE} , the base-emitter voltage; C_O , the gate oxide capacitance per unit area; μ , the mobility in the channel; V_{GS} , the gate-source voltage; L , the channel length, and Z its width.

Self-Consistency Criteria: As was noted, the ideal charge-control model of Fig. III-3(c) is based on the quasi-static approximation. Because this approximation is exact in the dc steady state, we take the following as the criterion for self-consistency: *that the configuration of current flow in the network representation during transients must correspond to only a small perturbation of the configuration existing in the dc steady state.* Hence, from Fig. III-3(c), the criterion is:

-
- [7] J. J. Ebers and J. L. Moll, *Proc. IRE*, 42, 1761 (1954); J. L. Moll, *Proc. IRE*, 42, 1761 (1954).
 - [8] C. T. Sah, *IEEE Trans. Electron Devices*, ED-11, 324 (1964).
 - [9] F. A. Lindholm, *IEEE J. Solid-State Circuits*, SC-6, 250 (1971); and SC-7, 322 (1972).

$$\frac{|C_{31}V_{31}| + |C_{32}V_{32}|}{|i_{21}|} \ll 1, \quad (\text{III-8})$$

or, in more general notation, the merit of which will be seen later,

$$\frac{\sum_{k \neq j} |C_{jk}V_{jk}|}{Q_{21}/t_{21}} \ll 1, \quad (\text{III-9})$$

where $j=3$ for the MOSFET of Fig. III-1(a). Expressions (III-8) and (III-9) constitute necessary and sufficient conditions for self-consistency. As we shall see, they form the basis for a self-consistency test that can be implemented in computer programs for circuit analysis.

Viewed from another standpoint, we see that a consistent but less demanding necessary condition is

$$\frac{|Q_3|}{|Q_{21}/t_{21}|} = \frac{|Q_{21}|}{|Q_{21}/t_{21}|} \ll 1. \quad (\text{III-10})$$

Let the applied excitation make Q_{21} vary in proportion to $e^{i\omega t}$. Then (III-10) becomes

$$\omega t_{21} \ll 1, \quad (\text{III-11})$$

which implies self-consistency with the quasi-static approximation provided the period of the applied signal much exceeds the transit time of the mobile carriers. If this holds, the mobile carriers can spatially distribute themselves fast enough to assume essentially the same relation to the boundary conditions as exists in the dc steady state.

The self-consistency criteria just developed have followed from the assertion that the quasi-static approximation is exact in the dc steady state. Despite the intuitive appeal of this assertion, and although it has been advanced by others, the recent observations of Cherry [10], and the earlier ones of Macdonald [11], about capacitance indicate that it is not strictly

[10] E. M. Cherry, *IEEE Trans. Electron Devices*, ED-18, 1166 (1971).

[11] J. R. Macdonald, *Solid-State Elec.*, 5, 11 (1962).

true. From a practical standpoint, no change in our conclusions results, however, from regarding the quasi-static approximation as exact in the dc steady state. The device models we consider consist of controlled sources and resistors, which represent the flow of conduction current, and other circuit elements [12], usually capacitors, which represent capacitive effects. That the magnitude of the capacitors calculated by quasi-static analysis may be in error, even for arbitrarily low frequency of excitation, does not affect our conclusions, because at such frequencies the currents flowing through these capacitors are most often negligible relative to the conduction currents. That the magnitude of these quasi-static (or charge-control) capacitors may be in error at higher frequencies contained in the excitation likewise does not affect our conclusions. Rather it influences only the strength of the inequalities appearing in such expressions as (III-9).

Effect of Nonidealities on Self-Consistency Criteria: The criteria expressed in (III-8) and (III-9) are based on the ideal charge-control model, which provides only a first-order approximation to the performance of actual devices. Nonetheless, these criteria stay valid in the presence of commonly occurring nonidealities, as we shall now show.

There are several kinds of such nonidealities. Consider first the kind that arises from net recombination or generation occurring within the intrinsic device. The drain current of the MOS transistor, for example, can increase its rate of growth with increasing drain voltage because of avalanche multiplication of

-
- [12] F. A. Lindholm and J. I. Arreola, "Network Representation of the Large-Signal Transient Response of MOS Transistors," *Late News Supplement of 1975 Int. Electron Devices Meeting*, paper 11.8, p. 6 (1975); and "Network Representation for Self-Consistent Transient Response of Electronic (and other) Devices, Including Multi-Terminal Effects," *7th Annual Pittsburgh Conference on Modeling and Simulation*, (April 1976).

mobile carriers--a form of net generation. One must account for this nonideality in the modeling of the mobile charge, Q_{21} , and of the (effective) transit time, t_{21} ; but the network representation remains that of the ideal charge-control model shown in Fig. III-3(c). Hence the self-consistency criteria of (III-8) and (III-9) remain valid for a three-terminal model of the MOSFET despite the occurrence of avalanche multiplication. The same conclusions hold for net recombination or generation in a MOSFET that arises from other processes, such as thermal emission or capture of carriers, provided the resultant current flowing in the dc steady state to the substrate is negligible.

In contrast, net generation or recombination in a bipolar transistor requires the addition of controlled current sources in parallel with each capacitor in the network representation of Fig. III-3(c). The time constants normally associated with these sources will differ markedly from the transit time for mobile carriers. From a practical viewpoint, however, except for operation near avalanche breakdown, the additional sources will carry small current relative to that of source i_{21} , and the self-consistency criteria for the quasi-static approximation remain as we have stated them.

In a previous section, we discussed a second kind of nonideality that arises because the controlling charge Q_3 fails to exercise full control over the controlled charge Q_{21} . One must attend to this nonideality in modeling the charges and the transit time; but, again, it affects only the functional dependencies of the circuit elements in the network representation of Fig. III-3(c), not the representation itself. Hence, this kind of nonideality also requires no change in the self-consistency criteria for a three-terminal model of device behavior.

Indeed, all three-terminal large-signal device models for the intrinsic bipolar transistor [4,7,13-27] or MOSFET [9,27-35] commonly used in computer-aided circuit analysis conform to the network representation of the ideal charge control model of

-
- [13] J. Logan, *Bell System Tech. J.*, 50, 1105 (1971).
 - [14] H. K. Gummel and H. C. Poon, *Bell System Tech. J.*, 49, 827 (1970).
 - [15] D. J. Hamilton, F. A. Lindholm, and J. A. Narud, *Proc. IEEE*, 239 (1964).
 - [16] C. Rosenberg, D. S. Gage, R. S. Caldwell and G. H. Hanson, *IEEE Trans. Nucl. Sci. (Special Issue on Nuclear Radiation Effects Conference)*, NS-10, 149 (1963).
 - [17] J. G. Linvill, *Proc. IRE*, 46, 1141 (1958).
 - [18] L. D. Miliman, W. A. Massena, and R. H. Dickhaut, "CIRCUS, digital computer program for transient analysis of electronic circuits - user's guide," *Harry Diamond Lab., Tech. Rep.*, 346-1, Jan. 1967.
 - [19] J. G. Bowers and S. R. Sedore, *SCEPTRE: A Computer Program for Circuits and Systems Analysis*. Englewood Cliffs, N.J.: Prentice-Hall, 1971.
 - [20] A. F. Malmberg, "NET-2 network analysis program - preliminary user's manual," *Harry Diamond Lab., Tech. Rep.*, May 1970.
 - [21] T. E. Idleman, F. F. Jenkins, W. J. McCala and D. O. Pederson, *IEEE J. Solid-State Circuits (Special Issue on Computer-Aided Circuit Analysis and Device Modeling)*, SC-6, 188-203 (1971).
 - [22] B. R. Chawla, *ibid.*, (Corresp.), SC-6, 262-264 (1971).
 - [23] J. G. Fossum, *Proc. IEEE (Lett.)*, 60, 756 (1972); *IEEE Trans. Electron Devices*, ED-20, 582 (1973).
 - [24] F. A. Lindholm, *Transistor Circuit Models*, Solid-State Materials and Devices, Final Rep. AFCRL-TR-74-0044, 1974.
 - [25] F. A. Lindholm, *IEEE Trans. Circuit Theory*, CT-18, 122 (1971).
 - [26] P. Rohr and F. A. Lindholm, *IEEE J. Solid-State Circuits*, SC-10, 65 (1975).
 - [27] See the categorized bibliography: A. H. Agajanian, *Solid-State Electronics*, 18, 917 (1975).
 - [28] H. Shichman and D. A. Hodges, *IEEE J. Solid-State Circuits*, SC-3, 285 (1968).
 - [29] D. Frohman-Bentchkowsky and L. Vadasz, *ibid.*, SC-4, 57 (1969).
 - [30] D. J. Hamilton, F. A. Lindholm, and A. H. Marshak, *Principles and Applications of Semiconductor Device Modeling*, Holt, Rinehart and Winston, Inc., New York, 1971, Ch. 5.

Fig. III-3(c), provided only that current sources are included to account for net recombination and generation in the bipolar transistor. The basic differences among the various models in common use occur in the functional dependencies of the circuit elements in the network representation. Apparent differences are suggested by the fact that circuit configurations other than Fig. III-3(c) are widely employed to represent the bipolar transistor; the transport Ebers-Moll configuration [13], is one example. However, a simple network transformation will change any of these into the circuit representation of the ideal charge-control model, modified, as we have stated, by the addition of a current source in parallel with each capacitor [24].

In MOS transistors, there is another kind of nonideality of central importance. Many large-signal circuit applications cause the potential of the substrate to change with time, rapidly sometimes, relative to the potentials of the source or gate or drain. Hence the potential, or charge, associated with the substrate can influence the charge Q_{21} of mobile carriers in the channel.

This necessitates an extension of our previous view of a charge-control device and of the ideal charge-control model. The extension is straightforward. To the device pictured in Fig. III-1 add terminal 4 to represent the substrate, and regard this terminal as another gate capable of influencing Q_{21} . Then the ideal charge-control model for the resulting four-terminal device involves these extensions of its defining properties:

-
- [31] J. E. Meyer, *RCA Review*, 32, 42 (1971).
 - [32] R.S.C. Cobbold, *Theory and Applications of Field-Effect Transistors*, Wiley, New York, 1970, Ch. 8.
 - [33] G. A. Armstrong and J. A. Magowan, *Electronics Letters*, 7, 282 (1971).
 - [34] J. J. Kalinowski, *Proc. IEEE*, 60, 1000 (1972).
 - [35] F. S. Jenkins, E. R. Lane, W. W. Lattin, and W. S. Richardson, *IEEE Trans. Circuit Theory*, CT-20, 649 (1973).

Property (b) and (III-2) become

$$dQ_{21} = -dQ_3 + dQ_4 . \quad (\text{III-2}')$$

Property (c) and (III-3) become

$$i_3 = \dot{Q}_3 \text{ and } i_4 = \dot{Q}_4 . \quad (\text{III-3}')$$

Property (a) and (III-1) remain unchanged.

Thus the reasoning we have followed using the three-terminal model and the conclusions derived from it apply equally when the fourth terminal must be taken into account. In particular, the self-consistency criterion for the quasi-static approximation expressed in (III-9) still holds if one extends the meaning of the notation so that

$$\frac{\sum_{k \neq j} C_{jk} \dot{v}_{jk}}{Q_{21}/t_{21}} \ll 1 \quad (\text{III-9}')$$

indicates the self-consistent validity of the quasi-static approximation, with $j = 3$ and 4 . Here, the C_{jk} 's designate capacitors others [31-35] have used to represent capacitive effects occurring in the four-terminal MOSFET.

Because of the asymmetries in physical structure present in the MOSFET, however, capacitors alone cannot completely represent these effects [12]. For broadest applicability of the criterion expressed above, therefore, one should interpret the terms in the numerator of (III-9') as designating the time-rate accumulation, within the device, of the components of mobile-carrier charge, by whatever circuit elements are needed to represent this accumulation. For the network representations of four-terminal behavior now used widely for circuit analysis [31, 32, 35], however, a literal interpretation of (III-9') applies: that is to say, the C_{jk} 's denote capacitors.

2.2 Model for the Extrinsic Device

The division shown in Fig. (III-1) determines that the intrinsic part is a charge-control device, whose behavior is approximated by properties of (III-1) through (III-3) given in the previous section. On the other hand, this same division determines that the behavior of the extrinsic part will not depend on

charge control. For example, the number of *mobile* carriers contained within the source or drain islands of a MOS transistor is fixed to a good approximation by the resistivity of these regions, independently of the charges associated with the terminals bordering the regions. Hence the modeling of the extrinsic device follows a different scheme than that described for the intrinsic device, which requires approximate solutions for the dc steady state of the basic differential equations [1]. The scheme for the extrinsic structure depends on intuitive reasoning, and much of the modeling is done by inspection [9]. As we shall see, however, the self-consistency criterion for the extrinsic part greatly resembles the criterion we have already developed.

Because treatment of the extrinsic part cannot be unified in the framework of the charge-control model, extrinsic network representation will depend strongly on the details of device structure. Even within one category of device--THE MOSFET, for example--differing structure may require differing network representation; and, typically, the representations for the MOS and bipolar transistors will be highly dissimilar. For concreteness, therefore, we shall restrict discussion here to one type of device: the enhancement-mode MOSFET indicated in Fig. (III-1). Nonetheless, the line of argument we shall follow, and the general statement of the self-consistency criterion deriving from it, are meant to apply widely: to the bipolar transistor and to other devices as well.

Consider first the dc steady state. The dominant process occurring within the extrinsic MOS transistor is the conduction of current through resistive material; and we model this, to the first order, by the three resistors connecting intrinsic terminals to actual terminals shown in Fig. (III-4).

Consider now the transient response. In addition to conduction, capacitive effects also occur, arising from the overlap of the gate metal above the source and drain islands and from the pn junctions of these islands. Because the capacitive effects are distributed, RC transmission lines, with nonlinear capaci-

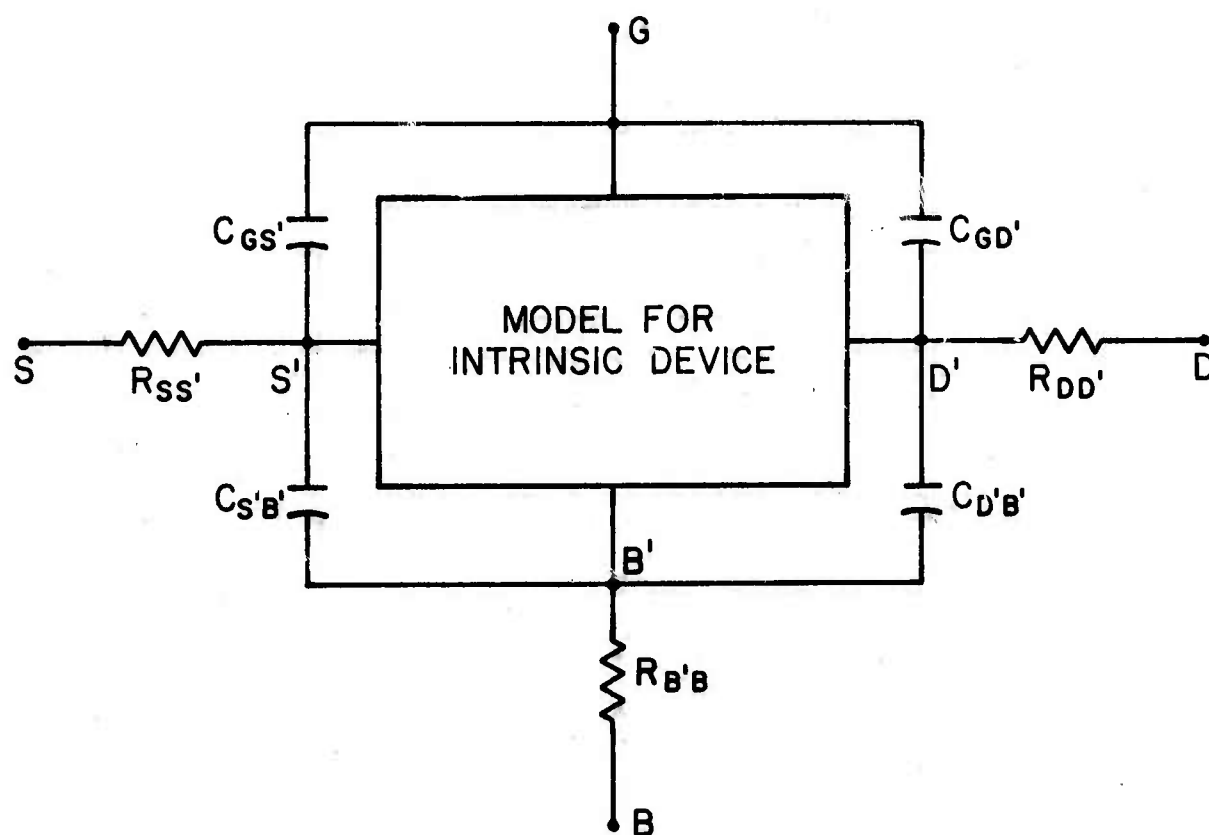


Fig. III-4. One possible network representation of the extrinsic part of the MOS transistor pictured in Fig. 1.

tances representing the pn junctions, offer a refined approximation of the behavior. From a computational standpoint, however, they pose such severe difficulties in circuit analysis that one resorts to coarser approximations [9]: for instance, the lumped circuits of Fig. III-4, consisting of one resistor and two capacitors. A further reduction to a two-element circuit, gained by letting the resistor have zero value, may additionally simplify computation [2].

Thus the model of Fig. III-4 for the extrinsic device comes from approximating a distributed network by a lumped network. This is a *quasi-static approximation*, which is exact in the dc steady-state. Therefore, as for the intrinsic part, we take the following as the self-consistency criterion for the network representation of the extrinsic device: *that the configuration of current flow in the network representation during transients must correspond to only a small perturbation of the configuration existing in the dc steady state.* Hence, from Fig. III-4, the criterion is

$$\frac{|C_{GS}, \dot{V}_{GS}| + |C_{S,B}, \dot{V}_{S,B}|}{|R_{SS}, V_{SS}|} \ll 1 \quad (\text{III-12})$$

for the network representation of the source island. An analogous criterion applies for the drain island. If more circuit elements are used to approximate the distributed line, the same criterion holds, with (III-12) now interpreted as applying to each three-element section of the network representation.

2.3 Self-Consistency Test

A circuit whose behavior is to be simulated by computer may involve many devices, each of which may have a network representation composed of the models described here for the extrinsic and intrinsic parts. It is routine in the simulation to compute the current in each branch of the network representations at each time step in the computation. Inserting the values of these currents in (III-9') for the intrinsic device and in (III-12),

or a suitable counterpart, for the extrinsic device then constitutes a test for the self-consistency of the quasi-static approximation and of the models that are based on it. Note that the intrinsic and extrinsic models of a device must each be tested for obedience to the self-consistency criteria.

The implementation of the self-consistency test in MOSFET circuit simulation will be treated elsewhere [2], together with discussions of the violations of self-consistency from a physical standpoint and of their influence on the system parameters of MOSFET circuits.

3.0 Commentary

The self-consistency criteria derived here apply generally to all electronic devices whose intrinsic structure operates in accord with charge control. The criteria hold for all regions of device operation and include consideration of the extrinsic part of the device. In previous related discussions [9,36,37], the scope was more limited. These treated the intrinsic device only and concentrated only on the region of operation in which terminal 3 of Fig. III-2 acts as the sole collector of the controlled current: that is, forward-active operation of the bipolar transistor and pinch-off operation of the MOSFET. Moreover, previous treatments have focussed only on a three-terminal model of device behavior, ignoring the role of a fourth terminal, which is a non-ideality of practical significance during large-signal transients in some MOSFET circuits.

In the present paper, nonidealities are *defined* as departures from the behavior predicted by the extended ideal charge-control model set forth in Section 2. Nonidealities receive careful attention. Despite their existence, the following statement, derived for the ideal charge-control model, holds generally:

-
- [36] P. E. Gray, D. DeWitt, A. R. Boothroyd, and J. F. Gibbons, *Physical Electronics and Circuit Models of Transistors*, John Wiley and Sons, Inc., New York, 1965, pp. 214-222.
 - [37] P. E. Gray and C. L. Searle, *Electronic Principles--Physics Models and Circuits*, John Wiley and Sons, Inc., (1969) pp. 293-296, 336-341.

the configuration of current flow in the network representation of large-signal device behavior during transients must correspond to only a small perturbation of the configuration existing in the dc steady state. Otherwise the quasi-static approximation, on which the network representation is based, will be violated. The statement applies individually to the representations of the intrinsic and the extrinsic part of a device.

The general statement above is phrased mathematically in expressions (III-9) and (III-9') for the intrinsic structure and in (III-12) for the extrinsic structure. It enables self-consistency to be tested by comparing currents in various branches of the network representation. Since these currents are available in any simulation of circuit behavior, the test is easily implemented in circuit-analysis programs [2].

3.1 Model Updating

If a particular device model fails to pass the test in a circuit simulation, a more complex model is needed to assure self-consistency with the quasi-static approximation.* To help determine the degree of added complexity that is necessary, one can again employ the self-consistency test, extended by the following strategy.

Recall that the test is to be applied individually to the extrinsic and intrinsic models of the device. Consider first the model for the extrinsic structure. If the simple three-element model of Fig. III-4 disobeys the inequality expressed in (III-12), and thus violates self-consistency, an additional three-element section can be added to give a better approximation of the distributed nature of the extrinsic structure. As was already discussed in Section 2.2, self-consistency is then tested again, with (III-12) now interpreted as applying for each section of the extrinsic model. This procedure is repeated until self-consistency

* It is possible for device models to radically fail the self-consistency test, and yet to serve adequately in the simulation of such system parameters as the turn OFF propagation delay time. This issue is explored in reference [2].

ultimately prevails in the extrinsic model. That it will prevail is assured because the capacitance and resistance per section both diminish as more sections are added; ultimately, as more sections are added, the resistive current in each section will much exceed the capacitive current and (III-12) will be obeyed. Notice that the self-consistency test will determine how many sections must be added to cause this to occur.

Adding complexity to the *extrinsic* model by itself tends to favor self-consistency of the model for the intrinsic structure. This occurs because the more complex the extrinsic model, the more effectively it filters out the higher frequencies contained in the excitation at the device terminals, preventing them from reaching the terminals of the intrinsic model. Hence, in accord with the alternative criterion for self-consistency expressed in (III-11), a detailed model for the extrinsic structure tends to protect the self-consistency of the model for the intrinsic structure.

Despite the complexity of the extrinsic model, the model for the intrinsic device may still violate self-consistency by failing to obey inequalities (III-9) or (III-9'). To remedy this, just as for the extrinsic model, one can complicate the network representation to give a better lumped network approximation of the distributed nature of the intrinsic structure. One simple example of this, which utilizes only models now in common use for MOSFET computer simulation, involves the addition of nonlinear capacitors to represent the transient currents flowing to the substrate terminals [2]. Notice that the self-consistency test implied by (III-9) and (III-9') would suggest the degree of added complexity needed.

Another strategy for model updating would require generating multisection models for the intrinsic device, similar to those for the extrinsic structure discussed immediately above and in

Section 2.2. Various techniques [30,33] might be used to furnish such models. The basic idea here is simple. Decreasing the size of the region being modeled will ultimately bring the boundaries close enough together to enable mobile carriers within the region to distribute themselves so rapidly that they assume essentially the same relationship to the boundary conductions during transients as exists in the dc steady state. Thus, use of enough sections in an intrinsic model will ultimately yield self-consistency with the quasi-static approximation. To determine the degree of complexity needed, one could employ the test implied in (III-9) and (III-9'), extended in interpretation to apply section-by-section.

Work on multisection models has started, and is meant to be the subject of future writings.

-
- [38] C. T. Sah, *Solid-State Electronics*, 13, 1547 (1970); *Physica Status Solidi (a)*, 7, 713 (1971); *Electronics Letters*, 3, 88 (1972).

Chapter IV

Carrier Mobility in an MOSFET

C. T. Hsing and D. P. Kennedy

	<u>Page</u>
1.0 Introduction.....	101
2.0 Solution for the Schroedinger's Equation for..... an MOS Structure Assuming a Constant Potential Gradient	103
2.1 Formulation of the Problem.....	105
2.2 Inversion Layer Carrier Distribution..... in a MOS Structure	109
2.3 Discussion.....	111
3.0 A Method of Solving the Schroedinger and..... Poisson Equations in a Simultaneous Manner	117
4.0 Research Plan.....	121

CHAPTER IV
Carrier Mobility in an MOSFET
C. T. Hsing and D. P. Kennedy

1.0 Introduction

Conventional theory of MOSFET operation [1,2] 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 voltage. It has been experimentally established that the inversion layer carrier mobility exhibits a large change, with a change of applied gate voltage, throughout the normal range of device operation. A theory for this phenomenon was first formulated by J. R. Schrieffer [3]; he proposed that diffuse (random) scattering of the carriers at the oxide-semiconductor interface was the basic source of this difficulty. Following his work, other researchers proposed theories to explain this change of mobility with gate voltage [4,5]. However, none were found to be in satisfactory agreement with experiment [6].

The reduction of inversion layer carrier mobility is believed to result from scattering at the oxide-semiconductor interface, yet the details of this scattering mechanism are not adequately understood. It is evident that this scattering process is intimately related to the average distance between these inversion layer carriers and the silicon surface: a decrease of this dis-

-
- [1] H.K.J. Ihantola and J. L. Moll, *Solid State Electronics*, 7, 423 (1964).
 - [2] R.S.C. Cobbold, *Theory and Application of Field-Effect Transistors*, Wiley-Interscience, N.Y. (1970).
 - [3] J. R. Schrieffer, *Phys. Rev.*, 97, 641 (1955).
 - [4] R. F. Greene, D. R. Frankl, and J. Zemel, *Phys. Rev.*, 118, 967 (1960).
 - [5] R. F. Pierret and C. T. Sah, *Solid State Electronics*, 11, 279 (1968).
 - [6] N.S.J. Murphy, F. Berz, and I. Flinn, *Solid State Electronics*, 12, 775 (1969).

tance 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. It is toward this goal the present research has been directed.

In an MOS structure, the inversion layer carriers are bounded within a potential well; on one side there is the oxide and semiconductor interface, and on the other side a large substrate electric field. For this reason, it was postulated that these inversion layer carriers would exhibit a quantized energy distribution, as in most problems of this type [7]. The proof was obtained for this quantization in the form of Shubneknov-de Hass oscillations [8]. Thereby, it was established that mechanisms other than traditional electrostatics determine the carrier distribution in the inversion layer of an MOSFET.

In traditional theory of MOSFET operation it was initially assumed that the inversion layer carrier distribution could be established from solutions of Poisson's equation. This proof of quantization in energy clearly established that a true evaluation of this carrier distribution required a simultaneous solution of both Poisson's equation and Schroedinger's equation [9,10]. To date, all available solutions for this problem have utilized a simplifying approximation that is unwarranted from a physical point of view, and which is inadequate for surface scattering calculations. Namely, it is assumed the oxide-semiconductor potential barrier is exceedingly large and, thereby, all eigenfunctions are zero at this location. Because tunneling has,

-
- [7] J. R. Schrieffer, *Semiconductor Surface Physics*, (edited by R. A. Kingston), Univ. of Pennsylvania Press (1957).
 - [8] A. B. Fowler, F. F. Fang, W. E. Howard and P. J. Stiles, *Phys. Rev. Letters*, 16, 901 (1966).
 - [9] F. Stern, *Phys. Rev. Letters*, 21, 1687 (1968).
 - [10] F. Stern, *Phys. Rev. B*, 5, 4891 (1972).

indeed, been observed into the gate oxide [11], a zero eigenfunction boundary is inconsistent with experiment. Further, previous calculations were based upon an equilibrium solution for Poisson's equation, which is inapplicable to studies of MOSFET operation.

In the present investigation we aim toward a rigorous numerical solution of this Schroedinger-Poisson problem. Clearly, all numerical solutions of this type require a first "guess"; the quality of this supposition will significantly influence the computer time required to attain a complete solution of this problem.

The present work is directed toward solving the Schroedinger's equation, based upon the inversion layer potential distribution obtained by a rigorous solution of Poisson's equation. From the solution of Schroedinger's equation, we can thereafter introduce into Poisson's equation the calculated inversion layer carrier distribution, and obtain an improved estimate of the inversion layer potential distribution. In this fashion, a Picard iteration between Poisson's equation and Schroedinger's equation can be used to obtain an accurate numerical evaluation of the inversion layer carrier distribution.

The following discussions outline the procedure of solving these two equations conjunctively. A first trial solution of Schroedinger's equation, assuming a constant potential gradient at the interface, is also presented.

2.0 Solution for the Schroedinger's Equation for an MOS Structure Assuming a Constant Potential Gradient [12]

During an early phase of this research, an analytical solution for the Schroedinger's equation was obtained by assuming a constant potential gradient at the interface. This solution can be stated analytically in terms of airy functions. This linear potential distribution is based upon an assumption that a large electric field of constant magnitude extends from the oxide-semiconductor interface into the substrate, see Figure IV-1.

[11] M. A. Green, F. D. King, and J. Shewchun, *Solid State Electronics*, 17, 551 (1974).

[12] C. T. Hsing and D. P. Kennedy, *HDL-CR-75-193-1 Final Report*, October (1975).

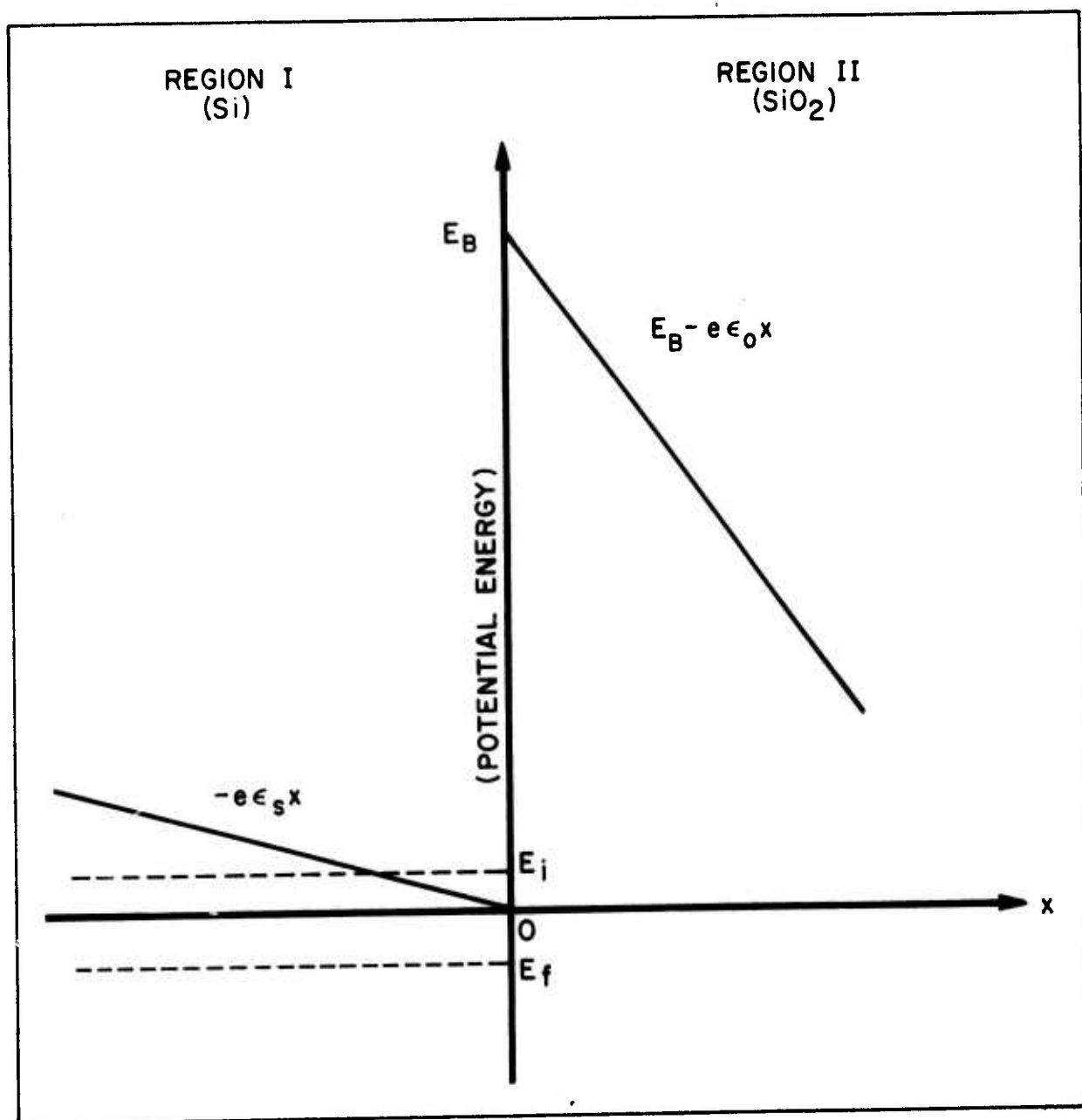


Fig. IV-1. Simplified energy diagram at the interface of Si-SiO₂.

2.1 Formulation of the Problem

In Figure IV-1 all potential energies are measured with respect to the conduction band edge of the semiconductor. For this reason, the Fermi level (E_F in Fig. IV-1) is a negative quantity, whereas the energy eigenvalues (E_i) are positive quantities.

Thus, separate forms of Schroedinger's equation must be specified for the two regions of this semiconductor structure. Thereafter, the solutions of these two equations can be matched at their boundary ($x=0$):

$$\text{Region I} \quad \frac{d^2 \psi_{Ii}}{dx^2} + \frac{2m}{h^2} (E_i + e\epsilon_s x) \psi_{Ii} = 0 \quad (IV-1)$$

($x \leq 0$)

$$\text{Region II} \quad \frac{d^2 \psi_{IIi}}{dx^2} + \frac{2m}{h^2} (E_i - E_B + e\epsilon_o x) \psi_{IIi} = 0 \quad (IV-2)$$

($x \geq 0$)

In (IV-1) and (IV-2), m represents the effective mass of electrons in a direction perpendicular to the semiconductor-oxide interface.

Substantial simplification of these equations is obtained through a change of variables. For this reason, we assume the following:

$$\psi_{Ii} = \phi_I(\alpha_i) \quad (IV-3)$$

$$\psi_{IIi} = \phi_{II}(\beta_i) \quad (IV-4)$$

where

$$\alpha_i = \left[x + \frac{E_i}{e\epsilon_s} \right] \kappa_\alpha \quad (IV-5)$$

$$\beta_i = - \left[x + \frac{E_i - E_B}{e\epsilon_o} \right] \kappa_\beta, \quad (IV-6)$$

and

$$\kappa_\alpha = \left[\frac{2me\epsilon_s}{h^2} \right]^{1/3} \quad (IV-7)$$

$$\kappa_\beta = \left[\frac{2me\epsilon_o}{h^2} \right]^{1/3} \quad (IV-8)$$

Introducing these expressions into (IV-1) and (IV-2) we obtain

$$x \leq 0 \quad \frac{d^2 \phi_I}{d\alpha_i} + \alpha_i \phi_I = 0 \quad (\text{IV-9})$$

$$x \geq 0 \quad \frac{d^2 \phi_{II}}{d\beta_i} + \beta_i \phi_{II} = 0. \quad (\text{IV-10})$$

It can be shown [13] that solutions for (IV-9) and (IV-10) have the form:

$$\phi_I(\alpha_i) = a \sqrt{\alpha_i} J_{1/3}(\zeta_i) + b \sqrt{\alpha_i} J_{-1/3}(\zeta_i) \quad (\text{IV-11})$$

$$\phi_{II}(\beta_i) = c \sqrt{\beta_i} K_{1/3}(\xi_i), \quad (\text{IV-12})$$

where $J_{1/3}$ and $J_{-1/3}$ are Bessel functions of the first kind and of orders $1/3$ and $-1/3$, respectively. Similarly, $K_{1/3}$ is a modified Bessel function of the second kind, and of order $1/3$. The parameters ζ_i and ξ_i are given by

$$\zeta_i = \frac{2}{3} \alpha_i^{3/2}; \quad \xi_i = \frac{2}{3} \beta_i^{3/2} \quad (\text{IV-13})$$

In (IV-11) and (IV-12), the terms a , b , and c arbitrary constants used to satisfy the particular boundary value problem under consideration.

The relative magnitudes of a and b in (IV-11) are readily established using (IV-3) in conjunction with Fig. IV-1. Because $x \leq 0$ in Region I, the variable α_i becomes negative when $x < -E_i/e\epsilon_s$ and, thereby, the Bessel functions of (IV-11) have a negative argument yielding

$$\phi_I(\alpha_i) = -a \sqrt{|\alpha_i|} I_{1/3}(\alpha_i) + b \sqrt{|\alpha_i|} I_{-1/3}(\alpha_i), \quad (\text{IV-14})$$

where $I_n(\alpha_i)$ is a modified Bessel function of the first kind.

[13] F. B. Hildebrand, *Advanced Calculus for Applications*, p. 156, Prentice-Hall (1962).

Insight is gained by considering an approximation for $I_{1/3}(\alpha_i)$ and $I_{-1/3}(\alpha_i)$ that is valid when α_i is very large [14]:

$$I_{1/3}(\alpha_i) = I_{-1/3}(\alpha_i) = \frac{e^{|\zeta_i|}}{2\pi|\zeta_i|} \quad (\text{IV-15})$$

Equation (IV-15) is unbounded when $\alpha_i = \infty$. Clearly, no valid solution for this problem can be unbounded; this situation is avoided in (IV-11) by setting $a = b$, thus

$$\begin{aligned} \phi_I(\alpha_i) &= a\sqrt{\alpha_i} \left[J_{1/3}(\xi_i) + J_{-1/3}(\zeta_i) \right] \\ &= a\text{Ai}(-\alpha_i) \end{aligned} \quad (\text{IV-16})$$

where $\text{Ai}(m)$ is the Airy function [15].

Similarly, (IV-12) can be written in terms of the Airy function:

$$\phi_{II}(\beta_i) = c\sqrt{\beta_i} K_{1/3}(\xi_i) = c\text{Ai}(\beta_i) \quad (\text{IV-17})$$

Thereby, the two solutions for this problem are obtained in one common functional expression.

Next we establish the arbitrary constants (a and b) associated with (IV-16) and (IV-17). This is readily accomplished by assuming continuity for both the magnitudes and derivatives of ϕ_I and ϕ_{II} at the semiconductor-oxide boundary ($x = 0$):

$$\phi_I(\alpha_i) \Big|_{x=0} = \phi_{II}(\beta_i) \Big|_{x=0} \quad (\text{IV-18})$$

$$\frac{d\phi_I}{d\alpha_i} \cdot \frac{d\alpha_i}{dx} \Big|_{x=0} = \frac{d\phi_{II}}{d\beta_i} \cdot \frac{d\beta_i}{dx} \Big|_{x=0} \quad (\text{IV-19})$$

[14] F. B. Hildebrand, *Advanced Calculus for Applications*, p. 151, Prentice-Hall (1962).

[15] M. Abramowitz and I. A. Stegien, *Handbook of Mathematical Functions*, U. S. Dept. of Commerce, Washington, D.C. (1964).

Substituting (IV-16) and (IV-17) into (IV-18) yields the equality

$$a\text{Ai}(-\alpha_{oi}) = c\text{Ai}(\beta_{oi}) \quad , \quad (\text{IV-20})$$

where α_{oi} and β_{oi} designate the magnitudes of α_i and β_i at $x=0$, respectively. Similarly, after substituting (IV-16) and (IV-17) into (IV-18), we obtain

$$a\text{Ai}'(-\alpha_{oi})\kappa_\alpha = -c\text{Ai}'(\beta_{oi})\kappa_\beta \quad , \quad (\text{IV-21})$$

where $\text{Ai}'(m)$ represents the first derivative of the Airy function.

From (IV-20) and (IV-21), we obtain the relation

$$\text{Ai}(\beta_{oi})\text{Ai}'(-\alpha_{oi}) + \frac{\kappa_\beta}{\kappa_\alpha} \text{Ai}(-\alpha_{oi})\text{Ai}'(\beta_{oi}) = 0 \quad , \quad (\text{IV-22})$$

and this relation establishes the energy eigenvalues (E_i) for this boundary value problem.

Thus (IV-16) and (IV-17) represent solutions for this problem when $x \leq 0$ and $x \geq 0$, respectively. Furthermore, (IV-18) and (IV-19) are continuous at $x = 0$ when the individual eigenvalues (E_i) satisfy the equality of (IV-22). In addition, the arbitrary constant of (IV-16) is (from (IV-20)) given by

$$c_i = a_i \frac{\text{Ai}(-\alpha_{oi})}{\text{Ai}(\beta_{oi})} \quad . \quad (\text{IV-23})$$

Therefore, the eigenfunctions of this problem are given by

$$x \leq 0 \quad \psi_{Ii}(x) = a_i \text{Ai}[-\alpha_i(x)] \quad , \quad (\text{IV-24})$$

$$x \geq 0 \quad \psi_{IIi}(x) = a_i \frac{\text{Ai}(-\alpha_{oi})}{\text{Ai}(\beta_{oi})} \text{Ai}[\beta_i(x)] \quad , \quad (\text{IV-25})$$

where a_i is determined from the normalizing requirement

$$|a_i|^2 \int_{-\infty}^{\infty} |\psi_i(x)|^2 dx = 1 \quad . \quad (\text{IV-26})$$

2.2 Inversion Layer Carrier Distribution in a MOS Structure [16]

In section 2.1 we established the energy eigenvalues (E_i) and the associated eigenfunction solutions for Schroedinger's equation [$\psi_i(x)$]. From these solutions, we can obtain the spatial probability for inversion layer carriers $|\psi_i(x)|^2$ at each allowed energy level (E_i). This probability distribution, in conjunction with the density of carriers (N_i) at each energy level (E_i), yields the spatial distribution of these particular energetic carriers $N_i |\psi_i(x)|^2$. After summing the carrier distributions from all energy eigenvalues, we obtain an expression for the total distribution of inversion layer carriers within this semiconductor device:

$$n(x) = \sum_{i=1}^{\infty} N_i |\psi_i(x)|^2 . \quad (\text{IV-27})$$

In this section we derive a rigorous mathematical expression for this inversion layer carrier distribution.

The density of states in an inversion layer can be written as:

$$D(E) = D_1(E) D_2(E) , \quad (\text{IV-28})$$

where $D_1(E)$ is the density of states in a plane parallel to the oxide-silicon interface, and $D_2(E)$ is the density of states in a direction perpendicular to this interface.

First we consider directions parallel to the oxide-semiconductor interface. The total number of states per unit area in this paln residing between E and $E+dE$ is given by

$$D_1(E) dE = d\Gamma(E) = \frac{k}{2\pi} \cdot dk . \quad (\text{IV-29})$$

The electron energy in this direction is

$$E = \frac{h^2}{2m^*} (k_y^2 + k_z^2) = \frac{h^2}{2m^*} k^2 , \quad (\text{IV-30})$$

[16] The methods used here are from Gnädingers and Talley, *Solid-State Electronics*, 13, 1301 (1970).

where m' is the effective electron density-of-states mass in the y-z direction. Thus, from (IV-29), in conjunction with (IV-30), we have

$$D_1(E) = \frac{d\Gamma(E)}{dE} = \frac{d\Gamma(E)/dk}{dE/dk} = \frac{k/2\pi}{h^2k/m} = \frac{m'}{2\pi h^2} \quad (\text{IV-31})$$

The total density of states perpendicular to the oxide-semiconductor interface $D_2(E)$ can be written as

$$D_2(E) = n_v \sum_{i=1}^{\infty} H(E - E_i) \quad , \quad (\text{IV-32})$$

where n_v is the degeneracy factor. This factor depends upon the crystal orientation, with respect to the semiconductor surface. In (IV-32), $H(E - E_i)$ is a step function, and is defined by:

$$\begin{aligned} H(\delta) &= 1 & \delta &\geq 0 \\ &= 0 & < 0 \end{aligned} \quad (\text{IV-33})$$

From (IV-28), (IV-31), and (IV-32), the density of states for each energy eigenvalue within the inversion layer of a MOS structure is given by

$$D_i(E) = \left[\frac{m' n_v}{\pi h^2} \right] H(E - E_i) \quad . \quad (\text{IV-34})$$

In (IV-34), a factor two has been introduced for spin degeneracy.

The total number of electrons at each energy level E_i is therefore given by

$$N_i = \int_0^{\infty} D_i(E) f(E) dE \quad , \quad (\text{IV-35})$$

where $f(E)$ is the probability an electron occupies the energy level E . Assuming Fermi-Dirac statistics, $f(E)$ is given by

$$f(E) = \{1 + \exp[(E - E_F)/kT]\}^{-1} \quad , \quad (\text{IV-36})$$

where E_F is the Fermi energy. After substituting into (IV-35) the expressions given in (IV-34) and (IV-36), we obtain the

carrier density at each energy eigenvalue:

$$N_i = \int_0^{\infty} D_i(E) f(E) dE = \frac{m' k T n_v}{\pi h^2} \log_e \{1 + \exp[(E_F - E_i)/kT]\} . \quad (IV-37)$$

Having calculated the carrier density N_i , (IV-27) yields the density distribution of these carriers within the inversion layer of a MOS structure:

$$n(x) = \frac{m' n_v k T}{\pi h^2} \sum_{i=1}^{\infty} \log_e \{1 + \exp[q(E_F - E_i)/kT]\} |\psi_i(x)|^2 . \quad (IV-38)$$

2.3 Discussion

Figs. IV-2 and IV-3 illustrate the calculated inversion layer energy levels for two different values of electric field at the semiconductor surface: $\epsilon_s = 1.25 \times 10^5$ v/cm and $\epsilon_s = 3.4 \times 10^4$ v/cm. These two illustrations establish a general trend that has thus far been verified in this analysis: an increased level of energy quantization is realized with an increase of electric field. From a practical point of view, implied here is an increase level of quantum mechanical mechanisms with an increase of electric field. Thus, in strong inversion we can assume the inversion layer carrier distribution will be poorly described by traditional MOS theory. As the gate voltage is reduced and the structure enters weak inversion, these quantum mechanical mechanisms should tend to disappear. Thus, it is suggested that in the normal range of MOSFET operation, the inversion layer carrier distribution will change between those regions where quantum mechanical mechanisms are significant (strong inversion), and those regions where traditional electrostatic mechanisms dominate (weak inversion).

This type of situation is generally consistent with other well known solutions of Schroedinger's equation for electrons in a potential well. Under conditions of weak inversion the substrate electric field is small and, hence, the interface potential well is very wide; little energy quantization is therefore observed. Contrasting with this weak inversion case, under strong

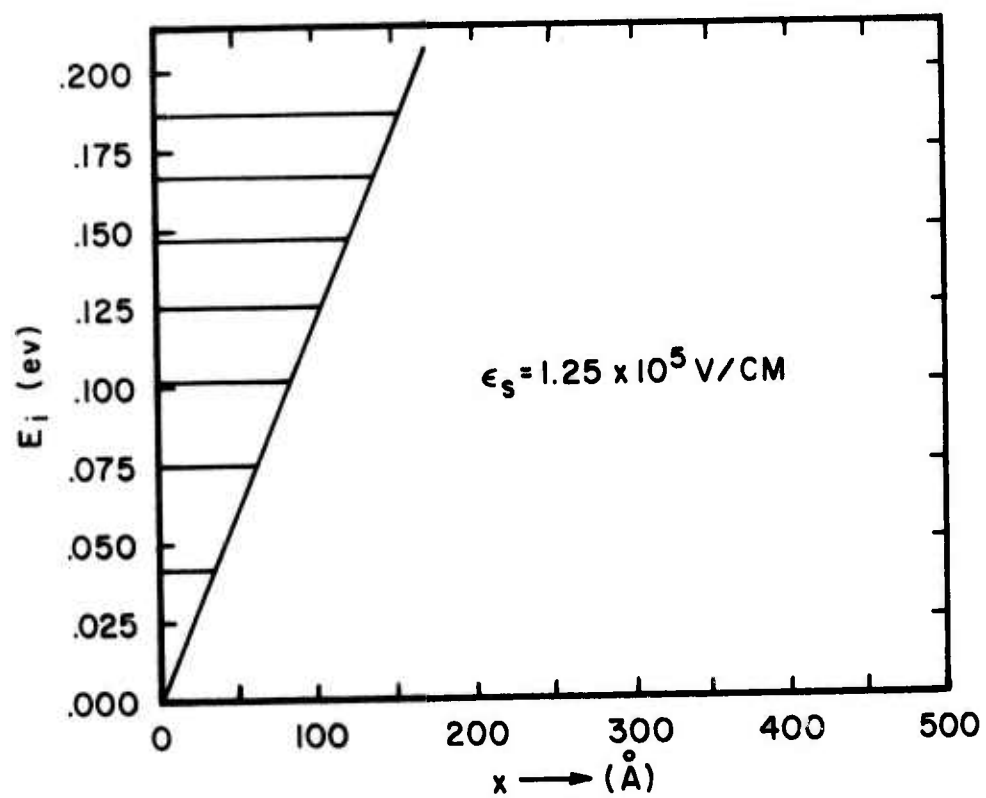


Fig. IV-2. Calculated electron energy levels in the surface potential well.

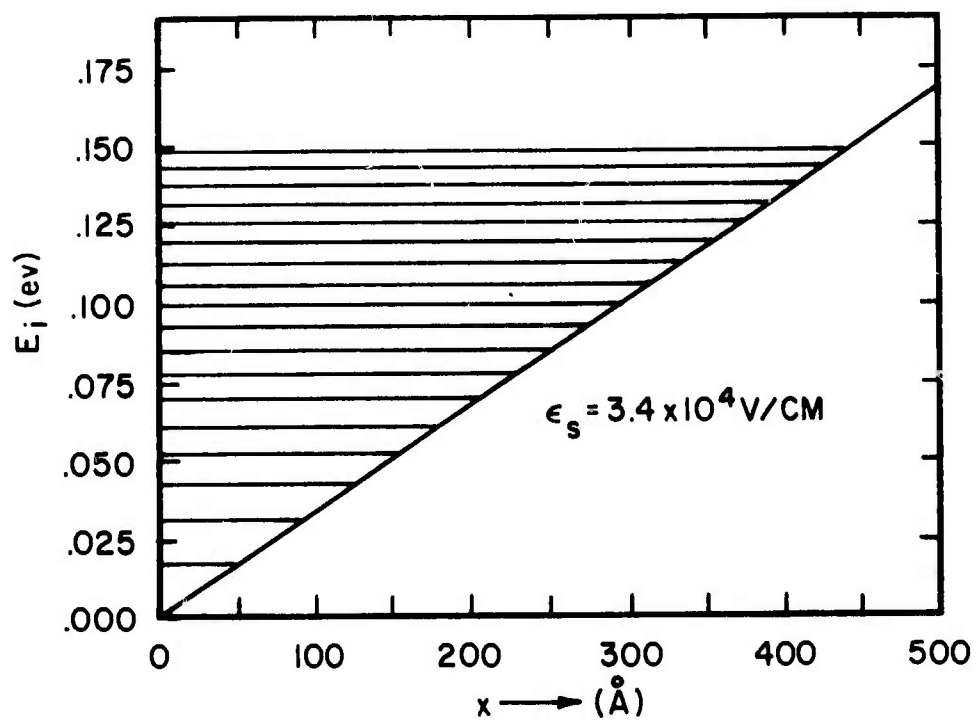


Fig. IV-3. Calculated electron energy levels in the surface potential well.

inversion the substrate electric field is very large and, hence, the potential well is very narrow. A narrow potential well tends to produce a large separation between the energy eigenvalues of Schroedinger's equation and, therefore, quantum mechanical mechanisms tend to dominate.

Figs. IV-4 and IV-5 present a comparison between the inversion layer mobile carrier distribution derived from this quantum mechanical solution and from traditional solutions of Poisson's equation.

There are several interesting aspects of the results shown in Fig IV-4 and Fig. IV-5:

- (1) The width of the inversion layer (channel) was much thinner than expected. This is clearly due to the assumption of an inaccurate constant electric field. The electric field in actual device should vary from maximum at the interface to nearly zero in the substrate;
- (2) The maximum carrier density was located at about $20 \text{ \AA}$ from the interface, and there was finite carrier density at the interface. Qualitatively, this situation is due to quantum mechanical reflection at the semiconductor-oxide boundary. The large coulomb forces attract inversion layer carriers toward the interface. Upon reaching this interface, these energetic carriers either reflect back into the semiconductor or, instead, tunnel into the oxide. This reflection mechanism produces a maximum carrier density at some location removed from the reflecting surface (or potential barrier); and
- (3) An increase in electric field will increase the maximum carrier density and decrease the inversion layer width: the carriers are pushed harder toward the interface.

Although this solution is not adequate for MOSFET analysis, it does show some qualitative insight of the characteristic of the inversion layer in an MOS structure. A rigorous approach will be discussed in the next section of this report.

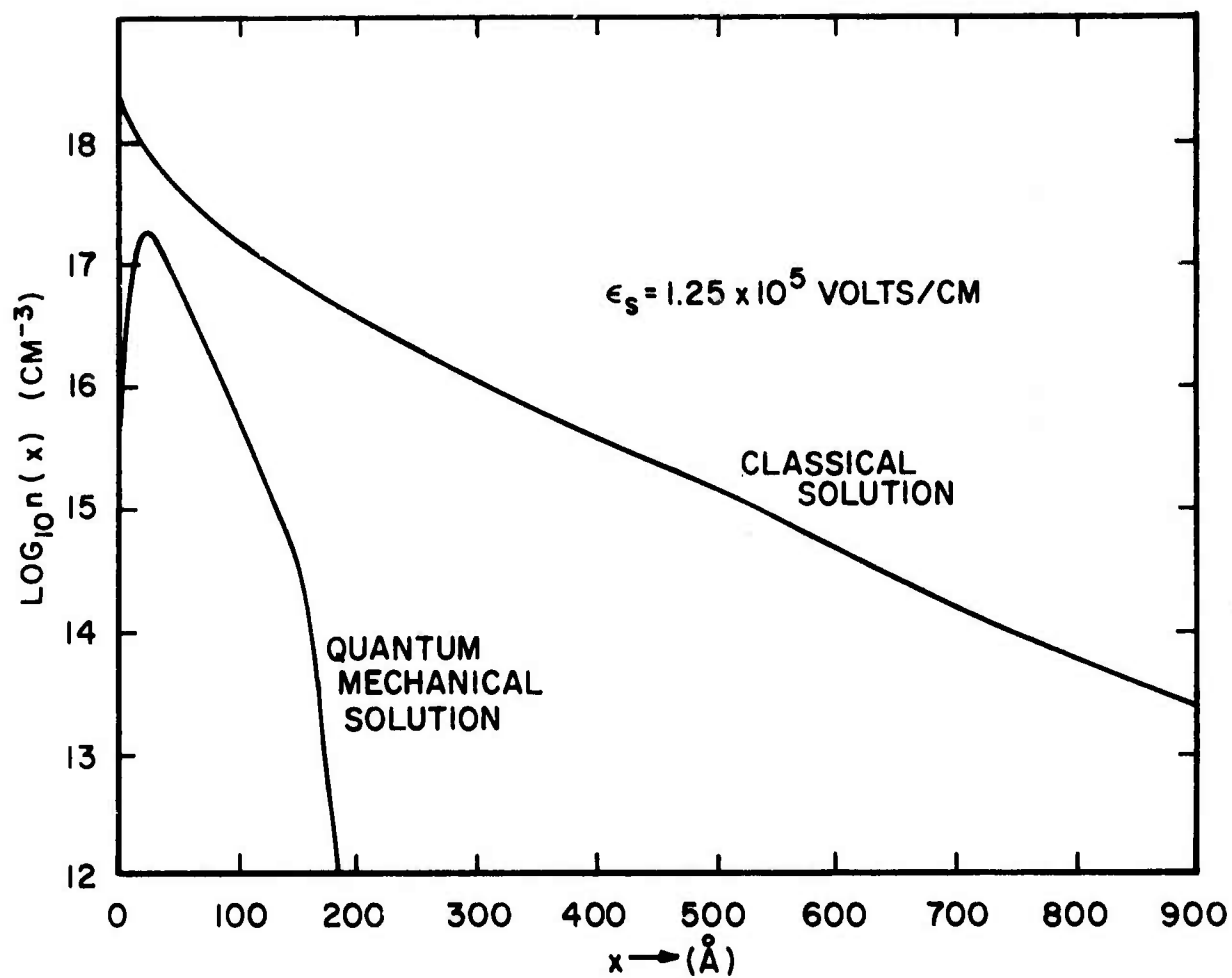


Fig. IV-4. Comparison between quantum mechanical solution and classical solution of the inversion layer carriers.

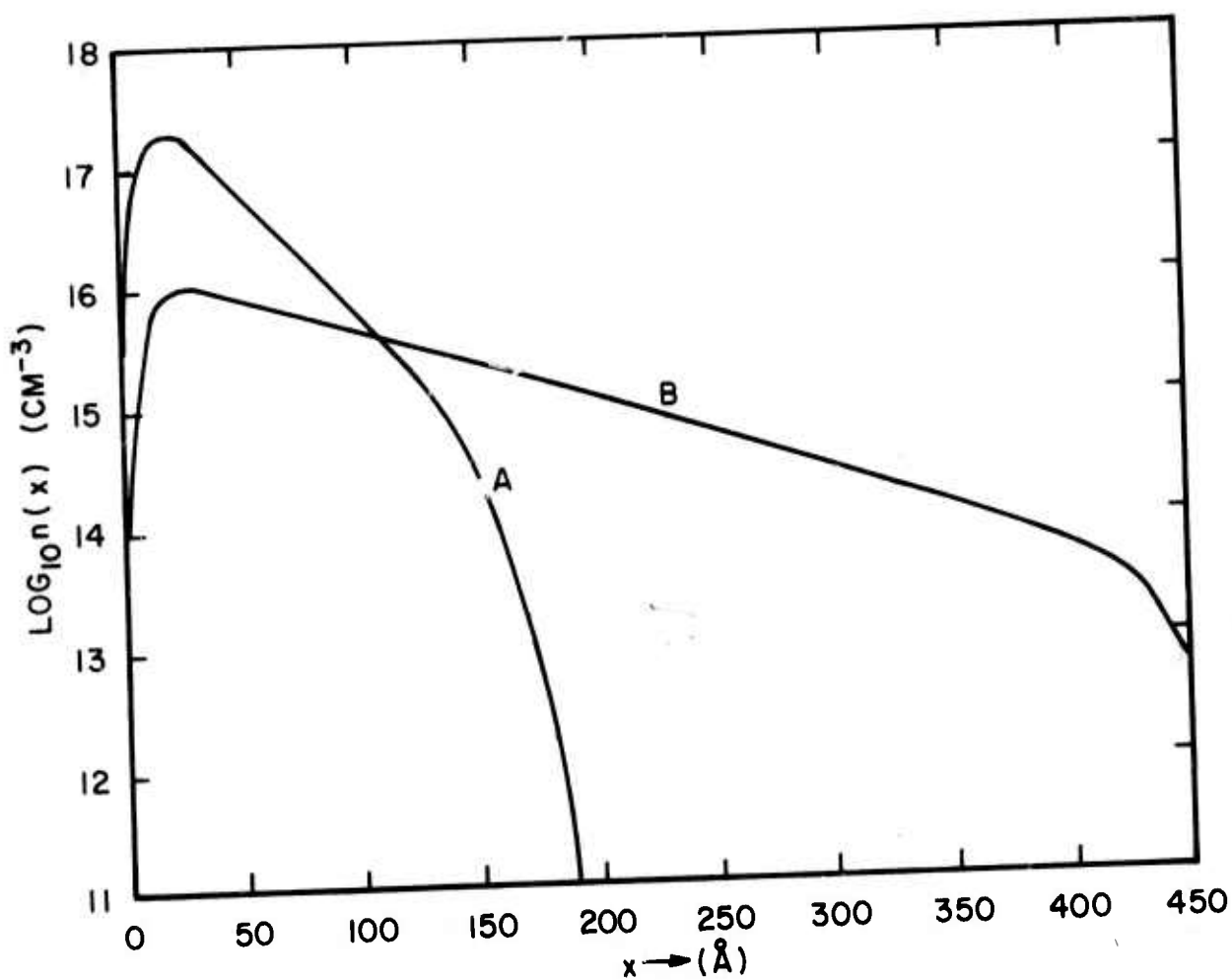


Fig. IV-5. Quantum mechanical solutions of inversion carrier distribution with different surface electric fields (ϵ_s).

(A) $\epsilon_s = 1.25 \times 10^5 \text{ V/cm}$

(B) $\epsilon_s = 3.4 \times 10^4 \text{ V/cm}$

3.0 A Method of Solving the Schroedinger and Poisson Equations in a Simultaneous Manner

A. Schroedinger's Equation

The Schroedinger's equation can be written as

$$H\psi = -\frac{\hbar^2}{2m} \frac{d^2\psi_i}{dx^2} + V(x)\psi_i = E_i\psi_i, \quad (\text{IV-39})$$

where

$$(i) \quad H \equiv -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x);$$

(ii) $V(x)$ = potential energy from the solution of Poisson's equation;

(iii) E_i, ψ_i are the eigenvalue and corresponding eigenfunction, respectively.

The boundary conditions for $\psi_i(x)$ are

$$\psi_i(\infty) = 0 \text{ and } \psi_i(-\infty) = 0.$$

There are two methods to solve (IV-39):

- (1) Green's Function Technique: This method starts with solving the differential equation for Green's function:

$$HG(x,t) = -\frac{\hbar^2}{2m} \frac{d^2G(x,t)}{dx^2} + V(x)G(x,t) = \delta(x-t) \quad (\text{IV-40})$$

with the boundary conditions

$$G(\infty, t) = G(-\infty, t) = 0.$$

After obtaining the Green's function $G(x,t)$, the wavefunction ψ_i can be obtained by the following integral equation:

$$\psi_i(x) = E_i \int_{-\infty}^x G(x,t)\psi_i(t)dt \quad (\text{IV-41})$$

Since the wavefunction ψ_i appears as a part of integrand in (IV-41), it requires an iterative process to find the wavefunction.

- (2) Rayleigh-Ritz Method: By expressing the wavefunction in terms of a series of linearly independent functions, the eigenvalues can be obtained with a very good accuracy. The desired wavefunction can be easily obtained once the eigenvalues are known. This method has important advantages: it does not require the time-consuming iterative process. Therefore, the Rayleigh-Ritz method [17] will be used in the present research, and it is described in detail in the following paragraphs.

The Rayleigh-Ritz method is used to calculate the upper bounds of the eigenvalues of a positive definite differential operator H^* . In the present problem, H is the differential operation defined in (IV-39). If we let the sequence of eigenvalues for which we seek upper bounds be denoted by $E_1, E_2, \dots, E_r, \dots$ which corresponding eigenfunctions $\psi_1, \psi_2, \dots, \psi_r, \dots$, then

$$E_r = \min[(H\psi, \psi) / (\psi, \psi)] = \min \left[\frac{\int_{-\infty}^{\infty} \psi H \psi dx}{\int_{-\infty}^{\infty} \psi \psi dx} \right] \quad (\text{IV-42})$$

with $(\psi, \psi_s) = 0; s = 1, 2, \dots, r-1$.

For any interger n , let us choose an arbitrary set of n linearly independent functions $f_i(x)$ such that $f_i(\infty) = f_i(-\infty) = 0$.

Now we can write $\psi(x)$ as:

$$\psi(x) = \sum_{i=1}^n C_i f_i(x). \quad (\text{IV-43})$$

- [17] S. H. Gould, *Variational Method for Eigenvalue Problems*, 2nd ed., p. 75, Univ. of Toronto Press, 1957.

* A differential operator H is called positive definite if

$$(f, Hf) = \int_a^b f(x) Hf(x) dx > 0$$

for all values of $f(x)$ except $f(x) = 0$. a and b are the end points of range under consideration.

For each function $\psi(x)$ written as (IV-43), we have

$$(H\psi, \psi) = (H \sum_{i=1}^n C_i f_i, \sum_{k=1}^n C_k f_k) = \sum_{i,k} (Hf_i, f_k) C_i C_k = \sum_{i,k} A_{ik} C_i C_k \quad (\text{IV-44})$$

and

$$(\psi, \psi) = \left(\sum_{i=1}^n C_i f_i, \sum_{k=1}^n C_k f_k \right) = \sum_{i,k} (f_i, f_k) C_i C_k = \sum_{i,k} b_{ik} C_i C_k, \quad (\text{IV-45})$$

$$\text{where } a_{ik} \equiv (Hf_i, f_k) = \int_{-\infty}^{\infty} f_k(x) Hf_i(x) dx$$

$$\text{and } b_{ik} \equiv (f_i, f_k) = \int_{-\infty}^{\infty} f_i f_k dx.$$

Then from (IV-42) we can obtain a set of n equations:

$$\sum_{i=1}^n C_i (a_{ik} - E b_{ik}) = 0 \quad k = 1, 2, \dots, n \quad (\text{IV-46})$$

For non-trivial solution of the simultaneous equations for C_i 's, the determinant

$$\det (a_{ik} - E b_{ik}) \equiv 0. \quad (\text{IV-47})$$

Equation (IV-47) can be expanded into a polynomial of n^{th} degree which, in general, have n distinct roots E 's. These n roots are the upper bounds of the eigenvalues. As n increases, the upper bounds of the eigenvalues decrease, i.e., they are getting closer to the true eigenvalues. Therefore, the accuracy of eigenvalues obtained by the Rayleigh-Ritz method depends upon

the number (n) of terms used in (IV-43). If we choose the set of linearly independent functions such that each of them has close resemblance to the wavefunction $\psi(x)$, then the resulting eigenvalues will be more accurate for a given n. From the solution stated in Section 2, we know that the Airy function is a good choice for this expansion and it will be used in the eigenvalue calculations.

By substituting each eigenvalue into (IV-46), the coefficients C_i 's of the corresponding eigenfunction can be found up to an arbitrary constant. This arbitrary constant is determined by the normalization condition of $\psi(x)$, i.e.,

$$(\psi, \psi) = 1 \quad (\text{IV-48})$$

After obtaining the wavefunctions, the carrier density distribution within the MOS inversion layer is given by (IV-38):

$$n(x) = \frac{m' n_v kT}{\pi h^2} \sum_{i=1}^n \log_e \{1 + \exp[q(E_F - E_i)/kT]\} |\psi_i(x)|^2, \quad (\text{IV-49})$$

where (i) m' $\equiv$ effective mass in the direction parallel to the interface;

(ii) n_v is the degeneracy factor;

(iii) E_F is the Fermi energy;

(iv) E_i and ψ_i are the eigenvalue and eigenfunction, respectively.

B. Poisson's Equation

The potential distribution within the MOS structure is given by the solution of the Poisson's equation:

$$\frac{d^2 U}{dx^2} = \frac{q^2}{\epsilon kT} [n(x) + N_A - p(x)] \equiv f(x), \quad (\text{IV-50})$$

where (i) $U \equiv \frac{qV}{kT}$ = normalized potential;

(ii) ϵ is the dielectric constant of silicon;

(iii) $n(x)$ and $p(x)$ are the density functions of electrons (inversion charge for p-type substrate) and

holes, respectively, and $p(x) = n_i/n(x)$, where n_i is the intrinsic carrier density in silicon;
 (iv) N_A is the acceptor density in the substrate.

The boundary conditions for (IV-50) are:

$$U(L) = U_s \text{ and } U(0) = 0,$$

where L is the width of space charge region (see Fig. IV-6) and U_s is the normalized surface potential. The solution of (IV-50) is

$$U = \frac{1}{L} (x-L) \int_0^x t f(t) dt + \frac{x}{L} \int_x^L (t-L) f(t) dt + \frac{x}{L} U_s. \quad (\text{IV-51})$$

The self-consistent solution of the Schroedinger's equation and the Poisson's equation can be obtained by calculating (IV-49) and (IV-51) iteratively.

4.0 Research Plan

The computer program is being written for above calculations. It is expected that the complete solution would be obtained within the next three months. Then, our effort will be directed toward the study of effective surface mobility of inversion charge carriers, which is the main object of the present research.

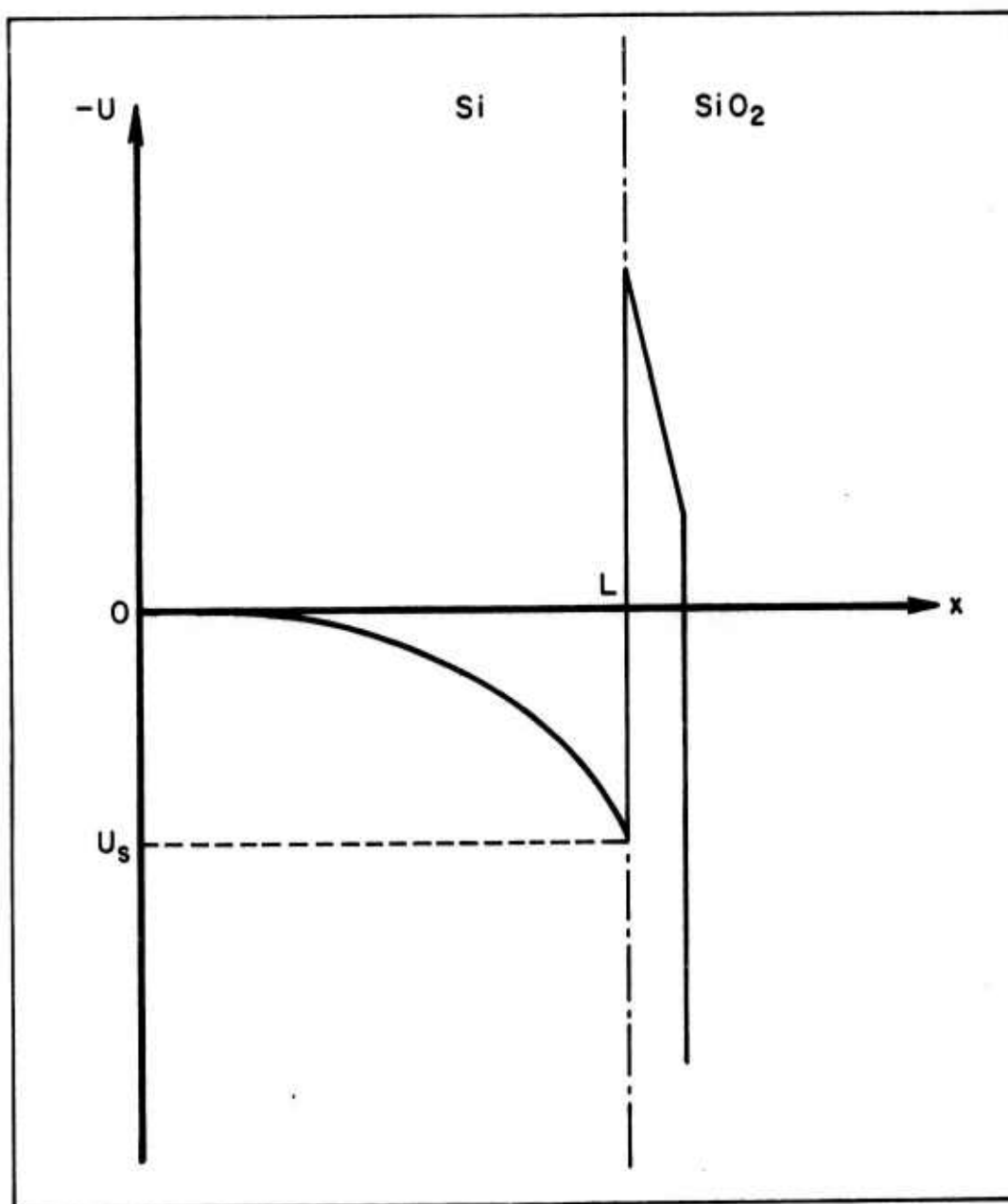


Fig. IV-6. Potential distribution at the interface of Si- SiO_2 .

Chapter V

Integral Methods for the Analysis of Semiconductor Devices

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

	<u>Page</u>
1.0 Introduction.....	125
2.0 Integral Solution Method.....	128
3.0 Results for the Constant Gradient Junction.....	132
4.0 Direction of Future Work.....	137
5.0 List of Symbols.....	140

Chapter V

Integral Methods for the Analysis of Semiconductor Devices

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

1.0 Introduction

One of the more important and yet incompletely solved problems in microelectronics is the determination of the potential and carrier density distributions in two dimensional semiconductor devices of specified geometry, impurity distribution, and applied external voltage. In principal, one should be able to determine these characteristics numerically by solving the governing nonlinear partial differential equations. In practice, however, such calculations, which have almost exclusively relied on the finite difference approach [1-4], require lengthy computation times and, therefore, are of questionable value from an economic point of view. The origin for these numerical difficulties lie in the nonlinear nature of the equations, and the need to use a large number of mesh points to adequately represent large potential gradients in certain regions of the semiconductor device. It is the objective of this mathematical study to investigate alternate methods of analysis for numerically determining the carrier and potential distributions in two dimensional semiconductor devices (such as the MOSFET), and to develop techniques which yield shorter overall computer times.

The coupled partial differential equations we wish to analyze can be found in the literature [5], and are given by:

-
- [1] H. K. Gummel, *IEEE Trans. Electron Devices*, 11, 455 (1964).
 - [2] M. S. Mock, *Solid State Electronics*, 16, 601 (1973).
 - [3] H. H. Heimeier, *IEEE Trans. Electron Devices*, 20, 708 (1973).
 - [4] O. Manck and W. L. Engl, *IEEE Trans. Electron Devices*, 22, 339 (1975).
 - [5] S. M. Sze, *Physics of Semiconductor Devices*, John Wiley and Sons, Inc., N.Y. (1969).

$$\nabla^2 V = - \frac{q}{\kappa \epsilon_0} \{N_D - N_A + p - n\} \quad (V-1)$$

$$\frac{\partial n}{\partial t} = - \frac{1}{q} \nabla \cdot \vec{J}_n + R \quad (V-2)$$

$$\frac{\partial p}{\partial t} = \frac{1}{q} \nabla \cdot \vec{J}_p + R \quad (V-3)$$

$$\vec{J}_p = -qD_p \nabla p - q\mu_p p \nabla V \quad (V-4)$$

$$\vec{J}_n = qD_n \nabla n - q\mu_n n \nabla V \quad (V-5)$$

$$0 = \nabla \cdot (\vec{J}_n + \vec{J}_p + \vec{J}_D) \quad (V-6)$$

where V is the potential, N_D and N_A the spatially dependent donor and acceptor impurity distributions, n and p the electron and hole number densities, and R the Shockley-Reed-Hall recombination factor. The remaining quantities in these equations have their standard meanings. These equations represent the Poisson equation, made nonlinear because of the approximate exponential dependence of n and p on V , coupled with continuity equations for the electron and hole currents.

Various limiting forms of these equations have been treated by earlier investigators. The first such finite-difference model, that due to Gummel [1], deals with a one-dimensional steady state problem. Mock [2] and Heimeier [3] treat two-dimensional geometries in the steady state, while Manck and Engl [4] deal with the time-varying problem in two dimensions. In Mock's model for MOSFETs the computer time is somewhat reduced by letting $R=0$ so that the continuity requirements (V-2) and (V-3) become equivalent to having divergence free hole and electron currents.

In the present study we set ourselves the goal of treating a simplified version of the governing equations by other than the finite difference technique, with the objective of reducing the required computer time. Further, we have the longer range aim of using such improved numerical approaches to solve both steady and time-dependent BIPOLAR and MOSFET problems. Drawing

See references [1]-[4] on page 125.

on some of our experience on the treatment of boundary value problems in the area of mechanics, we decided early in the program that a promising alternate numerical technique should be the integral equation method. This method consists essentially of converting a given differential equation and an appropriate set of boundary conditions (via a Green's function) to a Fredholm integral equation and then solving this equation by a Picard iteration. The advantages of such an integral approach include the elimination of the higher derivatives, and the ability to place the nonlinear terms under an integral.

The specific problem we wish to analyze below, using the proposed integral equation approach, is obtained from equations (V-1), (V-2) and (V-3) by assuming no electron or hole current, zero time dependence and recombination, and that all functions depend on one independent variable x only. In this limit the governing equations decouple and one is left with solving a one-dimensional Poisson equation, subject to an x dependent impurity distribution ($N_D(x)$ and $N_A(x)$). Our reason for considering this very simple version of the governing equations is to establish the numerical advantages of an integral approach, without obscuring this point by introducing unnecessary numerical complications. Our ultimate objective is to extend these one dimensional studies to two dimensions, and this will involve the need to generate Green's functions, which satisfy the mixed boundary conditions encountered there.

The order of our discussion below will be as follows. We begin with a derivation of the Fredholm integral equation corresponding to the nonlinear Poisson equation. This is followed by a description of the Picard iteration technique for solving such integral equations; this discussion also includes a consideration of the convergence criterion for this successive approximation approach. Next, a method of under-relaxation is discussed, and it is shown how such a method may be used (under certain circumstances) to obtain convergence when an unrelaxed Picard iteration is unstable. Finally, we present some calculations for the

constant gradient junction. Here we demonstrate the utility of the integral method and discuss briefly our proposed extension of the integral method to two dimensional problems.

2.0 Integral Solution Method

For previously mentioned reasons we chose to investigate the potential and carrier distribution densities in semiconductor devices using a Green's function-iterative solution technique. In view of the utility of integral equation approaches for solving boundary value problems in areas such as mechanics, it appeared that this approach is warranted for the problem under consideration. In order to simplify the calculations, we confined our initial attention, as previously mentioned, to the limiting form of (V-1), (V-2), and (V-3) for which there is no time dependence, the current is zero, there is zero recombination, and all quantities vary with only one independent variable x . In this limit the governing equations decouple, and one obtains the one-dimensional nonlinear Poisson equation

$$\frac{d^2U}{dx^2} = - \frac{q^2}{\epsilon_0 kT} [p(U) - n(U) + N_D(x) - N_A(x)] = F(x, U) \quad (V-7)$$

subjected to the non-homogeneous boundary conditions $U(0)=0$ and $U(W)=U_T$. Here the hole number density $p(U)$ and the electron number density $n(U)$ in (V-1) are given by the standard Boltzmann relations. Typically $x=0$ corresponds to the metallurgical junction and $x=W$ is of the order of one thousand Angstroms; this distance represents the approximate depletion layer half-width for a constant gradient function under equilibrium conditions.

This boundary value problem may be converted to a nonlinear Fredholm integral equation by means of the familiar triangular Green's function kernel as discussed by Tricomi [6]. After integrating (V-7) twice and using the left boundary condition one finds

[6] F. G. Tricomi, *Integral Equations*, p. 116, Interscience Publishers, Inc., N.Y. (1957).

$$U(x) = xU'(0) + \int_0^x dr \int_0^r F(t,U)dt \quad (V-8)$$

where r and t are dummy variables. Next, eliminating the derivative $U'(0)$ by use of the second boundary condition $U(W)=U_T$, and applying the Leibnitz rule to reduce the double integrals to single integrals, one obtains the desired nonlinear integral equation

$$U(x) = \frac{U_T x}{W} - \int_0^x \frac{t(W-x)}{W} F(t,U)dt - \int_x^W \frac{x(W-t)}{W} F(t,U)dt \quad (V-9)$$

Although equation (V-9) has no obvious analytic solution, even for very simple doping distributions $N_D(x)-N_A(x)$, it is always possible to solve this equation numerically. One of the better techniques for doing this is the method of successive approximations (Picard method) [6] used in conjunction with a trapezoidal or Simpsons rule for evaluating the integrals. In this method (V-9) is approximated by the iterative form

$$U_{n+1}(x) = \frac{U_T x}{W} - \int_0^x \frac{t(W-x)}{W} F(t,U_n)dt - \int_x^W \frac{x(W-t)}{W} F(t,U_n)dt \quad (V-10)$$

with $U_1, U_2, U_3, \dots$ forming a sequence converging to the solution $U(x)$, if the convergence criterion

$$|U_{n+1}(x) - U_n(x)| < \epsilon \text{ for } n > N \quad (V-11)$$

is met. This iteration method, when it converges, has advantages over other solution techniques of (V-9); the nonlinear term appears under the integrals and, thus, does not require linearization in determining the iterate $U_{n+1}(x)$. In actual calculations to be presented below, the solution is said to be obtained and the iteration terminated when ϵ in criterion (V-11) is less than 10^{-3} for all points in the range $0 < x < W$. As a starting value in this iteration procedure one typically uses the linear term $U_T x/W$ for $U_0(x)$.

[6] F. G. Tricomi, *Integral Equations*, p. 116, Interscience Publishers.

Unfortunately, the iterative procedure given by (V-10) converges only for restricted values of F and W . This fact is well established for Picard iterative solutions of the corresponding linear Fredholm equation, and can be partially inferred by the following modified argument to one given by Tricomi [7]. Consider the iterative form for a slightly generalized non-linear Fredholm equation

$$U_{n+1}(x) = f(x) + \lambda \int_0^W H[x, t, U_n(t)] dt \quad (V-12)$$

where $f(x)$ is a specified function of x , λ a constant and $H(x, t, U_n)$ is a nonlinear function of x, t and the n th iterate U_n of $U(x)$. Letting the maximum value of H be M for all x and t in $[0, W]$, and introducing the Lipschitz condition,

$$|H(x, t, U_{n+1}(t)) - H(x, t, U_n(t))| < C |U_{n+1}(t) - U_n(t)|, \quad (V-13)$$

where C is a constant, it follows from (V-12) that

$$U_1(x) - U_0(x) = \lambda \int_0^W H(x, t, U_0(t)) dt < \lambda MW \quad (V-14)$$

or, more generally, that

$$U_{n+1}(x) - U_n(x) < MC^n (\lambda W)^{n+1} \quad (V-15)$$

Comparing this last inequality with the convergence condition (V-11) one sees that convergence of the Picard iterative procedure is guaranteed when

$$\left| \lambda W \text{Max} \left\{ \frac{H(x, t, U_n) - H(x, t, U_{n-1})}{U_n - U_{n-1}} \right\} \right| < 1 \quad (V-16)$$

-
- [7] F. G. Tricomi, *Integral Equations*, p. 117, Interscience Publishers, Inc., N.Y. (1957).

where Max refers to the maximum value of the term in the inside brackets. We note that this last condition is quite stringent, suggesting in a qualitative sense that convergence is attained only when W remains small, and the variation in H between successive iterations is not too large. Unlike the corresponding convergence criterion for the associated linear problem [8], criterion (V-16) involves U_n explicitly and thus the above interpretations are necessarily of a qualitative nature as U_n is not known a priori.

For typical functions F and values W encountered solving (V-9) by Picard iteration one can expect the convergence criterion (V-16) to be violated and hence have the likelihood of convergence failure. Under such circumstances it is sometimes possible to obtain convergence by introducing a relaxation relation of the form

$$U_{n+1}^* = U_n + \frac{1}{v} (U_{n+1} - U_n) \quad (V-17)$$

where K is an under-relaxation parameter greater than one used to slow down the rate of variation between the iterates U_{n+1} and U_n . This procedure is quite successful in stabilizing solutions for linear problems (see Forsythe and Wasow [9]) and can be expected to aid in the convergence of the non-linear iteration problem (V-10) under conditions where the unrelaxed iteration sequence $U_n, U_{n+1}, \dots$ (corresponding to $K=1$) fails to converge. Mock [2] has also suggested improved relaxation procedures involving the use of a sequence of relaxation parameters for problems where the use of (V-17) still fails to produce convergence. It should be pointed out that when the relaxation procedure (V-17)

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

[8] F. G. Tricomi, *Integral Equations*, p. 51, Interscience Publishers, Inc., N.Y. (1957).

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

is used in conjunction with the iteration (V-10), the term U_n in F is replaced by U_n^* .

3.0 Results for the Constant Gradient Junction

To obtain some specific numerical results using the integral-iteration method described in the preceding section we considered the special and relatively elementary case of the constant gradient p-n junction where the impurity distribution near the metallurgical junction (at $x=0$) goes as $N_D(x) - N_A(x) = Ax$, with A being the grade constant having a magnitude 10^{22} atoms/cm⁴ in the present calculations. In this case the one-dimensional Poisson equation (V-7), in the presence of an applied forward potential V_A , has been shown by Morgan and Smitz [10] to have the form

$$\frac{d^2 U}{dx^2} = \frac{2n_i q^2}{\kappa \epsilon_0 kT} e^{\frac{qV_A}{2kT}} \left\{ \sinh U(x) - \frac{Ax}{n_i} \exp \left[-\frac{qV_A}{2kT} \right] \right\} \quad (V-18)$$

subject to the same boundary conditions used above, except that now $U_T = U_0 + \frac{qV_A}{2kT}$, with U_0 being the equilibrium junction potential. The corresponding integral equation follows directly from (V-9)

$$U(x) = \frac{U_T x}{W} - \gamma \int_0^x \frac{t(W-x)}{W} \left[\sinh U(t) - \frac{At}{2n_i} \exp - \frac{qV_A}{2kT} \right] dt \\ - \gamma \int_x^W \frac{x(W-t)}{W} \left[\sinh U(t) - \frac{At}{2n_i} \exp - \frac{qV_A}{2kT} \right] dt \quad (V-19)$$

where $\gamma = \left[2n_i q^2 \exp \frac{qV_A}{2kT} \right] / (\kappa \epsilon_0 kT)$. The integration interval $[0, W]$ here extends into the n-type region, with the resulting solution $U(x)$ having odd symmetry about $x=0$.

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

Equation (V-19) was programmed for solution using an IBM 370/65 computer. The program was designed to iterate (V-19) according to the successive approximation scheme shown in (V-10) until the convergence condition (V-11) was met for a value of $\epsilon = 10^{-3}$. The integrals occurring in the iteration were evaluated using Simpsons rule with 141 mesh points in the interval $[0, W]$. Our first computer run was for equilibrium conditions ($V_A = 0$), with an integration interval of 1200 Å. No convergence was obtained, under these conditions, without the use of under-relaxation. Instead, the output between successive iterations was found to oscillate between the two potential distributions shown in Figure V-1. This unstable behaviour suggested the introduction of the under-relaxation procedure (V-17) into the calculations, and this was done. Using under-relaxation, in conjunction with the Picard iteration method for the same problem, convergence was obtained for a relaxation parameter $K > 1.5$. The most rapid convergence, requiring only twelve iterations ($n=12$), occurring at $K=2$. This convergence behaviour is shown in Figure V-2. The corresponding equilibrium potential $U(x)$ is shown in Figure V-3 together with the potential distribution predicted by depletion layer theory. The two curves agree at points near the metallurgical junction, as expected, and differ near $x=W$ by about kT/q volts. Our computer solution agrees well with the earlier results of Kennedy and O'Brien [11] who employed the finite difference method. It should be stressed that the solution using the above integral method took approximately 7 seconds of computer time at $K=2$, compared to substantially greater times for the finite difference approach.

We next examined the case of a 0.5 volt applied forward voltage using the same 1200 Å width. Again, for convergence, under-relaxation was required to obtain stability, and the values of K had to be increased. The number of iterations, n , needed

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

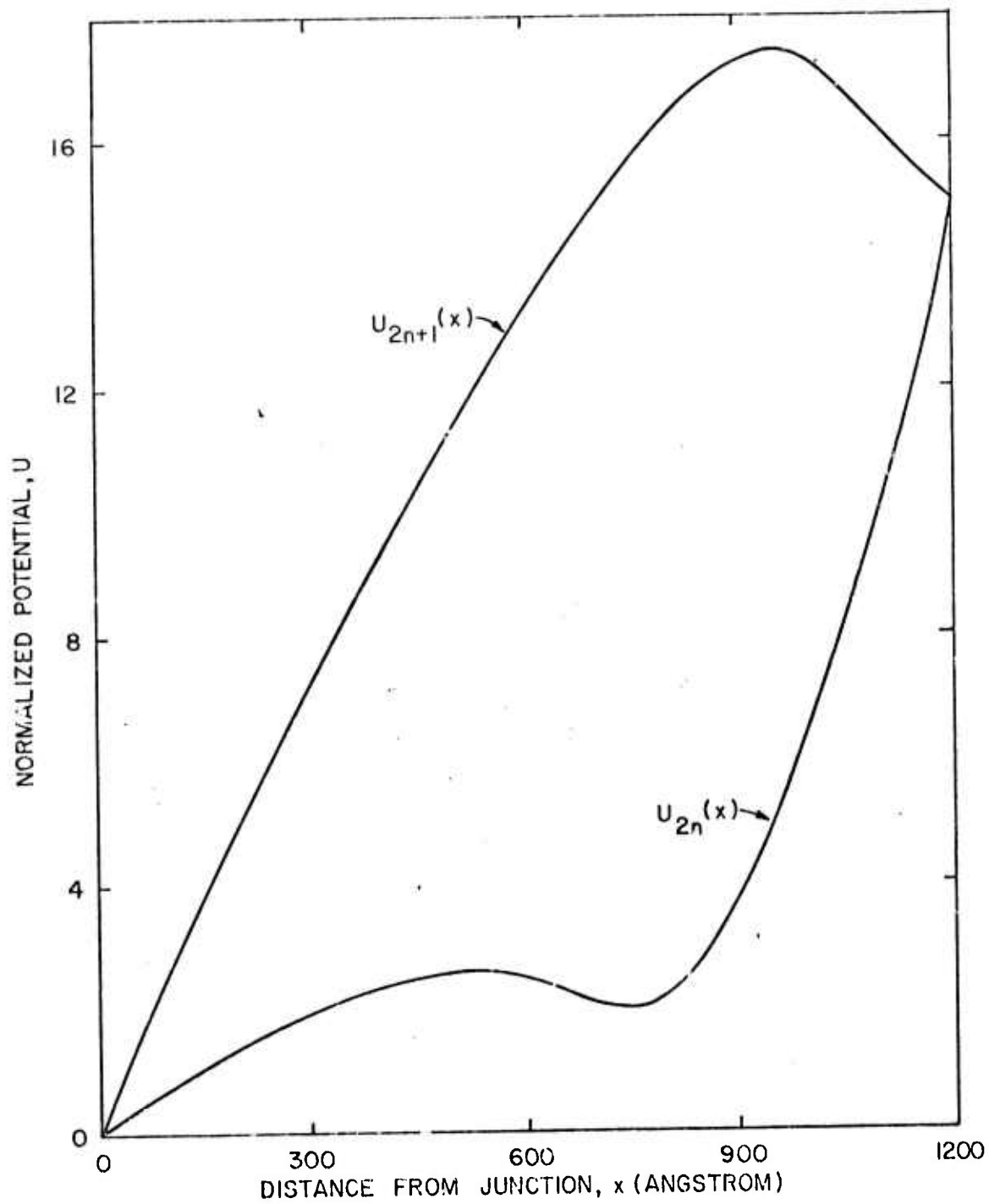


Fig. V-1. Computational instability in the absence of under-relaxation.

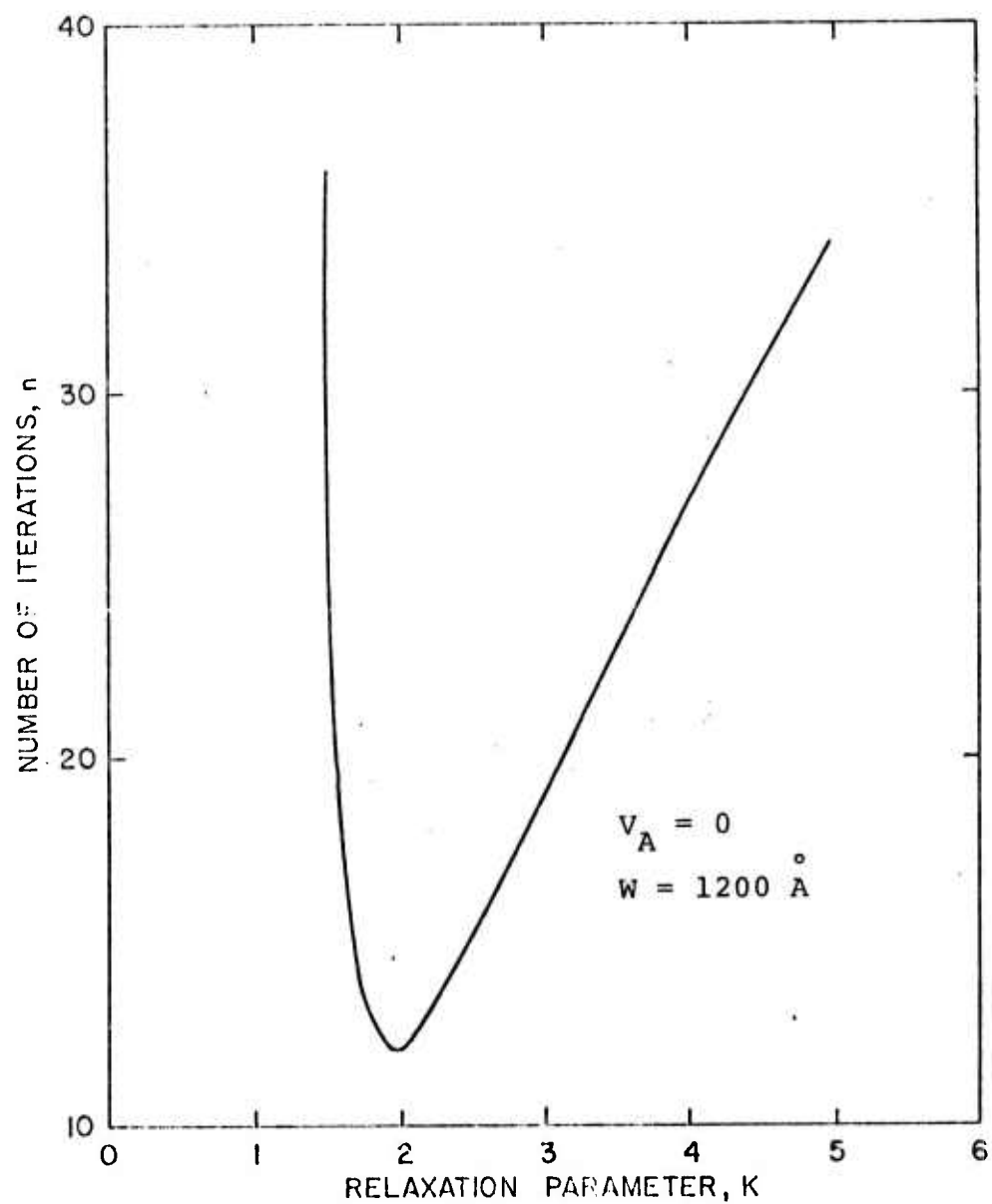


Fig. V-2. Number of Picard iterations vs. under-relaxation parameter.

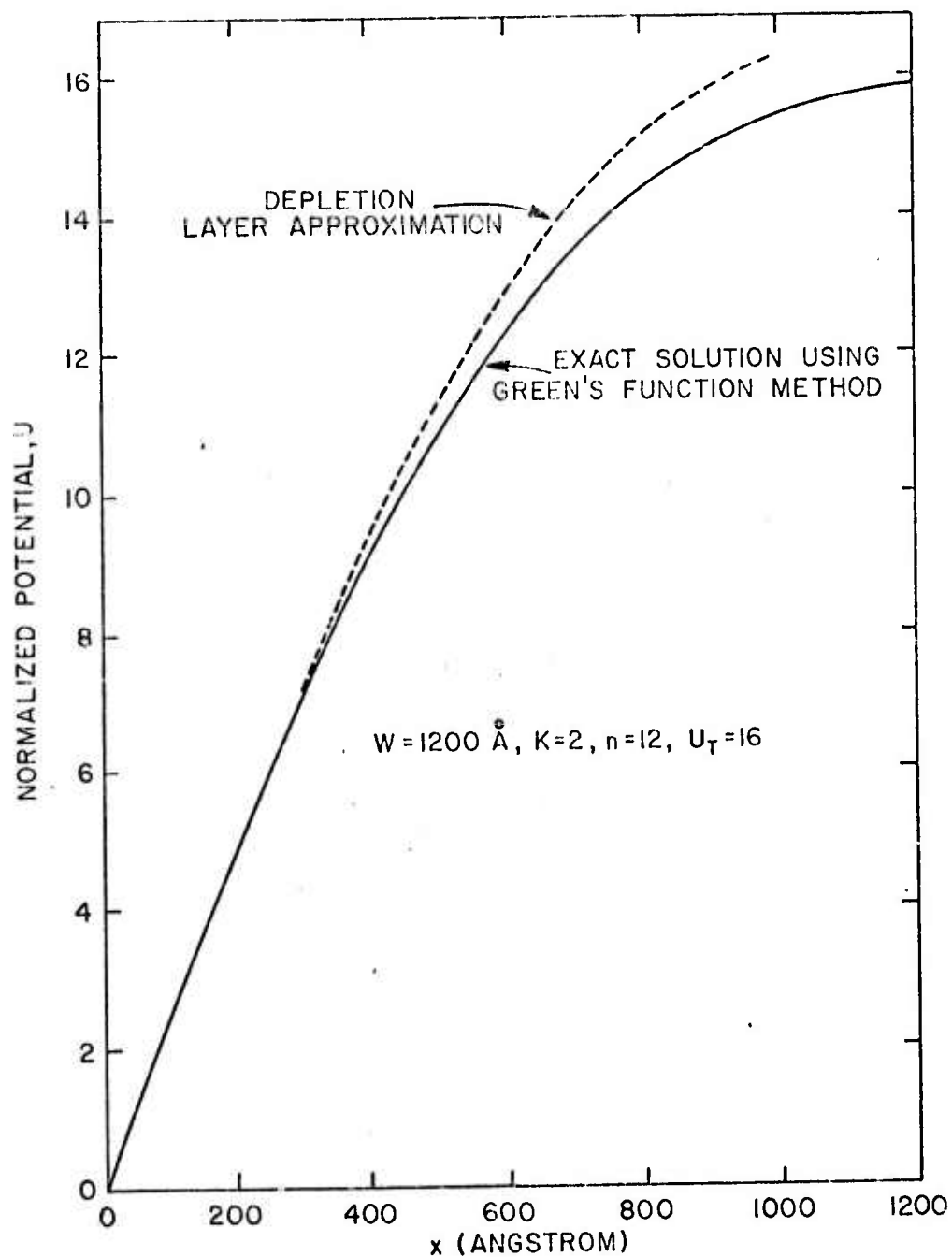


Fig. V-3. Solution of Poisson's equation using the Fredholm integral method.

for convergence at $\epsilon = 10^{-3}$, as a function of K , is shown in Figure V-4. From a comparison between the optimum value of the under-relaxation parameter in this figure with that obtained for equilibrium conditions, it is suggested that K may have to be determined for each new situation.

The calculated potential for this forward bias case is shown in Figure V-5. Our last computer run was done for zero applied bias, and for a wider integration range ($W=2500 \text{ \AA}$). This time, even with the use of under-relaxation for values as high as $K=75$, no convergence could be obtained. The calculated potentials, U_n , continually increased in magnitude, with each successive iteration, until the computer capabilities were exceeded (overflow). It is evident from this result that an improved relaxation procedure, such as used by Mock [2], is needed if convergence is to be obtained when very large violations of the stability criterion (V-16) can be expected. We are presently trying to suppress this observed numerical instability.

4.0 Direction of Future Work

It has been shown that the one-dimensional Poisson equation can be converted to a non-linear Fredholm equation and that the resultant equation can be solved iteratively using the Picard method in conjunction with an under-relaxation procedure. The solution method is found to be rapidly convergent for narrower integration ranges, but exhibits convergence problems for wider intervals. Our efforts at the moment (and in the immediate future) are directed toward improving the relaxation procedure used by allowing K in (V-17) to change with both x and with each iteration. In view of the success demonstrated by Mock [2] in the use of a changing sequence of relaxation parameter, this approach should lead to convergence over a wider range $0 < x < W$. Once having accomplished this, we plan to extend our calculations to more complex impurity profiles including the Gaussian and error function distributions associated with diffusion fabrication of semiconductor devices.

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

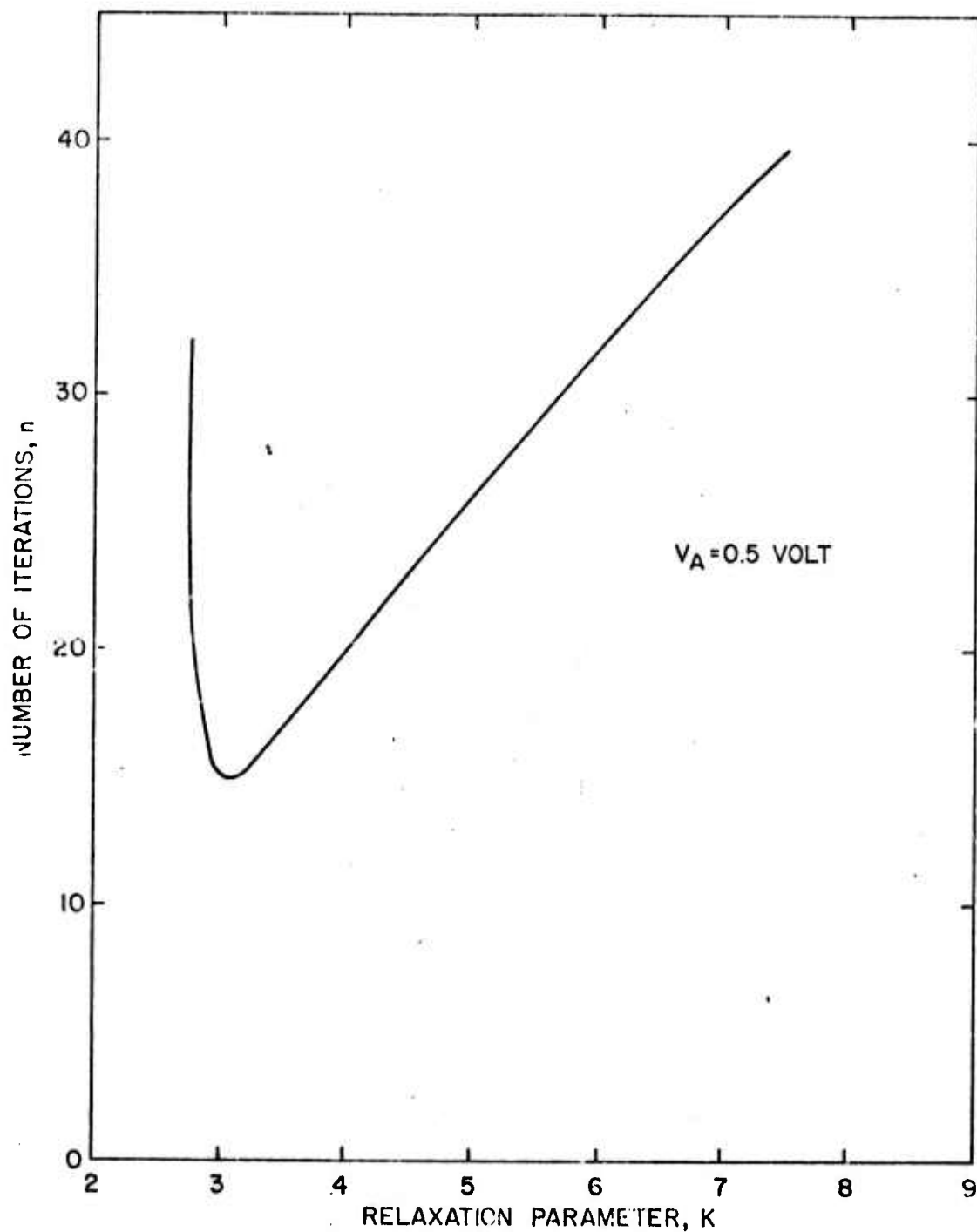


Fig. V-4. Number of Picard iterations vs. under-relaxation parameter for a forward biased, constant gradient, pn junction.

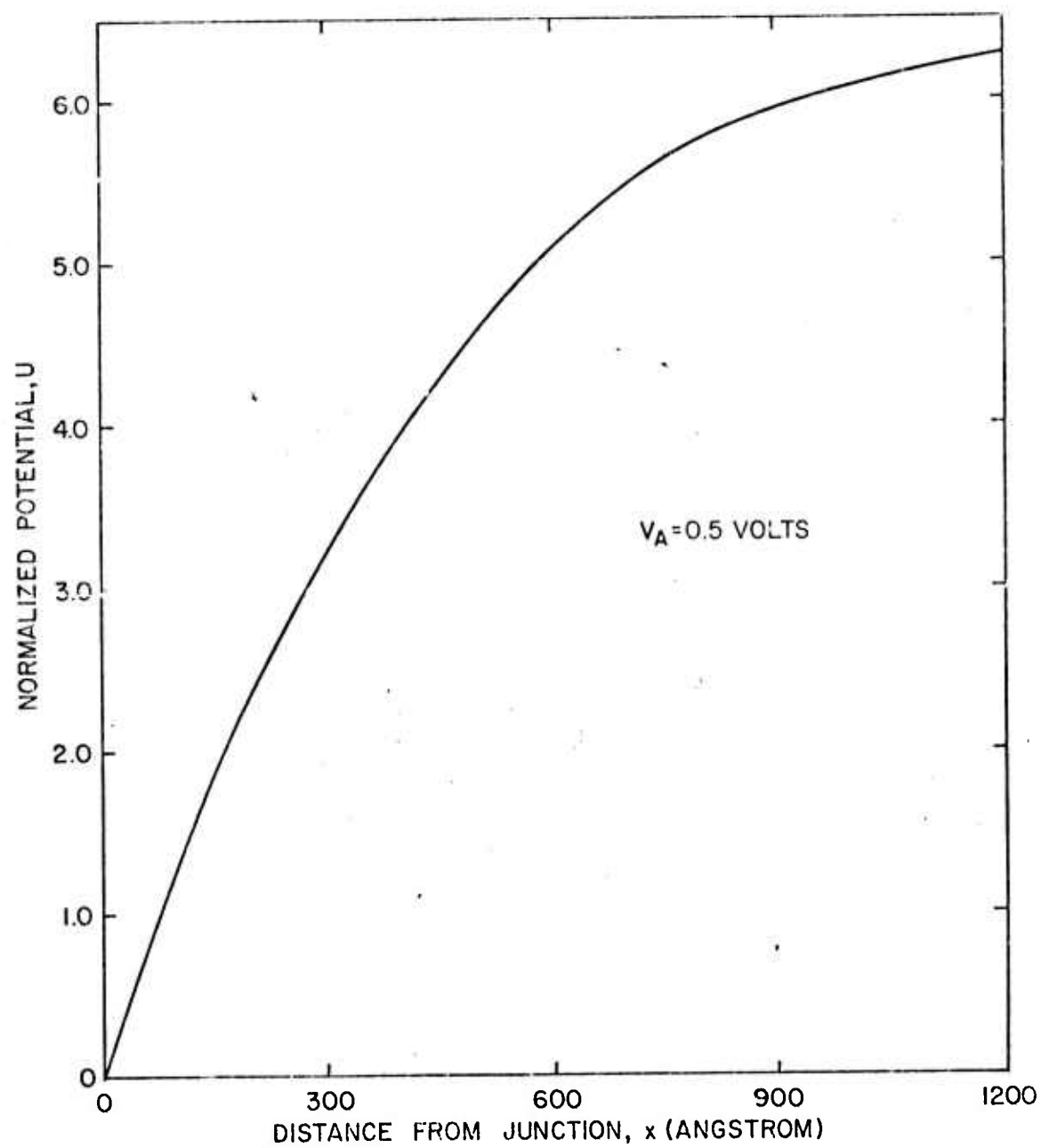


Fig. V-5. Calculated potential distribution in a forward biased, constant gradient, pn junction.

Some work has also been initiated on extending the above described integral approach to the two-dimensional geometry of a typical MOSFET. In this case the calculation of the Green's function is a formidable task; we encounter a more complicated geometry, and the introduction of mixed boundary conditions. Some preliminary studies conducted by us show that, even for a simple rectangular region subjected to homogeneous boundary conditions, the inversion of the Laplacian operator in (V-1) leads to a fairly cumbersome double infinite series form for the Green's function. It therefore appears that a better way to determine the Green's function for two-dimensional problems is to utilize numerical methods for its determination, in conjunction with tape or disc storage of this information. We plan to do this, and now have the required program (see Chapter 1).

Assuming that convergence problems associated with the Picard iteration can also be overcome for the two-dimensional case, where the iterative procedure also involves double integrals, we plan to use this method to examine the potential and carrier distributions in other two-dimensional devices such as bipolar transistors. It is envisioned that eventually adequate knowledge will be gained to solve equations (V-1) to (V-6) simultaneously, and with a substantial reduction of computer time over the finite difference method.

5.0 List of Symbols

κ	Relative dielectric constant of silicon (12.0)
ϵ_0	Permittivity of free space (8.854×10^{-14} Farad/cm)
k	Boltzmann constant
T	Temperature in degrees Kelvin
p	Hole concentration (cm^{-3})
n	Electron concentration (cm^{-3})
q	Charge on electron (1.6×10^{-19} coulombs)
n_i	Intrinsic carrier concentration ($1.4 \times 10^{10} \text{ cm}^{-3}$)
μ_p	Hole mobility ($\text{cm}^2/\text{volt-sec}$)

μ_n	Electron mobility ($\text{cm}^2/\text{volt-sec}$)
N_D	Donor concentration (cm^{-3})
N_A	Acceptor concentration (cm^{-3})
R	SRH recombination factor
U	Normalized potential ($U = qV/kT$)
U_T	Value of potential at $x=W$
W	Width of integration interval in $\text{\AA}$
V_A	Applied bias voltage in volts
K	Relaxation parameter
A	Grade constant (cm^{-4})

Chapter VI

Test Pattern Model Using Monte Carlo Methods

S. C. Taylor and D. P. Kennedy

	<u>Page</u>
1.0 Introduction.....	145
2.0 Approach to the Problem.....	148
3.0 Present Activities.....	150
4.0 Statistical Evaluation of the Harwell Generator...	151
5.0 Conclusions.....	158
6.0 Future Work.....	158

Chapter VI

Test Pattern Model Using Monte Carlo Methods

S. C. Taylor and D. P. Kennedy

1.0 Introduction

Sheet resistance measurements are used extensively for monitoring integrated circuit fabrication processes. This measurement technique offers many advantages: it is easily implemented, it is nondestructive, and sheet resistance test patterns can be measured (and tabulated) using high-speed automatic equipment. For this reason, a large segment of the semiconductor industry uses sheet resistance test patterns for production line monitoring and control.

Statistical studies have been undertaken on sheet resistance measurements derived from IC test sites. These studies show that a given type sheet resistance measurement, when viewed statistically on a given production line, produces a distribution, rather than one single value (see Fig. VI-1). Further, these studies also show that a large degree of correlation exists between the sheet resistance distribution and inadvertent processing variations in a given IC manufacturing process. As a consequence, contained within any statistical sample of sheet resistance measurements is process variability data. These data, in conjunction with other test pattern measurements, can be used in mathematical models of device operation to evaluate the manufacturability of a given IC design, assuming a specific manufacturing line.

The purpose of this modeling effort is to develop computational methods whereby this process variability data can be extracted from a sheet resistance frequency distribution. For illustrative purposes, Fig. VI-1 shows the frequency distribution of sheet resistance for a boron diffusion into silicon. Contained within Fig. VI-1 is the frequency distribution of boron surface concentration, Fig. VI-2, associated with this particular IC manufacturing line. Figures VI-1 and VI-2 were obtained using a mathematical model of the type presently under development.

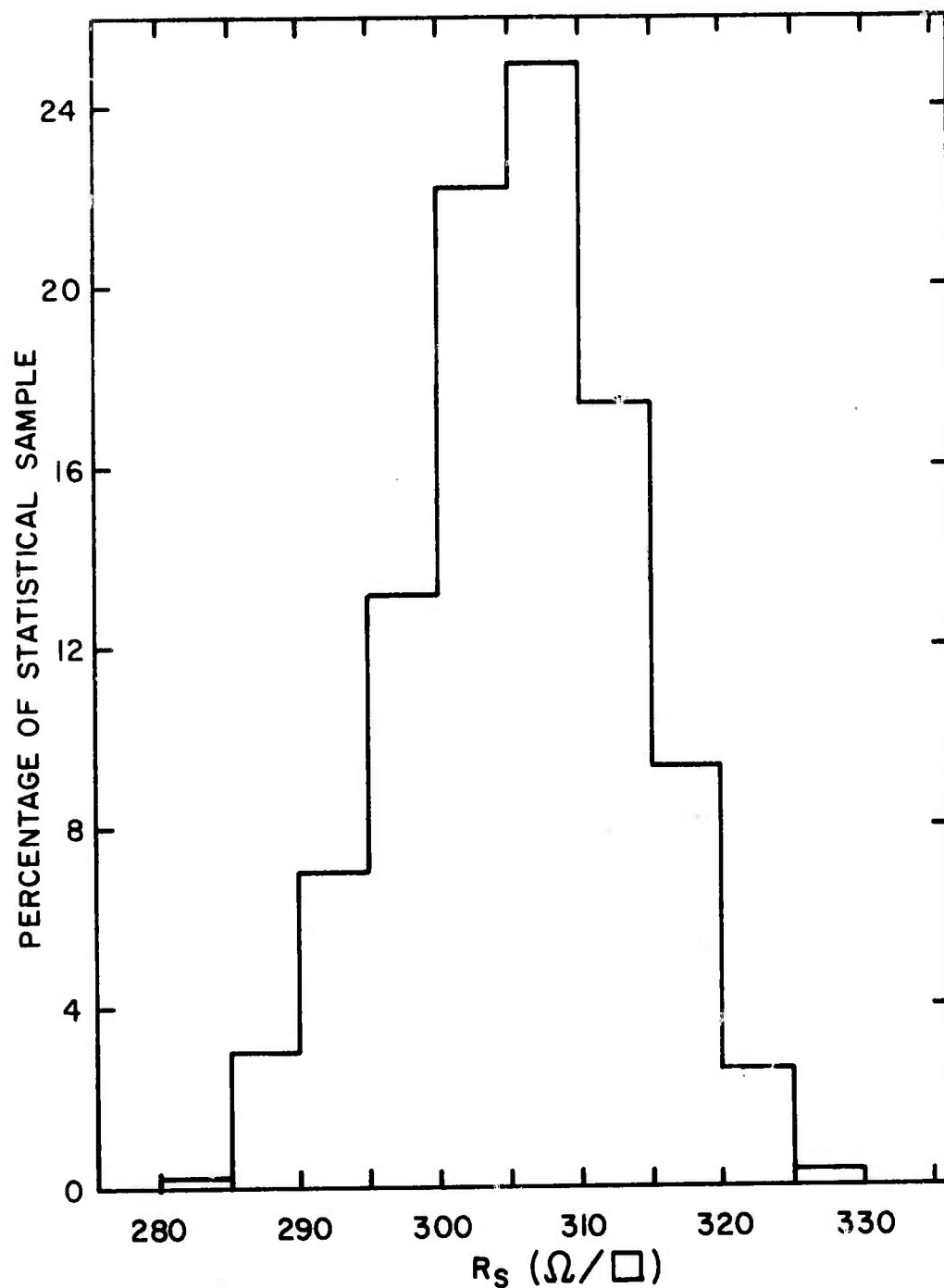


Fig. VI-1. Sheet resistance frequency distribution from a statistical sample of about 5000 test pattern measurements.

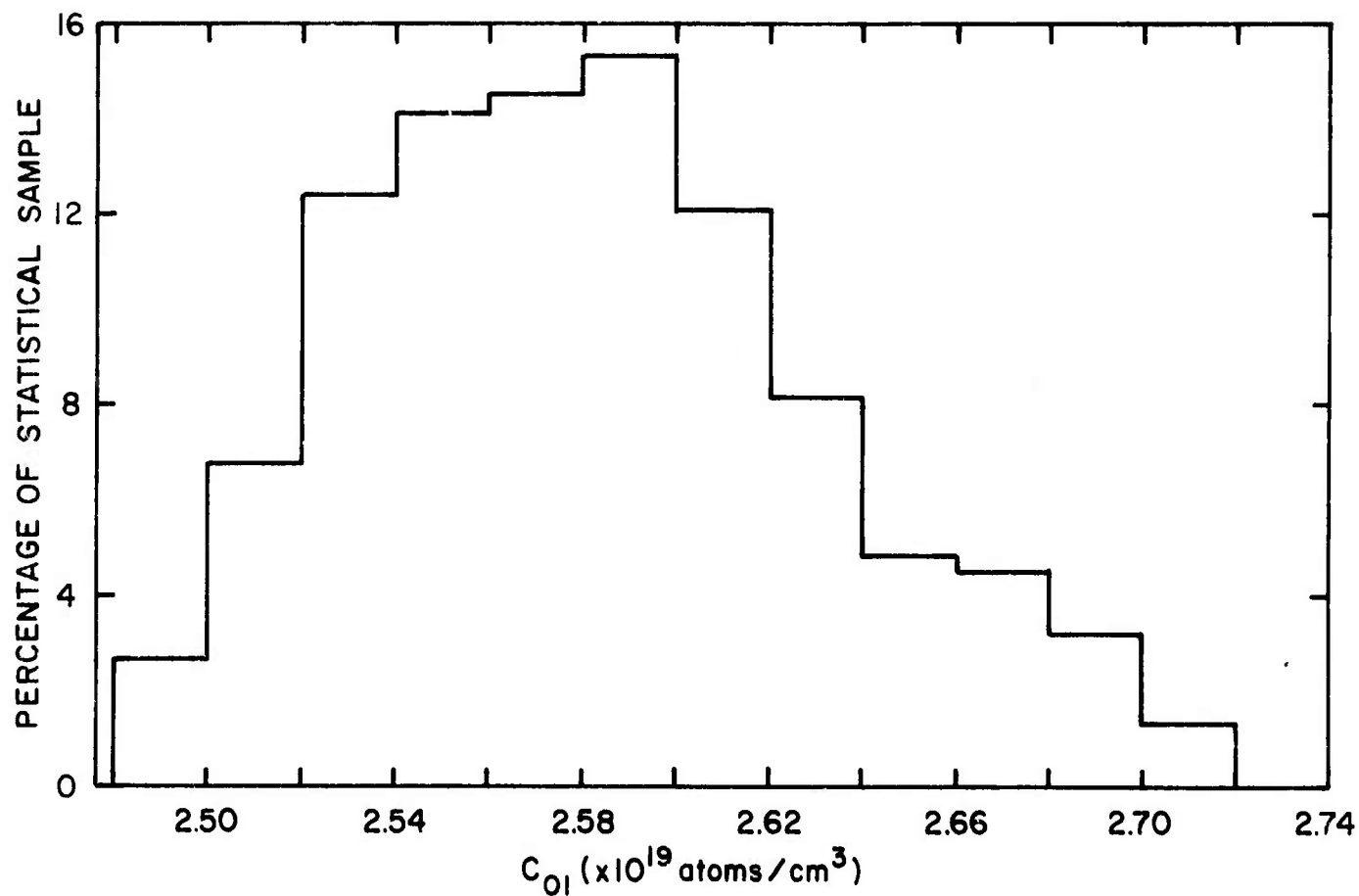


Fig. VI-2. Calculated frequency distribution for the boron surface concentration during a pre-dep. diffusion.

2.0 Approach to the Problem

The sheet resistance test-site model represents a simple numerical calculation of sheet resistance for a given impurity atom profile. This calculation must be accomplished in a manner requiring very little computer time. Further, the overall model must contain two distinctly different mathematical representations for the impurity atom distribution: one to represent impurity atoms after an initial deposition step, and a second to represent these impurity atoms after drive-in (or oxidation).

This first mathematical representation is easily implemented. During an initial impurity atom deposition negligible silicon is lost during oxidation, and the overall impurity profile can be described by

$$N_1(x,t) = C_{01} \operatorname{Erfc} \left(\frac{x}{2\sqrt{Dt}} \right) - C_B(x) , \quad (\text{VI-1})$$

where $C_B(x)$ represents a background impurity atom distribution of opposite type.

The second profile representation is adequately approximated by the integral [1]

$$N_2(x,t_1,t_2) = \int_0^\infty G(x,t_2;x^1,t^1) N_1(x^1,t_1) dx^1 , \quad (\text{VI-2})$$

The term $G(x,t_2;x^1,t_1)$ in (VI-2) represents a Green's Function relation of the form

$$G(x,t_2;x^1,t_1) = \frac{1}{\sqrt{\pi D(t_2-t^1)}} \left\{ \exp \left[-\frac{(x-x^1)^2}{4D(t_2-t^1)} \right] + \exp \left[-\frac{(x+x^1)^2}{4D(t-t^1)} \right] \right\} , \quad (\text{VI-3})$$

and $N_1(x^1,t_1)$ is given by (VI-1).

[1] D. P. Kennedy, *Proc. IEEE*, 52, 5 (1964).

For sheet resistance measurements, these impurity profiles must be modified to account for the fraction of impurities that are substitutional and interstitial within a semiconductor lattice. In addition, other corrections are needed to establish the fraction of substitutional impurities expected to be ionized at the measurement temperature, and the redistribution of impurities due to solubility differences between SiO_2 and silicon. The resulting mobile carrier distribution $n(x)$ can thereafter be used to establish the sheet resistance

$$\rho_s = \frac{1}{\bar{\sigma} x_j} \quad (\text{VI-4})$$

where

$$\bar{\sigma} = \frac{1}{x_j} \int_0^{x_j} q\mu(N)n(x)dx \quad (\text{VI-5})$$

Included in (VI-5) is the impurity dependent mobility of electrons/holes in silicon [2].

Both of these mathematical representations contain terms traditionally assumed constant although, in practice, are not. For example, the surface concentration (C_{01}) in (VI-1) undergoes variations from one location to another on a given silicon wafer. Similarly, this parameter exhibits variations from one silicon wafer to the next, in a given diffusion run, and from one diffusion run to the next, using a given diffusion furnace. Parameter variations are also found between various diffusion furnaces in a given manufacturing line.

Such variability is statistical in nature, when viewed in terms of individual IC structures produced by a given manufacturing line. Correlation does, indeed, exist from one location to another on a single silicon wafer, and from one location to another within a single diffusion run. Nevertheless, this correlation is lost if we view C_{01} variations upon the completed

[2] Irving, *Bell System Tech. J.*, 41, 387 (1962).

product. Measurements upon an arbitrary selection of completed integrated circuits show little (or no) correlation for a given process parameter and, therefore, randomness represents a reasonable approximation, with a specified frequency distribution.

In this particular calculation we find the following typical process parameters that must be assigned both a median value and a frequency distribution

- C_{01} Impurity atom surface concentration
- C_b Background impurity concentration
- T_1 Temperature of deposition diffusion
- T_2 Temperature of drive-in (oxidation) diffusion
- t_1 Deposition diffusion time
- t_2 Drive-in (oxidation) diffusion time
- W_0 Oxide width from oxidation cycle

In our model we approximate these process parameters, and their variabilities, by a psuedo-random number generator containing adjustable statistical filters in the output. Each parameter in (VI-1) and (VI-2), subject to variability, is given a specified median value, and a specified frequency distribution. Thereafter, the generator yields sets of numbers that are statistically typical of the fabrication parameter process found in any arbitrarily selected sheet resistance test-site. Using this number generator, in conjunction with the forementioned impurity atom profile representations, thousands of sheet resistance calculations can be performed and, thereby, we can generate a sheet resistance distribution similar to Fig. VI-1.

3.0 Present Activities

To perform this statistical calculation we must first require a suitable psuedo-random number generator. This generator must produce a sequence of uniformly distributed random numbers in the interval (0,1). Thereafter, the uniform, or rectangular, output from this generator is statistically filtered to provide the necessary distribution of process parameters: the normal distribution with a given mean and standard deviation, or a log-normal distribution with a specified skewness.

Such a random number generator was acquired from the Harwell Subroutine Library (Program FA01AS). This generator uses the linear congruential method for generating a set of uniformly distributed random numbers. These numbers, $\langle U_n \rangle$, are obtained through an application of the recursion relation

$$U_{n+1} = (aX_n + c) \text{Mod } m \quad n \geq 0 ,$$

which represent a linear congruential sequence. The Harwell generator uses the following numbers in the above recursion relation:

$$\begin{aligned} a &= 3^{15} \\ c &= 0 \\ m &= 2^{32} \\ X_0 &= 21845 \end{aligned}$$

Our first task is to establish the randomness of this Harwell number generator and, thereby, verify its applicability for particular problems at hand.

4.0 Statistical Evaluation of the Harwell Generator

The field of statistics offers 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 [3]) to test the Harwell generator sequence:

- a. The Equidistribution (Frequency) Test
- b. The Gap Test

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

- 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:

$$V = \sum_{s=1}^k \frac{(Y_s - np_s)^2}{np_s} \quad . \quad (VI-6)$$

Assuming k categories in (VI-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 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

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

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 $(U_1, U_2, U_3, \dots, U_n)$, and developing the distribution function $F_n(x)$ where

$$F_n(x) = \frac{\text{Quantity in sequence} \leq x}{n} \quad (\text{VI-7})$$

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:

$$K_n^+ = \text{Max} \{F_n(x) - F(x)\} \quad \text{where } -\infty < x < \infty \quad (\text{VI-8})$$

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

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:

$$U_n = U_1, U_2, \dots, U_n$$

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

$$\langle Y_n \rangle = Y_1, Y_2, Y_3, \dots, Y_n \quad (\text{VI-9})$$

and $Y_n = \lfloor dU_n \rfloor$, 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:

<u>Number Sequence</u>	<u>Value of (VI-6)</u>
1-1000	V = 75.58
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 (VI-7) and thereafter determining the magnitudes of K_n^+ and K_n^- (VI-8). This test was made for three sequences of numbers from the Harwell generator:

<u>Number Sequence</u>	<u>Max</u>	<u>Probability of exceeding</u>
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 [3].

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, $U_j, U_{j+1}, U_{j+2}, \dots, U_{j+r}$, in which U_{j+r} lies within the range $\alpha < U_{j+r} < \beta$ but the others do not. This sub-sequence U_{j+r} represents a gap of length r .

[3] Knuth, Donald E., *The Art of Computer Engineering, "Semi-numerical Algorithms,"* Addison-Wesley (1971).

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 $0, 1, 2, \dots, t-1$, and the number of gaps $\geq t$. Thereafter, the Chi-Square test is applied to this distribution of gap lengths, with suitable probabilities:

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

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

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 $(\beta-\alpha)$ "above the mean," $(.5, 1)$, and "below the mean," $(0, .5)$:

<u>Number Sequence</u>	<u>$(\beta-\alpha)$</u>	<u>Value of (VI-6)</u>	<u>Probability of Exceeding</u>
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 [3] 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 $\langle U \rangle$ 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:

[3] Knuth, Donald E., *The Art of Computer Engineering, "Semi-numerical Algorithms,"* Addison-Wesley (1971).

<u>Number Sequence</u>	<u>Value of (VI-6)</u>	<u>Probability of Exceeding</u>
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 [3] 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

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

and apply the Kolmogorov-Smirnov test to the resulting sequence $V_0, V_1, V_2 \dots V_{n-1}$ 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 $\max (U_1, U_2, U_3 \dots U_t) < x$ is the probability that $U_1 < x$ and $U_2 < x \dots$, and this is the product of each individual probability -- $x \cdot x \cdot x \dots = x^t$.

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

<u>Number Sequence</u>	<u>Max</u>	<u>Probability of Exceeding</u>
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%

[3] Knuth, Donald E., *The Art of Computer Engineering, "Semi-numerical Algorithms,"* Addison-Wesley (1971).

As before, it is suggested [3] 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 coefficient:

$$c = \frac{n(U_0U_1 + U_1U_2 + \dots + U_{n-2}U_{n-1} + U_{n-1}U_0) - (U_0 + U_1 + \dots + U_{n-1})^2}{n(U_0^2 + U_1^2 + \dots + U_{n-1}^2) - (U_0 + U_1 + \dots + U_{n-1})^2} \quad (\text{VI-10})$$

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 $\mu_n + 2\sigma$ [3] where

$$\mu_n = -\frac{1}{(n-1)} \quad \sigma_n = \frac{1}{n-1} \frac{n(n-3)}{n+1} \quad (\text{VI-11})$$

It has been found that when a normal distribution is assumed for (V-10), these are the values for (VI-11).

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

<u>Number Sequence</u>	<u>Values of (VI-11)</u>
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.

[3] Knuth, Donald E., *The Art of Computer Engineering, "Semi-numerical Algorithms,"* Addison-Wesley (1971).

5.0 Conclusions

The Harwell random number generator failed none of the above tests. In no way does this evaluation imply the numbers from this generator are truly random. Instead, implied here is that this generator is adequate for our purposes.

6.0 Future Work

Having completed evaluating this number generator, the next task is to obtain normal and log-normal distributions from its output. The present generator has a statistical filter for the normal distribution; this must be similarly tested, probably using a standard Chi-Square test. Thereafter, we can easily modify this normal distribution to obtain the needed log-normal distribution. It is emphasized that this number generator will also be used for other mathematical modeling tasks associated with this project.

After completing this number generator task, we will next undertake the development of a mathematical model for sheet resistance measurements. This is a relatively straightforward problem, and no technical difficulties are foreseen.

Chapter VII

Hot Carrier Mechanisms in Semiconductor Devices

J. T. Wang and D. P. Kennedy

	<u>Page</u>
1.0 Introduction.....	161
2.0 Approach to the Problem.....	164
3.0 Future Plans.....	168

Chapter VII

Hot Carrier Mechanisms in Semiconductor Devices

J. T. Wang and D. P. Kennedy

1.0 Introduction

The theory of semiconductor device operation [1-3] is based upon solutions of a system of differential equations first proposed by Van Roosbroeck [4] in his studies of mobile carrier transport in semiconductors. Contained within this system of equations is an expression for the electric current density produced by mobile electrons:

$$J_n = en(x)\mu_n \xi(x) + eD_n \frac{dn(x)}{dx} . \quad (\text{VII-1})$$

In (VII-1) it is assumed the total electric current, due to electrons, can be represented as the superposition of a drift term, and a diffusion term. Furthermore, Van Roosbroeck assumed the diffusivity of electrons (D_n) and their drift mobility (μ_n) are related by the Einstein relation [5],

$$\frac{D_n}{\mu_n} = \frac{k_o T_L}{e} . \quad (\text{VII-2})$$

In an early study of this subject, Shockley [6] proposed that (VII-1) would be applicable only when thermal equilibrium exists between mobile carriers and the semiconductor lattice, i.e., the carrier temperature T_e equals the lattice temperature T_L .

-
- [1] A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley and Sons, Inc., N.Y. (1967).
 - [2] S. K. Ghandi, *The Theory and Practice of Microelectronics*, John Wiley and Sons, Inc., N.Y. (1968).
 - [3] S. M. Sze, *Physics and Semiconductor Devices*, McGraw-Hill Book Co., N.Y. (1969).
 - [4] W. Van Roosbroeck, *Bell Sys. Tech. J.*, 29, 560 (1950).
 - [5] S. Wang, *Solid State Electronics*, McGraw-Hill Book Co., N.Y. (1966).
 - [6] W. Shockley, *Bell Syst. Tech. J.*, 30, 990 (1951).

Furthermore, he suggested that this equilibrium situation can be altered in the presence of a large electric field. During electric conduction, Shockley proposed that electrons gain energy from an externally applied source and, therefore, the carrier temperature T_e can differ from the lattice temperature T_L . Assuming acoustic intravalley scattering, Shockley suggested that these hot carriers lose energy to the semiconductor lattice and, thereby, the average drift velocity is no longer proportional to the electric field. Quantitative calculations based upon this proposed situation yielded a field dependent electron drift mobility ($\mu_n \propto \xi^{-1/2}$).

The first experimental verification of this theory was given by Ryder [7] in 1953. Ryder demonstrated that at large electric fields, the electric current in homogeneously doped semiconductor material did, indeed, exhibit non-ohmic behavior. In addition, Ryder experimentally demonstrated that $\mu_n \propto \xi^{-1/2}$ when the carrier velocity exceeded some critical value, v_{c1} , and at another critical velocity, v_{c2} , (where $v_{c1} < v_{c2}$) $\mu_n \propto \xi^{-1}$.

The success of Shockley's theory stimulated additional research in this area. Gunn [8] found he was unable to experimentally verify the relation $\mu \propto \xi^{-1/2}$ at small values of carrier velocity, although other workers did, indeed, verify Ryder's conclusions [9]. In addition, Morgan [10] suggested that the high current mobility, $\mu \propto \xi^{-1}$, for germanium could result through the mechanism of optical phonon scattering; later, Conwell [11] theoretically verified this concept. Thereafter Harrison [12], Long [13], and Duh [14] undertook a similar evaluation for silicon;

-
- [7] E. J. Ryder, *Phys. Rev.*, 90, 766 (1953).
 - [8] J. B. Gunn, *Progress in Semiconductors*, Vol. 2, p. 213, John Wiley & Sons, N.Y. (1957).
 - [9] A. C. Prior, *J. Phys. Chem.*, 12, 175 (1959).
 - [10] T. Morgan, *Bull. Am. Phys. Soc.*, 2, No. 2, 265 (1959).
 - [11] E. M. Conwell, *Phys. Chem. Solids*, 8, 234 (1959).
 - [12] W. A. Harrison, *Phys. Rev.*, 104, 1281 (1956).
 - [13] D. Long, *Phys. Rev.*, 120, 2044 (1960).
 - [14] C. Y. Duh and J. L. Moll, *Solid-State Elec.*, 11, 917 (1968).

it was concluded that Ryder's experiments could be explained through a combination of intervalley and intravalley scattering.

In general, the hot-carrier concept has qualitatively explained nonohmic behavior of electric conduction in homogeneously doped silicon, at large values of electric field. Following Ryder [7], many device designers [15-17] adopted his experimental results with little consideration of associated implications. A frequently used model for hot-carrier mechanisms is to assume that Ryder's critical velocities (v_{c1} , v_{c2}) are attained at specific values of electric field (ξ_{c1} , ξ_{c2}). Further, these models assume $\mu = \mu_0$ when $\xi < \xi_{c1}$, $\mu = \mu_0 (\xi_{c1}/\xi)^{1/2}$ when $\xi_{c1} < \xi < \xi_{c2}$, and $\mu = \mu_0 (\xi_{c2}/\xi)$ when $\xi_{c2} < \xi$. Although this model contains an obvious ambiguity in regions where drift and diffusion are both important ($D/\mu \neq k_0 T_L/e$), little consideration is given to problems produced by this model in regions of a device containing an inhomogeneous impurity atom distribution.

For example, it is generally accepted that large values of electric field (ξ) can reside in inhomogeneously doped silicon at equilibrium, and that this electric field can exceed ξ_{c1} . For such regions, the carrier temperature T_e must, from thermodynamic considerations, remain equal to the lattice temperature T_L . Despite this situation, the present device model predicts hot-electron mechanisms, and hence $\mu = f(\xi)$, at equilibrium. In fact, a theory for device operation is available [18] in which hot electron mechanisms are assumed within the charge neutral base region of a bipolar transistor.

-
- [7] E. J. Ryder, *Phys. Rev.*, 90, 766 (1953).
 - [15] G. C. Dacey, *Phys. Rev.*, 90, 759 (1953).
 - [16] G. C. Dacey and I. M. Ross, *Bell Syst. Tech. J.*, 1149 (1955).
 - [17] S. R. Hofstein and G. Warfield, *IEEE Trans. on Electron Devices*, 129, March (1965).
 - [18] L. K. Makeshivari and S. N. Ihanwar, *Int. J. Electronics*, Vol. 37, No. 3, 435 (1974).

A second approach to this modeling problem was suggested by Stratton [19-21], Goldberg [22], and Stokoe [23-24]. In this approach, it is assumed that large values of electric field do not alter the basic velocity distribution of mobile carriers (it remains Maxwellian). Instead, it is assumed these carriers can be described by the Maxwell velocity distribution at an effective electron temperature, T_e , where $T_e \neq T_L$. Through this assumption they eliminate a previously stated objection $D/\mu = k_o T/e$; in this model

$$\frac{n_n}{\mu_n} = \frac{k_o T_e}{e} \quad . \quad (VII-3)$$

Implied here is another unsatisfactory situation. If the electron distribution remains Maxwellian at a temperature T_e (where $T_e \neq T_L$), electron-electron scattering, alone, must predominate. This implication could be reasonable for regions containing a large electron density, but it is questionable to imply that electron-electron scattering predominates in all regions of a semiconductor device.

As a consequence of this situation, it is our aim to investigate hot-electron mechanisms in silicon, and to develop a model for this mechanism that is applicable to device analysis.

2.0 Approach to the Problem

In previous studies [25-28] researchers have investigated the validity of (VII-1) and (VII-2). A frequently used approach

-
- [19] R. Stratton, *J. Appl. Phys.*, 40, No. 11, 4582 (1969).
 - [20] R. Stratton, *J. Appl. Phys.*, 38, No. 12, 4536 (1967).
 - [21] R. Stratton, *Phys. Rev.*, 126, 2002 (1962).
 - [22] Colman Goldberg, *J. Appl. Phys.*, 40, No. 11, 4612 (1969).
 - [23] T. Y. Stokoe and J. E. Parrott, *Solid State Elec.*, 17, 477 (1974).
 - [24] T. Y. Stokoe and J. E. Parrott, *ibid.*, 18, 811 (1975).
 - [25] R. Stratton, *IEEE Trans. Elec. Dev.*, ED-19, No. 12 (1972).
 - [26] J. D. Gassaway, *IEEE Trans. Elec. Dev.*, ED-18, 175 (1971).
 - [27] A. H. Marshak and D. Assaf, *Solid State Elec.*, 16 (1973).
 - [28] B. R. Nag and A. N. Chakravarti, *ibid.*, 18, (1975).

to this problem is to expand the mobile carrier distribution function $f(\vec{r}, \vec{k})$ in a series of Legendre polynomials, and retain only the first two terms of the resulting series:

$$f(\vec{r}, \vec{k}) = f_0(\vec{r}, E) + kg(\vec{r}, E) \cos \phi \quad (\text{VII-4})$$

where $f_0(\vec{r}, E)$ is the spherically symmetrical part and $kg(\vec{r}, E) \cos \phi$ is the asymmetrical part (both in k -space), respectively. Thereafter, (VII-4) is substituted into the steady-state Boltzmann transport equation, and $g(\vec{r}, E)$ is established in terms of $f_0(\vec{r}, E)$. Because the drift component of electric current arises from the generation of asymmetry in $f(\vec{r}, k)$, a quantitative evaluation of $g(\vec{r}, E)$ provides a means to calculate the drift component of electric current, due to an assumed electric field.

Using this approach, it can be shown [25] that (VII-1) is applicable only when the symmetrical part of (VII-4) can be written as the product of a position dependent carrier density $n(x)$ and an energy dependent relation ($f_0(\vec{r}, E) = n(\vec{r})W(E)$). Further, it can also be shown that the Einstein relation (VI-2) is applicable only when $W(E)$ assumes the Maxwellian form

$$W(E) \propto \exp(-E/k_0 T) \quad , \quad (\text{VII-5})$$

assuming T is constant.

Our approach to this problem is to utilize the same simplified polynomial expansion (VII-4) for $f(\vec{r}, \vec{k})$, without introducing the other, traditional, simplifying assumptions. By retaining (VII-4) we require that a large electric field in silicon produces asymmetry in $f(\vec{r}, \vec{k})$, although this asymmetry must be small. Implied here is that the associated carrier collision mechanisms produce a large degree of velocity randomization, when compared with mechanisms tending to reduce this randomness [29]. Such a simplification appears reasonable, except at exceedingly large values of electric field where, for practical reasons, few questions arise in device modeling.

[25] R. Stratton, *IEEE Trans. Elec. Dev.*, ED-19, No. 12 (1972).

[29] E. M. Conwell, *High Field Transport in Semiconductors*. In *Solid State Physics*, Supplement No. 9, Academic Press, N.Y. (1967).

Thereafter, substituting (VII-4) into the steady-state form of Boltzmann's transport equation,

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = -\frac{\hbar \vec{k}}{m} \cdot \text{grad}_{\vec{r}} \left[f(\vec{r}, \vec{k}) \right] + \frac{e \xi(x) \vec{1}}{\hbar} \cdot \text{grad}_{\vec{k}} \left[f(\vec{r}, \vec{k}) \right] \quad (\text{VII-6})$$

we obtain

$$\left| \frac{\partial f(\vec{r}, \vec{k})}{\partial t} \right|_{\text{coll}} = \frac{\hbar k}{m} \left[\frac{\partial f_o}{\partial x} - e \xi(x) \frac{\partial f_o}{\partial E} \right] \cos \phi$$

$$- \frac{2eE\xi(x)}{3\hbar} \frac{\partial g}{\partial E} + \frac{2E}{3\hbar} \frac{\partial g}{\partial x} - \frac{e\xi(x)}{\hbar} g \quad (\text{VII-7})$$

The orthogonal properties of the Legendre polynomial permits us to separate (VII-7) into two separate equations.

$$\left| \frac{\partial f_o}{\partial t} \right|_{\text{coll}} = \frac{2E}{3\hbar} \frac{\partial g}{\partial x} - \frac{e\xi(x)}{\hbar} g - \frac{2eE\xi(x)}{3\hbar} \frac{\partial g}{\partial E} \quad (\text{VII-8})$$

and

$$\left| \frac{\partial g}{\partial t} \right|_{\text{coll}} = \frac{\hbar}{m} \frac{\partial f_o}{\partial x} - \frac{e\hbar\xi(x)}{m} \frac{\partial f_o}{\partial E} \quad (\text{VII-9})$$

In (VII-8) and (VII-9), the collision process in silicon can be written in integral form:

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = -\frac{1}{(2\pi)^3} \iint \left[P(\vec{k}, \vec{k}^1) f(\vec{r}, \vec{k}) - P(\vec{k}^1, \vec{k}) f(\vec{r}, \vec{k}^1) \right] d^3 k^1 \quad (\text{VII-10})$$

where $P(\vec{k}, \vec{k}^1)$ and $P(\vec{k}^1, \vec{k})$ denote the transition probabilities from the $\vec{k}$ to $\vec{k}^1$ states and from $\vec{k}^1$ to $\vec{k}$ states, respectively. From Fermi's rule, the transition probability from $\vec{k}$ to $\vec{k}^1$ is given by

$$P(\vec{k}, \vec{k}^1) = \frac{2\pi}{\hbar} |\langle \vec{k}^1 | H^1 | \vec{k} \rangle|^2 \delta(E_{\vec{k}} - E_{\vec{k}^1} + I) \quad (\text{VII-11})$$

In (VII-11) the term $|\langle \vec{k}^1 | H^1 | \vec{k} \rangle|$ is the transition matrix element, $\delta(E)$ is the Dirac delta function, and I is the phonon interaction term. In this expression, $I = \hbar\omega_q$ for phonon absorption,

$I = -\hbar\omega_q$ for phonon emission, and $I = 0$ for no phonon interaction (for example, impurity scattering).

Substituting our approximation (VII-4) for $f(\vec{r}, \vec{k})$ into (VII-10) this collision integral can be separated into two collision integrals: one integral for the symmetrical part of our distribution function,

$$\left(\frac{\partial f_o}{\partial t}\right)_{\text{coll}} = \frac{1}{(2\pi)^3} \int \left[P(\vec{k}^1, k) f_o(\vec{r}, E^1) - P(\vec{k}, \vec{k}^1) f_o(\vec{r}, E) \right] d^3k^1 \quad (\text{VII-12})$$

and another for the asymmetrical part

$$\left(\frac{\partial [k_x g]}{\partial t}\right)_{\text{coll}} = \frac{1}{(2\pi)^3} \int \left[P(\vec{k}^1, \vec{k}) k_x^1 g(\vec{r}, E^1) - P(k, k^1) k_x g(\vec{r}, E) \right] d^3k^1. \quad (\text{VII-13})$$

Classical methods are available to evaluate (VII-12) and (VII-13) for the numerous scattering mechanisms encountered in silicon:

- (a) Acoustic Intravalley Scattering,
- (b) Intervalley Scattering,
- (c) Impurity Scattering,

Thereby, expressions can be derived for the left side of (VII-8) and (VII-9).

After completing this sequence of calculations, one obtains from (VII-8) and (VII-9), a second order partial differential equation in $\vec{r}$ and E and is of an elliptic type, and it provides a means to calculate both $f_o(\vec{r}, E)$ and $kg(\vec{r}, E)\cos\phi$ in (VII-4). If, indeed, these equations are to be solved without introducing traditional simplifying assumptions, numerical techniques must be applied, using a computer. From such a calculation we thereby obtain the carrier distribution function for a specified field distribution $\xi(x)$.

For an inhomogeneously doped semiconductor we cannot assume that the electric field $\xi(x)$ is a constant. As a consequence, the Boltzmann equation must be solved subject to the constraint $\partial J / \partial x = 0$ where

$$J = - \frac{eh}{m} \int k_x f(\vec{r}, \vec{k}) d^3k, \quad (\text{VII-14})$$

and from the resulting distribution function $f(\vec{r}, \vec{k})$ we can determine the drift and diffusion components,

$$J_{\text{diff}} = \frac{8\sqrt{2\pi m}^{1/2} e}{3\hbar^3} \left[\frac{\partial}{\partial x} \int (\tau E) f_0 E^{1/2} dE - \int \left(\frac{\partial \tau}{\partial x} E \right) f_0 E^{1/2} dE \right] \quad (\text{VII-15})$$

$$J_{\text{drift}} = - \left(\frac{8\sqrt{2\pi m}^{1/2} e}{3\hbar^3} \right) e \xi(x) \int E^{3/2} \tau \frac{\partial f_0}{\partial E} dE \quad (\text{VII-16})$$

In this type semiconductor material an applied biasing voltage, in conjunction with a modified carrier distribution function, will necessitate a calculation of the electrical distribution $\xi(x)$ to be used in the solution of the Boltzmann equations (VII-8) and (VII-9). This is readily obtained from a numerical solution of Poisson's equation

$$\frac{d^2 \psi}{dx^2} = - \frac{e}{\kappa \epsilon_0} \left\{ N_D(x) - n(x) \right\}, \quad (\text{VII-16})$$

where $N_D(x)$ is the impurity ion distribution and

$$n(x) = \int f_0(\vec{r}, \vec{k}) d^3k \quad (\text{VII-18})$$

represents the electron density distribution. Equation (VII-17) must be solved simultaneously with (VII-8) and (VII-9); this is to be accomplished by numerical means, on a computer.

3.0 Future Plans

Present research is directed toward a numerical solution of this boundary value problem. A computer program is being developed to simultaneously solve Boltzmann's equation and Poisson's equation for inhomogeneously doped semiconductor material.

Part II

Integrated Circuit Process Modeling

Stanford University
Stanford, California

Introduction - Part II

Integrated Circuit Process Modeling

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 accurately predicting 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 deposition, (3) epitaxy, and (4) thermal diffusion. This report describes the progress which has been made in the first six months of this program in all four areas.

In ion implantation, initial studies have been done in two areas. Effects of implantation damage on impurity profiles in annealed Si are being investigated. Here, the effect of the vacancies generated by this damage on diffusion of implanted ions

is being studied. We are studying the "knock on" phenomena which occurs when a high dose implant is done through a mask. So far, arsenic implantation through SiO_2 has been studied.

In thermal oxidation, initial effort is being directed to achieve an accurate analytic prediction of oxide thickness for an arbitrary oxidation sequence. The effects of oxidation ambient (dry O_2 or wet O_2), crystal orientation, substrate doping, and chlorine oxidation are included in this study.

In epitaxy, preliminary effort has been directed toward understanding the kinetics of growth and the mechanism of dopant incorporation in the epitaxial layer. Work has begun to generate a system transfer function for the epitaxial reactor to relate the input parameters such as gas flow, time, temperature, etc. to the resulting profile in the epitaxial layer.

In thermal diffusion, a new mathematical model for the diffusion of impurities in Si is being generated. This model will consider the influence of the internal electric field on the motion of impurity ions at elevated temperatures.

Stanford University

J. D. Meindl
K. C. Saraswat
J. D. Plummer
J. F. Gibbons
R. W. Dutton
T. I. Kamins

Louisiana State University

Alan Marshak

Fairchild Camera and Inst. Corp.

Bruce Deal

Chapter VIII
Ion Implantation

	<u>Page</u>
1.0 Introduction.....	175
2.0 Effects of Implantation Damage on Impurity..... Profiles in Annealed Si	175
2.1 Objective.....	175
2.2 Approach.....	176
2.3 The Diffusion Model.....	179
2.4 The Diffusion Solver.....	180
2.5 Optimization.....	185
3.0 Defects Arising as a Result of Implantation..... through the Mask (Knock-on Phenomena)	185
3.1 Introduction.....	185
3.2 Experimental Measurements of Knock-on..... Atoms	186
3.3 Range Distribution Calculations of..... Implanted and Knock-on Ions Based on Energy Distribution	192

Chapter VIII

Ion Implantation

1.0 Introduction

The major effort of the research in ion implantation during the first six months of this program has been concentrated in two areas.

- (a) Because of the interaction between the vacancies produced by damage and the implanted ions, complicated diffusion phenomena occur during the annealing, with the result that actual impurity profiles may bear no simple relation to the original implanted profile. Thus, the objective here is to study the interaction described above and produce an analytical model to predict the profiles of the annealed dopant.
- (b) A study is being done of high dosage implants through the mask, including both how the mask itself is affected and to what extent secondary ions (knock-ons) implanted from the mask into the semiconductor produce constraints on subsequent processing steps.

2.0 Effects of Implantation Damage on Impurity Profiles in Annealed Si

2.1 Objective. A major unsolved problem in the field of ion implantation is concerned with the general question of how the annealing of implantation damage in a crystal affects the diffusion of the implanted species. Briefly, the problem here 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. The project consists of two phases; one is the development of an appropriate computer program for the simulation of the diffusion processes that take place during annealing, and the other phase is the experimental verification of the assumptions used in the models, such as diffusion mechanisms, parameters, etc.

2.2 Approach. The block diagram in Fig. VIII-1 illustrates the building blocks in our approach for accomplishing the objectives. The inputs to the Diffusion Solver are: As-implanted profiles, damage profiles, process parameters, and diffusion parameters. Since many parameters are unknowns or not known accurately, an optimization of these parameters has to follow. Inputs to the optimization routine are the computed and experimental profiles. The optimization will be carried out in the diffusion parameter space minimizing a cost function (to be defined) over the parameter space and additional variables such as depth into the material and temperature for the isochronal annealing case. Figure VIII-2 illustrates the experimental data from an 35-minute isochronal anneal in the temperature range of 700 to 1100°C [1]. These data are represented by a family of curves on a fictitious experimental surface to be approximated by the calculated surface. The cost function is a weighted measure of the difference between the surfaces to be minimized over the unknown parameters by the Optimization Routine.

The areas of contribution will be the determination of diffusion parameters in an initial phase, the prediction of annealed profiles subsequently, and ultimately-by inclusion of the process parameters in the optimization space-the feasibility analysis of desired profiles for a specific device application will be possible.

Basic requirements are: generality of the diffusion solver and the "correct" diffusion model. At this point, it is pertinent

[1] W. K. Hofke, et al., *Applied Physics*, 2, 1973, pp. 265-278.

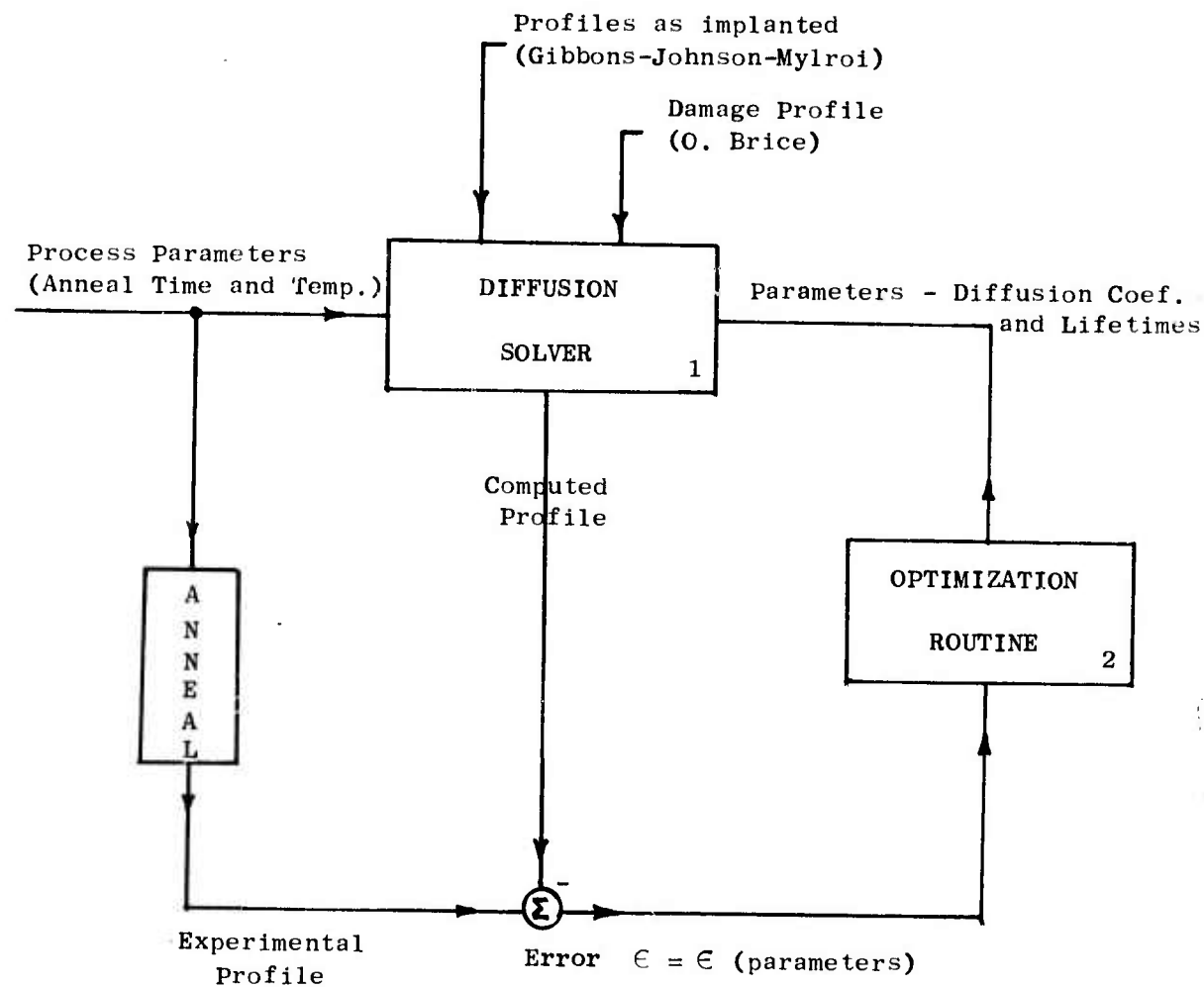


Figure VIII-1. Block diagram for study of implant annealing profiles in common semiconductor materials.

Duration: 35 min.

Temp.: {700 - 1100°C}

Dose: 10^{14} ion/cm²

Accel.: 70 KV

From: W. K. Hofke et al

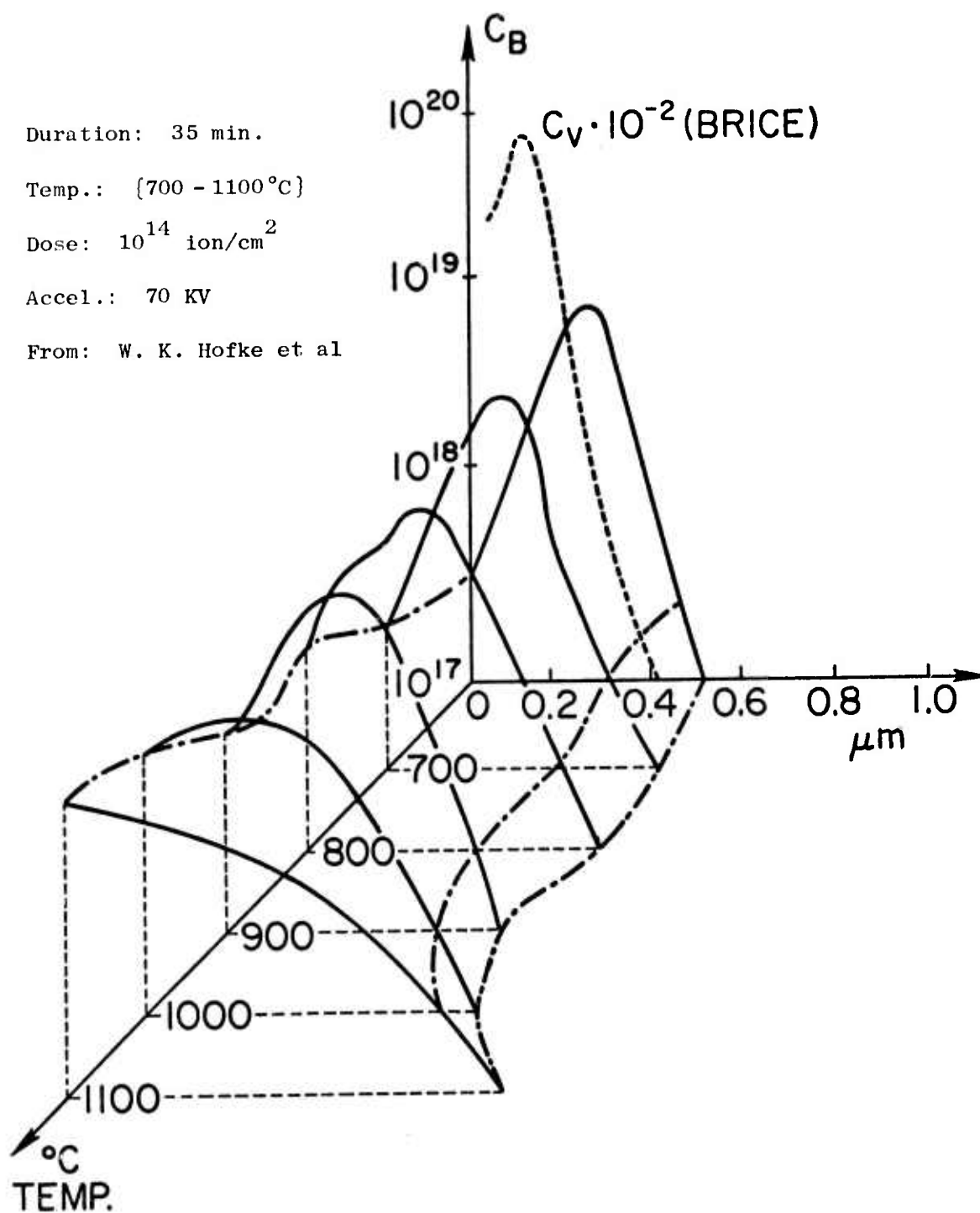


Figure VIII-2. Isochronal annealing of boron in silicon.

to state that the proof of correctness requires a diffusion solver and the comparison with experimental results which are precisely as shown in the block diagram we have outlined.

2.3 The Diffusion Model. Anderson and Gibbons have proposed a new model for the diffusion of boron in silicon [2]. This model has the merit of using concentration-independent diffusion coefficients for the species, namely, boron-substitutional and boron-vacancy pairs. This basic model will be used in the present work. However, in this study the diffusion equations for positive vacancies will be solved along with the boron and boron-vacancy pairs. This three-specie model is temporary; as we build up the block diagram, the neutral and negative vacancies will be incorporated. Hence, the model for boron in silicon in its general form will be formulated as follows:

The diffusing species, their concentrations, energy levels, and diffusion coefficients are:

boron substitutional	B	C_B	---	D_B
boron-vacancy pair	BV	C_{BV}	E_{BV}	D_{BV}
positive vacancy	V^+	C_{V^+}	E_{V^+}	D_{V^+}
neutral vacancy	V^0	C_V	E_V	D_V
negative vacancy	V^-	C_{V^-}	E_{V^-}	D_{V^-}
double-neg. vacancy	$V^{=}$	$C_{V^{=}}$	$E_{V^{=}}$	$D_{V^{=}}$

The reactions and concentrations at equilibrium are:

$$\begin{array}{lll}
 B^- + V^+ = B^-V^+ & C_{BV} = K_1 C_B C_{V^+} & K_1 = K_O \exp(E_{BV}/KT) \\
 V^0 = V^+ + e & C_{V^+} = C_{V^0} K_2 / n & K_2 / n = \exp(E_{V^+} - E_F) / KT \\
 V^0 + e = V^- & C_{V^-} = C_{V^0} K_3 n & K_3 n = \exp(E_F - E_{V^-}) / KT \\
 V^- + e = V^{=} & C_{V^{=}} = C_{V^-} K_4 n & K_4 n = \exp(E_F - E_{V^{=}}) / KT
 \end{array}$$

where E_F and n are the Fermi level and the electron concentration, respectively.

[2] J. R. Anderson and J. F. Gibbons, "A New Model for Boron Diffusion in Silicon" (to be published).

Departure from equilibrium produces a kinetic problem with generation and recombination of species. This is modeled by the following expression for the first reaction.

$$+ \left(\frac{K_1 C_B C_V - C_{BV}}{\tau} \right)$$

Notice that the first term in the sum is a nonlinear term, and the first order kinetic is modeled by a constant time constant.

The diffusion equations are a set of partial differential equations with nonlinear coupling through the generation and recombination terms.

$$\frac{\partial C_B}{\partial t} = D_B \frac{\partial^2 C_B}{\partial x^2} - K_O \frac{C_B C_{V+}}{\tau} + \frac{C_{BV}}{\tau}$$

$$\frac{\partial C_{V+}}{\partial t} = D_{V+} \frac{\partial^2 C_{V+}}{\partial x^2} - K_O \frac{C_B C_{V+}}{\tau} + \frac{C_{BV}}{\tau} - \frac{C_{V+}}{\tau_V}$$

$$\frac{\partial C_{BV}}{\partial t} = D_{BV} \frac{\partial^2 C_{BV}}{\partial x^2} + \underbrace{K_O \frac{C_B C_{V+}}{\tau}}_{\text{nonlinear coupling}} - \frac{C_{BV}}{\tau}$$

nonlinear
coupling

The recombination term C_V/τ_V accounts for recombination of vacancies with silicon interstitials or formation of vacancy loops.

2.4 The Diffusion Solver. Our approach is to use the transformation of the partial differential equations into a set of ordinary differential equations by partitioning the space coordinate. Proper ordering of species in space partitions yields a set of ordinary differential equations with a banded Jacobian, which can be solved very efficiently [3] since it does not require the storage of elements outside of the "band."

-
- [3] A. C. Hindmarsh, "Solution of Ordinary Differential Equations with Banded Jacobian," Lawrence Livermore Lab., University of California, Livermore, Calif., Mar. 1975.

$$\dot{y}_1 = f(y_1 \dots y_{99}) \quad 1 = (1 \dots 99)$$

TABLE VIII-1. DEFINITION OF THE VECTOR $\vec{y} = (y_1 \dots y_{99})$

Partition	1			2			3			1			...			33		
Specie	B	V	BV	B	V	BV	B	V	BV	B	V	BV	B	V	BV	B	V	BV
Concentration	$y_{(1)}$	$y_{(2)}$	$y_{(3)}$	$y_{(4)}$	$y_{(5)}$	$y_{(6)}$	$y_{(7)}$	$y_{(8)}$	$y_{(9)}$	$y_{(31-2)}$	$y_{(31-1)}$	$y_{(31)}$	$y_{(98)}$	$y_{(98)}$	$y_{(99)}$	$y_{(98)}$	$y_{(98)}$	$y_{(99)}$

For instance, for the first three equations in the diffusion model, we can use the Crank-Nicholson method [4] to approximate the space derivative as:

$$\frac{\partial^2 C}{\partial x^2} \approx \frac{C^l - 2C^c + C^r}{\Delta x^2}$$

where superscripts l, c, r designates partitions left, center, and right, respectively, and Δx is the partition width.

Ordering the species in each partition and taking 33 partitions, the set of three coupled partial differential equations becomes a set of 99 ordinary differential equations.

$$\dot{Y}_{(1)} = -2K_1 Y_{(1)} + K_1 Y_{(4)} - K Y_{(1)} Y_{(2)} + Y_{(3)}$$

$$\dot{Y}_{(2)} = (-2K_2 - K_V) Y_{(2)} + K_{(2)} Y_{(5)} - K Y_{(1)} Y_{(2)} + Y_{(3)}$$

$$\dot{Y}_{(3)} = (-2K_3 - 1) Y_{(3)} + K_{(3)} Y_{(6)} + K Y_{(1)} Y_{(2)}$$

$$i = (1, 2, \dots, 31)$$

$$\dot{Y}_{(3i+1)} = K_1 Y_{(3i-2)} - 2K_1 Y_{(3i+1)} + K_1 Y_{(3i+4)}$$

$$- K Y_{(3i+1)} Y_{(3i+2)} + Y_{(3i+3)}$$

$$\dot{Y}_{(3i+2)} = K_2 Y_{(3i-1)} - (2K_2 + K_V) Y_{(3i+2)} + K_2 Y_{(3i+5)}$$

$$- K Y_{(3i+1)} Y_{(3i+2)} + Y_{(3i+3)}$$

$$\dot{Y}_{(3i+3)} = K_3 Y_{(ei)} - (2K_3 + 1) Y_{(3i+3)} + K_3 Y_{(3i+6)}$$

$$- K Y_{(3i+1)} Y_{(3i+2)}$$

$$\dot{Y}_{(97)} = K_1 Y_{(94)} - 2K_1 Y_{(97)} - K_1 Y_{(97)} Y_{(98)} + Y_{(99)}$$

$$\dot{Y}_{(98)} = K_2 Y_{(95)} - (2K_2 + K_V) Y_{(98)} - K Y_{(97)} Y_{(98)} + Y_{(99)}$$

$$\dot{Y}_{(99)} = K_3 Y_{(96)} - (2K_3 + 1) Y_{(99)} + K Y_{(97)} Y_{(98)}$$

where

[4] Richtmeyer and Morton, *Difference Methods for Initial Value Problems*, Chapter 8.

$$K_1 = \frac{D_B^T}{\Delta x^2} \quad K - \text{Eq. Const.}$$

$$K_2 = \frac{D_V^T}{\Delta x^2} \quad K_V = \frac{\tau}{\tau_2}$$

$$K_3 = \frac{D_{BV}^T}{\Delta x^2}$$

The Jacobian of the set of ordinary differential equations $\dot{y}_i = f_i(y_1, y_2, \dots, y_N)$, $i(1 \dots N)$ has elements J_{ij} given by

$$J_{ij} = \frac{\partial f_i}{\partial y_j}$$

Hence:

$$J_{11} = -2K_1 - Ky(2) = \partial f_1 / \partial y_1$$

$$J_{12} = -Ky(1) = \partial f_1 / \partial y_2$$

$$J_{13} = 1 = \partial f_1 / \partial y_3$$

$$J_{14} = K_1 = \partial f_1 / \partial y_4$$

$$\begin{array}{c} \cdot \\ \cdot \\ \cdot \end{array} \quad \text{etc.}$$

And, in general, the nonzero elements of J are for $i = (0 \dots 32)$:

$$J(3i + 1, 3i + 1) = -2K_1 - Ky(3i + 2)$$

$$J(3i + 1, 3i + 2) = -Ky(3i + 1)$$

$$\longrightarrow J(3i + 1, 3i + 4) = \begin{cases} K_1 & i = 32 \\ 0 & i < 32 \end{cases}$$

$$J(3i + 1, 3i + 3) = 1$$

$$J(3i + 1, 3i - 2) = \begin{cases} 0 & i = 0 \\ K_1 & i > 0 \end{cases}$$

$$J(3i + 2, 3i + 1) = -K_y(3i + 2)$$

$$J(3i + 2, 3i + 2) = -2K_2 - K_y(3i + 1) - K_v$$

$$J(3i + 2, 3i + 3) = 1$$

$$J(3i + 2, 3i + 5) = \begin{cases} K_2 & i < 32 \\ 0 & i = 32 \end{cases}$$

$$J(3i + 2, 3i - 1) = \begin{cases} K_2 & 0 < i \\ 0 & i = 0 \end{cases}$$

$$J(3i + 3, 3i + 1) = K_y(3i + 2)$$

$$J(3i + 3, 3i + 2) = K_y(3i + 1)$$

$$J(3i + 3, 3i + 3) = -2K_3 - 1$$

$$J(3i + 3, 3i + 6) = \begin{cases} K_3 & i < 32 \\ 0 & i = 32 \end{cases}$$

$$J(3i + 3, 3i) = \begin{cases} K_3 & i > 0 \\ 0 & i = 0 \end{cases}$$

Schematically, the nonzero elements of the Jacobian are represented by "x" in Table VIII-2. The Banded Structure is apparent.

TABLE VIII-2. NONZERO ELEMENTS OF THE JACOBIAN .

y_1	y_2	y_3				y_{97}	y_{98}	y_{99}
x	x	x	x					
x	x	x		x				
x	x	x			x			
x			x	x	x	x		
	x		x	x	x		x	
		x	x	x	x			x
J =						...		
						x	x	x
						x	x	x
						x	x	x
						x		
						x	x	x
							x	x
							x	x

2.5 Optimization. A number of optimization routines not requiring derivatives have been located [5]. At this point, it is premature to compare their adequacy for solving our problem since the cost function has not yet been defined. We will defer this topic until the performance of some trial runs.

3.0 Defects Arising as a Result of Implantation through the Mask (Knock-on Phenomena)

3.1 Introduction. When an implantation is done through a thin mask such as SiO_2 or Si_3N_4 , the highly energetic ions of the

[5] M.J.D. Powell, *The Computer J.*, 7, July 1965, pp. 155-162.

W. J. Zangwill, *The Computer J.*, 10, 3, Nov. 1967.

Gill, Murray, and Pitfield, N.P.L., "Minimize a Real Function of N Real Variables Using a Quasi-Newton Method--Derivatives Not Required," (Subprogram in Stanford Library.).

main beam scatter. During this process, some of the mask atoms are given so much energy that they are dislodged from their lattice sites. If these recoiled ions have enough energy, they will have enough range to reach the silicon substrate. In the near vicinity of these knock-ons, the semiconductor quality will be very poor even after annealing. Thus, the objective of this work is to study this phenomenon and generate an analytical model of it. During the first six months, the work has been mainly concerned with the experimental verification of the knock-on phenomena in the case of arsenic implantation in Si through a mask of SiO_2 . A new method also has been developed to calculate the range distribution of incident ions as well as the knock-ons.

3.2 Experimental Measurements of Knock-on Atoms. The depth distribution of oxygen recoils 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 $217 \text{ \AA}$ for an As implant at 100 keV.

Moline et al [6] have shown that a recoil effect can be expected under these conditions. Moline and Cullis [7] have further demonstrated that recoil oxygen from an As implantation through an SiO_2 film causes residual lattice damage in Si even after annealing. Chu et al [8] found by channel experiments that implantation of 10^{16} cm^{-2} As through $480 \text{ \AA}$ SiO_2 results in about 4×10^{15} oxygen atoms in Si.

-
- [6] R. A. Moline, G. W. Reutlinger, and J. C. North, in *Atomic Collisions and Solids 1*, ed. by S. Datz, B. R. Appleton, and C. D. Moak, Plenum Press, New York, 1975.
 - [7] R. A. Moline and A. G. Cullis, *Appl. Phys. Lett.* 26, 551 (1975).
 - [8] W. K. Chu, H. Müller, and J. W. Mayer, *Appl. Phys. Lett.* 25, 297 (1974).

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 (Fig. VIII-3). The correlation of surface state density and oxygen concentration was established by implanting various doses of oxygen into test samples.

Fahrner and Goetzberger [9] have shown that ions implanted into an Si-SiO₂ interface produce characteristic surface state peaks that can be studied by MOS capacitance techniques. Fahrner [10] 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 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 $3.6 \times 10^{-3} \text{ cm}^2$ 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.

[9] W. Fahrner and A. Goetzberger, *Appl. Phys. Lett.* 21, 329 (1972).

[10] W. Fahrner, *J. Electrochem. Soc.*, 121, 784 (1974).

As IMPLANT 5×10^{13} , 100 KeV

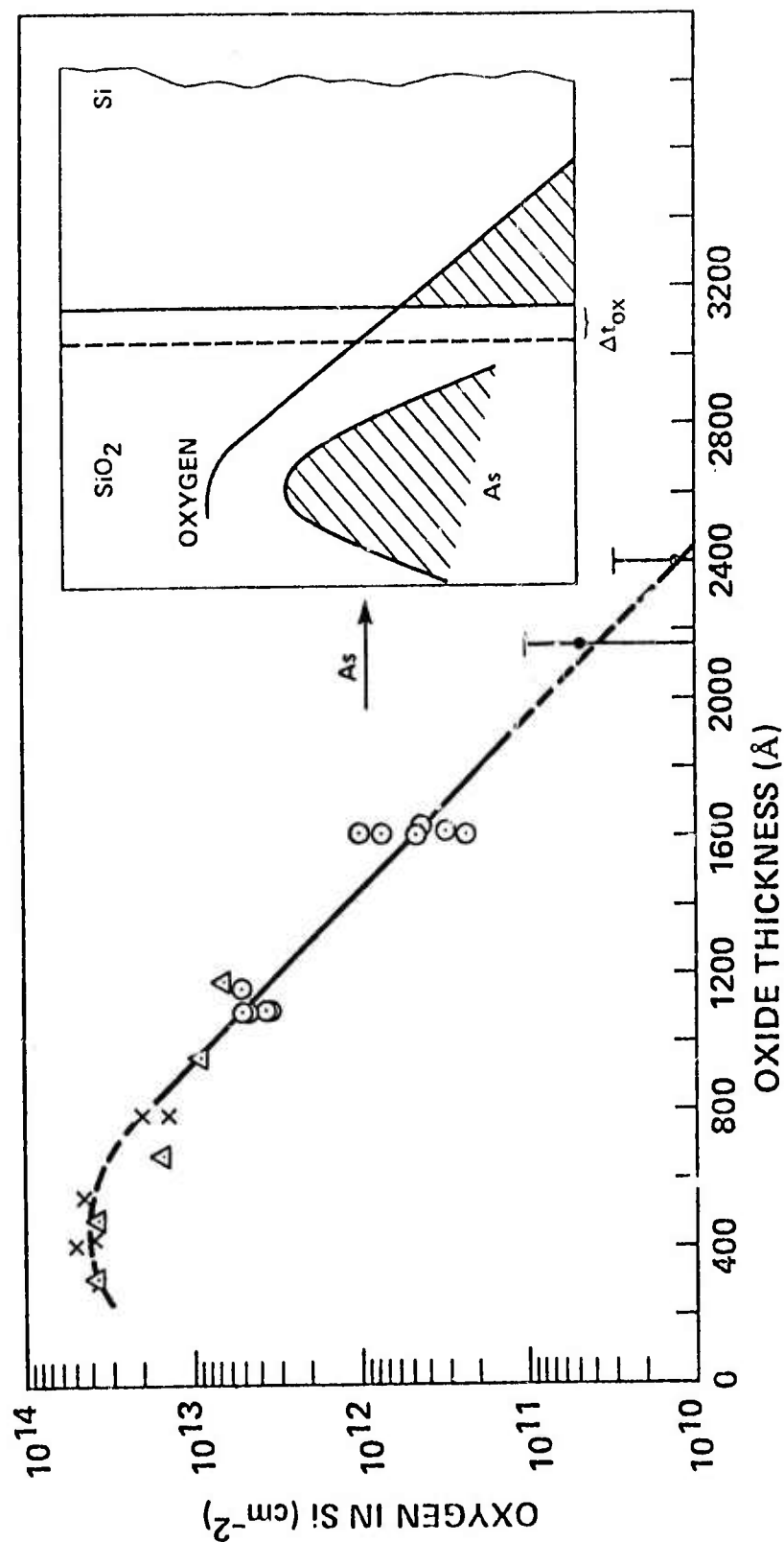


Figure VIII-3. Curve showing integral density of oxygen in silicon versus oxide thickness obtained from experiments for an implant dose of 5×10^{13} As cm⁻² at 100 keV. (Insert: sketch of method used for producing recoil oxygen in silicon.)

Electrical measurements of surface state density were carried out by employing the slow ramp technique [11,12] as in previous investigations [9,10]. 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 undertaken 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 $\text{cm}^{-2} \text{ eV}^{-1}$ 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 $764 \pm 10 \text{ \AA}$ (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 [13]. 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 VIII-4 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 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.

-
- [9] W. Fahrner and A. Goetzberger, *Appl. Phys. Lett.*, 21, 239 (1972).
 - [10] W. Fahrner, *J. Electrochem. Soc.*, 121, 784 (1974).
 - [11] R. Castagne, *Compt. Rend, Acad. Sci.*, Paris 267, Series B, 866 (1968).
 - [12] M. Kuhn, *Solid State Electronics*, 13, 873 (1970).
 - [13] J. F. Gibbons, W. S. Johnson, and S. W. Mylroie, *Projected Range Statistics, Semiconductors and Related Materials*, 2nd Ed., Halsted Press Division of John Wiley and Sons, Somerville, N.J.

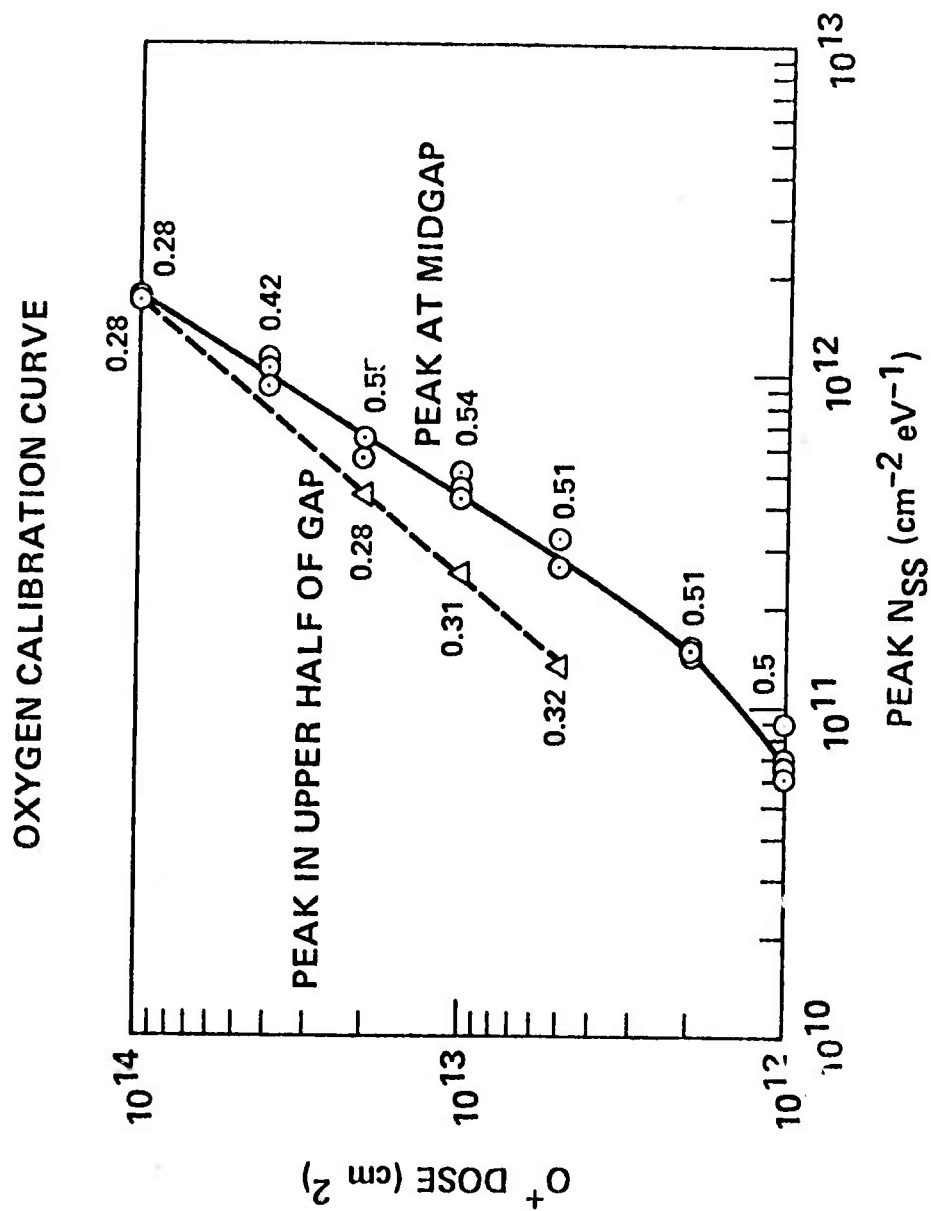


Figure VIII-4. Oxygen dose versus surface state density.

The dependence of energy of the two observed peaks on dose is indicated at each of the measured points in Fig. VIII-4. 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 Fig. VIII-4 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 SiO₂ was not expected to interfere because its recoil efficiency is much less.

The results of this experiment are shown in Fig. VIII-3. 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 of 100 keV. Recall, that although the data were obtained for oxygen in Si, Fig. VIII-3 represents the integral of the distribution of the recoil oxygens in SiO₂ 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 SiO₂. The exponential part gives $C(\text{Ox}) = 7.34 \times 10^6 \phi_{\text{As}} \exp(-x/217)$ 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 Fig. VIII-5, together with the distribution of 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 .

3.3 Range Distribution Calculations of Implanted and Knock-on Ions Based on Energy Distribution. In the Lindhard, Scharff, and Schiott (LSS) [14] theory, a statistical model has been used to arrive at integrals for moments of the range distribution. In most of the cases of interest, the first two moments of the range are calculated and the approximation of the Gaussian profile is assumed [13]. To obtain better accuracy (such as skewness) higher order moments have to be calculated [13]. Calculation of higher order moments is a tedious task. For implantation through a mask, the calculation of moments becomes an impossible task. Furthermore, calculations of the range distribution of the knock-ons becomes impossible using the LSS moment approach [15].

The purpose of this part of our report is to suggest a new technique to calculate the range distribution of the implanted ions as well as the knock-ons. This method is based on calculating the energy distribution of the ions in the substrate, whereas, in the LSS theory, the probability function corresponding to the energy distribution function disappears in the process to derive moments of the range. Because the energy distribution of incident ions can be considered as continuous at the boundaries of multiple layers in the target (in our case, a mask on silicon), the range distribution in each layer can easily be obtained if we calculate the energy distribution in the substrate.

[13] See page 189.

[14] J. Lindhart, M. Scharff, and H. E. Schiott, *Mat. Fys. Medd.*, 33, 14 (1963).

[15] S. Furukawa and H. Ishiwara, *J. Appl. Phys.*, 43, 3, Mar, 1972, pp. 1268-1273.

As IMPLANT 5×10^{13} , 100 KeV

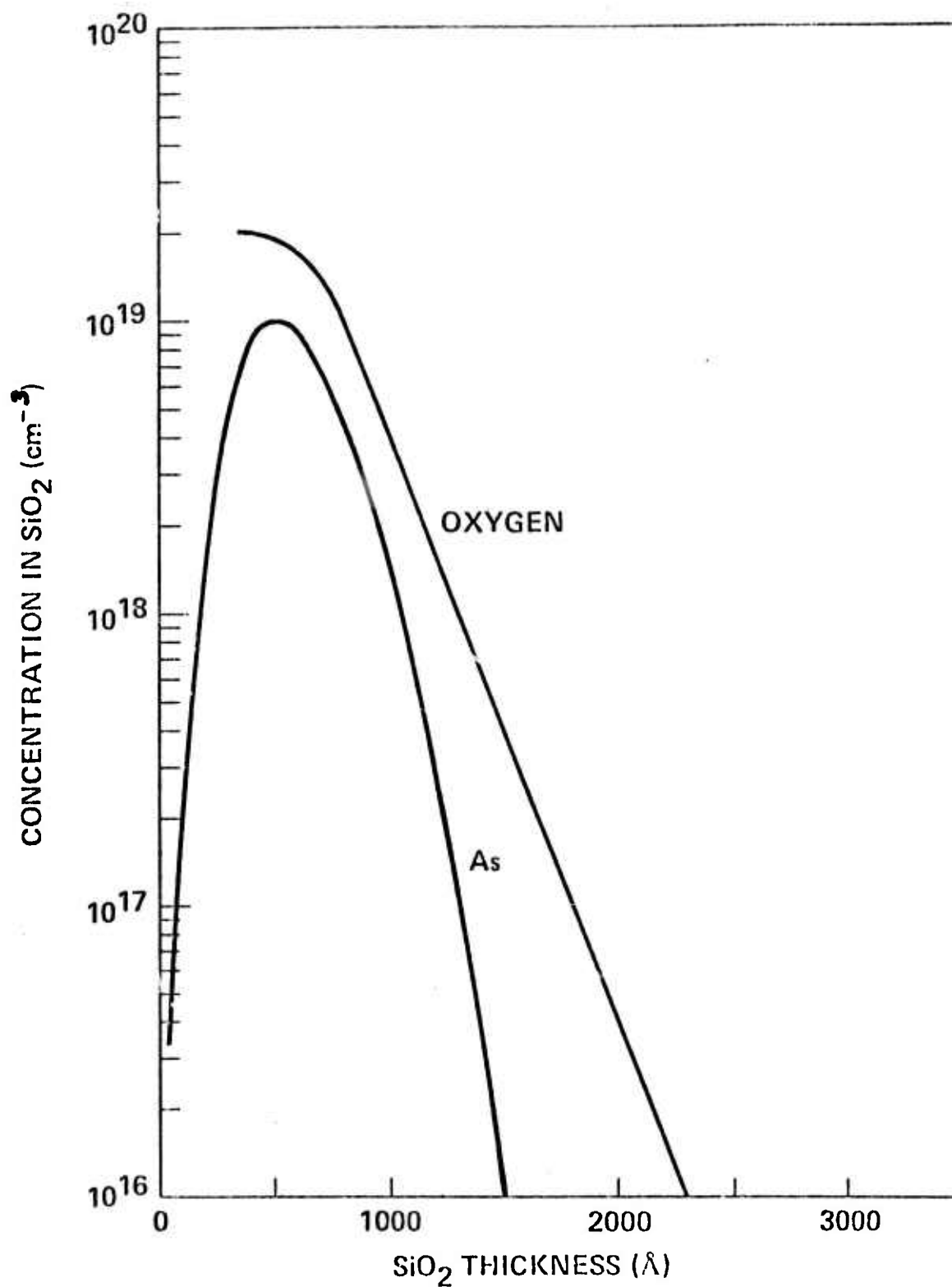


Figure VIII-5. Distribution of implanted As and recoil oxygen in SiO₂.

The substrate is divided into a large number of layers, each a few Å thick, as shown in Fig. VIII-6a. As the incident ion beam travels through each of these layers, its ions are scattered by the substrate atoms. A Boltzmann-type transport equation is used to simulate the scattering process. By knowing the energy loss parameters, the energy distribution in each layer can be numerically calculated. Because energy distribution of ions in each layer is statistically calculated, using this distribution, an ion concentration in terms of energy in each layer is determined. The principle of the technique is illustrated in Fig. VIII-6b. Where $Q(E,x)$ is the number of incident ions per unit energy interval plotted against the energy as a function of distance travelled into the substrate. Energy E_0 is considered to be so low that ions below this energy are assumed to be stopped. Thus, the range distribution can be calculated. While going from one type of medium to the other type (e.g., SiO_2 to Si), the only changes which must be made are the scattering parameters of the medium (such as stopping power, screening function, etc.). The power of this technique is realized when range distribution calculations of the knock-ons is considered. Using the LSS approach, it is impossible to do so. However, in our method, the calculations are based on a two body collision process, and it is easy to calculate, at each step, the energy imparted to the substrate atom by the incident ion. Once the energy of the knock-on atom is known, the range distribution can be calculated readily.

Preliminary calculations have been completed for the case of arsenic implanted in Si through SiO_2 . The model is being perfected at this point to include some second order effects so that the analytical predictions will match the experimental profiles.

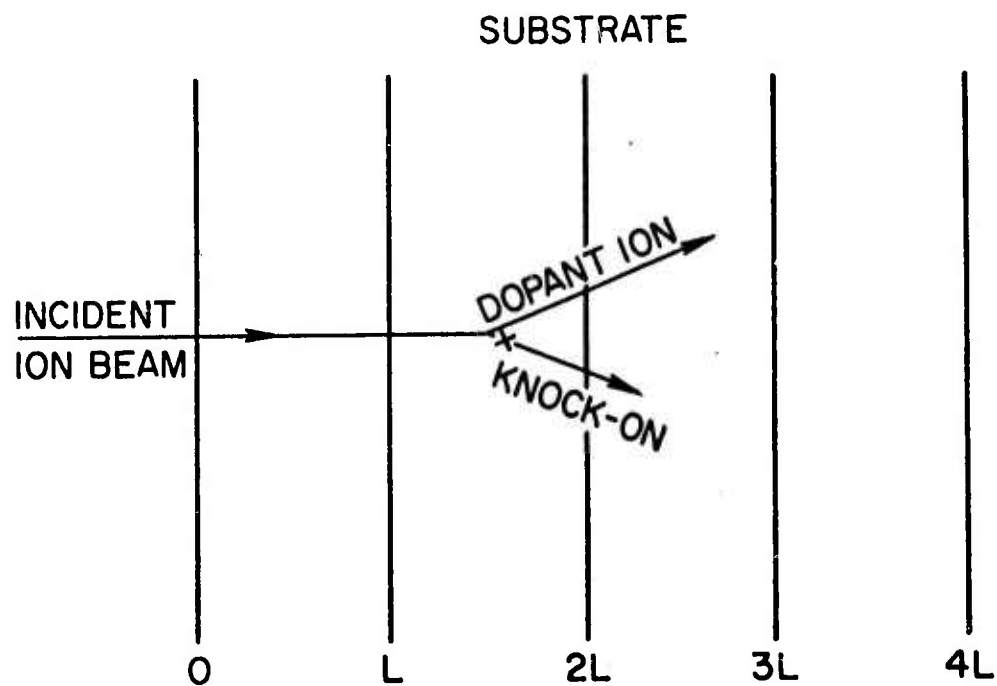


Figure VIII-6a. Division of substrate into thin layers to calculate energy distribution of the ions.

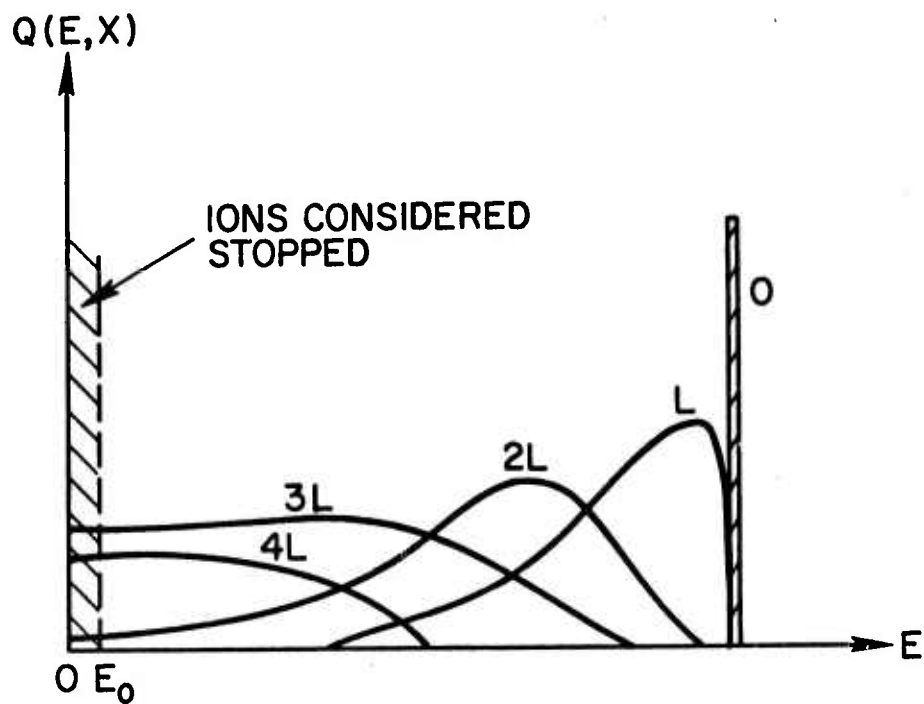


Figure VIII-6b. Total incident ion flux per unit energy as a function of distance traveled.

Chapter IX
Thermal Oxidation

	<u>Page</u>
1.0 Introduction.....	199
2.0 General Theory.....	199
3.0 Modifications to Account for Silicon Crystal..... Orientation	202
4.0 Modifications to Incorporate Impurity Doping..... Effects	206
5.0 Modifications to Incorporate HCl in the..... Oxidizing Ambient	208
6.0 Summary.....	210

Chapter IX

Thermal Oxidation*

1.0 Introduction

The principal goal of the first year of the program is to achieve accurate analytic prediction of oxide thickness for an arbitrary oxidation sequence or sequences. An equation will be derived which predicts oxide thickness and incorporates the following parameters to make it completely general.

- (a) Oxidation ambient (dry O₂, wet O₂)
- (b) Silicon crystal orientation
- (c) Impurity doping levels (phosphorus, arsenic, or boron) in the silicon substrate
- (d) Chlorine concentration (derived from an HCl source) present in the oxidizing ambient

At the present time, the general relationship for silicon oxidation incorporates only (a).

2.0 General Theory

The basic thermal oxidation model of Deal and Grove [1], based on the diffusion of the oxidant from the ambient through the existing oxide to react at the Si-SiO₂ interface, has been presented as:

$$X_O^2 + AX_O = B(t + \tau) \quad (\text{IX-1})$$

where

$$A = 2D_{\text{eff}}(1/k + 1/h)$$

$$B = 2D_{\text{eff}}C^*/N_1$$

$$\tau = (X_i^2 + AX_i)/B$$

-
- [1] B. E. Deal and A. S. Grove, *J. Appl. Phys.*, 36, Dec. 1965, pp. 3770-3778.

* This work represents a joint effort by the Stanford University Integrated Circuits Laboratory and Fairchild Camera and Instrument Corp. Research and Development Laboratory.

$$X_i = X_o \quad (t = 0)$$

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:

$$X_o = \frac{A}{2} \left[\left(1 + \frac{t + \tau}{A^2/4B} \right)^{1/2} - 1 \right] \quad \begin{array}{l} \text{time as input parameter} \\ \text{thickness as output} \end{array} \quad (\text{IX-2})$$

$$t = \frac{X_o^2 - X_i^2}{B} + \frac{X_o - X_i}{B/A} \quad \begin{array}{l} \text{desired thickness as input parameter} \\ \text{necessary time as output} \end{array} \quad (\text{IX-3})$$

For oxidation time t , long, relative to the characteristic time $A^2/4B$, the oxidation model reduces to a parabolic behavior $X_o^2 \approx Bt$ with B , therefore, termed the parabolic rate constant.

For oxidation t , short, relative to the characteristic time, a linear behavior $X_o \approx B/A[t + \tau]$ results, with $B/A = (kh/k + h) \cdot (C^*/N_1)$ 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 $h \gg k$). 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 $X_i =$

200 Å is found for dry O₂ oxidation (due to an initial phase of rapid oxidation by a different mechanism).

For moderately p-type boron doped ($C_B = 1.45 \times 10^{16} \text{ cm}^{-3}$) Si, with <111> surface orientation, it has been found experimentally [1] that:

$$B(T,p) [\mu^2/\text{hr}] = C_1 p e^{-E_1/kT} \quad \text{for } p[\text{atmospheres}], T[\text{K}^\circ] \quad (\text{IX-4})$$

$$\begin{aligned} \text{Dry O}_2: \quad E_1 &\approx 28.5 \text{ kcal/mole} = 1.23 \text{ eV/molecule} \\ C_1 &\approx 7.72 \times 10^2 \mu^2/\text{hr} \end{aligned}$$

$$\begin{aligned} \text{Wet O}_2: \quad E_1 &\approx 16.3 \text{ kcal/mole} = 0.71 \text{ eV/molecule} \\ C_1 &\approx 2.14 \times 10^2 \mu^2/\text{hr} \end{aligned}$$

$$A(T) [\mu] = C_2 e^{+E_2/kT} \quad \text{for } T[\text{K}^\circ] \quad (\text{IX-5})$$

$$\begin{aligned} \text{Dry O}_2: \quad E_2 &\approx 17.5 \text{ kcal/mole} = 0.76 \text{ eV/molecule} \\ C_2 &= 1.24 \times 10^{-4} \mu \end{aligned}$$

$$\begin{aligned} \text{Wet O}_2: \quad E_2 &\approx 29.0 \text{ kcal/mole} = 1.26 \text{ eV/molecule} \\ C_2 &\approx 2.39 \times 10^{-6} \mu \end{aligned}$$

It is expected that such factors as Si surface orientation, percentage HCl, and heavy impurity concentration level will be found to be parameters affecting the effective energies E_1 and E_2 and/or constants C_1 and C_2 through their effects on the diffusion process, surface-reaction rate, oxidant concentration, etc.

Experimental determination of the constants B and A is made easier by rewriting (IX-1) in the following form.

$$X_O = B \left(\frac{t + \tau}{X_O} \right) - A \quad (\text{IX-6})$$

Plotting X_O versus $(t + \tau)/X_O$ yields a straight line with slope B and intercept -A. Such an analysis therefore allows one to obtain both the parabolic (B) and linear (B/A) rate constants from one plot.

[1] B. E. Deal and A. S. Grove, *J. Appl. Phys.*, 36, Dec. 1965, pp. 3770-3778.

As indicated above, an initial increased oxidation rate is observed during the dry oxidation of silicon. To obtain the correction factor τ , which accounts for this initial oxidation, a linear plot can be made of oxide thickness versus time, and the curve extrapolated back to zero oxide thickness. Such a plot will generate only an approximate value for τ since the extrapolation is performed on a line with changing slope. Thus, in the analysis of experimental data using (IX-6), little confidence can be placed on values of $t+\tau$ in which the value of τ is more than 50% of t .

The basic oxidation relation has been implemented in form (IX-2) as a FORTRAN program on the HP-2100 system. At present, it assumes as constants the energies and constant coefficients found (for the specific experimental conditions noted) above. It accepts as input the desired ambient (dry or wet O_2), temperature (in $^{\circ}C$), pressure (in torr), and time (in minutes), and produces as output the oxide thickness (in Angstroms). The user may test either single oxidation steps or sequential (e.g., dry-wet-dry) oxidation cycles. Testing of the program against new experimental data and modifications for the additional implementation in form (IX-3) to give required oxidation time as output are being pursued at present.

3.0 Modifications to Account for Silicon Crystal Orientation

To date, the effect of silicon orientation on silicon oxidation kinetics has not been extensively investigated. It is known that this effect becomes significant at temperatures below $1100^{\circ}C$, but no data have been reported with regards to the magnitude of this effect. Thus, in the following experiments, $\langle 100 \rangle$ and $\langle 111 \rangle$ oriented silicon were used in each kinetic run.

A log-log plot of thickness versus time for the thermal oxidation of $\langle 100 \rangle$ and $\langle 111 \rangle$ 4-6 Ω -cm n-type silicon in dry oxygen at $1000^{\circ}C$ is shown in Fig. IX-1. It can be seen from this plot that 10 to 20% differences in the oxidation rate of the two orientations exist at $1000^{\circ}C$.

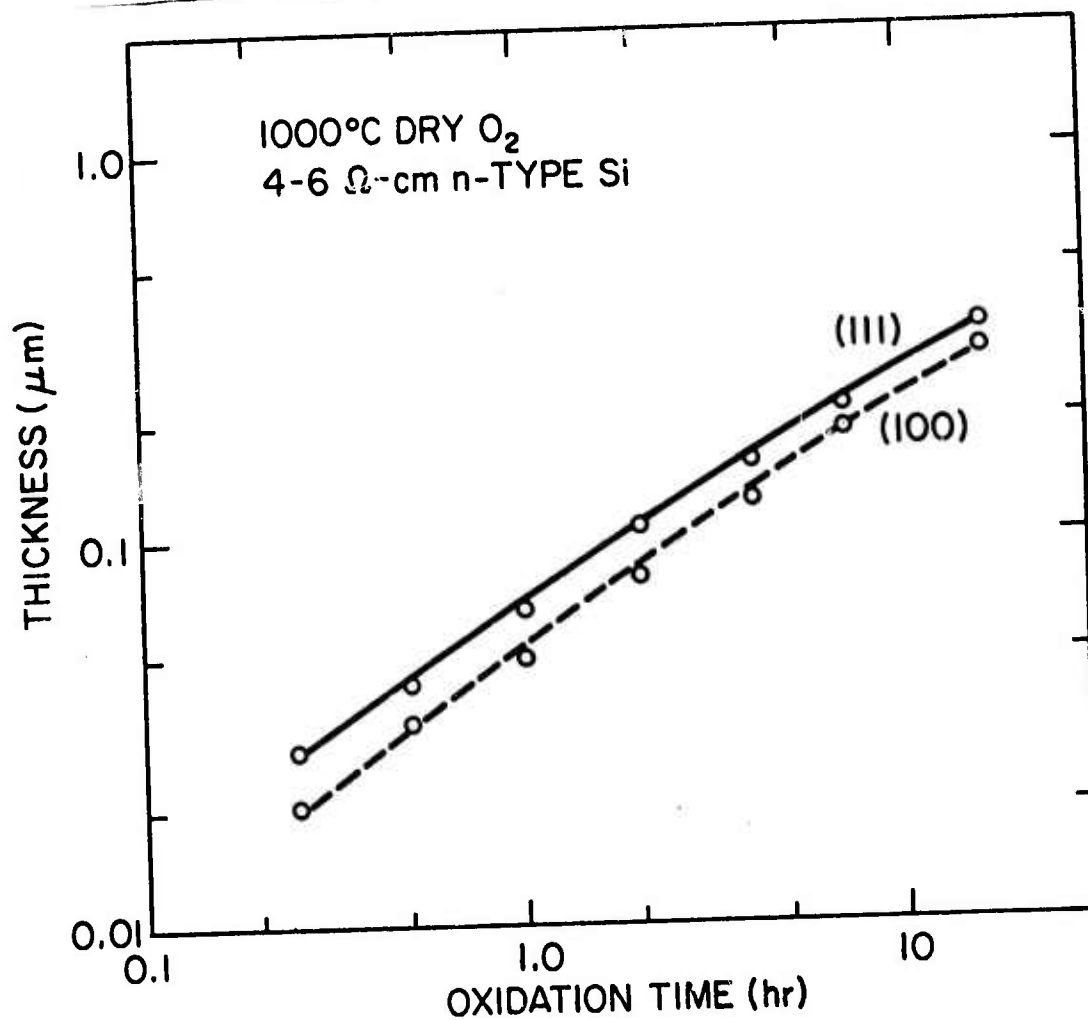


Figure IX-1. Oxide thickness versus time for the oxidation of (100) and (111) oriented N-type silicon in dry oxygen at 1000°C.

As indicated previously, a linear plot of X_o versus time gives a value for τ , which, for the above data is approximately 0.35 hr. Using this value, the resulting plot of X_o versus $(t+\tau)/X_o$ is shown in Fig. IX-2. It may be seen that the points for thicker oxides (longer oxidation times) lie reasonably close to a straight line, while most deviations from this line occur for the thinner oxides. Although some points appear to deviate significantly from the straight line, it must be realized that such a plot is extremely sensitive to small changes in X_o .

Silicon orientation is expected to affect surface reaction kinetics and thus to affect the linear rate constant B/A and not the parabolic rate constant B . Indeed, if we rewrite the general oxidation relationship in the form

$$\frac{X_o^2}{B} + \frac{X_o}{B/A} = t + \tau \quad (\text{IX-7})$$

where

$$B = C_1 p e^{-E_1/kT} \quad (\text{IX-8})$$

$$B/A = C_3 p e^{-E_3/kT}$$

then it is expected that the particular parameter affected by orientation is C_3 . Thus, the value of C_3 used in (IX-8) will depend on the orientation of the silicon being oxidized. Additional data are currently being taken on $\langle 111 \rangle$ and $\langle 100 \rangle$ samples to obtain the correct value of C_3 for these two orientations.

From a least squares analysis of the data presented in Fig. IX-2, the following parameters can be extracted.

TABLE IX-1. DRY O_2 OXIDATION PARAMETERS.

Ambient	Orientation	τ (hr)	B (μ_m^2/hr)	A (μ_m)
Dry O_2 , 1000°C	$\langle 111 \rangle$	0.35	0.009	0.101
Dry O_2 , 1000°C	$\langle 100 \rangle$	0.35	0.009	0.183

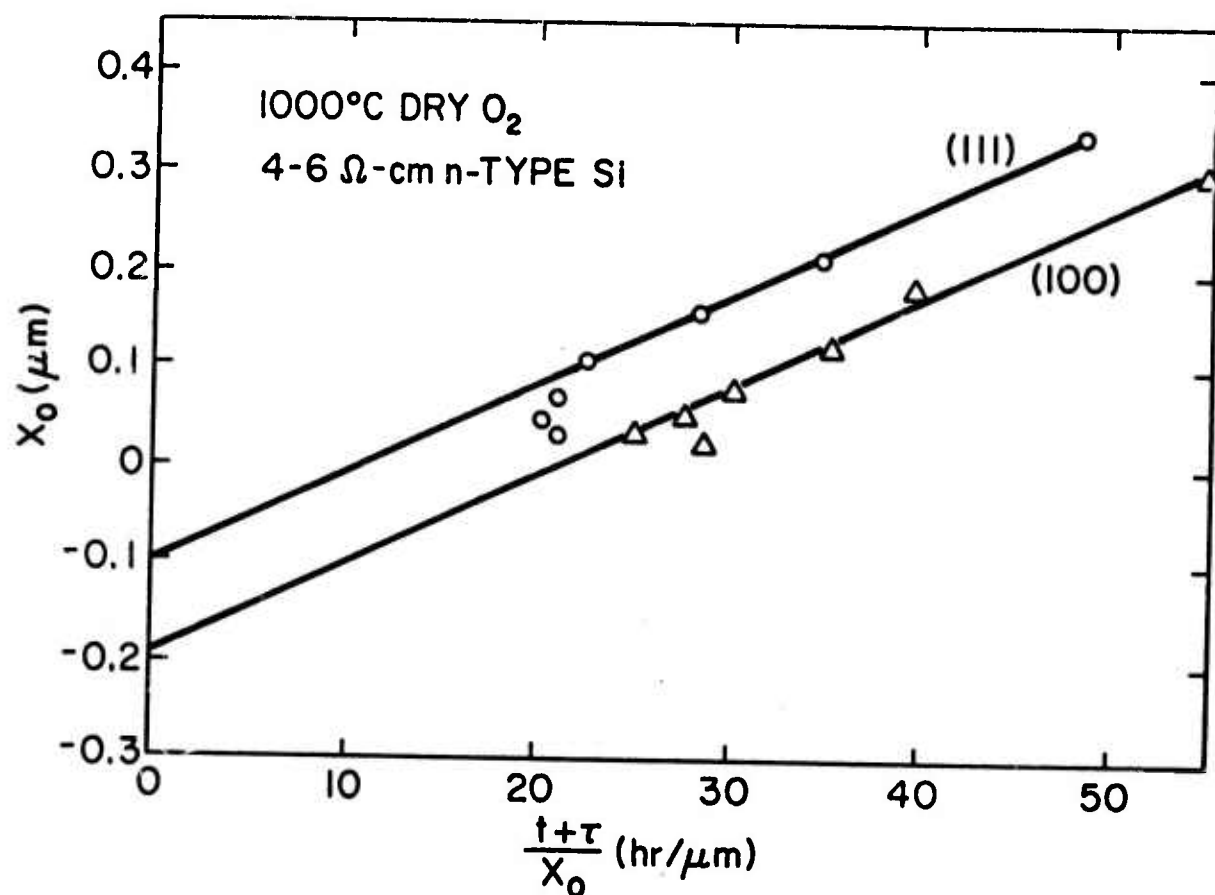


Figure IX-2. Determination of rate constants for the dry oxidation of (100) and (111) oriented N-type silicon at 1000°C. The slopes are equal to B and the intercepts to -A as per Eq. (IX-6).

As expected, no change in B is observed between the $\langle 111 \rangle$ and $\langle 100 \rangle$ silicon orientations. Also, A is larger for $\langle 100 \rangle$ than for $\langle 111 \rangle$, and thus B/A is smaller for $\langle 100 \rangle$ than for $\langle 111 \rangle$. This accounts for the increase in oxidation rate between the two orientations observed in Fig. IX-1.

4.0 Modifications to Incorporate 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 IX-3 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 [2] 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 Fig. IX-2, the two parameters that have been correlated [2] 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.

[2] B. E. Deal and M. Skalar, *J. Elec. Soc.*, 112, Apr. 1965, pp. 430-435.

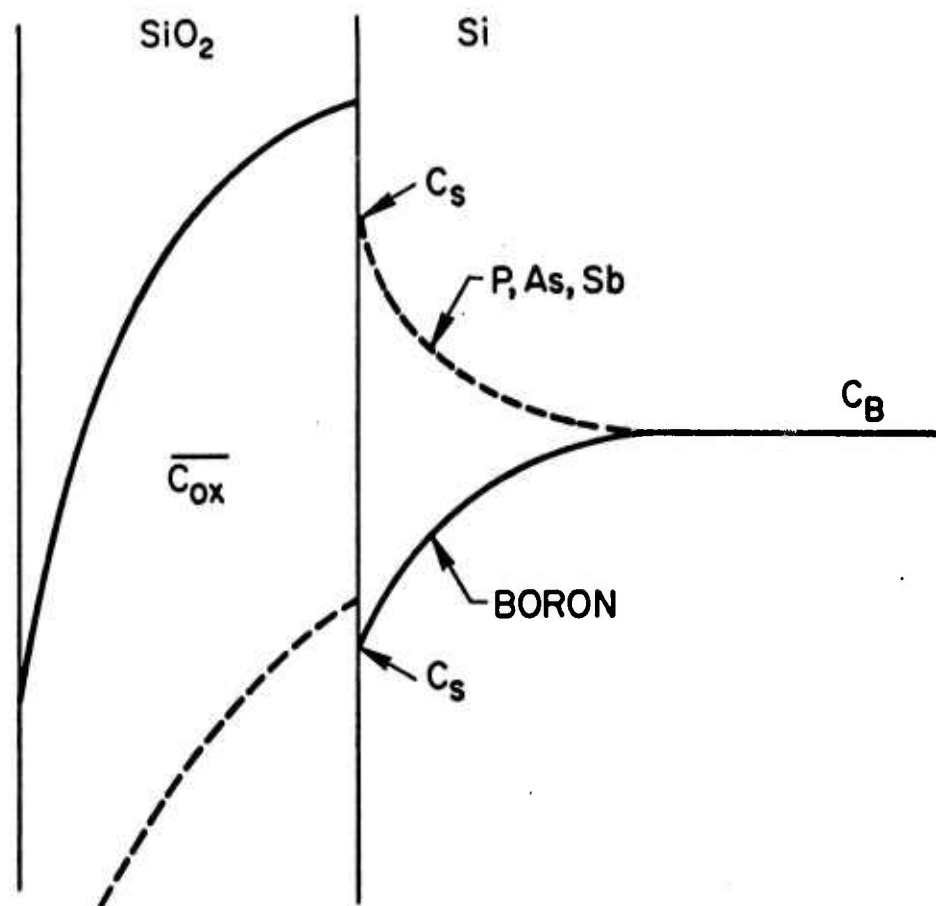


Figure IX-3. Effect of redistribution of Si and SiO₂ impurity profiles.

At the present time, work is beginning at Stanford to gather data on oxidation over heavily doped phosphorus, boron, and arsenic substrates. A small amount of published data exists in this area [2,3], and these data will be used where possible. This effort should be completed within the next several months. During the time the data are being gathered, the goal of the analytic portion of the work will be to incorporate the results into the general oxidation relationship, i.e., (IX-6).

5.0 Modifications to Incorporate HCl in the Oxidizing Ambient

Recently, it has been shown that addition of a chlorine species during silicon oxidation results in improved threshold stability and increased dielectric strength [4-7]. Thus, widespread use of this technique at Fairchild and other semiconductor companies ensued. Since it has also been found that chlorine addition increases the rate of silicon oxidation, it is proposed that the effect of chlorine on the thermal oxidation kinetics and on subsequent charge properties be established as part of the Fairchild portion of the ARPA program. The ultimate objective of this part of the program is to fit the chlorine oxidation rate data into the general relationship for the thermal oxidation of silicon and determine the effect of chlorine on the parabolic and linear rate constants.

-
- [2] B. E. Deal and M. Skalar, *J. Elec. Soc.*, 112, Apr. 1965, pp. 430-435.
 - [3] A. S. Grove, O. Leistiko, and C. T. Sah, *J. Appl. Phys.*, 35, 9, Sept. 1965, pp. 2695-2701.
 - [4] G. Baccarani, M. Severi, and G. Soncini, *J. Electrochem. Soc.*, 120, 1973, p. 1436.
 - [5] K. Hirabayashi and J. Iwamura, *J. Electrochem. Soc.*, 120, 1973, p. 1595.
 - [6] R. J. Kriegler, Y. G. Cheng, and D. R. Colton, *J. Electrochem. Soc.*, 119, 1972, p. 388.
 - [7] Y. J. Van der Meulen, C. M. Osburn, and J. F. Ziegler, *J. Electrochem. Soc.*, 122, 1975, p. 284.

At the present time, kinetic data have been taken on 1% HCl in dry O₂ at 1000°C. The calculated rate constants are shown in Table IX-2. These data will be extended during the next two quarters to cover the temperature range between 900°C and 1100°C and HCl concentrations between 1 and 10%.

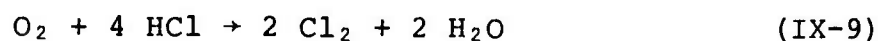
TABLE IX-2. HCl OXIDATION PARAMETERS.

Ambient	Orientation	τ (hr)	$B(\mu_m^2/hr)$	$A(\mu_m)$
Dry O ₂ , 1% HCl, 1000°C	$\langle 111 \rangle$	0.20	0.014	0.150
Dry O ₂ , 1% HCl, 1000°C	$\langle 100 \rangle$	0.20	0.014	0.144 to 0.176

A log-log plot of the 1% HCl data shows a 15 to 30% increase in oxidation rate as compared to dry oxidation at the same temperature.

It should be noted that a range is given for A in the case of O₂ + 1% HCl on $\langle 100 \rangle$ oriented silicon. This is due to the fact that for this oxidation ambient, B was experimentally determined to be 0.12 $\mu m^2/hr$ for $\langle 100 \rangle$ oriented silicon, instead of 0.14 $\mu m^2/hr$ as determined for $\langle 111 \rangle$ silicon. As discussed above, and indeed, as observed for the dry oxygen oxidation, B should be the same for both orientations. Thus, using 0.14 $\mu m^2/hr$ for B for $\langle 100 \rangle$ oriented silicon, we obtain different A values depending upon how the straight line (with fixed slope) is fitted to the experimental points. The above assumption of constant B for different orientations needs further investigation in the case of HCl oxidation since it is possible that other effects could enter the expression for B.

Further inspection of Table IX-2 shows that a decrease in τ and an increase in B result from the addition of 1% HCl to the oxidation ambient. Such changes are consistent with the generation of water vapor in the oxidation tube via the following reaction:



since it is known that τ is zero for steam oxidation, and also

that B is significantly larger for steam than for dry oxidation. At this time, no quantitative correlation between the amount of water vapor generated by the above reaction and the observed increase in oxidation rate for HCl addition to dry oxygen has been attempted.

It is expected, then, at the present time, that both B and B/A will be affected by the addition of HCl to the oxidizing ambient. (Intuitively, it seems reasonable that Cl may affect both the surface reaction rate and hence B/A and the diffusion of the oxidizing species through the oxide and thus B .) A joint effort between Stanford and Fairchild is currently under way and will continue to incorporate the HCl results into the generalized oxidation relationship.

6.0 Summary

In summary, the principal goal of the first year's work is the realization of an analytic expression which predicts accurately the oxide thickness resulting from an arbitrary sequence of oxidation steps. We are attempting to include the effects of silicon orientation, heavily doped regions, and the presence of chlorine by obtaining modified parabolic and linear rate constants which include these parameters.

Chapter X

Silicon Epitaxy

	<u>Page</u>
1.0 Introduction.....	213
2.0 Kinetics of Epitaxial Growth.....	213
3.0 Kinetics of Dopant Inclusion in the Epitaxial..... Layer	218
3.1 Doping of Epitaxial Layers.....	218
3.2 System Transfer Function of the Epitaxial.... Reactor-Basic Approach	222
3.3 Experimental Work.....	225
3.4 Systems Analysis.....	228
4.0 Deep Depletion MOS Capacitance Voltage..... Measurements to Obtain Semiconductor Impurity Profiles	229
4.1 Introduction.....	229
4.2 A Comparison of Measurement Techniques.....	229
4.3 The DDMOS Capacitor.....	231
5.0 Autodoping.....	235

Chapter X

Silicon Epitaxy

1.0 Introduction

The third part of this program deals with modeling of epitaxial growth of silicon on silicon. Initially, the work was divided into three areas:

- (a) kinetics of epitaxial growth
- (b) kinetics of dopant inclusion in the epitaxial layer
- (c) dopant profile measurement techniques

The progress in these areas during the first six months is reported here. During the next six months, we expect also to consider auto-doping. A discussion of these plans is also included.

All work presented here was done in the commercially available horizontal reactor* illustrated in Fig. X-1. The numbered features are:

- (1) water-cooled rf 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 gasses occurs
- (6) double-dilution doping system.

2.0 Kinetics of Epitaxial Growth

The generally accepted model for epitaxial growth is the stagnant layer model [1]. A horizontal reactor cross section is shown in Fig. X-2. Graphs of reactor variables pertinent to the stagnant layer model are also given. It has been shown both theoretically and empirically that during epitaxial growth in a horizontal reactor, the gas stream above the susceptor divides

[1] F. C. Eversteyn et al., *J. Electrochem. Soc.*, 117, 7, July 1970, pp. 925-931.

* Unicorp Model HIER II, Unicorp, Sunnyvale, California.

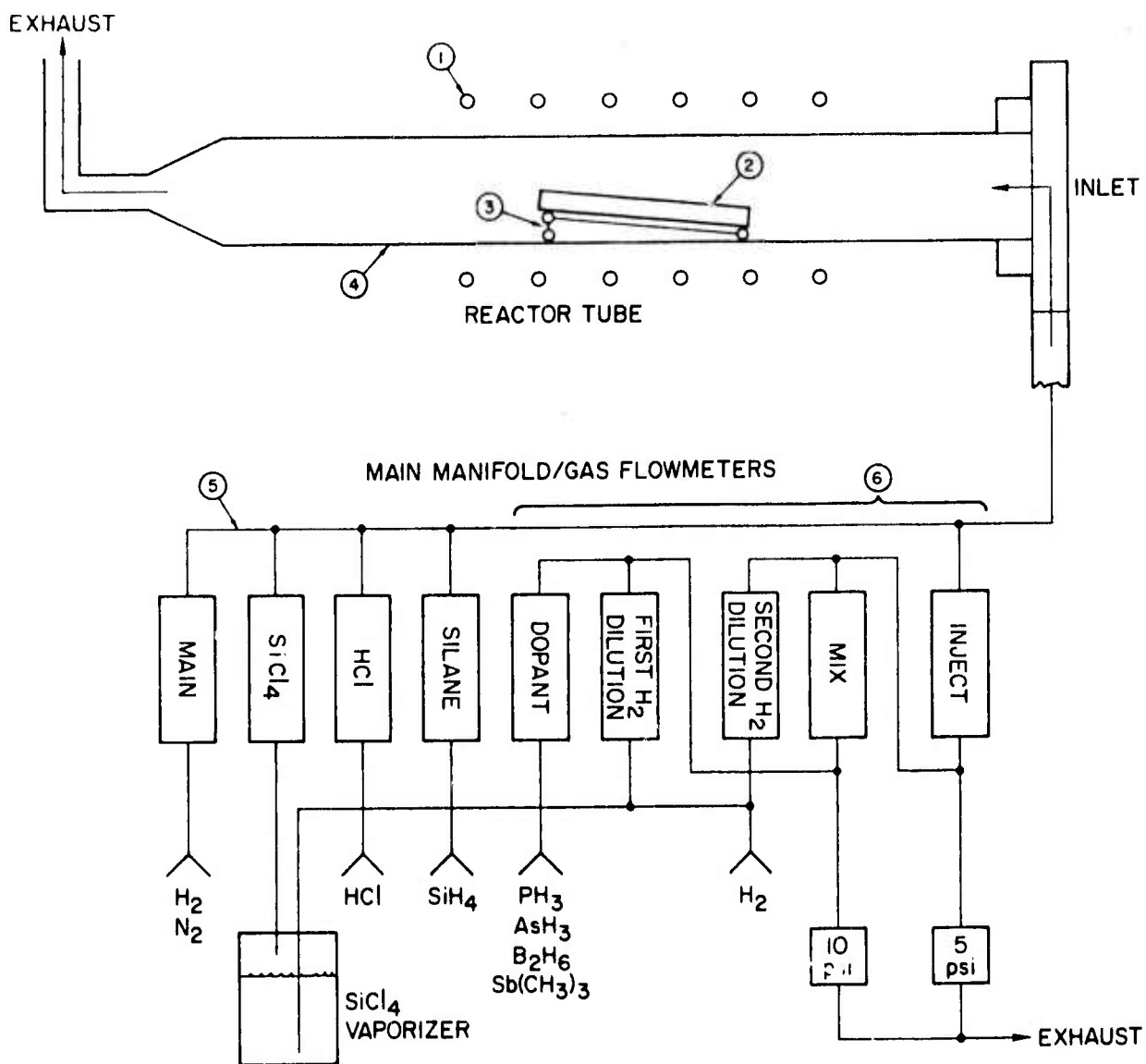


Figure X-1. Horizontal epitaxial reactor schematic.

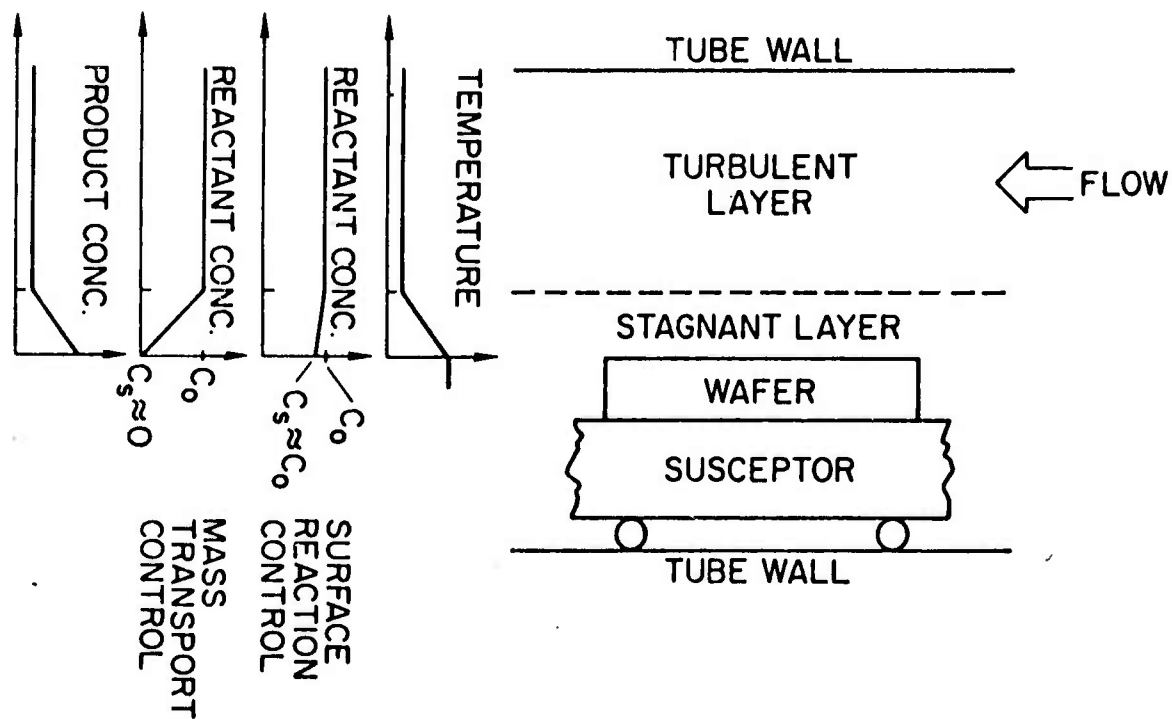


Figure X-2. Horizontal reactor cross section with thermal and chemical profiles during growth.

into two parts, a turbulent layer and a stagnant layer. The stagnant layer of about 1/2 cm in thickness is next to the susceptor. There is no displacement gas flow in the stagnant layer either vertically or horizontally across the wafer. The turbulent layer has horizontal flow corresponding to the main flow displacement and also vertical flow due to thermal convection tumbling. The turbulent layer 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 stagnant layer as graphed in Fig. X-2. Epitaxial growth proceeds by the following steps [2]:

- (1) Mass transfer of the reactant molecules (e.g., SiH_4) by diffusion from the turbulent layer reservoir across the stagnant layer to the silicon surface.
- (2) Adsorption of reactant atoms on the surface.
- (3) One or more chemical reactions at the surface.
- (4) Desorption of product molecules (e.g., H_2).
- (5) Mass transfer of the product molecules by diffusion through the stagnant layer, back to the turbulent layer.
- (6) Lattice arrangement of the adsorbed silicon atoms. This step may occur as part of (3).

Product and reactant concentrations as a function of distance from the gas-solid interface are also given in Fig. X-2. Concentration gradients across the stagnant layer are such that there is a diffusion flux of reactant molecules toward the surface and a flux of product molecules away from the surface. The turbulent layer is depicted as having no concentration or temperature gradient.

The surface-controlled and mass-transport controlled realms of SiH_4 deposition are easily described by the balancing of molecular fluxes because no reverse reactions complicate the SiH_4 problem. The growth rate v is given by the relation

[2] E. G. Bylander, *J. Electrochem. Soc.*, 109, 12, Dec. 1962, pp. 1171-1175.

$$v = K_s [\text{SiH}_4]_s \quad (\text{X-1})$$

$$= K_s [\text{SiH}_4]_t \left\{ 1 + \frac{N_{\text{Si}}}{\psi_{\text{SiH}_4}} K_s \right\}^{-1} \quad (\text{X-2})$$

The surface reaction rate constant K_s has the characteristic activation energy of 37 kcal/mole and the units of the growth rate (cm/sec). $[\text{SiH}_4]_t$ and $[\text{SiH}_4]_s$ are SiH_4 concentrations (ratio of silane to hydrogen) in turbulent layer and at the silicon surface, respectively. N_{Si} is the atomic density of silicon, $N_{\text{Si}} = 5.0 \times 10^{22} \text{ cm}^{-3}$. ψ_{SiH_4} is the mass transport coefficient for SiH_4 , a temperature insensitive constant which relates molecular flux to concentration gradient in the stagnant layer. ψ_{SiH_4} is proportional to the SiH_4 diffusion coefficient in H_2 and inversely proportional to the stagnant layer thickness. The flux of silicon atoms being incorporated into the lattice is (there are two limiting cases):

- (1) Surface reaction control. At low temperatures, K_s is small enough that $[\text{SiH}_4]_s \approx [\text{SiH}_4]_t$ (see Fig. X-4). In this case, from (X-1),

$$v = K_s [\text{SiH}_4]_t \quad (\text{X-3})$$

- (2) Mass transport control. At high temperatures, K_s becomes large, and $[\text{SiH}_4]_s \approx 0$. Then, from (X-2),

$$v = \frac{\psi_{\text{SiH}_4}}{N_{\text{Si}}} [\text{SiH}_4]_t \triangleq K_m [\text{SiH}_4]_t \quad (\text{X-4})$$

The rate constant for mass transport control is K_m (cm/sec).

Although both (X-3) and (X-4) predict a first order growth rate dependence on silane concentration, the surface reaction rate constant K_s is exponentially dependent on temperature while the mass transport rate constant K_m has been shown to have only a

linear dependence on temperature [1] (i.e., essentially no temperature dependence over the narrow 1000°C-1100°C range of temperatures commonly used for SiH₄ epitaxy).

Experimental SiH₄ growth rate is plotted versus temperature in Fig. X-3. The surface reaction region has an activation energy of $\Delta E = 39$ kcal/mole. Transfer to the mass transport region is nearly complete at 1000°C for all growth rates from 0.1 μ /min to 0.5 μ /min. From 1000°C to 1100°C, the maximum variation in growth rate observed is only about 10%. Figure X-4 is a graph of growth rate versus silane concentration for $T = 1050^\circ\text{C}$. The slope of the line is the mass transport rate constant value $K_m = 7.5 \times 10^{-4}$ cm/sec = 450 μ /min. These values for activation energy [3] and rate constant [1] are consistent with the literature.

3.0 Kinetics of Dopant Inclusion in the Epitaxial Layer

One of the major goals of this research is to develop a mathematical model for the kinetics of dopant inclusion in the epitaxial layer. Effort was divided into two areas initially. First, a model was developed to relate the gas phase dopant concentration to doping density in the epitaxial layer. In these experiments, the layers being grown were homogeneously doped. In the second research area, work is now being done to obtain a system transfer function of the epitaxial reactor by taking its transient response.

3.1 Doping of Epitaxial Layers. Epitaxial layers are most commonly doped by the incorporation of a small amount of dopant hydrides (PH₃, AsH₃, B₂H₆). Phosphine, arsine, and diborane are easily measured gaseous sources which can be diluted to the ppb level required for lightly doped epitaxial layers. Unfortunately, the gaseous antimony hydride SbH₃ has a half-life which is only on the order of one week, when it is stored as a 100 ppm H₂-diluted

-
- [1] F. C. Eversteyn et al., *J. Electrochem. Soc.*, 117, 7, July 1970, pp. 925-931.
 - [3] R.F.C. Farrow, *J. Electrochem. Soc.*, 121, 7, July 1974, pp. 899-907.

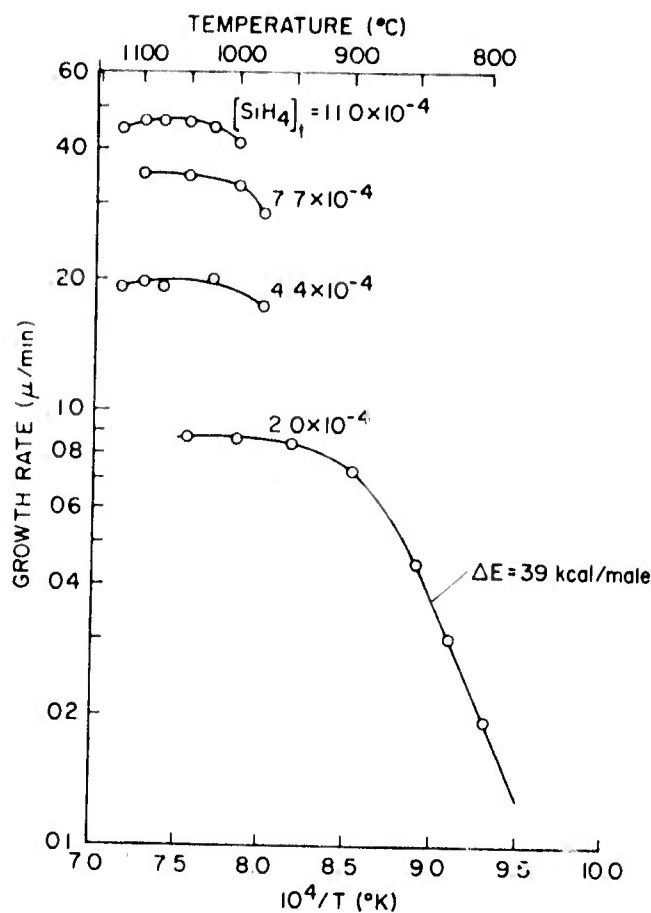


Figure X-3. Arrhenius plot of SiH_4 growth rate.

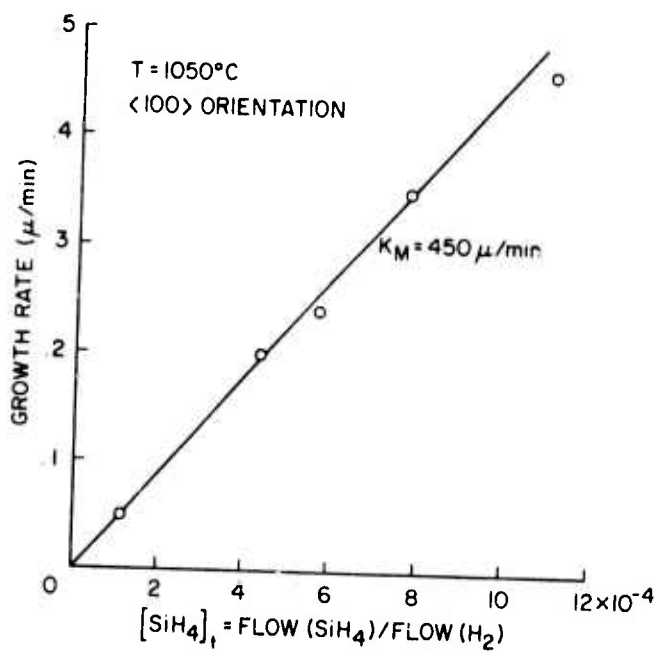


Figure X-4. SiH_4 growth rate versus SiH_4 concentration.

mixture. The liquid source trimethylstibine, $(\text{CH}_3)_3\text{Sb}$, is more stable than stibine, and may be prepared as a dilute vapor in H_2 for use as a gaseous antimony source.

The doping of SiH_4 epitaxial layers with dopant hydride sources has been described theoretically [4,5]. The dopant is transported across the stagnant layer by diffusion in the same manner as described for SiH_4 in the previous section. The analysis is complicated, however, by the fact that the incorporated dopant is nonuniformly distributed in the first few hundred angstroms near the surface of the growing layer. The relevant dopant concentrations in the reactor and in the wafer are depicted in Fig. X-5a for the case of PH_3 . The turbulent layer (input) dopant molar concentration $[\text{PH}_3]_t$ is the flow ratio PH_3/H_2 . The concentration of phosphine near the surface of the growing epitaxial layer is $[\text{PH}_3]_s$. The concentration of incorporated phosphorous atoms near the surface of the layer $[\text{P}]_s$ is higher than the phosphorous concentration deeper in the bulk material $[\text{P}]_b$. All of the incorporated phosphorous is considered to be ionized. The solid state phosphorous concentrations $[\text{P}]_s$ and $[\text{P}]_b$ have units of (atoms P/atoms Si), and must be multiplied by the atomic density of silicon $N_{\text{Si}} = 5.0 \times 10^{22}$ atoms/cm³ to get conventional doping concentrations.

The goal of the analysis is to predict the solid state doping $[\text{P}]_b$ from the turbulent layer phosphine concentration $[\text{PH}_3]_t$ as a function of temperature and growth rate. The effective segregation coefficient γ_{EFF} relates these two variables:

$$\frac{[\text{P}]_b}{[\text{PH}_3]_t} = \frac{1}{\gamma_{\text{EFF}}} \quad (\text{X-5})$$

It has been shown [4,5] for doping concentrations which are lower than the concentration of free electrons in silicon at epitaxial

[4] J. Bloem, *J. Crystal Growth*, 13/14, 1972, pp. 302-305.

[5] J. Bloem, *Semiconductor Silicon 1973*, op. cit., pp. 213-226.

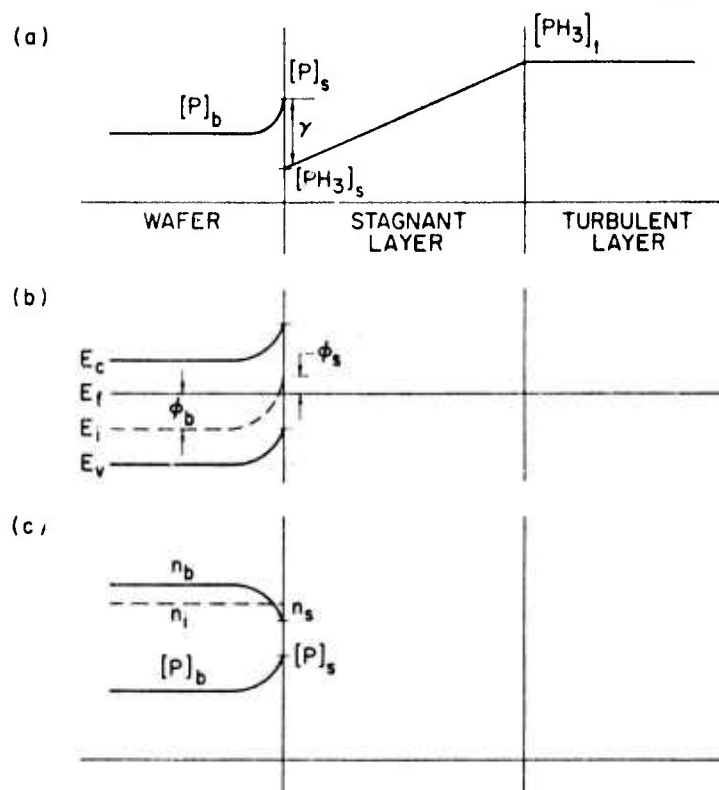


Figure X-5. (a) Dopant concentrations near an epitaxial interface.
 (b) Band bending near a growing silicon surface.
 (c) Dopant and free electron concentrations near a growing silicon surface.

temperatures ($\sim 10^{19} \text{ cm}^{-3}$) that the segregation coefficient is a function of only temperature and growth rate, but not doping level. That is, for a set temperature and growth rate, solid state doping is a linear function of gas phase dopant concentration.

The effective segregation coefficient has been measured experimentally for a wide range of conditions. Bulk doping level versus monatomic gas phase dopant concentration is given in Fig. X-6 for a combination of different dopant sources, silicon sources, growth rates, and temperatures. The doping concentration is calculated as the effective flow ratio of dopant gas to main hydrogen flow. With reference to the reactor schematic in Fig. X-1, this ratio is:

(X-6)

$$[DX_3]_t = \frac{f(\text{dopant})}{f(\text{dopant}) + f(1\text{st } H_2)} \cdot \frac{f(\text{mix})}{f(\text{mix}) + f(2\text{nd } H_2)} \cdot \frac{f(\text{inject})}{f(\text{main})}$$

f is flow in liters per minute. The concentration given in the right-hand scale is the usual concentration per cubic centimeter divided by N_{Si} . All the dopants have higher concentrations in the bulk than in the gas phase by a factor of 10-200. The effective partition coefficient for the PH_3/SiH_4 system of 0.01 is a factor of 10 larger than some reported values [4,5] but close to another [6]. All the dopants have about the same effective partition coefficient, about 0.01, except for AsH_3/SiH_4 at 1050°C and $(CH_3)_3Sb/SiH_4$.

3.2 System Transfer Function of the Epitaxial Reactor-Basic Approach

As stated earlier, the second and more important phase of this project is to develop a system transfer function of the epitaxial reactor from the point of view of dopant inclusion into the epitaxial layer. In this phase of the project, the overall characteristics of the dopant incorporation process are being modeled by a system approach: First, an epitaxial layer is grown

-
- [4] J. Bloem, *J. Crystal Growth*, 13/14, 1972, pp. 302-305.
 - [5] J. Bloem, *Semiconductor Silicon 1973*, op. cit., pp. 213-226.
 - [6] B. A. Joyce and R. R. Bradley, *J. Electrochem. Soc.*, 110, 12, Dec. 1963, pp. 1235-1240.

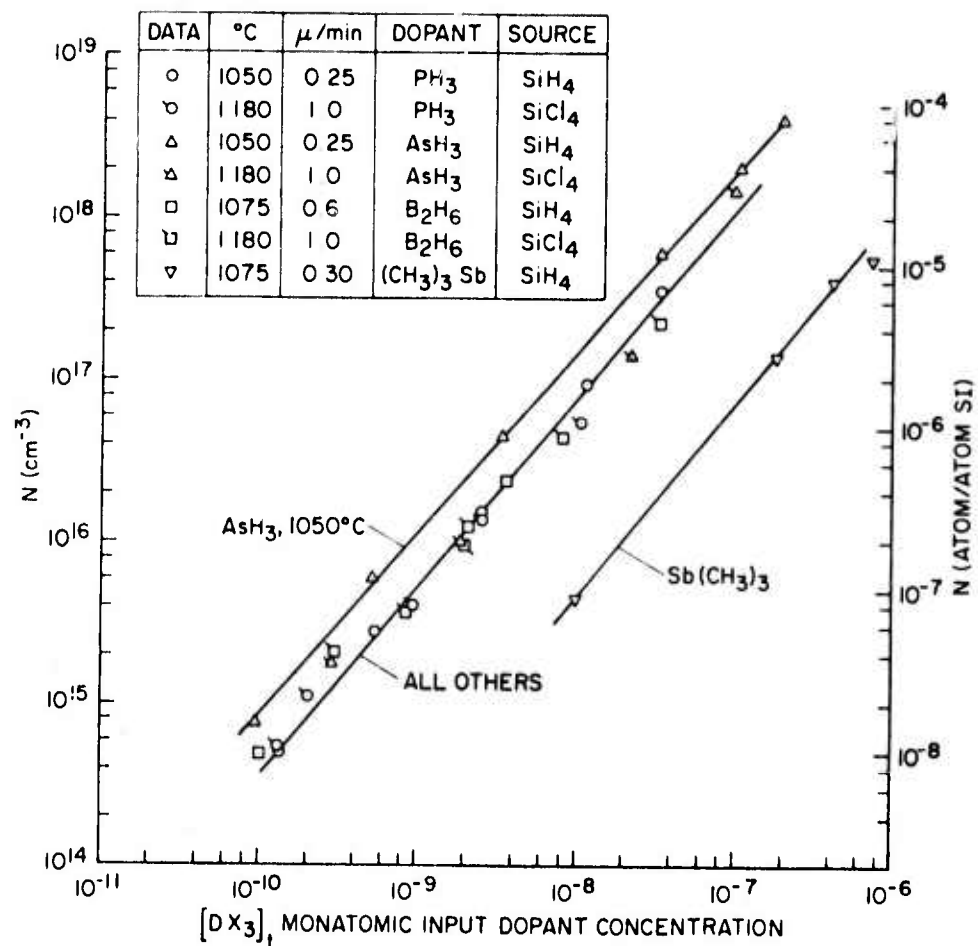


Figure X-6. Solid-state doping versus gas phase dopant concentration for various dopant sources, silicon sources, and temperatures.

with a dopant-gas flow which is a function of time. Then the resulting dopant profile in the epitaxial layer is determined. By relating the dopant profile in the epitaxial layer (system output) to the dopant-gas flow (system input), the system may be characterized by a transfer function. It is hoped that this transfer function can then be used to calculate the dopant flow as a function of time needed to realize a desired dopant profile in the epitaxial layer. In this manner, a controlled graded dopant concentration, which can be advantageous in numerous devices, can be obtained.

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 [7]. 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 [8]. 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 [9], transient studies may reveal the importance of

-
- [7] P. F. Deisler, Jr. and R. H. Wilhelm, *Ind. and Eng. Chemistry*, 45, 1953, p. 1219.
 - [8] J. B. Rosen and W. E. Winsche, *J. Chem. Phys.*, 18, 1950, p. 1587.
 - [9] H. Kobayashi and M. Kobayashi, *Catalysis Reviews--Science and Engineering*, 10, 1974, p. 139.

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.

3.3 Experimental Work

Using the results of the work described in the first two sections, the reactor was first optimized for a nominal deposition of approximately $0.5 \mu\text{m}/\text{min}$ and uniform doping during the entire deposition process. Silane was used as the source of Si and deposition was done at 1050°C . Arsine was used as the dopant gas and flow settings were found for the typical dopant concentrations of $1 \times 10^{15} \text{ cm}^{-3}$ and $3 \times 10^{15} \text{ cm}^{-3}$. Then layers were grown with a step-function change in the dopant gas flow during the continuous deposition of an epitaxial layer, that is, the dopant-gas flow was changed from one of the previously obtained flow settings to the other during the deposition.

The dopant profile in the epitaxial layer was obtained by capacitance-voltage measurements on deep-depletion MOS structures and p-n junctions. The thickness and dopant concentrations used in the samples were chosen to be compatible with the C-V technique. Figure X-6 shows the dopant profile in the step epitaxial layer. The indicated dopant concentration is actually closer to the majority-carrier concentration than to the dopant concentration since the correction between the two [10] was not applied in the

[10] D. P. Kennedy and R. R. O'Brien, *IBM J. Res. Develop.*, 13, 1969, p. 212.

data reduction procedure. Subsidiary analysis shows, however, that the difference between the two quantities is not large in the present case.

Measurements using the spreading-resistance technique indicate the same gross features shown by the C-V measurements and provide a cross-reference for the depth scale. However, the raw spreading-resistance data have not yet been analyzed to produce the dopant profile.

From Fig. X-7, we see that the transition from one dopant concentration to the other occurred in 1.2 μm , corresponding to a time of approximately 2 min. In addition, there is an initial time delay between the change of the dopant gas flow and the start of the transition region. This time can be attributed to delays in the reactor plumbing and diffusion through the stagnant boundary layer; the flow-meter which was changed during the deposition was chosen to minimize the former delay.

Further epitaxial layers have been prepared with step function changes in dopant-gas flow to study the details of the transition region more thoroughly. C-V analyses will be performed on these samples.

The heat cycling during the fabrication of the MOS and p-n junction C-V samples produced a $\sqrt{Dt}$ of only 0.07 μm 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 C-V sample fabrication. However, a mesa C-V technique is being developed so that other more-rapidly diffusing dopants may be used as desired.

After characterization of increasing step functions is complete, decreasing step functions will be studied. The system response will not necessarily be symmetric for the two cases because of the different mechanisms involved; in particular, evaporation of dopant from the already deposited layer and reincorporation into the growing layer may be significant in the second case. After this case is investigated, combinations of

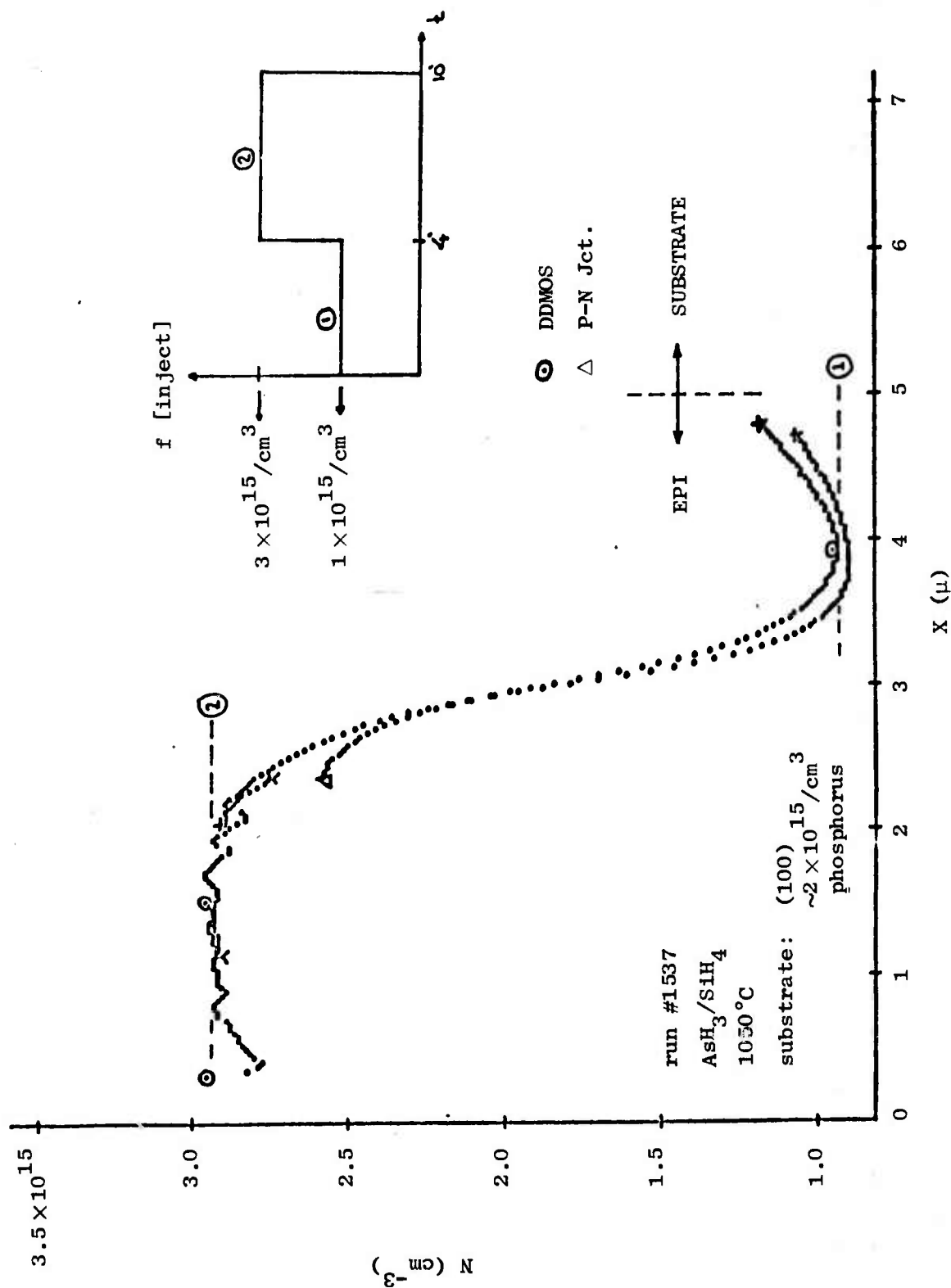


Figure X-7. Doping profile of the step epitaxial from C-V Measurements.

increasing and decreasing step functions will be used in an attempt to approximate an impulse in dopant-gas flow.

3.4 Systems Analysis

A trial function has been used for characterization of the epitaxial dopant. The dopant concentration in the step epitaxial layer has been modeled by quadratic functions, the simplest function which is continuous and has a continuous derivative. With the dopant concentration (system output) modeled by quadratic functions for a step-function change in the dopant-gas flow (system input), the system response can be found by dividing the Laplace transform of the two functions. For this simple case, the transfer function is analytic and can be transformed into the time domain as the impulse response of the system. As indicated above, in future work, increasing and decreasing step function changes of dopant gas flow will be combined to approach an impulse input.

After the limits and validity of the transfer-function approach have been established, more detailed characterization of the mechanisms involved in the overall transfer response will be attempted. Each process involved in the dopant inclusion will be investigated in order to find the limiting mechanisms. Some of these mechanisms will be dependent upon the particular epitaxial reactor employed while others will be more or less independent of the reactor itself and predominantly influenced by variables which can be kept constant from one reactor to another. One of the dominant mechanisms is expected to be diffusion of the dopant gas through the stagnant gas layer above the heated susceptor. Much recent work [11,12] has involved an investigation of the steady-state behavior of the dopant gas in this stagnant layer. We hope to extend this work to the transient case.

[11] T. Ishii et al., *J. Electrochem. Soc.*, 122, 1975, p. 1523.

[12] P. H. Langer and Joseph I. Goldstein, *J. Electrochem. Soc.*, 121, 4, April 1974, pp. 563-571.

4.0 Deep Depletion MOS Capacitance Voltage Measurements to Obtain Semiconductor Impurity Profiles

4.1 Introduction. The high frequency capacitance versus voltage characteristics of MOS capacitors, p-n junction diodes, and Schottky barrier diodes, have been used for some time to determine semiconductor impurity profiles. There are many problems associated with the measurement and analysis techniques used to obtain the C-V data and the corresponding impurity distribution. One purpose of this investigation was to consider the feasibility of collecting C-V data from a relatively new MOS capacitor structure. This device allows the retention of the deeply depleted state of the MOS capacitor under large bias conditions by preventing the formation of an inversion layer at the surface. Due to the nature of the operation of the device, we will subsequently refer to it as a deep depletion MOS (DDMOS) capacitor. In the process of examining the use of the DDMOS capacitor, the limitations of the analysis techniques were investigated in order to determine the accuracy and resolution which can be expected for various types of impurity distribution. The discussion to follow will refer to structures fabricated in an n-type substrate.

4.2 A Comparison of Measurement Techniques. Until the present time, the most common methods of obtaining high frequency C-V information for use in profiling has been through the use of reverse biased Schottky barrier diodes. In some cases, reverse biased Schottky barrier diodes and MOS capacitors are also utilized. However, factors such as the accuracy when profiling near the surface, the ease with which measurements can be performed, and the maximum depth at which a profile can be obtained have made the use of the Schottky barrier diode advantageous.

When it is desirable to measure the impurity profile near the surface of the substrate, the use of a p-n junction is obviously not acceptable. The surface is greatly modified by the p-type diffusion required to form the junction. Both the MOS capacitor and the Schottky barrier diode avoid this problem and allow profiling near the surface. In the case of the MOS capacitor, there

may be a loss of accuracy near the surface due to fast surface states at the oxide-silicon interface. This problem occurs if the interface states are able to follow the ac voltage used to measure the MOS capacitance. In this case, the C-V characteristics are modified for small values of applied bias, resulting in an incorrect profile for small depths. It is possible to minimize these effects by measuring the capacitance with a signal having a period which is small compared to the time constant of the interface states. In addition, some interface state errors can be corrected by combining the high frequency C-V information with a measurement of the low frequency C-V characteristics [13].

For all of the measurement methods, the maximum depth at which the doping profile may be obtained is the maximum depletion depth which can be reached before the surface undergoes avalanche breakdown. The p-n junction diode allows the achievement of the deepest depletion since the depletion starts at the junction, which is already located at some depth within the substrate. However, several factors make the use of the junction diode undesirable. The formation of the junction by diffusion causes compensation of the substrate at depths greater than that of the junction and leads to errors in the profile near the junction. The depletion of the diffused region and the capacitance of the junction side wall also introduce errors. Some of these errors can be corrected if the doping profile in the diffused region is known [14]. This information is not always easily acquired. The final objectionable aspect of the use of junction diodes is the necessity of using a destructive technique to measure the junction depth. The junction depth is needed because the analysis of the C-V information yields the doping profile as a function of the distance from the junction.

[13] J. R. Brews, *J. Appl. Phys.*, 44, July 1973, pp. 3228-3231.

[14] M. G. Buehler, *IEEE Trans. on Electron Devices*, ED-19, November 1972, pp. 1171-1178.

By using the MOS capacitor or the Schottky barrier diode, the maximum depletion depth can be attained without encountering the problems associated with the use of the junction diode. The standard MOS capacitor presents an additional difficulty. In order to achieve a deeply depleted condition, some method must be used to prevent the inversion of the semiconductor surface while the capacitance measurement is made. This generally requires the use of specialized equipment to measure the capacitance during the application of a bias pulse which depletes the substrate for a short time. The length of time before inversion occurs depends on the minority carrier generation rate. The use of the DDMOS capacitor simplifies matters a great deal because the capacitance is measured under static conditions. The only apparent disadvantage of this structure is the necessity of placing a p-type diffused region around the MOS capacitor. However, the extra processing involved is also required in the case of the Schottky barrier diode, since a guard ring is necessary to prevent reverse breakdown at the edge of the diode.

It is apparent that in order to achieve a simple and accurate measurement technique, either the DDMOS capacitor or the Schottky barrier diode should be used. Due to the concentration of this laboratory's efforts on MOS device technology and our ability to obtain interface state densities reliably as low as $10^{10}/\text{cm}^2$, it was felt that the use of the DDMOS capacitor should be investigated further.

4.3 The DDMOS Capacitor. The DDMOS capacitor structure is shown in Fig. X-8. It is identical to a standard MOS capacitor except for the inclusion of a p-type diffused region which surrounds the capacitor. When the annular p-n junction is reverse biased, the quasi-fermi level for holes under the MOS capacitor gate oxide is set by the potential of the p-type region. Figure X-9 illustrates the positions of the energy bands in the region under the gate oxide when the p-n junction is reverse biased. As long as the condition $|\psi_s| < |V_R|$ is satisfied, the valence band is far enough from the quasi-fermi level for holes to prevent

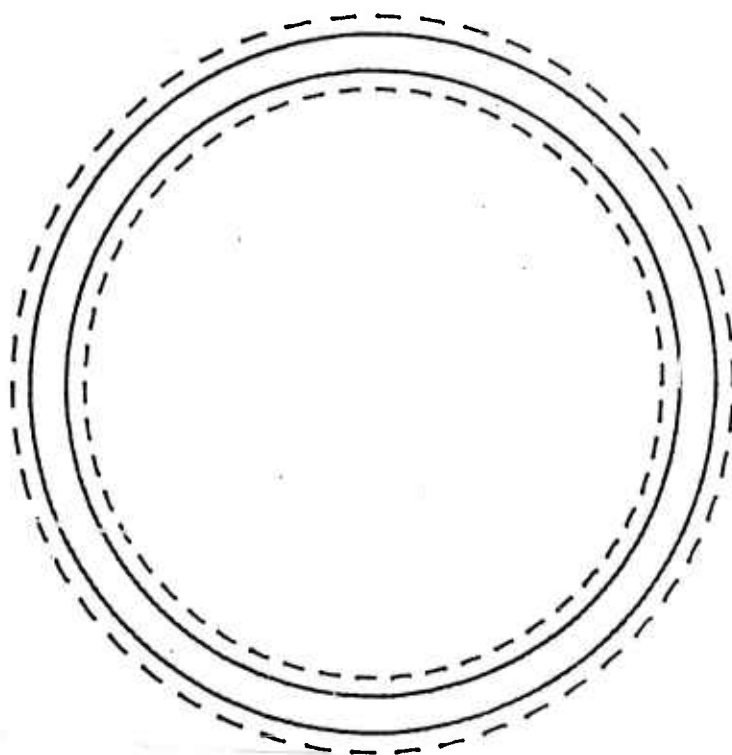
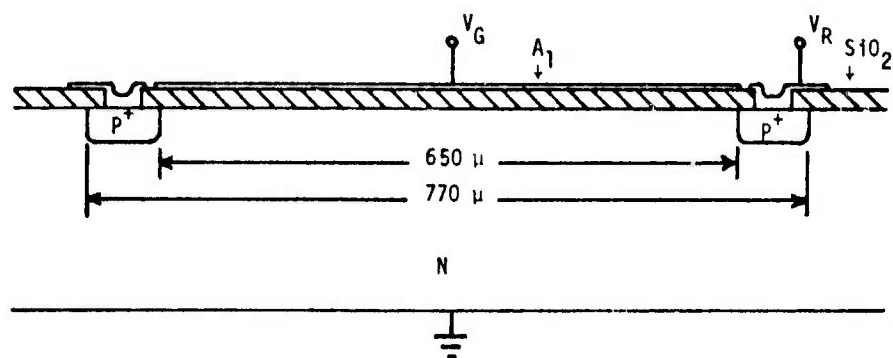


Figure X-8. The DDMOS capacitor structure.

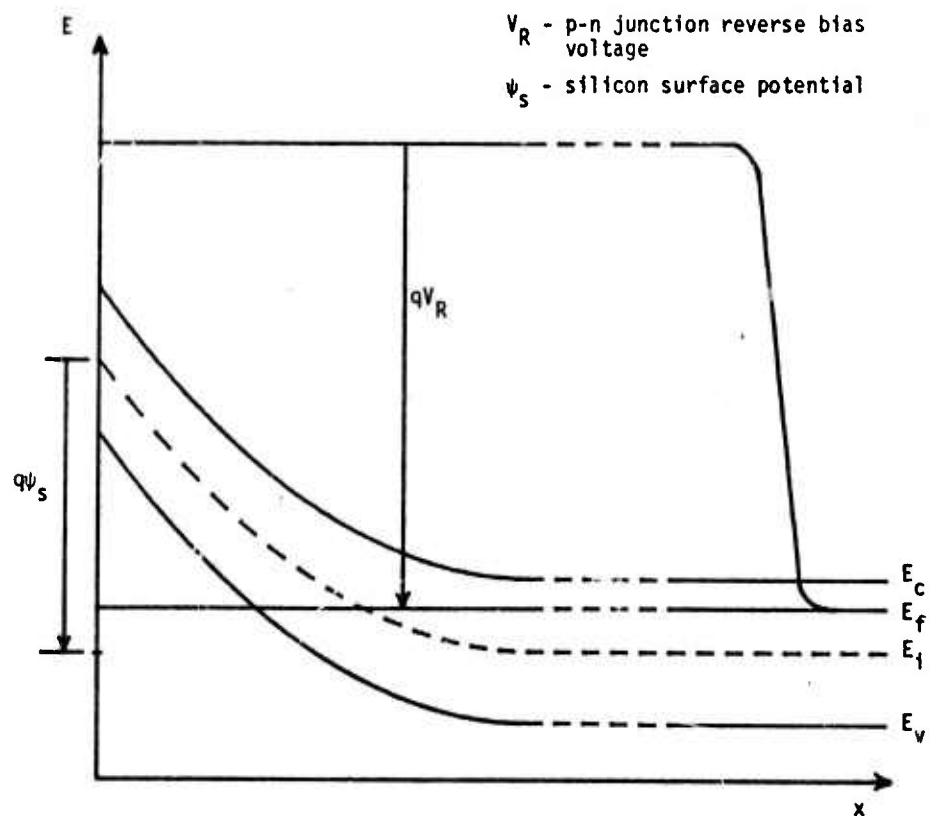


Figure X-9. Typical energy band diagram for the substrate under the gate oxide.

any significant accumulation of holes at the surface. Of course, in order to prevent surface inversion under static conditions, the reverse biased junction must be continually removing the thermally generated holes. This requires a lateral charge transfer under the gate oxide. One question which merits further investigation is whether the thermally generated charge is completely removed in the large area devices being used. The present study has concentrated on the limitations of the data analysis techniques.

The factor which limits the maximum depletion depth in the DDMOS capacitor is the onset of avalanche breakdown in the semiconductor. When breakdown occurs, a large number of minority carriers are generated. This inhibits the effective removal of the inversion layer charge by the reverse biased p-n junction. Avalanche breakdown may occur in two regions. As mentioned previously, the electric field at the oxide-silicon interface may reach the breakdown level. Another point where breakdown may occur is at the p-n junction. In order to prevent inversion, the reverse bias applied to the p-n junction must be greater than the maximum potential drop across the semiconductor (not including the potential drop across the oxide). Thus, the maximum depletion depth which can be obtained may be limited by the maximum reverse bias which the p-n junction can withstand. Factors such as junction curvature and the impurity gradient across the junction influence this maximum bias.

By applying the C-V analysis techniques to the theoretical C-V curves for various impurity distributions, we investigated the errors introduced by the depletion approximation and the use of the Ziegler [15] theory for the surface region. We also simulated the errors which result from inaccurate measurement of the C-V curve and the capacitor area. Results were obtained illustrating

[15] K. Ziegler, E. Klausmann, and S. Kar, *Solid State Electronics*, 18, February 1975, pp. 189-198.

the effects of these measurement errors on the impurity distribution produced by the C-V analysis techniques. The measurement errors were found to cause particularly large errors near the surface where the Ziegler theory is employed to obtain the impurity profile.

The results of this investigation indicate that the use of the Deep Depletion MOS capacitor can yield useful information about the impurity distribution in a nonuniformly doped substrate. However, the limitations described in the report must be kept in mind when evaluating the results obtained by using the C-V analysis techniques.

5.0 Autodoping

The subject of autodoping has been given substantial theoretical consideration during the current time period. It will be studied experimentally during the next period. One of the limiting and most important considerations in epitaxial deposition is the movement of substrate impurities into epitaxial layers. Two components of this movement are considered: outdiffusion, movement due to solid-state thermal diffusion in silicon, and autodoping, extra movement of substrate doping into an epitaxial layer which cannot be explained by diffusion effects. It has recently been shown [12] that there are three dopant fluxes from a wafer into an epitaxial layer that must be considered:

- (1) dopant "evaporated" from the back of the wafer which mixes into the turbulent layer.
- (2) dopant which "evaporates" from the front of the wafer into the stagnant layer, and
- (3) dopant which outdiffuses from the substrate due to thermal diffusion alone.

For the typical case of standard buried collector bipolar processing, a moderately doped ($10^{15} - 10^{16} \text{ cm}^{-3}$) epitaxial layer is grown over localized N^+ buried layers. Usually, the backside of the wafer is masked during the buried layer diffusion to eliminate this autodoping source. The wafer is then subjected to very

[12] P. H. Langer and Joseph I. Goldstein, *J. Electrochem. Soc.*, 121, 4, April 1974, pp. 563-571.

high temperature-time cycles to provide junction isolation. In this case, the scaled wafer backside does not provide an autodoping source, and front side autodoping is dominated by outdiffusion from the high temperature isolation step. Consequently, for standard buried-collector processing, outdiffusion alone can explain dopant migration from the buried layer into the epitaxial layer.

For low temperature silane epitaxial growth, especially when high temperature processing steps do not follow, there is a definite movement of dopant into the growing layer from the substrate which cannot be explained by outdiffusion.

Given the general description of the three components of dopant movement into an epitaxial layer-outdiffusion, front side autodoping, and backside autodoping-a conceptual picture of autodoping can be drawn. Figure X-10 shows that outdiffusion dominates nearest the interface, front side evaporation effects become apparent as soon as the outdiffusion tail dies out, and backside evaporation dominates the autodoping profile farthest from the interface because the backside autodoping source remains relatively constant while the front side source is continuously buried under growing silicon. Figure X-10 should be considered as being highly idealized. For example, the analysis does not consider lateral autodoping variations which include systematic worse autodoping at wafer edges due to the backside source. Any of the three regions in Fig. X-10 may not be present under certain conditions, e.g., a large outdiffusion tail may go completely to the surface of a thin epitaxial layer.

Aside from rather elaborate computer calculation of autodoping profiles [12], only empirical results can be used with

[12] P. H. Langer and Joseph I. Goldstein, *J. Electrochem. Soc.*, 121, 4, April 1974, pp. 563-571.

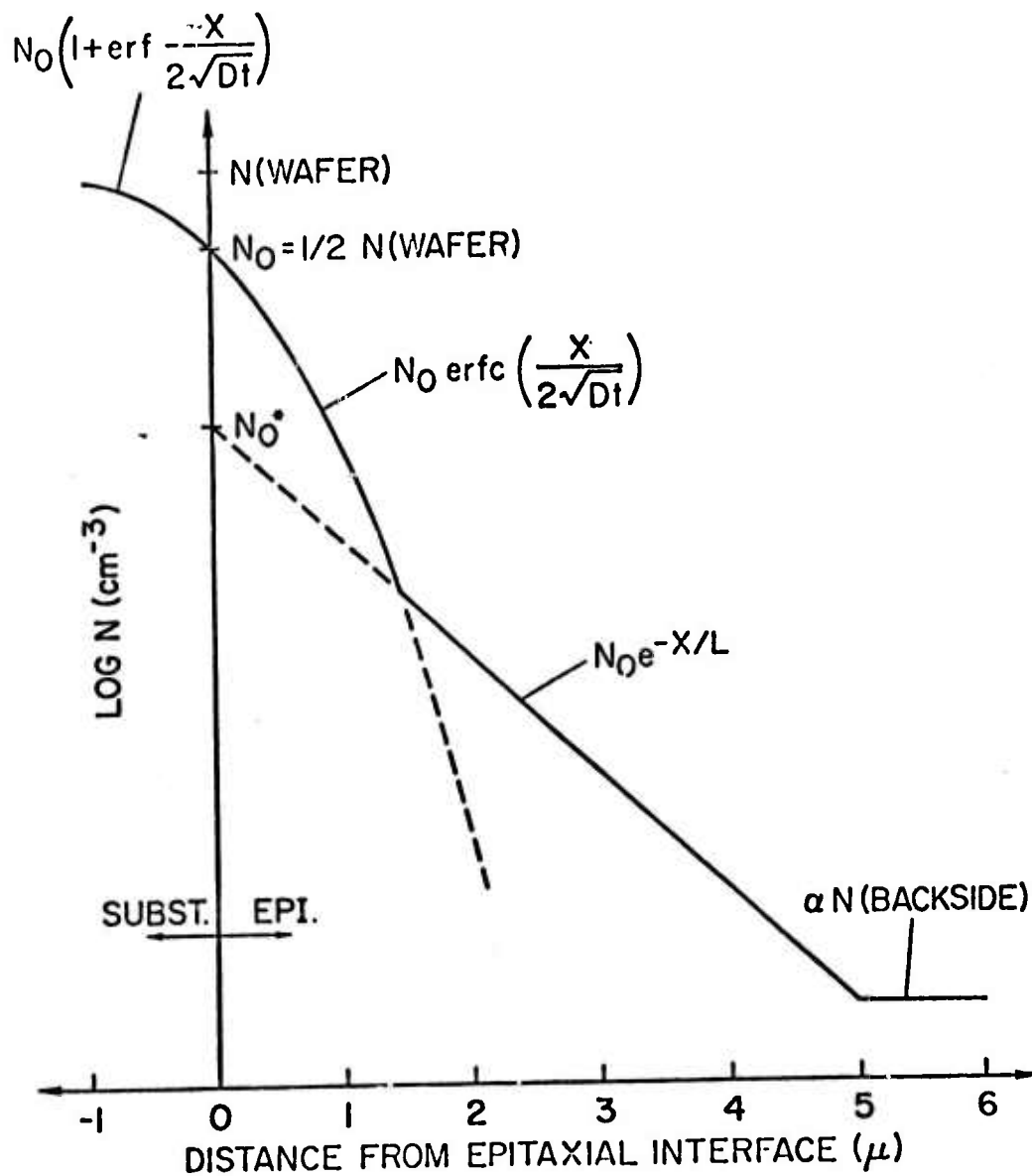


Figure X-10. Conceptual diagram of autodoping and outdiffusion profiles.

confidence, and even empirical autodoping data varies with reactor geometry and growth conditions. During the next few months, we will study more closely the problem of autodoping. Experiments will be done to obtain an analytical model to explain autodoping. The two-step epitaxial growth technique [11] to minimize autodoping will be investigated. In our preliminary studies, arsenic-doped silane epitaxial layers will be grown on arsenic doped substrates.

[11] T. Ishii et al., *J. Electrochem. Soc.*, 122, 1975, p. 1523.

Chapter XI

Mathematical Model of Impurity Diffusion

	<u>Page</u>
1.0 Introduction.....	241
2.0 Technical Approach.....	241
3.0 Numerical Analysis.....	244
4.0 Results.....	246
5.0 Conclusions and Recommendations.....	248

Chapter XI

Mathematical Model of Impurity Diffusion*

1.0 Introduction

To obtain lower cost and improved reliability of integrated circuits, the ability to simulate fabrication processes is mandatory. Process simulation will also allow the optimization of some aspects of device performance in terms of process variables. Simulation, however, cannot be accomplished unless adequate models of the integrated circuit fabrication process are developed.

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. The model will consider the influence of the internal electric field on the motion of impurity ions at elevated temperatures. The electric field is due to the charge density produced by the ionization of impurities present in the material.

2.0 Technical Approach

Diffusion in semiconductor material is an ambipolar process involving mobile holes, electrons, and ionized impurity atoms. In the fabrication of modern devices and integrated circuits, plane

* This work is exclusively the work of the Louisiana State University, Department of Electrical Engineering, Baton Rouge, Louisiana 70803.

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

$$f(x,t) = -D_c \frac{\partial c}{\partial x} + Z\mu_c cE \quad (XI-1)$$

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

$$\frac{\partial c}{\partial t} + \frac{\partial f}{\partial x} = 0 \quad (XI-2)$$

Substitution of (XI-1) into (XI-2) yields the transport equation which determines the ionized impurity distribution in the wafer; thus,

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x} \left(D_c \frac{\partial c}{\partial x} - Z\mu_c cE \right) \quad (XI-3)$$

Holes and electrons also satisfy similar equations, and it follows that

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial x} \left(D_p \frac{\partial p}{\partial x} - \mu_p pE \right) - R \quad (XI-4)$$

$$\frac{\partial n}{\partial t} = \frac{\partial}{\partial x} \left(D_n \frac{\partial n}{\partial x} + \mu_n nE \right) - R \quad (XI-5)$$

Gauss' law is used to relate the field to the charge density

$$\frac{\partial E}{\partial x} = \frac{e}{\epsilon} (p - n + Zc) \quad (XI-6)$$

The net recombination term is represented by

$$R = \frac{pn - n_i^2}{\tau_n (p + n_i) + \tau_p (n + n_i)} \quad (XI-7)$$

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.

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.

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

$$E = - \frac{\partial \phi}{\partial x} \quad (XI-8)$$

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 (XI-4) under these conditions yields

$$p = n_i e^{-\phi/V_T} \quad (XI-9)$$

where $V_T = D_p/\mu_p = KT/e$. In a similar manner, (XI-5) yields

$$n = n_i e^{\phi/V_T} \quad (XI-10)$$

The hole and electron density can now be eliminated from the problem statement by substituting (XI-9) and (XI-10) into (XI-6). Thus, the quasi-static problem can be specified by

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x} \left(D_c \frac{\partial c}{\partial x} + z \mu_c c \frac{\partial \phi}{\partial x} \right) \quad (\text{XI-11})$$

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{e}{\epsilon} (2n_i \sinh \phi/V_T - zc) \quad (\text{XI-12})$$

The first part of this research considered a vapor phase deposition where an inert carrier gas containing a single impurity species flows over an intrinsic semiconductor wafer. 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 gaseous diffusion process are given by

$$c(0, t) = C_0 \quad (\text{XI-13a})$$

$$\frac{\partial \phi(0, t)}{\partial x} = 0 \quad (\text{XI-13b})$$

$$c(\infty, t) = 0 \quad (\text{XI-13c})$$

$$\phi(\infty, t) = 0 \quad (\text{XI-13d})$$

$$c(x, 0) = 0, \quad x > 0 \quad (\text{XI-13e})$$

The quasi-static problem is specified by two nonlinear 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.

3.0 Numerical Analysis

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

$$c = n_i \hat{c} \quad (\text{XI-14a})$$

$$\phi = V_T \hat{\phi} / Z \quad (\text{XI-14b})$$

$$x = \sqrt{\frac{\epsilon V_T}{e n_i}} \hat{x} \quad (\text{XI-14c})$$

$$t = \frac{\epsilon V_T}{e D_c n_i} \hat{t} \quad (\text{XI-14d})$$

In order to simplify notation for this section, the normalized variables will be represented by c , ϕ , x , and t . Substituting the normalized variables (XI-11) and (XI-12) yields

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x} \left(\frac{\partial c}{\partial x} + c \frac{\partial \phi}{\partial x} \right) \quad (\text{XI-15})$$

$$\frac{\partial^2 \phi}{\partial x^2} = 2 \sinh \phi - c \quad (\text{XI-16})$$

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

$$t_i = (i - 1) \Delta t + t_0, \quad i = 1, \dots, n \quad (\text{XI-17})$$

$$x_j = (j - 1) \Delta x, \quad j = 1, \dots, m \quad (\text{XI-18})$$

where Δt and Δx were chosen to be fixed for simplicity.

A nonzero starting time t_0 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_0 varied and will be discussed later.

The time derivative was approximated by a two-point implicit scheme

$$\frac{\partial f(t_i)}{\partial t} \approx \frac{f(t_i) - f(t_{i-1})}{\Delta t} \quad (\text{XI-19})$$

Evaluating (XI-15) and (XI-16) at $t = t_i$ yields

$$\frac{d^2\phi}{dx^2} - 2 \sinh \phi + c = 0 \quad (\text{XI-20a})$$

$$\frac{d}{dx} \left[\frac{dc}{dx} + c \frac{d\phi}{dx} \right] - \frac{c - N}{\Delta t} = 0 \quad (\text{XI-20b})$$

where the notation

$$c = c(x, t_i) \quad (\text{XI-21a})$$

$$N = c(x, t_{i-1}) \quad (\text{XI-21b})$$

has been used. Notice that this constitutes a set of ordinary differential equations with a driving function N and boundary conditions

$$c(0) = \beta = C_o/n_i \quad (\text{XI-22a})$$

$$\frac{d\phi(0)}{dx} = 0 \quad (\text{XI-22b})$$

$$c(\infty) = 0 \quad (\text{XI-22c})$$

$$\phi(\infty) = 0 \quad (\text{XI-22d})$$

The procedure is straightforward. Once the distribution at some time t_{i-1} is known, (XI-20) and (XI-22) 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 Fig. XI-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.

4.0 Results

The input data used corresponded to the case for arsenic diffusion in silicon for a gaseous process:

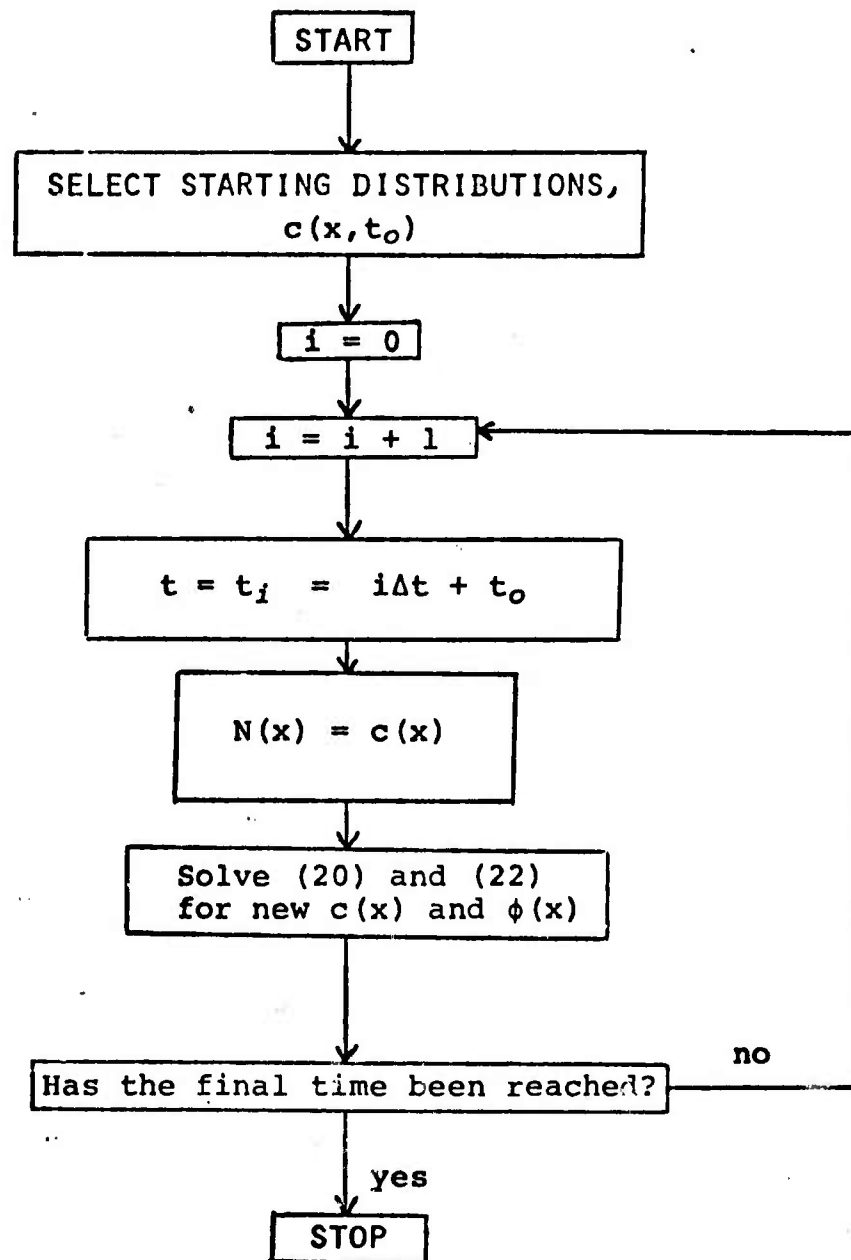


Figure XI-1. Flow diagram of the numerical method.

$$T = 1050^{\circ}\text{C}$$

$$C_o = 1.6 \times 10^{21} \text{ cm}^{-3}$$

$$D_c = 6.44 \times 10^{-15} \text{ cm}^2/\text{sec}$$

$$n_i = 1.17 \times 10^{19} \text{ cm}^{-3}$$

The surface concentration used represents solid solubility, and thus the effect of the internal electric field on the diffusion process should be maximum. The diffusion constant is the average value of data given by Masters and Fairfield with Chiu and Chosh. The intrinsic electron density was calculated from Morin and Maita.

Several starting conditions were used, the most prominent of which was the complementary error function, given by

$$c(x, t_o) = C_o \operatorname{erfc} \left[x / (4D_c t_o)^{1/2} \right] \quad (\text{XI-23})$$

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

$$\phi(x, t_o) = V_T \sinh^{-1} [c(x, t_o) / 2n_i] \quad (\text{XI-24})$$

Typically, t_o was chosen as 2 minutes with $\Delta x = 0.001 \mu$ and $\Delta t = 0.5 \text{ sec}$.

Figure XI-2 shows the results for arsenic in silicon for a gaseous diffusion process. Also shown is the complementary error function which is the exact solution for negligible electric field. Inspection of the curve indicates significant deviation in the impurity profiles from those predicted by simple theory. These results are typical for diffusion with high surface concentration.

5.0 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.

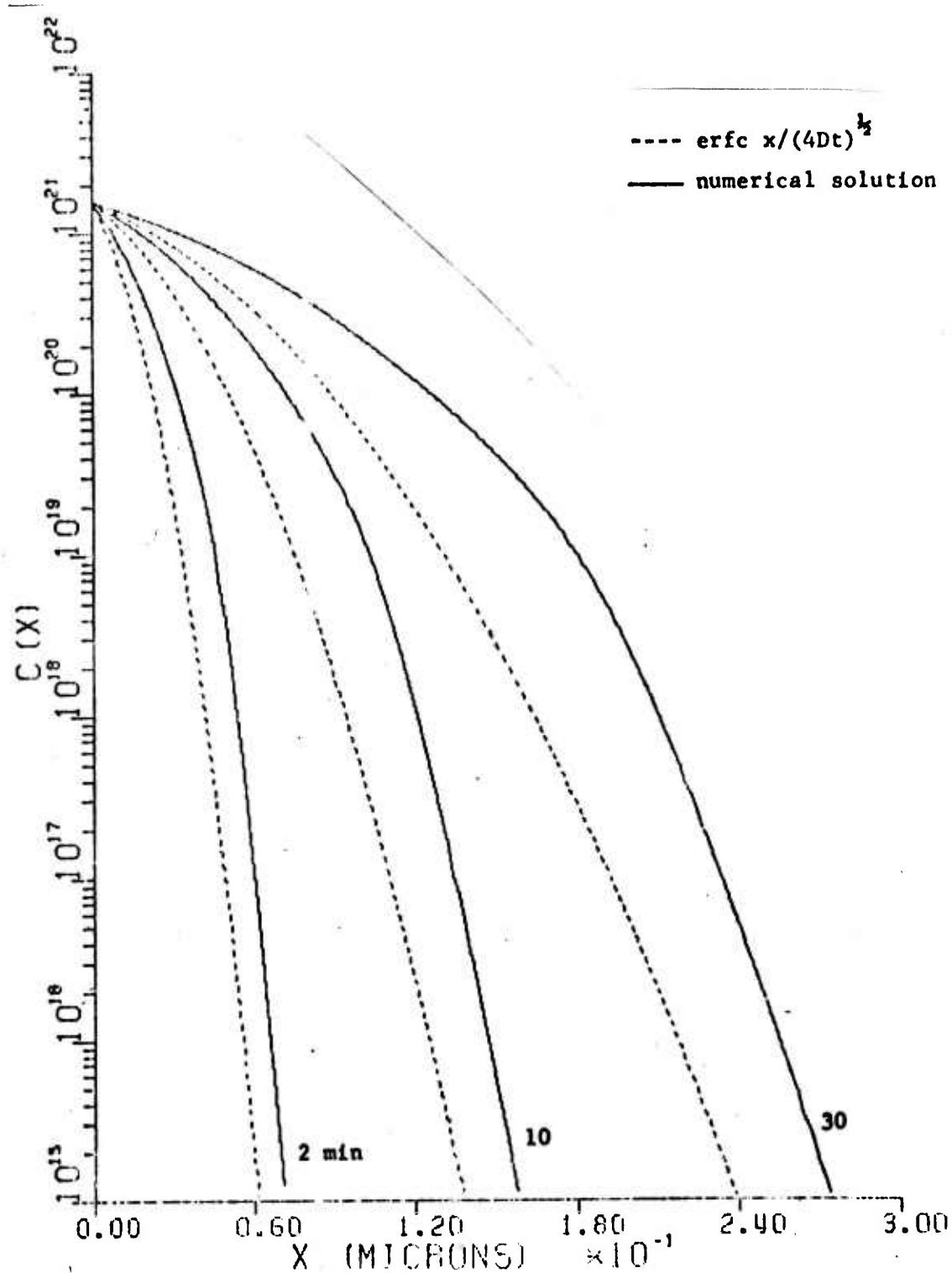


Figure XI-2. Diffusion profiles of arsenic in silicon at 1050°C with $C_0 = 1.6 \times 10^{21}$.

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 0.005 μ . 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 $\Delta t = 0.005 \text{ s}$ until 10 seconds and $\Delta x = 0.001 \mu$ and $\Delta t = 0.5 \text{ s}$ 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 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 program developed for calculating impurity profiles for a gaseous diffusion process will be extended to include a drive-in process where diffusion occurs from an initial distribution. Thus, the redistribution of impurities from an ion implantation can be studied. These computer programs will 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, which indicates an exponential profile near the surface for high concentration. This discrepancy between theory and experiment must be due to using a diffusion model which includes only the field-aided effect. Thus, the numerical technique must be generalized to include a variable diffusion coefficient so that excess vacancy generation and plastic deformation effects can be incorporated into the diffusion model.

References Cited in This Report - Part I

Chapter I -- Two-Dimensional Engineering Models for MOS and Bipolar Transistor Operation.

- [1] M. S. Mock, *Solid State Electronics*, 16, 601 (1973).
- [2] D. P. Kennedy, private communication.
- [3] S. M. Sze, *Physics of Semiconductor Devices*, p. 26, Wiley, N.Y. (1969).
- [4] H. K. Gummel, *IEEE Trans. Electron Devices*, ED-11, 455 (1964).
- [5] D. P. Kennedy, *IEEE Trans. Electron Devices*, ED-22, 988 (1975).
- [6] H. L. Stone, *SIAM Jour. Numer. Anal.*, 5, 530 (1968).

Chapter II -- A One-Dimensional Model for MOSFET Operation.

- [1] H.K.J. Ihantola and J. L. Moll, *Solid State Electronics*, 7, 423 (1964).
- [2] C. T. Sah and H. C. Pao, *IEEE Trans. Electron Devices*, ED-13, 393 (1966).
- [3] R. H. Crawford, *MOSFET in Circuit Design*, McGraw-Hill Company, N.Y. (1967).
- [4] A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley and Sons, Inc., N.Y. (1967).
- [5] S. M. Sze, *Physics of Semiconductor Devices*, Wiley-Interscience, N.Y. (1969).
- [6] D. P. Kennedy and F. A. Lindholm, *Physical Models of MOSFET Devices*, Final Report, HDL-CR-75-193-1 (1975).
- [7] D. P. Kennedy and P. C. Murley, *IBM Jour. of Research and Development*, 17, 2 (1973).
- [8] M. B. Barron, *Solid State Electronics*, 15, 293 (1972).
- [9] R. M. Swanson and J. D. Meindl, *IEEE Journal of Solid-State Circuits*, SC-7, 146 (1972).

Chapter III - Equivalent Circuit Studies.

- [1] W. Shockley, *Bell System Tech. J.*, 28, 435 (1949).
- [2] D. L. Fraser, Jr., and F. A. Lindholm, "Violations of the Quasi-Static Approximation in Large-Signal MOSFET Models," to be published.
- [3] A. Rose, "La Photoconductivite," *L'Onde Electrique*, 34, 645-651 (1954).
- [4] J. J. Sparkes and R. Beaufoy, *Proc. IRE*, 45, 1740 (1957); *ATE J.*, 13, 310 (1957).

- [5] E. O. Johnson and A. Rose, *Proc. IRE*, 47, 407 (1959).
- [6] R. D. Middlebrook, *Proc. IRE*, 106, Part B, Suppl. No. 17, 887 (1959).
- [7] J. J. Ebers and J. L. Moll, *Proc. IRE*, 42, 1761 (1954);
J. L. Moll, *Proc. IRE*, 42, 1761 (1954).
- [8] C. T. Sah, *IEEE Trans. Electron Devices*, ED-11, 324 (1964).
- [9] F. A. Lindholm, *IEEE J. Solid-State Circuits*, SC-6, 250 (1971); and SC-7, 322 (1972).
- [10] E. M. Cherry, *IEEE Trans. Electron Devices*, ED-18, 1166 (1971).
- [11] J. R. Macdonald, *Solid-State Elec.*, 5, 11 (1962).
- [12] F. A. Lindholm and J. I. Arreola, "Network Representation of the Large-Signal Transient Response of MOS Transistors," *Late News Supplement of 1975 Int. Electron Devices Meeting*, paper 11.8, p. 6 (1975); and "Network Representation for Self-Consistent Transient Response of Electronic (and other) Devices, Including Multi-Terminal Effects," *7th Annual Pittsburgh Conference on Modeling and Simulation*, (April 1976).
- [13] J. Logan, *Bell System Tech. J.*, 50, 1105 (1971).
- [14] H. K. Gummel and H. C. Poon, *Bell System Tech. J.*, 49, 827 (1970).
- [15] D. J. Hamilton, F. A. Lindholm, and J. A. Narud, *Proc. IEEE*, 239 (1964).
- [16] C. Rosenberg, D. S. Gage, R. S. Caldwell and G. H. Hanson, *IEEE Trans. Nucl. Sci. (Special Issue on Nuclear Radiation Effects Conference)*, NS-10, 149 (1963).
- [17] J. G. Linvill, *Proc. IRE*, 46, 1141 (1958).
- [18] L. D. Miliman, W. A. Massena, and R. H. Dickhaut, "CIRCUS, digital computer program for transient analysis of electronic circuits - user's guide," *Harry Diamond Lab., Tech. Rep.*, 346-1, Jan. 1967.
- [19] J. G. Bowers and S. R. Sedore, *SCEPTRE: A Computer Program for Circuits and Systems Analysis*. Englewood Cliffs, N.J.: Prentice-Hall, 1971.
- [20] A. F. Malmberg, "NET-2 network analysis program - preliminary user's manual," *Harry Diamond Lab., Tech. Rep.*, May 1970.
- [21] T. E. Idleman, F. F. Jenkins, W. J. McCala and D. O. Pederson, *IEEE J. Solid State Circuits (Special Issue on Computer-Aided Circuit Analysis and Device Modeling)*, SC-6, 188-203 (1971).
- [22] B. R. Chawla, *ibid.*, (Corresp.), SC-6, 262-264 (1971).

- [23] J. G. Fossum, *Proc. IEEE (Lett.)*, 60, 756 (1972); *IEEE Trans. Electron Devices*, ED-20, 582 (1973).
- [24] F. A. Lindholm, *Transistor Circuit Models*, Solid-State Materials and Devices, Final Rep. AFCRL-TR-74-0044, 1974.
- [25] F. A. Lindholm, *IEEE Trans. Circuit Theory*, CT-18, 122 (1971).
- [26] P. Rohr and F. A. Lindholm, *IEEE J. Solid-State Circuits*, SC-10, 65 (1975).
- [27] See the categorized bibliography: A. H. Agajanian, *Solid-State Electronics*, 18, 917 (1975).
- [28] H. Shickman and D. A. Hodges, *IEEE J. Solid-State Circuits*, SC-3, 285 (1968).
- [29] D. Frohman-Bentchkowsky and L. Vadasz, *ibid.*, SC-4, 57 (1969).
- [30] D. J. Hamilton, F. A. Lindholm, and A. H. Marshak, *Principles and Applications of Semiconductor Device Modeling*, Holt, Rinehart and Winston, Inc., New York, 1971, Ch. 5.
- [31] J. E. Meyer, *RCA Review*, 32, 42 (1971).
- [32] R.S.C. Cobbold, *Theory and Applications of Field-Effect Transistors*, Wiley, New York, 1970, Ch. 8.
- [33] G. A. Armstrong and J. A. Magowan, *Electronics Letters*, 7, 282 (1971).
- [34] J. J. Kalinowski, *Proc. IEEE*, 60, 1000 (1972).
- [35] F. S. Jenkins, E. R. Lane, W. W. Lattin, and W. S. Richardson, *IEEE Trans. Circuit Theory*, CT-20, 649 (1973).
- [36] P. E. Gray, D. DeWitt, A. R. Boothroyd, and J. F. Gibbons, *Physical Electronics and Circuit Models of Transistors*, John Wiley and Sons, Inc., New York, 1965, pp. 214-222.
- [37] P. E. Gray and C. L. Searle, *Electronic Principles--Physics Models and Circuits*, John Wiley and Sons, Inc., (1969) pp. 293-296, 336-341.
- [38] C. T. Sah, *Solid-State Electronics*, 13, 1547 (1970); *Physica Status Solidi (a)*, 7, 713 (1971); *Electronics Letters*, 8, 88 (1972).

Chapter IV - Carrier Mobility in an MOSFET

- [1] H.K.J. Ihantola and J. L. Moll, *Solid State Electronics*, 7, 423 (1964).
- [2] R.S.C. Cobbold, *Theory and Application of Field-Effect Transistors*, Wiley-Interscience, N.Y. (1970).
- [3] J. R. Schrieffer, *Phys. Rev.*, 97, 641 (1955).
- [4] R. F. Greene, D. R. Frankl, and J. Zemel, *Phys. Rev.*, 118, 967 (1960).
- [5] R. F. Pierret and C. T. Sah, *Solid State Electronics*, 11, 279 (1968).

- [6] N.S.J. Murphy, F. Berz, and I. Flinn, *Solid State Electronics*, 12, 775 (1969).
- [7] J. R. Schrieffer, *Semiconductor Surface Physics*, (edited by R. A. Kingston), Univ. of Pennsylvania Press (1957).
- [8] A. B. Fowler, F. F. Fang, W. E. Howard and P. J. Stiles, *Phys. Rev. Letters*, 16, 901 (1966).
- [9] F. Stern, *Phys. Rev. Letters*, 21, 1687 (1968).
- [10] F. Stern, *Phys. Rev. B*, 5, 4891 (1972).
- [11] M. A. Green, F. D. King, and J. Shewchun, *Solid State Electronics*, 17, 551 (1974).
- [12] C. T. Hsing and D. P. Kennedy, *HDL-CR-75-193-1 Final Report*, October (1975).
- [13] F. B. Hildebrand, *Advanced Calculus for Applications*, p. 156, Prentice-Hall (1962).
- [14] F. B. Hildebrand, *Advanced Calculus for Applications*, p. 151, Prentice-Hall (1962).
- [15] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, U.S. Dept. of Commerce, Washington, D.C. (1964).
- [16] The methods used here are from Gnädingers and Talley, *Solid State Electronics*, 13, 1301 (1970).
- [17] S. H. Gould, *Variational Method for Eigenvalue Problems*, 2nd ed., p. 75, Univ. of Toronto Press, 1957.

Chapter V -- Integral Methods for the Analysis of Semiconductor Devices

- [1] H. K. Gummel, *IEEE Trans. Elec. Devices*, 11, 455 (1964).
- [2] M. S. Mock, *Solid State Elec.*, 16, 601 (1973).
- [3] H. H. Heimeir, *IEEE Trans. Elec. Devices*, 20, 708 (1973).
- [4] O. Manck and W. L. Engl, *IEEE Trans. Elec. Dev.*, 22, 339 (1975).
- [5] S. M. Sze, *Physics of Semiconductor Devices*, John Wiley and Sons, Inc., N.Y. (1969).
- [6] F. G. Tricomi, *Integral Equations*, p. 116, Interscience Publishers, Inc., N.Y. (1957).
- [7] F. G. Tricomi, *Integral Equations*, p. 117, Interscience Publishers, Inc., N.Y. (1957).
- [8] F. G. Tricomi, *Integral Equations*, p. 51, Interscience Publishers, Inc., N.Y. (1957).
- [9] G. E. Forsythe and W. R. Wasow, *Finite Difference Methods for Partial Differential Equations*, p. 246, John Wiley and Sons, Inc., N.Y. (1960).

- [10] S. P. Morgan and F. M. Smitz, *Bell Sys. Tel. J.*, 29, 1573 (1960).
- [11] D. P. Kennedy and R. R. O'Brien, *IBM J. Res. Dev.*, 11, May 1967.

Chapter VI - Test Pattern Model Using Monte Carlo Techniques

- [1] D. P. Kennedy, *Proc. IEEE*, 52, 5 (1964).
- [2] Irving, *Bell System Tech. J.*, 41, 387 (1962).
- [3] Donald E. Knuth, *The Art of Computer Programming*, "Semi-numerical Algorithms", Addison-Wesley (1971).

Chapter VII - Hot Carrier Mechanisms in Semiconductor Devices

- [1] A. S. Grove, *Physics and Technology of Semiconductor Devices*, John Wiley and Sons, Inc., N.Y. (1967).
- [2] S. K. Ghandi, *The Theory and Practice of Microelectronics*, John Wiley and Sons, Inc., N.Y. (1968).
- [3] S. M. Sze, *Physics and Semiconductor Devices*, McGraw-Hill Book Co., N.Y. (1969).
- [4] W. Van Roosbroeck, *Bell Sys. Tech. J.*, 29, 560 (1950).
- [5] S. Wang, *Solid State Electronics*, McGraw-Hill Book Co., N.Y. (1966).
- [6] W. Shockley, *Bell Sys. Tech. J.*, 30, 990 (1951).
- [7] E. J. Ryder, *Phys. Rev.*, 90, 766 (1953).
- [8] J. B. Gunn, *Progress in Semiconductors*, Vol. 2, p. 213, John Wiley & Sons, N.Y. (1957).
- [9] A. C. Prior, *J. Phys. Chem.*, 12, 175 (1959).
- [10] T. Morgan, *Bull. Am. Phys. Soc.*, 2, No. 2, 265 (1959).
- [11] E. M. Conwell, *Phys. Chem. Solids*, 8, 234 (1959).
- [12] W. A. Harrison, *Phys. Rev.*, 104, 1281 (1956).
- [13] D. Long, *Phys. Rev.*, 120, 2044 (1960).
- [14] C. Y. Duh and J. L. Moll, *Solid-State Elec.*, 11, 917 (1968).
- [15] G. C. Dacey, *Phys. Rev.*, 90, 759 (1953).
- [16] G. C. Dacey and I. M. Ross, *Bell Sys. Tech. J.*, 1149 (1955).
- [17] S. R. Hofstein and G. Warfield, *IEEE Trans. on Electron Devices*, 129, March (1965).
- [18] L. K. Makeshivari and S. N. Ihanwar, *Int. J. Electronics*, Vol. 37, No. 3, 435 (1974).

- [19] R. Stratton, *J. Appl. Phys.*, 40, No. 11, 4582 (1969).
- [20] R. Stratton, *J. Appl. Phys.*, 38, No. 12, 4596 (1967).
- [21] R. Stratton, *Phys. Rev.*, 126, 2002 (1962).
- [22] Colman Goldberg, *J. Appl. Phys.*, 40, No. 11, 4612 (1969).
- [23] T. Y. Stokoe and J. E. Parrott, *Solid State Elec.*, 17, 477 (1974).
- [24] T. Y. Stokoe and J. E. Parrott, *ibid.*, 18, 811 (1975).
- [25] R. Stratton, *IEEE Trans. Elec. Dev.*, ED-19, No. 12 (1972).
- [26] J. D. Gassaway, *IEEE Trans. Elec. Dev.*, ED-18, 175 (1971).
- [27] A. H. Marshak and D. Assaf, *Solid State Elec.*, 16, (1973).
- [28] B. R. Nag and A. N. Chakravarti, *ibid.*, 18, (1975).
- [29] E. M. Conwell, *High Field Transport in Semiconductors*. In *Solid State Physics*, Supplement No. 9, Academic Press, N.Y. (1967).

Chapter VIII -- Ion Implantation

- [1] W. K. Hofke, et al., *Applied Physics*, 2, 1973, pp. 265-278.
- [2] J. R. Anderson and J. F. Gibbons, "A New Model for Boron Diffusion in Silicon" (to be published).
- [3] A. C. Hindmarsh, "Solution of Ordinary Differential Equations with Banded Jacobian," Lawrence Livermore Lab., University of California, Livermore, Calif., March 1975.
- [4] Richtmeyer and Morton, *Difference Methods for Initial Value Problems*, Chapter 8.
- [5] M.J.D. Powell, *The Computer J.*, 7, July 1965, pp. 155-162; W. J. Zangwill, *The Computer J.*, 10, 3, November 1967; Gill, Murray, and Pitfield, N.P.L., "Minimize a Real Function of N Real Variables Using a Quasi-Newton Method--Derivatives Not Required," (Subprogram in Stanford Library).
- [6] R. A. Moline, G. W. Reutlinger, and J. C. North, in *Atomic Collisions and Solids 1*, ed. by S. Datz, B. R. Appertson, and C. D. Moak, Plenum Press, New York, 1975.
- [7] R. A. Moline and A. G. Cullis, *Appl. Phys. Lett.*, 26, 551 (1975).
- [8] W. K. Chu, H. Müller, and J. W. Mayer, *Appl. Phys. Lett.*, 25, 297 (1974).
- [9] W. Fahrner and A. Goetzberger, *Appl. Phys. Lett.*, 21, 329 (1972).
- [10] W. Fahrner, *J. Electrochem. Soc.*, 121, 784 (1974).

- [11] R. Castagne, *Compt. Rend, Acad. Sci., Paris* 267, Series B, 866 (1968).
- [12] M. Kuhn, *Solid State Electronics*, 13, 873 (1970).
- [13] J. F. Gibbons, W. S. Johnson, and S. W. Mylroie, *Projected Range Statistics, Semiconductors and Related Materials*, 2nd Ed., Halsted Press Division of John Wiley and Sons, Somerville, N.J.
- [14] J. Lindhart, M. Scharff, and H. E. Schiott, *Mat. Fys. Medd.*, 33, 14 (1963).
- [15] S. Furukawa and H. Ishiwara, *J. Appl. Phys.*, 43, 3, March, 1972, pp. 1268-1273.

Chapter IX -- Thermal Oxidation

- [1] B. E. Deal and A. S. Grove, *J. Appl. Phys.*, 36, December 1965, pp. 3770-3778.
- [2] B. E. Deal and M. Skalar, *J. Elec. Soc.*, 112, April 1965, pp. 430-435.
- [3] A. S. Grove, O. Leistiko, and C. T. Sah, *J. Appl. Phys.*, 35, 9, September 1965, pp. 2695-2701.
- [4] G. Baccarani, M. Severi, and C. Soncini, *J. Electrochem. Soc.*, 120, 1973, p. 1436.
- [5] K. Hirabayashi and J. Iwamura, *J. Electrochem. Soc.*, 120, 1973, p. 1595.
- [6] R. J. Kriegler, Y. G. Cheng, and D. R. Colton, *J. Electrochem. Soc.*, 119, 1972, p. 388.
- [7] Y. J. Van der Meulen, C. M. Osburn, and J. F. Ziegler, *J. Electrochem. Soc.*, 122, 1975, p. 284.

Chapter X -- Silicon Epitaxy

- [1] F. C. Eversteyn et al., *J. Electrochem. Soc.*, 117, 7, July 1970, pp. 927-931.
- [2] E. G. Bylander, *J. Electrochem. Soc.*, 109, 12, December 1962, pp. 1171-1175.
- [3] R.F.C. Farrow, *J. Electrochem. Soc.*, 121, 7, July 1974, pp. 899-907.
- [4] J. Bloem, *J. Crystal Growth*, 13/14, 1972, pp. 302-305.
- [5] J. Bloem, *Semiconductor Silicon 1973*, op. cit., pp. 213-226.
- [6] B. A. Joyce and R. R. Bradley, *J. Electrochem. Soc.*, 110, 12, December 1963, pp. 1235-1240.
- [7] P. F. Deisler, Jr. and R. H. Wilhelm, *Ind. and Eng. Chemistry*, 45, 1953, p. 1219.
- [8] J. B. Rosen and W. E. Winsche, *J. Chem. Phys.*, 18, 1950, p. 1587.

- [9] H. Kobayashi and M. Kobayashi, *Catalysis Reviews--Science and Engineering*, 10, 1974, p. 139.
- [10] D. P. Kennedy and R. R. O'Brien, *IBM J. Res. Develop.*, 13, 1969, p. 212.
- [11] T. Ishii et al., *J. Electrochem. Soc.*, 122, 1975, p. 1523.
- [12] P. H. Langer and Joseph I. Goldstein, *J. Electrochem. Soc.*, 121, 4, April 1974, pp. 563-571.
- [13] J. R. Brews, *J. Appl. Phys.*, 44, July 1973, pp. 3228-3231.
- [14] M. G. Buehler, *IEEE Trans. on Electron Devices*, ED-19, November 1972, pp. 1171-1178.
- [15] K. Ziegler, E. Klausmann, and S. Kar, *Solid State Electronics*, 18, February 1975, pp. 189-198.

Appendix A
Nonequilibrium Solution of Poisson's Equation
in One Spatial Dimension

A. Solution for $Q_i(V_s)$

In the source-drain inversion layer of a MOSFET, the electron distribution must satisfy a Poisson equation of the form

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = \frac{-1}{\kappa_s \epsilon_0} \rho, \quad (A-1)$$

where ρ represents the total density of electrostatic charges (ions, holes, and electrons) residing within this semiconductor material. Near the source end of this structure, $\partial^2 V / \partial x^2 \ll \partial^2 V / \partial y^2$ and, therefore, (A-1) has the approximate form

$$\frac{d^2 V}{dy^2} = \frac{-1}{\kappa_s \epsilon_0} \rho. \quad (A-2)$$

Assuming a p-type semiconductor substrate, the electrostatic charge density can be written

$$\rho = -q(N_A - p + n). \quad (A-3)$$

Because we have a source-drain current through this inversion layer, this semiconductor structure is in nonequilibrium. Therefore, the densities of electrons and holes within this structure cannot be described by equilibrium distribution functions. However, assuming that we have a relatively small perturbation from equilibrium, we can approximate these densities by "quasi-equilibrium" distributions of the form

$$n = n_i \exp \{q(V - \phi_n) / kT\} \quad (A-4)$$

$$p = n_i \exp \{-q(V - \phi_p) / kT\}, \quad (A-5)$$

where ϕ_n and ϕ_p are quasi-Fermi potentials defined by

$$\phi_n \equiv V - \frac{kT}{q} \log_e (n / n_i) \quad (A-6)$$

$$\phi_p \equiv V + \frac{kT}{q} \log_e (p / n_i). \quad (A-7)$$

A physically meaningful interpretation of these quasi-Fermi potentials can be made by introducing the substitutions

$$n_q = n_i \exp(-q\phi_n/kT) \quad (A-8)$$

$$p_q = n_i \exp(q\phi_p/kT) \quad , \quad (A-9)$$

whereby (A-4) and (A-5) have the simplified form

$$n = n_q \exp(qV/kT) \quad (A-10)$$

$$p = p_q \exp(-qV/kT) \quad . \quad (A-11)$$

From (A-10) and (A-11), we see that n_q and p_q represent nonequilibrium densities of electrons and holes, respectively, in charge neutral regions of this semiconductor material. Thus, from (A-8) and (A-9), these quasi-Fermi potentials account for modified carrier densities in charge-neutral semiconductor material under nonequilibrium conditions. Substituting (A-10) and (A-11) into (A-3) yields

$$\rho = -q[N_A - p_q e^{-\beta V} + n_q e^{\beta V}] \quad . \quad (A-12)$$

Since $d^2V/dy^2 = -dE_y/dy$, (A-2) can be written

$$\frac{dE_y}{dy} = \frac{1}{\kappa_s \epsilon_o} \rho \quad (A-13)$$

and, therefore,

$$dE_y = \frac{1}{\kappa_s \epsilon_o} \rho dy \quad . \quad (A-14)$$

Further, since $E_y = -dV/dy$, we have

$$E_y dE_y = \frac{-1}{\kappa_s \epsilon_o} \rho dV \quad . \quad (A-15)$$

Substituting (A-12) into (A-15), we obtain

$$E_y dE_y = \frac{q}{\kappa_s \epsilon_o} [N_A - p_q e^{-\beta V} + n_q e^{\beta V}] dV \quad . \quad (A-16)$$

Equation (A-16) can be integrated from the oxide-semiconductor interface to a point within the substrate:

$$\int_{E_{ys}}^{E_y} n dn = \frac{q}{\kappa_s \epsilon_o} \int_{V_s}^V [N_A - p_q e^{-\beta \xi} + n_q e^{\beta \xi}] d\xi. \quad (A-17)$$

In (A-17), E_{ys} and V_s represent the magnitudes of the electric field and electrostatic potential, respectively, at the oxide-semiconductor interface. Thus, from (A-17),

$$E_y^2 = E_{ys}^2 + \frac{2q}{\kappa_s \epsilon_o} \{ N_A (V - V_s) + \frac{kT}{q} p_q (e^{-\beta V} - e^{-\beta V_s}) + \frac{kT}{q} n_q (e^{\beta V} - e^{\beta V_s}) \} . \quad (A-18)$$

In charge neutral regions of the substrate, $E_y=0$ and $V=0$; therefore, from (A-18)

$$0 = E_{ys}^2 - \frac{2q}{\kappa_s \epsilon_o} \{ N_A V_s + \frac{kT}{q} p_q (e^{-\beta V_s} - 1) + \frac{kT}{q} n_q (e^{\beta V_s} - 1) \} . \quad (A-19)$$

Further, from (A-12), space-charge neutrality requires that

$$0 = \rho = -q(N_A - p_q + n_q) \quad (A-20)$$

and, therefore

$$p_q = N_A + n_q . \quad (A-21)$$

At the oxide-semiconductor interface, a one-dimensional application of Gauss' law yields

$$E_{ys} = \frac{\kappa_i}{\kappa_s} \left(\frac{V_G - V_s}{t_{ox}} \right) . \quad (A-22)$$

Substituting (A-21) and (A-22) into (A-19) and thereafter solving for n_q , we obtain

$$n_q(V_s) = \frac{\frac{\kappa_s \epsilon_o}{2q} \left[\frac{\kappa_i}{\kappa_s} \left(\frac{V_G - V_s}{t_{ox}} \right) \right]^2 - N_A \left[V_s + \frac{kT}{q} (e^{-\beta V_s} - 1) \right]}{\frac{kT}{q} \left[(e^{-\beta V_s} - 1) + (e^{\beta V_s} - 1) \right]} . \quad (A-23)$$

In a similar fashion, substituting (A-21) and (A-22) into (A-18) and, thereafter, solving for E_y yields

$$E_y(V, V_s) = \left\{ \left[\frac{\kappa_i}{\kappa_s} \left(\frac{V_G - V_s}{t_{ox}} \right) \right]^2 + \frac{2q}{\kappa_s \epsilon_0} \left[N_A (V - V_s) + \frac{kT}{q} (N_A + n_q) \right. \right. \\ \left. \left. \cdot (e^{-\beta V} - e^{-\beta V_s}) + \frac{kT}{q} n_q (e^{\beta V} - e^{\beta V_s}) \right] \right\}^{1/2}, \quad (A-24)$$

where n_q is given by (A-23).

To calculate the inversion charge, $Q_i(V_s)$, we integrate the electron density from the oxide-semiconductor interface (where $y=0$) to a depth y_i (where $n=n_i$) within this semiconductor material:

$$Q_i = -q \int_0^{y_i} n(y) dy. \quad (A-25)$$

Since $E_y = -dV/dy$, (A-25) can be rewritten

$$Q_i(V_s) = -q \int_{V_i}^{V_s} \frac{n}{E_y(V, V_s)} dV, \quad (A-26)$$

where $V_i \equiv V(y_i)$. By definition, $n=n_i$ when $V=V_i$; therefore, from (A-10),

$$n_i = n_q \exp(qV_i/kT). \quad (A-27)$$

Solving (A-27) for V_i yields

$$V_i = \frac{kT}{q} \log_e (n_i/n_q). \quad (A-28)$$

Substituting (A-10) into (A-26), we obtain

$$Q_i(V_s) = -q \int_{V_i}^{V_s} \frac{n_q(V_s) \exp(qV/kT)}{E_y(V, V_s)} dV. \quad (A-29)$$

B. Solution for dQ_i/dV_s

Differentiating (A-29) with respect to V_s yields

$$\begin{aligned} \frac{dQ_i}{dV_s} = -q \int_{V_i}^{V_s} \exp(qV/kT) \left\{ n_q(V_s) \left[\frac{-dE_y/dV_s}{E_y^2(V, V_s)} \right] + \frac{1}{E_y(V, V_s)} \left[\frac{dn_q}{dV_s} \right] \right\} dV \\ -q \left(\frac{kT}{q} \right) \left[\frac{n_i}{n_q(V_s)} \right] \frac{dn_q}{dV_s} \left[E_y(V_i, V_s) \right]^{-1} \\ -q \left(n_q(V_s) \left[\exp(qV_s/kT) \right] E_{ys}^{-1} \right), \end{aligned} \quad (A-30)$$

where

$$\frac{dn_q}{dV_s} = \frac{-2n_q(V_s) \sinh(\beta V_s) - (C_{ox}/q) E_{ys}^{-N_A} (1 - e^{-\beta V_s})}{2 \left(\frac{kT}{q} \right) [\cosh(\beta V_s) - 1]} \quad (A-31)$$

$$\begin{aligned} \frac{dE_y}{dV_s} = \left(\frac{1}{2} \right) \frac{1}{E_y(V, V_s)} \left\{ \left(\frac{-2\kappa_i}{\kappa_s t_{ox}} \right) E_{ys} \right. \\ \left. + \frac{2q}{\kappa_s \epsilon_o} \left[N_A (e^{-\beta V_s} - 1) - 2n_q \sinh(\beta V_s) \right] \right. \\ \left. + 2 \left(\frac{kT}{q} \right) \frac{dn_q}{dV_s} [\cosh(\beta V) - \cosh(\beta V_s)] \right\} \end{aligned} \quad (A-32)$$

and where E_{ys} , n_q , and E_y are given by (A-22), (A-23), and (A-24), respectively.

C. Derivation of V_{sc}

From (A-10), the electron density at the oxide semiconductor interface, n_s , is given by

$$n_s = n_q(V_s) \exp(qV_s/kT) \quad (A-33)$$

and, therefore,

$$V_s = \frac{kT}{q} \log_e \left[\frac{n_s}{n_q(V_s)} \right] . \quad (A-34)$$

When V_s attains a critical magnitude, V_{sc} , we have a negligible density of electrons within the inversion layer ($n_s = n_i$):

$$V_{sc} = \frac{kT}{q} \log_e \left[\frac{n_i}{n_q(V_{sc})} \right] . \quad (A-35)$$

Substituting (A-23) into (A-35) yields (A-36)

$$V_{sc} = \frac{kT}{q} \log_e \left\{ \frac{n_i \left(\frac{kT}{q} \right) \left[(e^{-\beta V_{sc}} - 1) + (e^{\beta V_{sc}} - 1) \right]}{\left[\frac{\kappa_s \epsilon_o}{2q} \left[\frac{\kappa_i}{\kappa_s} \left(\frac{V_G - V_{sc}}{t_{ox}} \right) \right]^2 - N_A \left[V_{sc} + \frac{kT}{q} (e^{-\beta V_{sc}} - 1) \right] \right]} \right\}$$

Clearly, (A-36) has the approximate form (A-37)

$$V_{sc} = \frac{kT}{q} \log_e \left\{ \frac{e^{\beta V_{sc}}}{\frac{1}{n_i} \left(\frac{q}{kT} \right) \left\{ \frac{\kappa_s \epsilon_o}{2q} \left[\frac{\kappa_i}{\kappa_s} \left(\frac{V_G - V_{sc}}{t_{ox}} \right) \right]^2 - N_A \left[V_{sc} - \frac{kT}{q} \right] \right\}} \right\}$$

Solving (A-37) for V_{sc} yields (A-38)

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

DISTRIBUTION LIST

	<u># of Copies</u>
Defense Documentation Center ATTN: DDC-TCA Cameron Station (Bldg. 5) Alexandria, VA 22314	12
Office of Naval Research Code 427 Arlington, VA 22217	1
Naval Ships Engineering Ctr. ATTN: Code 6157D Prince Georges Center Hyattsville, MD 20782	1
Commander Naval Electronics Lab Ctr. ATTN: Library San Diego, CA 92152	1
Commander US Naval Ordnance Lab ATTN: Technical Library White Oak Silver Spring, MD 20910	1
Rome Air Development Ctr. ATTN: Documents Library (TILD) Griffiss AFB, NY 13441	1
HQ ESD (XRRI) L. G. Hanscom Field Bedford, MA 01730	1
OSASS-RD Washington, DC 20310	1
Commander US Army Missile Command Redstone Scientific Info Ct. ATTN: Chief, Document Section Redstone Arsenal, AL 35809	2
Director Defense Communication Agency Technical Library Center Code 205 Washington, DC 20305	1

	<u># of Copies</u>
Commander Harry Diamond Laboratories ATTN: Library 2800 Powder Mill Road Adelphi, MD 20783	1
Commander USA Mobility Eqmt R&D Cmd. ATTN: SMEFB-R Fort Belvoir, VA 22060	2
Commander US Army Materiel Development and Readiness Command ATTN: DRCRD-R 5001 Eisenhower Ave. Alexandria, VA 22333	1
NASA Scientific & Tech Info Facility ATTN: Acquisitions Br (S-AK/DL) PO Box 33 College Park, MD 20740	2
Advisory Gp on Electron Devices 201 Varick Street, 9th Floor New York, NY 10014	2
Director Naval Research Laboratory ATTN: Code 2627 Washington, DC 20375	1
Director of Defense Res & Engrg. ATTN: Technical Library Rm 3E-1039, The Pentagon Washington, DC 20301	1
USA Security Agency Combat Dev Activity ATTN: IACDA-P(T) Arlington Hall Station, Bldg 420 Arlington, VA 22212	1
Reliability Analysis Center RADC (RBRAC) ATTN: I. L. Krulac Griffiss AFB, NY 13441	1

	<u># of Copies</u>
Naval Research Laboratory ATTN: L. J. Palkutti Code 5216 Washington, DC 20375	1
Bell-Northern Research IC Design Aids ATTN: D. M. Cauchey PO Box 3511, Station C Ottawa, Ontario K1Y4H7 Canada	1
Commander Air Force Avionics Lab ATTN: Dr. Ronald Belt AFAL/TEA-3 Wright Patterson AFB, OH 45433	1
Brown Engineering Cummings Research Park ATTN: Mr. Leonard Owen MS 15 Huntsville, AL 35807	1
Fairchild Semiconductor ATTN: Dr. G. Amelio 4001 Miranda Avenue MIS 30-0213 R&D Palo Alto, CA 94304	1
Commander Naval Research Laboratory ATTN: Dr. D. F. Barbe Washington, DC 20375	1
Rockwell International Electronics Research Div. ATTN: B. T. French MS HB27 3370 Miraloma Ave. Anaheim, CA 92803	1
Westinghouse Defense & Electronics Systems Center ATTN: Dr. Marvin White PO Box 746 Baltimore, MD 21203	1

	<u># of Copies</u>
Honeywell Systems & Research Center ATTN: Dr. G. Weber 2600 Ridgway Parkway Minneapolis, MI 55413	1
Reticon Corp. ATTN: Dr. E. Snow 910 Benica Avenue Sunnyvale, CA 94040	1
Raytheon Missile Sys. Div. ATTN: Mr. Howard Klemmer Hartwell Road Bedford, MA 01730	1
General Electric Corporate Res. & Dev. ATTN: Dr. J. Tiemann Schenectady, NY 12301	1
Institute of Defense Analysis Arlington, VA 22209	1
Mr. Jack Kilby 5924 Royal Lane Suite 150 Dallas, TX 75230	1
Dr. Gordon E. Moore Intel Corporation 3065 Bowers Road Santa Clara, CA 95951	1
Commander Harry Diamond Laboratories ATTN: Mr. A. J. Baba 2800 Powder Mill Road Adelphi, MD 20783	1
Naval Electronics Systems Cmd. ATTN: Mr. L. W. Sumney (Code 3042) Washington, DC 20360	1
Commander Naval Electronic Laboratory Ctr ATTN: Mr. C. E. Holland, Jr. Code 4300 271 Catalina Blvd. San Diego, CA 92152	1

	<u># of Copies</u>
Air Force Avionics Lab ATTN: Mr. H. H. Steenbergen Wright-Patterson AFB, OH 45433	1
Commander Rome Air Development Center ATTN: Mr. J. Brauer (RBRM) Griffiss AFB, NY 13441	1
Dr. Gerald B. Herzog Solid-State Technology Center RCA David Sarnoff Research Ctr Princeton, NJ 08540	1
Dr. George E. Smith Bell Telephone Labs, Inc. Room 2A-323 Murray Hill, NJ 07974	1
Mr. Harold D. Toombs Texas Instruments, Inc. P.O. Box 5012 Dallas, TX 75222	1
Dr. Barry Dunbridge TRW Systems Group One Space Park Redondo Beach, CA 90278	1
Commander ATTN: DRSEL-MS-TI Fort Monmouth, NJ 07703	2
Commander ATTN: DRSEL-TL-DT Fort Monmouth, NJ 07703	1
Commander ATTN: DRSEL-PL-ST Fort Monmouth, NJ 07703	1
Commander ATTN: DRSEL-TL-I Fort Monmouth, NJ 07703	1
Commander ATTN: DRSEL-TL-I (CPC File) Fort Monmouth, NJ 07703	3
Commander ATTN: DRSEL-TL-IG (Reitmeyer) Fort Monmouth, NJ 07703	1