Research and Development Technical Report

DELET-TR-77-2684

# COMPUTER-AIDED ENGINEERING OF SEMICONDUCTOR INTEGRATED CIRCUITS

prepared by

J.D. Meindl
R.W. Dutton
J.F. Gibbons
J.D. Plummer
W.A. Tiller
K.C. Saraswat
B.E. Deal
T.I. Kamins

INTEGRATED CIRCUITS LABORATORY
STANFORD UNIVERSITY
Stanford, Ca. 94305

July 1979

Final Report for Period 1 March 1978-28 February 1979

Distribution Statement
Approved for public release;
distribution unlimited.

Prepared for:
ADVANCED RESEARCH PROJECTS AGENCY
1400 Wilson Boulevard
Arlington, Va. 22209

US Army Electronics Technology & Devices Laboratory

**ERADCOM**

U S ARMY ELECTRONICS RESEARCH AND DEVELOPMENT COMMAND
FORT MONMOUTH, NEW JERSEY 07703

AD A117982

DTIC FILE COPY

DTIC ELECTE AUG 9 1982 S D A

82 08 09 096

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>DELET-TR-77-2684-F | 2. GOVT ACCESSION NO.<br>AD-A117982 | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>Computer-Aided Engineering of Semiconductor Integrated Circuits | | 5. TYPE OF REPORT & PERIOD COVERED<br>Final Report<br>1 March 1978 - 28 February 1979 |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s) J.D.Meindl, J.F.Gibbons, L.A.Christel, B.E.Deal, C.P.Ho, H.Z.Massoud, J.D.Plummer, R.R.Razouk, W.A.Tiller, K.C.Saraswat, R.Reif, T.I.Kamins, M.M.Mandurah, F.Mohammadi, (over) | | 8. CONTRACT OR GRANT NUMBER(s)<br>DAAB07-77-C-2684<br>P00003 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Integrated Circuits Laboratory<br>Stanford University<br>Stanford, Ca. 94305 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>6.1 1T1 61101 A 00A 00<br>690000.211.13.27.00 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Advanced Research Projects Agency ( ARPA )<br>1400 Wilson Blvd.<br>Arlington, Virginia 22209 | | 12. REPORT DATE<br>June 1979<br>13. NO. OF PAGES |
| | | 15. SECURITY CLASS. (of this report)<br>UNCLASSIFIED |
| 14. MONITORING AGENCY NAME & ADDRESS (if diff. from Controlling Office)<br>ERADCOM<br>ATTN: DELET-IS<br>Fort Monmouth, New Jersey 07703 | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release: distribution unlimited.

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

18. SUPPLEMENTARY NOTES

This contract is awarded under Basic Agreement N00014-76-A-0029 dated 18 Nov.'75 as amended by modification P00001 & P00002, issued by the Office of Naval Research. This Research is Sponsored by Defense Advanced Research Projects Agency (DARPA)

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

MODELING, SEMICONDUCTOR DEVICES, SEMICONDUCTOR PROCESSING, COMPUTER DESIGN CIRCUITS CIRCUIT FABRICATION PROCESSES

ION IMPLANTATION, THERMAL OXIDATION, CHEMICAL VAPOR DEPOSITION, DEVICE SIMULATION

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

Economical procurement of small quantities of high performance custom integrated circuits for military systems is severely impeded by inadequate process, device and circuit models that handicap accurate computer-aided design at low cost. The salient objective of this program is to formulate physical models of fabrication processes, devices and circuits to allow total computer-aided design of custom large-scale integrated circuit subsystems to reduce development cycle time and cost. The basic areas under investigation are: (1) ion implantation and diffusion of dopants, (2) thermal oxidation, (3) chemical vapor deposition of silicon, (4) device simulation and (5) analytical measurements. (Continued other side)

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19. KEY WORDS (Continued)

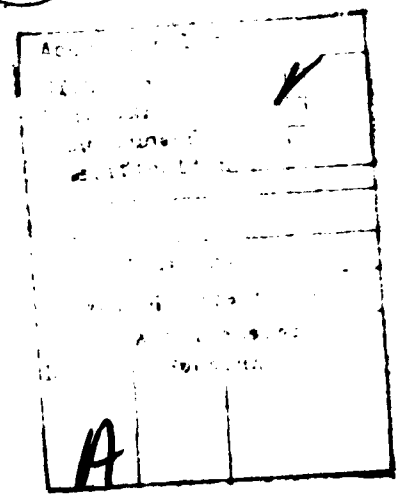
20 ABSTRACT (Continued)

7. authors

R.W.Dutton, D.C.D'Avanzo, M.Vanzi, J.A.Greenfield, S.Y.Oh,
C.H.Price, H.G.Lee, D.B.Estreich, D.Ward, D.Divekar, W.J.McCalla

20. This report provides a detailed development of research results in developing computer-aids for modeling the IC fabrication process. The effort is vertically integrated and the outcome is the widely distributed SUPREM program. Moreover device analysis capabilities have been developed based on SUPREM and applications have been demonstrated. The impact of this effort on the design of custom VLSI is apparent and future research will explore, define and model the limits of technology for VLSI.

DTIC
COPY
INSPECTED
2



TR DXG569

ICL 7-77

# SECOND

# INTERIM REPORT

# ON

# COMPUTER-AIDED ENGINEERING OF

# SEMICONDUCTOR INTEGRATED CIRCUITS

1 March 1978 -- 28 February 1979

This work was supported by
a contract from the

ADVANCED RESEARCH PROJECTS AGENCY

DAAB07-77-C-2684

ARPA Order No. 2985

Approved for public release;
distribution unlimited.

J.D. Meindl, Director
Stanford Electronics Laboratories
Stanford University Stanford, California

TABLE OF CONTENTS


| | Page |
|---|---|
| INTRODUCTION | 1 |
| PART 1 - ION IMPLANTATION | 4 |
| 1.1 Implantation of Ions into Layered Targets | 4 |
| References | 18 |
| PART 2 - THERMAL OXIDATION | 19 |
| 2.1 Introduction | 19 |
| 2.2 Summary of Principal Accomplishments to Date | 23 |
| 2.3 A "Working Model" for the Growth of $SiO_2$ Layers on Silicon | 27 |
| 2.4 Detailed Discussion of Results | 36 |
| References | 88 |
| PART 3 - CHEMICAL VAPOR DEPOSITION OF SILICON | 124 |
| 3.1 Introduction | 124 |
| 3.2 A Model for Dopant Incorporation into Growing Silicon Epitaxial Films | 127 |
| 3.3 Low-Pressure Chemical-Vapor-Deposition of Silicon: Phosphorus Doping | 170 |
| 3.4 Silicides: Annealing of Sputtered $WSi_2$ | 180 |
| References | 189 |
| Appendix A | 192 |
| Appendix B | 194 |
| PART 4 - DEVICE SIMULATION AND STATISTICAL CIRCUIT MODELING | 223 |
| 4.1 Introduction | 223 |
| 4.2 One Dimensional SEmiconductor Device ANalysis (SEDAN) | 226 |
| Appendix 4.2.1 | 251 |

TABLE OF CONTENTS (cont'd) Page


Appendix 4.2.2 . . . . . 256

4.3 Two Dimensional Modeling of Nonplanar Devices on Minicomputers. . . . . 270

4.4 A Simplified Two Dimensional Numerical Analysis of MOS Device - DC Case . . . . . 301

Appendix 4.4.1 . . . . . 316

Appendix 4.4.2 . . . . . 318

Appendix 4.4.3 . . . . . 324

4.5 Two Dimensional Grid Generation for Semiconductor Device Simulation . . . . . 342

4.6 Test Structures for Two-Dimensional Profile Measurements . . . 353

4.7 Latch-Up in CMOS Integrated Circuits . . . . . 374

4.8 A Charge Oriented MOS Capacitance Model. . . . . 387

4.9 DC Statistical Circuit Analysis for Bipolar IC's Using Parameter Correlations -- An Experimental Example. . . . . 400

References. . . . . 419

Appendix I. . . . . 424

# COMPUTER AIDED ENGINEERING OF SEMICONDUCTOR INTEGRATED CIRCUITS

## Introduction

Custom Integrated Circuits for small volume production of electronic equipment is of primary value to the military although the importance of custom circuitry is growing rapidly for commercial applications. The impact of the custom design approach on electronics is growing primarily because of two factors:

1) the reliability of monolithic systems compared with MSI circuitry is well known,
2) the performance of architectures specialized to a given application can be increased many fold compared with general purpose electronics.

The major draw-back of custom LSI is that the design cycle is expensive and it is difficult to insure proper functionality and reliability with a single pass design. Computer-aided design is now recognized as the only viable means to cope with the growing complexity of custom LSI. Research efforts reported here focus on the technology controlled aspects of VLSI design. The purpose of this report is to summarize results of a vertically integrated program to predict directly the performance and reliability limiting factors of VLSI devices directly from the processing steps used in the IC fabrication process.

The backbone of this effort is a multidiscipline effort to understand and model IC fabrication processes from first principles. The first three sections of this report present results of ongoing research to develop models of the ion implantation, oxidation and chemical vapor deposition processes used in silicon IC fabrication. In each of these

sections the consolidated efforts of many individuals is evident. Figure 1 shows schematically how measurement technologies, industrial collaboration and material science personnel impact each subgroup effort. While the greatest number of the research personnel come from the Integrated Circuits Laboratory, the efforts in ion implantation and measurements technology are spearheaded by members of the Solid State Laboratory. As noted, the deepest roots of this effort come from material science and the impact of this discipline on the modeling effort is pivotal.

The cutting edge of computer-aided process and device modeling is the SUPREM program and its continued evolution based on new and improved process models. Figure 1 shows the confluence of the subgroup activities as inputs to SUPREM and the output is the program distribution to more than 114 government agencies, contractors, industrial groups and universities. The wide distribution and use of SUPREM continues to provide this effort with feedback. Moreover a one-day seminar held annually at Stanford provides a forum for review of the overall program. Appendix I contains a list of current SUPREM users, typical feedback comments and the program for this years review seminar.

The viability of process modeling for IC design hinges on prediction of device performance. The final subsection of this report provides a comprehensive review of device modeling results based on SUPREM as the cornerstone. SEmiconductor Device ANalysis both for one-dimension (SEDAN) and for Two-dimensional ANalysis (TANDEM) is described. Moreover, the applications of process and device analysis to test structures and circuit simulator model development are presented.

In summary, this report provides a detailed development of research results in developing computer-aids for modeling the IC fabrication process. The effort is vertically integrated as described in Figure 1 and the outcome is the widely distributed SUPREM program. Moreover device analysis capabilities have been developed based on SUPREM and applications have been demonstrated. The impact of this effort on the design of custom VLSI is apparent and future research will explore, define and model the limits of technology for VLSI.

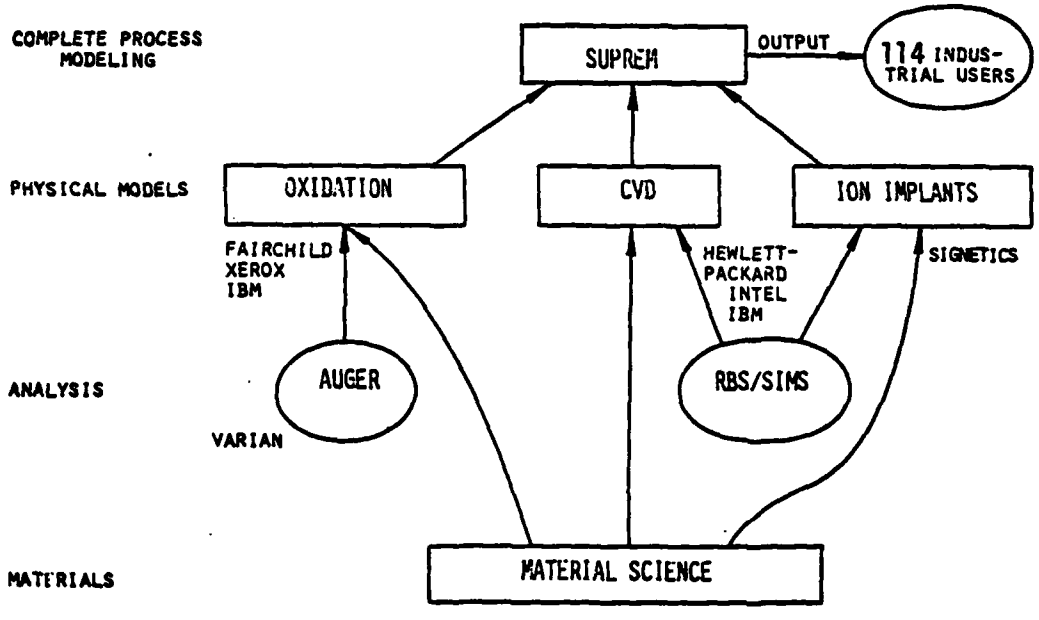



FIGURE 1 VERTICAL INTEGRATION AND INDUSTRIAL COUPLING OF RESEARCH PROGRAM

# PART 1

# ION IMPLANTATION

J. F. Gibbons, L. A. Christel

## 1.1 IMPLANTATION OF IONS INTO LAYERED TARGETS

### The Boltzmann Equation

Semiconductor device processing often involves ion implantation into multi-layered substrates. In such cases, it is desirable to answer the following questions: (1) what is the distribution of the implanted ion in the layered structure, (2) what are the distributions of the various atoms which are recoil implanted from one layer to the next and (3) what is the effect of channeling on the range distributions?

The most commonly used model for ion implantation is LSS [1.1] theory. This method works quite well for semi-infinite targets but is not applicable to the problem we are considering. To answer the questions posed above it is necessary to know the momentum distributions of each of the various projectiles (primary ions and recoiled atoms) at each depth $z$ in the target. The range distribution of the primary ions is then determined by assuming that any ion with momentum smaller than some value $p_o$ is effectively stopped.

The momentum distribution of the primary ions together with a description of the scattering process can be used to calculate the number and momenta of recoiled atoms which are generated at each depth in the sample. The range distribution of the recoiled atoms is then determined in the same way as that of the primary atoms.

It is possible to calculate range profiles in amorphous multi-layered structures by numerical integration of the linearized Boltzmann transport equation. Consider a collection of a large number of identical particles described by a phase space distribution function F

$$F(\bar{x},\bar{v},t)d\bar{x}d\bar{v} = dN(\bar{x},\bar{v},t) \tag{1.1}$$

where dN is the number of particles in the differential element $d\bar{x}d\bar{v}$ about $\bar{x}$ and $\bar{v}$ at time t.

In the absence of interactions, conservation of the total number of particles implies that F satisfies a continuity equation

$$\frac{\partial F}{\partial t} + \bar{\nabla} \cdot (F\bar{v}) = 0 \tag{1.2}$$

In the phase space formulation, $\bar{v}$ and $\bar{x}$ are independent variables, div $\bar{v}$=0 so we have

$$\frac{\partial F}{\partial t} + \bar{v} \cdot \bar{\nabla} F = 0 \tag{1.3}$$

Equation (1.3) describes the collisionless evolution of the system.

Now we assume that particles may be scattered from one region of phase space to another. We assume that there exists a constant density of scattering centers, $N_s$ per unit volume and that the interaction of particles with one such center is described by a differential scattering cross-section.

$$d\sigma(\bar{v}\rightarrow\bar{v}') = \frac{\begin{array}{l}\text{number of particles scattered into} \\ d\bar{v}' \text{ about } \bar{v}' \text{ per unit time}\end{array}}{\begin{array}{l}\text{number of particles with velocity } \bar{v} \\ \text{incident per unit time per unit area} \\ \text{transverse to } \bar{v}\end{array}} \tag{1.4}$$

The transfer of particles from one region of phase space to another at a given space point $\bar{x}$ is then deduced as follows. In the time interval dt, all particles with velocity $\bar{v}$ will move a distance $dx=|\bar{v}|dt$. In the volume element traversed by these particles are contained precisely $N_s dx=N_s|\bar{v}|dt$ scattering centers per unit area transverse to $\bar{v}$. From (1.4) it follows that the number of particles scattered out of $d\bar{v}$ in the time dt is given by

$$\Delta N(d\bar{v}\to d\bar{v}') = dF^{-}(\bar{x},\bar{v},t)d\bar{v} = F(\bar{x},\bar{v},t)d\bar{v}d\sigma(\bar{v}\to\bar{v}')N_s|\bar{v}|dt \quad (1.5)$$

The change in the distribution function due to scattering out of the element $d\bar{v}$ is determined by integrating equation (1.5) over all final velocities $\bar{v}'$ with the result

$$\left(\frac{\partial F}{\partial t}\right)^{-} = N_s \int_{\bar{v}'} |\bar{v}| F(\bar{x},\bar{v},t)d\sigma(\bar{v}\to\bar{v}') \quad (1.6)$$

Similarly the change in the distribution function due to scattering from all other velocities into $\bar{v}$ is

$$\left(\frac{\partial F}{\partial t}\right)^{+} = N_s \int_{\bar{v}'} |\bar{v}'| F(\bar{x},\bar{v}',t)d\sigma(\bar{v}'\to\bar{v}) \quad (1.7)$$

If in addition we include the possibility of new particles being generated at a rate $Q(\bar{x})$, the equation governing the distribution function F at each point in space becomes

$$\frac{\partial F}{\partial t} + \bar{v}\cdot\bar{\nabla}F = Q(\bar{x}) + N_s \int_{\bar{v}'} d\sigma(\bar{v}'\to\bar{v})F(\bar{v}')\ |\bar{v}'|$$

$$- d\sigma(\bar{v}\to\bar{v}')f(\bar{v})|\bar{v}| \quad (1.8)$$

This is the Boltzmann equation.

## Simplification of the Transport Equation

In the case we wish to consider, namely implants into layered structures which possess translational symmetry in the x-y plane, equation (1.8) can be simplified. We define a second distribution function f by

$$f(z,p_z,p_\perp) = \int F(\bar{x},\bar{v},t)|\bar{v}|v_\perp \, dxdydt \qquad (1.9)$$

where $\bar{p} = m\bar{v}$ and $p_\perp \equiv \bar{p}-(\bar{p}\cdot\hat{z})$.

The quantity $f(z,p_z,p_\perp)dzdp_z\, dp_\perp$ is then the total fluence (time integrated flux) of particles in the interval $dzdp_z\, dp_\perp$. It follows that the distribution function f obeys

$$\frac{\partial f}{\partial t} = N_s \int_{\bar{p}'} \frac{d\sigma(\bar{p}'\to\bar{p})}{\cos\theta_{\bar{p}}} f(z,\bar{p}') - \frac{d\sigma(\bar{p}\to\bar{p}')}{\cos\theta_{\bar{p}}} f(z,\bar{p}) \qquad (1.10)$$

where we've let Q=0, and $\theta_{\bar{p}} = \tan^{-1}\frac{p_\perp}{pz}$

In the case of heavy ions into lighter substrates, the assumption that the implanted ions travel in essentially the $\hat{z}$ direction ($p_\perp$=0) allows a further simplification. This has been done by other authors [1.2] but we will use equation (1.10) as the basis of our analysis.

## The Scattering Cross Section

Before describing the numerical integration of equation (1.10) we must examine the form of the differential cross section $d\sigma$. The cross section can be logically divided into two parts,

$$d\sigma = d\sigma_n + d\sigma_e \qquad (1.11)$$

where $d\sigma_n$ is known as the nuclear part and $d\sigma_e$ the electronic part. This separation stems from the fact that to a good approximation in the

energy range of interest, there exist two independent mechanisms by which ions lose energy as they travel through solids.

The first energy loss mechanism is that of the elastic two body interaction. We assume that each projectile interacts with only one atom of the substrate at a time, and that such interactions conserve energy and momentum. The assumption of two body interactions is justified by the fact that the electrostatic potentials of both particles are considerably screened by the presence of the many electrons in the solid. We thus expect the interaction to be substantial only when the two are in very close proximity. If there is more than one type of atom in the substrate, this assumption allows us to simply sum the contributions from each type. The particulars of this interaction are contained in the nuclear part of the cross section $d\sigma_n$. It is this part of the scattering which accounts for nearly all of the angular deflections of the ions.

Lindhard [1.3] et al have shown that the nuclear cross section for the two body interaction has a universal form

$$d\sigma_n (E \to E') = \frac{\pi a^2}{2} \frac{f(t^{\frac{1}{2}})dt}{t^{3/2}} \qquad (1.12)$$

where

$a = 0.8853 a_o (Z_1^{2/3} + Z_2^{2/3})^{-\frac{1}{2}}$

$t = TE\left[\frac{M_2}{4M_1}\left(\frac{a}{Z_1 Z_2 e^2}\right)^2\right]$

$a_o = \frac{\hbar}{me^2} = 0.529\ \text{Å}$

$T = E - E'$ = energy transferred

$M_1$ = ion mass

$M_2$ = target mass

$Z_1$ = ion atomic number

$Z_2$ = target atomic number

e = charge of the proton (e.s.u.)

Winterbon [1.4], Sigmund [1.5] and others have suggested an analytic form for the function $f(t^{\frac{1}{2}})$:

$$f(t^{\frac{1}{2}}) = \frac{\lambda t^{\frac{1}{2}} - m}{[1+(2\lambda t^{1-m})^q]^{1/q}} \quad T < T_{max} \equiv \frac{4M_1 M_2}{(M_1+M_2)^2} E \qquad (1.13)$$

where $\lambda$, m, and q are constants which depend on the model chosen for the interatomic potential. The Thomas-Fermi statistical model for example corresponds to $(\lambda, m, q) = (1.309, 0.333, 0.667)$.

The description of the scattering given by equation (1.12) is incomplete in that it specifies the cross section only as a function of energy, not of momentum. This inadequacy is remedied by again appealing to the elastic two body interaction. For such an interaction, conservation of energy and momentum imply that an ion which has initial energy E and transfers energy T to a target atom is deflected by an angle $\Phi$ given by [1.6]

$$\cos \Phi (E,T) = \sqrt{1 - \frac{T}{E}} + \frac{1-\mu}{2} \cdot \frac{T/E}{\sqrt{1 - \frac{T}{E}}} \qquad (1.14)$$

where $\mu = M_2/M_1$. Equation (1.14) together with equation (1.12) provides a complete description of the nuclear scattering.

The second energy loss mechanism is inelastic electronic scattering. In thiscase the ion loses energy in small amounts as it creates ionization

and other electronic excitations in the solid. Because the interaction is a many body effect, the ion is not deflected substantially in direction. As a result, these losses can be taken into account by knowing only the electronic stopping power defined by

$$S_e = \int T \cdot d\sigma_e(t) \tag{1.15}$$

The expression

$$S_e(E) = k \cdot E^p \tag{1.16}$$

where k is a constant and $p \approx 0.5$ is a good approximation at reasonable energies. The energy loss of an ion as it travels a distance $\Delta x$ in the solid due to such processes is then simply

$$\Delta E = N_s \Delta x \cdot S_e \tag{1.17}$$

Numerical Solution

For simplicity in what follows, we assume that at a given depth z the target is composed of a single atomic species. For more complex targets, the independent nature of the interactions involved allows us to simply sum the various contributions. In addition, we refer all quantities to unit area of the target.

Because the nuclear scattering is described in terms of the variables E and $\theta$ rather than $P_z$ and $P_\perp$, it is convenient to use a distribution function which is also a function of these former variables

$$f = f(z,E,\theta) \tag{1.18}$$

where again E and $\theta$ are defined by

$$E = \frac{p_2^2 + p_1^2}{2M_1} \; ; \; \theta = \tan^{-1} \frac{p_1}{p_2} \tag{1.19}$$

The solution to equation (1.10) requires that we specify an initial condition on f. At the surface (z=0) the ions which strike the target are assumed to be of unique energy $E=E_0$ and trajectory $\theta=0$ so that the initial condition

$$f(0,E,\theta) = D\delta(E-E_0)\delta(\theta) \qquad (1.20)$$

where D is the total dose of ions per unit area, is satisfied. Equations (1.10) and (1.20) then constitute a well defined mathematical problem.

There is a small discrepancy between this problem and the true physical system. An ion which is backscattered at z=0 is in reality lost to the vacuum whereas in the mathematical formulation it has a finite probability of being forward scattered back into the sample. The error introduced is small because the probability of wide angle scattering is very small and any ion which is scattered through a large angle must lose a large amount of energy. Hence such ions would not be expected to travel very far from the surface z=0 before coming to rest. We therefore assume this discrepancy has a negligible effect.

Accepting equations (1.10) and (1.20) as defining the problem at hand, the evolution of the momentum distribution is generated as follows. First, the quantities $E,\theta,z$ and f are discretized. We begin by defining the energy array $E_i$ $(i=1,.....M+1)$ with $E_1=0$ and $E_{M+1}=E_0$. Similarly we define the angle array $\theta_j(j=1, ...N+1)$ with $\theta_1=0$ and $\theta_{N+1}=\pi/2$.

At a given depth z, the system is then characterized by a two dimensional array $f_{ij}$ $(i=1, ...M;j=1,...N)$ where $f_{ij}(z)$ is the number of ions at depth z which have energies between $E_i$ and $E_{i+1}$ and

trajectories between $\theta_j$ and $\theta_{j+1}$. The initial condition (1.20) then implies that

$$f_{ij}(0) = D\delta_{iN}\,\delta_{j1} \tag{1.21}$$

As one moves in depth from z to $z + \Delta z$ the change in the distribution function is manifested in two distinct ways. First, particles may be scattered from one element of the array to another consistent with the relations (1.12) and (1.14). These events correspond to nuclear collisions in which the energy loss and angle of deflection are greater than the extent of the particular element in which the particle initially resides.

Secondly, particles lose energy due to small angle nuclear and electronic scattering, i.e. they scatter from the higher energy region of a particular element to the lower energy region of the same element. These losses must be accounted for by altering the energies $E_i$ which characterize the average energy associated with each row of the array. We will consider these two effects one at a time.

(1) Wide Angle Nuclear Scattering

In order to describe the consequences of wide angle scattering on the values of $f_{ij}$ it is necessary to investigate further the implications of the relation (1.14). Equation (1.14) states that given an initial energy E and energy transfer T, a unique scattering angle $\Phi$ is determined.

$$\Phi = \Phi(E,T) \tag{1.22}$$

In the above formulation, a particle is characterized by the angle

$\gamma_i$ which it makes with the surface normal $\hat{z}$. If the particle is scattered directly away from the $\hat{z}$ direction, its final trajectory will be $\gamma_f = \gamma_i + \Phi$. On the other hand if it is scattered directly toward the $\hat{z}$ direction, we have $\gamma_f = |\gamma_i - \Phi|$. These two cases correspond to the condition that the direction of the particle's final momentum $\bar{p}_f$ lies in the plane determined by $\bar{p}_i$ and $\hat{z}$. In general this will not be the case and we have

$$|\gamma_i - \Phi| \leq \gamma_f \leq \gamma_i + \Phi \tag{1.23}$$

We therefore define a function G such that the probability of scattering to the interval $d\gamma_f$ about $\gamma_f$ from the initial trajectory $\gamma_i$ is given by

$$P(\gamma_i \rightarrow \gamma_f)\, d\gamma_f = G(\gamma_i, \gamma_f, \Phi)\, d\gamma_f \tag{1.24}$$

If one assumes that the scattering is azimuthally symmetric about the initial direction, one can show that the function G is given by

$$G(\gamma_i, \gamma_f, \Phi) = \begin{cases} \dfrac{\sin\gamma_f}{A}\left[1 - \left(\dfrac{\cos\gamma_f - B}{A}\right)^2\right]^{-\frac{1}{2}} & \text{if } \gamma_f \text{ satisfies (1.23)} \\ 0 \qquad \text{otherwise} \end{cases} \tag{1.25}$$

where $A = \sin(\gamma_i)\sin(\Phi)$ and $B = \cos(\gamma_i)\cos(\Phi)$. G satisfies the normalization condition:

$$\int G(\gamma_i, \gamma_f, \Phi) = 1 \tag{1.26}$$

The integral of G is expressible in terms of the inverse cosine function.

Now consider a transition from the element (ij) to the element (i'j'). Since the energy of a particle can only decrease, we

consider only $i' < i$. In practice the initial energy and trajectory are characterized by average values

$$\bar{E} = \tfrac{1}{2}(E_i + E_{i+1})$$
$$\bar{\theta}_j = \tfrac{1}{2}(\theta_j + \theta_{j+1}) \tag{1.27}$$

The energy transfer lies in the interval

$$\bar{E} - E_{i+1} \equiv T_{min} \leq T \leq T_{max} \equiv \bar{E} - E_i \tag{1.28}$$

The scattering angle for the transition is determined using an average value for the energy transfer

$$\Phi = \Phi(\bar{E},\bar{T}) \tag{1.29}$$

where

$$\bar{T} \equiv \tfrac{1}{2}\,(T_{min} + T_{max}) \tag{1.30}$$

The probability of a transition involving an energy transfer in the interval dT about T is proportional to $d\sigma(T)$. Using a finite difference approximation for $\frac{\partial f}{\partial z}$, the number of particles which are transferred from the element (ij) to the element (i'j') due to collisions in the interval $\Delta z$ is given by

$$\Delta f_z(ij \to i'j') = N_s \Delta z f_{ij} \int_{T_{min}}^{T_{max}} \frac{d\sigma_n(T)}{\cos\bar{\theta}_j} \int_{\bar{\theta}_{j'}}^{\theta_{j'+1}} G(\bar{\theta}_j, \gamma_f, \Phi(\bar{E},\bar{T}))\, d\gamma_f \tag{1.31}$$

Repeated application of (1.31) for all i, i', j, j' determines the evolution of the array $f_{ij}(z)$.

$$f_{ij}(z+\Delta z) = f_{ij}(z) + \sum_{\substack{i'>i \\ j'}} \Delta f_z(i'j' \to i\ j) - \sum_{\substack{i'<i \\ j'}} \Delta f_z(ij \to i'j') \qquad (1.32)$$

One special consideration must be mentioned. The above formulation neglects all transfers with final trajectories $\gamma_f > \pi/2$. In practice half of such ions are considered to have been stopped, the other half are transferred to the element (i,N+1) where i is known from energy consideration as above.

(2) Small angle and electronic scattering

For each row i of the array, these losses are calculated using

$$\Delta\bar{E} = N_s \Delta z \, (S_e(E) + S_n(T_{\frac{1}{2}})) \sum_j \frac{f_{ij}}{\cos \theta_j} \qquad (1.33)$$

where $S_e(E)$ is given by (1.16) and

$$S_n(T_{\frac{1}{2}}) = \int_0^{T_{\frac{1}{2}}} T d\sigma_n(T) \qquad T_{\frac{1}{2}} = \tfrac{1}{2}(E_{i+1} - E_i) \qquad (1.34)$$

The energy array $E_i$ and elements $f_{ij}$ are then altered so that the losses are taken into account but the values of $E_i$ remain equally spaced.

Profile Generation

The range distribution of an implanted ion is calculated by assuming that any particle which is scattered to an energy less than $E_2$ is stopped. In practice, the energy $E_2$ is fixed at 1 Kev. At each step of the integration the number of ions stopped in the interval $\Delta z$ is then simply

$$N_{stopped}(z, z+\Delta z) = \sum_j f_{1j}(z) \qquad (1.35)$$

These particles are removed from the array (set to zero) and the concentration profile is then described by

$$C(z) = \frac{N_{stopped}(z, z+\Delta z)}{\Delta z} \qquad (1.36)$$

Multi-layer targets merely require that at each interface we alter the parameters which define the differential cross section. Integration then continues just as before.

The Knock-on Problem

At each step in the integration, two body collision theory can again be used to determine the number, energies and trajectories of the recoiled substrate atoms which are involved in collision processes. These recoiled particles can be placed in a second array $g_{ij}$, and the range distribution determined either as above, or perhaps with a simpler one dimensional stopping theory.

Channeling

Because of the presence of detailed information about the trajectories of the particles in the above formulation, it is possible to investigate the effects of channeling in crystalline materials.

Imagine for example, that an ion is being implanted into a silicon target which has its <111> axis misaligned by eight degrees with respect to the beam direction. Let $\Psi(E)$ be defined such that ions of energy E which have trajectories within $\Psi(E)$ of the <111> direction are channeled. The number of ions which enter the <111> channel at depth z with energies between $E_i$ and $E_{i+1}$ can then be approximated by

$$N_{channeled}(E,z) = \sum_{j'} f_{ij'} \frac{\Psi(E)}{\pi} \qquad (1.37)$$

where the sum includes those values of j' such that

$$8° - \Psi(\bar{E}) \leq \bar{\theta}_{j'} \leq 8° + \Psi(\bar{E}) \qquad (1.38)$$

The factor $\frac{\Psi(E)}{\pi}$ is necessary because only this fraction of the total azimuthal angle lies within $\Psi(E)$ of the <111> direction.

Thus the number and energy of ions satisfying the channeling condition are known at each depth z. Once in the channel, ions are assumed to be described by pure electronic stopping. If the electronic stopping coefficient is k, an ion of energy E as a range

$$R_E = \int_0^E \frac{dE'}{KE'^p} = \frac{E^{1-p}}{K(1-p)} \sim \frac{2E^{\frac{1}{2}}}{K} \qquad (1.39)$$

Since the point of origin and the range of each channeled ion is known, the effect of channeled ions on the final distribution can be taken into account. In addition, planar channeling and various dechanneling effects can be included.

## Conclusions

A computer program employing the above formulation has been written and debugged. Preliminary results are encouraging. In the future we plan to study implants into silicon targets coated with various thicknesses of silicon dioxide and silicon nitride. The distributions of recoiled nitrogen and oxygen atoms will also be investigated. Finally, an attempt will be made to characterize the experimentally observed exponential tails which occur during implants into single crystal targets. Hopefully, this new method of calculation will be a powerful addition to current theoretical attempts to predict the range distributions which occur in ion implantation.

## References


[1.1] J. Lindhard, M. Scharff, Phys. Rev. 124, p 128 (1961).
H. E. Schiott, Mat. Fys. Medd. Dan Vid. Selsk. 35, No. 9 (1966).
Lindhard, Scharff, and Schiott, Mat. Fys. Medd. Dan. Vid. Selsk. 33, No. 14 (1963).

[1.2] Gibbons and Smith, Ion Implantation in Semiconductors 1976 (Plenum Publishing Corp. N.Y. 1977)

[1.3] Lindhard, Nielsen and Scharff, Mat. Fys. Medd. Dan. Vid. Selsk. 36 No. 10 (1968)

[1.4] Winterbon, Sigmund and Sanders, Mat. Fys. Medd. Dan. Vid. Selsk. 37, No. 14 (1970)

[1.5] Sigmund, Rev. Roumanian Physics, 17, p 823 ff. (1972).

[1.6] Goldstein, Classical Mechanics (Addison-Wesley, Reading Mass., 1950) p 58 ff.

PART 2

# THERMAL OXIDATION*

B. E. Deal, C. P. Ho, H. Z. Massoud, J. D. Plummer, R. R. Razouk, and W. A. Tiller

## 2.1 INTRODUCTION

Thermally grown silicon dioxide remains one of the cornerstones of modern integrated circuit technology. $SiO_2$ is commonly used as a mask against dopant diffusion, to passivate active device regions and junctions, to insulate field regions and provide isolation between active components and, as the gate dielectric, is an actual component in MOS devices. As a result, the control and predictability of oxide growth and the resulting electrical properties are critical if reproducible device performance is to be achieved.

Despite the fact that thermally grown $SiO_2$ has been used in the production of discrete devices and integrated circuits for almost twenty years, considerable gaps remain in our understanding of its growth kinetics, the charges associated particularly with the $Si/SiO_2$ interface, and the interaction of oxidation with other important processes such as dopant segregation and diffusion. The recent resurgence of research activity in laboratories throughout the country in these areas is a direct indication of these gaps in our knowledge.

Spurring the activity towards better understanding of $SiO_2$ growth kinetics and oxide properties is the continuing decrease in active device

*This work represents a joint effort by the Stanford University Integrated Circuits Laboratory, Fairchild Camera and Instrument Corporation Research and Development Laboratory, the Stanford University Department of Materials Science and the Stanford University Solid State Laboratory.

dimensions in integrated circuits. Device dimensions well under 5 μ are routinely used in production today; projections for the next several years suggest the emergence of devices with active dimensions close to 1 μ. In fact, submicron device structures have already been built in a number of research laboratories.

As device dimensions shrink in an effort to achieve higher performance, the models for oxidation, diffusion, and other processes, many of which were developed ten to fifteen years ago, become inadequate. Interactions between oxidation and diffusion, for example, which can largely be neglected in large geometry devices with relatively deep junctions, can no longer be neglected as lateral and vertical dimensions shrink. Modeling of process phenomena can no longer be considered one dimensional. An oxidation on one portion of an integrated circuit structure can significantly affect nearby diffusions and devices.

If the integrated circuit industry is to fabricate VLSI structures with good reproducibility and high yields, better models for the basic processes used must be developed and an understanding of the interaction of various process phenomena must be achieved. Such are the specific goals of the work described here. In summary form, the goals of this program are as follows:

1. To achieve accurate analytic prediction of oxide thickness under all process conditions encountered in modern technologies. These include the effects of high substrate doping levels, arbitrary silicon substrate orientation, the presence of a chlorine species during the oxidation, the use of multiple species (such as $O_2/H_2O$ mixtures) during oxidation, and ambient pressures greater than or less than one atmosphere.

2. To achieve accurate analytic prediction of oxide charge densities, also under all process conditions important to modern devices. These conditions include arbitrary silicon substrate orientation, ambient conditions during oxidation, and high and low temperature anneals.

3. To achieve accurate analytic prediction of segregation and redistribution effects at the Si-$SiO_2$ interface. This includes accurate prediction of dopant profiles in both the $SiO_2$ and Si.

4. To develop a microscopic, electrically accurate model of the Si-$SiO_2$ interface. Such a model is believed to be essential for achieving physical understanding of oxidation kinetics, the origin of oxide charges and segregation and redistribution.

5. To predict quantitatively the effect of an oxidation process on other important process phenomena such as diffusion and the growth of oxidation-induced stacking faults. Point defects (vacancies and interstitials) in the $SiO_2$ and Si layers appear to be the basic physical microstructures which are responsible for, and will enable quantitative prediction of oxidation kinetics, oxide charges, enhanced dopant diffusion during oxidation, the growth of oxidation-induced stacking faults and other process phenomena. A "working model" we have developed to explain these processes will be described later in this report.

Realization of these goals will allow accurate prediction of oxide and surface properties and the influence of oxidation on other processes for an arbitrary device structure. This, in turn, will minimize costly iterative, empirical techniques in the development and characterization of new and improved technologies.

An illustration of the importance of these goals is shown in Fig. 2-1. The device shown here is an oxide isolated NMOS transistor, similar in configuration to many devices currently in production. The numbers indicated on the figure refer to practical, current problems in oxidation kinetics, $SiO_2$ charge densities, and other process phenomena which are addressed by the above goals. Referring to the figure:

1,5: Over the $N^+$ source and drain diffusions, enhanced oxidation rates due to high dopant concentrations are important as are impurity redistribution and segregation effects. In addition, the oxidation of such heavily doped regions is well known to affect such things as diffusion coefficients in the underlying silicon.

2,3: In the thin gate oxide, the kinetics of multiple species oxidations ($O_2/H_2O$ for example) and the use of a Cl species may be important.

4: The three-dimensional nature of the device is illustrated in the oxide isolation region where the effective surface crystal orientation changes with position and affects the oxidation kinetics and charge densities.

5: Redistribution and segregation are also extremely important under the thick field oxide where a lightly doped P-type layer is often ion implanted to prevent surface inversion. In addition, the growth of the thick field oxide over this $P^-$ region is known to enhance the diffusion coefficient of the P-type impurity (boron), presumably through the generation of point defects (interstitials) during the oxidation process.

6: Fixed oxide charges ($Q_{ss}$) and interface state charges ($N_{st}$) located near the Si-$SiO_2$ interface affect device threshold voltage, carrier mobilities, junction leakage currents and breakdown voltage, and numerous other important device properties. These cannot at present be predicted analytically.

Similar arguments could be made with respect to other technologies and other devices currently being manufactured. The major point to be made here is that the goals presented above for this program are addressed to important practical problems. Understanding on a more basic physical level of oxidation kinetics, charges, redistribution, and their influence on other processs phenomena will permit optimization of present device structures and will minimize the amount of empiricism needed to develop new structures.

## 2.2 SUMMARY OF PRINCIPAL ACCOMPLISHMENTS TO DATE

The thermal oxidation work to be described here is the result of an extensive interdisciplinary effort including four main groups--the Stanford University Integrated Circuits Laboratory, Fairchild Camera and Instrument's Research and Development Laboratory, the Department of Materials Science at Stanford, and the Stanford University Solid State Laboratory.

At the beginning of this program, understanding of silicon oxidation kinetics was based largely upon the general oxidation relationship developed in 1965 by Deal and Grove [2.1] and given by

$$\frac{X_o^2}{B} + \frac{X_o}{B/A} = t + \tau \qquad (2.1)$$

where $X_o$ = oxide thickness, t = oxidation time, and A, B, and $\tau$ are constants which are functions of oxidation conditions. The parabolic rate constant B dominates the overall reaction at high temperatures and for thick oxides and includes such factors as solubility of the oxidant in the oxide, oxidant diffusion rate in the $SiO_2$, and the partial pressure of the oxidant. The linear rate constant B/A dominates the reaction for low temperatures and short times and includes oxidant solubility, partial pressure, and the reaction at the Si-$SiO_2$ interface. Such surface variables as silicon orientation and dopant concentration can affect this surface oxidation reaction and hence B/A.

The use of this first-order kinetic relationship allows prediction of oxide thickness given that B and B/A may be expressed as functions of orientation, doping level, ambient conditions, and temperature. Since B and B/A are based on a model in which oxidant diffusion through an existing oxide layer and Si-$SiO_2$ interface reaction kinetics, respectively, determine the overall oxidation rate, some understanding of the underlying physical mechanisms involved may also be obtained from this model.

An even more basic physical understanding of the mechanisms involved in thermal oxidation will be needed, we feel, to understand fully oxidation kinetics and, in particular, to predict analytically oxide charge densities and the influence of oxidatior on other process phenomena. A "working model" based largely upon point defects in the $SiO_2$ and Si layers will be described later in this report.

Many of the important oxidation kinetic processes and most of the charges associated with $SiO_2$ are determined by the 5-40 Å transition region between Si and $SiO_2$. A clear understanding on an atomic level of this region appears to be necessary for a physical understanding of the kinetic

and charge density experimental data obtained here and elsewhere.

Toward these ends, a large portion of the initial effort in this program was aimed at gathering sufficient kinetic data to make modeling of the physical processes involved possible. More recently, with this kinetic data as a basis, considerable effort has been devoted to developing underlying physical models to explain the data. The following list summarizes our efforts in both these areas.

1. Complete characterization of (111) and (100) orientation kinetics for dry $O_2$ and wet $O_2$ between 700° and 1200°C has been completed. These results were described in Ref. [2.2].

2. Complete experimental characterization of $O_2$/HCl oxidation kinetics for 0-10% HCl and T = 900-1100°C has been completed. These results were also described in Ref. [2.2] and are the subject of journal publication [2.3].

3. In an effort to understand physically the observed $O_2$/HCl kinetics, oxidations in $O_2/H_2O$, $O_2/Cl_2$, and $H_2O/Cl_2$ mixtures have been investigated. The results were described in Ref. [2.4] and are the subject of a journal article [2.5].

4. The kinetics of $H_2O$ and $H_2O$/HCl mixtures using a pyrogenic system have been studied, and the results were described in last year's annual report [2.4] and in a journal publication [2.6].

5. Complete experimental characterization of heavily doped phosphorus substrate oxidation kinetics for $N_D = 10^{15}$ to solid solubility, T = 800°-1100°C in dry $O_2$ has been completed. Empirical relationships for both the linear and parabolic rate constants as a function

of phosphorus chemical concentration have been derived. These results were described in last year's annual report [2.4] and are the subject of a journal article [2.7].

6. A physical model based entirely upon electrical effects (higher equilibrium vacancy concentrations at high doping levels) has been derived to explain the observed enhanced oxidation rate in heavily phosphorus doped substrates. This model explains our and other experimental data extremely well and allows prediction of oxidation kinetics for B, As, Sb, and other impurity doped substrates. It was described in detail in last year's annual report [2.4] and is the subject of two journal articles to be published later this year in The Journal of The Electrochemical Society [2.8,2.9]. The essence of this model is reviewed later in this report along with results demonstrating its agreement with experiment.

7. Within the past year, significant results have been achieved in experimentally correlating the two principal oxide charges, $Q_{ss}$ and $N_{st}$, with process conditions. Initial results were presented in last year's annual report [2.4] and are described in detail in a journal article [2.10]. More recent results are described later in this report.

8. Installation and programming of a flexible, computer controlled C-V measurement system for characterization of oxide charge densities has been completed.

9. Implementation of computer prediction in the SUPREM program of first-order oxidation kinetics has been completed, including the point defect (vacancy) model for oxidation of heavily doped substrates.

10. A system for corona charging of Si wafers in a high temperature oxidizing environment has been constructed. It is being used in an experimental investigation of the effects of electric fields on oxidation kinetics.

11. Initial experimental measurements on the growth kinetics and other properties of thin (0-500 Å) oxide layers on silicon have been made and will be described later in this report.

12. A conceptual framework relating oxidation kinetics and other process phenomena to basic physical mechanisms has been developed. This "working model" is an attempt at providing a consistent understanding of many interrelated phenomena. It is described in detail below.

## 2.3 A "WORKING MODEL" FOR THE GROWTH OF $SiO_2$ LAYERS ON SILICON

As an introduction to the remainder of this report which describes our work in the last year, we present here a "working model," based largely on point defects (vacancies and interstitials) for the growth of $SiO_2$ on silicon and the relationship between oxidation kinetics and oxide charges. We do not claim that this model is complete in all aspects but it will be shown to be consistent with our experimental observations over the past several years, and with the experimental results of many other workers. We present it as an attempt at a unifying picture of the physical basis for oxidation kinetics, oxide charges, and the influence of oxidation on other processes. A key feature of this model is that it shows promise of allowing quantitative prediction of not only oxidation kinetics but also of oxide charge densities and other phenomena such as enhanced diffusion coefficients during oxidation and the growth rate of oxidation induced

stacking faults. The model may be viewed as an extension of the basic Deal-Grove model [2.1], with particular attention paid to the reaction at the $Si/SiO_2$ interface and to the need for "free volume" at this interface in order for the reaction to proceed.

The basis for the model is shown in Fig. 2.2. We consider first the growth of relatively thick layers of $SiO_2$ (>500 Å) in dry $O_2$ and will consider later the initial growth kinetics in dry $O_2$ and $H_2O$.

$O_2$ from the gas phase is incorporated into the surface of the $SiO_2$ layer and is known to diffuse down to the $Si/SiO_2$ interface where the reaction converting silicon to $SiO_2$ takes place. The work of Doremus [2.11], Jorgensen [2.12], Rayleigh [2.13], and Collins and Nakayama [2.14] presented conflicting evidence concerning the presence or absence of an important charged oxygen ion in the diffusion process. The work of Mills and Kroger [2.15] clearly showed that a doubly negative oxygen interstitial ion was the dominant species in the high temperature conductivity of $SiO_2$. The first-order pressure dependence of the parabolic rate constant B in the Deal and Grove model [2.1] seems to indicate that the dominant diffusing species is molecular oxygen. For the present, we regard the question of the particular oxygen species involved in the diffusion as incompletely resolved and, for the purposes of this working model, will consider $O_2$ as the dominant species for thick oxide layers. (The system for high temperature corona charging of Si wafers during oxidation which we have recently completed should help to resolve this question.) In any event, the model described here deals primarily with the interface reaction.

At the $Si/SiO_2$ interface, dissociation of the $O_2$ occurs as a first step in the oxidation reaction. We now address the question of "free volume" which is at the heart of the "working model" proposed here.

When silicon and an oxygen species react in the interface region to form an $SiO_2$ molecule, a necessary quantity of "free volume" must also be supplied so that the molecule can fit into the normal $SiO_2$ network structure. From density considerations, the average spacing between Si atoms in $SiO_2$ is about 1.3 times the average spacing between Si atoms in the Si lattice. This would lead to a 70% strain in the $SiO_2$ if the lattice continued in a coherent way into the $SiO_2$ film. If the interface moves into the Si at V cm/sec, the oxide thickens at 2.25 V cm/sec, which means that the "free volume" must flow in at a rate of 1.25 V cm/sec per square centimeter of interface area to produce an unstrained film. If no "free volume" is supplied, then the excess free energy stored as strain in the $SiO_2$ film is sufficient to strongly retard the oxidation reaction at normal driving forces.

A portion of the required "free volume" for the reaction may be supplied as shown in the top reaction in Fig. 2.2. In general, we might expect some lattice mismatch to occur at the interface, stored as a cross-grid of dislocation lines (dangling bonds) and some of the mismatch to be stored as strain in the $SiO_2$ film. The grid of dangling bonds may be partially responsible for the interface states $N_{st}$ observed at the interface. The strain energy builds up in the film until the stress is sufficient to cause viscous flow in the oxide at some thickness $X^*$ so that the $SiO_2$ material at $X > X^*$ is in the unstrained state. This yields a distributed array of crossed dislocation-type lines, with their associated dangling bonds, spread throughout the layer of thickness $X^*$ to satisfy some of the mismatch. This array of dangling bonds may be partially responsible for the fixed charge ($Q_{ss}$) observed near the interface.

Two alternative means of providing the necessary free volume at the interface are shown in the middle and bottom of Fig. 2.2 and are based upon point defects in the $SiO_2$ and silicon layers. In the middle reaction, vacancies ($Si_V$) from the silicon substrate provide the reaction sites or the required "free volume." In the bottom reaction, silicon atoms are removed from lattice sites to create interstitials ($Si_I$) and the required reaction sites for the growth of the $SiO_2$. Note that possible charge states of the point defects have been neglected in the reactions proposed here. Si vacancies are known to exist in at least three charge states, $Si_V^+$, $Si_V^-$, and $Si_V^=$, in addition to neutral vacancies; very little is known at the present time about possible $Si_I$ charge states.

It is believed that all three of these mechanisms play a role in the oxidation process. Each of them likely dominates under specific process conditions. For example, the reaction involving $Si_V$ is believed to dominate in the case of heavily doped $N^+$ substrates when the position of the Fermi level is near the conduction band and the equilibrium concentration of vacancies is dramatically increased. This mechanism is the basis for our quantitative model explaining the enhanced oxidation rates of $N^+$ substrates [2.4].

The bottom mechanism in Fig. 2.2 has been postulated by Dobson [2.16, 2.17] and the excess interstitial flows produced during thermal oxidation may be the mechanism of oxidation enhanced dopant diffusion effects [2.16, 2.17,2.18]. In addition, the interstitials produced during the reaction have been linked to oxidation-induced stacking faults [2.18]. Finally, it might be conjectured at this point that some of the excess interstitials move into the $SiO_2$ layer. Such a flow would then very probably be closely related to the observed oxide charge $Q_{ss}$ that has been discussed as incompletely oxidized silicon atoms [2.19].

If we assume for the moment the role of point defects in the oxidation process, then some interesting observations may be made about the origin of the various oxide charges ($Q_{ss}$ and $N_{st}$). Consider, for example, the well known $Q_{ss}$ triangle [2.19] shown in schematic form in Fig. 2.3. If the density of these charges resulting from a particular oxidation process is proportional to the concentration of excess silicon interstitials ($Si_I$) produced by the oxidation, then the temperature behavior of $Q_{ss}$ may be rationalized.

In the dry $O_2$ case, for example, as temperature increases, the oxidation rate increases and more $Si_I$ are produced per unit time. However, the equilibrium concentration of vacancies also increases [2.20]. The activation energy associated with vacancy generation is believed to be >2 eV, whereas the activation energy of the interface oxidation reaction is well known to be quite close to 2 eV [2.1]. Van Vechten and Thurmond [2.21] conclude from bulk quenching data that $[Si_V] \alpha e^{-2.4/kT}$, while Seidel and MacRae [2.22] propose that $[Si_V] \alpha e^{-3.4/kT}$. In either case, behavior similar to that shown in Fig. 2.4 would be expected with the concentration of excess $Si_I$ decreasing with temperature. (Excess $Si_I$ are assumed to be eliminated by available $Si_V$.) $Q_{ss}$ would therefore be expected to decrease with temperature, exactly as observed, in an oxidizing environment.

A similar argument involving $Si_V$ and $Si_I$ has been proposed to account for the growth rate of oxidation induced stacking faults [2.23]. An intriguing relationship between $Q_{ss}$ and oxidation induced stacking faults may now be observed. In dry $O_2$, it is well known that the temperature at atmospheric pressure at which retrograde growth (shrinkage) of oxidation-induced stacking faults occurs is ≃1240°C [2.23]. This is also the approximate temperature at which the dry $O_2$ $Q_{ss}$ curve in Fig. 2.3 reaches its

"minimum." We hypothesize that both of these phenomena have the same origin; at this temperature, the concentration of excess $Si_I$ drops to $\simeq 0$ because sufficient $Si_V$ are present to eliminate them.

In the case of $H_2O$, shown in Fig. 2.3, the interface reaction rate is faster at all temperatures (although the interface reaction rate activation energy is still $\simeq 2$ eV). As a result, more excess $Si_I$ are generated per unit time at any given temperature, and therefore the temperature required to generate sufficient $Si_V$ to bring the excess $Si_I$ concentration to $\simeq 0$ should be higher. In the case of oxidation-induced stacking faults, this temperature is $\simeq$1350°C [2.23]. In the case of the $Q_{ss}$ triangle, the high temperature behavior of $Q_{ss}$ in $H_2O$ is not well known but the point of intersection of the $H_2O$ curve in Fig. 2.3 with the "minimum" $Q_{ss}$ value is certainly >1300°C and may well correspond to the 1350°C temperature for retrograde growth of stacking faults.

It is also interesting that the $H_2O$ $Q_{ss}$ curve in Fig. 2.3 has a smaller slope vs temperature than does the dry $O_2$ curve. This is in spite of the fact that more $Si_I$ are generated per unit time in the $H_2O$ case. One might, therefore, expect a higher $Q_{ss}$ value for $H_2O$ at all temperatures; this is not observed as Fig. 2.3 shows. The reason for this is not understood at present. Perhaps the $H_2$ released at the $Si/SiO_2$ interface in the case of $H_2O$ oxidation bonds with some of the excess $Si_I$, effectively eliminating some fraction of them and therefore lowering $Q_{ss}$.

One can predict several other interesting features of $Q_{ss}$ behavior under various process conditions, based upon this model. In heavily doped $N^+$ silicon in which the Fermi level position generates more $Si_V$ at any given temperature [2.20], the value of $Q_{ss}$ should be lower. (Measurement difficulties make this difficult to verify.) At low oxygen or $H_2O$ partial

pressures, the oxide growth rate is slower, hence fewer $Si_I$ are generated per unit time and therefore $Q_{ss}$ should be lower. A recent publication confirms this prediction [2.24]. In fact, the inert gas curve in Fig. 2.3 might be regarded simply as the limiting case for low pressure oxidation. Annealing in an inert gas allows time for the excess $Si_I$ concentration to drop to $\simeq 0$ and therefore $Q_{ss}$ drops to its "minimum" value. (It is well known [2.23] that the growth rate of oxidation-induced stacking faults decreases in low partial pressures of $O_2$ or $H_2O$, in agreement with the model described here.)

Interestingly, a further prediction of the model is that high pressure oxidation ought to generate more $Si_I$ per unit time and therefore produce faster growth rates for oxidation-induced stacking faults and higher $Q_{ss}$. We are not aware of any measurements of $Q_{ss}$ in high pressure oxidized samples; measurement of this effect would be difficult because most high pressure systems return the samples to atmospheric pressure (and therefore "anneal" $Q_{ss}$) before they are removed from the furnace.

One final interesting observation follows from the discussion above. The "minimum" value of $Q_{ss}/q$ shown in Fig. 2.3 is generally shown as a constant value, independent of temperature [2.19]. If this minimum is really a result of the equilibrium (not excess) concentration of $Si_I$, then it should decrease with temperature since the generation of $Si_I$ is believed to be an activated process with an Arrhenius temperature dependence. Minimum values for $Q_{ss}/q$ have been reported between $\simeq 10^{10}/cm^2$ and $2 \times 10^{11}/cm^2$, depending upon the silicon orientation, the laboratory in which the experiments are performed, etc. These values may represent nothing more or less than the residual contamination levels ($Na^+$) in the oxides or the $Q_{ss}$ caused by trace amounts of $O_2$ or $H_2O$ in the furnace gases. In fact,

there have been a few reports [2.25] of very low $Q_{ss}/q$ values on ⟨111⟩ silicon ($\simeq 10^{10}/cm^2$) achieved by inert gas annealing under very well controlled conditions. It may well be that very low $Q_{ss}$ values can be achieved by relatively low temperature inert gas anneals; experiments of this type would be an interesting test of the model presented here.

We can also speculate about the often observed enhanced growth rates for very thin (<200 Å) oxide layers in dry $O_2$. During the initial stages of the oxidation, an excess of $Si_V$ would be expected because of the presence of the silicon surface. These vacancies might well diffuse into the $SiO_2$ layer where they become oxygen vacancies $O_V$. In the $SiO_2$ layer, these vacancies provide reaction sites for the downward diffusing $O_2$. The reaction shown in Fig. 2.5 may therefore take place. (Again, possible charge states of $Si_V$ and $O_V$ have been neglected.) The oxygen molecule dissociates, creating an interstitial ($O_I^=$) and filling the vacancy ($O_V$). The ionized oxygen interstitial ion now diffuses in a coupled manner with the positive holes, effectively increasing the diffusion coefficient of the oxygen because of the electric field. (The holes diffuse more rapidly than the $O_I^=$ and "pull" the $O_I^=$ along behind them.) Such a mechanism of coupled diffusion has been proposed previously [2.26]; here, we show that it is possibly consistent with the point defect model of oxidation. The enhanced growth rate caused by this process would be expected to be effective over an oxide thickness roughly corresponding to the extrinsic Debye length. Values of ≃150 Å have been calculated for dry $O_2$ and ≃5 Å for wet $O_2$ [2.26].

Interestingly, this model would predict a new "linear-parabolic" growth law for thin oxides in dry $O_2$ with $O_I^=$ as the diffusing species. This growth law would asymptotically join the familiar Deal-Grove model

once the above mechanism becomes negligible ($t_{ox} > \simeq 150$ Å in dry $O_2$). Such a linear-parabolic growth law (with higher effective B and B/A rate constants than are appropriate for thicker layers) is not inconsistent with some recently reported experimental data [2.27]. In addition, this model would predict a $P^{0.5}$ pressure dependence for very thin oxide layers in dry $O_2$, which gradually shifts over to the $P^{1.0}$ dependence of the Deal-Grove model as the dominant diffusing species shifts from $O_I^=$ to $O_2$. Pressure dependencies between $P^{0.5}$ and $P^{1.0}$ have been recently observed for thin oxide layers [2.28] grown in dry $O_2$.

In summary, we propose here a "working model" for the thermal oxidation of silicon. It may be regarded as an extension of the Deal-Grove model with particular attention paid to the reaction occurring at the Si/$SiO_2$ interface. Heavy emphasis is placed on the role of point defects in the oxidation reaction although we also consider the possibility that some of the "free volume" required for the interface reaction to occur may be provided by ingrown strain in the $SiO_2$ layer, which may be partially relieved by a network of dislocations. The attraction to this model is primarily that it agrees with a wide body of experimental results and it may allow quantitative prediction of oxidation kinetics, oxide charge densities, and other related process phenomena such as enhanced diffusion coefficients during oxidation and the growth of oxidation-induced stacking faults. We propose this model as a means of structuring and unifying our research program. We now turn to the specific results achieved in the program in the past year.

## 2.4 DETAILED DISCUSSION OF RESULTS

### A. Physical Modeling of Oxidation Kinetics

The physical model shown in Fig. 2.2 provides a perspective overview of the partitioning of the total driving force for the oxidation process into key parallel subprocesses and delineates the boundary value problems needed to quantitatively connect these components. In our earlier work, the free volume supply condition required to sustain such a transformation was explicitly given so that attention was focussed on Frenkel defect formation in the substrate as well as on viscous flow in the oxide [2.29, 2.30]. An atomistic level model was given for the oxidation process which appeared consistent with much of the earlier data. In addition, five possible paths for influencing the rate and character of the oxidation process were described.

During the present reporting period, a steady-state analysis which included electric-field effects was completed and this predicted orientation-dependent linear and parabolic rate constants. A free energy budget accounting for the system indicated that solely neutral oxygen diffusion in $SiO_2$ cannot account for the oxidation reaction. An analysis of the system strain energy associated with the formation of a coherent $Si/SiO_2$ interface revealed that ~2/3 of the total driving force for the reaction would be consumed by this energy storage mechanism. Analysis of the system energetics associated with forming an incoherent $Si/SiO_2$ interface indicates that this is the preferred condition. The interface disregistry associated with this system relaxation indicates that a pseudodislocation network forms at the $Si/SiO_2$ interface to reduce the stored strain energy and this network is probably the origin of the Nst states at the interface.

Movement of the interface by continued oxidation requires the influx of vacancies from the Si, free volume from the $SiO_2$, or interstitial formation at the interface to remove these extra half-plane segments on the Si-side of the interface. At high temperatures, the $SiO_2$ flows readily to relieve the stresses whereas, at low temperatures, the interstitial formation and vacancy influx mechanisms tend to dominate. Overall, this strain relaxation process tends to become the dominating feature controlling the kinetics of oxidation. Molecular beam deposition of the Si with simultaneous in-situ oxidation should greatly reduce this driving force and enhance the oxidation rate at low temperatures.

The interface attachment mechanism of the oxidation process was also considered from a simple point of view and the rate limiting step found to be the breaking of the Si-Si bonds prior to an actual oxidation event. About 10% to 25% of the overall driving force is expected to be consumed by this process. It is proposed that the driving force needed for this process can be reduced by exposure of the oxidizing interface to ultraviolet light or to laser light because of the increased populations of Si-Si broken bonds under these illumination conditions.

By considering the oxidation of (100) silicon at low temperatures, it was possible to show that $\alpha$-cristobalite should be the preferred $SiO_2$ form. However, because only ~2 Kcal/mole energy difference exists between the different forms of $SiO_2$ with an overall driving force of ~150 Kcal/mole, only a thin layer is expected to form before it transforms for strain reasons to the vitreous form. The (100) $\alpha$-cristobalite phase matches the (100) Si phase quite well if 4 silicon atoms per unit cell become interstitial atoms and diffuse either into the bulk Si or into the bulk $SiO_2$. Those that go into the $SiO_2$ (the majority) oxidize to form interstitial

Si-O-Si-O polymer chains. These chains swell the α-cristobalite lattice, causing it to be under a state of compressive stress and a state of inhomogeneous distortion so that it naturally transforms to the vitreous state. In addition, the chain ends are charged and become the residual $Q_{ss}$ at the end point of the oxidation process.

The interstitials that flow into the silicon side of the interface annihilate with the local vacancies, causing long range vacancy influx towards the interface. This "vacancy wind" produces the oxidation-enhanced diffusion effects that have been so important in device processing considerations. The excess interstitials increase in concentration with time and eventually exceed the supersaturation needed for the nucleation of interstitial clusters which produces stacking fault formation in the silicon at distances up to microns from the $Si/SiO_2$ interface. The exact calculation of these conditions requires the solution to the diffusion problem in the presence of an interface field which is made up of several components: (a) electrostatic field variations, (b) dielectric constant variations, (c) thermodynamic activity coefficient variations, (d) stress field variations, and (e) chemical bond type variations.

A method has been devised for solving the field-dependent time-dependent diffusion equation subject to the proper boundary conditions. It copies some of the procedures used in nuclear scattering theory, and it provides us with a set of successively better approximations to the solution in the region close to the interface and in the region far from the interface. Calculations are being made for the constant velocity regime of oxidation using an interface field that decays exponentially with distance from the interface.

In general, the theoretical work during the past year strengthens the model depicted in Fig. 2.2 and makes it apparent that all of the mechanisms indicated in the figure play important roles in the overall growth process.

## B. Process Dependence of Oxide Charges

Characterization and understanding of the process dependence of oxide charges, particularly interface states or traps, is very important in the fabrication of modern devices. The reductions in device size brought about by advancing technology have made the control of oxide charges and the fabrication of $Si-SiO_2$ interfaces with low electrical charge levels a necessity. Furthermore, serious attempts at comprehensive process and device modeling will have to include the contributions of oxide charges. The characterization of fixed oxide charge levels and their dependence on process parameters has been underway for some time [2.31-2.35], but a systematic characterization for interface states has not been carried out.

Interface states of various kinds are located at the $Si-SiO_2$ interface with energies in the silicon band gap. The total charge due to interface states at any given instant depends on the type of states (donor-like or acceptor-like), their energy distribution, and the surface potential. In the work reported here, only intrinsic interface states (i.e., those caused by the silicon oxidation process itself) are considered. In previously reported work partly funded by this contract, the effects of oxidation temperature, oxidation ambient, and in situ anneal and cool ambients were discussed. For oxidation and cool in dry $O_2$, midgap interface state densities were found to be dependent on oxidation temperature in a manner similar to that reported by Deal [2.31] for oxide fixed charge.

This result is in agreement with the work of Montillo and Balk [2.33] and Breed and Kramer [2.35]. The decrease in interface state density with increasing temperature is illustrated in Fig. 2.6 for both n- and p-type silicon of both (100) and (111) orientations. It should be noted that great care is necessary to obtain these and other results because of the sensitivity of interface state density levels to trace amounts of hydrogen (or moisture) in the oxidizing or anneal ambients.

Other results from previous work include the verification of the approximate 3:1 ratio of interface state density between (111) and (100) silicon, which is also characteristic of oxide fixed charge. Figure 2.7 shows the measured values of midgap interface state density for (111) vs that for (100) for various material and process variations, with the solid line showing an exact 3:1 ratio. Although some scatter is evident in the data, the general tendency is for an approximate 3:1 ratio. Similar observations were made by other investigators [2.31,2.35,2.36,2.37], and the similarity in results between oxide fixed charge and interface state densities points to a similar origin for both. This origin is intrinsic to the thermally oxidized Si-$SiO_2$ interface.

The relationship between oxide fixed charge ($Q_{ss}/q$) and interface state density at midgap ($N_{st}$) is further illustrated in Fig. 2.8. The data shown are for oxidation in a dry $O_2$ ambient at temperatures of 1000° and 1200°C. Variations in $Q_{ss}/q$ values are a result of varying material parameters such as (100) and (111) silicon, and process parameters such as different oxidation temperatures and cool conditions. The results indicated a correlation of $Q_{ss}$ and $N_{st}$ data prior to, as well as, following a low temperature hydrogen anneal. The actual hydrogen anneal parameters (temperature, ambient, and time) are critical to the determination of

final device parameters and the majority of fabricated devices are subjected to a low temperature hydrogen anneal in the final processing stages. Typical effects of variations in anneal parameters are illustrated in Fig. 2.9 where the p-type (100) wafers (from Fig. 2.6) are measured following hydrogen anneal for two different anneal conditions. The general tendency of lower $N_{st}$ values for higher oxidation temperatures is preserved with the anneal at 450°C clearly more efficient in the removal of interface states.

A detailed study to increase our understanding of the hydrogen anneal mechanisms was therefore deemed necessary in order to more fully characterize the interface state density dependence on process parameters. In the work reported here, the effects of low temperature anneal parameters (temperature: 350° to 500°C, ambient: 0% to 100% hydrogen in nitrogen, and time: 5 to 240 min) were investigated. In addition, two different cooling conditions were used to examine the effect of high temperature processing on the subsequent low temperature hydrogen anneal. N-type (111) silicon wafers were used in this investigation in order to obtain $N_{st}$ values sufficiently above the system's resolution estimated at approximately $10^{10}/cm^2$-eV at midgap.

Two major areas were investigated. One deals primarily with the interaction between hydrogen in the anneal ambient and the Si-$SiO_2$ interface for unmetallized structures. In this case, hydrogen in some active, atomic form (originating from the dissociation of molecular hydrogen) is thought to be the major factor in the annealing of interface states. The effect of process conditions, such as temperature, on the amount of atomic hydrogen present as well as the reaction between hydrogen and the interface states were investigated. Another area of interest deals with annealing

in the presence of an aluminum field plate. In this case, moisture at the Al-$SiO_2$ interface is thought to be the source of the active hydrogen [2.38] and very effective annealing can be obtained even in ambients containing no hydrogen.

## B.1 Experimental Procedure

N-type (111) silicon wafers, 4 to 6 Ω-cm resistivity were cleaned in sulfuric peroxide or hot sulfuric acid, aqua regia, 10:1 water:hydrofluoric acid, and isopropanol vapor, with appropriate deionized water rinses. They were then loaded into an oxidation furnace in dry $O_2$ and oxidized for a given time. Following oxidation, the wafers were annealed in situ in nitrogen, subsequently cooled in the anneal ambient (typical pull from the furnace was about 2 min). Where no anneal was required, the wafers were pulled in the oxidizing ambient (typically 1 to 3 sec). Two major experiments were carried out as follows.

1. To investigate the annealing of a bare Si-$SiO_2$ interface when exposed to a hydrogen/nitrogen ambient, the wafers were annealed using the temperature, ambient, and time of interest prior to metallization. Metal deposition of aluminum dots approximately 1 μm thick and 750 μm in diameter was carried out in vacuum at 25°C. To avoid radiation effects, a nonelectron beam flash evaporation system was employed.

2. To investigate the annealing of an Si-$SiO_2$ interface covered by an aluminum fieldplate either in nitrogen or in hydrogen/nitrogen mixtures, the wafers were metallized following oxidation (also by cold flash) and subsequently annealed using the desired anneal parameters.

Values of $N_{st}$, the interface state density, were determined by the quasistatic C-V technique [2.39,2.40]. This method is based on the proportionality that exists between the incremental MOS capacitance and the charging current in the structure when it is subjected to a linear voltage ramp. As a result of this proportionality, a low frequency thermal equilibrium MOS capacitance-voltage curve can be obtained. The quasistatic and high frequency curves are then used to extract the interface state density distribution in the band gap. This method is valid in the interval of the forbidden gap extending from the inversion threshold to a position about 200 mV from the majority carrier band edge. The relationship between the silicon surface potential and the applied voltage is obtained by the integration of the quasistatic C-V curve as proposed by Berglund [2.41]. The accuracy of the quasistatic method is estimated to be $10^{10}$/ $cm^2$-eV at midgap.

Measurements of interface state density are reported either in terms of midgap $N_{st}$ values or by showing the entire interface state density distribution. Midgap values will be used primarily when they are indicative of the major trends related to process variations.

## B.2 Results and Discussion

The results presented here will be divided in two sections, one dealing with premetallization low temperature hydrogen anneals (bare oxide anneals) and the other dealing with postmetallization hydrogen anneals. In each case, typical interface state density curves will be presented and their dependence on anneal temperature, ambient, and time will be discussed.

(a) Premetallization Anneals

The importance of studying the annealing behavior of bare oxides is that it allows one to examine the behavior of the Si-$SiO_2$ interface without the added complexity of the presence of a second metal-$SiO_2$ or poly-$SiO_2$ interface. Results obtained in this investigation will hopefully yield increased understanding of the hydrogen annealing of poly-Si field plates which are quite prevalent in modern devices.

Figure 2.10 shows typical interface state density energy distributions obtained from wafers oxidized in dry $O_2$, cooled in the oxidizing ambient, and subsequently annealed for 10 min (a) at various temperatures in a 100% hydrogen ambient, (b) at 400°C, in ambients containing various percentages of hydrogen in nitrogen. The experiments that were carried out included four temperatures (350°, 400°, 450°, and 500°C) and five different anneal ambients at each temperature (0%, 10%, 25%, 50%, and 100% hydrogen in nitrogen). The major conclusions are summarized below.

1. An optimum temperature of 450°C is observed for all hydrogen concentrations in nitrogen larger or equal to 10% for a 10 min anneal.

2. The most efficient annealing is observed for 100% hydrogen anneal ambients at all temperatures examined.

3. The annealing of interface states above midgap proceeds at a much faster rate than that for states below midgap. A peak of interface states is observed at approximately 0.2 eV below midgap which anneals rather slowly with increasing hydrogen concentration in the anneal ambient, and which reaches a minimum at 450°C.

4. At an anneal temperature of 500°C, a reinducement of interface states can be observed particularly below midgap.

The data showing a best anneal temperature of 450°C are in agreement with previous work published by Castro and Deal [2.42] in which surface recombination velocity was determined using gate controlled diodes. The fast annealing of interface states above midgap may be due to the nature of the states (donor-like or acceptor-like) present. It was previously observed [2.10] that interface states in wafers cooled rapidly ($O_2$ FP) in the oxidizing ambient in an open tube are predominantly donor-like and show a peak of states located below midgap, while wafers cooled in nitrogen indicated the presence of both donor- and acceptor-like states. The rapid anneal of acceptor-like states above midgap in the presence of a small amount of hydrogen may explain these observations. In the case of cooling procedures during which the wafers are in contact with moisture, a rapid anneal of acceptor-like states above midgap occurs and the presence of predominately donor-like states below midgap is observed.

The mechanism proposed for interface state annealing in hydrogen at low temperature is illustrated in Fig. 2.11 and for bare oxides is characterized by the two reactions:

$$H_2 \underset{\leftarrow}{\overset{k_1}{\rightleftarrows}} 2(H\cdot) \qquad (2.2)$$

and

$$H\cdot + \begin{array}{c} | \\ -Si \\ | \end{array} \underset{\leftarrow}{\overset{k_2}{\rightleftarrows}} \begin{array}{c} | \\ -SiH \\ | \end{array} \qquad (2.3)$$

The first reaction most likely goes to the right with increasing temperature, while reaction (2.3) goes to the left at temperatures above

450°C. This would explain the "reinducement" of interface states observed by heating previously annealed samples at 500°C in nitrogen [2.43,2.44] as well as the reinducement observed when heating in vacuum [2.33]. An increase in interface state density was also noted during this investigation when annealing in a 100% hydrogen ambient.

The effect of anneal time can be summarized as follows:

1. For anneals in a 100% hydrogen ambient, interface state density levels were observed to decrease with increasing time at lower temperatures (300° or 350°C) and for times exceeding 4 hours.

2. For samples annealed in hydrogen/nitrogen mixtures, increasing time did not lead to a reduction in interface states.

(b) Postmetallization Anneals

This investigation deals with the characteristics of postmetallization anneals. In these experiments, the wafers were annealed following cold flash aluminum deposition. Typical interface state density energy distributions are flat in the band gap portion where the measurement is valid, except for some wafers receiving a less than optimum anneal where a small peak below midgap is still observed.

Figure 2.12 shows the values of interface state density at midgap for various low temperature hydrogen anneal parameters. The data are for n-type (111) wafers, 4 to 6 Ω-cm resistivity, oxidized in dry $O_2$ at 1000°C and cooled in the oxidizing ambient ($O_2$ FP). The optimum anneal temperature is clearly 400°C for any percentage of hydrogen in nitrogen as well as strictly nitrogen anneal ambients. Similar results have been reported previously [2.42]. No basic differences were observed between $O_2$ cooled

and $N_2$ annealed/cooled wafers although a lower level of interface states was obtained from $N_2$ annealed/cooled samples. These results have also been observed previously [2.10].

The mechanism for interface state annealing in this case is quite similar to that described for premetallization anneal except for the source of the active hydrogen which is postulated to come from the interaction between the aluminum field plates and moisture in the following manner:

$$2Al + 3H_2O \overset{k_3}{\underset{}{\rightleftharpoons}} Al_2O_3 + 6H\cdot \qquad (2.4)$$

The atomic hydrogen generated contributes to the annealing of interface states as described by reaction (2.3). The optimum temperature where both reactions (2.3) and (2.4) are favorable to the annealing of interface states seems to be 400°C. From data obtained from premetallization anneals, reaction (2.3) was observed to be clearly unfavorable at temperatures exceeding 450°C while, at low temperature, the reaction rate $k_3$ may be too low. Long anneal times carried out at 400° to 500°C do not show a reduction but rather a slight increase of interface states for ambients containing hydrogen in nitrogen, particularly at higher temperatures (500°C).

## B.3 Summary

The dependence of interface state densities on low temperature hydrogen anneal parameters has been investigated. The results indicate that the best annealing condition for premetallization anneal is at 450°C and in a 100% hydrogen ambient. In the case of postmetallization anneals,

the best parameters are 10 to 15 min, 10 to 25% hydrogen in nitrogen at 400°C.

## C. Experimental Evaluation of the Optical and Structural Properties of Ultrathin Layers of $SiO_2$ on Silicon

The structural properties of silicon dioxide films thermally grown on single crystal silicon substrates have been the subject of many investigations [2.45-2.51]. Atalla et al studied $SiO_2$ films grown at 1000°C in dry and wet oxygen using electron diffraction [2.45]. No crystalline structure was observed, and they assumed an essentially continuous and amorphous film. Edagawa et al studied $SiO_2$ films grown on (111)-oriented silicon over the temperature range of 1000° to 1300°C in various oxidizing gases such as saturated water vapor, wet oxygen, wet nitrogen, dry oxygen, and wet hydrogen, using the electron diffraction method [2.46]. They also studied $SiO_2$ films formed by thermal decomposition of ethyltriethoxysilane. Diffuse halo patterns were observed by the transmission electron diffraction, indicating that the oxide films consisted of small crystallites of $\alpha$-cristobalite. They also observed that the crystalline state of the oxide is unaffected by the oxidizing ambient and that the grain size depends on the temperature, increasing with annealing.

Knopp and Stickler studied thermal oxide films grown on (111)-oriented silicon by wet or wet-dry oxidation processes using x-ray, transmission and reflection electron diffraction [2.47]. Uniform amorphous oxide films were obtained by oxidation in open-tube systems at temperatures between 990° and 1200°C. A phosphorus-glass deposition treatment at temperatures below 1150°C did not affect the amorphous state of the

oxide. Heat treatments of the phosphorus-deposited oxide film at temperatures above 1200°C resulted in nucleation and growth of crystallized islands in the amorphous film. These islands consisted mainly of a mixture of high ($\beta$) and low ($\alpha$) temperature cristobalite and their sizes varied with heat treatment conditions. Uniform crystalline films were obtained in sealed quartz ampoules containing oxygen at 1235°C. These films consisted of $\alpha$-cristobalite.

Sugano et al studied thermally grown silicon dioxide films on silicon using electron diffraction and examined the effect of impurities intentionally added during oxidation [2.48]. They found that among the undoped-$SiO_2$ samples, oxides grown in wet oxygen tended to crystallize more than in dry oxygen. This led to the assumption that some hydrogen-associated species play a significant role in the crystallization. This was confirmed by the structural analysis of hydrogen-annealed samples. Impurity-related crystallization showed boron enhancing it, phosphorus and antimony suppressing it, and gallium having no apparent effect. The crystallization of $SiO_2$ observed was localized.

Nagasima analyzed $SiO_2$ films grown at 1100°C in dry $O_2$ or wet $O_2$ at 1100°C by electron diffraction and infrared adsorption [2.49]. Applying the Bragg particle size equation, he found that the average $\alpha$-cristobalite microcrystal dimension (as proposed by Edagawa et al) was comparable to the size of the unit cell of $\alpha$-crystobalite. Such particles can hardly be called crystals. He concluded, from the above consideration and the infrared adsorption spectra, that the thermal oxide films on silicon have short range order similar to that of fused silica in which the $SiO_4$ tetrahedra are three-dimensionally joined with each other lacking long range order.

Alessandrini and Campbell found that the transformation from the amorphous to a crystalline phase was dependent on the catalytic behavior of phosphorus [2.50].

In the present work, x-ray diffraction and transmission electron diffraction were used to study ultrathin layers of $SiO_2$ (<100 Å) thermally grown on (100) orientation single crystal silicon. Results indicate, under some oxidation conditions, a 30-40 Å crystalline layer of $\alpha$-cristobalite at the Si-$SiO_2$ interface. The remainder of the $SiO_2$ film is assumed to be amorphous. The transmission electron diffraction samples were prepared for measurement and characterized by Dr. Tom Magee of Advanced Research and Applications Corp. partially under support from Stanford. His cooperation and substantial help with these measurements is gratefully acknowledged. The remainder of the measurements were performed at Stanford.

## C.1 Sample Preparation

$SiO_2$ was thermally grown at temperatures ranging from 700° to 1100°C in dry oxygen in a resistance-heated furnace. The single silicon substrate was Czochralski-grown, (100)-oriented, phosphorus doped in the 5-10 Ωm range. Immediately prior to the oxidation, the wafers were cleaned using the following procedure: heat in $H_2O_2$:$H_2SO_4$ - DI rinse - heat in 5:1:1 $H_2O$:$H_2O_2$:$NH_4OH$ - DI rinse - heat in 5:1:1 $H_2O$:$H_2O_2$:HCl - DI rinse - dip in 50:1 $H_2O$:HF - thorough DI rinse - blow dry using nitrogen. The samples used in transmission electron diffraction study were cut into pellets 2.5 mm × 2.5 mm prior to the cleaning which included an additional careful degreasing step using ultrasonic agitation.

After oxidation, the samples used in the transmission electron diffraction study were then mounted on a glass slide with paraffin and subjected to conventional jet thinning from the back surface. An $NHO_2$:HF (5:1) solution was used for thinning at a normal flow rate of ~6 ml/min. Under these conditions, a depth removal rate of ≃75 μm/min is attained, producing a bowl-shaped depression at the back-surface and a thin electron transparent area in the central region of the front surface. However, in the present experiments, after thinning for approximately 3 min at 6 ml/min, the solution flow rate was reduced to 3 ml/min for more adequate control during the final stages of etching. The slide was periodically removed during this stage of jet thinning and examined in an optical microscope. When the silicon attained a thickness of several thousand Angstroms, interference colors from the $SiO_2$/Si layer were observed. Thinning was terminated when the yellow-green color from the Si layer (≤1000 Å) disappeared, leaving only the thin $SiO_2$ film at the surface. Under these conditions, $SiO_2$ films of 50 to 100 Å thickness were obtained suspended across a central hole of 50 to 100 μm diameter. The specimens were then removed from the slide by immersing in a warm benzene solution followed by a DI water rinse. A series of benzene/DI water rinses were used to remove any contamination from the $SiO_2$ film and the samples subsequently blow-dried for examination in the TEM.*

*The procedures described in this and the following several paragraphs were carried out by Dr. Tom Magee of Advanced Research and Applications Corp. [2.51].

## C.2 Transmission Electron Diffraction Study

### (a) Measurement of Plane Spacings and Calibration Procedures

To obtain information on the structure of "d" spacings in a diffraction pattern, it is essential to use a standard or reference sample for calibration. If a material such as Au, either in the form of thin films or microcrystals, is used, the radii, R (mm), of polycrystalline rings or single crystal spots can be converted to d-values (in Å) using the relation:

$$d = \frac{L\lambda}{R} \tag{2.5}$$

where L (mm) is the object to focal plane distance and $\lambda$ is the electron wavelength in Å. The quantity, $L\lambda$ (mm-Å), is called the camera constant and is used routinely in diffraction pattern measurements. Since the d-spacings of Au are very well known, measurements of R and d give an accurate value for the camera constant which is thus experimentally determined for a fixed lens setting and electron accelerating voltage.

In these experiments, it was desirable to obtain a very accurate calibration of the camera constant since extremely thin $SiO_2$ layers of unknown phase on bulk Si were to be identified. Consequently, thin (≃100 Å thick) gold microcrystals deposited from solution on top of the $SiO_2$ layer were used for TED measurements. The Au microcrystals were prepared by a chemical reduction of chlorauric acid with salicylic acid. The crystals were grown at room temperature by adding 4 ml of 0.1% salicylic acid to 98 mls of chlorauric acid so that the final solution contained a 0.5 millimole concentration of Au. After 100 hours of growth,

the microcrystals were ≃100 Å thick and ≃1.2 μm across. The crystals were transferred in solution to the surface of the $SiO_2$ film on the prepared TEM sample. After drying, diffraction patterns were obtained from the Au microcrystallites and the camera constant determined from measurements of the spot distances using the previous equation.

(b) Radioisotopic Image Enhancement and Diffraction Analysis

In an extremely thin layer containing more than one phase, the diffraction intensities from the constituent containing the smaller number of crystallites are generally low and experimental resolution is often obscured or absent in routine analysis procedures. In the current application where the $SiO_2$ layer is largely amorphous, the presence of a thin crystalline layer at the $SiO_2$/Si interface is not readily detected in normal exposures on a photographic plate because of the dominant broad amorphous rings that mask any weak lines from a crystalline phase. Underexposing diffraction plates will reduce the background intensity from amorphous rings or "halos", but will simultaneously reduce the intensity of weak lines from the crystalline material, thereby eliminating practical detection or identification by conventional film development procedures.

To retrieve diffraction data from crystalline $SiO_2$ layers, a radioisotope image enhancement technique developed from a modified procedure of Thackray for photographic film intensification was used [2.52, 2.53]. This technique has been previously used to identify gold-silicide reactant layers at the Au/Si interface and in later studies of reactions in the Au/GaAs system [2.54, 2.55]. Briefly, the technique involves exposing

a diffraction plate in a solution containing a radioisotope. The radioisotope attaches to exposed silver grains of the film and becomes a spatially defined radioactive source. The film is then used to produce an autoradiograph of the toned film.

In this application, the sulphur-35 isotope ($E_\beta$ = 0.167 MeV, t 1/2 = 87 days) was used. The radioactive sulphur was mixed in a 1:3 ratio with anhydrous potassium carbonate and subsequently heated to 280°C in a paraffin bath under a stream of nitrogen. After cooling, the residue was taken up in a 2% solution of $K_2CO_3$ and stored under nitrogen for use as the stock solution. Approximately 1 μCi of the sulphur-35 was used for each square centimeter of film to be intensified. For toning of diffraction plates, the solution was stirred in contact with the film for periods of 3 to 9 days. After removal from the toning solution, the film was washed with 2% $K_2CO_3$, followed by exposure to an inactive polysulphide solution and a subsequent rinse in a 2% $K_2CO_3$ solution. The film was washed in running water and dried in the final steps.

To produce autoradiographs, the toned film was placed in direct contact with an unexposed electron image plate or film and sealed in a lead-lined light-tight box for periods of 3 to 24 hours, depending upon the level of intensification required. The exposed plate was removed and developed by conventional procedures and the autoradiograph carefully washed to remove any residual gelatin or adhering radioactive material possibly transferred from the toned plate during exposure. From the autoradiographs an intensification factor of x10 - x12 relative to conventionally processed plates was obtained, and diffraction lines not observed on normal plates are readily detected.

(c) Electron Diffraction Results

Examination of $SiO_2$ films of 50 Å and 100 Å thickness with the TEM showed the surfaces to be free of contamination, cracks, or other defects. In all cases, we observe a relatively structure-free surface and an absence of preferential growth or nucleation zones typically noted at defect sites within the substrate at the oxide/semiconductor interface.

Figure 2.13 shows representative bright-field transmission electron micrographs of a 50 Å oxide layer grown at 800°C. The low magnification micrograph in Fig. 2.13a shows the morphology of the thinned sample and the suspended $SiO_2$ film across the hole produced by the jet thinning process. Figure 2.13b is a high magnification bright-field micrograph obtained at the center of the region in Fig. 2.13a showing the absence of structure and macroscopic growth defects within the film. Shown in the inset is a selected area transmission electron diffraction pattern obtained from the 50 Å layer. The plate was slightly underexposed to show the position of crystalline spots and arcs (arrowed) relative to the amorphous diffraction rings produced by the matrix. As can be observed, diffraction lines are typically weak and largely obscured in conventional analyses, making possible identification of crystalline phases extremely difficult in thin $SiO_2$ layers.

Table 1 lists representative data obtained from a number of $SiO_2$ films of varying thickness, $t_{ox}$, prepared at growth temperatures, $T_g$, of 700°C and 800°C. For reference, we also list d-values for $\alpha$-cristobalite (ASTM-11-695) and corresponding (hkl) indices. Measurements obtained from plates processed by conventional techniques yield d-values consistent with the tetragonal $\alpha$-cristobalite phase. In many

cases, however, the strongest reflection (101) is obscured by the central spot and the scattered background intensity of the amorphous matrix. Using plates that were deliberately underexposed to reduce the background fog level, radioisotopic image enhancement techniques were applied and autoradiographs obtained. From the intensified plates, we were able to clearly identify the three major lines of α-cristobalite. In over 100 diffraction patterns examined, we found only tetragonal α-cristobalite and were unable to detect any evidence of the cubic (α-cristobalite) form or other phases of $SiO_2$. Additional TED analyses obtained in the near surface region of the substrate (after stripping the $SiO_2$ layers) also showed no evidence of polycrystallinity of the Si or prominent alteration in the single crystal structure.

## C.3 X-Ray Diffraction Study

### (a) Experimental Procedure

Single crystal silicon wafers were oxidized as described previously. X-ray diffraction spectra were obtained using a Norelco diffractometer equipped with a CuKα x-ray source [$\lambda(K\alpha_1) = 1.54051$ Å]. The intensity of the diffracted beam is measured directly (by means of counting the number of current pulses caused by the ionization due to a single x-ray quantum) and is recorded as the angle $2\theta$ is swept from ~20° to ~100°. The sample and detector supports are mechanically coupled so that a rotation of the counter through $2\theta$ degrees is automatically accomplished by rotation of the wafer by $\theta$ degrees. This ensures that the angles of incidence and reflection will be equal at all times and equal to half the

total angle of diffraction. This arrangement preserves the focusing conditions of the reflected beam at the detector.

Transmission electron diffraction (TED) and x-ray diffraction (XRD) are complementary techniques in that TED is sensitive to all planes except those parallel to the silicon substrate and XRD, in the diffractometer described above, is only sensitive to those planes parallel to the silicon substrate.

(b) Diffraction by Very Small Crystals

Consider a crystal consisting of $N$ lattice planes of spacing $d$, such that its thickness is $L$ [2.56]. If $A$ is the amplitude diffracted by a single lattice plane, then the reflected ray has an amplitude $NA$ if the incident ray of wavelength $\lambda$ is incident at the Bragg angle of incidence $\theta_0$ (such that $\lambda = 2d \sin \theta_0$). For angles of incidence $\theta$ differing from $\theta_0$ by a small angle $\epsilon$, it can be shown that the spectrum of the resulting reflections is given by [2.56].

$$A \frac{\sin (2\pi ND\epsilon \cos \theta_0/\lambda)}{\sin (2\pi d\epsilon \cos \theta_0/\lambda)} \tag{2.6}$$

The integral line width $\Delta(2\theta)$ is defined as the width of a line of rectangular profile which would have the same maximum and integral values of the observed line. Using the Scherrer formula, we get

$$\Delta(2\theta) = \frac{\lambda}{L \cos \theta_0} \tag{2.7}$$

This equation can be applied to crystals of approximately 10 to 1000 Å in size [2.56].

(c) Diffraction Results

Figures 2.14 and 2.15 show typical x-ray spectra obtained on (100) and (111)-oriented silicon substrates, respectively. The spectra were independent of the temperature at which the oxide was grown (in the 700° to 1000°C range). For the (100)-oriented substrate, the (004) reflection of Si appears at 69.2°, and a $\lambda/2$ reflection of the same spacing occurs at 32.95° superimposed on a broad low-intensity peak at 33.8°. Tungsten peaks at 66.4° and 65.8° were also observed, and they resulted from tungsten related lines in the $CuK\alpha$ radiation.

The ratio of the intensity of the broad peak at 33.8° to that of the silicon (004) peak at 69.2° is approximately $1:10^5$, and its $\Delta(2\theta)$ line width corresponds to a minimum thickness of 30 Å.

The intensity of the x-ray is known to obey a simple exponential relation

$$\begin{aligned} I &= I_0 e^{-\mu x} \\ &= I_0 e^{-(\mu/\rho)\rho x} \end{aligned} \tag{2.8}$$

where $(\mu/\rho)$ is the mass absorption coefficient. For $CuK\alpha$ radiation

$$\left(\frac{\mu}{\rho}\right)_{Si} = 60.6 \text{ cm}^2/\text{g} \tag{2.9}$$

and

$$\left(\frac{\mu}{\rho}\right)_{SiO_2} = 43.22 \text{ cm}^2/\text{g} \tag{2.10}$$

The corresponding characteristic attenuation length is 70.82 and 101.9 μm in Si and $SiO_2$, respectively. Assuming no practical attenuation in

~100 Å of $SiO_2$ and approximately 5 to 10 crystalline layers of $SiO_2$, the ratio of $SiO_2$ to Si signals is approximately $1:10^5$ (as observed).

It was noticed that changing the orientation of the oxidized silicon wafer in its plane affected both the Si peak at 69.2° and the weak broad peak at 33.8°. For some orientations, the 33.8° peak disappeared and the Si signal was correspondingly at its minimum value. A spectrum identical to that of Fig. 2.14 was obtained on the same sample immediately after etching the $SiO_2$ layer using 50:1 $H_2O$:HF solution and rinsing in DI $H_2O$. The time required to obtain the x-ray spectrum was sufficient to grow a native oxide at room temperature. The 33.8° peak corresponds to a "d" spacing of 2.6496 Å. The (200) peak for Si is a forbidden reflection with a spacing of 2.714 Å, but the slight misorientation of the silicon substrate from (100) must be taken into consideration. This small misorientation of the (100) orientation is responsible for the variation of the Si signal as the wafer is rotated in its plane. By changing the lower level and window width settings on the diffractometer, it was possible to eliminate the 33.8° peak completely without altering the intensity of the silicon signal substantially.

As suggested from the TED study, a layer of α-cristobalite exists at the interface. The only α-cristobalite plane arrangement that can be detected by the x-ray diffractometer corresponds to the (004) plane. The spacing between (004) planes is 1.73 Å in the α-cristobalite form of $SiO_2$ and that corresponds to a peak at 52.9°. Such a peak was never observed in the x-ray study done here. As a result, we were unable to confirm the presence of the α-cristobalite crystalline layer with the TED measurements. This may be simply a result of the difficulty of the experimental technique.

## C.4 Ellipsometry Study

Ellipsometric measurements $(\Delta,\Psi)$ were made on (100) and (111)-oriented silicon substrates which had been oxidized in a dry oxygen ambient at 900°C. The optical properties of thin layers of $SiO_2$ set a limit on the validity of the use of ellipsometry in determining such constants. It has been regularly assumed that the optical properties of ultrathin layers remain unchanged from those of much thicker oxides [2.57]. Index of refraction values $(n_{SiO_2})$ of 1.45 to 1.47 have been routinely used in calculating oxide thicknesses using ellipsometry. However, if the obtained ellipsometric parameters $\Delta$ and $\Psi$ are used to find both the index of refraction $(n^*_{SiO_2})$ and the thickness $(X^*_{ox})$ of such layers, the results obtained in Fig. 2.16 are found.

It is noted that the index of refraction is enhanced over its thick oxide value, and the enhancement increases as the thickness decreases. Correspondingly, the thickness $(X^*_{ox})$ obtained by using $n^*_{SiO_2}$ instead of 1.46 is monotonically lower than that obtained using 1.46. The difference between $X^*_{ox}$ and $X_{ox}$ increases as the thickness decreases. For all practical purposes, it is observed that for oxide thicknesses greater than ~500 Å, the optical properties assume the thick oxide value of 1.46. It should be noticed that the above mentioned results were based on the assumption that there is no appreciable absorption in thin layers of $SiO_2$ (i.e., $K^*_{SiO_2}$ was assumed to be zero). The above results could be interpreted as the presence of optical layers at the Si-$SiO_2$ interface whose optical properties are distinguishly different from thick layers of $SiO_2$.

As an example, the following simulation was carried out using McCraken's NBS ellipsometry program [2.58]. Assume a layer with $\Delta$ = 121.366 and $\Psi$ = 14.131. A search of both $n^*_{SiO_2}$ and $X^*_{SiO_2}$ results in

$$X^*_{ox} = 200 \text{ Å}$$

$$n^*_{SiO_2} = 1.8$$

If $n_{SiO_2} = 1.457$ were used as an index of refraction, the thickness obtained would have been 240.9 Å. These overall $\Delta$ and $\Psi$ were used in modeling this layer with a two layer oxide of different optical properties as shown in Fig. 2.17.

The result for $n_2$ to match both $\Delta$ and $\Psi$ is approximately 2.75. Therefore, we can model the observed results on $n^*_{SiO_2}$ by the presence of an optical layer at the interface. It is interesting to note that the index of refraction of $\alpha$-quartz is 2.4 at the energy of the He-Ne laser used in the ellipsometry measurements. Similar results were recently reported by Taft and Cordes [2.59]. They found that the thickness of this optical layer depended on the temperature at which the oxide was grown.

Ellipsometry was repeated after successive etching of layers of $SiO_2$ using a dilute solution of 50:1 $H_2O$:HF for 20 sec each on (100) silicon wafers oxidized at 700°C for 9 and 18 hours. In calculating the oxide thickness, the optical properties of thick oxides were used ($n = 1.457$). It was expected that, with successive etching, the error caused by using 1.457 in the ellipsometry calculations would increase monotomically. However, the trend in the error reversed as the thickness decreased

below 30 to 40 Å. These results are in accord with the thickness evaluation by using the chemically shifted Auger spectra [2.60]. This study showed a thickness constantly thinner than that obtained by ellipsometry assuming the optical properties of thick $SiO_2$ layers.

It has also been shown [2.61] that the refractive index of $SiO_x$ layers measured at 0.8 eV photon energy range from 3.3 to 2.3 as x changes from 0.1 to 1.2. According to these results, this optical interface layer should therefore be a layer of $SiO_{0.5}$. These results should be carefully interpreted, however, because they are based, at present, on a small sample of measurements.

## C.5 Conclusions

The atomic structure and the optical properties of the Si-$SiO_2$ interface were studied by transmission electron microscopy, x-ray diffraction, and incremental ellipsometry. TED showed that the interface is crystallized in an $\alpha$-cristobalite structure on the oxide side of the interface. X-ray diffraction results were inconclusive in determining whether the interface has long or only short range atomic ordering. Incremental ellipsometry showed that the interface has an optical layer of 5 to 10 Å thickness with a large index of refraction compared to that of thick $SiO_2$ layers. The possibility of having any crystallization at the interface was tested by trying to deposit an epitaxial layer on top of a possibly crystalline oxide; the result was however negative.

The above studies were done in order to evaluate the physical properties of thin layers of $SiO_2$. These ultrathin layers exhibit high growth

rates (predominantly logarithmic), enhanced optical properties, and an electric breakdown field higher than that of thicker layers of silicon dioxide [2.62]. These studies were essential before a systematic evaluation of growth kinetics (ellipsometry thickness evaluation) and charge characterization (using CV techniques) could be attempted.

It has been shown [2.63] that point defects in the oxide layers (namely oxygen vacancies and silicon interstitials) play an important role in the initial fast oxidation regime at low partial pressures. We expect these defects to play important roles also in our investigation of the oxide charges associated with thin $SiO_2$ layers.

## D. Physical Modeling of the Enhanced Oxidation of Heavily Doped Silicon and Its Implications

Enhanced thermal oxidation of heavily doped silicon [2.64] remains a phenomenon commonly encountered in integrated circuit processing and, therefore, continues to be of importance and interest in process technologies. A physical model has been developed [2.65,2.8,2.9] to explain the increase of the linear rate constant B/A (directly proportional to the $Si/SiO_2$ interface oxidation reaction rate constant $k_s$) of the Deal and Grove oxidation model [2.1] resulting from high dopant levels. The model enables calculation of the effective B/A under a wide variety of oxidation conditions.

In this report, the model is first reviewed and then quantitative predictions of the model are compared to experimental data. The effects of the oxidation ambient, silicon substrate orientation, specific donor and acceptor impurities, and compensating dopants are tested. The

physical significance of the model and possible implications for oxide-related and other process phenomena are examined.

## D.1 Summary of Model

It has been suggested [2.4] that the interface reaction rate constant $k_s$ and therefore the linear rate constant B/A commonly used to describe silicon oxidation kinetics [2.1] could include several "additive" components. These components result from different contributions of the silicon substrate to the $Si/SiO_2$ interface oxidation reaction; a silicon vacancy contribution was postulated to be one such term. The electrical effects of high dopant levels enhance the vacancy contribution to increase B/A. The effective linear rate constant may be expressed as the product of the intrinsic $(B/A)^i$ for lightly doped silicon and a vacancy-dependent multiplicative factor,

$$(B/A) = (B/A)^i \left[1 + \left(\frac{K}{C_1}\right) C_{V_T}^i e^{+2.0/kT} \left(\frac{C_{V_T}}{C_{V_T}^i} - 1\right)\right] \qquad (2.19)$$

where normalized quantities may be defined,

$$(B/A)/(B/A)^i \equiv (B/A)'$$

$$K/C_1 \equiv K'$$

$$C_{V_T}/C_{V_T}^i \equiv C_{V_T}'$$

and other quantities are defined as in [2.4]. The vacancy-related factor is calculated directly from vacancy statistics, using published physical parameter values, except for the single unknown $K'C_{V_T}^i$. This unknown was determined empirically (see Case A below) [2.4] to be,

$$K'C_{V_T}^{i} \approx 2.62 \times 10^{3}\, e^{-3.10/kT} \qquad (2.12)$$

Analysis in [2.4] suggested that the vacancy contribution (designated $R_2$ in [2.4]) is negligible for low substrate doping less than the intrinsic carrier concentration at oxidation temperatures (e.g., $n_i \approx 10^{19}$ cm$^{-3}$ at 1100°C); other components (designated $R_1$ in [2.4]) of the linear rate constant dominate to produce $(B/A)^{i}$ for lightly doped silicon. In addition, the modeling in [2.8] indicated that the $(B/A)'$ resulting from high dopant levels is likely independent of the oxidation ambient ($O_2$ or $H_2O$). The impact of substrate orientation is uncertain. Heavy donor doping should substantially increase B/A throughout the oxidation temperature range; however, high acceptor levels should raise B/A only at higher temperatures. The validity and effectiveness of Eqs. (2.11) and (2.12) in calculating B/A under a variety of oxidation conditions will now be tested.

## D.2 Comparison With Experiment

### (a) Case A: Phosphorus (Donor) Doping, (111) Orientation, $O_2$ Ambient

An extensive set of experimental data was described in [2.7] for the thermal oxidation of heavily phosphorus-doped (111) silicon in a dry $O_2$ ambient in the temperature range of 800° to 1100°C. A substantial enhancement of effective B/A was observed although with little change in the apparent activation energy. This case became the first comprehensive test for the vacancy-contribution model introduced in [2.4].

The experimental $(B/A)'$ values were matched to the corresponding effective $n > n_i$ values listed in Table 1 of [2.4]. The model, as

expressed in Eq. (2.11), was then fitted to the experimental (B/A)' values to obtain the effective $K'C_{V_T}^{i}$ in Eq. (2.12).

With this one fitted parameter, the (B/A)' dependence on heavy phosphorus-doping levels through $n > n_i$ could be calculated. Comparisons of these calculated (B/A)' curves to experimental (B/A)' data points as functions of n are plotted in Fig. 2.18 at oxidation temperatures of 800° to 1100°C, respectively; agreement is observed to be good. More detailed discussion of these experimental results is contained in [2.4].

(b) Case B: Phosphorus (Donor) Doping, (100) Orientation, $O_2$ Ambient

Possible influences of substrate orientation on the interface-oxidation effects may not be obvious and should be verified empirically. Limited data are available [2.66] for the oxidation of phosphorus-doped (100) silicon in dry $O_2$ at 780° to 1150°C. These data are restricted to a doping level measured at 0.0009 Ω-cm which corrects [2.67] to a carrier concentration of $n = 7 \times 10^{19}$ $cm^{-3}$. Although limited to one doping level, comparison of these data to Case A should indicate orientation effects.

Figure 2.19 replots the (100) B/A values [2.66] for lightly and heavily doped silicon vs 1/T. The lowest temperature data, at 780°C, presented in [2.66] have not been included because of apparent deviation of the lightly doped $(B/A)^i$ at that temperature from Arrhenius behavior. Other reports [2.68,2.69] have indicated that Arrhenius behavior extends at least to 700°C. This discrepancy is not understood at present.

The (100) data may be compared qualitatively to the corresponding (111) data of Case A (see Fig. 2.13 of [2.4]). The enhancement of B/A

through heavy phosphorus doping with little change of apparent activation energy can be seen to be similar for both orientations. Direct quantitative comparison must take into account the orientation dependence of $(B/A)^i$ in lightly doped substrates [2.68-2.70]

$$\frac{(B/A)^i(111)}{(B/A)^i(100)} = \frac{C_1(111)e^{-2.0/kT}}{C_1(100)e^{-2.0/kT}} \approx 1.7 \qquad (2.13)$$

where Arrhenius behavior has been assumed for $(B/A)^i$, with pre-exponential $C_1$ and activation energy of 2.0 eV. Comparison of the normalized $(B/A)'$ should indicate any orientation dependence in the vacancy-dependent multiplicative factor in Eq. (2.11), or specifically in the $K/C_1$ ratio. If the vacancy contribution contains the same orientation dependence as the other additive components of B/A, then the $K/C_1$ ratio and the vacancy factor should be identical for (100) and (111) silicon at all doping levels.

Figure 2.20 compares the $(B/A)'$ values for the two orientations. The (100) results of [2.66] for lightly doped Si and the single heavy phosphorus doping level are plotted as normalized $(B/A)'$ data points vs 1/T. Considerable scatter in the data is present (apparent also in Fig. 2.19). Also plotted is the $(B/A)'$ vs 1/T curve calculated for that heavy doping level, using the $K'C_{V_T}^i$ in Eq. (2.12) as determined from the (111) results of Case A. The agreement between the (100) points and the (111) calculated curve is not unreasonable; the (100) points may be slightly lower than the (111) curve. The $K/C_1$ ratio and the vacancy factor in Eq. (2.11), or the enhancement by heavy phosphorus doping of the linear rate constant relative to the lightly doped silicon value, are at least comparable for (111) and (100) orientations.

(c) Case C: Arsenic (Donor) Doping, (111) Orientation, $H_2O$ Ambient

Since the postulated dependence of B/A on heavy doping is electrical in nature, different donor impurities yielding the same carrier concentration n should produce similar effects on B/A and identical (B/A)' values. Also, ambient effects on B/A in Eq. (2.11) are likely contained entirely within the $(B/A)^i$ factor. As a result, the (B/A)' values should be identical for $O_2$ and $H_2O$ oxidations [2.8].

Both of these assumptions can be tested by studying data reported [2.71] for the oxidation in $H_2O$ (wet oxygen, 95°C water bubbler) of (111) silicon doped with arsenic to surface chemical concentrations of $2.5 \times 10^{20}$ to $2.2 \times 10^{21}$ $cm^{-3}$. Qualitatively, the arsenic $H_2O$ ambient data are very similar to the phosphorus dry $O_2$ ambient data in Case A--an enhancement of overall B/A with little change of apparent activation energy. Quantitatively, the (B/A)' values for similar n values should be identical. The chemical As levels of [2.71] must be corrected first to the resulting resistivity values [2.72] and then from resistivities to carrier concentrations [2.73].

Figure 2.21 compares the normalized experimental As, $H_2O$ ambient (B/A)' data points to the calculated (B/A)' curves (based on the phosphorus, dry $O_2$ ambient analysis in Case A) at 750° and 850°C, respectively. (No lightly doped control $(B/A)^i$ data at 650°C were reported in [2.71].) Again, the agreement is good.

(d) Case D: Boron (Acceptor) Doping, (100) Orientation, $O_2$ Ambient

To this point, emphasis has focused on high donor impurities. Vacancy charge states have been neglected, and total vacancy concentration

has been used as the relevant parameter in altering the interface-reaction rate constant $k_s$. This has been done because charge-state effects may not be straightforward [2.8]. With this simplistic formulation, the modeling for donor impurities has been successful.

If only total vacancy concentration is considered and the value of $K'C_{V_T}^{i}$ in Eq. (2.12) is assumed to be relevant for acceptor impurities as well as for donors, then Eqs. (2.11) and (2.12) can be extended to heavy substrate doping by such acceptor impurities as boron. Calculated results based upon this assumption are shown in Fig. 2.22 [2.8]. Data for (100) silicon samples of a single boron-doping level were included with the (100) phosphorus dry $O_2$ oxidation work (Case B) in [2.66]. This boron level produced a measured resistivity of 0.0016 Ω-cm which corrects [2.74] to a carrier concentration of $p = 1 \times 10^{20}$ cm$^{-3}$.

Figure 2.23 replots the (100) B/A values [2.66] for lightly doped and heavily boron-doped silicon vs 1/T. Again, as in Case B, the lowest temperature (780°C) data have not been included because of apparent deviation of the lightly doped $(B/A)^{i}$ from Arrhenius behavior. Qualitatively, the data closely resemble the expected behavior for acceptor doping shown theoretically in Fig. 2.22; at lower oxidation temperatures, the vacancy term should have little impact at most large p values, and consequently $B/A = (B/A)^{i}$ for the data in Fig. 2.23. At higher temperatures and for sufficiently large p values, the total B/A should increase significantly above $(B/A)^{i}$ as the vacancy contribution becomes comparable to the other components of B/A grouped in $R_1$. Correspondingly, the highest temperature (1150°C) data in Fig. 2.23 indicate that the effective B/A at $p = 1 \times 10^{20}$ cm$^{-3}$ has increased to twice the

$(B/A)^i$ value. The B/A values for the heavily boron-doped silicon display the resulting curvature with 1/T (rather than a single effective activation energy) that was predicted theoretically in Fig. 2.22.

As was discussed in Case B, the orientation problem can be resolved and possible orientation effects may be tested by direct comparison of the normalized (B/A)' values rather than the total B/A values. Figure 2.24 presents the (100) experimental (B/A)' data obtained in [2.66] for the one p value and the (B/A)' curve calculated for that doping level using $K'C_{V_T}^i$ of Eq. (2.12) derived from the (111) n results in Case A.

It is evident that direct extension from the donor- to the acceptor-model may not be completely satisfactory. The calculated curve for this p value adequately represents the (100) data for most lower temperatures. This may indicate, as in Case B, that orientation effects in the normalized (B/A)' may be minor, but the data are not conclusive because the vacancy contribution determining (B/A)' in Eq. (2.11) is expected to be insignificant at these lower temperatures. At this p value, however, the calculated (B/A)' has not yet begun to deviate significantly from unity at higher temperatures, whereas the data apparently have begun to deviate at the highest temperature point.

If this incongruence does not indicate some inherent failure in the physical modeling, it may be indicative of vacancy charge-state effects thus far neglected [2.8]; that is, the $V^+$ charge state dominant for heavy acceptor doping may be more effective in the interface oxidation reaction than are the $V^-$ and $V^=$ states of donor doping. At higher temperatures, therefore, total B/A begins to vary significantly from

$(B/A)^i$ at a lower p value and corresponding $V^+$ concentration than is considered necessary based on the direct extension of the donor and $V^-$, $V^=$ modeling. Although this evidence is tenuous (only one data point), the possibility of vacancy charge-state effects warrants further investigation.

(e) Case E: Phosphorus and Boron (Donor/Acceptor Compensation) Doping

In the modeling of heavy doping effects on oxidation, effects such as impurity-created strain have been implicitly assumed to be of less importance than electrical effects in enhancing $k_s$ and B/A [2.7].

Heavily compensated phosphorus and boron doped silicon should provide an intriguing test of this assumption. Both phosphorus and boron have covalent radii smaller than that of silicon and cause the substrate lattice to contract on doping [2.75,2.76]. Silicon doped with both impurities to specific levels should be under greater compressive strain than a substrate doped with only one dopant to the same level of that impurity. If strain-related effects are of more than secondary importance, material doped heavily with both boron and phosphorus should exhibit an oxidation interface-reaction rate enhanced as much as, if not more than, silicon with only one dopant.

Boron also has a greater size mismatch relative to silicon than does phosphorus:

$$d_B = 0.88 \text{ Å}$$

$$d_p = 1.10 \text{ Å}$$

$$d_{Si} = 1.17 \text{ Å}$$

A specific boron level should produce more lattice contraction than would a comparable phosphorus level [2.76]. The data in Cases B and D do not indicate a greater boron effect in the interface oxidation-reaction rate; actually, boron has less effect than phosphorus, which is contrary to these strain-related observations.

Compensation of phosphorus donors with boron acceptors should reduce the electrical doping level and force the Fermi level back toward its intrinsic value. If electrical effects dominate, such heavily compensated material should yield a lower oxidation rate than would silicon heavily doped with the phosphorus alone.

As a test of this model, a set of increasingly heavily phosphorus-doped (100) samples at electrically active levels comparable to the (111) substrates in Case A [2.4,2.65] and a lightly doped wafer were subjected to a 60 min boron "predeposition" diffusion from a diborane source at 1100°C. The boron glass was then stripped. Auger measurement indicated a boron surface concentration of $\sim 10^{21}$ $cm^{-3}$ which implies an electrically active level exceeding $10^{20}$ $cm^{-3}$. The net doping level of the samples will vary, as the "background" phosphorus level increases, from heavily p-type to heavily compensated (lower net electrical doping) and then toward heavily n-type. On the other hand, the lattice strain is increased monotonically by the increasing phosphorus level. Thus, the electrical and strain forces are diverging in the samples, and a dominance may be inferred from trends in the resulting oxidation rate.

A control set of phosphorus-doped samples and a lightly doped wafer were given a 2 min $O_2$ followed by 58 min $N_2$ heat treatment at 1100°C but no boron diffusion. Residual oxide was removed by chemical etching. The

two sets of samples were then oxidized together in a dry $O_2$ ambient at 900°C for 4 hours; the increasingly heavy phosphorus-doped samples were placed further downstream on the boat. The resulting oxide thicknesses were measured ellipsometrically, and the data are plotted vs electrically active phosphorus concentrations (prior to boron diffusion or heat treatment and oxidation) in Fig. 2.25.

The heat-treated uncompensated data set shows the expected steady increase of the oxidation rate with heavier phosphorus doping due to increasing strain and/or electrical dopant effects.

For the compensated data set, as the degree of compensation increases, the growth rate first holds steadily at the value characteristic of the high p level. This rate is somewhat faster than that of the lightly doped control wafer to which no boron or phosphorus was added. As the compensating donor level approaches, becomes comparable to, and eventually exceeds that p value, the oxide growth rate correspondingly remains constant, perhaps decreases slightly, and then increases. (Because exact compensation is exceedingly unlikely due to experimental uncertainties, the observed growth rates for the nominally compensated samples should not be expected to decrease completely to that of the lightly doped control wafer; instead, their behavior as observed in Fig. 2.25 may indicate that the boron diffusion yielded an electrically active boron surface level of $\sim 2 \times 10^{20}$ $cm^{-3}$.)

It can be observed that the growth rate in the compensated samples does not rise steadily as demanded by a dominant strain-related mechanism, and that it does not follow the monotonic increase seen in the uncompensated data set. The data show that the electrical influence of heavy

doping must dominate over the strain-related effects to produce the observed enhanced interface oxidation rates.

The parabolic rate constant B, related to diffusion of the oxidizing species through the growing oxide [2.64], also contributes to oxidation but has not been included in the above interpretation. Boron accumulating in the growing oxide from a heavily boron-doped substrate has been shown to produce significant enhancement of B only at higher temperatures (exceeding 1040°C), with relatively little effect on B at 900°C [2.66]. The boron diffusion of the compensated samples should, if anything, only enhance B slightly relative to the uncompensated phosphorus-doped samples. (Enhanced oxidation of the substrate doped only with boron, therefore, should result from a slight increase of both B/A and B above the values for the lightly doped control wafer.) Yet, overall oxidation of the most heavily phosphorus-doped compensated samples is retarded relative to the uncompensated controls, indicating that the effective B/A is indeed reduced in the compensated samples, in agreement with the above interpretation.

## D.3 Discussion

The vacancy-contribution model developed in [2.4,2.8] and compared to experimental data above appears to be adequate for determining an effective B/A resulting from high doping levels. Possible deviation from experimental results in heavy acceptor doping at high temperatures is not a critical limitation because the influence of B/A on the oxidation rate at these B-dominated higher temperatures is relatively minor [2.64]. This model may have immediate practical application under most device processing conditions in integrated circuit technology [2.77].

Modeling, however, has raised numerous questions, particularly concerning the role of silicon point defects in oxidation and other high temperature processes that should be pursued. Point defect streams may prove to be a unifying factor.

(a) Nature of the Vacancy Contribution

The nature of the interface oxidation reaction and the proposed vacancy contribution to that reaction rate are of great interest. The observation that the temperature behavior of $(B/A)^i$ resembles an activated process with an activation energy of 2.0 eV has led to the speculation that the limiting energy represents that required to break a Si-Si bond [2.64]. Other factors, however, could become significant. As discussed in [2.8], movement of the $SiO_2$/Si interface into the silicon as the interface oxidation reaction proceeds may be considered as a conversion of the silicon crystal lattice into an $SiO_2$ network possibly with local, as-grown order "matched" to the Si substrate. Based on volumetric, strain, and lattice-size considerations, it is unlikely that an oxygen atom is inserted between every two Si atoms bonded together at the interface; instead, an overabundance of Si atoms could be present, and space or sites must be created for oxygen atoms to occupy. Silicon-lattice point defects should be significant in providing these reaction sites. Interstitials can be created or vacancies consumed in the interface reaction, and the oxidizing interface becomes an interstitial source or vacancy sink.

It has been suggested [2.78-2.81] that such an excess of interstitials and/or depletion of vacancies could be created by the oxidation

process. Actually, interstitial flows may have some bearing on a possible orientation dependence of the vacancy contribution and the normalized $(B/A)'$ at high doping levels. Interstitials may combine with vacancies to annihilate both point defects and thus serve as another potential sink for the vacancies. It has been proposed [2.81,2.82] that the interstitial concentration present during oxidation is larger for (100) than for (111) silicon, which explains the greater enhancement of boron diffusion (postulated to be interstitial-related) under (100) Si oxidizing conditions. A greater interstitial concentration presumably could reduce the magnitude of the vacancy contribution to the oxidation reaction simply by annihilating more vacancies; the normalized $(B/A)'$ would then be less for (100) than for (111) silicon. The comparisons of (100) data to calculated $(B/A)'$ in Figs. 2.20 and 2.24 do not discount this likelihood.

With such a multitude of possible mechanisms (point-defect related or otherwise) involved in the oxidation reaction, the assumption of multiple contributions to the effective interface-reaction rate constant $k_s$ and linear rate constant $B/A$ should not be unreasonable. The modeling described here has employed this assumption, plus the proposal that the contribution most significantly affected by heavy substrate doping is vacancy-related.

All other possible contributions (whether interstitial-related or the breaking of a Si-Si bond) are grouped into $R_1$ and assumed to be unaffected or influenced to a much lesser extent. This last assumption is largely a matter of convenience because little quantitative information is available in the literature concerning interstitials. The

empirical results in Case A do not indicate a significant change of B/A activation energy with heavy donor doping that should be observable if a bond-breaking energy mechanism is influenced substantially by dopant-induced strain. The order of magnitude increase exhibited by B/A with heavy phosphorus doping would require a reduction of the activation energy by 0.25 eV at oxidation temperatures; however, as indicated in Fig. 2.13 of [2.4], relatively little effect is observed on the associated activation energy and surely not a 0.25 eV decrease (in fact, a slight increase may possibly be present). Also, a Morse potential calculation [2.83] indicates that the change in Si-Si bond energy to be expected from the lattice constant change [2.76] produced by even a substitutional phosphorus level of $\sim 10^{21}$ $cm^{-3}$ is less than 1 percent. Such a direct strain-related bond energy change seems unlikely to be a major contributing factor. Other possible pathways remain equally doubtful or unmanageable.

A vacancy contribution, on the other hand, can be determined numerically with the additional simplifying assumption of thermal or quasi-equilibrium with respect to vacancy statistics and concentrations. Vacancy lifetime has been estimated [2.84] from vacancy diffusion-length data at 750°C [2.85] to be on the order of $10^{-4}$ sec. Because this appears small relative to other time constants associated with the oxidation process, the assumption of equilibrium vacancy contributions should not be an unreasonable first-order approximation.

Within this framework, the vacancy contribution is characterized by the fitted parameter $K'C_{V_T}^{i}$ in Eq. (2.12) based on the $n$ values for the heavily phosphorus-doped samples reported in the study described in

Case A [2.8,2.65]. However, degeneracy effects, nonionized substitutional phosphorus, and other possible phosphorus species were neglected, and room-temperature measurements were applied to estimate these carrier concentrations. Such values may underestimate the concentrations applicable at oxidation temperatures, particularly in the most heavily doped samples. As a result, in Fig. 2.18, the experimental $(B/A)'$ value for the most heavily doped sample (designated F) appears greater than the calculated value at the $n$ value assumed for F in [2.8,2.65]. The actual $n$ value may be higher, and the measured $(B/A)'$ point would then be shifted to a higher $n$ value for which the calculated $(B/A)'$ would be in closer agreement.

Because the true concentrations may approach the total chemical phosphorus $C_{BC}$ values also listed in [2.7], these values may be used to approximate the actual $n$ values. Figure 2.26 plots the results of applying the vacancy-contribution model to the phosphorus data obtained in Case A at 800°C, assuming that $n \approx C_{BC}$. Equation (2.11) was fitted to the data by following the procedure described in [2.8]. The offset proportionality parameter $K'C_{V_T}^{i}e^{2.0/kT}$ was reduced to $5.0 \times 10^{-3}$ from the earlier value of $1.8 \times 10^{-2}$ at 800°C derived in Eq. (2.12). The agreement may be better than in Fig. 2.18a where the apparent deviation of the F point may partly result from an n-value underestimation which is most significant for the F point. (Additional factors could contribute to this deviation. Other mechanisms grouped in $R_I$ and previously assumed unaffected by heavy doping may be influenced at the highest phosphorus levels. For example, dopant-related strain could become significant, and error in the B/A rate-constant extraction is inherently greatest for the F points [2.64,2.7].)

Similar considerations would apply to modeling the other experimental cases above. Ultimately, they all employ room-temperature measurements to estimate and, perhaps, undervalue carrier concentrations applicable at the oxidation temperatures. The actual carrier concentration at a specific doping level should vary only slightly throughout the limited oxidation temperature range (~750° to 1150°C). If corrections are made in the n values matched to the phosphorus data points in Case A, each sample type would still have a "constant" n value at all oxidation temperatures studied. As a result, refitting Eq. (2.11) to the data to determine a new offset parameter analogous to $K'C_{V_T}^i$ in Eq. (2.12) would yield a different pre-exponential factor; however, the ~3.1 eV activation energy in Eq. (2.12) would remain unchanged.

Although the precise value of the pre-exponential in Eq. (2.12) may be questioned, the ~3.1 eV activation energy should be meaningful. The vacancy contribution has been determined [2.8] as

$$KC_{V_T} = \left(K'C_{V_T}^i\right) C_1 C'_{V_T} \propto C'_{V_T} \exp\left(\frac{-3.1}{kT}\right) \qquad (2.14)$$

where $C'_{V_T}$ indicates the dependence of the total vacancy contribution on the doping level. With heavy doping, Fermi-level shifts in the band gap vary $C'_{V_T}$ and, through Eq. (2.14), also the apparent activation energy of the vacancy component of the interface reaction (Figs. 8 and 9 of [2.8]). It appears reasonable, therefore, that the vacancy contribution at low doping levels $(K'C_{V_T}^i)C_1$ or $KC_{V_T}^i$ must reflect the nature or mechanisms of the vacancy contribution unconcealed by Fermi-level shifts.

The ~3.1 eV activation energy of the fitted parameter $K'C^{i}_{V_T}$ may be revealing. As discussed in [2.8], $C^{i}_{V_T}$ is calculated from and follows very closely the Arrhenius temperature dependence of the concentration of neutral vacancies $C_{V^x}$. [This quality has not been required numerically in the analysis; instead, the actual calculated variable in Eq. (2.11) for B/A is $C'_{V_T}$ from which $C_{V^x}$ is canceled by normalization.] Physical interpretation of the vacancy contribution will be dependent, therefore, on the precise determination of the behavior of $C_{V^x}$ with temperature or the neutral vacancy formation energy $E_x$.

A survey of the literature indicated that there is some controversy over vacancy properties. Van Vechten and Thurmond [2.86] concluded that $E_x = 2.4 \pm 0.2$ eV. The observed behavior of $K'C^{i}_{V_T}$ would then suggest an additional $K'$ temperature dependence. Because Van Vechten [2.87] also found that vacancy migration requires an additional migration energy $E_m = 1.2 \pm 0.3$ eV regardless of the charge state, $K'$ may be related to vacancy diffusivity. [The concept that $E_m$ is not dependent on vacancy charge states further supports the contention that $K'C^{i}_{V_T}$ should indicate the limiting process in the vacancy contribution undisguised by changes in the Fermi level. These effects then influence only the vacancy concentrations through $C'_{V_T}$.] The vacancy contribution reflected in $K'C^{i}_{V_T}$, therefore, may consist of a vacancy flux to the interface or a silicon self-diffusion via vacancies into the bulk substrate. Van Vechten's total vacancy-diffusion energy of ~3.6 eV, Swalin's ~3.38 eV [2.83], and Bennemann's ~3.22 eV [2.88] are in good agreement. Vacancies may diffuse from such substrate sources as bulk defects or vacancy clusters frozen into the bulk when the silicon is pulled from the melt [2.89, 2.90].

Fairfield and Masters [2.91] propose a substantially greater formation energy of $E_x \approx 3.4$ eV and a correspondingly larger self-diffusion energy of $5.13 \pm 0.1$ eV; $K'$ would then appear to be temperature insensitive, and the vacancy contribution may possibly be linked to vacancy generation at or near the interface. Alternatively, the contribution could be directly proportional to the total vacancy concentration at all temperatures. As a result, vacancies may simply produce a greater effective "surface area" or larger number of available Si-Si bonds at the "rougher" interface [2.7,2.66]. The 3.1 eV activation energy of $K'C^i_{V_T}$ in Eq. (2.12) again could be reasonably related to the relevant independently reported physical parameters. The model described here is therefore consistent with either of the above models for vacancy generation and diffusion.

(b) Effects of the Vacancy Charge State

The dominant vacancy charge state will change as $E_F$ shifts due to heavy doping [2.92]. Because charge states have been neglected [2.8], however, the Fermi-level position influences "directly" the magnitude and apparent activation energy of the vacancy contribution via Eq. (2.14). As $E_F$ varies with heavy doping, so do the vacancy contribution and its effective activation energy. The agreement with experimental results above appears to justify regarding the charge-state effects as secondary in importance.

Charge-state effects, however, may provide additional insight into the physical mechanisms involved. For example, the results obtained from boron doping in Case D indicated that the effects of the $V^+$ charge state

and the $V^-$, $V^=$ states on the oxidation reaction may differ. If the charge state became a significant parameter, the linear rate constant becomes, in analogy with [2.4,2.8],

$$\frac{B}{A} = R_1 + K_{V^x} C_{V^x} + K_{V^+} C_{V^+} + K_{V^-} C_{V^-} + K_{V^=} C_{V^=} \qquad (2.15)$$

The neutral vacancy contribution could then be grouped in $R_1$ with all other possible contributions dependent on temperature, unaffected by heavy doping, and independent of heavy-doping electrical effects. Conceivably, the neutral vacancy contribution could even dominate $R_1$ if $C_{V^x}$ can be shown to have an activation energy of ~2.0 eV. Van Vechten and Thurmond's [2.86] value of 2.4 $\pm$ 0.2 eV applies to bulk vacancies and may not allow for a reduction in the energy required because of strain at the $SiO_2$/Si interface resulting from mismatch of the Si and $SiO_2$ lattice parameters. Van Vechten [2.93] has also estimated ~1.8 eV as the corresponding formation energy for a surface vacancy.

The $K'C^i_{V_T}$ in Eq. (2.11) and its associated 3.1 eV activation energy may represent a weighted average of the $V^-$ and $V^=$ contributions dominant in heavy phosphorus (donor) doping from which it was deduced. It is not unlikely that the $V^+$ contribution could reveal a different proportionality and effective activation energy as a result of the modified charge-balance requirements in the detailed oxidation reaction at the interface. In view of the possible negatively charged nature of the diffusing oxygen species in the oxide [2.64], Collins and Nakayama [2.94] suggested that holes or positive charge are consumed or electrons are liberated in the interface reaction to maintain charge balance. A positively charged vacancy may then be more "effective" in

reacting with a negative oxidant, as was implied by the boron data obtained in Case D. If an electric field should be present at or near the interface, the diffusion of positively vs negatively charged vacancies may also be affected.

Such charge-state effects would be expected to differentiate between positively and negatively charged vacancies to a greater degree than between $V^-$ and $V^=$. The $V^-$ and $V^=$ differences, therefore, may be small enough so that, for example, the phosphorus data in Case A would not be a sufficiently sensitive test to reveal them.

In the absence of such factors as electric field effects, the highly delocalized nature of the charge associated with ionized vacancies [2.92] may reduce the significance of such charge-state effects. If the principal requirement of the interface reaction is to provide reaction sites for the oxidant, the vacancy charge state may actually be a second-order consideration.

(c) Redistribution and the Vacancy Contribution

Dopant redistribution has been neglected in the analysis in [2.8] and above. Serious questions may be raised concerning its manageability and/or usefulness in this investigation.

Observed phosphorus redistribution profiles [2.95] for the heavily doped samples used in Case A differ in several respects from predictions of simple theory [2.96]. For example, the ratio of pileup to bulk phosphorus concentration ($C_S/C_B$) for dry $O_2$ oxidation at 900°C is expected to be ~1.5. The observed ratios may be over an order of magnitude greater for the oxidized heavily doped samples. In theory, the segregation

coefficient for phosphorus is assumed to be ~10. The actual ratio of the observed pileup to concentration in the oxide ($C_S/C_{OX}$), or an effective segregation coefficient, again is considerably greater--possibly by more than an order of magnitude. From theory, the extent of the redistribution disturbance into the silicon from the interface, or the width of the pileup peak, should be approximately one phosphorus-diffusion length $2\sqrt{Dt}$; for 200 min at 900°C, this length for low phosphorus-doping levels should be ~500 Å. The peak widths observed in the heavily doped samples are an order of magnitude narrower (~30 to 40 Å). As a result, the existing theory of impurity redistribution appears inadequate, at least for the oxidation of heavily phosphorus-doped silicon.

Even if agreement between redistribution theory and experiment for oxidation of heavily doped silicon were reasonable, note that simple theory assumes parabolic oxide growth to find dopant concentrations at the $Si/SiO_2$ interface should reach a steady-state constant value [2.96]. Oxidation proceeds through a linear growth period before achieving parabolic growth. Thus, redistribution should be subject to a finite transient period before reaching the steady-state. Because B/A is most relevant during this linear growth transient, the effective B/A may correspond to a time-averaged dopant concentration in the silicon which differs from the steady-state value.

Also, simple theory predicts that phosphorus pile-up should be greater at lower temperatures. Yet, for a given initial phosphorus level in Case A, the observed enhancement of B/A is no greater at 800°C than at 1100°C and does not appear to track the theoretical temperature behavior of the phosphorus pile-up.

Even if the numerical problem becomes tractable, the usefulness of applying redistribution to the analysis would depend on the physical mechanism by which heavy doping alters the oxidation rate. The heavy-doping phosphorus pileup peaks of Case A may well include precipitates [2.97] or clusters [2.67] that are electrically inactive; such inactive phosphorus would not be crucial to the proposed electrical nature and vacancy mechanism of the dopant effects on the interface oxidation reaction.

(d) Point Defects, Oxidation, Oxide Charge, and Related Process Phenomena

Thermal oxidation and oxide properties are closely connected to the interface reaction. Vacancies and interstitials participating in this reaction, therefore, may be connected to the resulting oxide and interface charges. As noted in [2.8] and earlier in this report, perhaps a small fraction of interstitials produced in the oxidation process may move into the oxide. Such a flow very probably would then be directly related to the observed oxide charge $Q_{ss}$ which has been described as incompletely oxidized silicon atoms [2.98]. Calculations based on perturbation theory have demonstrated that localized levels in the silicon band gap may result from the dangling bonds of Si atoms associated with oxygen or silicon vacancies at the $Si/SiO_2$ interface, which suggests a physical origin for interface states [2.99]. Vacancies may also be related to surface recombination at the interface via these states which may serve as recombination-generation centers and, therefore, should be directly proportional to the surface recombination velocity. Recombination velocity values have been shown to increase dramatically with heavy-doping levels [2.100] as should vacancy concentrations.

Oxidation and point defects may be related to many other processing phenomena for which point-defect streams may prove to be a unifying factor. The postulated excess interstitial flows produced during thermal oxidation may be the mechanism of oxidation-enhanced dopant diffusion [2.78,2.79,2.81]. Vacancy-related models abound for impurity diffusion [2.77,2.82,2.84,2.101,2.102]. These models that employ vacancy streams are particularly necessary at high doping levels to resolve "nonideal" diffusion profiles and to explain the interaction between sequential dopant diffusions [2.103] such as the often-observed emitter push [2.104] and emitter-pull [2.105]. As a result, these phenomena may impact and be impacted by the oxidation process in general and by the oxidation of heavily doped silicon in particular. Clearly linked to oxidation are oxidation-induced stacking faults and other defects that have been attributed to excess interstitials [2.81,2.82,2.106] and to vacancy flows [2.80] and clusters [2.107].

This investigation has postulated that large vacancy concentrations resulting from the heavy doping of silicon may increase oxidation rates; such enhanced oxidation appears to extend directly to heavily doped polycrystalline silicon [2.108,2.109]. Greater concentrations of vacancies in heavily doped silicon may also produce faster silicon self-diffusion [2.91] and solid-phase epitaxial regrowth of amorphous silicon layers on single-crystal silicon [2.110,2.111]. The more rapid self-diffusion could actually be linked directly to the enhanced oxidation. The mechanisms may all be identical--the vacancies provide reaction sites into which the appropriate reactant atom may move.

With oxidation and Si point-defect streams potentially so critically involved in these and other phenomena, additional investigation to understand better the oxidation process and proposed vacancy (and interstitial) contributions to the interface oxidation reaction may prove rewarding.

## REFERENCES


[2.1] B.E. Deal and A.S. Grove, "General Relationship for the Thermal Oxidation of Silicon," J. Appl. Phys., 36, 3770-3778, Dec 1965.

[2.2] J.D. Meindl et al., "Final Report on Computer-Aided Semiconductor Process Modeling," Stanford Electronics Laboratories, TR-4969-73-F, Oct 1976.

[2.3] D.W. Hess and B.E. Deal, "Kinetics of the Thermal Oxidation of Silicon in $O_2$HCl Mixtures," J. Electrochem. Soc., 124, 735-739, May 1977.

[2.4] J.D. Meindl et al.,"Interim Report on Computer-Aided Semiconductor Process Modeling," Stanford Electronics Laboratories, TR-4969-3, Feb 1978.

[2.5] B.E. Deal, D.W. Hess, J.D. Plummer, and C.P. Ho, "Kinetics of the Thermal Oxidation of Silicon in $O_2/H_2O$ and $O_2/Cl_2$ Mixtures," J. Electrochem. Soc., 125, 339-346, Feb 1978.

[2.6] B.E. Deal, "Thermal Oxidation Kinetics of Silicon in Pyrogenic $H_2O$ and 5% $HCl/H_2O$ Mixtures," J. Electrochem. Soc., 125, 576-579, Apr 1978.

[2.7] C.P. Ho, J.D. Plummer, B.E. Deal, and J.D. Meindl, "Thermal Oxidation of Heavily Phosphorus Doped Silicon," J. Electrochem. Soc., 125, 665-671, Apr 1978.

[2.8] C.P. Ho and J.D. Plummer, "$Si/SiO_2$ Interface Oxidation Kinetics: A Physical Model for the Influence of High Substrate Doping Levels. I. Theory," J. Electrochem. Soc., to be published Sep 1979.

[2.9] C.P. Ho and J.D. Plummer, "Si/SiO Interface Oxidation Kinetics: A Physical Model for the Influence of High Substrate Doping Levels. II. Comparison with Experiment and Discussion," J. Electrochem. Soc., to be published Sep 1979.

[2.10] R.K. Razouk and B.E. Deal, "Dependence of Interface State Density on Silicon Thermal Oxidation Process Variables," J. Electrochem. Soc., to be published.

[2.11] R.H. Doremus, "Oxidation of Silicon by Water and Oxygen and Diffusion in Fused Silica," J. Phys. Chem., 80, 1773, 1976.

[2.12] P.J. Jorgensen, "Effect of Electric Field on Silicon Oxidation," J. Chem. Phys., 37, 874, 1962.

[2.13] D.O. Rayleigh, "Transport Processes in the Thermal Oxidation of Silicon," J. Electrochem. Soc., 113, 782, 1966.

[2.14] F.C. Collins and T. Nakayama, "Transport Processes in the Thermal Growth of Metal and Semiconductor Oxide Films," J. Electrochem. Soc., 114, 162, 1967.

[2.15] T.G. Mills and F.A. Kroger, "Electrical Conduction at Elevated Temperatures in Thermally Grown Silicon Dioxide Films," J. Electrochem. Soc., 120, 1582, 1973.

[2.16] P.S. Dobson, "The Effect of Oxidation on Anomalous Diffusion in Silicon," Philosophical Mag., 24, 567-576, 1971.

[2.17] P.S. Dobson, "The Mechanism of Impurity Diffusion in Silicon," Philosophical Mag., 26, 1301-1306, 1972.

[2.18] S.M. Hu, "Formation of Stacking Faults and Enhanced Diffusion in the Oxidation of Silicon," J. Appl. Phys., 45, 1567, 1974.

[2.19] B.E. Deal, "The Current Understanding of Charges in the Thermally Oxidized Silicon Structure," J. Electrochem. Soc., 121, 198C, 1974.

[2.20] W. Shockley and J.L. Moll, "Solubility of Flaws in Heavily-Doped Semiconductors," Phys. Rev., 119, 1480, 1960.

[2.21] J.A. Van Vechten and C.D. Thurmond, "Comparison of Theory with Quenching Experiments for the Entropy and Enthalpy of Vacancy Formation in Si and Ge," Phys. Rev. B, 14, 3551, 1976.

[2.22] T.E. Seidel and A.V. MacRae, "Some Properties of Ion Implanted Boron in Silicon," Trans. Metall. Soc. AIME, 245, 491, 1969.

[2.23] S.P. Murarka, "Role of Point Defects in the Growth of Oxidation-Induced Stacking Faults in Silicon," Phys. Rev. B, 16, 2849, 1977.

[2.24] S.P. Murarka, "Oxygen Partial Pressure Dependence of the Fixed Surface-State Charge $Q_{ss}$ Due to Thermal Oxidation of n-(100) Silicon," App. Phys. Letters, 34, 587-588, May 1979.

[2.25] B.E. Deal, private communication.

[2.26] A.S. Grove, Physics and Technology of Semiconductor Devices, John Wiley and Sons, New York, 1967.

[2.27] R. Ghez and Y.J. van der Meulen, "Kinetics of Thermal Growth of Ultra-Thin Layers of $SiO_2$ on Silicon. II. Theory," J. Electrochem. Soc., 119, 1100, 1972.

[2.28] Y.J. van der Meulen, "Kinetics of Thermal Growth of Ultra-Thin Layers of $SiO_2$ on Silicon. I. Experiment," J. Electrochem. Soc., 119, 530, 1972.

[2.29] W.A. Tiller, "On the Kinetics of the Thermal Oxidation of Silicon Part I: A Theoretical Perspective," submitted to J. Electrochem. Soc.

[2.30] W.A. Tiller, "On the Kinetics of the Thermal Oxidation of Silicon Part II: Some Theoretical Evaluation," submitted to J. Electrochem. Soc.

[2.31] B.E. Deal, M. Sklar, A.S. Grove, and E.H. Snow, "Characteristics of the Surface-State Charge ($Q_{ss}$) of the Thermally Oxidized Silicon," J. Electrochem. Soc., 114, 266, 1967.

[2.32] D.R. Lamb and F.R. Badcock, "The Effect of Ambient, Temperature and Cooling Rate, on the Surface Charge at the Silicon/Silicon Dioxide Interface," Int. J. Electronics, 24, 11, 1968.

[2.33] F. Montillo and P. Balk, "High Temperature Annealing of Oxidized Silicon Surfaces," J. Electrochem. Soc., 118, 1463, 1971.

[2.34] J.L. Pautrat and J.C. Pfister, "Electronic Properties of the Si-$SiO_2$ Interface as a Function of Oxide Growth Conditions," Phys. Status Solidi A, 11, 669, 1972.

[2.35] D.J. Breed and R.P. Kramer, "Stable and Unstable Surface State Charge in Thermally Oxidized Silicon," Solid-State Electron., 19, 897, 1976.

[2.36] F.M. Fowkes and D.W. Hess, "Control of Fixed Charge at Si-$SiO_2$ Interface by Oxidation-Reduction Treatments," Appl. Phys. Lett., 22, 377, 1973.

[2.37] R.F. Pierret and B.B. Roesner, "Photo-Thermal Probing of Si-$SiO_2$ Surface Centers II," Solid-State Electron., 19, 593, 1974.

[2.38] P. Balk, "Effects of Hydrogen Annealing on Silicon Surfaces," The Electrochemical Society Meeting at San Francisco, Extended Abstract No. 109, May 1965.

[2.39] M. Kuhn, "A Quasi-Static Technique for MOS C-V and Surface State Measurements," Solid-State Electron., 13, 873, 1970.

[2.40] R. Castagne and A. Vapaille, "Description of the Si-$SiO_2$ Interface Properties by Means of Very Low Frequency MOS Capacitance Measurements," Surface Science, 28, 157, 1971.

[2.41] C.N. Berglund, "Surface States in Steam-Grown Silicon-Silicon Dioxide Interfaces," IEEE Trans. Electron. Devices, ED-13, 701, 1966.

[2.42] P.L. Castro and B.E. Deal, "Low-Temperature Reduction of Fast Surface States Associated with Thermally Oxidized Silicon," J. Electrochem. Soc., 118, 280, 1971.

[2.43] Y.T. Yeow, D.R. Lamb, and S.D. Brotherton, "An Investigation of the Influence of Low-Temperature Annealing Treatments on the Interface State Density at the Si-$SiO_2$ Interface," J. Phys. D.: Appl. Phys., 8, 1495, 1975.

[2.44] B.E. Deal, "Charge Effects and Other Properties of the Si-$SiO_2$ Interface: The Current Understanding," Proceedings of the Third International Symposium on Silicon Material Science and Technology, Electrochemical Society, Philadelphia, Pa., May 8-13, 1977.

[2.45] M.M. Atalla, E. Tannenbaum, and E.J. Scheibner, "Stabilization of Silicon Surfaces by Thermally Grown Oxides," Bell Syst. Tech. J., 38, 749, 1959.

[2.46] H. Edagawa, Y. Morita, S. Mackawa, and Y. Inuishi, "Growth and Structure of Si Oxide Films on Si Surface," Jap. J. Appl. Phys., 2, 765, 1963.

[2.47] A.N. Knopp and R. Stickler, "The Structure and Perfection of Thermally Grown Oxide Films on Silicon," Electrochem. Tech., 5, 37, 1967.

[2.48] T. Sugano, K. Hoh, K. Kudo, and N. Nishinuma, "Ordered Structure and Ion Migration in Silicon Dioxide Films," Jap. J. Appl. Phys., 7, 715, 1968.

[2.49] N. Nagasima, "Structure Analysis of Thermal Oxide Films of Silicon by Electron Diffraction and Infrared Absorption," Jap. J. Appl. Phys., 9, 879, 1970.

[2.50] E.I. Alessandrini and D.R. Campbell, "Catalyzed Crystallization in $SiO_2$ Thin Films," J. Electrochem. Soc., 121, 1115, 1974.

[2.51] T.J. Magee, J. Peng, and R. Ormond, "Identification of Crystalline Phases in $SiO_2$ Films," unpublished.

[2.52] M. Thackray, "Autoradiography of Radioactive Photographic Images," Australian Atomic Energy Comm. Tech. Rep. No. AAEC/E317, Sep 1974.

[2.53] M. Thackray, D. Roman, E.L.R. Hetherington, and H.R. Brian, "Investigation of Solid-Solid Reactions of Au Films on Silicon," Intl. J. Appl. Rad. and Isotopes, 23, 79, 1972.

[2.54] T.J. Magee and J. Peng, "Intensification of Photographs by Means of Autoradiography," Phys. Stat. Sol., A 49, 313, 1978.

[2.55] T.J. Magee and J. Peng, ARO Contract Rep. DAAG29-77-C-0024, 1977, unpublished.

[2.56] A. Guinier, X-ray Diffraction in Crystals, Imperfect Crystals, and Amorphous Bodies, W.H. Freeman & Co., San Francisco, Ch. 5, 1963.

[2.57] Y.J. van der Meulen, "Kinetics of Thermal Growth of Ultra-Thin Layers of $SiO_2$ on Silicon. I. Experiment," J. Electrochem. Soc., 119, 530, 1972.

[2.58] F.L. McCrackin, NBS Technical Note 479, "A FORTRAN Program for the Analysis of Ellipsometer Measurements," U.S. Government Printing Office, Washington, D.C.

[2.59] E. Taft and L. Cordes, "Optical Evidence for a Silicon-Silicon Oxide Interlayer," J. Electrochem. Soc., 126, No. 1, 131, 1979.

[2.60] C.C. Chang and D.M. Boulin, "Oxide Thickness Measurements up to 120 A on Silicon and Aluminum Using the Chemically Shifted Auger Spectra," Surface Science, 69, 385-402, 1977.

[2.61] A.M. Goodman, "Optical Properties of Amorphous and Recrystallized $SiO_x$ Layers," Inst. Phys. Conf. Ser., No. 43, 805, 1979.

[2.62] E. Harari, "Dielectric Breakdown in Electrically Stressed Thin Films of Thermal $SiO_2$," J. Appl. Phys., 49, 2478, Apr 1978.

[2.63] V.A. Arslambekov and A. Safarov, "Some Details of the Mechanism and Kinetics of Silicon Oxidation," Mikroelecktronika, Vol. 6, No. 1, 75-81, Jan-Feb 1977.

[2.64] B.E. Deal and M. Sklar, "Thermal Oxidation of Heavily Doped Silicon," J. Electrochem. Soc., 112, 430-435, Apr 1965.

[2.65] C.P. Ho, "Silicon Oxide Studies: Physical Modeling and Device Applications of Thermal Oxidation of Heavily Doped Silicon," Ph.D. Dissertation, SEL-78-033, TR No. 4970-2, Stanford Electronics Laboratories, Stanford University, Stanford, California, Nov 1978.

[2.66] E.A. Irene and D.W. Dong, "Silicon Oxidation Studies: The Oxidation of Heavily B- and P-Doped Single Crystal Silicon," J. Electrochem. Soc., 125, 1146-1151, Jul 1978.

[2.67] R.B. Fair and J.C.C. Tsai, "A Quantitative Model for the Diffusion of Phosphorus in Silicon and the Emitter Dip Effect," J. Electrochem. Soc., 124, 1107-1118, Jul 1977.

[2.68] D.W. Hess and B.E. Deal, "Kinetics of the Thermal Oxidation of Silicon in $O_2$/HCl Mistures," J. Electrochem. Soc., 124, 735-739, May 1977.

[2.69] B.E. Deal, "Thermal Oxidation Kinetics of Silicon in Pyrogenic $H_2O$ and 5% HCl/$H_2O$ Mixtures," J. Electrochem. Soc., 125, 576-579, Apr 1978.

[2.70] J.R. Ligenza, "Effect of Crystal Orientation on Oxidation Rates of Silicon in High Pressure Steam," J. Phys. Chem., 65, 2011-2014, Nov 1961.

[2.71] S. Ohkawa and Y. Nakajima, "Thermal Oxidation of Arsenic-Diffused Silicon," J. Electrochem. Soc., 125, 1997-2002, Dec 1978.

[2.72] R.B. Fair and G.R. Weber, "Relationship Between Resistivity and Total Arsenic Concentration in Heavily Doped n- and p-type Silicon," J. Appl. Phys., 44, 280-282, Jan 1973.

[2.73] R.B. Fair and J.C.C. Tsai, "The Diffusion of Ion-Implanted Arsenic in Silicon," J. Electrochem. Soc., 122, 1689-1696, Dec 1975.

[2.74] R.B. Fair, "Recent Advances in Implantation and Diffusion Modeling for the Design and Process Control of Bipolar ICs," Extended Abstracts of the Spring Meeting of the Electrochem. Soc., Philadelphia, Pa., Abs. 235, 598-600, May 8-13, 1977.

[2.75] T.I. Kamins and E.S. Meieran, "X-Ray Measurements of Stress in Thin Single-Crystal Silicon Films," J. Appl. Phys., 44, 5064-5066, Nov 1973.

[2.76] Y.T. Lee, N. Miyamoto, and J. Nishizawa, "The Lattice Misfit and Its Compensation in the Si-Epitaxial Layer for Doping with Germanium and Carbon," J. Electrochem. Soc., 122, 530-535, Apr 1975.

[2.77] C.P. Ho and J.D. Plummer, "Improved MOS Device Performance through the Enhanced Oxidation of Heavily Doped $N^+$ Silicon," IEEE Trans. Electron Devices and J. Solid State Circuits Joint Special Issue on VLSI, Apr 1979.

[2.78] P.S. Dobson, "The Effect of Oxidation on Anomalous Diffusion in Silicon," Philosophical Mag., 24, 567-576, 1971.

[2.79] P.S. Dobson, "The Mechanism of Impurity Diffusion in Silicon," Philosophical Mag., 26, 1301-1306, 1972.

[2.80] I.R. Sanders and P.S. Dobson, "Oxidation, Defects and Vacancy Diffusion in Silicon," Philosophical Mag., 20, 881-893, 1969.

[2.81] S.M. Hu, "Formation of Stacking Faults and Enhanced Diffusion in the Oxidation of Silicon," J. Appl. Phys., 45, 1567-1573, Apr 1974.

[2.82] R.B. Fair, "Boron Diffusion in Silicon-Concentration and Orientation Dependence, Background Effects, and Profile Estimation," J. Electrochem. Soc., 122, 800-805, Jun 1975.

[2.83] R.A. Swalin, "Theoretical Calculations of the Enthalpies and Entropies of Diffusion and Vacancy Formation in Semiconductors," J. Phys. Chem. Solids, 18, 290-296, 1961.

[2.84] A. Chu, "A Theoretical Approach to the Calculation of Annealed Impurity Profiles of Ion Implanted Boron into Silicon," Ph.D. Dissertation, TR No. 4969-2, Stanford Electronics Laboratories, Stanford University, Stanford, California, 1977.

[2.85] T. Tsuchimoto and T. Tokuyama, Proc. of the First International Conference on Ion Implantation in Semiconductors, Thousand Oaks, California, Gordon and Breach, New York, 1970.

[2.86] J.A. Van Vechten and C.D. Thurmond, "Comparison of Theory with Quenching Experiments for the Entropy and Enthalpy of Vacancy Formatio- in Si and Ge," Phys. Rev. B, 14, 3551-3557, Oct 1976.

[2.87] J.A. Van Vechten, "Enthalpy of Vacancy Migration of Si and Ge," Phys. Rev. B, 10, 1482-1506, Aug 1974.

[2.88] K.H. Bennemann, "New Method for Treating Lattice Point Defects in Covalent Crystals," Phys. Rec., 137, A1497-A1514, Mar 1965.

[2.89] A.J.R. de Kock, "The Elimination of Vacancy-Cluster Formation in Dislocation-Free Silicon Crystals," J. Electrochem. Soc., 118, 1851-1856, Nov 1971.

[2.90] L. Elstner and W. Kamprath, "Quenched-In Levels in p-type Silicon," Phys. Stat. Sol., 22, 541-547, 1967.

[2.91] J.M. Fairfield and B.J. Masters, "Self-Diffusion in Intrinsic and Extrinsic Silicon," J. Appl. Phys., 38, 3148-3154, Jul 1967.

[2.92] J.A. Van Vechten and C.D. Thurmond, "Entropy of Ionization and Temperature Variation of Ionization Levels of Defects in Semiconductors," Phys. Rev. B, 14, 3539-3550, Oct 1976.

[2.93] J.A. Van Vechten, "Effect of Reconstruction during Epitaxial Growth," J. Vac. Sci. Technol., 14, 992-998, Jul/Aug 1977.

[2.94] F.C. Collins and T. Nakayama, "Transport Processes in the Thermal Growth of Metal and Semiconductor Oxide Films," J. Electrochem. Soc., 114, 167-171, Feb 1967.

[2.95] S.A. Schwarz, C.R. Helms, W.E. Spicer, and N.J. Taylor, "High Resolution Auger Sputter Profiling Study of the Effect of Phosphorus Pile-Up on the Si-$SiO_2$ Interface Morphology," J. Vac. Sci. Technol., 15, 227-230, Mar/Apr 1978.

[2.96] A.S. Grove, O. Leistiko, Jr., and C.T. Sah, "Redistribution of Acceptor and Donor Impurities during Thermal Oxidation of Silicon," J. Appl. Phys., 35, 2695-2701, Sep 1964.

[2.97] A. Armigliato, D. Nobili, M. Servidori, and S. Solmi, "SiP Precipitation within the Doped Silicon Lattice, Concomitant with Phosphorus Predeposition," J. Appl. Phys., 47, 5489-5491, Dec 1976.

[2.98] B.E. Deal, "The Current Understanding of Charges in the Thermally Oxidized Silicon Structure," J. Electrochem. Soc., 121, 198C-205C, Jun 1974.

[2.99] T. Lizuka and T. Sugano, "Calculations of Energy Levels of Oxygen Silicon Vacancies at the Si-$SiO_2$ Interface," Japan J. Appl. Phys., 12, 73-79, Jan 1973.

[2.100] D.J. Fitzgerald and A.S. Grove, "Surface Recombination in Semiconductors," Surface Science, 9, 347-369, 1968.

[2.101] D. Shaw, "Self- and Impurity Diffusion in Ge and Si," Phys. Stat. Sol. (b), 72, 11-39, 1973.

[2.102] R.B. Fair and J.C.C. Tsai, "The Diffusion of Ion-Implanted Arsenic in Silicon," J. Electrochem. Soc., 122, 1689-1696, Dec 1975.

[2.103] A.F.W. Willoughby, "Interactions between Sequential Dopant Diffusions in Silicon - A Review," J. Phys. D: Appl. Phys., 10, 455-480, 1977.

[2.104] S. Matsumoto, E. Arai, H. Nakamura, and T. Niimi, "The Distribution of the Excess Vacancies in the Bulk at the Diffusion of Phosphorus into Silicon," Japan J. Appl. Phys., 16, 1177-1185, Jul 1977.

[2.105] R.B. Fair, "Quantitative Theory of Retarded Base Diffusion in Silicon n-p-n Structures with Arsenic Emitters," J. Appl. Phys., 44, 283-291, Jan 1973.

[2.106] S.P. Murarka, "The Role of Point Defects in the Growth of the Oxidation-Induced Stacking Faults in Silicon," Phys. Rev. B., 16, 2849-2857, Sep 1977.

[2.107] K.V. Ravi and C.J. Varker, "Oxidation-Induced Stacking Faults in Silicon. I. Nucleation Phenomenon," J. Appl. Phys., 45, 263-271, Jan 1974.

[2.108] H. Sunami, "Thermal Oxidation of Phosphorus-Doped Polycrystalline Silicon in Wet Oxygen," J. Electrochem. Soc., 125, 892-897, Jun 1978.

[2.109] T.I. Kamins, "Oxidation of Phosphorus-Doped Low-Pressure and Atmospheric-Pressure CVD Polycrystalline-Silicon Films," J. Electrochem. Soc., 126, pp. 838-844, May 1979.

[2.110] L. Csepregi, E.F. Kennedy, T.J. Gallagher, J.W. Mayer, and T.W. Sigmon, "Reordering of Amorphous Layers of Si Implanted with $^{31}P$, $^{75}As$, and $^{11}B$ Ions," J. Appl. Phys., 48, 4234-4240, Oct 1977.

[2.111] Y. Tamminga and W.J.M.J. Josquin, "The Regrowth of Amorphous Layers Created by High-Dose Antimony Implantation in (100) Silicon," Appl. Phys. Lett., 32, 13-15, Jan 1978.

Table 1. IDENTIFICATION OF CRYSTALLINE PHASE IN $SiO_2$ FILMS. Spacings from transmission electron diffraction patterns obtained at 100 kV; designated patterns intensified using radioisotopic image enhancement techniques.

| Sample | Description | $d_1$(Å) | $d_2$(Å) | $d_3$(Å) | $d_4$ (Å) | Comments |
|---|---|---|---|---|---|---|
| | ASIM (11-695) Reference data | 4.05 (101)* | 2.841 (102) | 2.485 (200) | 2.019 (202) | Low temperature tetragonal phase |
| A501 | $T_g$ = 700°C, $t_{ox}$ = 50 Å | --- | 2.85 | 2.48 | 2.019 | (101) ring obscured |
| A501R | $T_g$ = 700°C, $t_{ox}$ = 50 Å | 4.03 | 2.828 | 2.483 | 2.018 | Radioisotopic enhancement of diffraction pattern |
| A503 | $T_g$ = 800°C, $t_{ox}$ = 50 Å | --- | 2.83 | 2.46 | 2.009 | (101) ring obscured |
| A100 | $T_g$ = 800°C, $t_{ox}$ = 100 Å | 4.059 (weak) | 2.89 (very weak) | 2.49 (weak) | | |
| A101R | $T_g$ = 800°C, $t_{ox}$ = 100 Å | 4.02 | 2.89 | 2.45 | | Radioisotopic enhancement of diffraction pattern |

*(hkl) values given in parentheses.

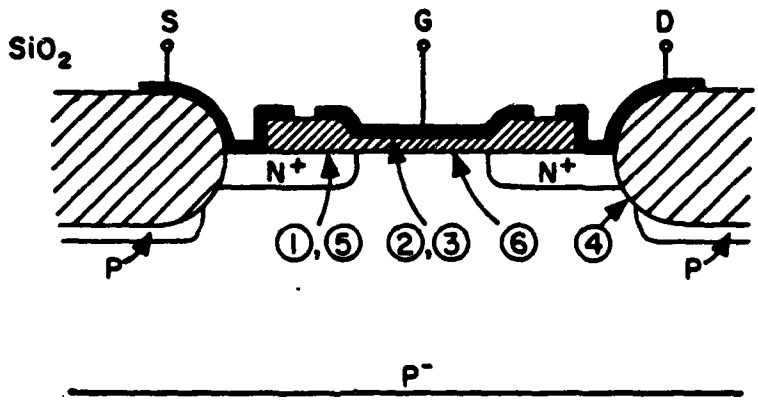



Fig. 2.1. Cross-section of oxide isolated NMOS transistor.

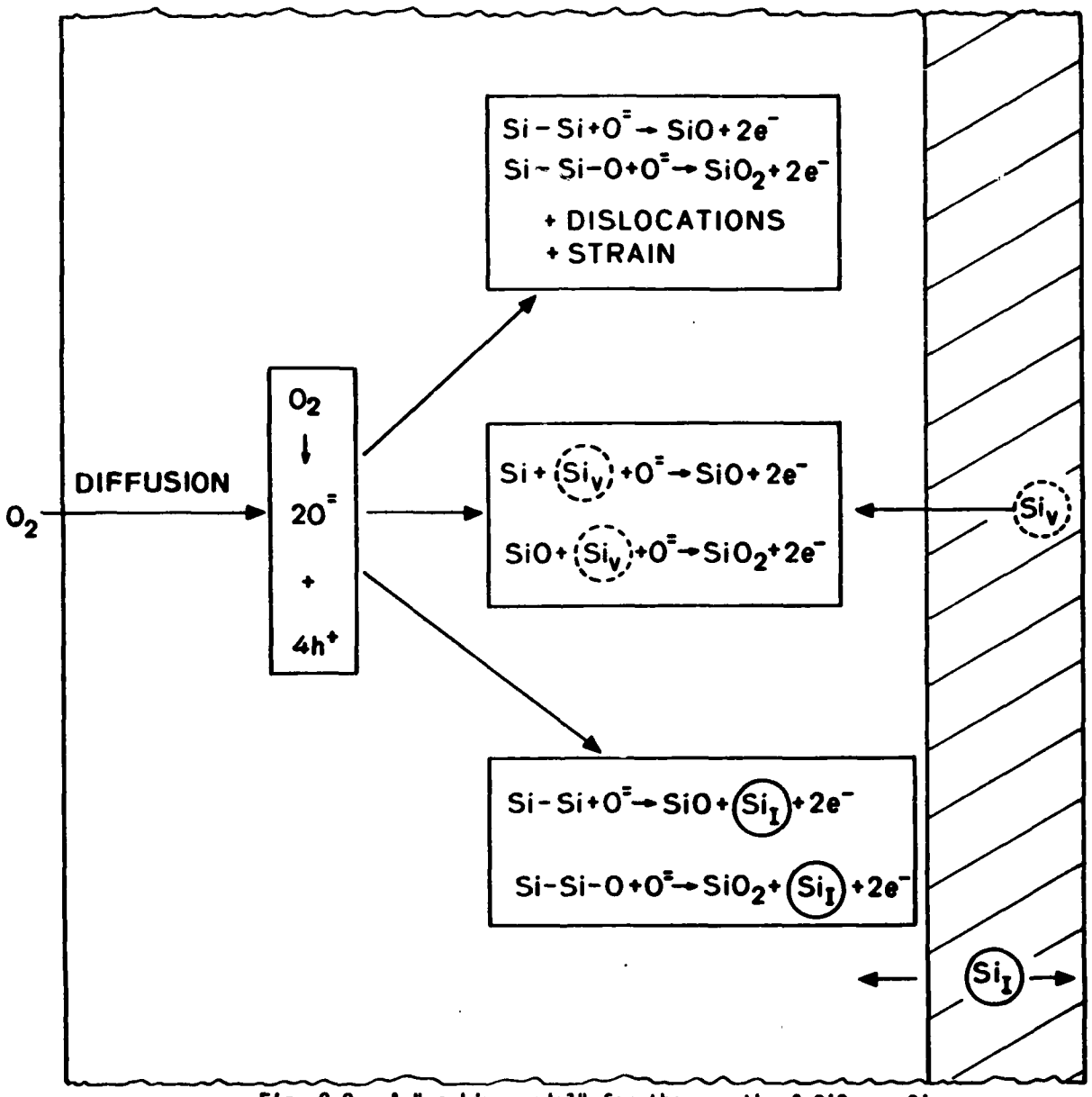



Fig. 2.2. A "working model" for the growth of $SiO_2$ on Si.

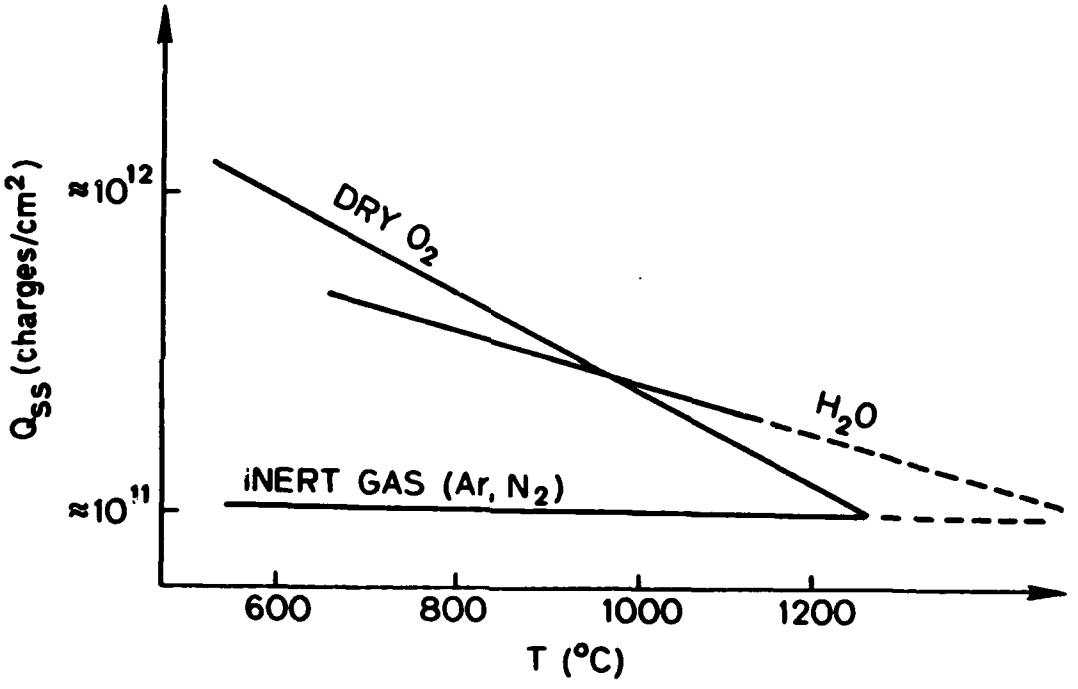



Fig. 2.3. "$Q_{ss}$ triangle" after Deal [2.19].

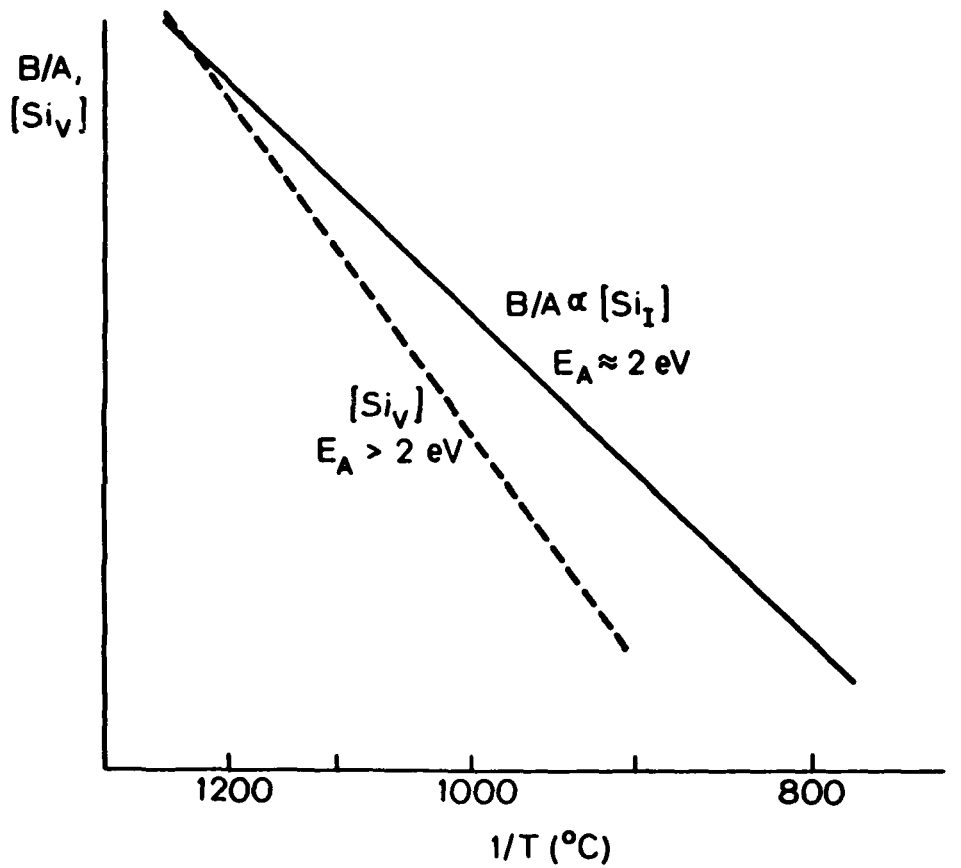



Fig. 2.4. Relative activation energies of interface reaction rate B/A and generation rate of $Si_V$.

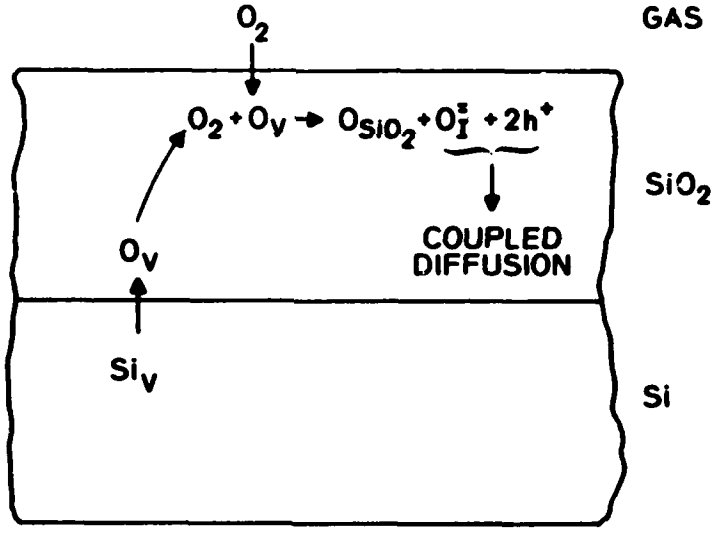



Fig. 2.5. Possible mechanism for enhanced growth rate of thin $SiO_2$ layers in dry $O_2$.

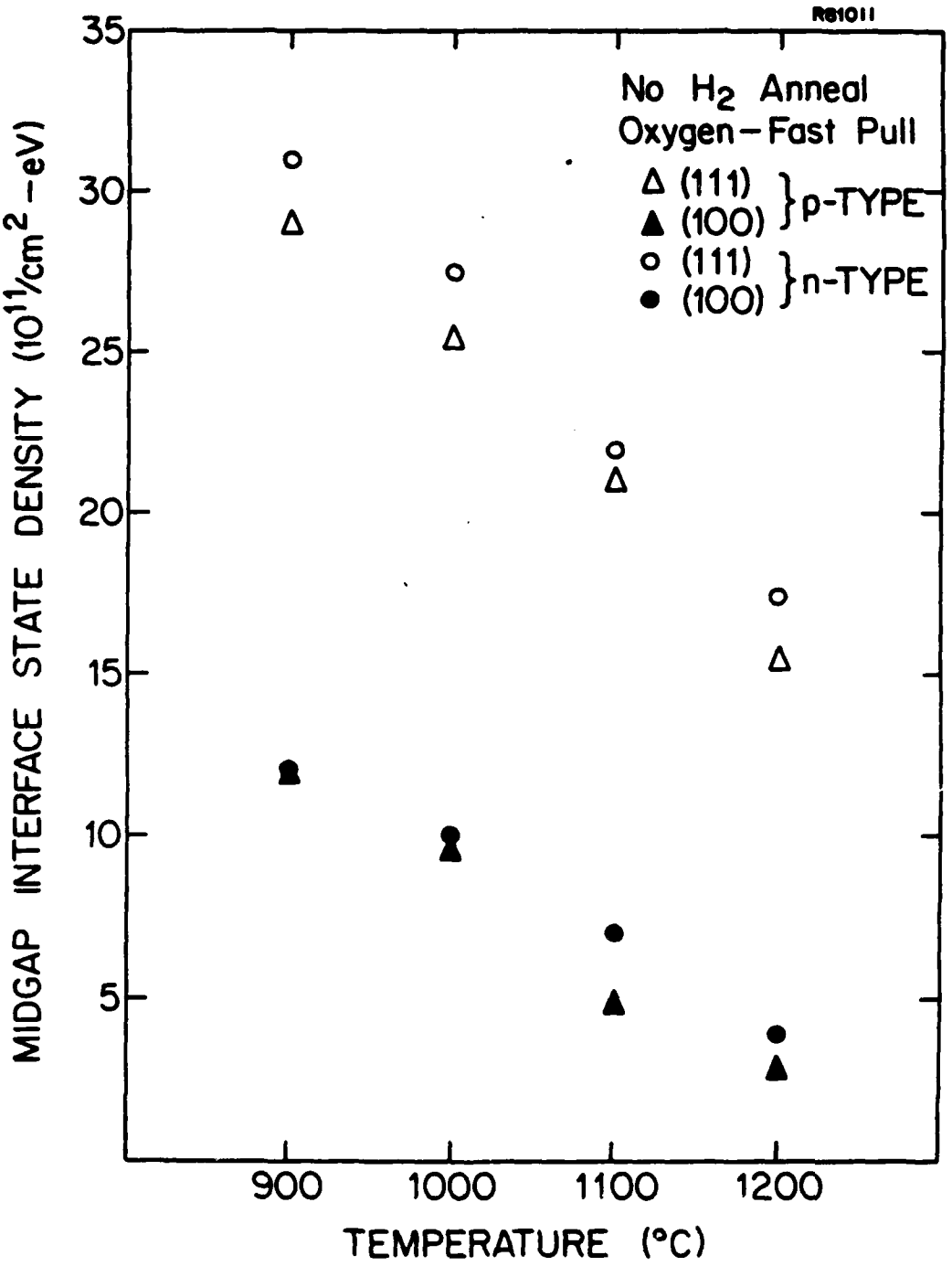



Fig. 2.6. Interface state density at midgap for p and n type silicon, (100) and (111), oxidized in dry $O_2$ and cooled in $O_2$ (<3 see pull). Data are for samples with no hydrogen anneal.

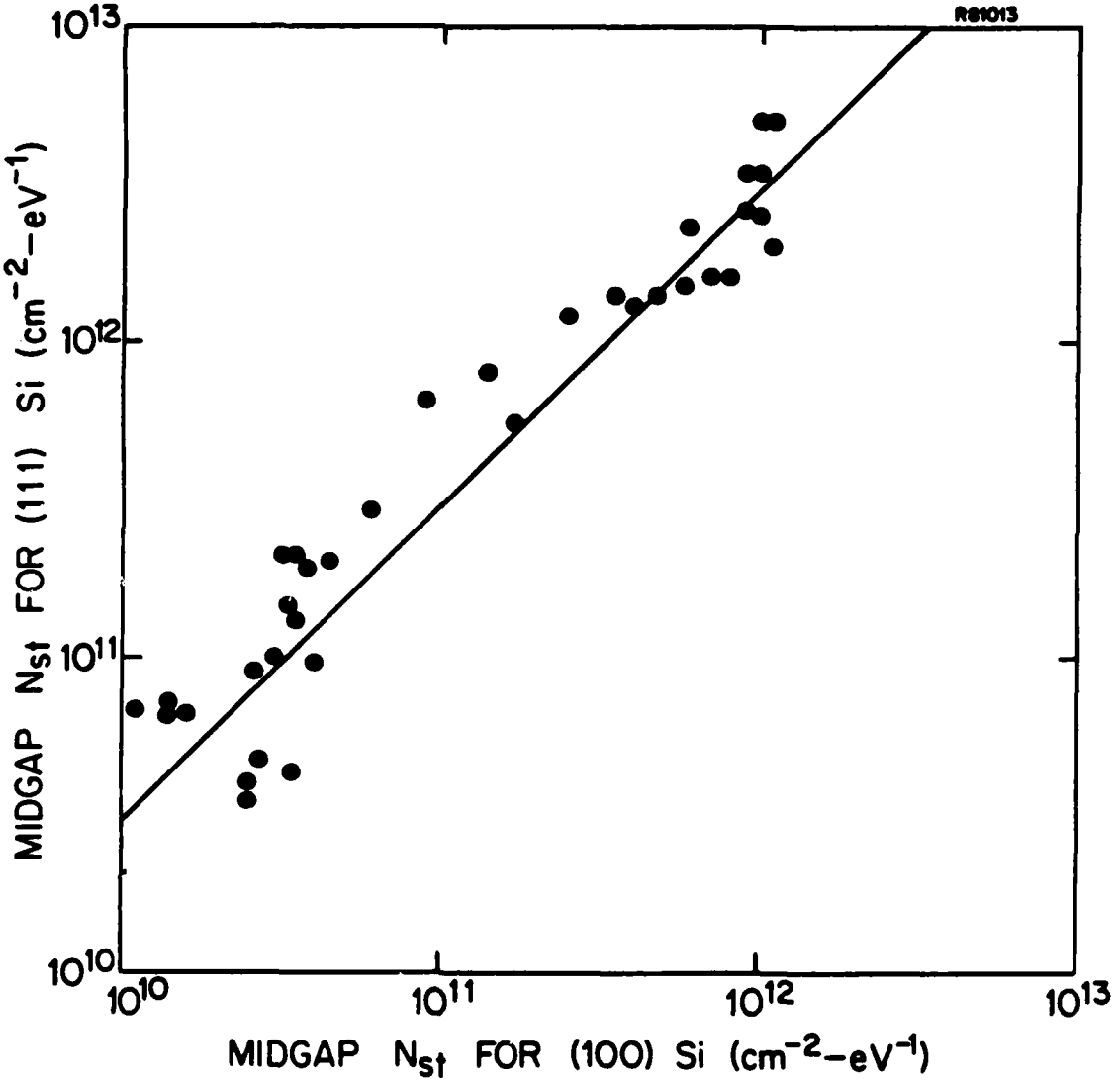



Fig. 2.7. Interface state density at midgap for (100) and (111) silicon as measured from $H_2$ annealed and non-$H_2$ annealed samples for various oxidation ambients, anneal ambients, and cooling rates.

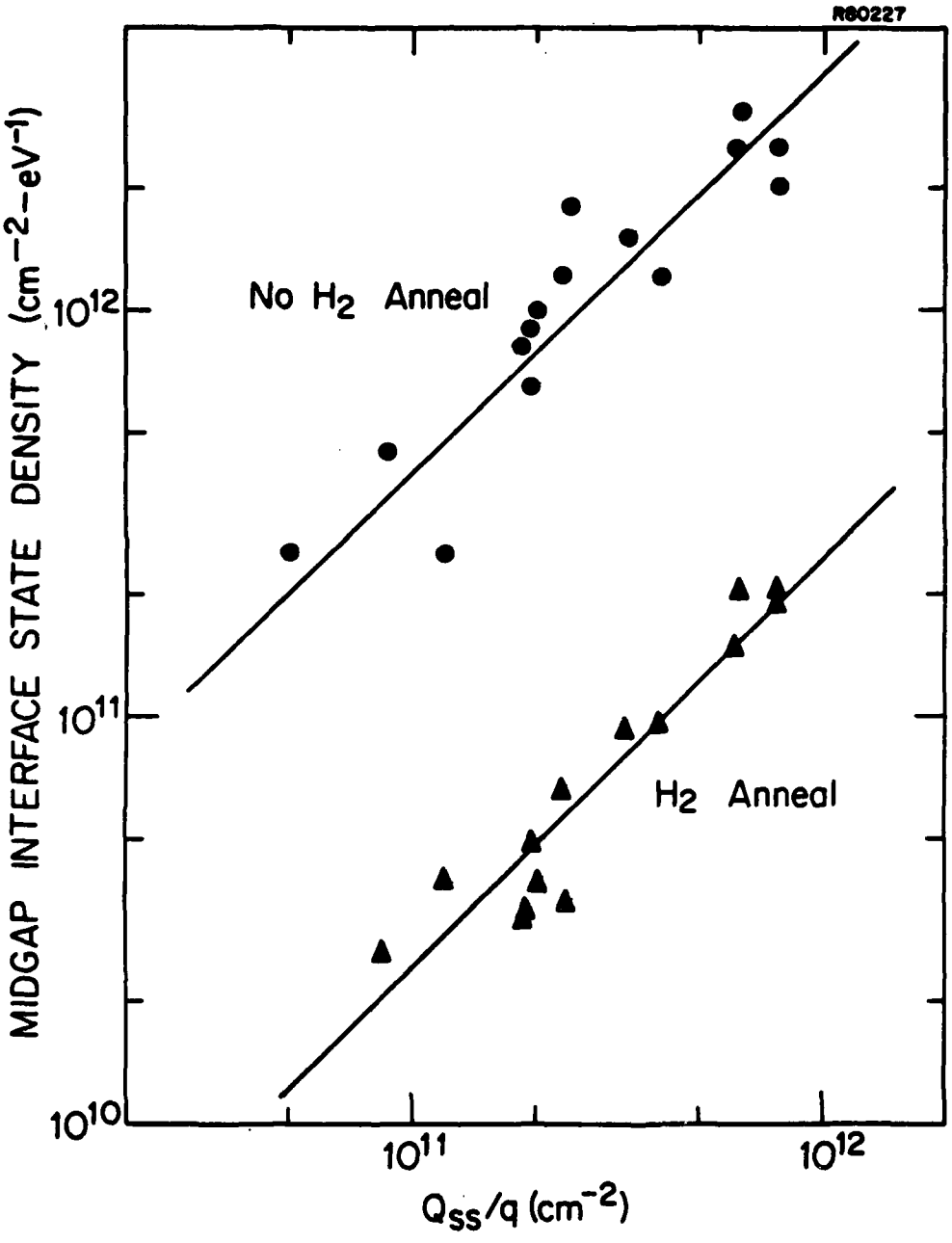



Fig. 2.8. Fixed oxide charge density vs interface state density at midgap for unannealed and $H_2$-annealed n-type and p-type silicon samples, (100) and (111) orientation, oxidized in dry $O_2$ at 1000° and 1200°C. (Hydrogen anneal: 2 l/min, 10% $H_2$ in $N_2$, 10 min, 500°C.)

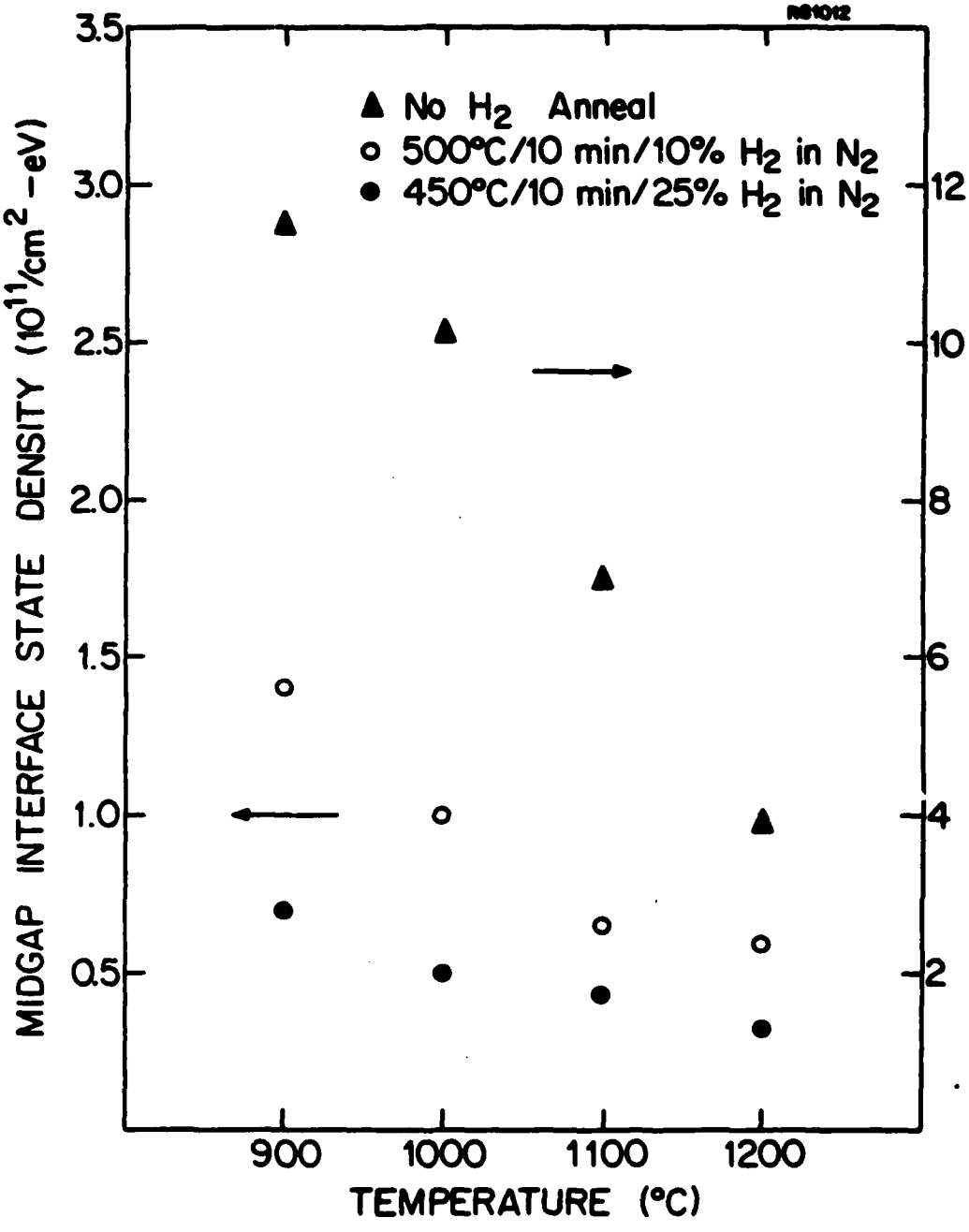



Fig. 2.9. Interface state density at midgap for n-type (100) silicon, oxidized and cooled in dry $O_2$ prior to low temperature $H_2$ annealing and after two different annealing treatments.

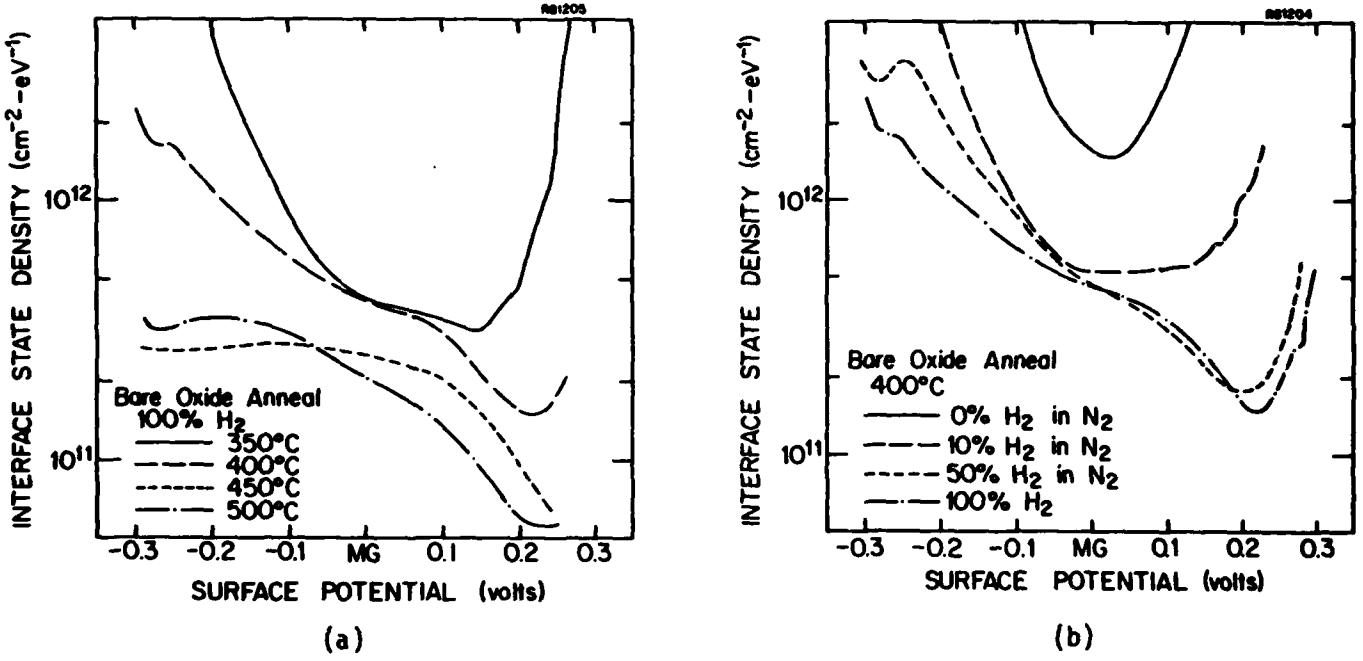
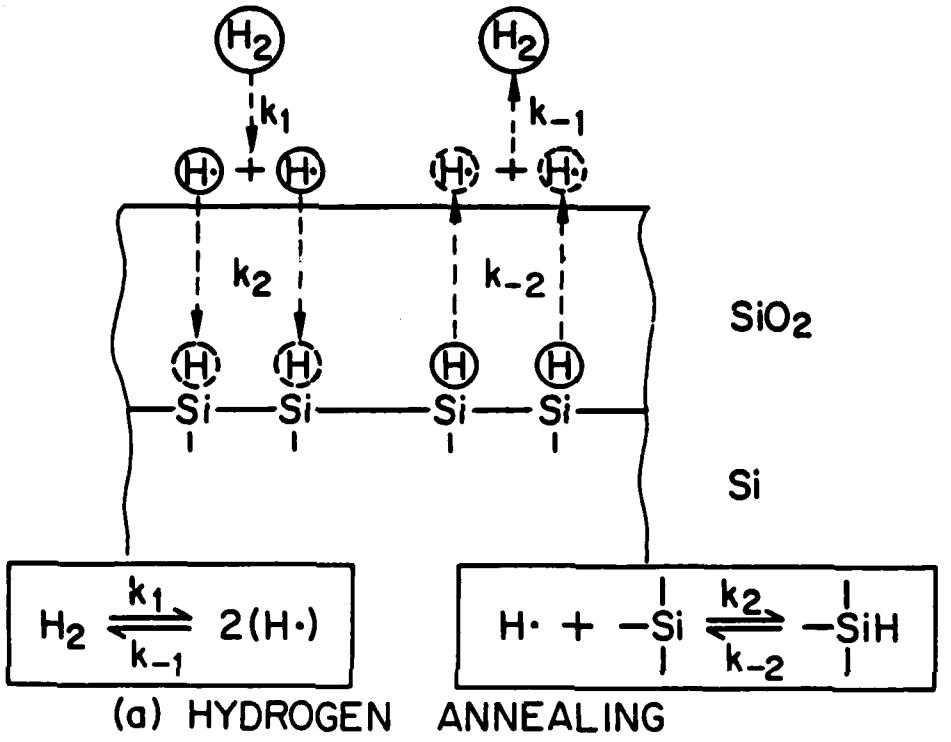



Fig. 2.10. Dependence of interface state density on low temperature annealing parameters for n-type (111) silicon samples, 4-6 Ω-cm, oxidized in dry $O_2$ ($O_2$ FP) and annealed for 10 min: (a) in 100% hydrogen at various temperatures, and (b) at 400°C in various hydrogen/nitrogen ambients.

# INTERFACE STATE ANNEALING MECHANISMS
## [VARIABLES: TEMP., TIME, $H_2$ CONC.]



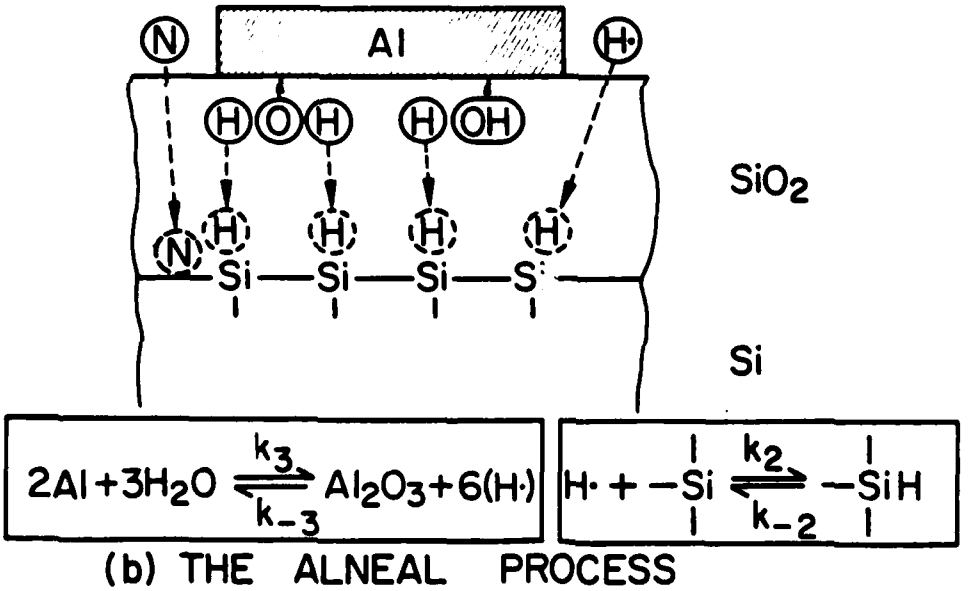




Fig. 2.11. Proposed mechanism for interface state annealing in hydrogen at low temperature with and without the presence of aluminum field plates.

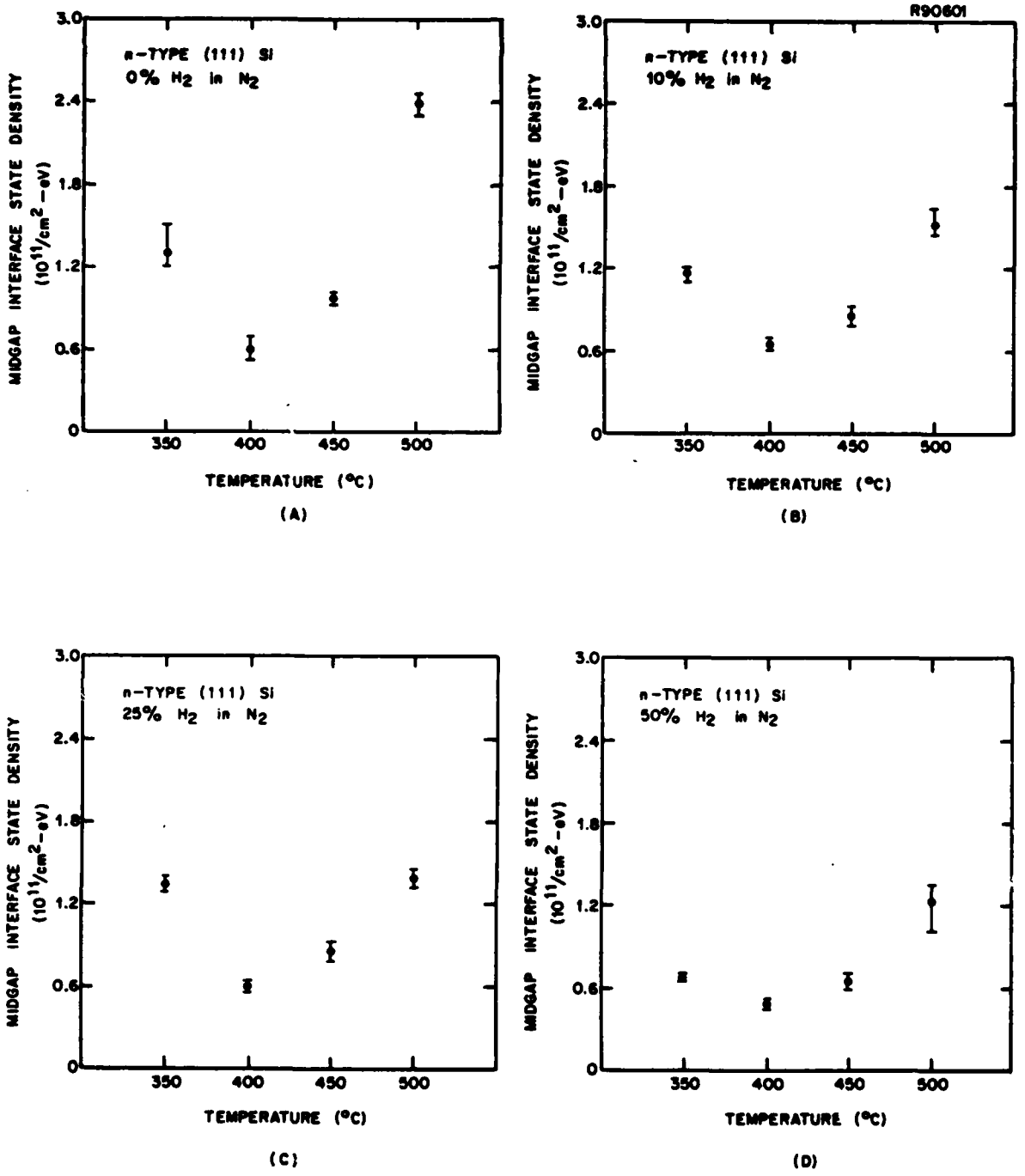



Fig. 2.12. Midgap interface state density for n-type (111) wafers oxidized in dry $O_2$ at 1000°C and pulled in oxygen ($O_2$ FP) for various post-metallization anneal treatments. All anneals are 10 min. The percentage of hydrogen in nitrogen is (a) 0%, (b) 10%, (c) 35%, and (d) 50%.

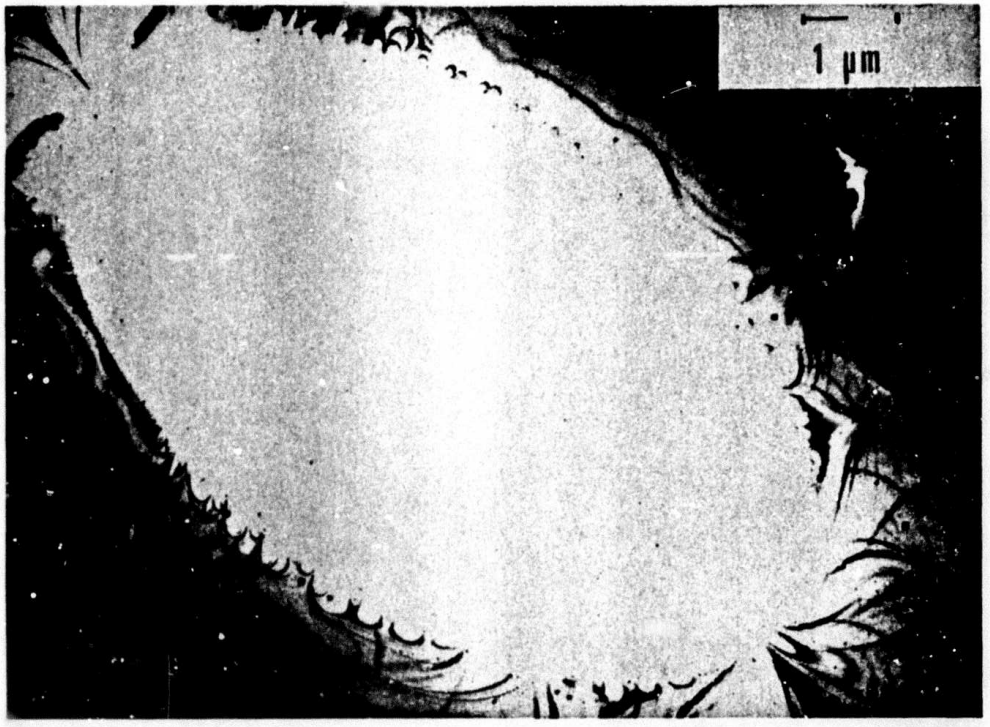



Fig. 2.13a. Morphology of a thinned sample. An $SiO_2$ film is suspended across the hole produced by the jet thinning process.

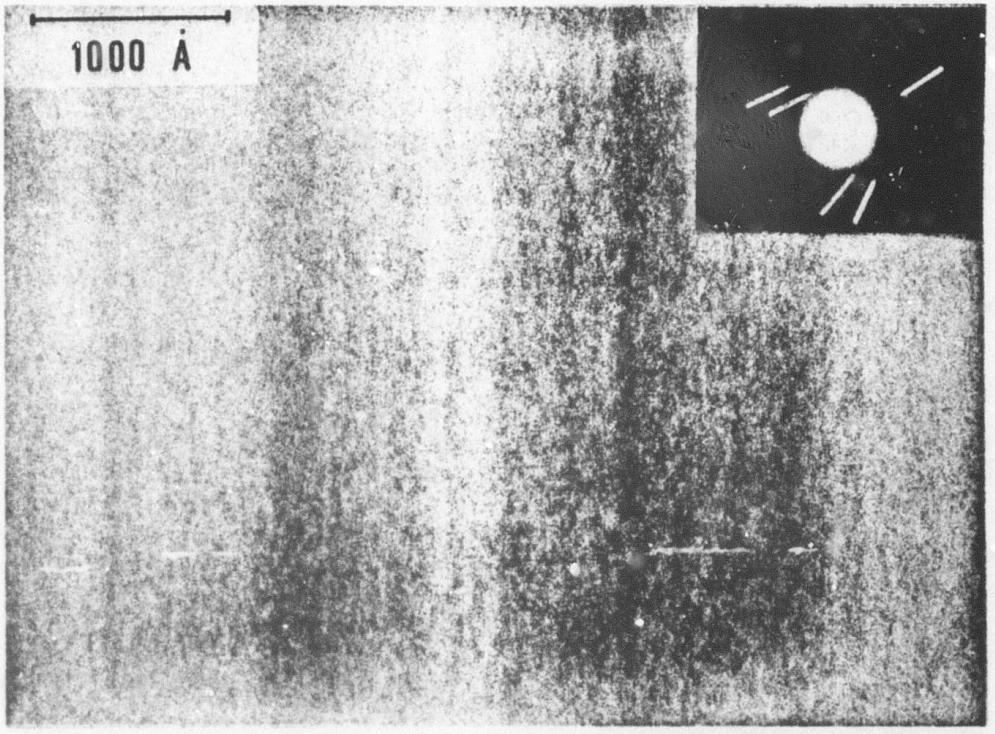



Fig. 2.13b. High magnification bright-field micrograph obtained at the center of the region in Fig. 2.13a showing the absence of structure and macroscopic growth defects within the film. The inset shows a selected area transmission electron diffraction pattern obtained from the 50 Å layer.

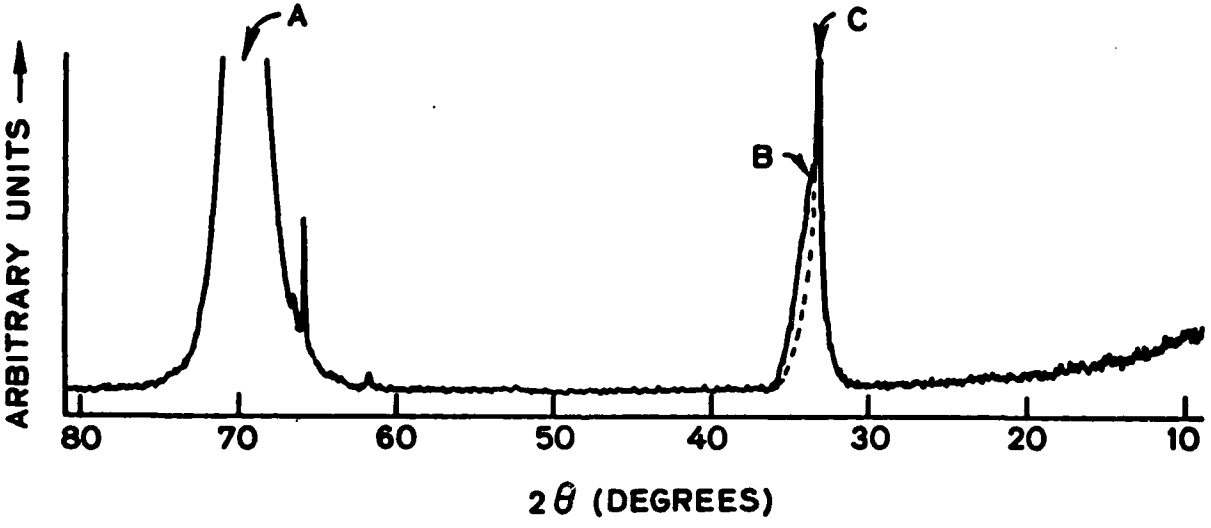



Fig. 2.14. X-ray diffraction spectrum of a Si-$SiO_2$ (~100 Å) interface using CuKα radiation.

A. Strong peak 69.2° (1.3565 Å) - λ reflection of (004) planes of Si (1.357 Å).
B. Weak broad peak 33.8° (2.6496 Å). This peak disappeared (dashed line) after changing the radiation window width and lower limit.
C. Weak narrow peak 32.95° (2.716 Å) - π/2 reflection of (004) planes of Si (2.719 Å).

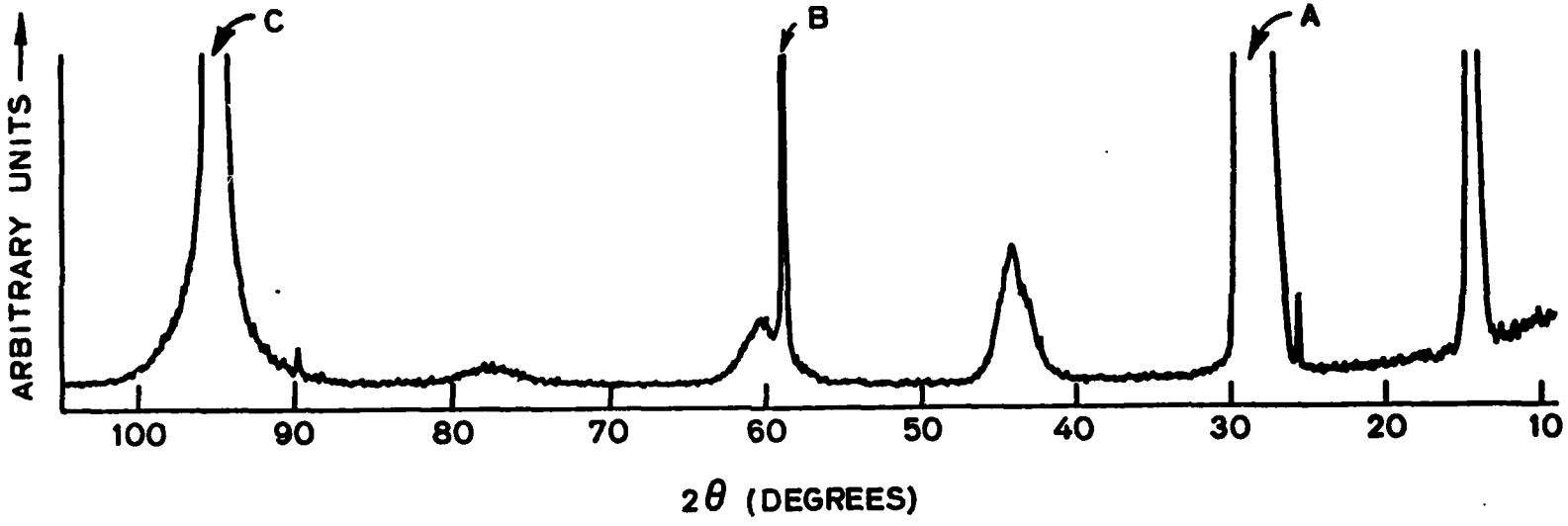



Fig. 2.15. X-ray diffraction spectrum of a Si-$SiO_2$ (~100 Å) interface using CuKα radiation.
A. Strong peak 28.5° (3.1292 Å) λ reflection of (111) planes of Si.
B. Weak peak 58.8° (1.5691 Å) λ reflection of (222) planes of Si.
C. Strong peak 95.0° (1.0447 Å) λ reflection of (333) planes of Si.

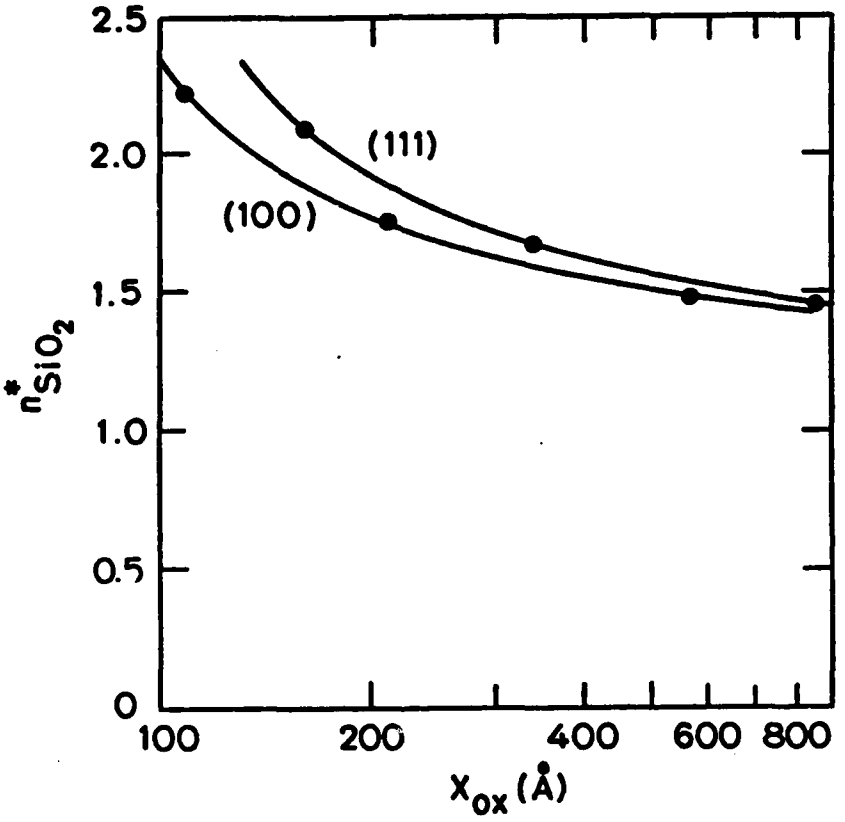



Fig. 2.16. Effective index of refraction of thermally grown $SiO_2$ layers, grown at 900°C in dry oxygen ambient on (100) and (111)-oriented single-crystal silicon substrates.

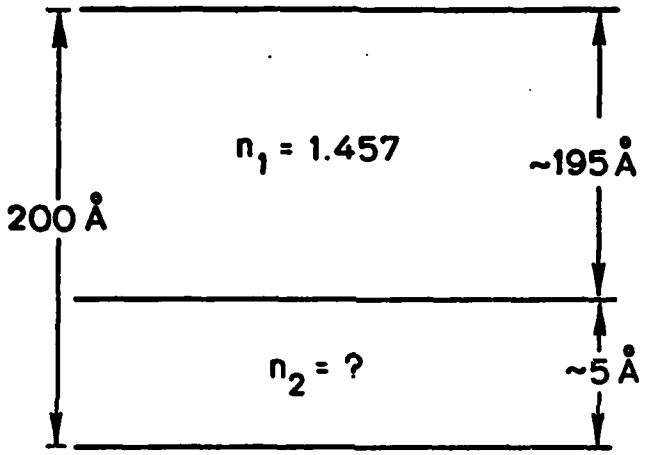



Fig. 2.17. Optical interface layer at the Si-$SiO_2$ interface. Its thickness is 4-7 Å and its index of refraction is 2.5-2.8.

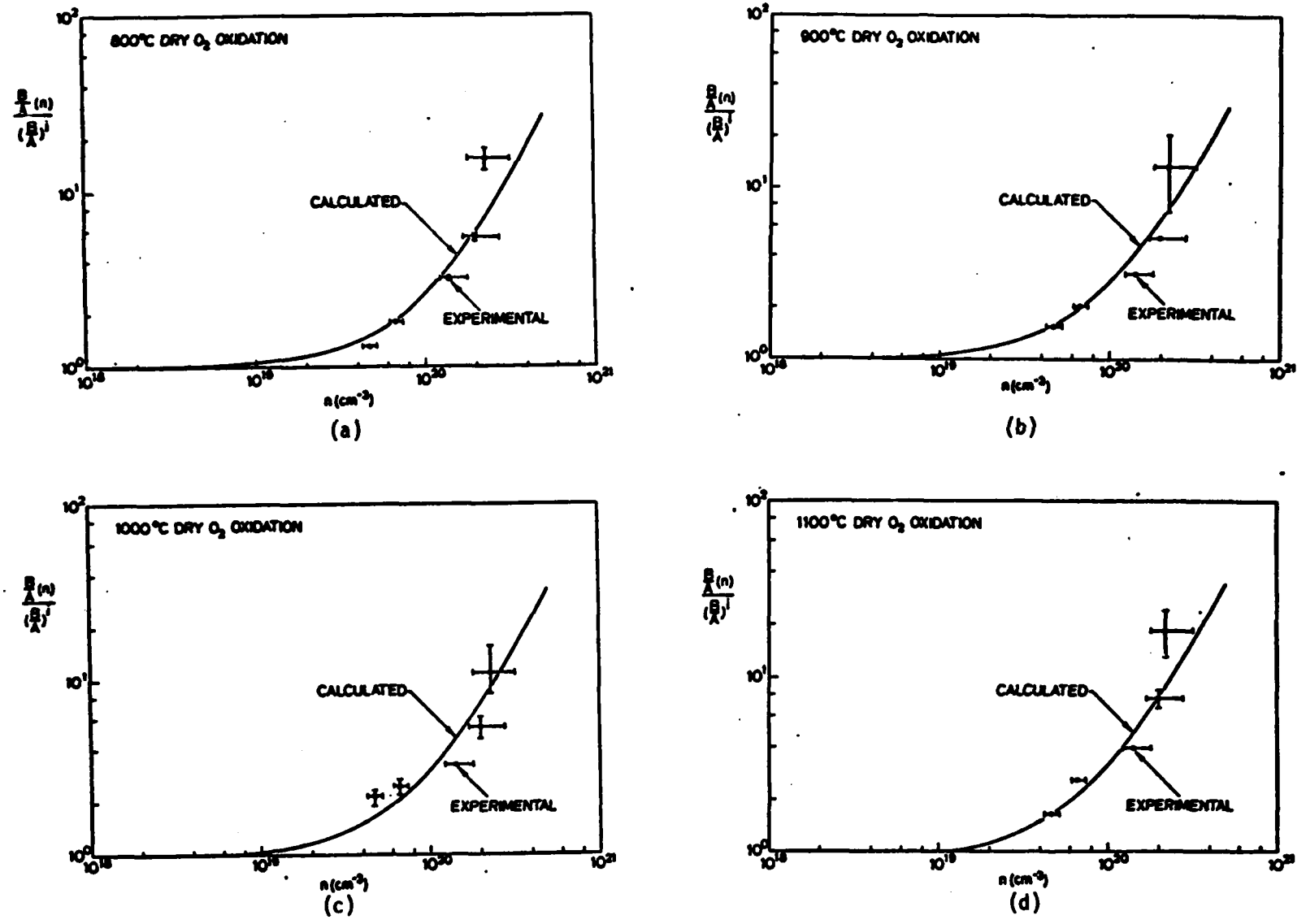



Fig. 2.18. Comparison of case A phosphorus experimental normalized linear rate constant to calculated values at (a) 800°C, (b) 900°C, (c) 1000°C, and (d) 1100°C.

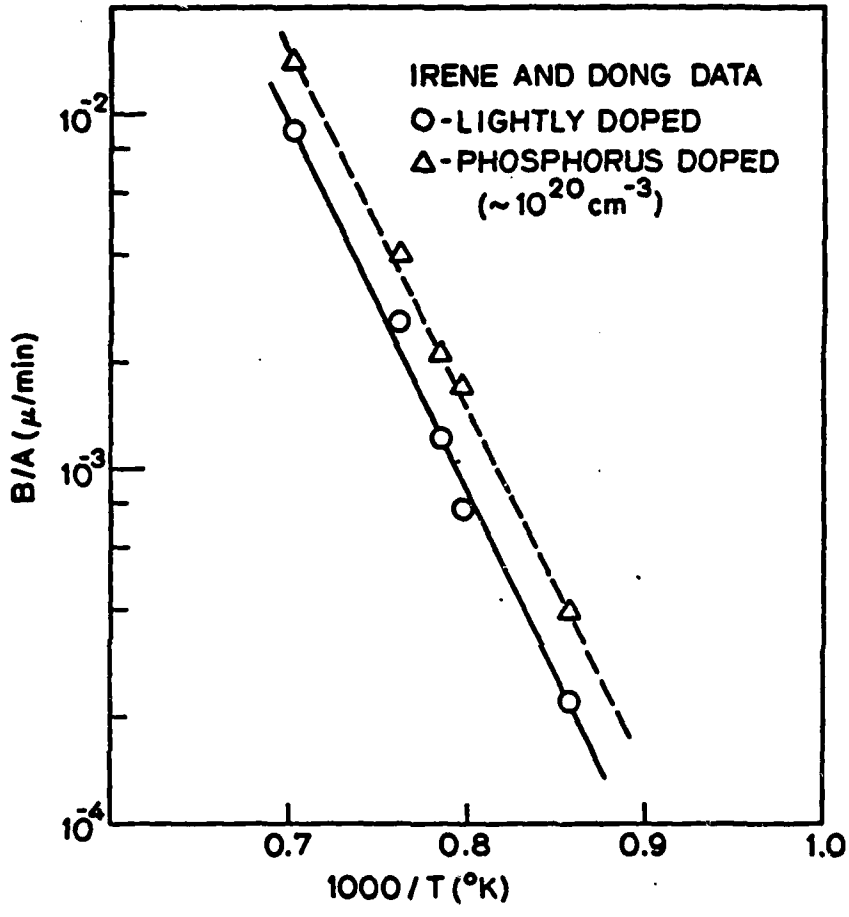



Fig. 2.19. Temperature dependence of case B phosphorus experimental linear rate constant.

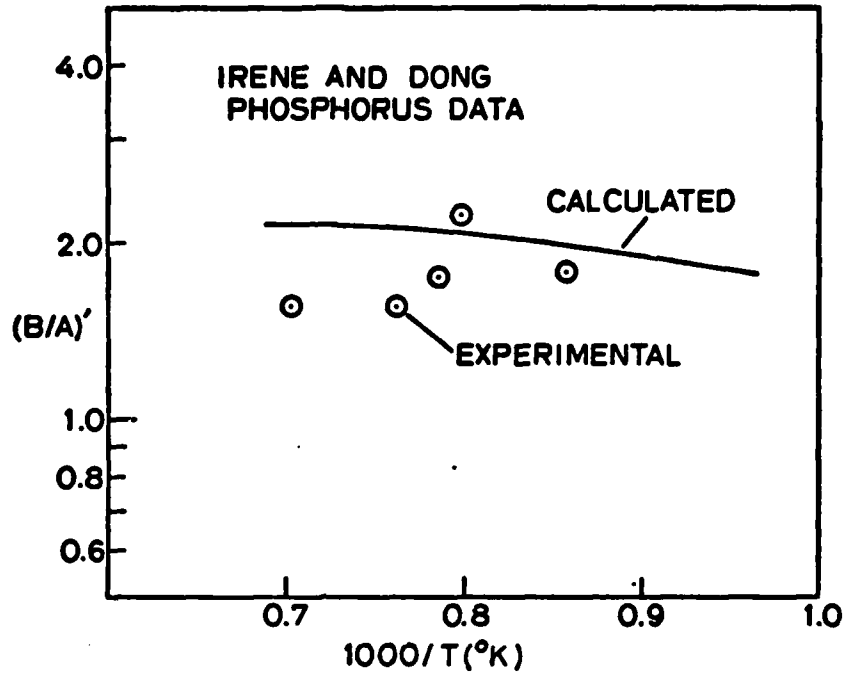
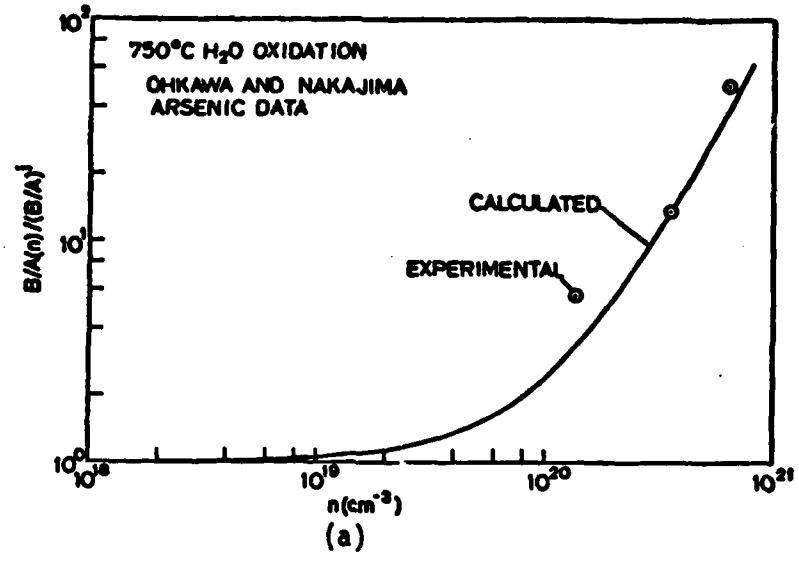



Fig. 2.20. Comparison of case B phosphorus experimental normalized linear rate constant to calculated values.



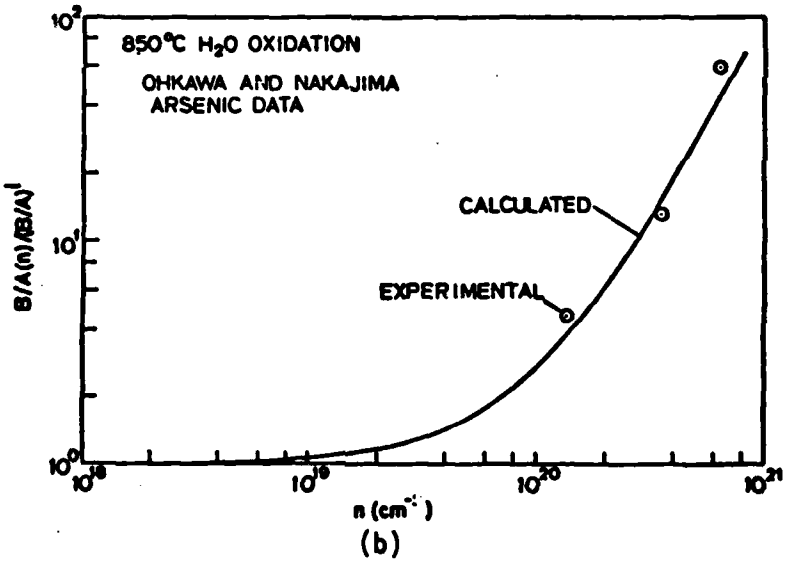




Fig. 2.21. Comparison of case C arsenic experimental normalized linear rate constant to calculated values at (a) 750°C and (b) 850°C.

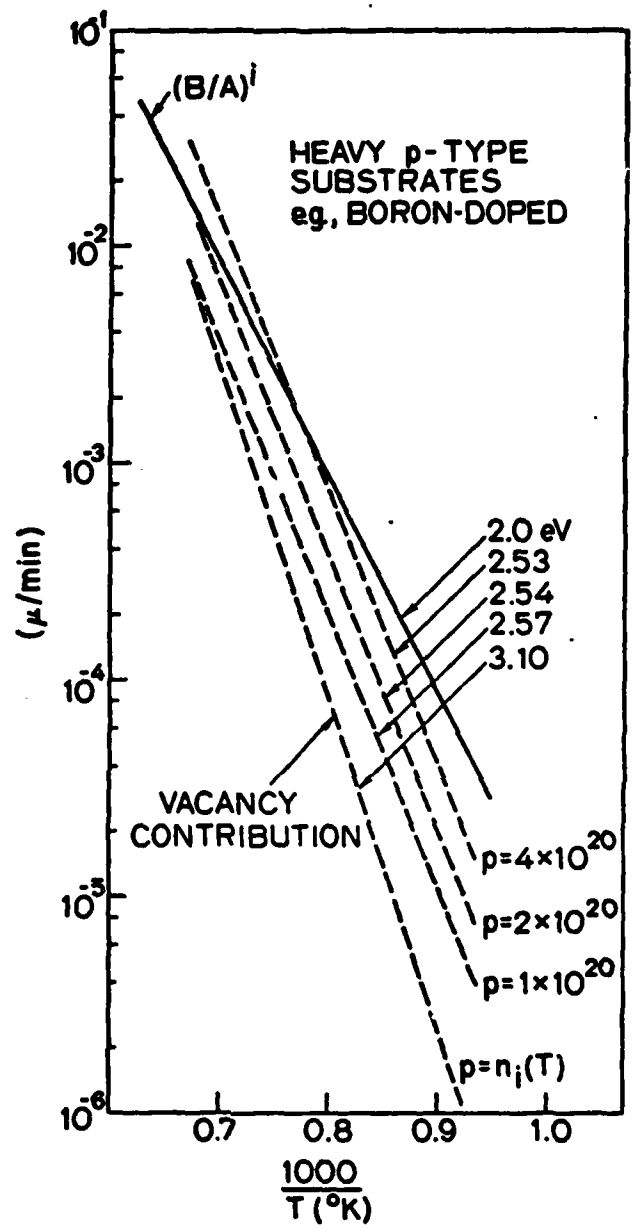



Fig. 2.22. Comparison of calculated vacancy contribution magnitude and temperature dependence for heavily doped p-type substrates (e.g. boron doped) with $(B/A)^1$.

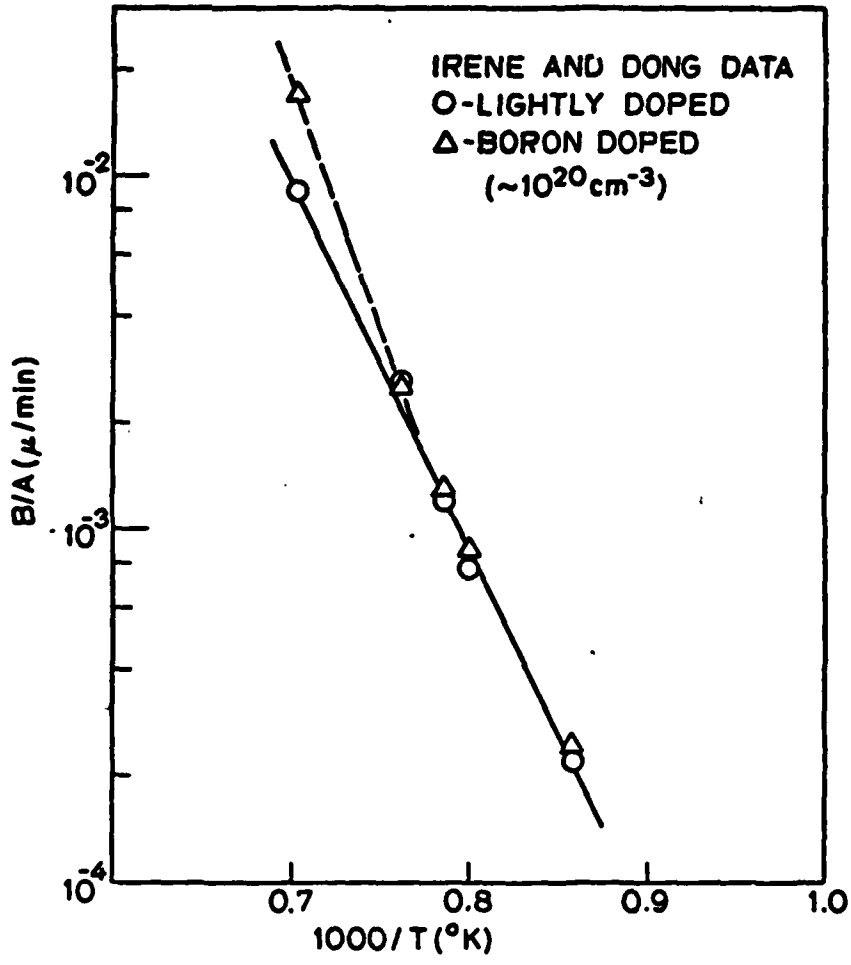



Fig. 2.23 Temperature dependence of case D boron experimental linear rate constant.

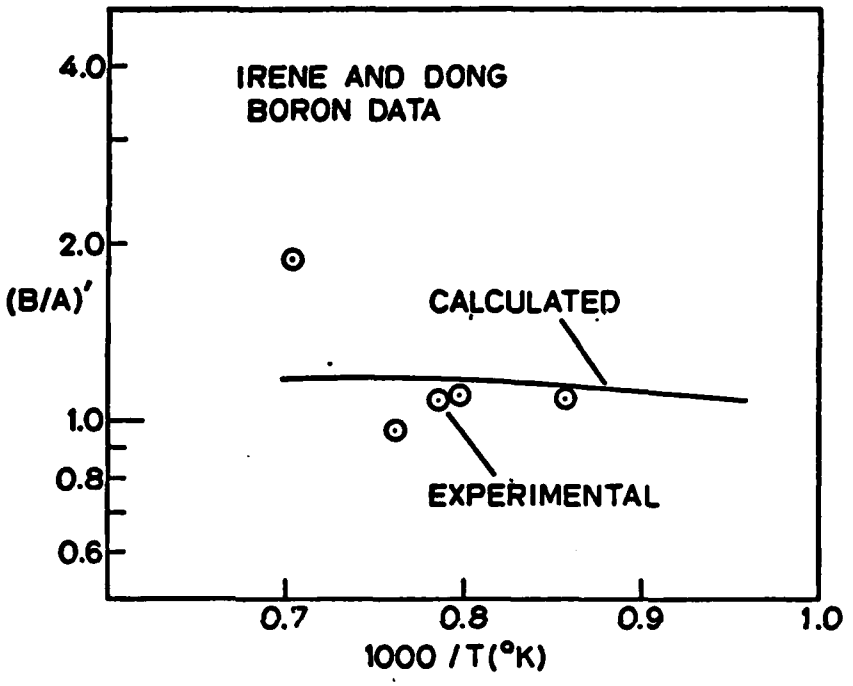



Fig. 2.24. Comparison of case D boron experimental normalized linear rate constant to calculated values.

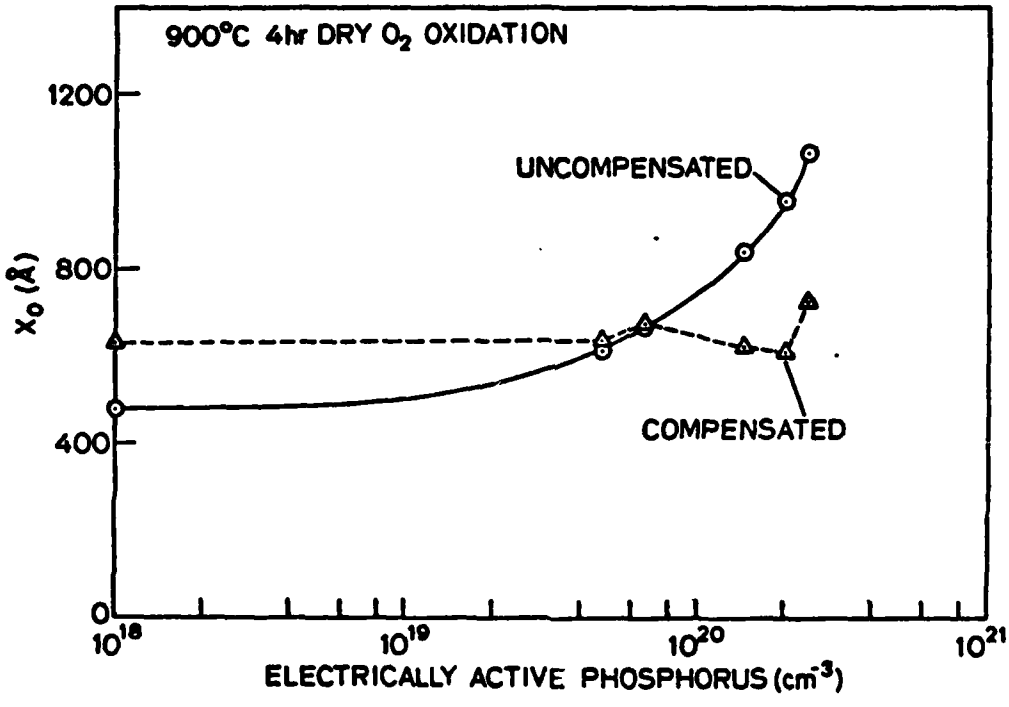



Fig. 2.25. Oxide thickness (oxidation rate) vs phosphorus concentration of phosphorus-doped silicon compensated and uncompensated by constant boron diffusion.

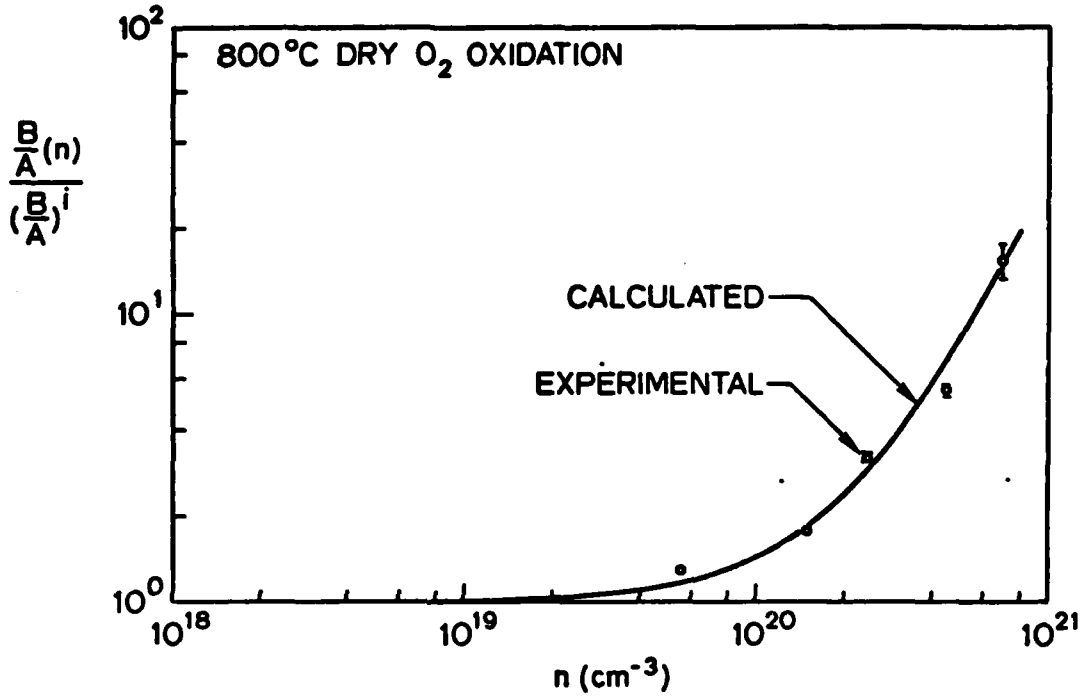



Fig. 2.26. Adjusted comparison of case A phosphorus experimental normalized linear rate constant with calculated values; T = 800°C and $n \cong C_{BC}$ is assumed.

Part 3

# CHEMICAL VAPOR DEPOSITION OF SILICON

K.C. Saraswat, R. Reif, T.I. Kamins,
M.M. Mandurah and F. Mohammadi

## 3.1 INTRODUCTION

Epitaxial deposition of crystalline silicon on single crystal substrates is heavily used in integrated circuit technologies, such as bipolar, $I^2L$, DMOS, VMOS, oxide isolation, and SOS. One of the major areas in silicon epitaxy, which has been so far relatively poorly understood, is dopant inclusion. A model for the dopant inclusion into silicon epitaxial films has been developed at the Stanford Integrated Circuits Laboratory. The physiochemical model presented here describes the incorporation of dopant atoms into silicon epitaxial films during deposition from a $SiH_4$-$AsH_3$-$H_2$ mixture in a horizontal atmospheric-pressure, epitaxial reactor. The model considers a sequence of processes occurring in the gas phase and at the surface. In order to properly describe the doping process under both transient and steady-state conditions, mass-balance of the As-containing species is considered at each important point in the epitaxial system. The detailed representation of these mass-balance equations is obtained, and seven first-order linear differential equations containing the mathematical description of the doping process result. In order to conveniently handle this set of equations, an equivalent electric circuit represented by an analogous set of equations is found. This RC circuit representation provides insight into the different mechanisms taking part in the doping process and their relative importance.

Both the steady-state and the transient behavior of the epitaxial doping process predicted by the model are then described and compared to previously

presented experimental results using the equivalent circuit. In steady-state the model indicates that the doping process is controlled by thermodynamics at very low growth rates and by reaction kinetics at high growth rates. The decrease in the epitaxial-layer arsenic concentration observed with increasing deposition temperature at high growth rates is used to show that the dominant mechanism occurs at the growing surface. According to the model, at low growth rates the transient response is controlled by the RC time constant associated with the relaxation time of the slowest step in the kinetic sequence. At high growth rates it is controlled by the rate at which arsenic atoms occupying incorporation sites on the growing surface are covered by subsequently arriving silicon atoms. The decay time associated with the transient is independent of the growth rate at very low rates and is inversely proportional to the growth rate at high rates, in agreement with experimental results.

Polycrystalline silicon has been largely responsible for the success of silicon-gate MOS technology. Therefore it is important to understand the deposition and doping of polycrystalline silicon. Characterization of the structure and stability of films deposited by low-pressure chemical vapor deposition was described in the last year's final report of this program [3.15]. In this report, conduction in polycrystalline-silicon films deposited by low-pressure chemical vapor deposition has been investigated and compared to that in films deposited at atmospheric pressure. Low-pressure films were deposited at 580° and 620°C and doped with phosphorus by ion implantation. Films deposited at 620°C were polycrystalline while those deposited at 580°C were initially amorphous but crystallized readily on further heat treatment. The effect of annealing temperature on resistivity was studied in the low-pressure films for two phosphorus doses, and the resistivity was found to decrease with

increasing annealing temperature. After annealing the films deposited at 580°C always had lower resistivity than did those deposited at 620°C, with the most marked differences seen at lower annealing temperatures. In a second set of experiments, phosphorus was implanted with a wide range of doses corresponding to average dopant concentrations of $2 \times 10^{15}$ to $2 \times 10^{20}/cm^3$. The resistivity was only a slow function of dopant concentration below $6 \times 10^{16}/cm^3$ and above $2 \times 10^{18}/cm^3$, while in the intermediate range slight changes in concentration caused large changes in resistivity. As before, films deposited at 580°C always had the lowest resistivity, especially in the intermediate doping range. The Hall mobility was measured and found to be maximum near a dopant concentration of $6 \times 10^{18}/cm^3$ and to decrease rapidly at lower dopant concentrations. The observed behavior is consistent with that expected from a film composed of small crystallites surrounded by grain boundaries containing large numbers of carrier traps.

Use of silicides of refractory metals to form gates and interconnections in MOS integrated circuits have been suggested because of their higher conductivity as compared to polycrystalline silicon. However, the compatibility of silicides with present processing technology has to be investigated in detail before they can be used to fabricate integrated circuits. In this report we have investigated the effects of high temperature annealing on the structure and resistivity, and low temperature annealing on the stability of Al contacts to thin sputtered films of $WSi_2$. $WSi_2/SiO_2/Si$ and $WSi_2$/poly-Si/$SiO_2$/Si structures were annealed in $N_2$ in the temperature range of 500°C to 1200°C. Changes in the resistivity of the $WSi_2$ films due to annealing were measured and related to changes in structure of the films as determined by optical microscope, SEM, and X-ray diffraction measurements. The above two structures were coated with Al and

annealed in $N_2$ in the temperature range of 400°C to 700°C and stability of $Al/WSi_2$ interface was determined by studying the structure of the films using the techniques listed above. From these studies, $WSi_2$ appears to be highly attractive to form electrode and interconnection layers in integrated circuits.

## 3.2 A MODEL FOR DOPANT INCORPORATION INTO GROWING SILICON EPITAXIAL FILMS

### 3.2.1 Introduction

Transient and steady-state experimental studies of the dopant system of horizontal, silicon epitaxial reactor have been reported earlier [3.1-3.5]. One of the objectives of that work was to find a transfer function relating the time-variation of the input dopant gas concentration to the spatial-variation of the epitaxial-layer dopant concentration. Arsine ($AsH_3$) was the dopant species considered, and silicon growth was from silane ($SiH_4$) in a hydrogen ambient in an atmospheric-pressure reactor. In this report, a model will be presented to describe physically and mathematically the mechanisms by which dopant atoms are incorporated into growing Si epitaxial films. The model will then be compared with the experimental results.

The work presented in this report was motivated by the results obtained in the transient study of the doping process. As was reported earlier [3.4,3.5], after the dopant gas flow was abruptly changed, 2-3 min elapsed before a new steady-state condition was achieved. This time appears to be too long to be associated with a gas-phase mechanism (Fig. 3.1). The time to mix the reactant species in the reactor plumbing system is relatively short, and the time spent in the main gas stream of the reactor by an element of the reacting gas is only about 4 sec. On the other hand, Shepherd [3.6] and Bloem [3.7] have proposed that the rate-limiting step in the doping process is the mass transport of the dopant compound across the boundary layer

separating the well-mixed gas stream from the growing surface. These models assume that surface reactions occur rapidly compared to mass transport. A rough estimate of the associated diffusion time may be obtained from

$$t = \delta^2 / \left[ D_0 \, x (T/T_0)^m \right]$$

where $\delta$ is the boundary-layer thickness, $D_0$ is the diffusion coefficient of $AsH_3$ in $H_2$ at $T_0$ = 273 K, T is the average boundary-layer temperature, and m is normally taken to be between 1.75 and 2.00 [3.8]. With $\delta$ = 0.5 cm, $D_0$ = 0.31 $cm^2$/sec [3.8], T = 1100 K, and m = 1.75, the resulting diffusion time is 0.07 sec, several orders of magnitude shorter than the experimentally observed transient times. Therefore, mechanisms in addition to gas-phase processes (e.g. mechanisms occurring at the surface) should be considered to explain the relatively long transient times. It has also been reported [3.4] that the transient response is a strong function of the epitaxial growth rate, becoming faster at higher growth rates. This fact also cannot be explained solely by gas-phase processes. Therefore, in order to understand the physical mechanisms responsible for the transient behavior, the model for the doping process presented in this report was developed. Although this model is analyzed for an atmospheric system using $AsH_3$ as the dopant source, it can be readily extended to other dopant species and reduced-pressure systems.

The sequence of steps taking part in the doping process will be described in Section 3.2.2. In Section 3.2.3 mass-balance of the dopant species is applied at each point in the doping process in order to obtain a set of equations physically describing the doping process under transient and steady-state conditions. In Sections 3.2.4 and 3.2.5 representative equations

are derived, providing a mathematical description of the doping process. Finally, in Section 3.2.6 an equivalent RC circuit representing the doping process is derived from the equations obtained in Section 3.2.4. This circuit representation is useful in providing insight into the mechanisms described in Section 3.2.2 and assessing their relative importance.

In the following sections, the equivalent circuit will be used to study the steady-state and transient behavior of the epitaxial doping process and to show that the physiochemical model developed in Section 3.2.2 explains the experimental behavior reported earlier [3.4]. Steady-state conditions will be considered in Section 3.2.7, and then the more general time-varying case will be discussed in Section 3.2.8. The equivalent circuit will also be used in Section 3.2.9 to derive the "transfer function" of the dopant system [3.4], i.e. an expression relating the time variation of the doping gas concentration to the spatial variation of the epitaxial-layer dopant concentration. The expression obtained using the equivalent circuit will be shown to be in agreement with the transfer function obtained experimentally [3.4].

### 3.2.2 Sequence of Steps in the Doping Process

The doping process can be divided into a number of steps occurring sequentially [3.9]. A similar approach has been taken by Faktor and Garrett [3.10] and by Shaw [3.11], among others, in discussing crystal growth from the vapor. The present work will consider mechanisms occurring both in the gas phase and at the solid surface.

The sequence of steps occurring in the gas phase is shown in Fig. 3.1. (1) Forced-convection mass transport considers the transport of the dopant compound ($AsH_3$) from the reactor-tube entrance to the deposition region. (2) Boundary-layer mass transport considers the transport of $AsH_3$ from the

well-mixed main gas stream through the boundary layer to the growing surface. (3) Gas-phase chemical reactions. Arsine ($AsH_3$) may dissociate into several different As-containing species at elevated temperatures. These chemical reactions may occur throughout the boundary layer, although they are more important close to the growing surface. Duchemin [3.12] showed that $AsH_3$ and $AsH_2$ are the most abundant species above the surface, and in the present treatment, these species are taken to be the most important As-containing compounds just above the gas-solid interface. Other As-containing species such as AsH, As, $As_2$ and $As_4$ may be incorporated into the analysis similarly, but they are omitted from the present treatment for simplicity.

The discussion of mechanisms occurring at the growing surface uses the so-called terrace-ledge-kink model [3.13] for the surface of a crystal. This model divides the surface into three different types of sites: adsorption or terrace sites, step or ledge sites, and kink sites. In this work, the lower-energy step and kink sites will be termed incorporation sites, and atoms occupying those sites will be called incorporated atoms to differentiate them from atoms occupying adsorption sites, which will be called adsorbed atoms. The sequence of steps occurring at the surface is shown in Fig. 3.2. (4) Adsorption of the As-containing species at an adsorption site on the growing surface. (5) Chemical dissociation into As and H in the adsorbed layer. At this point, a population of $AsH_3$, As, H, etc., occupies adsorption sites, and is able to move on the surface. (6) Surface diffusion and incorporation of adsorbed As at step and kink sites on the surface. (7) "Burying" of the incorporated As by subsequently arriving Si atoms during epitaxial growth. (8) Desorption of hydrogen from the surface. As will be discussed in Section 3.2.5, the population of adsorbed hydrogen is assumed to be in equilibrium with the $H_2$ in the gas

phase because of the large hydrogen partial pressure (approximately 1 atm.).

The importance of step 7 to the doping process should be emphasized since it provides the driving force for the overall doping process. In the absence of epitaxial growth but with a constant $AsH_3$ flow, steps 1-6 will be in thermodynamic equilibrium due to the lack of a surface "sink" for the adsorbed As atoms. However, the process of burying incorporated As atoms during epitaxial growth has the effect of unbalancing the equilibrium between As atoms occupying incorporation sites and As atoms occupying adsorption sites, thereby forcing a net flow of adsorbed As into incorporation sites (step 6) in order to approach equilibrium. This in turn unbalances each of the previous steps in the doping process, so that the burying of As atoms occupying incorporation sites provides the driving force for the doping process.

The rate at which incorporated As atoms are covered by subsequently arriving silicon atoms (step 7) determines whether the doping process occurs under quasi-equilibrium conditions, or is controlled by the reaction kinetics. If the As atoms are covered very slowly, the doping process is in quasi-equilibrium. If they are rapidly covered, reaction kinetics dominate. In the latter case the doping density in the growing film depends on the rate at which the As atoms arrive at the incorporation sites, which in turn is controlled by the slowest step in the kinetic sequence (steps 1-6). Note that As does not play an active role in step 7, in contrast with steps 1-6. Step 7 depends on Si atoms arriving at incorporation sites and burying the incorporated As.

Steps 1-7 occur consecutively. Since the doping process does not occur under thermodynamic equilibrium, a net flux of As-containing species flows through the chain of steps and the slowest step in the chain is the rate-limiting step [3.10,3.11]. If the doping process were controlled by thermodynamics,

each of the steps 1-7 would be in equilibrium, and the flow of As-containing species through the chain of steps would be strictly zero. It will be shown later that at very low epitaxial growth rates, the doping process is very near equilibrium, while at high growth rates, the doping process is kinetically controlled, and the flow of As-containing species is limited by one of the steps 1-6 in the sequence.

### 3.2.3 Mass Balance in the Doping Process

By analyzing the sequence of steps discussed above, a model can be derived to mathematically describe the doping process. Under steady-state conditions, i.e. when the input dopant gas flow is constant, all of the steps occur at the same rate since they are in series, and the resulting dopant distribution in the epitaxial layer is uniform. However, when the dopant gas flow is a function of time, a transient occurs during which all of the steps involved in the doping process no longer occur at the same rate, and the correct physical picture is properly described by considering mass-balance of the As-containing species at each important point in the process. For example, mass-balance of $AsH_3$ in some specified region is given by

$$\begin{pmatrix}\text{rate of input}\\ \text{of } AsH_3\end{pmatrix} - \begin{pmatrix}\text{rate of depletion}\\ \text{of } AsH_3\end{pmatrix} = \begin{pmatrix}\text{rate of change of}\\ \text{the quantity of } AsH_3\end{pmatrix}$$

This approach was used by Kobayashi and Kobayashi to describe the transient-response method in heterogeneous catalysis [3.14]. For reference in later sections, the equations for mass balance at the important points are schematically outlined below.

Mass-balance of $AsH_3$ in the main gas stream within the deposition region:

$$\left[\begin{array}{l}\text{rate at which}\\ AsH_3 \text{ enters the}\\ \text{deposition region}\\ \text{from the inlet}\end{array}\right] - \left[\begin{array}{l}\text{rate at which}\\ AsH_3 \text{ leaves the}\\ \text{deposition region}\\ \text{to the exhaust}\end{array}\right] - \left[\begin{array}{l}\text{rate at which}\\ AsH_3 \text{ leaves the}\\ \text{main gas stream}\\ \text{toward the wafer}\\ \text{surface}\end{array}\right] = \left[\begin{array}{l}\text{rate of change}\\ \text{of } AsH_3 \text{ concen-}\\ \text{tration in the}\\ \text{main gas stream}\\ \text{within the depo-}\\ \text{sition region}\end{array}\right] \tag{3.1}$$

Mass-balance of $AsH_3$ just above the gas-solid interface:

$$\left[\begin{array}{l}\text{rate at which}\\ AsH_3 \text{ arrives at}\\ \text{the gas-solid}\\ \text{interface}\end{array}\right] - \left[\begin{array}{l}\text{net rate of chemical}\\ \text{dissociation of } AsH_3\\ \text{just above the gas-}\\ \text{solid interface}\end{array}\right] - \left[\begin{array}{l}\text{net rate of}\\ \text{adsorption of}\\ AsH_3 \text{ on the}\\ \text{growing sur-}\\ \text{face}\end{array}\right] = \left[\begin{array}{l}\text{rate of change}\\ \text{of } AsH_3 \text{ concen-}\\ \text{tration just}\\ \text{above the gas-}\\ \text{solid interface}\end{array}\right] \tag{3.2}$$

Mass-balance of $AsH_3$ adsorbed on the surface (adsorbed $AsH_3$):

$$\left[\begin{array}{l}\text{net rate of adsorption}\\ \text{of } AsH_3 \text{ on the growing}\\ \text{surface}\end{array}\right] - \left[\begin{array}{l}\text{net rate of chemical}\\ \text{dissociation of}\\ \text{adsorbed } AsH_3\end{array}\right] = \left[\begin{array}{l}\text{rate of change of the}\\ \text{surface density of}\\ \text{adsorbed } AsH_3\end{array}\right] \tag{3.3}$$

Mass-balance of As adsorbed on the surface (adsorbed As):

$$\left[\begin{array}{l}\text{net rate of forma-}\\ \text{tion of adsorbed As}\end{array}\right] - \left[\begin{array}{l}\text{net rate of incorpor-}\\ \text{ation of adsorbed As}\\ \text{into step or kink}\\ \text{sites on the surface}\end{array}\right] = \left[\begin{array}{l}\text{rate of change of the}\\ \text{surface density of}\\ \text{adsorbed As}\end{array}\right] \tag{3.4}$$

Mass-balance of As atoms occupying incorporation sites (incorporated As):

$$\left[\begin{array}{l}\text{net rate of incor-}\\ \text{poration of}\\ \text{adsorbed As into}\\ \text{step or kink sites}\end{array}\right] - \left[\begin{array}{l}\text{rate of burying the}\\ \text{incorporated As atoms}\\ \text{by subsequently arriv-}\\ \text{ing silicon atoms}\end{array}\right] = \left[\begin{array}{l}\text{rate of change of the}\\ \text{surface density of}\\ \text{incorporated As}\end{array}\right] \tag{3.5}$$

The following two equations are analogous to Eqs. (3.2) and (3.3), but refer to $AsH_2$ instead of $AsH_3$:

Mass-balance of $AsH_2$ just above the gas-solid interface:

$$\left[\begin{array}{l}\text{net rate of forma-}\\ \text{tion of } AsH_2 \text{ above}\\ \text{the gas-solid inter-}\\ \text{face}\end{array}\right] - \left[\begin{array}{l}\text{net rate of adsorp-}\\ \text{tion of } AsH_2 \text{ on the}\\ \text{growing surface}\end{array}\right] = \left[\begin{array}{l}\text{rate of change of } AsH_2\\ \text{concentration just above}\\ \text{the gas-solid interface}\end{array}\right] \tag{3.6}$$

Mass-balance of $AsH_2$ adsorbed on the surface (adsorbed $AsH_2$):

$$\left[\begin{array}{l}\text{net rate of adsorp-}\\ \text{tion of } AsH_2 \text{ on the}\\ \text{growing surface}\end{array}\right] - \left[\begin{array}{l}\text{net rate of chemical}\\ \text{dissociation of adsorbed}\\ AsH_2\end{array}\right] = \left[\begin{array}{l}\text{rate of change of the}\\ \text{surface density of}\\ \text{adsorbed } AsH_2\end{array}\right] \tag{3.7}$$

Under steady-state conditions the right-hand side of Eqs. (3.1)-(3.7) becomes zero, and each of the steps in the sequence proceeds at the same rate.

### 3.2.4 Mathematical Description of the Doping Process

In order to obtain a mathematical description of the doping process, a detailed representation of each of the schematic equations outlined above must be obtained. Representative equations are described in detail below, and the others are found analogously. First, rate equations for some of the mechanisms shown in Figs. 3.1 and 3.2 will be derived, and then they will be used in the corresponding mass-balance equation.

A. Mass-Balance of $AsH_3$ in the Main Gas Stream Within the Deposition Region [Eq. 3.1]

1. The flux of $AsH_3$ entering the deposition region from the inlet is given by:

$$N_{300} \times v_{H_2} \times X^{i}_{AsH_3} \tag{3.8}$$

where $N_{300}$ is the total gas-phase molecular concentration per unit volume at 300 K, $v_{H_2}$ is the hydrogen velocity at 300 K, and $X^{i}_{AsH_3}$ is the $AsH_3$ molar fraction in the main gas upstream of the deposition region. This molar fraction can be easily adjusted by controlling the $AsH_3$ gas flow. Similarly, the flux of $AsH_3$ leaving the deposition region toward the exhaust is given by

$$N_{300} \times v_{H_2} \times X^{o}_{AsH_3} \tag{3.9}$$

where $X^{o}_{AsH_3}$ is the $AsH_3$ molar fraction in the main gas downstream of the deposition region.

2. The rate at which $AsH_3$ within the deposition region leaves the main gas stream toward the wafer surface ($r_2$) can be taken to be proportional to the $AsH_3$ concentration gradient at the upper edge of the boundary layer. Since the time associated with changing the $AsH_3$ concentration within the boundary layer is short compared to the times of interest here (see Section 3.2.1),

a constant gradient across the boundary layer may be used in the expression for $r_2$:

$$r_2 = k_m \times \left[ \bar{X}_{AsH_3} - X^s_{AsH_3} \right] \tag{3.10}$$

where $k_m$ is the boundary-layer mass transport coefficient of $AsH_3$ in $H_2$, $\bar{X}_{AsH_3}$ is the average molar fraction of $AsH_3$ in the main gas stream within the deposition region, and $X^s_{AsH_3}$ is the $AsH_3$ molar fraction just above the gas-solid interface.

The molar fraction of $AsH_3$ in the main gas stream within the deposition region is a function of position due to diffusion to $AsH_3$ toward the susceptor. Nevertheless, under the deposition conditions typically used [3.4], less than 5% of the incoming $AsH_3$ molecules take part in the doping process. Therefore, an average $AsH_3$ molar fraction ($\bar{X}_{AsH_3}$) independent of position can be used without causing serious error.

3. The rate of change of $AsH_3$ concentration in the main gas stream within the deposition region is given by

$$L \times N_T \times \frac{d\bar{X}_{AsH_3}}{dt} \tag{3.11}$$

where L is the length of the susceptor, and $N_T$ is the total molecular concentration per unit volume in the main gas stream within the deposition region, which is at a temperature T.

4. By substituting Eqs. (3.8), (3.9), (3.10), and (3.11), into Eq. (3.1) and using the fact that $\bar{X}_{AsH_3} \approx X^o_{AsH_3}$, an equation describing the mass balance of $AsH_3$ in the main gas stream within the deposition region

is obtained

$$N_{300} \times v_{H_2} \times \left[ X^{i}_{AsH_3} - \overline{X}_{AsH_3} \right] - k_m \times \left[ \overline{X}_{AsH_3} - X^{s}_{AsH_3} \right] = L \times N_T \times \frac{d\overline{X}_{AsH_3}}{dt} \quad (3.12)$$

B. Mass-Balance of $AsH_3$ Just Above the Gas-Solid Interface [Eq. (3.2)]

1. As explained in the previous section, for the times of interest the concentration gradient is approximately constant throughout the boundary layer, and the rate at which $AsH_3$ arrives at the gas-solid interface is nearly the same as the rate at which $AsH_3$ leaves the main gas stream toward the wafer surface [Eq. (3.10)].

2. As stated previously, the only gas-phase chemical reaction that will be considered is the dissociation of $AsH_3$ into $AsH_2$

$$AsH_3(g) \underset{k_{r3}}{\overset{k_{f3}}{\rightleftharpoons}} AsH_2(g) + \frac{1}{2} H_2(g) \qquad K_3 = \frac{\left( X^{s}_{AsH_2} \times P_{H_2} \right) \times P^{\frac{1}{2}}_{H_2}}{\left( X^{s}_{AsH_3} \times P_{H_2} \right)} = \frac{k_{f3}}{k_{r3}} \quad (3.13)$$

where $K_3$ is the equilibrium constant for the chemical reaction, $X^{s}_{AsH_2}$ is the $AsH_2$ molar fraction just above the gas-solid interface, $P_{H_2}$ is the hydrogen pressure (1 atm.), and $k_{f3}$ ($k_{r3}$) is the forward (reverse), first-order, reaction rate constant for the chemical reaction. It will be assumed that the reactions involved in the doping process are reversible, and that the forward and reverse reaction rates are first order with respect to reactants and products.

The net rate of chemical dissociation of $AsH_3$ into $AsH_2$ just above the gas-solid interface $r_3$ is given by the difference between the rate of the forward ($r_{f3}$) and the reverse ($r_{r3}$) reactions

$$r_3 = r_{f3} - r_{r3} = k_{f3} \times X^{s}_{AsH_3} - k_{r3} \times X^{s}_{AsH_2} \qquad (3.14)$$

($P_{H_2}$ = 1 atm. is omitted for simplicity in most of the equations.) If the chemical reaction is in thermodynamic equilibrium, i.e. if the gas-phase $AsH_3$ concentration is in equilibrium with the gas-phase $AsH_2$ concentration, then $r_{f3} = r_{r3}$, and $r_3 = 0$.

3. The net rate of adsorption of $AsH_3$ on the growing surface can be analyzed similarly. In the adsorption process, an $AsH_3$ molecule in the gas-phase finds a vacant adsorption site on the growing surface and occupies it for a certain length of time

$$AsH_3(g) + s \underset{k_{r4}}{\overset{k_{f4}}{\rightleftharpoons}} AsH_3\text{-}s \qquad K_4 = \frac{\left(\theta_{AsH_3} \times N_s\right)}{\left(X^{s}_{AsH_3} \times P_{H_2}\right) \times \left(\theta \times N_s\right)} = \frac{k_{f4}}{k_{r4}} \qquad (3.15)$$

where s represents a vacant adsorption site on the surface, $AsH_3$-s represents an $AsH_3$ molecule occupying an adsorption site, $K_4$ is the equilibrium constant for the adsorption reaction, $\theta_{AsH_3}$ is the fraction of adsorption sites occupied by $AsH_3$ molecules, $N_s$ is the surface density of adsorption sites (occupied and unoccupied) per unit area, $\theta$ is the fraction of adsorption sites which are vacant, and $k_{f4}$ and $k_{r4}$ are the corresponding forward and reverse reaction rate constants, respectively.

The rate of adsorption $r_{f4}$ is proportional to the number of $AsH_3$ molecules just above the gas-solid interface and the fraction of the surface available for adsorption, while the rate of desorption $r_{r4}$ is proportional

to the amount of $AsH_3$ present on the surface. The net rate of adsorption of $AsH_3$ on the growing surface is then

$$r_4 = r_{f4} - r_{r4} = k_{f4} \times X^{s}_{AsH_3} \times \theta - k_{r4} \times \theta_{AsH_3} \quad (3.16)$$

4. The rate of change of $AsH_3$ concentration just above the gas-solid interface is given by

$$\epsilon \times N_T \times \frac{d X^{s}_{AsH_3}}{dt} \quad (3.17)$$

where $\epsilon \times N_T$ is the total number of molecules per unit area in a thin layer thickness $\epsilon$ at temperature T' just above the gas-solid interface.

5. By substituting Eqs. (3.10), (3.14), (3.16), and (3.17) into Eq. (3.2), an equation is obtained which mathematically describes the mass-balance of $AsH_3$ just above the gas-solid interface

$$k_m \cdot \left[\bar{X}_{AsH_3} - X^{s}_{AsH_3}\right] - \left[k_{f3} \cdot X^{s}_{AsH_3} - k_{r3} \cdot X^{s}_{AsH_2}\right] - \left[k_{f4} \cdot X^{s}_{AsH_3} \cdot \theta - k_{r4} \cdot \theta_{AsH_3}\right] = \epsilon \cdot N_T \cdot \frac{dX^{s}_{AsH_3}}{dt} \quad (3.18)$$

C. The mathematical representation of Eqs. (3.3), (3.4), (3.6), and (3.7) can be derived by applying the same techniques used above for Eq. (3.2) [3.15], and the chemical reactions and mass-balance equations are summarized without detailed derivation in Appendix A.

D. Mass-Balance of As Atoms Occupying Incorporation Sites [Eq. (3.5)]

This equation will be derived because it is slightly different from the mass-balance expressions considered above.

1. The net rate of incorporation of adsorbed As into step or kink sites on the surface is given by (see Appendix A)

$$r_6 = k_{f6} \times \theta_{As} - k_{r6} \times \theta_{As}^{i} \times N_s^{i} \quad \theta \tag{3.19}$$

where $\alpha_{As}^{i}$ is the fraction of incorporation sites occupied by As and $N_s^{i}$ is the surface density of incorporation sites.

2. The rate at which the incorporated As atoms are covered by subsequently arriving Si atoms is given by

$$r_7 = g \times N_{As} \tag{3.20}$$

where g is the epitaxial growth rate, and $N_{As}$ is the As concentration in the epitaxial film. During epitaxial growth, the As concentration in the film is proportional to the density of As atoms occupying the incorporation sites, i.e.

$$As\text{-}s^{i} + s^{i} \rightleftharpoons As(ss) + s^{i} \quad K_7 = \frac{\alpha_{As}}{\left(\theta_{As}^{i} \times N_s^{i}\right)} = \frac{k_H \times N_{As}}{\left(\theta_{As}^{i} \times N_s^{i}\right)} \tag{3.21}$$

where $As\text{-}s^{i}$ represents an As atom occupying an incorporation site, $\alpha_{As}$ is the activity of As in the solid solution, and $k_H$ is Henry's constant. Equation (3.21) arises since, when a Si atom "buries" an incorporated As atom, a new incorporation site is generated. By using Eq. (3.21) in Eq. (3.20)

$$r_7 = g \times \theta_{As}^{i} \times N_s^{i} \times \frac{K_7}{k_H} \tag{3.22}$$

3. By substituting Eqs. (3.19) and (3.22) into Eq. (3.5), the following mass-balance equation results

$$\left(k_{f6} \times \theta_{As} - k_{r6} \times \theta_{As}^{i} \times N_s^{i} \times \theta\right) - \left(g \times \theta_{As}^{i} \times N_s^{i} \times \frac{K_7}{k_H}\right) = N_s^{i} \times \frac{d\theta_{As}^{i}}{dt} \tag{3.23}$$

where the expression to the right of the equality sign represents the rate of change of the surface density of incorporated As.

In this section rate equations for the mechanisms shown in Figs. 3.1 and 3.2 were derived and then used in the schematic mass-balance equations outlined in Section 3.2.3. The result is a set of seven, first order, linear differential equations containing the mathematical description of the doping process. As is shown in Sections 3.2.7 and 3.2.8, these equations can be used to explain both the steady-state dopant concentration in the epitaxial film as a function of gas-phase dopant molar fraction, epitaxial growth rate, and temperature, and also the transient response of the dopant system and its growth-rate dependence [3.4]. They can also be used to derive the "transfer function" of the dopant system (Section 3.2.9) [3.2,3.4].

In order to see the significance of the various mechanisms entering into the doping process, an equivalent electric circuit representing the doping process will be derived in Section 3.2.6.

### 3.2.5 Determination of $\theta$ and $\theta_H$

During epitaxial growth, the adsorption sites on the growing surface are occupied by several species: H, $SiH_4$, Si, $AsH_3$, $AsH_2$, and As. The fraction of adsorption sites that remains vacant is given by

$$\theta = 1 - \theta_H - \theta_{SiH_4} - \theta_{Si} - \theta_{AsH_3} - \theta_{AsH_2} - \theta_{As} \tag{3.24}$$

Under normal operating conditions in the atmospheric-pressure system

under consideration, the input partial pressures of the relevant species are [3.4] $P_{SiH_4} \sim 10^{-3}$ atm., $P_{AsH_3} \sim 10^{-10}$ atm., and $P_{H_2} \simeq 1$ atm. Therefore, it is not unreasonable to assume that the population of hydrogen in the adsorbed layer is much larger than the population of any other species, so that Eq. (3.24) can be approximated by $\theta \simeq 1 - \theta_H$. For the same reason, it can be assumed that the population of hydrogen in the adsorbed layer is in thermodynamic equilibrium with the hydrogen atmosphere, so that $\theta_H$ can be assumed to be independent of $P_{SiH_4}$ and $P_{AsH_3}$. (Note that these two assumptions do not hold for low-pressure systems.)

The desorption of hydrogen can be analyzed as follows

$$H\text{-}s \rightleftharpoons \frac{1}{2} H_2(g) + s \qquad K_8 = \frac{P_{H_2}^{\frac{1}{2}} \times (\theta \times N_s)}{(\theta_H \times N_s)} \tag{3.25}$$

Since $\theta_H$ is in thermodynamic equilibrium with $P_{H_2}$

$$\theta_H = \frac{P_{H_2}^{\frac{1}{2}} \times \theta}{K_8} = \frac{\theta}{K_8} \simeq \frac{1}{1 + K_8} \tag{3.26}$$

### 3.2.6 Circuit Representation of the Doping Process

In Section 3.2.4, a set of seven, first order, linear differential equations containing the mathematical description of the doping process was derived. One convenient way to gain insight into the relative importance of the processes represented by this set of equations is by studying an equivalent electric circuit represented by an analogous system of equations. An R-C circuit corresponding to the set of equations derived for the doping process

is shown in Fig. 3.3. Each "resistor" represents the impedance presented by one of the mechanisms entering into the doping process, and the "current" represents the net flow of As-containing species. The "voltage" which drives the current through the network of resistors represents the driving force for the doping process, which is supplied by the dopant molar fraction present in the main gas stream. The "capacitors" account for the accumulation of dopant species at various locations during transients.

In order to find the circuit representing the doping process from the system of equations derived in Section 3.2.4, each of the seven differential equations can be rearranged into an equation resembling one associated with an electric circuit. As an illustration, the circuit elements corresponding to Eq. (3.18) will be derived. The same method can be applied to each of the other six differential equations to obtain the circuit shown in Fig. 3.3.

Equation (3.18) is rewritten below in a more compact form

$$r_2 - r_3 - r_4 = \varepsilon \times N_T \times \frac{dX^s}{dt} \qquad (3.27)$$

where $r_2$, $r_3$, and $r_4$ are given by Eqs. (3.10), (3.14), and (3.16), respectively, and $X^s \equiv X^s_{AsH_3}$. The subscript "$AsH_3$" in this and the following equations are omitted for simplicity, and all unsubscripted quantities refer to $AsH_3$.

The expression for $r_2$ [Eq. (3.10)] does not require any rearrangment. In order to place the expressions for $r_3$ and $r_4$ in a useful form, all As-containing species must be expressed in terms of molar fractions X of $AsH_3$ analogous to voltages. From Eqs. (3.14) and (3.13)

$$r_3 = k_{f3} \times \left[ X^s - \frac{1}{K_3} \times X^s_{AsH_2} \right] \qquad (3.28)$$

which can be rewritten as

$$r_3 = k_{f3} \times \left[ X^s - (X)^e_{XAsH_2} \right] \tag{3.29}$$

with

$$(X)^e_{XAsH_2} \equiv \frac{1}{K_3} \times X^s_{AsH_2} \tag{3.30}$$

From Eq. (3.13), which relates the equilibrium constant of the reaction to the equilibrium molar fractions of $AsH_3$ and $AsH_2$, $(X)^e_{XAsH_2}$ is defined as the $AsH_3$ molar fraction that would be in equilibrium with $X^s_{AsH_2}$ if the reaction were in thermodynamic equilibrium. Although both $X^s$ and $(X)^e_{XAsH_2}$ refer to $AsH_3$, $X^s$ is the actual molar fraction of $AsH_3$ just above the interface, while $(X)^e_{XAsH_2}$ is obtained from Eq. (3.30) using the actual molar fraction of $AsH_2$. In general, $X^s \neq (X)^e_{XAsH_2}$; they become equal only when the dissociation reaction is at thermodynamic equilibrium. The term in brackets in Eq. (3.29) indicates how far the reaction is from equilibrium, and represents the driving force which forces the reaction back toward equilibrium.

The equation representing $r_4$ can be rearranged in an analogous manner. From Eq. (3.16) and $K_4 = k_{f4}/k_{r4}$

$$r_4 = k_{f4} \times \theta \times \left[ X^s - \frac{1}{K_4} \times \frac{\theta_{AsH_3}}{\theta} \right] = k_{f4} \times \theta \times \left[ X^s - (X)^e_{\theta AsH_3} \right] \tag{3.31}$$

with

$$(X)^e_{\theta AsH_3} \equiv \frac{1}{K_4} \times \frac{\theta_{AsH_3}}{\theta} \tag{3.32}$$

From Eqs. (3.15) and (3.32), $(X)^{e}_{\theta AsH_3}$ is defined as the $AsH_3$ molar fraction that would be in equilibrium with $\theta_{AsH_3}$ if the adsorption process were in thermodynamic equilibrium.

By substituting Eqs. (3.10), (3.29), and (3.31), into Eq. (3.27), the mass-balance of $AsH_3$ just above the gas-solid interface can be expressed as

$$k_m x\left(\overline{X} - X^s\right) - k_{f3} x\left(X^s - (X)^{e}_{XAsH_2}\right) - k_{f4} x \theta x\left(X^s - (X)^{e}_{\theta AsH_3}\right) = \epsilon x N_T \cdot x \frac{dX^s}{dt} \tag{3.33}$$

Equation (3.33) can also be written as

$$\frac{\overline{X} - X^s}{R_2} - \frac{X^s - (X)^{e}_{XAsH_2}}{R_3} - \frac{X^s - (X)^{e}_{\theta AsH_3}}{R_4} = C_2 x \frac{dX^s}{dt} \tag{3.34}$$

where $1/R_2 \equiv k_m$, $1/R_3 \equiv k_{f3}$, $1/R_4 \equiv k_{f4} x \theta$, and $C_2 \equiv \epsilon x N_T$. Equation (3.34) relates the mass-balance equation to an equivalent electric circuit with R's representing resistors, C's representing capacitors, and X's analogous to voltages, as shown in Fig. 3.4. Each resistor corresponds to one mechanism in the doping process, and each node in the circuit represents a physical location in the dopant system. The voltage at each node is equivalent to the gas-phase $AsH_3$ molar fraction which would exist in equilibrium with the As-containing species at that point. For example, the upper-right node corresponds to $AsH_3$ in the adsorbed layer, and the voltage at this node is analogous to the $AsH_3$ molar fraction in equilibrium with $\theta_{AsH_3}$. Current through a resistor is analogous to the net reaction rate of the step associated with that resistor. In thermodynamic equilibrium, $\overline{X} = X^s = (X)^{e}_{XAsH_2} = (X)^{e}_{\theta AsH_3}$, and no current flows through the resistors. The capacitors incorporate time-varying effect related to the accumulation of dopant species, e.g., in a transition from one steady-state

operating condition to another, they account for the time required for the population of the As-containing species at each point in the dopant system to attain its new steady-state value.

By applying the method above, each of the seven differential equations can be restated in a form associated with an equivalent circuit similar to that shown in Fig. 3.4. The resulting equations are summarized in Appendix B. Each equation is analogous to that describing one node of an equivalent circuit, and the elements surrounding each node may be assembled into one overall equivalent circuit as shown in Fig. 3.5. Since each node corresponds to one point in the doping process, it is associated with one As-containing species (c.f. Fig. 3.3); $(X)_i^e$ is the $AsH_3$ molar fraction corresponding to the As-containing species associated with the $i^{th}$ node.

Some of the analogies mentioned above between kinetic, thermodynamic, and circuit terms, may be easier to understand with the aid of Figs. 3.6 and 3.7, which refer to one mechanism (in this case, adsorption of $AsH_3$). The circuit representation of the forward reaction rate $r_{f4}$, reverse reaction rate $r_{r4}$ (desorption of $AsH_3$), and net reaction rate $r_4$ are shown in Figs. 3.6a, b, and c, respectively. Note the direction of the current and the applied voltage in each case. In Fig. 3.7, the length of the arrows is proportional to the corresponding reaction rate (i.e. current through the resistor). Three cases are shown: (a) $X^s = (X)^e_{\theta AsH_3}$; the forward reaction rate is equal to the reverse reaction rate and the net reaction rate is zero: equilibrium. (b) $X^s \gtrsim (X)^e_{\theta AsH_3}$; the forward reaction rate is slightly larger than the reverse reaction rate, and both are much larger than the net reaction rate: quasi-equilibrium. (c) $X^s >> (X)^e_{\theta AsH_3}$; the forward reaction rate is much larger than the reverse reaction rate, indicating that the resistor represents a kinetic limitation to the net reaction rate: kinetic control.

In the overall circuit of Fig. 3.5, the time constants $R_i \times C_i$ correspond to the relaxation time of the $i^{th}$ step in the doping process. For example, $R_4 \times C_4$ corresponds to the relaxation time of the adsorption of $AsH_3$ on the growing surface. Table 3.1 shows the physiochemical mechanism corresponding to each resistor and the location in the epitaxial system corresponding to each capacitor.

Under steady-state conditions (i.e. for a constant $X^i$), the right-hand sides of Eq. (3.34) and the corresponding equations for the other nodes (Appendix B) are equal to zero; hence the capacitors disappear from the circuit, and the equivalent circuit reduces to the steady-state representation shown in Fig. 3.8.

The current $r_7$ in Figs. 3.5 and 3.8 represents the actual flow of As atoms into the growing epitaxial film. In the absence of epitaxial growth (i.e when $g = 0$), $R_7 = k_H/(g \times K_p) = \infty$ and no current flows in the circuit for a constant $X^i$. The voltage $(X)^e_{\theta^i As}$ is then equal to $X^i$, which indicates that the system is in equilibrium; i.e. the As population at the surface is in equilibrium with the $AsH_3$ present in the gas phase. On the other hand, during epitaxial growth $(X)^e_{\theta^i As}$ falls below $X^i$, providing the driving force for the flow of net dopant atoms into the growing film. This "current" is limited by the largest resistor in the sequence, and the doping process is then controlled by the reaction kinetics.

### 3.2.7 Steady-State

The term "steady-state" will be used to describe the condition for which the input dopant molar fraction has remained unchanged for at least several decay times of the dopant system [3.2]. For this case the equivalent circuit of Fig. 3.5 reduces to that shown in Fig. 3.8. Resistors $R_1$-$R_6$

represent the kinetic limitation presented by the corresponding physiochemical mechanism, and the capacitors incorporate time-varying effects related to the accumulation of dopant species at various locations in the epitaxial system. In the equivalent circuit of Fig. 3.8, force-convection mass transport is represented by $R_1$, boundary-layer mass transport by $R_2$, gas-phase chemical reactions by $R_3$, adsorption of $AsH_3$ ($AsH_2$) on the growing surface by $R_4$ ($R_{4'}$), surface chemical dissociation of $AsH_3$ ($AsH_2$ by $R_5$ ($R_{5'}$), surface diffusion and site incorporation of arsenic atoms by $R_6$, and the burying of arsenic atoms at incorporation sites by silicon atoms during epitaxial growth of $R_7$. Each node in the circuit represents a physical location in the dopant system, and the voltage at each node is equivalent to the gas-phase $AsH_3$ molar fraction which would exist in equilibrium with the arsenic-containing species at that point.

This circuit indicates the two parallel paths for the flow of arsenic-containing species towards the growing film which arise because of $AsH_3$ dissociation above the gas-solid interface. The $AsH_3$ path ($R_4$ and $R_5$) represents $AsH_3$ adsorption on the surface and decomposition in the adsorbed layer, while the $AsH_2$ path ($R_3$, $R_{4'}$, $R_{5'}$) describes gas-phase dissociation of $AsH_3$ (primarily into $AsH_2$), adsorption of $AsH_2$ on the surface, and further decomposition in the adsorbed layer. Both branches contribute to the flow of arsenic atoms into incorporation sites (current through $R_6$). The largest contribution to the total flow is provided by the path with the lowest resistance. To simplify the analysis the $AsH_2$ path is assumed to be dominant (Fig. 3.9) because it contains all of the possible mechanisms involved in the doping process, while the resistor $R_3$, representing the gas-phase dissociation of $AsH_3$, is absent from the $AsH_3$ path. Analogous results can be obtained for the more general case,

which includes both the $AsH_2$ path and the simplier $AsH_3$ path.

From Fig. 3.10 the dopant concentration $N_{As}$ in the epitaxial film can be related to the input $AsH_3$ molar fraction $X^i$ by using common circuit techniques

$$(X)^e_{\theta^i_{As}} = \frac{R_7}{R + R_7} \times X^i \tag{3.35}$$

where

$$R \equiv R_1 + R_2 + R_3 + R_4, + R_5, + R_6 \tag{3.36}$$

describes the overall limitation imposed by the kinetic sequence, steps 1-6 in the doping process [3.16]. Equation (3.35) relates the steady-state input $AsH_3$ *molar fraction* $X^i$ *to the gas-phase* $AsH_3$ *molar fraction* $(X)^e_{\theta^i_{As}}$ that would be in equilibrium with the arsenic atoms occupying the incorporation sites. It is obvious from Eq. (3.35) that $X^i \geq (X)^e_{\theta^i_{As}}$ in steady-state. From Eq. (B.12)

$$(X)^e_{\theta^i_{As}} = \frac{k_H \times N_{As}}{K_P} \tag{3.37}$$

Equation (3.37) arises because $(X)^e_{\theta^i_{As}}$ is proportional to the population of arsenic atoms occupying incorporation sites $(N^i_s \times \theta^i_{As})$, which in turn is proportional to $N_{As}$. Therefore, to obtain $N_{As}$ it is only necessary to calculate the "voltage" $(X)^e_{\theta^i_{As}}$. From Eqs. (3.35) and (3.37),

$$N_{As} = \frac{R_7}{R + R_7} \times \frac{K_P}{k_H} \times X^i \tag{3.38}$$

which relates the dopant concentration $N_{As}$ in the epitaxial film to the input

arsine molar fraction $X^i$ through kinetic and thermodynamic parameters. Since $R_7$ is a function of the epitaxial growth rate [Eq. (B.21)] while R is independent of it, two limiting regions of operation can be described depending on the growth rate.

1. Low growth-rate region. At very low growth rates $R_7 = k_H/(g \times K_p) \gg R$, and Eq. (3.38) reduces to

$$N_{As} \simeq \frac{K_p}{k_H} \times X^i \tag{3.39}$$

Equation (3.39) indicates that at very low growth rates, the dopant concentration in the epitaxial film is independent of the growth rate and is controlled by the thermodynamics of the overall doping reaction [see Eq. (B.29)]. From Eq. (3.35) $(X)^e_{\theta^i_{As}} \simeq X^i$, implying that very little current flows in the circuit of Fig. 3.9, again emphasizing the importance of thermodynamics. The lack of dependence of dopant concentration on growth rate is consistent with the experimental data presented in Fig. 3.10 [3.4], which shows that at growth rates lower than about 0.5 μm/min the arsenic concentration in the epitaxial film tends to become constant, independent of the growth rate. The value of $K_p/k_H$ can be obtained by extrapolating the results shown in Fig. 3.10 to zero growth rate to get $N_{As}$ at thermodynamic equilibrium. This number is then used in Eq. (3.39) to obtain $K_p/k_H \simeq 3.8 \times 10^{25}$ cm$^{-3}$ for $X^i = 6.2 \times 10^{-10}$ and T = 1050°C in the $SiH_4/AsH_3$ epitaxial system studied.

2. High growth-rate region. At high growth rates $R_7 = k_H/(g \times K_p) \ll R$, and Eq. (3.38) reduces to

$$N_{As} \simeq \frac{R_7}{R} \times \frac{K_p}{k_H} \times X^i \tag{3.40}$$

By substituting the expression for $R_7$ into Eq. (3.40),

$$N_{As} \simeq \frac{1}{g \times R} \times X^{i} \tag{3.41}$$

Equation (3.41) indicates that at high growth rates the doping density in the epitaxial film is inversely proportional to the epitaxial growth rate. This agrees with the results presented in Fig. 3.10, which show that at growth rates higher than approximately 0.5 μm/min the arsenic concentration in the film is proportional to the reciprocal of the growth rate. From Eq. (3.35) $(X)^{e}_{\theta^{i}As} \ll X^{i}$, and the current in the circuit of Fig. 3.9 is limited by the chain of resistors representing steps 1-6 of the doping process; i.e. the doping process is controlled by the reaction kinetics. If one of the resistors forming R is much larger than the others, the current flowing in the circuit for a given $X^{i}$ will be controlled mainly by that resistor, and most of the "voltage" available will be dropped across it. The kinetic process represented by the resistor will then be the rate-limiting step. There will be little voltage drop across any of the other resistors, indicating that the other steps in the doping process are near equilibrium. In this case the forward reaction rate of the other steps is only slightly larger than the reverse reaction rate, both being much larger than the net reaction rate. The value of R can be obtained by fitting Eq. (3.41) to the data of Fig. 3.10 to find that $R = 1.9 \times 10^{-19}$ cm$^2$·sec for $X^{i} = 6.2 \times 10^{-10}$ and T = 1050°C.

After some mathematical manipulation, Eq. (3.38) can be written as

$$N_{As} = \frac{R^{-1}}{g + \left[R\,(K_p/k_H)\right]^{-1}} \times X^{i} \tag{3.42a}$$

and after substituting the values found above, the equation describing the data of Fig. 3.10 is obtained

$$N_{As} = \frac{5.2 \times 10^{18}}{g + 1.4 \times 10^{-7}} \times X^{i} \text{ cm}^{-3} \qquad (3.42b)$$

with g in cm/sec.

A. Potential and Reaction Rates in the Doping Process

Figure 3.11 shows schematically the potential available to drive the dopant species towards the growing film at low and high growth rates. At very low growth rates (Fig. 3.11a) little potential is dropped across the kinetic chain because the rate at which silicon atoms bury arsenic atoms during epitaxial growth is too low to cause significant departure from equilibrium. At very high growth rates, however, the arsenic atoms are covered rapidly, equilibrium no longer exists, and almost all the available potential will be dropped across the slowest step in the sequence. As an example, Fig. 3.11b depicts the situation assuming $AsH_2$ adsorption on the growing surface to be the rate-limiting step, with all other kinetic steps near equilibrium. Figures 3.12a and 3.12b show the forward and reverse reaction rates $r_f$ and $r_r$ for each step in the doping process corresponding to the potential diagrams in Figs. 3.11a and 3.11b, respectively. The length of each arrow is proportional to the corresponding reaction rate. In Fig. 3.12a the forward and reverse reaction rates of each kinetic step are both much larger than the net incorporation rate $r_7$. In Fig. 3.12b this is no longer true for the rate limiting step $r_{4'}$, in which the forward reaction rate is much larger than the reverse reaction rate and, therefore, almost equal to the net incorporation rate. As explained in Section 3.2.2, the rate $r_7$ at which silicon atoms bury arsenic atoms during epitaxial growth

determines whether the doping process is near thermodynamic equilibrium (Fig. 3.12a) or is controlled by the kinetics of a rate-limiting step (Fig. 3.12b). In this example, the dopant incorporation rate $r_7$ cannot exceed $r_{f4'}$. It can be anywhere between zero and $r_{f4'}$, depending on the growth rate.

It is important to distinguish the physical nature of the chain of resistors $R_1$-$R_6$ from that of resistor $R_7$ in the circuit of Fig. 3.9. The arsenic-containing species play the active role in steps 1-6 of the doping process, while they play a passive role in step 7. In this step the active role is played by the silicon atoms which cover arsenic atoms occupying incorporation sites. Resistors $R_1$-$R_6$ represent kinetic limitations to the flow of arsenic-containing species toward incorporation sites, while resistor $R_7$ represents a thermodynamic limitation. Therefore, the "potential" drop across $R_7$ is not of the same nature as the potential drop across any of the resistors $R_1$-$R_6$. The latter potential drop is that required to drive the dopant species across the kinetic limitation represented by the resistors $R_1$-$R_6$. The former is a thermodynamic potential, and provides the "reference" potential $(X)^e_{\theta^i As}$. As indicated by Eq. (3.35), a slow burying rate (low growth rate) provides a high reference potential, and little potential is left to drive the dopant species toward incorporation sites: $X^i - (X)^e_{\theta^i As} \approx 0$. On the other hand, a fast burying rate (high growth rate) reduces the reference potential, and almost all the potential provided by $X^i$ will be available to drive the dopant species toward incorporation sites: $X^i - (X)^e_{\theta^i As} \approx X^i$. This distinction between the "kinetic" resistors $R_1$-$R_6$ and the "thermodynamic" resistor $R_7$ will be particularly useful in the discussion of the transient behavior of the dopant system in Section 3.2.8.

## B. Rate-Limiting Step in the High Growth-Rate Region

It is important to determine whether the rate-limiting step in the high growth-rate region is a gas-phase or a surface mechanism. In [3.4] the effect of the deposition temperature on the epitaxial-layer arsenic concentration was obtained experimentally and briefly analyzed in order to differentiate between the possible mechanisms. It was shown in Fig. 10 of that paper that the arsenic concentration in the epitaxial film decreases when the deposition temperature increases, showing an apparent enthalpy change of approximately 30 kcal/mole. In the following discussion, expressions for the epitaxial-layer arsenic concentration will be derived for the different possible mechanisms that may control the doping process; these expressions will then be compared with the experimental temperature dependence of dopant concentration described above. It will be shown that the process limiting the arsenic incorporation rate is a surface, rather than a gas-phase, mechanism.

1. Gas-phase mechanisms. Three processes are considered [3.16]: forced-convection mass transport, boundary-layer mass transport, and gas-phase chemical reactions.

a. Force-convection mass transport is represented by $R_1$ in the circuit of Fig. 3.9. By substituting the expression for $R_1$ [Eq. (B.13)] in Eq. (3.41)

$$N_{As} \simeq \frac{N_{300} \times v_{H_2}}{g} \times X^1 \tag{3.43}$$

where $N_{300}$ is the total gas-phase molecular concentration at 300 K and $v_{H_2}$ is the hydrogen velocity at 300 K. Under normal operating conditions [3.4], the epitaxial deposition of silicon from $SiH_4$ is controlled by mass transport through

the boundary layer. Since the boundary-layer mass-transport coefficient usually increases slowly with temperature, showing an apparent activation energy of 3-8 kcal/mole [3.16], the epitaxial growth rate g does also. Therefore Eq. (3.43) indicates that if forced-convection mass transport, which is independent of temperature, dominated the doping process, the doping density in the epitaxial film would decrease with increasing temperature with an apparent activation energy of 3-8 kcal/mole. This does not agree with the experimental observation [3.4], indicating that forced-convection mass transport is not the dominant mechanism. The same conclusion can be reached by comparing the values of the overall kinetic resistor R with the forced-convection mass-transport resistor $R_1$. From Eq. (B.13) and $N_{300} = P/[k \times (300\ K)]$ [3.17],

$$R_1 = \frac{k \times (300\ K)}{P \times v_{H_2}} \tag{3.44}$$

where P is the total pressure and k is the Boltzmann constant. With P = 1 atm, $v_{H_2}$ = 21.3 cm/sec, and $k = 1.37 \times 10^{-22}$ cm$^3$·atm/K, $R_1 = 1.93 \times 10^{-21}$ cm$^2$·sec, which is much smaller than the value of $R = 1.9 \times 10^{-19}$ cm$^2$·sec obtained earlier.

b. Boundary-layer mass transport is represented by $R_2$ in the circuit of Fig. 3.9. By using the expression for $R_2$ [Eq. (B.14)] in Eq. (3.41),

$$N_{As} \simeq \frac{k_m}{g} x\ X^i \tag{3.45}$$

where $k_m$ is the mass-transport coefficient of $AsH_3$ through the boundary layer. As discussed above, the deposition of silicon from $SiH_4$ is normally controlled by mass transport through the boundary layer, and the mass transport coefficients

of $AsH_3$ and $SiH_4$ are expected to have a similar temperature dependence. Therefore, the ratio $k_m/g$ will remain relatively constant with temperature, and if boundary-layer mass-transport dominated the doping process, the dopant concentration in the epitaxial film would not change significantly with varying deposition temperature. Since this disagrees with the experimental results, boundary-layer mass transport cannot be the dominant mechanism either. Again, the same conclusion can be reached by estimating the value of $R_2$ and comparing it to R. Since the deposition of silicon is controlled by mass transport through the boundary layer, an expression similar to Eq. (3.45) describes the deposition process

$$N_{Si} \simeq \frac{k_m'}{g} \times X^i_{SiH_4} \tag{3.46}$$

where $N_{Si}$ is the density of silicon atoms in the solid, $k_m'$ is the mass transport coefficient of $SiH_4$ through the boundary layer, and $X^i_{SiH_4}$ is the input gas-phase molar fraction of $SiH_4$. The value of $k_m'$ obtained from Eq. (3.46) should be close to the value of $k_m$ in Eq. (3.45) [3.6,3.12]. With $N_{Si} \doteq 5 \times 10^{22}$ cm$^{-3}$, $g = 0.6$ μm/min, and $X^i_{SiH_4} = 1.1 \times 10^{-3}$ atm in Eq. (3.46), $k_m' = 4.6 \times 10^{19}$/cm$^2$·sec $\simeq k_m$; using this value to find $R_2 = 1/k_m$, results in $R_2 = 2.2 \times 10^{-20}$ cm$^2$·sec. Clearly, $R \gg R_2$, confirming that boundary-layer mass transport is not the dominant mechanism. (Also note that $R_2 \gg R_1$.)

c. Gas-phase chemical reactions are represented by $R_3$. By substituting the expression for $R_3$ [Eq. (B.15)] in Eq. (3.41),

$$N_{As} \simeq \frac{k_{f3}}{g} \times X^1 \tag{3.47}$$

where

$$k_{f3} = a \times \exp(-E_a/RT) \tag{3.48}$$

is the corresponding forward reaction rate constant. Since activation energies are usually larger than 10 kcal/mole [3.18], Eq. (3.47) indicates that, if gas-phase chemical reactions were dominant, the doping density in the epitaxial film would increase with increasing temperature. The Arrhenius activation energy would be slightly smaller than $E_a$ since g(T) increases slightly with increasing temperature. This behavior disagrees with the experimental results, ruling out gas-phase chemical reactions as the dominant mechanism in the doping process.

From the discussion above, it can be concluded that none of the gas-phase mechanisms dominate the arsenic incorporate rate. The surface mechanisms are analyzed next.

2. Surface mechanisms. If a surface mechanism dominates the doping process, either adsorption, surface chemical reactions, or surface diffusion and site incorporation may be the rate-limiting step. Since the temperature dependence is similar to all three cases, only one mechanism will be considered. It is important to emphasize, however, that similar conclusions will be obtained for the other two mechanisms. In this analysis, surface diffusion and incorporation of adsorbed arsenic atoms, represented by $R_6$ in the circuit of Fig. 3.9, will be assumed to be the rate-limiting step. By substituting the expression for $R_6$ [Eq. (B.20)] in Eq. (3.41),

$$N_{As} \simeq \frac{k_{f6} \times K_{5'} \times K_{4'} \times K_3}{g} \times \frac{\theta^3}{\theta_H^2} \times X^i \tag{3.49}$$

where $\theta$ is the fraction of adsorption sites which are vacant, and $\theta_H$ is the fraction of adsorption sites occupied by hydrogen atoms. $K_3$, $K_{4'}$, and $K_{5'}$ are the equilibrium constants corresponding to steps 3, 4, and 5, respectively, in the doping sequence, and the term $\theta^3/\theta_H^2$ is a function of $K_8$, the equilibrium constant of the hydrogen-desorption process.

The temperature dependence of the dopant concentration in Eq. (3.49) is more complex than that of the gas-phase cases. The temperature dependence of $k_{f6}$ is given by an expression similar to Eq. (3.48), while the temperature dependence of each equilibrium constant is obtained by integrating van't Hoff's equation [3.19]

$$K = c \ x \ exp(-\Delta H/RT) \tag{3.50}$$

where $\Delta H$ is the heat of reaction at constant pressure, and c is a constant. Consequently, the temperature dependence of Eq. (3.49) depends on the endothermic or exothermic nature of the reactions involved and on the relative magnitudes of $E_a$ and $\Delta H$. Adsorption and site incorporation are exothermic mechanisms by nature, as is the dissociation of $AsH_3$ [3.20]. Therefore, $K_3$, $K_{4'}$, and $K_{5'}$ in Eq. (3.49) all decrease with increasing temperature and, unless $E_a$ is larger than $|\Sigma\Delta H_1|$, the doping density would decrease with increasing temperature if surface diffusion and site incorporation were dominant. A similar dependence is obtained for the other two surface mechanisms. Since this behavior agrees with experimental observation, it can be concluded that (a) $|\Sigma\Delta H_1| > E_a$, and (b) surface mechanisms dominate the doping process in the epitaxial system studied.

Table 3.2 summarizes the different expressions for $N_{As}$ in the high growth-rate region obtained for each controlling mechanism and the temperature dependence expected in each case.

C. Crystallographic Orientation

To further confirm the conclusions reached above, the influence of the crystallographic orientation of the substrate on the doping process was analyzed. As discussed above, adsorption, surface chemical reaction, surface migration, and site incorporation occur on the growing surface and, therefore, their kinetics depend on the nature of the surface. Consequently, if the doping process is controlled by surface mechanisms, it must be affected by crystallographic orientation.

The horizontal RF-heated epitaxial reactor used in these experiments was a commercial Unipak VI (Sola Basic-Tempress) that operates at atmospheric pressure. Substrates of two orientations [(111) $\pm$0.5° and (100)] were alternately placed on the susceptor (see Fig. 3.13). The arsine partial pressure was held at approximately $5 \times 10^{-10}$ atm, and the silane partial pressure was adjusted to produce growth rates of approximately 0.3 or 0.4 μm/min in different experiments. The substrates were boron-doped with resistivities ranging from 1 to 4 Ω-cm. The thicknesses of the epitaxial layers were approximately 7 um and were measured by a groove and stain technique. A four-point probe determined the resistivity of the layers. The results, summarized in Table 3.3, indicate that epitaxial layers grown on (111) substrates have a higher resistivity than those grown on (100) substrates, and this confirms conclusively that surface mechanisms dominate the doping process. On the other hand, the epitaxial growth rate is unaffected by the crystallographic orientation, consistent with mass transport control of the deposition process.

D. Summary

In Section 3.2.7 the equivalent circuit shown in Fig. 3.9 was used to analyze the model described in Section 3.2.6 under steady-state conditions. The analysis indicated the presence of two regions of operation depending on the growth rate, in agreement with experimental data. At very low growth rates the doping process is controlled by the thermodynamics of the overall doping reaction, and the doping density in the epitaxial layer is independent of the growth rate. At high growth rates the doping process is controlled by reaction kinetics, and the doping density is inversely proportional to the growth rate. Values for R and $K_p/k_H$ were obtained by fitting equations derived from the model to the experimental data. Expressions were derived for the arsenic concentration expected in the epitaxial film for different possible gas-phase and surface rate-limiting mechanisms. By comparing these expressions with the experimental temperature dependence of the epitaxial-layer arsenic concentration in the high growth-rate region [3.4]; it was concluded that the rate-limiting step is associated with a surface mechanism rather than with the gas phase. This was further confirmed by the observation that the crystallographic orientation of the substrate influences the doping process.

3.2.8 Transient Behavior

The doping process under time-varying conditions can be analyzed using the circuit shown in Fig. 3.14a, again assuming the more complex $AsH_2$ path to be dominant. This circuit is similar to that shown in Fig. 3.9, with the corresponding capacitors added to account for transient effects. In this section some considerations which simplify the circuit will be examined first, and then the simplified circuit will be analyzed.

As mentioned earlier, when one of the resistors in the chain is much larger than the others, most of the voltage is dropped across that resistor,

and the voltage drop across any other resistor will be relatively small. If resistor $R_{4'}$ is assumed to be the largest resistor, the circuit of Fig. 3.14a can be simplified to that shown in Fig. 3.14b. Since most of the voltage is dropped across $R_{4'}$, the nodes to the right of $R_{4'}$ are at approximately the same potential, and consequently the capacitors to the right of $R_{4'}$ can be lumped together; similarly for the nodes and capacitors to the left of $R_{4'}$. In Fig. 3.14b, $C_s$ represents capacitive effects associated with the adsorbed layer and incorporation sites on the growing surface, and $C_g$ represents capacitive effects associated with the gas phase (see Table 3.1).

An estimate of the values or $R_1 \times C_1$ and $R_2 \times C_2$ can be obtained as follows. From Eqs. (B.13) and (B.22) and $N_T = P/kT$,

$$R_1 x C_1 = \frac{L \times 300K}{v_{H_2} \times T} \tag{3.51}$$

where L is the length of the susceptor and T is the temperature in the main gas stream within the deposition region. With L = 22.8 cm, $v_{H_2}$ = 21.3 cm/sec, and assuming T = 873 K [3.12], $R_1 \times C_1$ = 0.37 sec. Similarly, from Eqs. (B.14) and (B.23) and $N_{T'} = P/k \times T'$,

$$R_2 C_2 = \frac{\varepsilon \times P}{k_m \times k \times T'} \tag{3.52}$$

where $\varepsilon$ is the thickness of a thin layer at temperature T' just above the gas-solid interface. With $\varepsilon$ = 0.1 cm, P = 1 atm, $k_m = 4.6 \times 10^{19}/cm^2 \cdot sec$, k = $1.37 \times 10^{-22}$ $cm^3$ x atm/K, and T' - 1273 K, $R_2 \times C_2$ = 0.013 sec. $R_1 \times C_1$ and $R_2 \times C_2$ are decay times associated with mass transport mechanisms, and are much smaller than the experimental decay times reported in [3.4] (Fig. 3.15).

Therefore, if any of the time constants resulting from elements to the left of $R_4$, dominates the experimental results, it must be $R_3 \times C_3$.

A circuit analogous to that shown in Fig. 3.14b may be obtained if any of the other resistors in the kinetic chain, $R_1$-$R_6$, were largest. If the largest resistor were either $R_5$, or $R_6$, $C_g$ would represent capacitive effects associated with both the gas phase and the adsorbed layer, while if $R_1$, $R_2$, or $R_3$ dominated, $C_s$ would include contributions from both regions. Since the analysis is similar in each case, only the circuit shown in Fig. 3.14b will be considered. It is important to emphasize, however, that similar conclusions will be obtained if any one of the other kinetic resistors is dominant.

The transient study presented in [3.4] showed that the transient response of the doping process was basically exponential with a single decay time, i.e. the $AsH_3$ dopant system contains one dominant pole. Therefore, either $C_g$ or $C_s$ must be responsible for this exponential. From Fig. 3.14b the corresponding decay times are given by

$$\tau_g \simeq C_g \times R_g \tag{3.53}$$

and

$$\tau_s \simeq C_s \times \frac{R_7 \times R_4'}{R_7 + R_4'} \tag{3.54}$$

$\tau_g$ is independent of the growth rate and is the decay time associated with the gas phase, while $\tau_s$ depends on the epitaxial growth rate through $R_7$ and is associated with mechanisms occurring at the growing surface. It was reported in [3.4] that the experimental decay time depends strongly on the epitaxial growth rate. Therefore, $\tau_g$, which is independent of growth rate, cannot be responsible for the exponential in the transient response. Gas-phase processes

do not control the transient behavior; and surface mechanisms must be dominant, in agreement with the conclusions reached independently in Section 3.2.7 from the steady-state behavior.

To continue the analysis, Eq. (3.54), which describes surface mechanisms, must be compared to the experimental results. Two limiting regions of growth rate can be considered.

1. Low growth-rate region. At very low growth rates $R_7 = k_H/(g \times K_p) >> R_{4'}$, and Eq. (3.54) reduces to

$$\tau_s \simeq C_s \times R_{4'} \tag{3.55}$$

Equation (3.55) indicates that, at very low epitaxial growth rates, $\tau_s$ is independent of the rate. This agrees with the experimental data presented in Fig. 3.15 [3.4], which shows that at rates lower than approximately 0.5 μm/min the dependence of the decay time on growth rate decreases.

2. High growth-rate region. At high growth rates $R_7 = k_H/(g \times K_p) << R_{4'}$, and Eq. (3.54) reduces to

$$\tau_s \simeq C_s \times R_7 = C_s \times \frac{k_H}{g \times K_p} \tag{3.56}$$

Equation (3.56) indicates that, at high rates, $\tau_s$ is inversely proportional to the growth rate, also in agreement with the experimental data in Fig. 3.15.

The value of $C_s$ can be obtained by using either Eq. (3.55) or (3.56) in conjunction with the data in Fig. 3.15 and the values calculated in Section 3.2.7. By fitting Eq. (3.56) to the data for $g > 0.5$ μm/min, a value of $C_s/(K_p/k_H) = 4.1 \times 10^{-5}$ cm is obtained. $C_s$ is now found by substituting in

this expression the value of $K_p/k_H = 3.8 \times 10^{25}$ cm$^{-3}$ obtained in Section 3.2.7; the resulting value of $C_s$ is $1.6 \times 10^{21}$ cm$^{-2}$. The same value is found when Eq. (3.55) is used.

After some mathematical manipulation, Eq. (3.54) can be rearranged as

$$\tau_s = \frac{C_s \times \left[K_p/k_H\right]^{-1}}{g + \left[R_4, \times (K_p/k_H)\right]^{-1}} \qquad (3.57a)$$

and after substituting the proper values into Eq. (3.57a), the equation describing the data in Fig. 3.15 is obtained

$$\tau_s = \frac{4.1 \times 10^{-5}}{g + 1.4 \times 10^{-7}} \text{ sec} \qquad (3.57b)$$

*with g given in cm/sec.* Note that the denominators or Eqs. (3.53) and (3.57) are the same.

In both the transient and steady-state cases, two regions of operation were seen. The two regions arise from the interrelation between the rate-limiting step and the step corresponding to the covering of the incorporated arsenic atoms (step 7). The following discussion emphasizes the importance of this interrelation to the transient behavior of the dopant system.

At low growth rates, Eq. (3.55) indicates that the decay time of the transient response is related to the relaxation time of the slowest step in the kinetic sequence. This is not surprising, since in this region the covering of arsenic atoms occupying incorporation sites by silicon atoms occurs so slowly that it only weakly influences the transient response. For example, if the dopant gas $X^1$ is abruptly decreased, the concentrations of arsenic-containing species in the dopant system, represented by charge on $C_g$ and $C_s$, will have to

decrease until they reach new steady-state values. The dopant species on $C_g$ will be quickly removed through $R_g$, since $R_g << R_{4'}$ at all growth rates. On the other hand, the dopant species on $C_s$, representing arsenic-containing species in the adsorbed layer and at incorporation sites, will be removed either by being buried in the growing epitaxial film (current through $R_7$) or by reaching the gas phase. This latter process is represented by current flowing backwards through $R_{4'}$. At low growth rates the covering rate is so slow that the removal of arsenic-containing species from the surface will be preferentially directed towards the gas phase [$R_7 = k_H/(g \times K_p) >> R_{4'}$]. The transient response will be characterized by Eq. (3.55) and will be independent of growth rate since the covering of arsenic atoms by silicon atoms will have little influence. On the other hand, at high growth rates arsenic atoms are rapidly covered by silicon atoms, and the removal of the arsenic-containing species from the adsorbed layer occurs much faster by this covering step than by flowing backwards through the kinetic chain [$R_7 = k_H/(g \times K_p) << R_{4'}$]. Therefore, in this high growth-rate region, the transient response will be controlled by the rate at which silicon atoms cover arsenic atoms. The higher the growth rate, the shorter the decay time, as expressed in Eq. (3.56).

Note that if $X^i$ is abruptly decreased in the low growth-rate region, most of the dopant species on $C_s$ will flow away from the growing layer during the transient, while if it occurs in the high growth-rate region, most of the dopant species will be incorporated into the growing film. Therefore, the spatial transition between the two steady-state doping levels in the epitaxial layer may be expected to be more abrupt at low growth rates, in agreement with the experimentally measured doping profiles [3.4].

In this section, the growth-rate dependence of the experimental

decay time was used to show that surface mechanisms, rather than gas-phase mechanisms, are responsible for the transient behavior of the doping process. At very low growth rates the transient response is controlled by the relaxation time of the slowest step in the kinetic sequence. At high growth rates the transient response is controlled by the epitaxial growth rate, which determines the rate at which arsenic atoms are covered. In both cases the surface capacitors play an important role. They are associated with the adsorbed layer and, in a transition from one steady-state condition to another, account for the time taken by the dopant species in the adsorbed layer to reach a new steady-state value. The value of $C_s$ was calculated by fitting equations derived from the model to the experimental data.

3.2.9 Transfer Function

The equivalent circuit shown in Fig. 3.14b can also be used to derive the "transfer function" H(s) of the dopant system, which relates the time variation of the doping gas concentration to the resulting epitaxial-layer dopant distribution [3.4]. [Recall that $N_{As}(t)$ is related to $N_{As}(x)$ by means of the epitaxial growth rate $g = x/t$.] By using Eq. (3.37)

$$H(s) \equiv \frac{N_{As}(s)}{X^i(s)} = \frac{K_p}{k_H} \times \frac{(X)_\theta^e i_{As}(s)}{X^i(s)} \tag{3.58}$$

where s is the frequency-domain variable. From the circuit of Fig. 3.14b (neglecting the effect of $\tau_g = C_g \times R_g$),

$$H(s) = \frac{K_p}{k_H} \times \frac{R_7}{R_7 + R_4} \times \frac{1}{1 + s\tau_s} \tag{3.59}$$

where $\tau_s$ is given by Eq. (3.54). Equation (3.59) has the same form as the

expression for the transfer function obtained experimentally in [3.4]. By substituting the values obtained in Sections 3.2.7 and 3.2.8 into Eq. (3.59) and appropriate mathematical manipulation,

$$H(s) = \frac{5.2 \times 10^{18}}{g + 1.4 \times 10^{-7}} \quad \frac{1}{1 + s \times \frac{4.1 \times 10^{-5}}{g + 1.4 \times 10^{-7}}} \tag{3.60}$$

with g in cm/sec. The transfer function is useful for predicting the spatial variation of the dopant profile expected for a given time-varying dopant gas flow, as was shown in [3.4], or alternatively for determining the dopant gas flow needed to obtain a desired dopant profile.

3.2.10 Summary and Conclusions

A physiochemical model describing the incorporation of dopant atoms into silicon epitaxial films during deposition from a $SiH_4$-$AsH_3$-$H_2$ mixture in a horizontal, atmospheric-pressure, epitaxial reactor was presented. The model considers sequential processes occurring both in the gas phase are (a) mass transport of $AsH_3$ to the deposition region, (b) mass transport of $AsH_3$ across the boundary layer to the growing surface, and (c) gas-phase chemical reactions. The sequence of steps occurring at the surface are (d) adsorption of the As-containing species on the growing surface, (e) chemical dissociation in the adsorbed layer, (f) surface diffusion and incorporation of As at incorporation sites on the surface, (g) burying of the incorporated As atoms by Si atoms during epitaxial growth, and (h) desorption of hydrogen from the surface. Special consideration was given to step 7, since it provides the driving force for the overall doping process, and also determines whether the doping process occurs under quasi-equilibrium conditions (slow burying rate), or is controlled by the reaction kinetics (fast burying rate).

In order to properly describe the doping process under both transient and steady-state conditions, mass-balance of the As-containing species was applied at each of the important points in the dopant system, resulting in seven, first-order linear differential equations containing the mathematical description of the doping process. In order to conveniently handle this set of equations, an equivalent electric circuit represented by an analogous set of equations was found. In this RC circuit, each resistor represents the impedance presented by one of the mechanisms entering into the doping process; each node represents a physical location in the dopant system; each capacitor incorporates time-varying effects related to the accumulation of dopant species at a point in the epitaxial system; current represents the net flow of As-containing species through the chain of steps; and the voltage represents the driving force for the doping process. The capacitors are required only during time-dependent conditions; e.g., in a transition from one steady-state operation condition to another, they account for the time required for the population of the As-containing species at each point in the dopant system to attain its new steady-state value. This circuit representation provides insight into the different mechanisms taking part in the doping process and their relative importance.

The equivalent circuit was then used to study both the steady-state and transient behavior of the epitaxial doping process. The model shows the presence of two regions of operation depending on the growth rate, in agreement with experimental data [3.4]. The two regions arise from the interrelation between the rate-limiting step of dopant movement, which is independent of the growth rate, and the covering of the incorporated arsenic atoms, which is growth-rate dependent. Both the steady-state and the transient analysis show, independently, that the mechanisms dominating the doping process occur at the

growing surface.

In steady-state the model indicates that the doping process is controlled by the thermodynamics of the overall doping reaction at very low growth rates, and the doping density in the epitaxial layer is independent of the growth rate. At high growth rates the doping process is controlled by reaction kinetics, being rate limited by one or more of the individual steps in the doping process, and the doping density is inversely proportional to the growth rate. The decrease observed in the epitaxial-layer arsenic concentration with increasing deposition temperature [3.4] was used to show that mechanisms occurring at the growing surface (adsorption, surface chemical dissociation, or site incorporation) dominate the doping process. This was further confirmed by the observation that the doping process depends on the crystallographic orientation of the substrate -- the resistivity of epitaxial layers grown on (111) substrates is higher than that of layers grown on (100) substrates.

In the transient study, the growth-rate dependence of the experimental decay time indicates that surface mechanisms, rather than gas-phase mechanisms, are also responsible for the transient behavior of the doping process. At very low growth rates the transient response is controlled by the RC time constant associated with the slowest step in the kinetic sequence, and the decay time associated with the transient is independent of the growth rate. At high growth rates the transient response is controlled by the rate at which silicon atoms cover arsenic atoms, which is determined by the epitaxial growth rate, and the decay time is inversely proportional to the growth rate. In both cases the surface capacitors play an important role. They incorporate time-varying effects related to the accumulation of dopant species in the adsorbed layer.

Since the equivalent circuit also relates the spatial variation of the dopant profile to the time-varying dopant gas flow, it was used to derive the transfer function of the dopant system.

## 3.3 LOW-PRESSURE CHEMICAL-VAPOR-DEPOSITION OF SILICON: PHOSPHORUS DOPING

### 3.3.1 Introduction

Polycrystalline silicon has many important applications in integrated circuit technology. Heavily doped films of polycrystalline silicon are used as the gate electrode in silicon-gate integrated circuits [3.21]. Lightly doped films are frequently used for high-value resistors in integrated circuits [3.22-2.24]. Until recently, these films have been generally deposited in atmospheric-pressure, cold-wall reactors, in which the silicon substrates lie flat on an externally heated susceptor. The electrical properties of such films have been reported by many researchers [3.25-3.29].

Within the last several years, low pressure, chemical-vapor-deposition (LPCVD) systems have been developed [3.30]. The high capacity and low cost of these systems led to their rapid acceptance in integrated circuit manufacturing facilities. Since the deposition rate, temperature and pressure in these reactors are markedly different from those conventionally employed, a study of the electrical properties of the material deposited in LPCVD reactors is needed. The resistivity of the films with a wide range of dopant concentrations and its stability during temperature cycling are especially important.

A study is being conducted at Stanford to investigate the properties of polycrystalline silicon deposited at low pressure. In the first phase of this study, we have reported the structure of low-pressure, polycrystalline-silicon films and their stability during further heat treatment [3.31]. In the present paper the electrical properties of phosphorus-implanted, low-

pressure polycrystalline silicon are discussed and compared to those of films deposited at atmospheric pressure. The effect of different annealing temperatures on resistivity is also reported.

### 3.3.2 Sample Fabrication

Three types of polycrystalline-silicon films were used in this study. Two types were prepared by thermal decomposition of silane on oxidized silicon wafers in a Tempress (Unicorp) LPCVD reactor with no intentionally added dopant. Nominal deposition temperatures of 580° and 620°C were used with 100% silane gas and a reactor pressure of approximately 0.5 Torr. The thickness of the films was about 0.47 μm, as measured with an automated spectrophotometer. The third type of film was prepared undoped by silane decomposition on oxidized wafers in a conventional, horizontal, atmospheric-pressure, cold-wall reactor. The deposition temperature was 960°C (corrected), and the films were about 0.63 μm thick.

In the first part of this study, the effect of annealing temperature on resistivity was investigated. Phosphorus doses of $3 \times 10^{15}$ and $1 \times 10^{16}$ ions/cm$^2$ were implanted into the low-pressure (LPCVD) films at an energy of 100 keV. The samples were annealed for one hour in dry $N_2$ at temperatures of 900, 1000, 1100 and 1200°C with different samples being used at each different temperature. Before annealing, the samples were coated with a thin layer of low-temperature deposited oxide to protect their surfaces from the furnace ambient and to prevent loss of dopant during the high-temperature processing. The grain structure shown by x-ray diffraction was the same after the anneal whether or not the coating was used [3.31]. After stripping the oxide, sheet resistance was measured using a four-point probe. Since the heat cycling was sufficient to distribute the dopant uniformly through the polycrystalline film [3.29], resistivity was obtained by multiplying the value of sheet resistance

by the thickness of each film.

In the second part of the study, low-pressure and atmospheric-pressure films were implanted with phosphorus doses ranging from $5 \times 10^{11}$ to $1 \times 10^{16}$ ions/cm$^2$ at an energy of 100 keV with two doses per decade. The samples were annealed for one hour at 1100°C; the first 16 min in dry $O_2$ and the remainder in dry $N_2$. In order to measure resistivity and mobility accurately on the more lightly doped samples (dose $\leq 10^{15}$ cm$^{-2}$) diffused-contact concentric rings, four-point probe, and cloverleaf-shaped, Van der Pauw structures were fabricated. After the annealing, photolithography was performed and the polycrystalline silicon was selectively etched to define these structures. A 0.5 μm thick layer of $SiO_2$ was deposited on the samples at 480°C and densified in wet oxygen for 20 min at 1000°C. A second photolithography step was performed to open contact windows in the oxide and the exposed polycrystalline silicon was doped with phosphorus from a $POCl_3$ source for 30 min at 900°C. Aluminum was deposited using an electron gun and the metal pattern was defined photolithographically. A 30 min anneal at 450°C in $N_2$ completed the devices. The silicon consumed during the two thermal oxidation steps was calculated, and the final thickness of each sample was used in the calculations. All measurements were made at room temperature, and the results reported here represent the average of measurements on at least three different samples.

Figure 3.16 shows the variation of resistivity with annealing temperature from 900 to 1200°C for the two types of LPCVD silicon deposited at 620°C (LP-620) and at 580°C (LP-580). The phosphorus doses of $3 \times 10^{15}$ and $1 \times 10^{16}$ ions/cm$^2$ are equivalent to dopant concentrations of $6.4 \times 10^{19}$ cm$^{-3}$ and $2.1 \times 10^{20}$ cm$^{-3}$, respectively.

For both types of films, the resistivity decreased with increasing annealing temperature, especially for the lower dose of $3 \times 10^{15}$ ions/cm$^2$

and in the LP-620 films. For the lower dose the resistivity of the LP-620 films decreased from $1.5 \times 10^{-2}$ Ω-cm after annealing at 900°C to $2.7 \times 10^{-3}$ Ω-cm after annealing at 1200°C, representing more than an 80% reduction in resistivity. On the other hand, the resistivities of the LP-580 films were $4.2 \times 10^{-3}$ Ω-cm and $2.5 \times 10^{-3}$ Ω-cm after annealing at the same temperatures - a reduction in resistivity of only 40%. For the higher dose of $1 \times 10^{16}$ ions/cm$^2$, the resistivities of the LP-620 films and LP-580 films decreased only 45% and 21%, respectively, over the same range of annealing temperatures. For both doses and at all annealing temperatures, the LP-580 films had lower resistivities than did the LP-620 films, especially at the lower annealing temperatures.

Figure 3.17 shows the resistivity of the three types of polycrystalline silicon as a function of the average dopant concentration. The resistivity of n-type single-crystal silicon is plotted on the same figure for comparison.

At the lowest dopant concentrations, the polycrystalline-silicon films had a resistivity of about 6-8 $\times 10^5$ Ω-cm, approximately six orders of magnitude higher than the resistivity of correspondingly doped single-crystal silicon. As the dopant concentration was increased, only small changes in resistivity were observed until the dopant concentration reached the range $2 \times 10^{17}$-$1 \times 10^{18}$ cm$^{-3}$; within this range the resistivity decreased sharply, finally falling to within half an order of magnitude of the single-crystal resistivity at high dopant concentrations. Throughout the whole dopant-concentration range, the LP-580 films had lower resistivity than the other samples. Also, the rapid decrease in resistivity occurred at a lower dopant concentration for the LP-580 samples. The resistivities of undoped polycrystalline-silicon films were measured and found to be approximately 8-9 $\times 10^5$ Ω-cm for all three types of film.

Hall mobility was measured on the van der Pauw structures by injecting a current, I, between two opposite contacts and measuring the resulting Hall voltage, $V_H$, when a magnetic field, B, was applied perpendicular to the plane of the sample. The equation used to calculate the average sample mobility, $\bar{\mu}$, is

$$\bar{\mu} = \frac{tV_H}{BI\rho} \tag{3.61}$$

where $\rho$ and t are the sample resistivity and thickness, respectively. A permanent magnet with a magnetic field of one kilogauss was used. For each measurement, the polarities of the injected current and of the magnetic field were reversed, and the value of Hall voltage was averaged over the four readings. The linearity of the measurement was tested frequently; Hall voltages were always found to be proportional to the injected current. All the measurements were made on Van der Pauw structures except for the highest two phosphorus doses where measurements were made on unpatterned square and rectangular-shaped samples. Because of the high dopant concentrations at these doses, good ohmic contact between the probes and the sample was obtained.

Figure 3.18 shows the carrier mobilities for the three types of polycrystalline silicon as functions of dopant concentration. The electron mobility in single-crystal silicon is also plotted for comparison. The mobility in polycrystalline-silicon films has been found to be a function of the thickness of the film [3.28]; consequently, the values reported here are averages over the thickness. At the highest dopant concentration, the carriers have a mobility of about 38 $cm^2/V\text{-}sec$ in all three types of polycrystalline silicon. As the dopant concentration was decreased, the mobility of each type of film increased slightly and then dropped sharply at a dopant concentration of about $1 \times 10^{18}$ $cm^{-3}$. Except at the highest doses, the Hall mobilities in the LP-580 films were significantly

higher than those in the other films. The Hall mobility was only measured at high and medium dopant concentrations in this study. At very low dopant concentrations, the resistance between opposite contacts of the Van der Pauw structures could greatly exceed $10^6$ ohms, degrading the accuracy of the measurements below acceptable levels.

The carrier concentration, n, can be calculated from the Hall data using the equation

$$n = \frac{BI}{qtV_H} \tag{3.62}$$

Figure 3.19 shows the carrier concentration as a function of dopant concentration for the three types of polycrystalline-silicon film.

3.3.4 Discussion

In a generally accepted model of polycrystalline silicon [3.29], the material is viewed as composed of small crystallites joined together by grain boundaries. Inside the crystallites, the atoms are arranged in a periodic manner, forming small single crystals, while the grain boundaries are composed of disordered atoms and contain large numbers of defects due to incomplete bonding. These defects cause trapping states at the grain boundaries capable of immobilizing the carriers, thus reducing the number of free carriers available for electrical conduction. Furthermore, once these traps capture carriers, they become electrically charged, creating potential-energy barriers which impede the motion of carriers from one crystallite to another. This reduces the mobility and creates regions of high resistivity at the grain boundaries. The model is illustrated in Fig. 3.20 for n-type polycrystalline material. To simplify the model, the

polycrystalline material is assumed to be composed of identical crystallites of length L cm. The impurity atoms within the crystallites are taken to be totally ionized and uniformly distributed with a concentration $N/cm^3$. The traps at the grain boundaries are presumed to be initially neutral with a concentration $N_t/cm^2$, and the band structure inside the crystallites is assumed to be that of single-crystal silicon.

Most of the trends observed in this experiment can be qualitatively explained by this model. The high resistivity observed at low dopant concentrations is explained by the trapping of most of the carriers at grain boundaries, leaving few free to contribute to the conduction. As the doping concentration is increased, the number of trapped carriers will increase in a manner determined by the energy distribution of the traps. Eventually the traps will approach saturation. Upon further increase in the doping concentration, the number of trapped carriers will not increase appreciably; however, the potential energy barrier will decrease and the space-charge region will narrow. Because of the exponential dependence of the resistivity of a grain boundary on the height of the energy barrier, this resistivity will decrease sharply with small increases in dopant concentration as the traps become saturated. This explains the observed abrupt reduction in resistivity at intermediate dopant concentrations. Finally, at high dopant concentrations, the region near the grain boundaries will no longer limit the conductivity of the samples, and the properties of the material will approach those of the crystallites.

This model has been used extensively to explain trends in the electrical properties of polycrystalline silicon [3.28,3.29]. Seto's analysis [3.29] showed that the conductivity and mobility in polycrystalline silicon are linearly dependent on the size of the crystallites L. He

observed a minimum in the carrier mobility at intermediate dopant concentrations and found that the dopant concentration at which the minimum should occur is inversely proportional to the size of the crystallites L. At higher dopant concentrations the mobility increases rapidly until it becomes limited by ionized impurity scattering. The mobility then decreases with increasing dopant concentration, as in single-crystal silicon, so that a maximum is observed [3.28].

In a previous study [3.31], in which the structure of LPCVD silicon films was investigated, we have shown that the LPCVD films grown at 580°C are initially amorphous. Such films are highly unstable and rapidly crystallize when annealed at even moderate temperatures. Transmission electron micrographs of the annealed samples showed that the recrystallized films which were initially amorphous have bigger grains than similarly annealed films grown at 620°C, which are initially polycrystalline. Furthermore, it has also been shown that the grains or crystallites grow bigger as the annealing temperature is increased. Based on this fact and on the model discussed above, the differences in electrical properties between the different types of polycrystalline-silicon films can be directly related to the size of the crystallites in each type of film.

Since the effect of grain boundaries on the electrical properties of polycrystalline silicon is especially important at intermediate dopant concentrations, we expect the differences in resistivity between films with different crystallites sizes to be maximum at these dopant concentrations. The lower resistivity of the recrystallized films grown at 580°C is then explained by their larger grain size. The films grown at atmospheric pressure have a slightly lower resistivity than do the films grown at 620°C, probably indicative of their slightly larger crystallites. At high dopant concentrations, the grain boundaries no longer limit the conduction [3.29],

which is dominated by the resistivity of the crystallites themselves; hence the resistivities of the three types of polycrystalline silicon approach a common value close to that of single-crystal silicon.

The same reasoning can be applied to the observed behavior of the mobility. The films grown at 580°C have higher mobilities at moderate dopant concentrations because they have larger crystallites, while at higher dopant concentrations the mobilities of the three types approach each other as the properties of the crystallites begin to dominate. The films grown at atmospheric pressure showed a slightly higher mobility than did the films grown at 620°C and low pressure -- another indication of slightly larger crystallites in the atmospheric-pressure films. Also, the mobility appears to approach its minimum value at lower dopant concentrations in the larger-grained LP-580 films, in full agreement with the model. At high dopant concentrations the Hall carrier concentrations in the three types of film are close to the single-crystal values, while the concentrations deviate from the single-crystal values at lower dopant concentrations where grain-boundary carrier trapping and resistivity begin to dominate. This deviation is most noticeable in the films with smaller crystallites.

The results of the annealing experiments can also be explained by this model. As the annealing temperature is increased, the crystallites grow larger, reducing the resistivity. This reduction should be especially strong at intermediate dopant concentrations, where the resistivity is a sensitive function of the crystallite size, consistent with the trends in Fig. 3.15 and the marked reduction in resistivity at the lower dose of $3 \times 10^{15}$ ions/cm$^2$. The amorphous films grown at 580°C recrystallize at a temperature much lower than 900°C [3.31], and apparently the crystallite size is not appreciably affected by further annealing, leading to only small changes in resistivity at higher annealing temperatures.

### 3.3.5 Summary and Conclusions

This study investigated conduction in polycrystalline-silicon films deposited at 580° and 620°C by low-pressure chemical vapor deposition and doped with phosphorus by ion implantation. Films deposited at 620°C were polycrystalline while those deposited at 580°C were initially amorphous but crystallized on further heat treatment. The electrical properties of the films were compared to those of polycrystalline-silicon films deposited at atmospheric pressure. The effect of annealing temperature on resistivity was studied first for phosphorus doses of $3 \times 10^{15}$ and $1 \times 10^{16}$ ions/cm$^2$. The resistivity was found to decrease with increasing annealing temperature. The films deposited at 580°C always had lower resistivity than those deposited at 620°C, with a greater difference appearing at lower annealing temperatures. In the second series of experiments, phosphorus was implanted with doses corresponding to average dopant concentrations ranging from $2 \times 10^{15}$ to $2 \times 10^{20}$/cm$^3$. The resistivity was only a slow function of the dopant concentration below $6 \times 10^{16}$/cm$^3$ and above $2 \times 10^{18}$/cm$^3$; however, for concentrations in the intermediate range, slight changes in concentration caused large changes in resistivity. As before, the films deposited at 580°C always showed the lowest resistivity of the three types of film investigated, especially in the intermediate doping range. The Hall mobility was measured and found to have a maximum near a dopant concentration of $6 \times 10^{18}$/cm$^3$, with the mobility being higher in films deposited at 580°C. The observed behavior is consistent with that expected from a film composed of small crystallites surrounded by grain boundaries containing large numbers of carrier traps.

## 3.4 SILICIDES: ANNEALING OF SPUTTERED $WSi_2$

### 3.4.1 Introduction

With the advances in the integrated circuit technology the device dimensions are continuously decreasing. This is beginning to pose new material problems. In MOS circuits poly-Si is widely used to interconnect the devices. In general the requirements for the interconnecting material are low resistivity, ability to withstand various chemicals and high temperatures encountered during fabrication process, and the capability to define fine patterns. Poly-Si is beginning to limit the performance of the circuits because of its poor conductivity. Although the grain size in as deposited poly-Si can be very small, however subsequent high temperature processing increases it markedly [3.32]. This causes a problem in defining very fine lines. Refractory metal silicides have been proposed as an alternative to poly-Si [3.33]. Most silicides have very low resistivity. Thermal oxidation of $MoSi_2$ [3.34] and $WSi_2$ [3.35] has been performed, making it possible to grow a continuous, electrically insulating $SiO_2$ overlayer, nearly of the same quality as thermally grown $SiO_2$ on Si [3.35].

The impurity masking ability of refractory metal silicides have been reported [3.33]. Silicides are stable against many acids and show the chemical properties similar to poly-Si [3.33,3.36]. Because of the extremely fine grain size of silicides it is possible to define very fine features [3.37]. Silicides have been used for a variety of other applications for a long time, e.g. to obtain ohmic contacts and Schottky barriers to Si and other semiconductors. The problem of "aluminum penetration" which can cause electrical shorts in shallow junction devices is eliminated using silicide barriers between aluminum and Si [3.38].

Intermetallic refractory metal silicides can be obtained by the deposition of the refractory metal on silicon and subsequent conversion of the

refractory metal to the desired intermetallic by reaction of the refractory metal with silicon at elevated temperatures [3.39]. Formation of silicides by chemical vapor deposition [3.40], coevaporation of metal and Si [3.35], sputtering from a target made of silicide [3.33] and laser irradiation of metal deposited on Si [3.41] have been investigated. The work presented in this paper was mainly centered around thin sputtered films of tungsten disilicide ($WSi_2$). Deposition of $WSi_2$ was performed on two different types of substrates, one set on oxidized single crystal Si and the other set had an additional layer of poly-Si in between $WSi_2$ and $SiO_2$. Both doped and undoped poly-Si were used. In the first set of experiments thermal annealing of these films was done for different temperatures in nitrogen ambient for different periods of annealing. The effect of these treatments on the resistivity and structure of the $WSi_2$ films was studied. In the second set of experiments Al was deposited on both types of films and low temperature annealing was performed to determine the structural stability of Al contacts to $WSi_2$. Compatibility of Al contacts to silicides of Mo, Pt/Ni and Co has been investigated by Van Gurp and Reukers [3.42] for different temperatures of annealing. Apparently after some critical temperature the silicide layer is dissociated by the reaction of the metal with aluminum and the precipitates consisting mainly of silicon and aluminum are formed. The critical temperature lies very close to the temperature commonly used for sintering. Therefore it is important to determine this temperature for Al/$WSi_2$ interface.

### 3.4.2 Experimental

#### A. Sample Preparation

Single crystal Si wafers with (100) and (111) orientation were thermally oxidized to grow 1000 Å $SiO_2$. Two types of poly-Si depositions

were performed on samples with (100) Si substrates. On the first set of wafers 4000 Å thick undoped poly-Si was low pressure chemically vapor deposited (LPCVD) in a Tempress LPCVD reactor. Nominal deposition temperature of 620°C was used with 100% $SiH_4$ gas as the source of Si at a reactor pressure of approximately 0.5 Torr. On the second set of wafers 5000 Å thick undoped poly-Si was deposited by $SiH_4$ decomposition in a conventional, horizontal, atmospheric-pressure cold wall reactor at a temperature of 960°C.

For part of the study a set of 4000 Å poly-Si films were implanted with phosphorus, with dose varying from $10^{14}$ to $10^{16}$ ions/cm$^2$, at an energy of 100 keV. Then samples were annealed for one hour at 1100°C, the first 15 min in dry $O_2$ and 45 min in dry $N_2$ ambient [3.32]. Another set of 4000 Å poly-Si films were implanted with boron, with dose varying from $10^{14}$ to $5 \times 10^{15}$ ions/cm$^2$, at an energy of 100 keV. The samples were thermally oxidized at 900°C wet oxygen for 30 min and then annealed at 1100°C for 30 min in dry $N_2$ ambient. The 5000 Å thick poly-Si films were doped by thermal diffusion. The first group was doped with boron using $B_2H_6$ as a source of impurity and the second group was doped with phosphorus using $POCl_3$ as a source of phosphorus. Both sets were doped for 30 min at 1000°C. A 4000 Å thick $WSi_2$ film was sputter deposited on top of the $SiO_2$/Si and poly-Si/$SiO_2$/Si structures, using a Perkin-Elmer sputtering system at a pressure of 20 μ in argon ambient. The substrate temperature was kept less than 300°C and the rate of deposition was 360 Å/min with 1.5 kV between electrodes and RF power of 280 watts. The target was made of hot pressed $WSi_2$. These samples were then annealed between 500 to 1200°C in dry $N_2$ ambient. Most of the samples were annealed for 30 min, however, some samples were annealed for times as long as 180 min.

The results of annealing at different temperatures (described in detail in Section 3.4.3) identified a temperature range in which cracks in

$WSi_2$ films were observed. In order to verify that internal stress in the films is causing cracks, as deposited films of $WSi_2$ were coated with 5000 Å thick $SiO_2$ at 480°C and then annealed between 600-950°C for 30 min in $N_2$ ambient.

Finally, the stability of Al contacts to $WSi_2$ films was studied by low temperature annealing. After the deposition of $WSi_2$ samples of $WSi_2/SiO_2/Si$ and $WSi_2$/undoped poly-Si/$SiO_2$/Si were annealed at 1000°C for two hours in $N_2$. Following the anneal, 1.5 μm thick Al was deposited on top of the $WSi_2$ layers using an electron-beam evaporation system. The resulting $Al/WSi_2/SiO_2/Si$ and $Al/WSi_2$/poly-Si/$SiO_2$/Si structures were annealed in the temperature range of 400 to 700°C for 30 min in dry $N_2$.

The annealed films in all of the experiments were examined to study their electrical properties and structure. A four point probe was used to determine resistivity, optical microscope and scanning electron microscope (SEM) were used to determine surface quality, and glancing x-ray diffraction technique was used to determine crystal structure. Auger electron spectroscopy and electron microprobe were used to determine W to Si ratio in as deposited $WSi_2$ films.

### 3.4.3 Results

#### A. Properties of as Deposited WSi2 Films

As measured with Auger electron spectroscopy and electron microprobe, as deposited films had the 2:1 ratio between the number of silicon atoms as deposited films had the 2:1 ratio between the number of silicon atoms and the number of the tungsten atoms. With atomic percentage of 66%, weight percent of silicon was 23.4%, the x-ray diffraction showed no specific peaks of any of the silicides of tungsten, indicating that the unannealed films were amorphous. An average resistivity of $6.5 \times 10^{-4}$ Ωcm was measured for the films with an underlayer of undoped poly-Si and $4.8 \times 10^{-4}$ Ωcm for the films without it.

The films were not attacked by chemicals generally used in fabrication of ICs, e.g. $H_2SO_4$, $HNO_3$, HCl, Hf and mixtures of $H_2SO_4 + H_2O_2$. $HCl + H_2O_2$ and $NH_4OH + H_2O_2$; however they were heavily attacked by mixtures of $HF + HNO_3$ and $NH_4F + HNO_3$, the common etchants for Si. Therefore in all of the experiments described here the samples were cleaned using normal cleaning procedures commonly used for Si.

B. Properties of Annealed $WSi_2$ Films

The resistivity measurements have been done to investigate the effect of annealing on electrical properties of $WSi_2$/poly-Si/$SiO_2$/Si and $WSi_2$/$SiO_2$/Si films. The temperature was varied between 500 to 1200°C with 100°C intervals in $N_2$ ambient for 30 min. Between the temperature range of 650°C to 900°C the resistivity showed wide variation and it was not possible to measure a meaningful value of the resistivity. Closer examination showed the existance of micro cracks in the films, which might be causing interruption in the current flow. The change of resistivity of the samples for annealing temperatures between 900°C-1200°C has been plotted in Fig. [3.21]. The resistivity decreases as a function of temperature. In order to investigate the effect of the annealing time on resistivity variations, samples were annealed for prolonged times at 650°C and 975°C. The results of change in resistivity by annealing $WSi_2$/poly-Si/$SiO_2$/Si film at 650°C for 15-60 min has been shown in Fig. 3.22. There is no specific trend of decrease in the resistivity with time. The results of variation in resistivity by annealing the $WSi_2$/$SiO_2$/Si and $WSi_2$/poly-Si/$SiO_2$/Si films at 975°C for 30-180 min have been plotted in Fig. 3.23 and a definite reduction in resistivity is observed.

X-ray diffraction was performed to study the structure of the annealed films. The x-ray diffraction of annealed $WSi_2$/poly-Si/$SiO_2$/Si

film for two hours at 975°C shows that the crystallized $WSi_2$ film does not have a preferred orientation. X-ray diffraction of $WSi_2$/poly-Si/$SiO_2$/Si samples which were annealed at 800°C and 900°C for 30 min were compared. The results were identical except at 900°C peaks of $WSi_2$ were intensified. Comparing this result to the x-ray diffraction of the sample which were annealed at 975°C for two hours shows that the $WSi_2$ peaks of 800°C and 900°C annealed samples have shifted within $\pm$0.2 degrees (2$\theta$) with respect to the peaks of 975°C annealed samples. This can be attributed to internal stress within the films [3.43].

Finally, the resistivity variations of $WSi_2$/doped poly-Si/$SiO_2$/Si vs. doping density of poly-Si have been plotted in Fig. 3.24 and for comparison resistivity of poly-Si films doped with boron and phosphorus are included. As previously mentioned between the temperature range of 650-900°C all the annealed films developed cracks. To further investigate this problem the samples were coated with $SiO_2$ at a temperature of 480°C and annealed afterwards. In case of $SiO_2$ deposited/$WSi_2$/$SiO_2$/Si films no cracks were observed for temperature range between 600-950°C annealing at $N_2$ for 30 min. However, the results of $SiO_2$ deposited/poly-Si/$SiO_2$/Si films show that for films that have been annealed at 850°C and 975°C for 30 min, cracks were observed for the first film only. In comparison of the x-ray diffraction patterns the same result was observed for both cases except peak intensification of the second film.

Average resistivity for $SiO_2$/$WSi_2$/poly-Si/$SiO_2$/Si annealed at 975°C for 30 min is about $5.2 \times 10^{-4}$ $\Omega$cm while for $WSi_2$/poly-Si/$SiO_2$/Si annealed for two hours at the same temperature average resistivity is about $1.2 \times 10^{-4}$ $\Omega$cm. X-ray diffraction of these two films showed that for $SiO_2$

covered films, peaks of $W_5Si_3$ also exist which are not present at the second film.

C. Properties of Annealed Al/$WSi_2$ Films

The Al/$WSi_2$/poly-Si/$SiO_2$/Si and Al/$WSi_2$/$SiO_2$/Si structures were annealed in $N_2$ ambient for 30 min in the temperature range of 400°C to 700°C with 50° intervals. The quality of the surface was examined by optical microscope and SEM. The temperature at which damage to surface occurred was above 550°C for the first structure and above 600°C for the second structure. At and below these temperatures annealing did not affect the chemical and electrical properties of Al/$WSi_2$ layers. Average resistivity of the films was about $3 \times 10^{-6}$ Ωcm dominated by Al. X-ray diffraction on all films were performed. For Al/$WSi_2$/poly-Si/$SiO_2$/Si films annealing at temperatures lower than 600°C, intensity of Si (110) remains constant. This is the peak of dominant recrystallized LPCVD 620°C poly-Si [14]. Intensity of the $WSi_2$ peaks of different orientations and aluminum remain constant. After annealing at 600°C, x-ray diffraction patterns showed decreased intensity of peak of (110) Si and as temperature increased the intensity of the peak further decreased. The intensity of peaks of (101) and (002) $WSi_2$ increased and aluminum peaks vanished as temperature increased above 600°C.

For Al/$WSi_2$/$SiO_2$/Si structures, annealing less than 650°C intensities of aluminum and $WSi_2$ peaks remain constant however at or above this temperature, although $WSi_2$ peaks remain unaltered, the Al peak completely vanishes and new peaks of $WAl_{12}$ and (111) Si appear indicating the existence of some chemical reaction. SEM and optical photograph of the surface of Al/$WSi_2$/$SiO_2$/Si annealed below and above the critical temperature are shown in Fig. 3.25.

### 3.4.4 Discussion

The improvement of electrical characteristics of tungsten disilicide over poly-Si can be mentioned as lower resistivity for $WSi_2$ films; and, because of a layer of small grains, possibility of fine pattern definition is a very important factor for high density integrated circuits. This is a serious problem for poly-Si layers because of large grain existence. The advantage over metal interconnection and electrodes can be mentioned as improvement on thermal, chemical and oxidation resistance using silicide films and, finally, another unique property of silicide films is masking the impurities.

As deposited films of $WSi_2$ were amorphous and applications of annealing produces homogenity and formation of crystallized grains which provide further reduction on resistivity. In the course of annealing at certain temperature ranges, cracks are developed within the $WSi_2$ films which are due to the stress exerted within the $WSi_2$ film and its interface with other materials. This internal stress decreases as temperature increases. We also showed that this stress can be deleted by using $SiO_2/WSi_2/SiO_2$ structure. The formation of cracks is not correlated to the existence of another phase of tungsten silicide. Finally, annealing above 900°C causes formation of randomly oriented crystallites. Annealing of $WSi_2$/doped poly-Si/$SiO_2$/Si films shows a decrease in resistivity of the whole structure as doping density of poly-Si film increases. This can be described as two parallel resistors, the combination of them providing lower resistivity.

As shown Al/$WSi_2$/$SiO_2$ and Al/$WSi_2$/poly-Si/$SiO_2$/Si layers are stable up to 600°C and 550°C annealing temperatures, respectively. No chemical degradation was observed at these ranges of temperatures. Both these temperatures are well above the temperatures commonly used for sintering the Al contacts to Si. Above these temperatures, for Al/$WSi_2$/$SiO_2$, one possible

mechanism of degradation is formation of $WAl_{12}$ and free silicon because of tungsten consumption from $WSi_2$ films and for $Al/WSi_2/poly\text{-}Si/SiO_2/Si$ diffusion of aluminum through a silicide layer and formation of aluminum-silicon precipitations.

### 3.4.5 Summary and Conclusion

$WSi_2$ films are suitable for interconnections and electrodes for integrated circuit technology because of improvement in electrical as well as chemical properties over poly-Si and metals. The effect of thermal annealing provided improvements in conductivity because of formation of homogenous and crystallized layer. This annealing should be done above certain temperatures to avoid formation of cracks which exist because of internal stress within the film layer. Compatibility of $Al/WSi_2$ layer has been verified for cases of available free silicon atoms as well as no source of silicon atoms.

## References


3.1 T.I. Kamins, R. Reif and K.C. Saraswat, Electrochem. Soc. Fall Meeting, Abs. 230, Las Vegas, Nev., Oct. 17-22, 1976, pp. 601-603.

3.2 J.D. Meindl, K.C. Saraswat and J.D. Plummer, Semiconductor Silicon 1977, Princeton, New Jersey, The Electrochem. Soc. Inc., 1977, pp. 894-909.

3.3 R. Reif, T.I. Kamins and K.C. Saraswat, Electrochem. Soc. Fall Meeting, Abs. 350, Atlanta, Ga., Oct. 9-14, 1977, pp. 921-923.

3.4 R. Reif, T.I. Kamins and K.C. Saraswat, J. Electrochem. Soc., 125 (1978) 1980.

3.5 J.D. Meindl, et al., "Final Report on Computer-Aided Semiconductor Process Modeling," Stanford Electronics Laboratories TR No. 4969-73-F, Stanford University, Stanford, Calif., Oct. 1976.

3.6 W.H. Shepherd, J. Electrochem. Soc., 115 (1968) 541.

3.7 J. Bloem, J. Crystal Growth, 13/14 (1972) 302.

3.8 W.H. Shepherd, J. Electrochem. Soc., 112 (1965) 988.

3.9 R. Reif, T.I. Kamins and K.C. Saraswat, Electrochem. Soc. Spring Meeting, Abs. 208, Seattle, Wash., May 21-26, 1978, pp. 509-511.

3.10 M.M. Faktor and I. Garrett, Growth of Crystals from the Vapor, Chapman and Hall, 1974, p. 191.

3.11 D.W. Shaw, in Crystal Growth-Theory and Techniques, (C.H.L. Goodman, ed.), Vol. I, Plenum, London (1974), pp. 1-48.

3.12 J.P. Duchemin, Revue Technique, Thomson-CSF, 9 (1977) 411.

3.13 M.M. Faktor and I. Garrett, op cit., p. 95.

3.14 H. Kobayashi and M. Kobayashi, Catalysis Reviews - Science and Engineering, 10 (1974) 139.

3.15 J.D. Meindl, et al., Stanford Electronics Laboratory TR No. 78-011, Stanford University, Stanford, Calif., Feb. 1978.

3.16 D.W. Shaw, op cit., p. 11.

3.17 A.S. Grove, Physics and Technology of Semiconductor Devices, John Wiley and Sons, New York (1967), p. 19.

3.18 K.G. Denbigh and J.C.R. Turner, Chemical Reactor Theory, Cambridge University Press, 2nd edition (1971), p. 24.

3.19 D.W. Shaw, op cit., p. 14.

3.20 C.L. Yaws and J.R. Hopper, Solid State Technology 17 (Jan. 1974) 47.

3.21 F. Faggin and T. Klein, Solid State Electron., 13 (1970) 1125.

3.22 T.R. O'Connel, et al., IEEE Intl. Solid State Circuits Conf., Philadelphia, Feb. 16-18, 1977, Paper WAM 1.2.

3.23 V.G. McKenny, IEEE Intl. Solid State Circuits Conf., Philadelphia, Feb. 16-18, 1977, Paper WAM 1.3.

3.24 L. Gerzberg and J.D. Meindl, DRC78 Abstracts, IEEE Trans. Electron Devices, ED-25 (Nov. 1978) 1357.

3.25 A.L. Fripp and L.H. Slack, J. Electrochem. Soc., 120 (1973) 145.

3.26 P. Rai-Choudhury and P.L. Hower, J. Electrochem. Soc., 120 (1973) 1761.

3.27 M.E. Cowher and T.O. Sedgwick, J. Electrochem. Soc., 119 (1972) 1565.

3.28 T.I. Kamins, J. Appl. Phys., 42 (1971) 4357.

3.29 J.Y.W. Seto, J. Appl. Phys., 46 (1975) 5247.

3.30 R.S. Rosler, Solid State Tech., 20(4) (1977) 63.

3.31 T.I. Kamins, M.M. Mandurah and K.C. Saraswat, J. Electrochem. Soc., 125 (1978) 927.

3.32 T.I. Kamins, M.M. Mandurah and K.C. Saraswat, J. Electrochem. Soc., 125(6) (June 1978) 932-937.

3.33 T. Mochizuki, K. Shibata, T. Inoue, K. Ohuchi and M. Kashiwagi, Electrochem. Soc. Meeting, Atlanta, Abs. 331, Oct. 1977, pp. 871-783.

3.34 T. Inoue and K. Koike, Appl. Phys. Lett., 33(9) (Nov. 1978) 826-827.

3.35 S. Zirinsky, W. Hammer, F. d'Heurle and J. Baglins, IBM Research Report, 2 (1978) 21.

3.36 T. Mochizuki and M. Kashiwagi, Electrochem. Soc. Meeting, Pennsylvania, Abs. 201, (1978) 537-539.

3.37 T. Mochizuki and K. Ohuchi, IEEE Trans. on Electron Devices, ED-25(12) (Dec. 1978) 1409-1411.

3.38 U. Koster, K.N. Tu and P.S. Ho, IBM Research Report, 5 (1977) 6.

3.39 L.D. Locker and C.D. Capio, J. Appl. Phys., 44(10) (Oct. 1973) 4366-69.

3.40 Dietrich E.R. Kehr, Proc. 6th Intl. Conf. on CVD, Electrochem. Soc., 77-5 (1977) 511-528.

3.41 F. Mohammadi, K.C. Saraswat, J. Beaudouin and J.D. Meindl, Extended Abstracts, Fall Meeting of Electrochem. Soc., Boston, Vol. 79-1, Abs. 146, May 1979.

3.42 G.J. Van Gurp and W.M. Reukers, Electrochem. Soc. Meeting, Pennsylvania Abs. 215, 1978, pp. 576-577.

3.43 B.D. Cullity, Elements of X-Ray Diffraction, Addison-Wesley, Reading, Mass., 1956, pp. 285-287.

## Appendix A

The following chemical reactions enter into the mass-balance Eqs. (3.3), (3.4), (3.6), and (3.7).

Adsorption of $AsH_2$ on the surface:

$$AsH_2\,(g) + s \underset{k_{r4'}}{\overset{k_{f4'}}{\rightleftharpoons}} AsH_2\text{-}s \qquad K_{4'} = \frac{\left(\theta_{AsH_2} \cdot N_s\right)}{\left(X_{AsH_2} \cdot P_{H_2}\right) \cdot \left(\theta \cdot N_s\right)} = \frac{k_{f4'}}{k_{r4'}}$$

Chemical dissociation of $AsH_3$ in the adsorbed layer:

$$AsH_3\text{-}s + 3s \underset{k_{r5}}{\overset{k_{f5}}{\rightleftharpoons}} As\text{-}s + 3H\text{-}s \qquad K_5 = \frac{\left(\theta_{As} \cdot N_s\right) \cdot \left(\theta_H^3 \cdot N_s^3\right)}{\left(\theta_{AsH_3} \cdot N_s\right) \cdot \left(\theta^3 \cdot N_s^3\right)} = \frac{k_{f5}}{k_{r5}}$$

Chemical dissociation of $AsH_2$ in the adsorbed layer:

$$AsH_2\text{-}s + 2s \underset{k_{r5'}}{\overset{k_{f5'}}{\rightleftharpoons}} As\text{-}s + 2H\text{-}s \qquad K_{5'} = \frac{\left(\theta_{As} \cdot N_s\right) \cdot \left(\theta_H^2 \cdot N_s^2\right)}{\left(\theta_{AsH_2} \cdot N_s\right) \cdot \left(\theta^2 \cdot N_s^2\right)} = \frac{k_{f5'}}{k_{r5'}}$$

Incorporation of adsorbed As into step or kink sites on the surface:

$$As\text{-}s + s^i \underset{k_{r6}}{\overset{k_{f6}}{\rightleftharpoons}} As\text{-}s^i + s + s^i \qquad K_6 = \frac{\left(\theta_{As}^i \cdot N_s^i\right) \cdot \left(\theta \cdot N_s\right)}{\left(\theta_{As} \cdot N_s\right)} = \frac{k_{f6}}{k_{r6}}$$

In these equations, $K_i$ is the equilibrium constant corresponding to step $i$ in the doping process (cf. Section II), $k_{fi}$ and $k_{ri}$ are the forward and reverse reaction rate constants, respectively, $\theta_{As}\left(\theta_H, \theta_{AsH_2}\right)$ is the fraction of adsorption sites occupied by $As(H, AsH_2)$, $\theta_{As}^i$ is the fraction of incorporation sites occupied by As, and $N_s^i$ is the surface density of incorporation sites. With the above expressions, the mass-balance equations may be written as follows:

Equation (3.3): mass-balance of adsorbed $AsH_3$ on the surface.

$$\left(k_{f4} \cdot X^{s}_{AsH_3} \cdot \theta - k_{r4} \cdot \theta_{AsH_3}\right) - \left(k_{f5} \cdot \theta_{AsH_3} \cdot \theta^3 - k_{r5} \cdot \theta_{As} \cdot \theta_H^3\right) = N_s \cdot \frac{d\theta_{AsH_3}}{dt} \tag{A.1}$$

Equation (3.4): mass-balance of adsorbed As on the surface.

$$\left[\left(k_{f5} \cdot \theta_{AsH_3} \cdot \theta^3 - k_{r5} \cdot \theta_{As} \cdot \theta_H^3\right) + \left(k_{f5'} \cdot \theta_{AsH_2} \cdot \theta^2 - k_{r5'} \cdot \theta_{As} \cdot \theta_H^2\right)\right] - \left(k_{f6} \cdot \theta_{As} - k_{r6} \cdot \theta_{As}^{i} \cdot N_s^{i} \cdot \theta\right) = N_s \cdot \frac{d\theta_{As}}{dt} \tag{A.2}$$

Equation (3.6): mass-balance of $AsH_2$ just above the gas-solid interface.

$$\left(k_{f3} \cdot X^{s}_{AsH_3} - k_{r3} \cdot X^{s}_{AsH_2}\right) - \left(k_{f4'} \cdot X^{s}_{AsH_2} \cdot \theta - k_{r4'} \cdot \theta_{AsH_2}\right) = \varepsilon \cdot N_{T'} \cdot \frac{dX^{s}_{AsH_2}}{dt} \tag{A.3}$$

Equation (3.7): mass-balance of adsorbed $AsH_2$ on the surface.

$$\left(k_{f4'} \cdot X^{s}_{AsH_2} \cdot \theta - k_{r4'} \cdot \theta_{AsH_2}\right) - \left(k_{f5'} \cdot \theta_{AsH_2} \cdot \theta^2 - k_{r5'} \cdot \theta_{As} \cdot \theta_H^2\right) = N_s \cdot \frac{d\theta_{AsH_2}}{dt} \tag{A.4}$$

## Appendix B

$$\frac{X^i - \bar{X}}{R_1} - \frac{\bar{X} - X^s}{R_2} = C_1 \cdot \frac{d\bar{X}}{dt} \tag{B.1}$$

$$\frac{\bar{X} - X^s}{R_2} - \frac{X^s - (X)^e_{XAsH_2}}{R_3} - \frac{X^s - (X)^e_{\theta AsH_3}}{R_4} = C_2 \cdot \frac{dX^s}{dt} \tag{B.2}$$

$$\frac{X^s - (X)^e_{\theta AsH_3}}{R_4} - \frac{(X)^e_{\theta AsH_3} - (X)^e_{\theta As}}{R_5} = C_4 \cdot \frac{d(X)^e_{\theta AsH_3}}{dt} \tag{B.3}$$

$$\frac{(X)^e_{\theta AsH_3} - (X)^e_{\theta As}}{R_5} + \frac{(X)^e_{\theta AsH_2} - (X)^e_{\theta As}}{R_{5'}} - \frac{(X)^e_{\theta As} - (X)^e_{\theta^i As}}{R_6} = C_5 \cdot \frac{d(X)^e_{\theta As}}{dt} \tag{B.4}$$

$$\frac{X^s - (X)^e_{XAsH_2}}{R_3} - \frac{(X)^e_{XAsH_2} - (X)^e_{\theta AsH_2}}{R_{4'}} = C_3 \cdot \frac{d(X)^e_{XAsH_2}}{dt} \tag{B.5}$$

$$\frac{(X)^e_{XAsH_2} - (X)^e_{\theta AsH_2}}{R_{4'}} - \frac{(X)^e_{\theta AsH_2} - (X)^e_{\theta As}}{R_{5'}} = C_{4'} \cdot \frac{d(X)^e_{\theta AsH_2}}{dt} \tag{B.6}$$

$$\frac{(X)^e_{\theta As} - (X)^e_{\theta^i As}}{R_6} - \frac{(X)^e_{\theta^i As}}{R_7} = C_6 \cdot \frac{d(X)^e_{\theta^i As}}{dt} \tag{B.7}$$

where:

$$(X)^{e}_{XAsH_2} \equiv \frac{X^{s}_{AsH_2}}{K_3} \tag{B.8}$$

$$(X)^{e}_{\theta AsH_3} \equiv \frac{\theta_{AsH_3}}{K_4 \cdot \theta} \tag{B.9}$$

$$(X)^{e}_{\theta As} \equiv \frac{\theta_H^{2} \cdot \theta_{As}}{K_3 \cdot K_4 \cdot K_5 \cdot \theta^{3}} \tag{B.10}$$

$$(X)^{e}_{\theta AsH_2} \equiv \frac{\theta_{AsH_2}}{K_3 \cdot K_4 \cdot \theta} \tag{B.11}$$

$$(X)^{e}_{\theta^{1}As} \equiv \frac{k_H \cdot N_{As}}{K_P} \tag{B.12}$$

$$1/R_1 \equiv N_{300} \cdot v_{H_2} \tag{B.13}$$

$$1/R_2 \equiv k_m \tag{B.14}$$

$$1/R_3 \equiv k_{f3} \tag{B.15}$$

$$1/R_4 \equiv k_{f4} \cdot \theta \tag{B.16}$$

$$1/R_{4'} \equiv k_{f4'} \cdot K_3 \cdot \theta \tag{B.17}$$

$$1/R_5 \equiv k_{f5} \cdot K_4 \cdot \theta^4 \tag{B.18}$$

$$1/R_{5'} \equiv k_{f5'} \cdot K_{4'} \cdot K_3 \cdot \theta^3 \tag{B.19}$$

$$1/R_6 \equiv k_{f6} \cdot K_{5'} \cdot K_{4'} \cdot K_3 \cdot \frac{\theta^3}{\theta_H^2} \tag{B.20}$$

$$1/R_7 \equiv g \cdot \frac{K_p}{k_H} \tag{B.21}$$

$$C_1 \equiv L \cdot N_T \tag{B.22}$$

$$C_2 \equiv \varepsilon \cdot N_{T'} \tag{B.23}$$

$$C_3 \equiv \varepsilon \cdot N_{T'} \cdot K_3 \tag{B.24}$$

$$C_4 \equiv N_s \cdot K_4 \cdot \theta \tag{B.25}$$

$$C_{4'} \equiv N_s \cdot K_3 \cdot K_{4'} \cdot \theta \tag{B.26}$$

$$C_5 \equiv N_s \cdot K_4 \cdot K_5 \cdot \frac{\theta^4}{\theta_H^3} \tag{B.27}$$

$$C_6 \equiv \frac{K_p}{K_7} \tag{B.28}$$

$(X)_j^e$ is the arsine molar fraction that would be in equilibrium with the As-containing species j, if the corresponding mechanisms were in thermodynamic equilibrium. $K_p$ is the equilibrium constant for the overall doping reaction, i.e.

$$AsH_3\,(g) \rightleftharpoons As(ss) + \tfrac{3}{2} H_2(g) \qquad K_p = \frac{\alpha_{As} \cdot P_{H_2}^{3/2}}{(X_{AsH_3} \cdot P_{H_2})} = \frac{k_H \cdot N_{As}}{X_{AsH_3}} =$$

$$K_4 \cdot K_5 \cdot K_6 \cdot K_7 \cdot \frac{\theta^3}{\theta_H^3} \qquad (B.29)$$

and g is the epitaxial growth rate.

Table 3.1. Physicochemical mechanism corresponding to each resistor and location in the epitaxial system corresponding to each capacitor.

| Resistor | Physicochemical Mechanism | Capacitor | Species | Location in the Epitaxial System |
|---|---|---|---|---|
| $R_1$ | Forced-Convection Mass Transport | $C_1$ | $AsH_3$ | Deposition region |
| $R_2$ | Boundary-Layer Mass Transport | $C_2$ | $AsH_3$ | Thin layer just above the growing surface |
| $R_3$ | Gas-Phase Chemical Reaction | $C_3$ | $AsH_2$ | |
| $R_4$ ($R_{4'}$) | $AsH_3$ ($AsH_2$) adsorption on the growing surface | $C_4$ ($C_{4'}$) | $AsH_3$ ($AsH_2$) | Adsorbed layer |
| $R_5$ ($R_{5'}$) | $AsH_3$ ($AsH_2$) dissociation in the adsorbed layer | $C_5$ | As | |
| $R_6$ | Surface diffusion and site incorporation of As | $C_6$ | As | Incorporation sites on the surface |
| $R_7$ | Covering of As atoms by Si atoms during epitaxial growth | | | |

Table 3.2

Expressions for $N_{As}$ in the high growth-rate region for each possible controlling mechanism and the temperature dependence expected in each case

| | Controlling Mechanism | Expression | Expected Temperature Dependence of Dopant Concentration |
|---|---|---|---|
| GAS PHASE | Forced-Convection Mass Transport | $N_{As} \simeq \frac{N_{300} \cdot v_{H_2}}{g} \cdot x^i$ | Decreases with increasing temperature with an apparent activation energy of 3-8 kcal/mole |
| | Boundary-Layer Mass Transport | $N_{As} \simeq \frac{k_m}{g} \cdot x^i$ | Relatively unaffected by changes in the deposition temperature |
| | Gas-Phase Chemical Reaction | $N_{As} \simeq \frac{k_{f3}}{g} \cdot x^i$ | Increases with increasing temperature showing an Arrhenius activation energy |
| SURFACE | Adsorption | $N_{As} \simeq \frac{k_{f4'} \cdot K_3 \cdot \theta}{g} \cdot x^i$ | Decreases with Increasing Temperature (unless $E_a \geq \lvert\Sigma\Delta H_i\rvert$) |
| | Surface Chemical Dissociation | $N_{As} \simeq \frac{k_{f5'} \cdot K_{4'} \cdot K_3 \cdot \theta^3}{g} \cdot x^i$ | |
| | Surface Diffusion and Site Incorporation | $N_{As} \simeq \frac{k_{f6} \cdot K_{5'} \cdot K_{4'} \cdot K_3}{g} \cdot \frac{\theta^3}{\theta_H^2} \cdot x^i$ | |

Table 3.3

EXPERIMENTAL DEPENDENCE OF THE DOPING PROCESS
ON SUBSTRATE CRYSTALLOGRAPHIC ORIENTATION
(The position of the substrates refer to Fig. 3.13)

| Position | ① | ② | ③ | ④ |
|---|---|---|---|---|
| Orientation | (111) | (100) | (111) | (100) |
| Growth rate (μm/min) | 0.30 | 0.30 | 0.29 | 0.30 |
| Resistivity (Ω-cm) | 1.57 | 1.16 | 1.47 | 1.11 |
| Orientation | (100) | (111) | (100) | (111) |
| Growth rate (μm/min) | 0.43 | 0.41 | 0.41 | 0.40 |
| Resistivity (Ω-cm) | 1.55 | 1.86 | 1.41 | 1.79 |

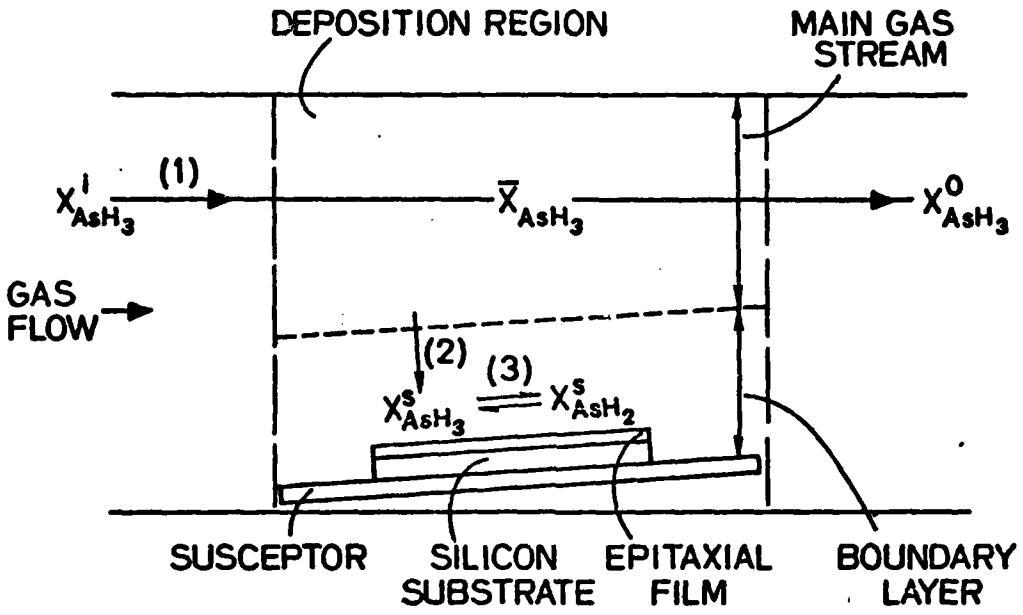



Fig. 3.1. Sequence of steps occurring in the gas phase: (1) forced-convection mass transport, (2) boundary-layer mass transport, (3) gas-phase chemical reactions. $X^{I}_{AsH_3}$: input $AsH_3$ molar fraction; $X^{O}_{AsH_3}$: output $AsH_3$ molar fraction; $\bar{X}_{AsH_3}$: average $AsH_3$ molar fraction in the well-mixed main gas stream within the deposition region; $X^{s}_{AsH_3}\left(X^{s}_{AsH_2}\right)$: gas-phase molar fraction of $AsH_3$ ($AsH_2$) just above the gas-solid interface.

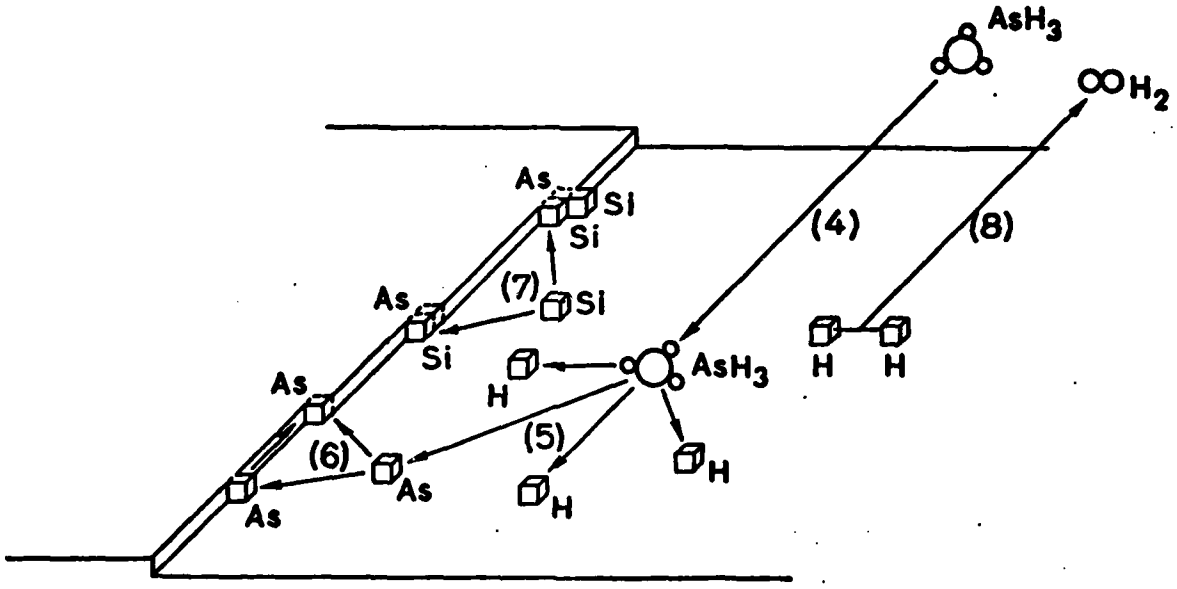



Fig. 3.2. Sequence of steps occurring on the surface: (4) adsorption of the As-containing compound; (5) surface chemical dissociation; (6) surface diffusion and site incorporation; (7)"burying" of As by subsequently arriving Si atoms; (8) desorption of hydrogen.

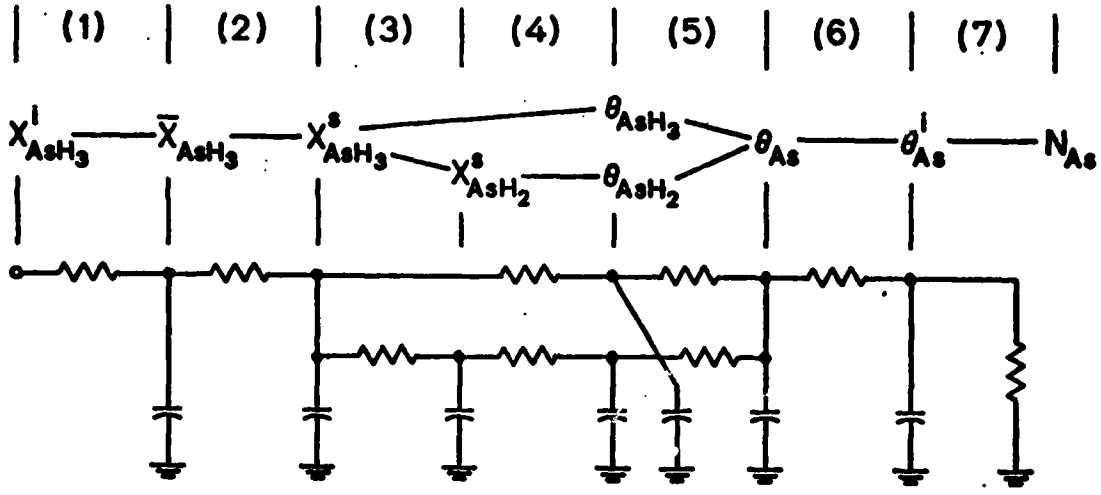



Fig. 3.3. Circuit representation of the dopant system. Each resistor represents one of the steps in the doping process. Steps (1)-(7) correspond to the mechanisms shown in Figs. 1 and 2.

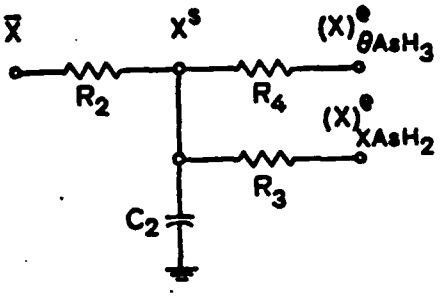



Fig. 3.4. Circuit representation of the mass-balance of $AsH_3$ just above the gas-solid interface [Eqs. (2), (18), (34)]. The capacitor incorporates time-varying effects related to the accumulation of $AsH_3$.

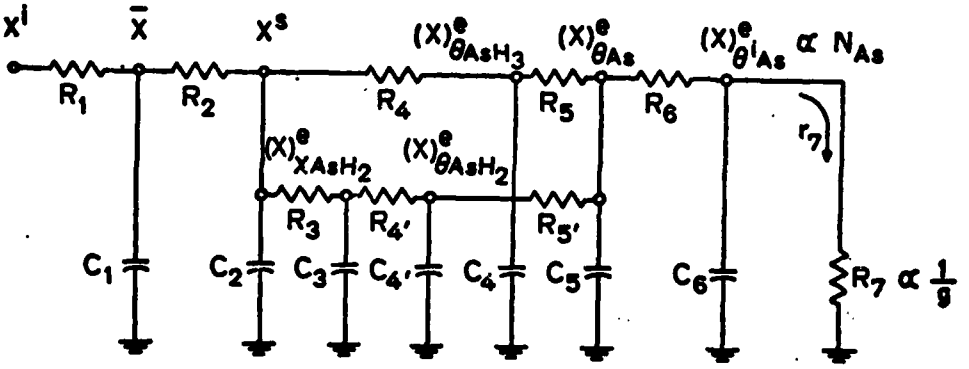



Fig. 3.5. Equivalent circuit representing the total doping process. Each node corresponds to one point in the dopant system, and it is associated with one As-containing species (c.f. Fig. 3); $(X)^e_i$ is the arsine molar fraction corresponding to the As-containing species associated with the $i^{th}$ node.

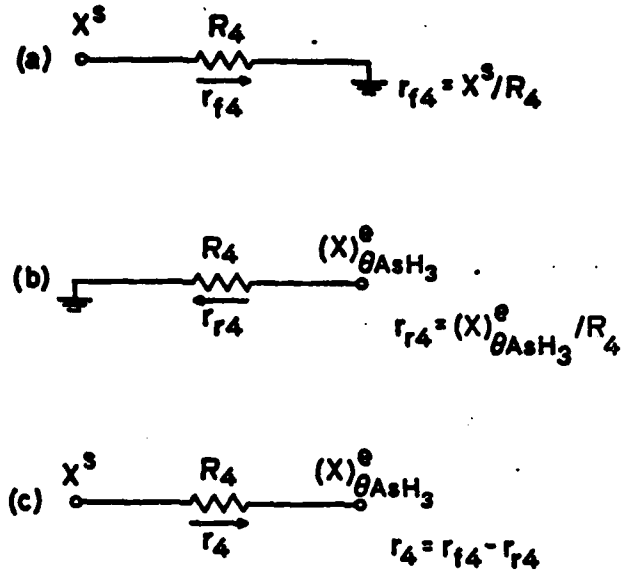



Fig. 3.6. Circuit representation of reaction rates associated with $AsH_3$ adsorption (step 4). (a) forward reaction rate $r_{f4}$; (b) reverse reaction rate $r_{r4}$; (c) net reaction rate $r_4$.

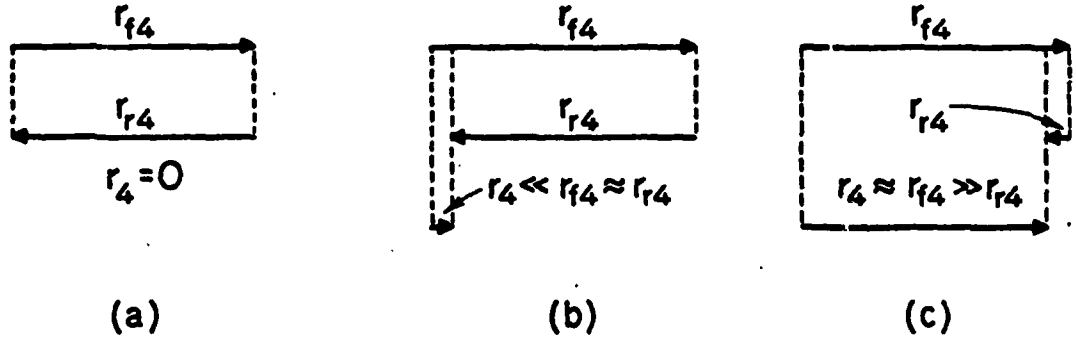



Fig. 3.7. Schematic representation of a reaction (a) in thermodynamic equilibrium, (b) in quasi-equilibrium, and (c) under kinetic control. The length of each arrow is proportional to the corresponding reaction rate.

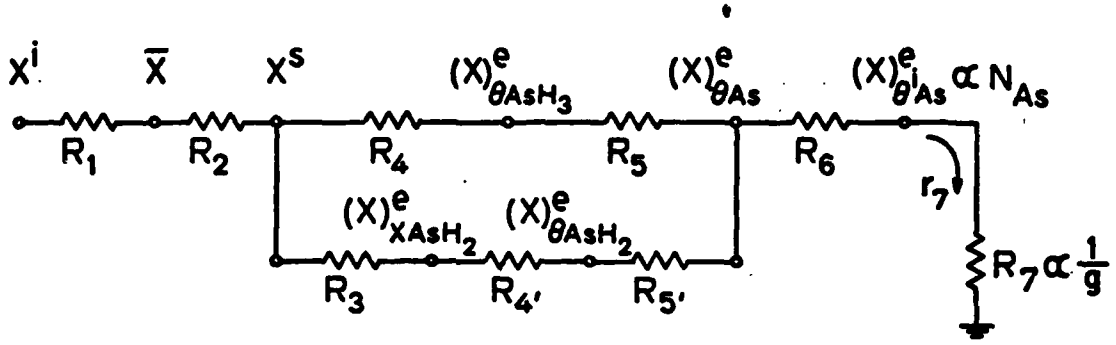



*Fig. 3.8. Steady-state circuit representation of the doping process.*

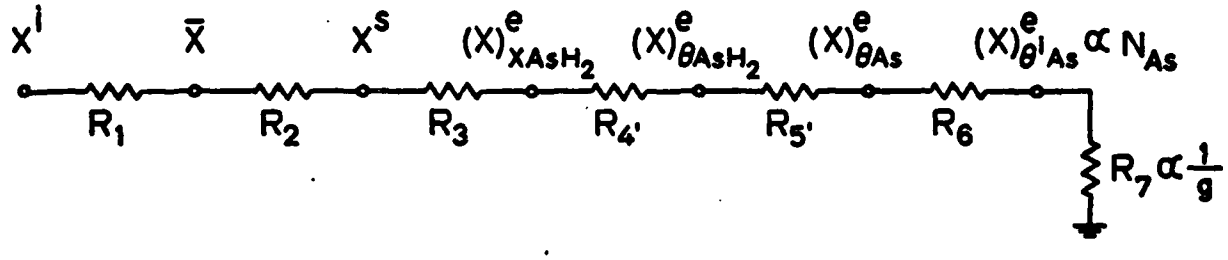



Fig. 3.9. Steady-state equivalent circuit assuming the $AsH_2$ path to be dominant. This path contains all of the possible mechanisms involved in the doping process.

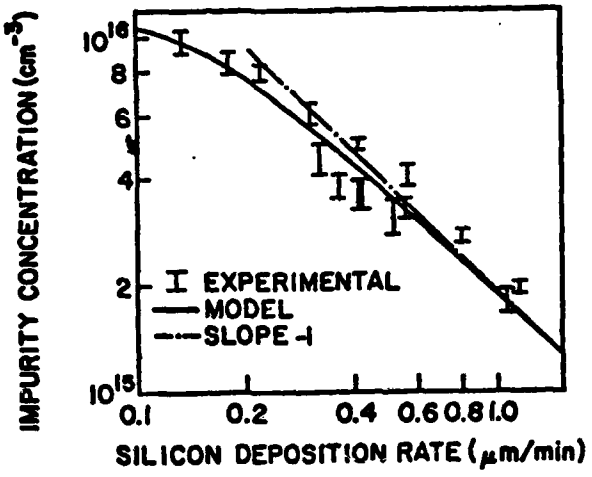



Fig. 3.10. Arsenic concentration of uniformly doped epitaxial layers as a function of growth rate.

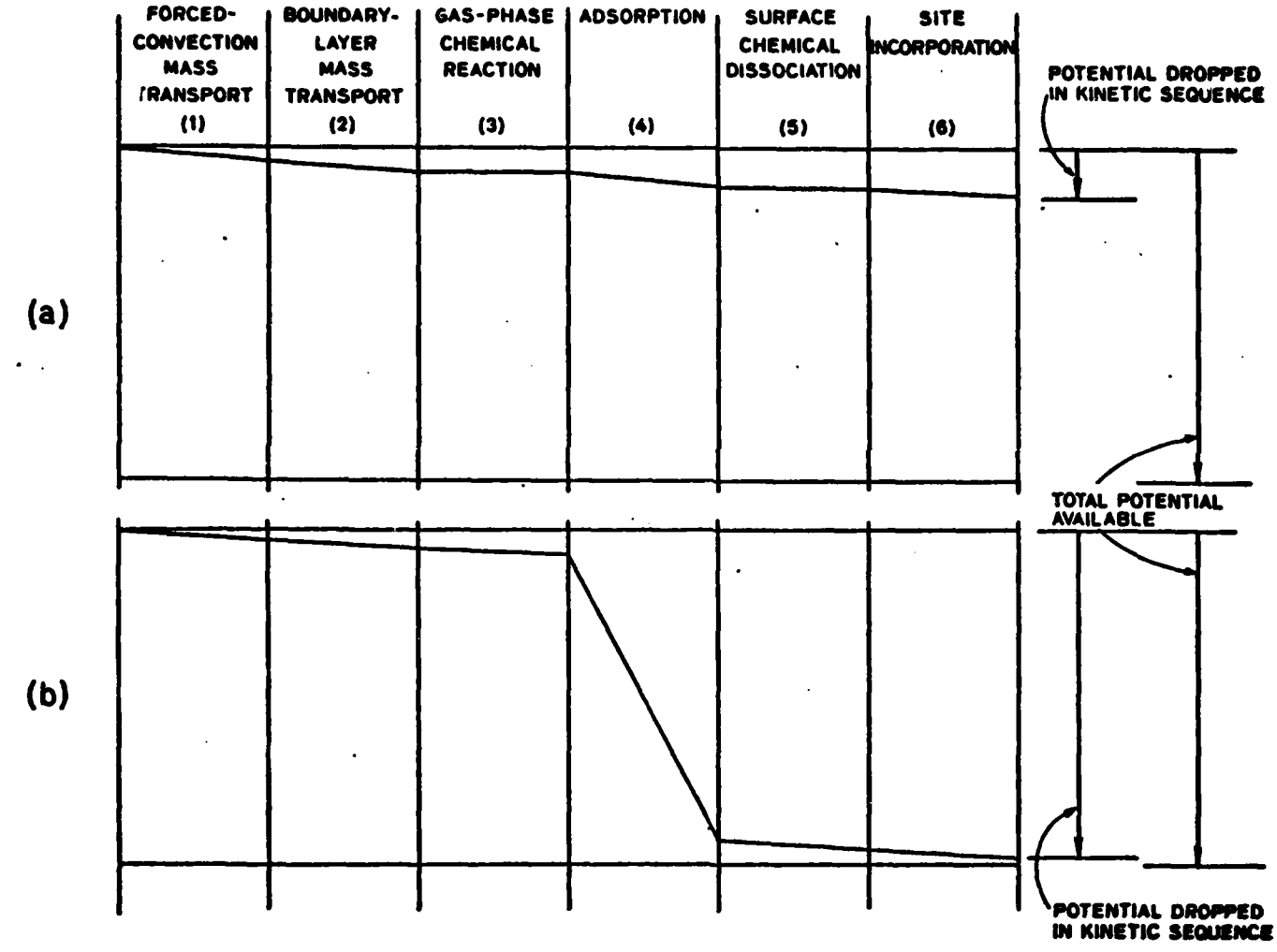



Fig. 3.11. Potential available to drive the dopant species towards the growing surface and potential dropped across each step of the kinetic sequence. (a) Low growth rates. (b) High growth rates.

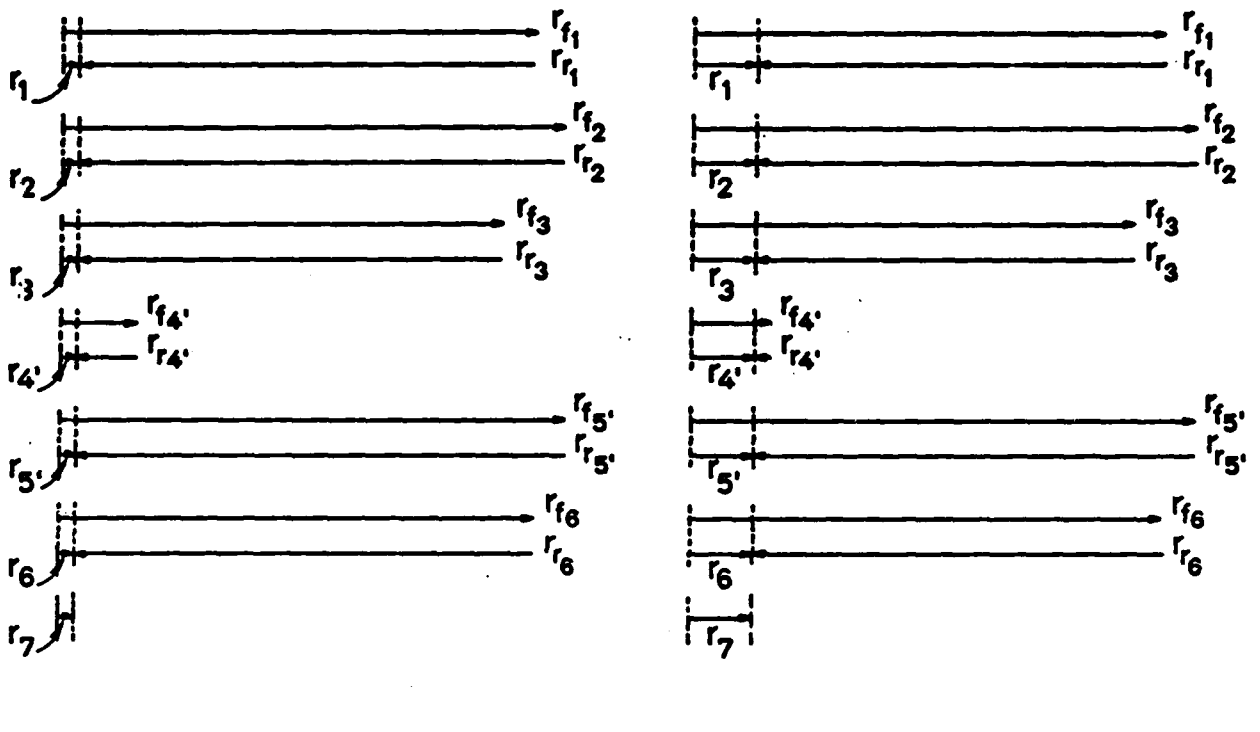



Fig. 3.12. Forward, reverse and net reaction rates for each step in the doping process. (a) Low growth rates. (b) High growth rates.

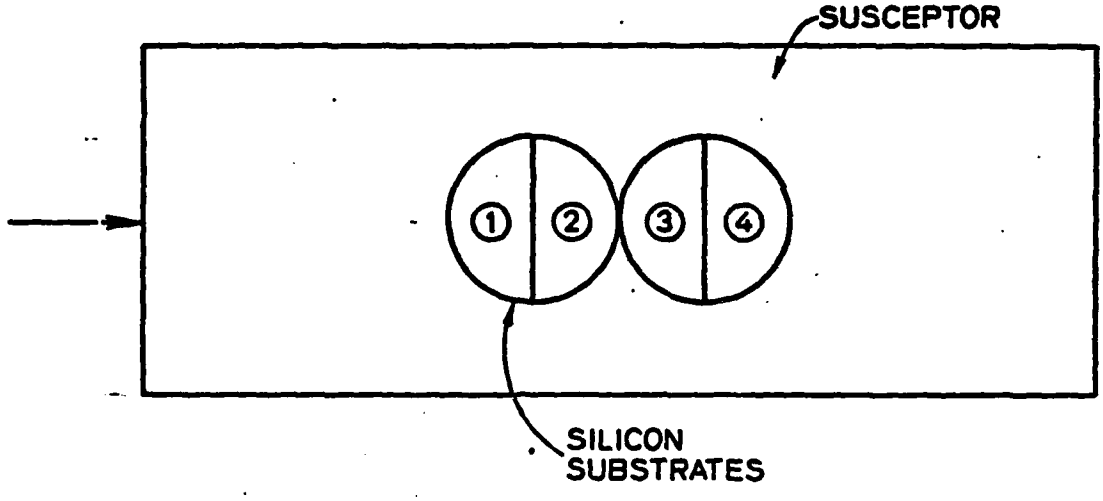
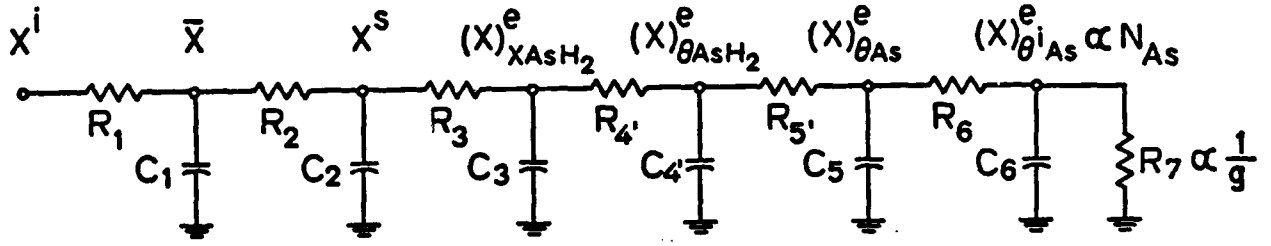
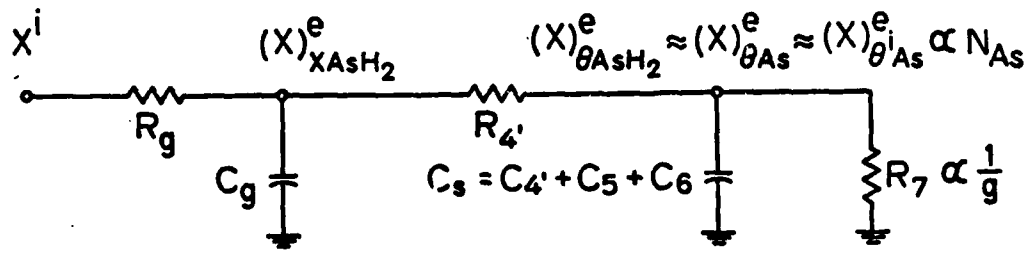



Fig. 3.13. Experimental arrangement. Substrates of (111) and (100) orientations were alternately placed on the susceptor. The arrow indicates the direction of gas flow.




(a)




(b)

Fig. 3.14. (a) Equivalent circuit of the doping process assuming the $AsH_2$ path to be dominant. (b) Simplified equivalent circuit assuming $R_{4'}$ to be the largest resistor.

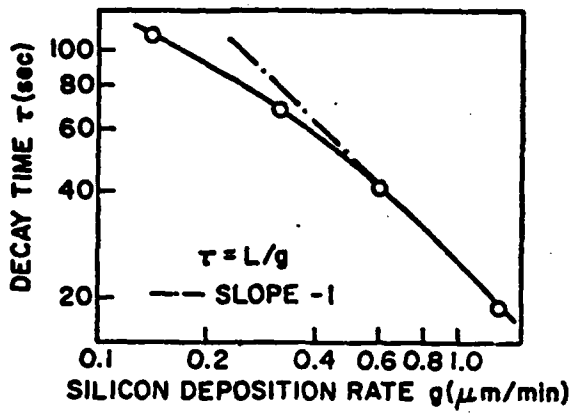



Fig. 3.15. Decay time of the transient response as a function of growth rate.

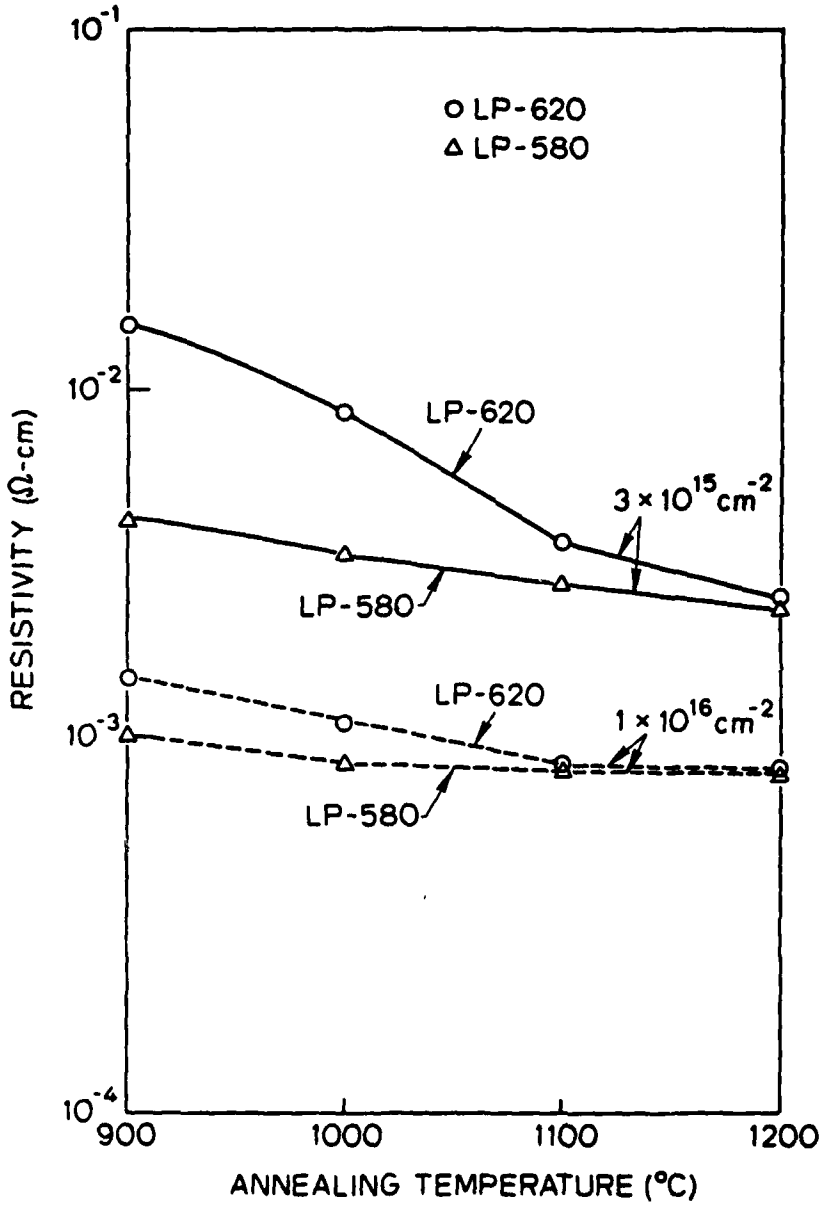



Fig. 3.16. Resistivity of low-pressure films deposited at 580° and 620°C as functions of annealing temperature for implanted phosphorus doses of $3 \times 10^{15}$ and $1 \times 10^{16}/cm^2$.

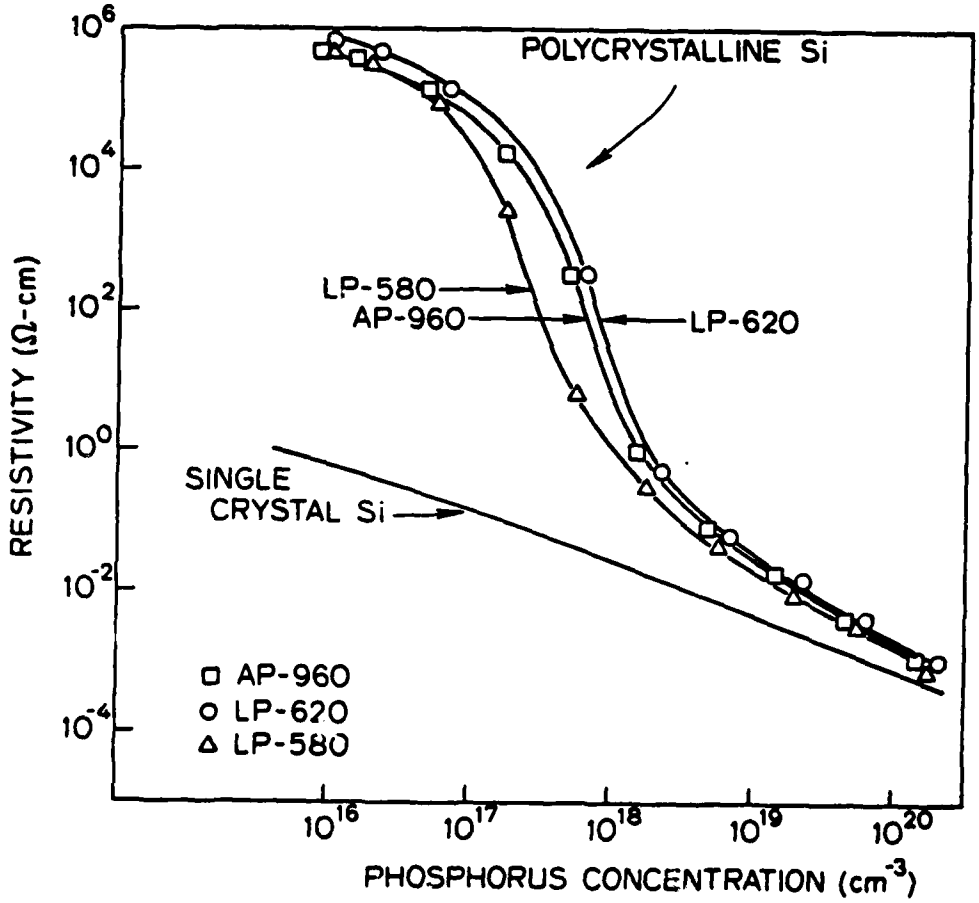



Fig. 3.17. Resistivity of phosphorus-implanted, low-pressure and atmospheric-pressure films as functions of average dopant concentration. The resistivity of n-type, single-crystal silicon is shown for comparison.

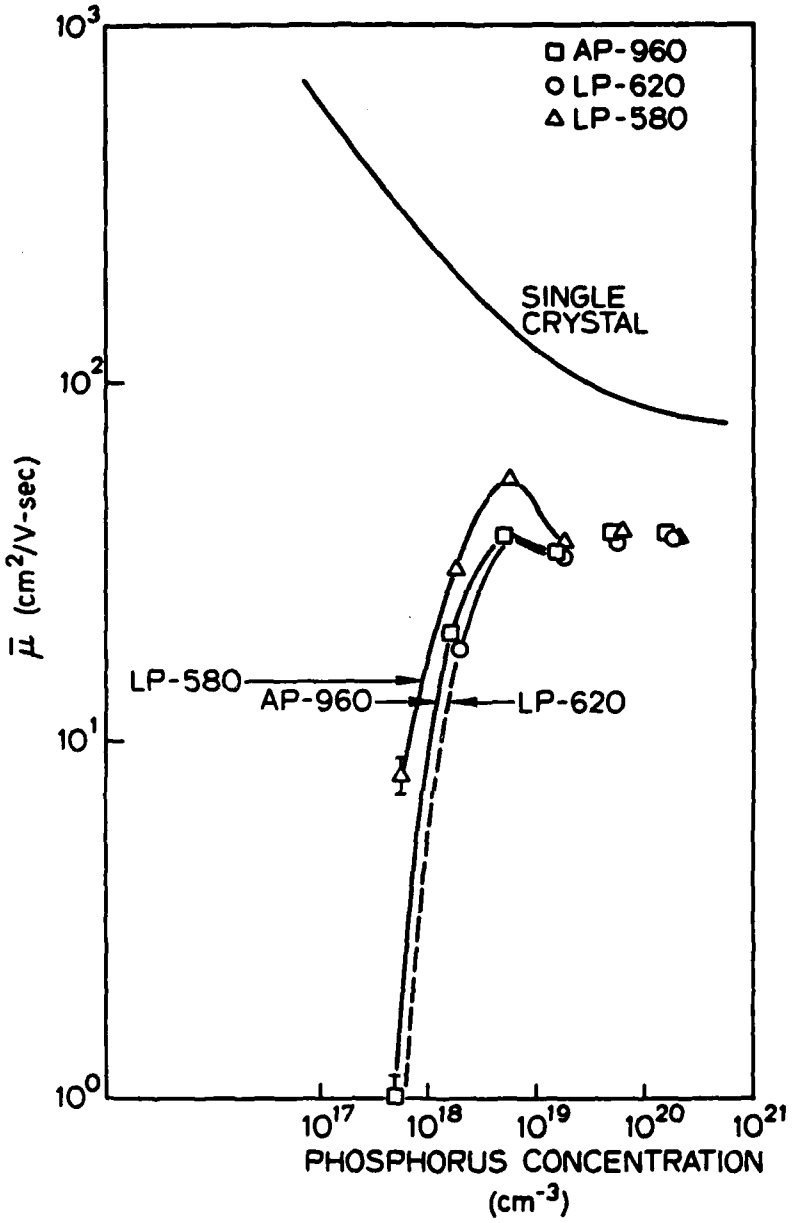



Fig. 3.18. Hall mobility of phosphorus-implanted, low-pressure and atmospheric-pressure films as functions of the average dopant concentration. The electron mobility in n-type, single-crystal silicon is shown for comparison. At a dopant concentration of $6.5 \times 10^{17}/cm^2$ the mobility in the LP-620 films was found to be less than 1 $cm^2$/V-sec as indicated by the dashed line.

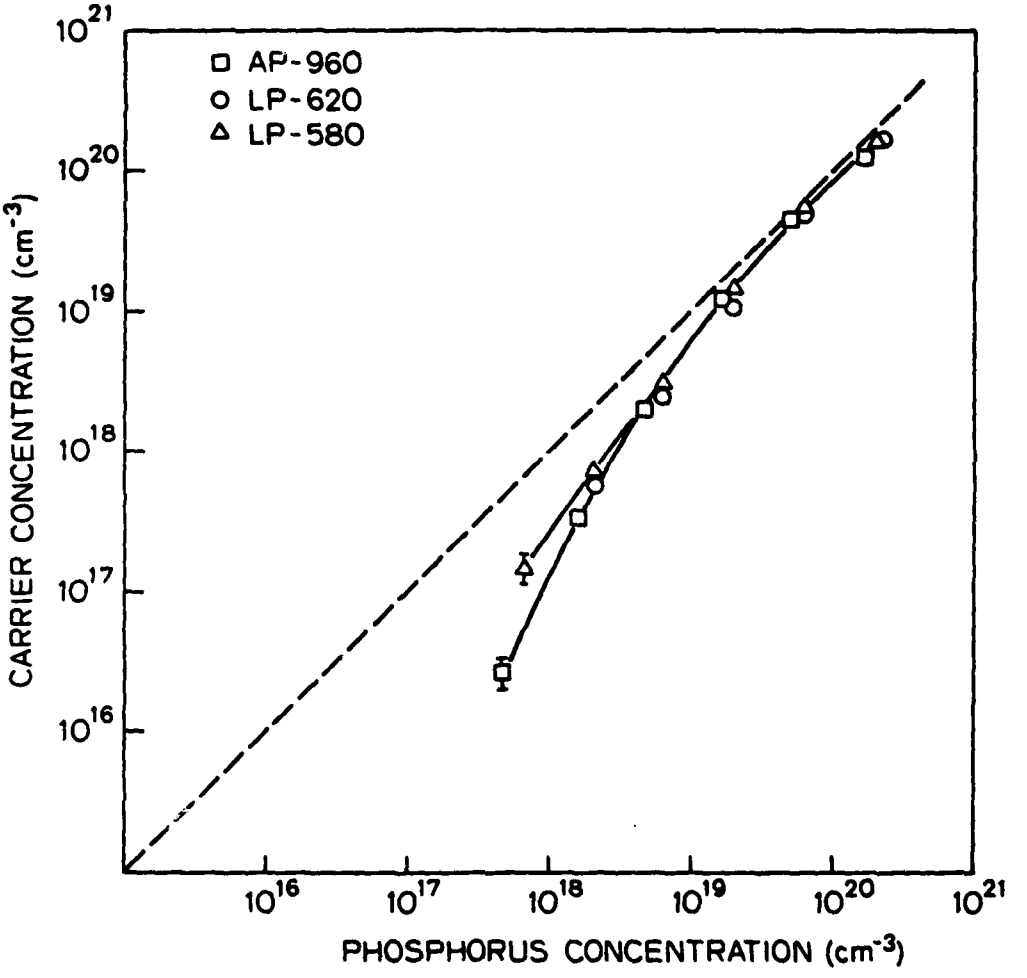



Fig. 3.19. Carrier concentrations from Hall measurements as functions of average dopant concentration in the films.

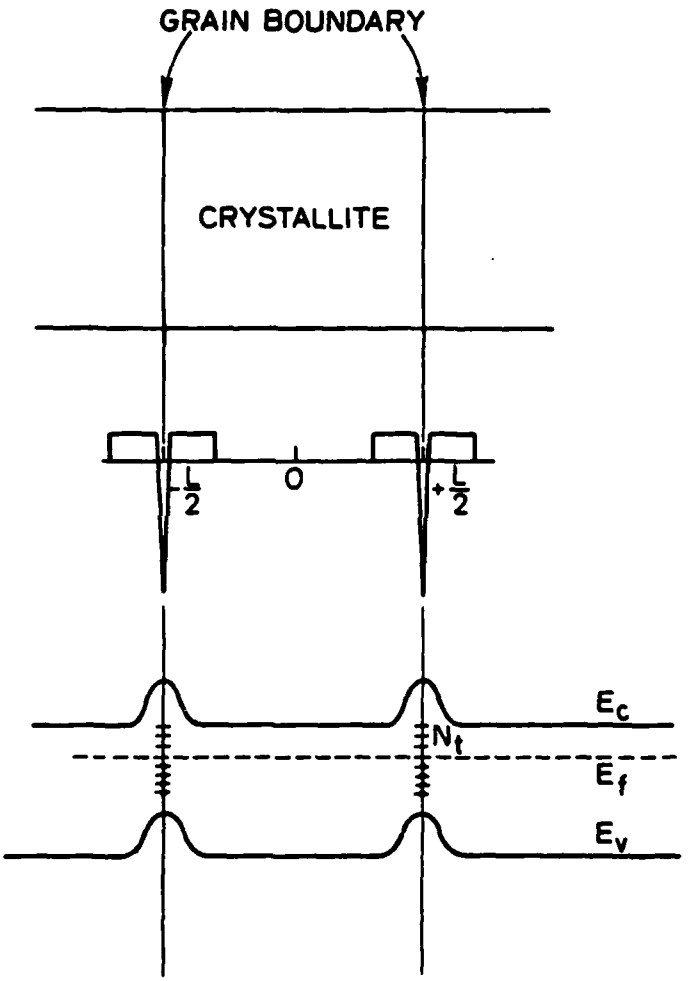



Fig. 3.20. Model of n-type polycrystalline material showing (a) small crystallites surrounded by grain boundaries containing large numbers of traps, (b) resulting space charge, and (c) corresponding energy-band diagram with potential barriers surrounding the grain boundaries.

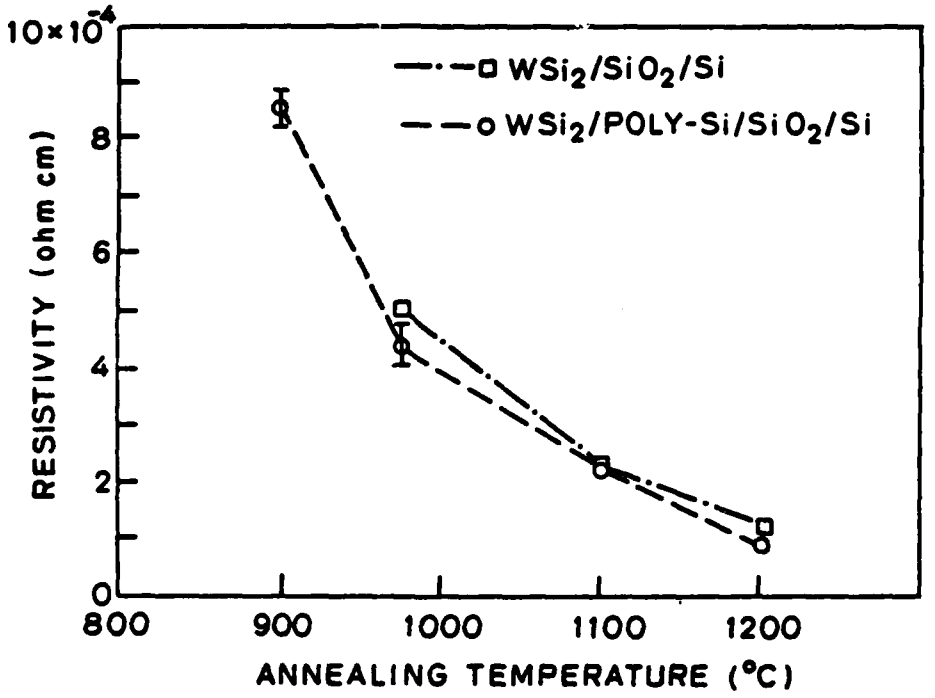



Fig. 3.21. Plot of variation of resistivity vs. annealing temperature for 30 min at $N_2$ ambient.

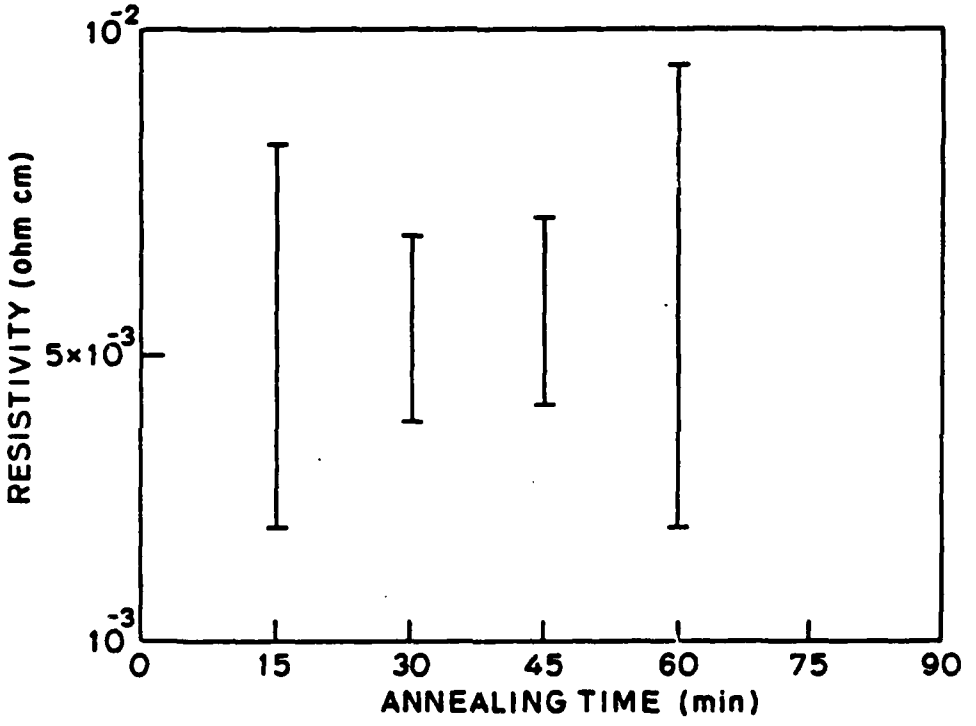



Fig. 3.22. Plot of variation in $WSi_2$/poly-Si/$SiO_2$/Si film resistivity vs. different annealing time for 650°C annealing temperature and $N_2$ ambient.

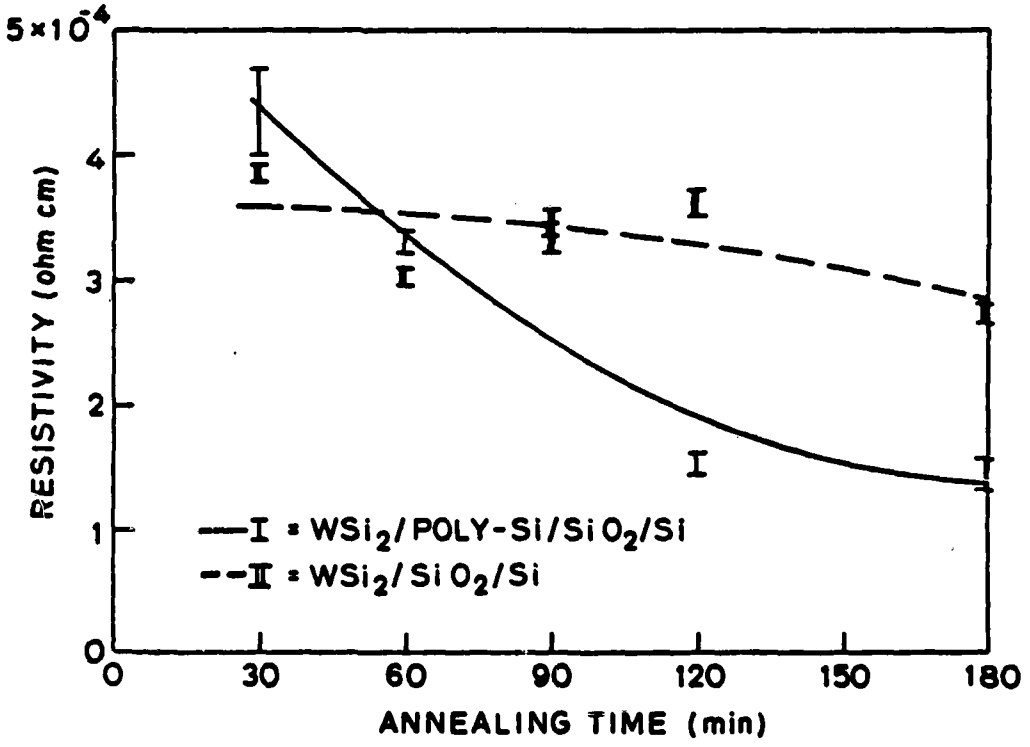



Fig. 3.23. Plot of variation in resistivity of $WSi_2$/poly-Si/$SiO_2$/Si and $WSi_2$/$SiO_2$/Si films vs. different annealing time for annealing temperature of 975°C at $N_2$ ambient.

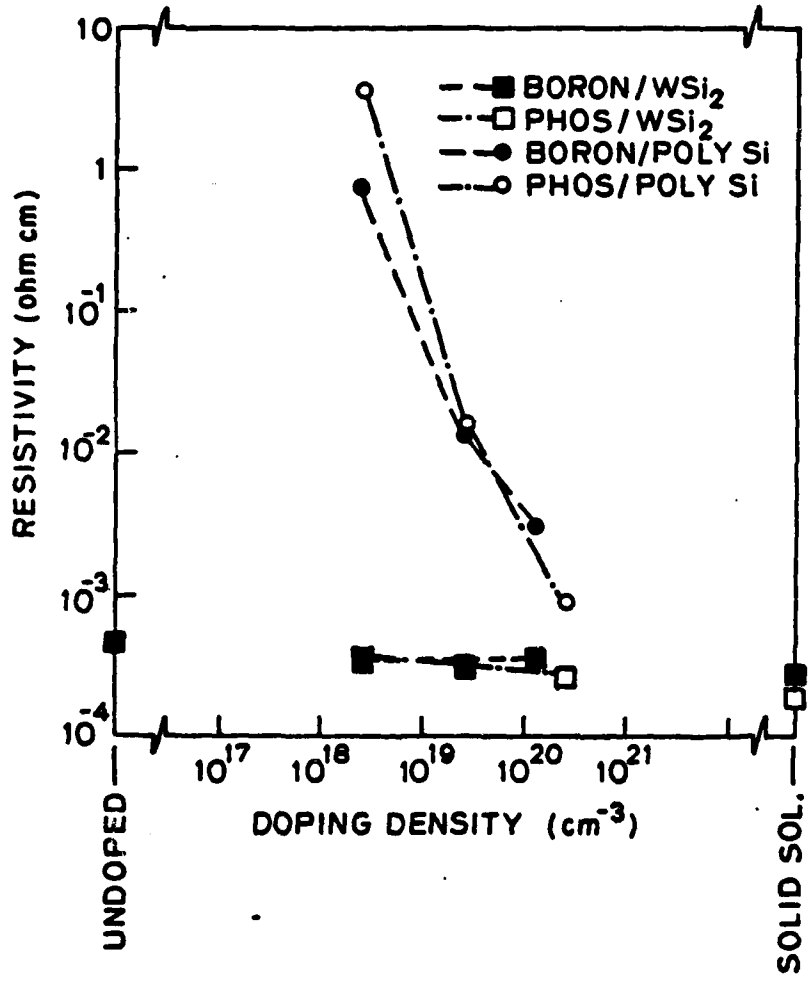
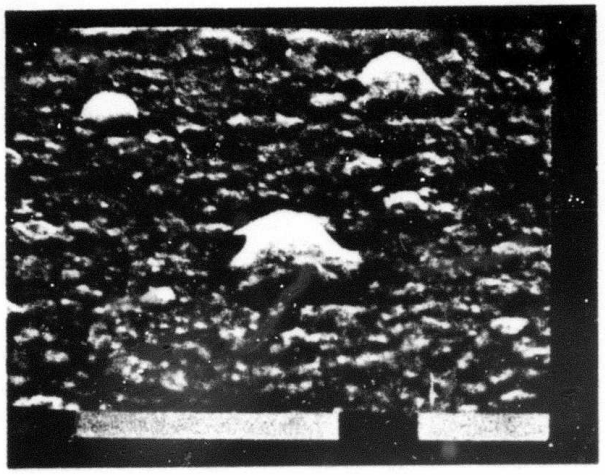
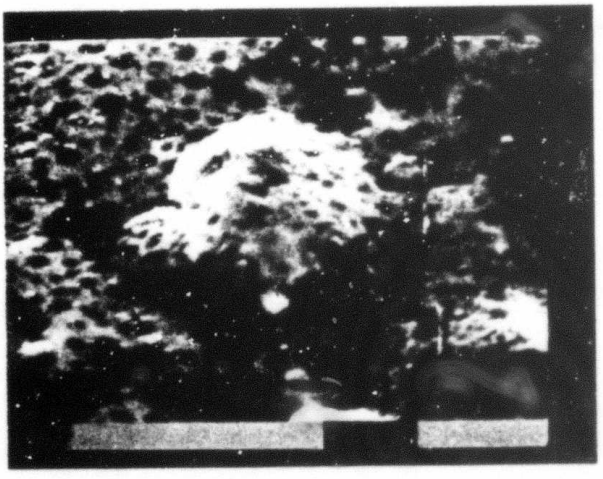



Fig. 3.24 Plot of variation in resistivity of $WSi_2$/doped poly-Si/$SiO_2$/Si and doped poly-Si/$SiO_2$/Si films vs. doping concentration.

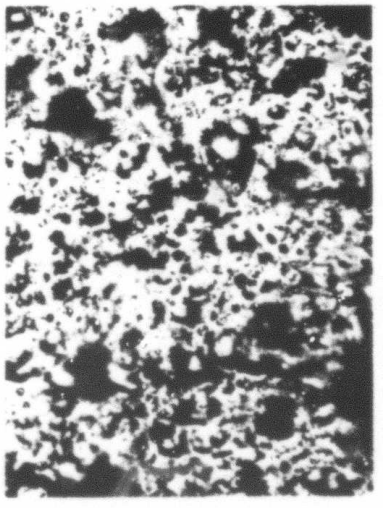


10000X
(a)

860X
(b)



10000X
(c)



860X
(d)

Fig. 3.25. SEM and optical photographs of Al/WSi2/SiO2/Si films annealed (a,b) below and (c,d) above critical temperature.

PART 4

# DEVICE SIMULATION AND STATISTICAL CIRCUIT MODELING

Robert W. Dutton

## 4.1 Introduction

Meindl et al [4.1] have outlined the overall objectives of the Computer Aided Engineering of Semiconductor Integrated Circuits activity within this project. Stated briefly the goals of the activities reported in this section are:

1) to couple into SUPREM [4.2] all process models discussed in the three preceding sections.
2) develop the analysis methodology and models needed to predict VLSI device performance -- both for CAD based circuit design and in the context of the statistical environment of production.

The general features of process simulation and advances in modeling capabilities are described elsewhere [4.1] [4.2]. However in the context of developments reported in the previous sections it is appropriate that major accomplishments in specific process modeling efforts will be reviewed in the next section. Within the second item of device analysis and modeling three major thrusts can be identified in subsequent sections -- the coupling of process and device analysis tools, test structures for characterization and modeling efforts oriented towards CAD.

### 4.1.2 Recent Advances in Process Modeling

The spectrum of activities described above in Sections 1-3 can be classified in two general categories -- those models which have been incorporated into SUPREM and those models still under development.

In the first category major accomplishments to be cited during this contract period include:

1) implementation of the point defect model for oxidation of heavily doped substrates
2) implementation of the surface kinetic dopant inclusion during epitaxial deposition

In the category of models under development several advances deserve honorable mention. The work in polycrystalline silicon is proceeding rapidly and suitable process models are expected within the next year. The multifaceted area of oxidation interface kinetics will shortly spawn several models. First order approximations for both stacking fault growth and unannealed values $Q_{ss}$ are soon to be reported. Moreover the partial pressure dependence of oxidation kinetics at both high and low pressures will also be modeled.

Turning to the device analysis and modeling efforts, the subsequent sections present specific accomplishments in tools, test structures and models. Sections 4.2 - 4.5 describe both one and two-dimensional device analysis. The one dimensional SEmiconductor Device ANalysis program (SEDAN) has clearly reached the point of a mature coupling to SUPREM. Subsequent sections dealing with two-dimensional poisson analysis, dc transport and grid generation each represent successively more formative efforts. Sections 4.6 - 4.7 describe details of two test chips and associated analytic approaches to determine performance limits as a function of geometry and technology variables. Sections 4.8 - 4.9 summarize major accomplishments in CAD modeling -- a capacitance model for MOS as well as a statistical bipolar device model -- in both cases experimental evidence validates the novel concepts which are reported.

Before embarking on the detailed discussions the following highlight contributions can be cited:

1. a complete and working one-dimensional SEDAN program which can use SUPREM data as input.
2. major benchmark results using two-dimensional MOS poisson solutions including technology dependent punchthrough characterization [4.14].
3. demonstration of a novel dc transport analysis method which offers an order-of-magnitude performance advantage over CADDET.
4. a CMOS latch-up test chip with structures and a new analysis strategy for field-aided lateral pnp parasitic devices [4.42].
5. continued improvement of lateral profile measurement methods and analysis tools [4.36].
6. experimental and 2D analytic corroboration of the change-oriented capacitance model
7. circuit and device level confirmation of the statistical modeling results based on model parameter correlations [4.3]

Having cited these accomplishments it is appropriate to unfold the details as given below.

## 4.2 ONE DIMENSIONAL SEMICONDUCTOR DEVICE ANALYSIS (SEDAN)

D. C. D'Avanzo and M. Vanzi

The following describes the formulation and results of a one-dimensional, numerical, semiconductor analysis program. First, the differential equations governing semiconductor physics are presented and normalized. The physical models for device parameters are described. Simulation results are presented for NPN bipolar transistor structure. The results dependent on several physical models are presented and compared among them and with measured data. The results of a time dependent analysis are shown, starting from $0^-$ to steady state after having applied a bias to the base and collector terminals and for several time increments. Finally a typical bias condition for electron and hole concentrations and a high level injection condition are compared and discussed.

### 4.2.1 Fundamental Equations and Normalization

The electrical properties of a semiconductor device can be completely specified by five physical relationships; Poisson's equation, electron and hole transport equations and electron and hole continuity equations. With the appropriate boundary conditions the coupled equations can be solved for carrier concentrations, current densities and electrostatic potential. The five differential equations are:

Poisson

$$\frac{d^2\psi}{dx^2} = \frac{-q}{\epsilon_{si}} \{p - n + N\} \qquad (4.2.1)$$

Transport

$$J_n = q \mu_n n E_x + q D_n \frac{dn}{dx} \quad (4.2.2)$$

$$J_p = q \mu_p p E_x - q D_p \frac{dp}{dx} \quad (4.2.3)$$

Continuity

$$\frac{dn}{dt} = \frac{1}{q} \frac{dJ_n}{dx} - U \quad (4.2.4)$$

$$\frac{dp}{dt} = -\frac{1}{q} \frac{dJ_p}{dx} - U \quad (4.2.5)$$

The one-dimensional Poisson equation relates the second derivative of the electrostatic potential, $\psi$, to the hole, p, electron, n, and net impurity, N (donor minus acceptor) concentrations, where q is electronic charge and $\varepsilon_{si}$ is the dielectric constant of silicon. The transport equations include both drift and diffusion in the total electron, $J_n$, and hole $J_p$, current densities. The electron, $\mu_n$, and hole, $\mu_p$, mobilities are related to their respective diffusion constants, $D_n$ and $D_p$ by the Einstein relation,

$$\mu_n = D_n/V_{kT} \quad (4.2.6)$$

$$\mu_p = D_p/V_{kT} \quad (4.2.7)$$

where $V_{kT}$ is the thermal voltage. The mobilities are dependent on total impurity concentration and electric field. The x component of the electric field is defined as the negative spatial derivative of the electrostatic potential,

$$E_x = -\frac{d\psi}{dx} \quad (4.2.8)$$

The continuity equations are time dependent and the recombination rate, U, includes both single level Shockley-Read-Hall and Auger recombination mechanisms.

For convenience in the derivation and numerical calculation all quantities are normalized to dimensionless form by appropriate constants. The normalization factors, and their values and units are listed in Table 4.2.1. The normalized device equations are:

$$\frac{d^2\psi}{dx^2} = - p + n - N \tag{4.2.9}$$

$$J_n = \mu_n \, n \, E + \mu_n \frac{dn}{dx} \tag{4.2.10}$$

$$J_p = \mu_p \, p \, E + \mu_p \frac{dp}{dx} \tag{4.2.11}$$

$$\frac{dn}{dt} = \frac{dJ_n}{dx} - U \tag{4.2.12}$$

$$\frac{dp}{dt} = - \frac{dJ_p}{dx} - U \tag{4.2.13}$$

where all variables and parameters are dimensionless. (For simplicity the subscript on the x-component of the electric field has been dropped.

The discussion will be limited to the npn bipolar transistor whose impurity profile and potential diagrams are shown in Fig. 4.2.1. However the method is applicable to any number and combination of impurity layers with two or three contacts. For the bipolar structure the positions of the contacts are x=0 for the emitter, x=B for the base and x=L for the collector, where L is the total length of the device being analyzed. B must be located in the neutral region of the base where the majority carrier quasi-fermi level is relatively constant.

### 4.2.3 Physical Models

#### A. Mobility

The mobilities of electrons and holes decrease as the probability of scattering increases. In bulk semiconductors two mechanisms contribute to increases in scattering probabilities; the presence of impurity atoms, including donors and acceptors, and the increase in mobile carrier energy due to high electric fields. To account for these effects hole and electron mobilities are expressed as empirical functions of the total impurity concentration and the electric field,

$$(x) = \left\{ \mu_{min} + \frac{\mu_{max} - \mu_{min}}{1 + \left(\frac{N_T(x)}{N^-}\right)^K} \right\} \left\{ 1 + E(x)/E_c \right\}^{-1} \quad (4.2.14)$$

where $N_T$ is the total impurity concentration, values of the empirical parameters, $\mu_{min}$, $\mu_{max}$, $N^-$, K and $E_c$, the critical field for velocity saturation, are listed in Table 4.2.2.

#### B. Band Gap Narrowing

When the impurity concentration is greater than $10^{17} cm^{-3}$ electron wave functions associated with the impurity energy levels begin to overlap forming bands which in turn overlap the intrinsic conductions and valence band edges. The spread of energy levels causes localized band gap narrowing and as a result an increase in the effective intrinsic electron concentration, $n_{ie}$. Slotboom [4.4] has shown that this increase can be accurately modeled by,

$$n_{ie}(x) = n_i \exp\left\{ \frac{V_1}{2V_{kT}} \left( \ell n \frac{N(x)}{N_o} + \sqrt{\left( \ell n \frac{N(x)}{N_o} \right)^2 + C} \right) \right\} \quad (4.2.15)$$

where $n_i$ is the intrinsic electron concentration in the absence of band gap narrowing, $V_1$ = 9.0 mV, $N_o = 10^{17} cm^{-3}$, and C = 0.5.

The presence of non-uniform band edges introduces an additional force on mobile carriers which must be included in the transport equations. Van Overstraeten and others have shown that this effect can be incorporated as an additional quasi-electric field component [4.5]. The expressions for the electron and hole current densities become,

$$J_n = q n \mu_n \left\{ \frac{-d\psi}{dx} - \frac{1}{q} \frac{d\,DEC}{dx} \right\} + q D_n \frac{dn}{dx} \qquad (4.2.16)$$

$$J_p = q p \mu_p \left\{ - \frac{d\psi}{dx} + \frac{1}{q} \frac{d\,DEV}{dx} \right\} - q D_p \frac{dp}{dx} \qquad (4.2.17)$$

where DEC and DEV are in general complicated functions of the band structure. However, if Boltzmann statistics apply, the DEC and DEV can be interpreted as the changes in the conduction and valence band edges due to heavy doping effects.

In equilibrium the electron concentration is given by,

$$n = n_{ie} \exp(\psi / V_{kT}) \qquad (4.2.18)$$

since the fermi level is constant and equal to zero for the particular voltage reference adopted in this work. Eq (4.2.18) can be substituted into (4.2.16) with $J_n$ = 0 in equilibrium. Applying the Einstein relation and rearranging results in a convenient expression for the gradient of DEC,

$$\frac{d\,DEC}{dx} = \frac{1}{n_{ie}} V_{kT}\, q \frac{d\, n_{ie}}{dx} \qquad (4.2.19)$$

since the change in band edge due to heavy doping is independent of bias voltage, the previous result can be substituted directly in (4.2.16)

to obtain,

$$J_n = q \mu_n n E_{tot_n} + q D_n \frac{dn}{dx} \tag{4.2.20}$$

where

$$E_{tot_n} = - \frac{d}{dx}\left(\psi + V_{kT} \ell n \frac{n_{ie}}{n_i}\right) \tag{4.2.21}$$

A similar expression applies for holes where,

$$E_{tot_p} = - \frac{d}{dx}\left(\psi - V_{kT} \; n \frac{n_{ie}}{n_i}\right) \tag{4.2.22}$$

In summary, two important physical effects arise from narrowing of the band gap by heavy doping. First, the effective intrinsic concentration becomes a function of the net impurity concentration. Second, an additional force acts on mobile carriers as a result of the non-uniform band edges. When Boltzmann statistics are valid this force is simply related to the gradient of $n_{ie}$ and can be incorporated in the transport equation as an additional field component.

C. Recombination

The recombination rate includes both trapping of hole and electron pairs in recombination centers, Shockley-Read-Hall effect, and the three particle process known as Auger recombination. The total recombination rate, U, can be expressed as the sum of the rates contributed by the two mechanisms,

$$U = U_{SRH} + U_A \tag{4.2.23}$$

where $U_{SRH}$ and $U_A$ are the recombination rates for Shockley-Read-Hall and Auger processes respectively.

For single level traps the Shockley-Read-Hall recombination rate is,

$$U_{SRH} = \frac{(np - n_{ie}^2)}{\tau_p (n + n_1) + \tau_n (p + p_1)} \tag{4.2.24}$$

where $n_1 = n_{ie} \exp [(E_t/q - \psi)/V_{kT}]$ and $p_1 = n_{ie} \exp [(\psi - E_t/q)/V_{kT}]$. $E_t$ is the energy level of the trapping center referenced to the intrinsic fermi level and $\tau_p$ and $\tau_n$ are the hole and electron lifetimes.

The Auger process, which is the reverse of impact ionization, consists of a hole and electron pair recombining, with the excess energy transferred to a third carrier as kinetic energy. The three particle process can be modeled by, the expression,

$$U_A = C_n(n^2p - n\, n_{ie}^2) + C_p\,(p^2n - p\, n_{ie}^2) \qquad (4.2.25)$$

where $C_n$ and $C_p$ are experimentally determined coefficients for n-type and p-type material, respectively.

Recent experimental results have demonstrated that the overall minority carrier lifetime is a strong function of impurity concentration. (For a summary of experimental results see reference [4.6 ]. This effect can be attributed at least partially if not entirely to the Auger process. However, since the SRH rate is proportional to trap density some SRH lifetime dependence on impurity concentration can be expected. In this case the hole and electron lifetimes can be expressed as,

$$\tau_p = \frac{\tau_{po}}{1 + \frac{N_T}{N_{op}}} \qquad (4.2.26)$$

$$\tau_n = \frac{\tau_{no}}{1 + \frac{N_T}{N_{on}}} \qquad (4.2.27)$$

The empirical constants, $\tau_{po}$, $\tau_{no}$, $N_{op}$, $N_{on}$, will generally depend on the specific process. Typical values for the recombination constants obtained from a recent literature search, [4.7 ], are tabulated in Table 4.2.3.

### 4.2.4 Results

The previous sections have described the numerical techniques and physical models implemented in the one-dimensional device analysis program SEDAN. The general formulation allows the analysis of a wide variety of one-dimensional device structures with either two or three contacts. The information obtained from the device analysis includes values of the dependent variables, p, n, $\psi$, $J_n$ and $J_p$, as well as any parameters which can be derived from the independent variables such as electron and hole quasi-Fermi levels, terminal currents, junction capacitance, integrated charge etc. The following examples demonstrate the utility of the program as applied to an npn transistor and an MOS capacitor.

A. Npn Transistor

The first example is of an epitaxial transistor with a 0.5 μm base. The impurity profile is contructed analytically from three Gaussian diffusions and a constant epitaxial layer. The impurity profile and carrier concentrations for moderate, $V_{BE}$ = 0.8 V, and high level, $V_{BE}$ = 1.0 V, injection are shown in Fig. 4.2.2. For the high level case, the concentration of electrons injected into the base greatly exceeds the impurity concentration, resulting in a fairly uniform potential distribution across the transistor. To maintain the condition of quasi-neutrality the hole concentration increases to the electron level. The ability to simulate the high level injection results from the flexibility of the base boundary condition at the base contact where only the majority carrier quasi-fermi level is specified.

The effect of high level injection on terminal currents is demonstrated in Fig. 4.2.3. When carrier concentrations exceed the base impurity concentration the effective base width increases, commonly known as base pushout, resulting in a decrease in collector current, Fig. 4.2.3a. The base current, which is determined by injection of holes into the emitter, continues to increase exponentially so that the normalized current gain decreases, Fig. 4.2.3b.

The sensitivity of the terminal currents to the physical models is demonstrated in Fig. 4.2.4. The collector, $J_C$, and base, $J_B$, current densities are plotted as functions of the base emitter voltage, $V_{BE}$, in Fig. 4.2.4a, while the current gain, $\beta = J_C/J_B$, versus collector current is plotted in Fig. 4.4.4b. The results have been simulated with and without band gap narrowing and with three different recombination mechanisms; Shockley-Read-Hall with constant lifetime, SRH; Shockley-Read-Hall with constant lifetime plus Auger, SRH-A; and Shockley-Read-Hall with concentration dependent lifetime, SRH - $\tau(N_t)$.

As can be seen in Fig. 4.4.4a the collector current is insensitive to the physical perturbations since it is mainly determined by injection of electrons into the relatively lightly doped base where both band gap narrowing and concentration dependent recombination are insignificant. The changes in current gain are directly attributable to changes in the base current. The effects of the physical models can be easily interpreted in reference to the two major components of base current; injection of holes into the emitter which displays a slope proportional to $1/V_{kT}$ on the semi-log scale, and recombination in the base-emitter space charge region with a slope proportional to $1/2V_{kT}$. The lowest values of $J_B$

and highest values of β are obtained with no band gap narrowing and SRH recombiantion. When Auger is added to SRH, $J_B$ increases and β decreases at high bias voltages. This results from an increase in the injection of holes into the emitter due to the increased recombiantion rate associated with high emitter carrier densities. At low bias levels recombination in the base-emitter space charge region dominates the base current. Since carrier concentrations are low, the effect of Auger at low bias is insignificant. However when concentration dependent lifetime is included with SRH, $J_B$ increases for all bias voltages. In this case the lifetime depends on the total impurity concentration so that recombination increases significantly even in the base-emitter space charge region.

When band gap narrowing is included the base currents increase for all values of bias voltage and the slopes of the curves approach the ideal limit proportional to $1/V_{kT}$. Both observations can be attributed to a significant increase in hole injection into the emitter due to the large values of $n_{iE}$. Sensitivity to recombination mechanisms is reduced in the presence of band gap narrowing since injection of holes into the emitter dominates the base current over most of the bias range. Similar conditions can be drawn from corresponding decreases in current gain.

The transient responses of electron and hole distributions to instantaneous increases in base and collector voltages are shown in Fig. 4.2.5. Electrons in the base region reach a steady state within approximately 0.1 μsec. while holes and electrons in the collector and holes in the emitter require 1.0 μsec. The results indicate that the time limiting mechanism in response to an increase in base voltage is

accumulation of stored charge in the emitter, which is hindered by a retarding field. On the other hand minority carrier storage in the base is enhanced by an aiding field. The transient response to increases in collector voltage is limited by depletion of both holes and electrons in the collector.

B. MOS Capacitor

The analysis of an MOS capacitor requires only the solution of Poisson's equation perpendicular to the oxide interface since steady state current can be neglected. The numerical analysis avoids the assumptions often included in analytical methods such as the depletion and strong inversion approximation. Accurate calculations of mobile and depletion charges are possible for arbitrarily doped substrates.

There are several considerations in the numerical implementation of an MOS structure which differ from the bipolar example. First, the dissimilarity in dielectric constants across the oxide-silicon interface ($x=t_{ox}$ in Fig. 4.2.6) must be accounted for by including a normalized relative permittivity, $E_r$, in Poisson's equation. The spatial variation of $E_r$ can be defined by,

$$E_r(i) = \begin{cases} \dfrac{\varepsilon_{ox}}{\varepsilon_{si}} & \text{for } i < I_t \\ 1 & \text{for } i \geq I_t \end{cases} \qquad (4.2.28)$$

where $I_t$ is the grid node corresponding to the oxide thickness, and $\varepsilon_{ox}$ and $\varepsilon_{si}$ are the permittivities of the oxide and silicon respectively.

The boundary conditions are also modified for the MOS structure. The metal or gate contact is placed at x=0 and the substrate contact at x=L. The reference voltage is taken as the majority carrier quasi-

fermi level in the substrate at x=L. The effect of the bias voltages on the quasi-fermi levels in a p-type substrate is shown schematically in the energy band diagram of Fig. 4.2.6. The gate to source voltage, $V_{GS}$, displaces the metal fermi level with respect to the minority carrier quasi-fermi level while the source to bulk substrate $V_{SB}$ displaces the minority level with respect to the majority level. The potential at the substrate contact is determined from equilibrium. The boundary conditions are summarized by,

$$\phi_p(x) = 0 \tag{4.2.29}$$

$$\phi_n(x) = V_{SB} \tag{4.2.30}$$

$$\psi(0) = \psi(L) + V_{GB} \tag{4.2.31}$$

$$\psi(L) = \ell n \left\{ \sqrt{1.0 + \frac{N(L)^2}{4n_i^2}} + \frac{N(L)}{2n_i} \right\} \tag{4.2.32}$$

where $V_{GB}$ is the gate to bulk voltage.

The one-dimensional analysis is especially useful for capacitors. on non-uniformly doped substrates, which cannot be described analytically. An important practical example is the threshold shifting implant in an MOS transistor. Fig. 4.2.7 shows the electron and hole concentrations for several gate voltages in a typical surface implanted, p-substrate transistor. Fig. 4.2.8 shows the total electron charge, $Q_n$, and the total net charge, $Q_{NET}$, as a function of surface potential for the same structure. The total charge is calculated by integrating the charge densities from the surface to the bulk. The net charge is the absolute value of $N_A$ + n-p. The three operational regions, accumulation, depletion and strong inversion, are clearly visible in Fig. 4.2.8. Both results

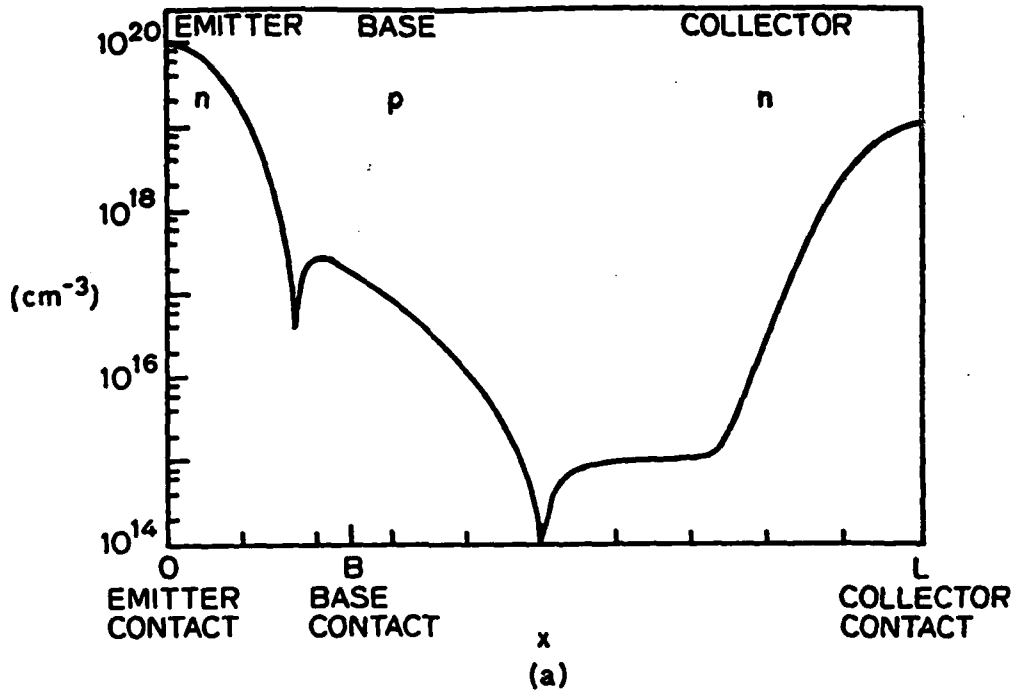
will aid in the prediction of a device threshold as a function of the implanted profile.

### 4.2.5 Summary

A detailed description of a one-dimensional, transient, semiconductor analysis program has been presented. Fundamental equations were described and numerical procedures were discussed. Applications were presented including high level and transient simulations of a bipolar transistor and threshold analysis of a non-uniformly doped MOS-capacitor.

The program is capable of analyzing any one dimensional structure with two or three contacts. In addition arbitrary profile shapes and extensive ranges of bias voltages can be accomodated. Second order physical models, such as band gap narrowing, concentratrion and field dependent mobility, and Shockley-Read-Hall and Auger recombination are included.

The general formulation facilitates the addition and modification of physical models, boundary conditions and physical parameters. For example avalanche multiplication, radiation effects, and photo-response could be simply included as additional terms in the recombination-generation rate in the continuity equations. The restrictions on the external boundary conditions can be relaxed by including a current boundary condition with finite recombination velocity. Non-uniformly distributed base current could be simulated by inserting position dependent majority carrier generation sites throughout the base region. In conclusion, the program has been designed to evaluate the effects of physical models and parameters, and device structures on the performance of semiconductor transistors and diodes.



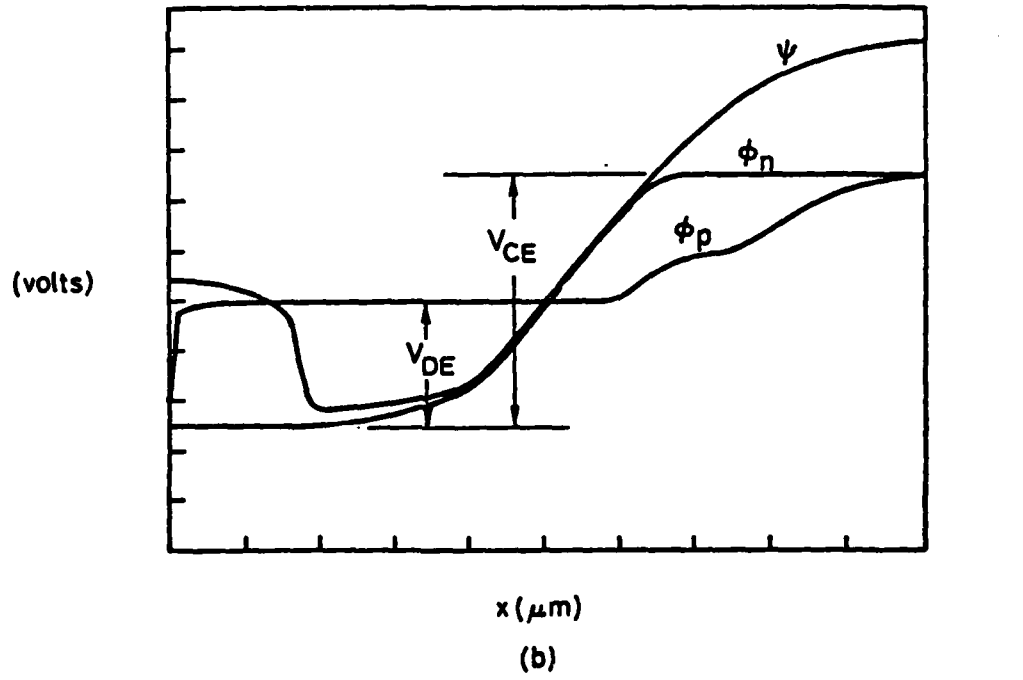




Fig. 4.2.1 Impurity profile, (a), and potential distributions, (b), for a typical npn bipolar transistor.

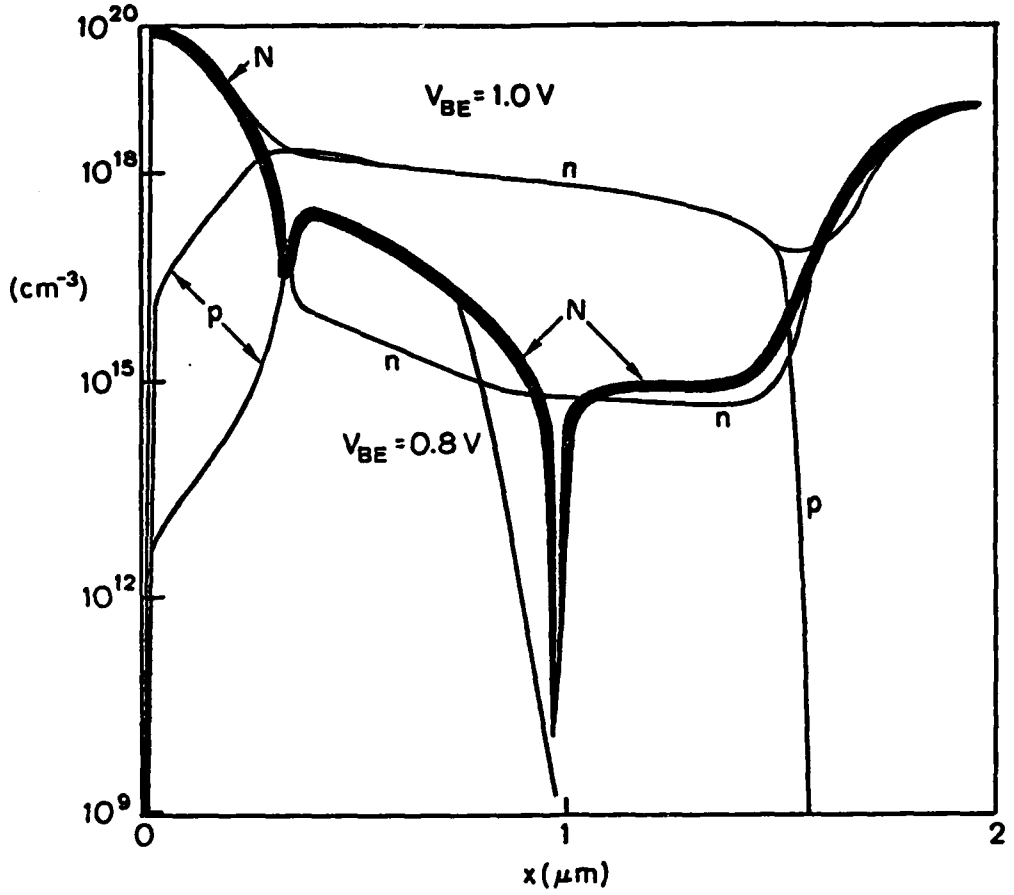
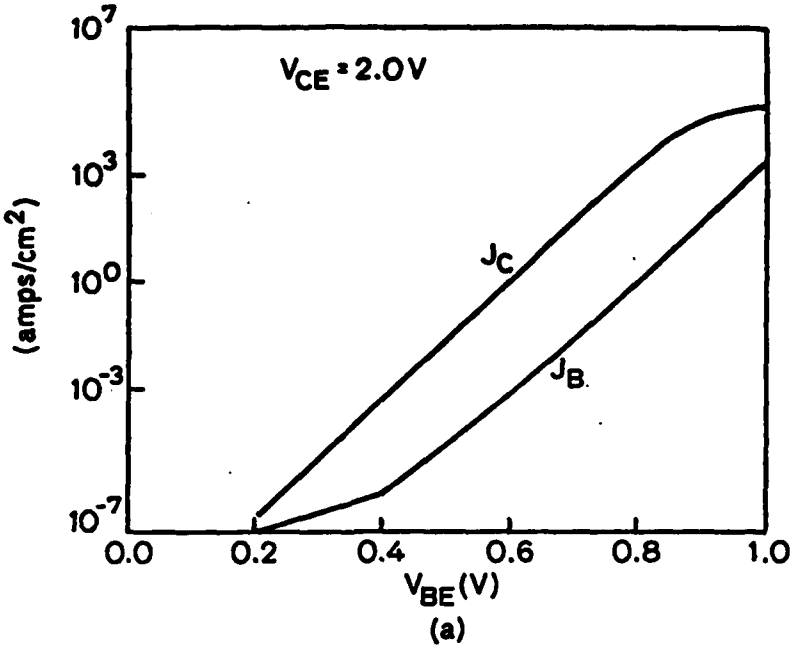



Fig. 4.2.2 Electron, n, and hole, p, concentrations for moderate and high level injection.



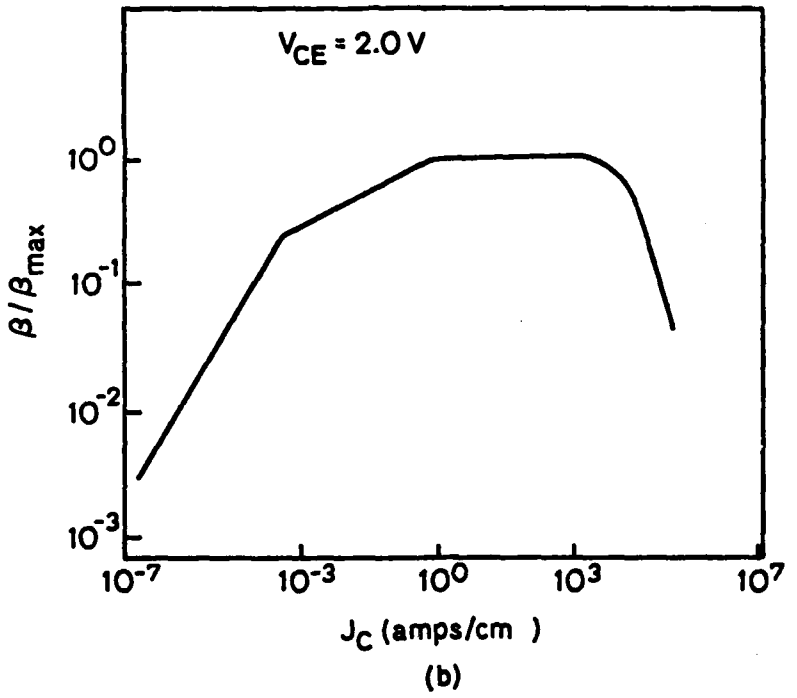
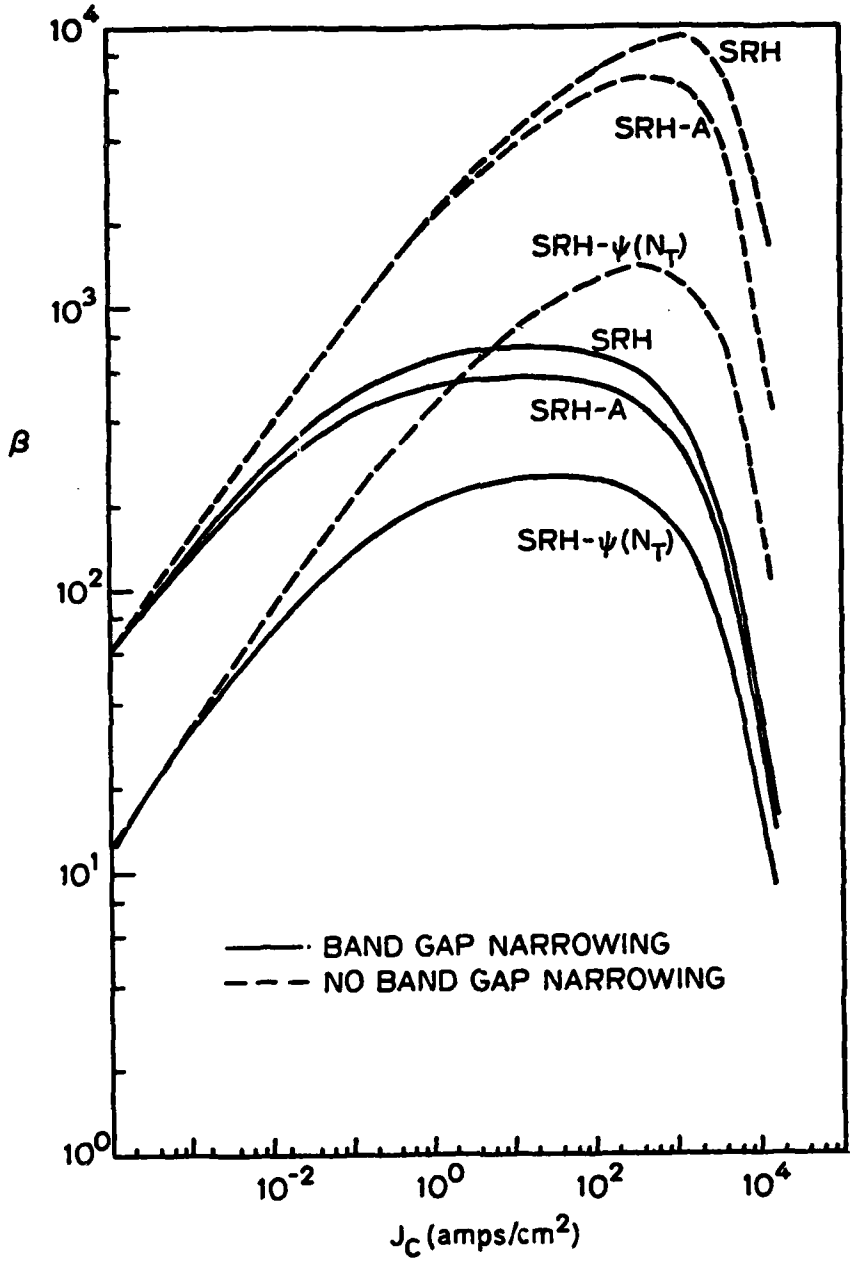




Fig. 4.2.3 Collector, $J_c$, and base, $J_B$, current densities as a function of base emitter voltage, (a), and normalized current gain, $\beta/\beta_{max}$, versus collector current density.




(a)

Fig . 4.2.4a Effect of band gap narrowing and recombination mechanisms on collector and base current densities.

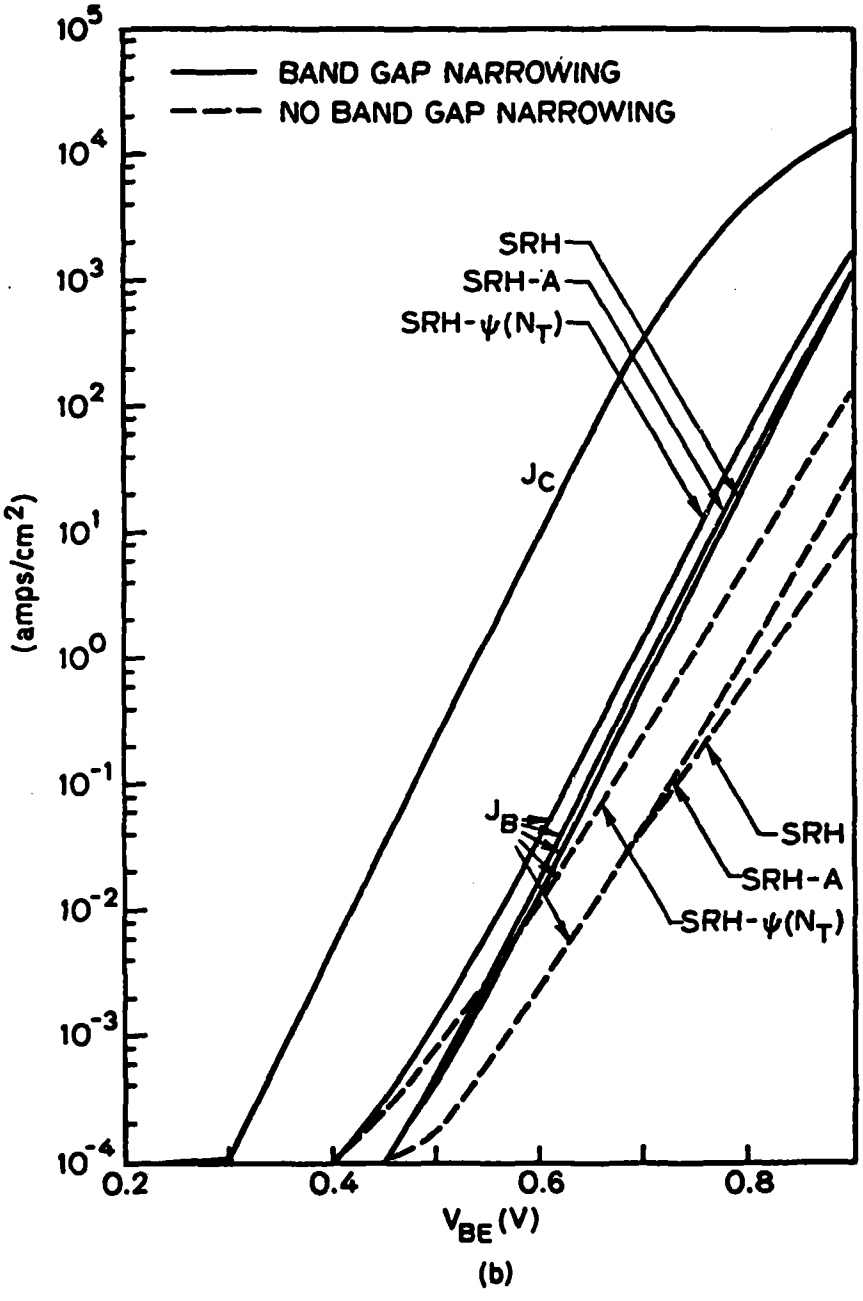
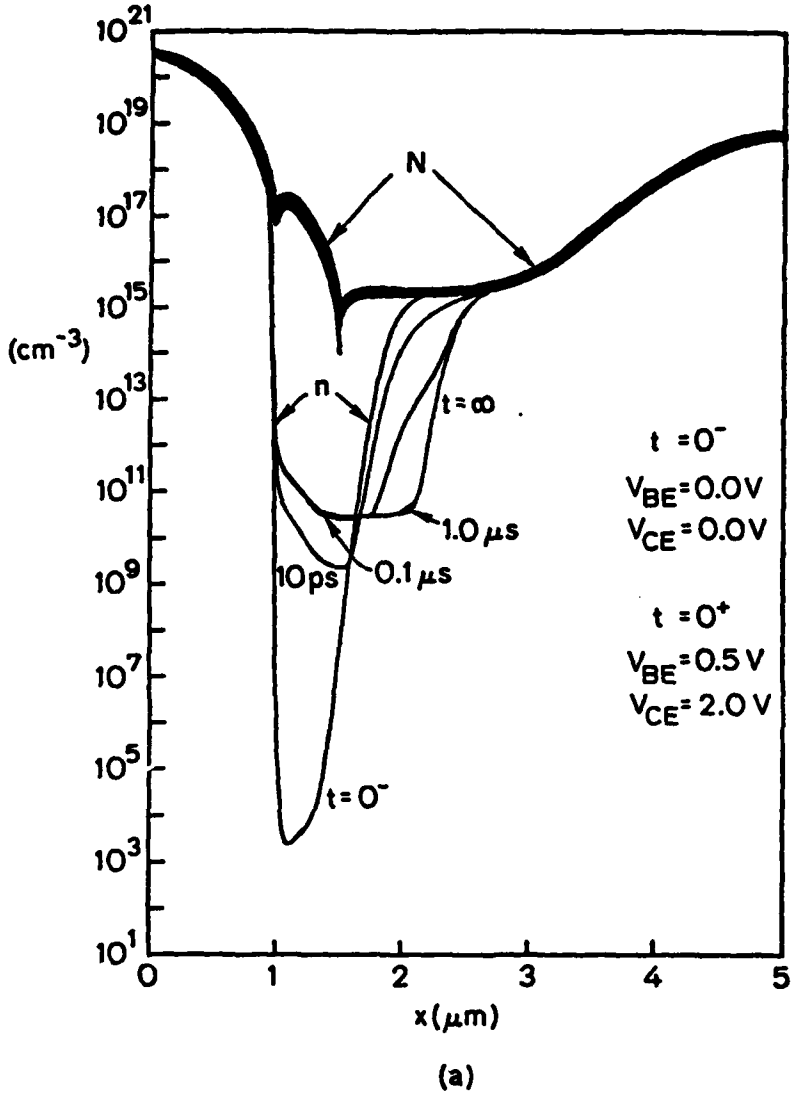
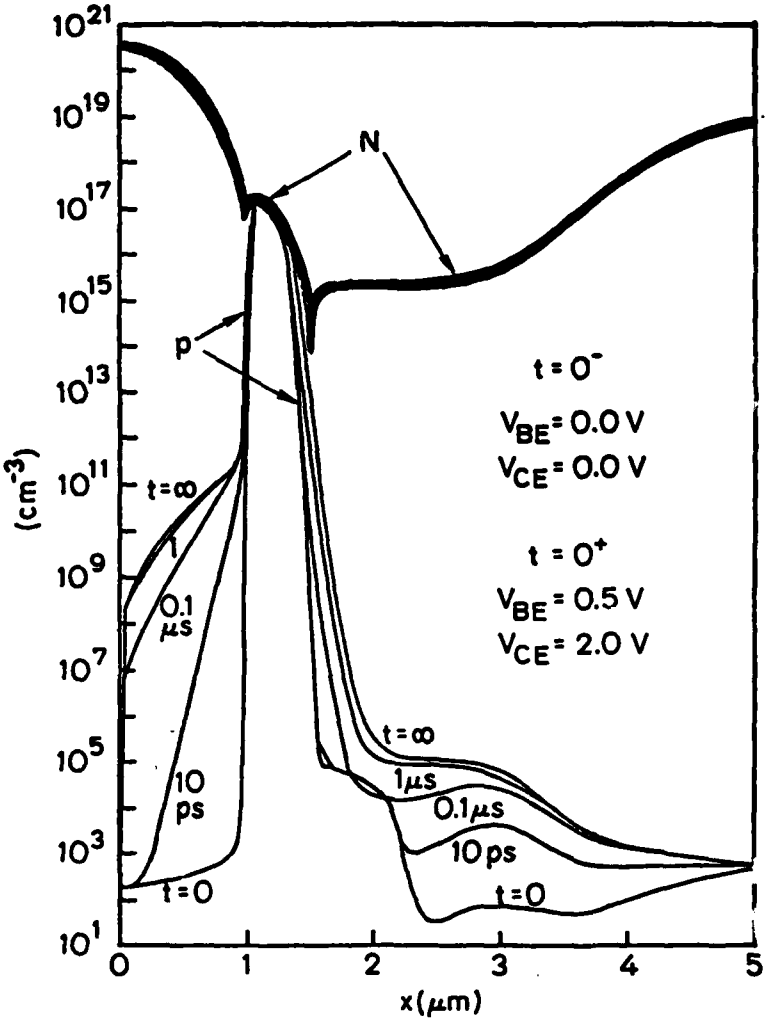



Fig. 4.2.4b Effect of band gap narrowing and recombination mechanisms on current gain.




(a)

Fig.4.2.5a Transient responses of a bipolar transistor to an abrupt change in base and collector voltages; electron concentration.




(b)

Fig. 4.2.5b Transient response of a bipolar transistor to an abrupt change in base and collector-voltages; hole concentration.

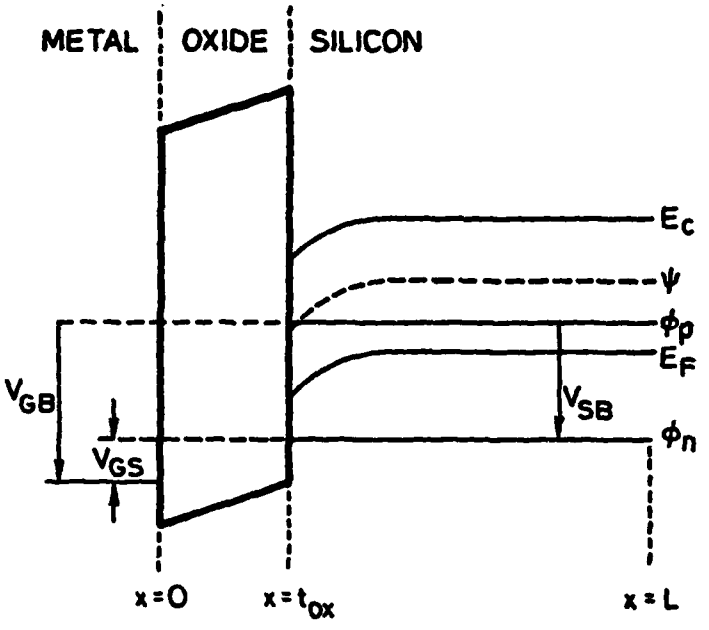



Fig. 4.2.6 Band diagram for an MOS capacitor with p-type substrate.

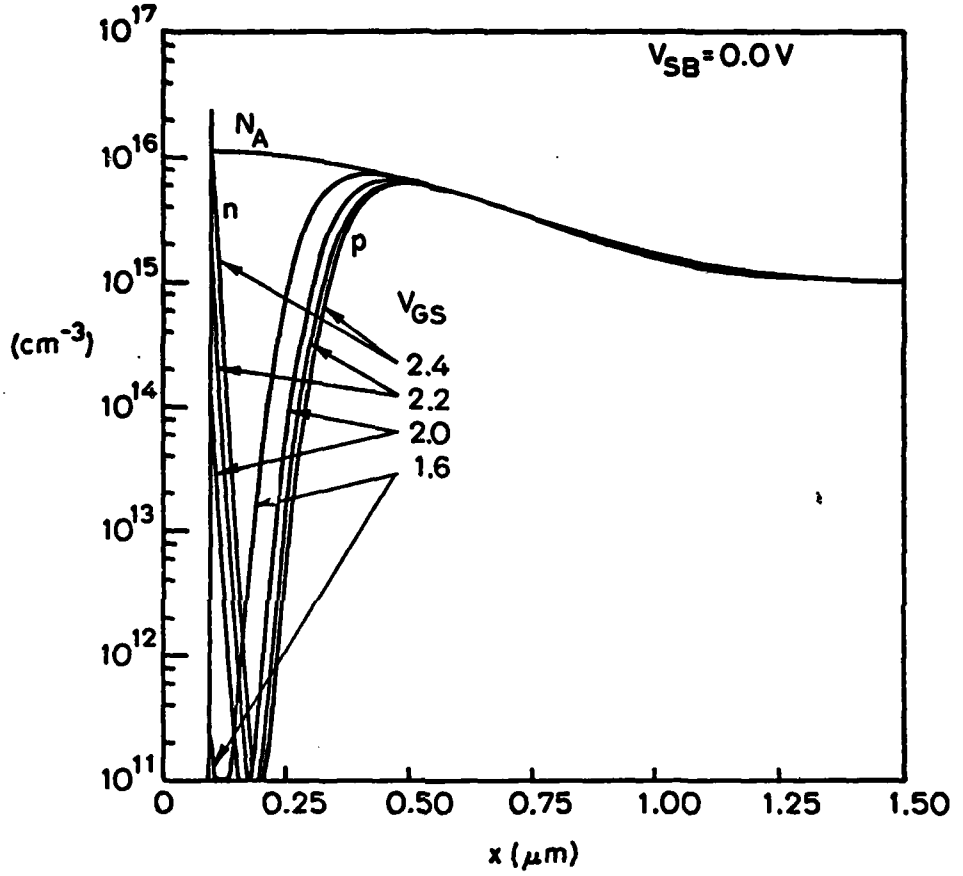



Fig. 4.2.7 Electron and hole distributions as functions of gate to source bias for a non-uniformly doped, p-type, MOS capacitor.

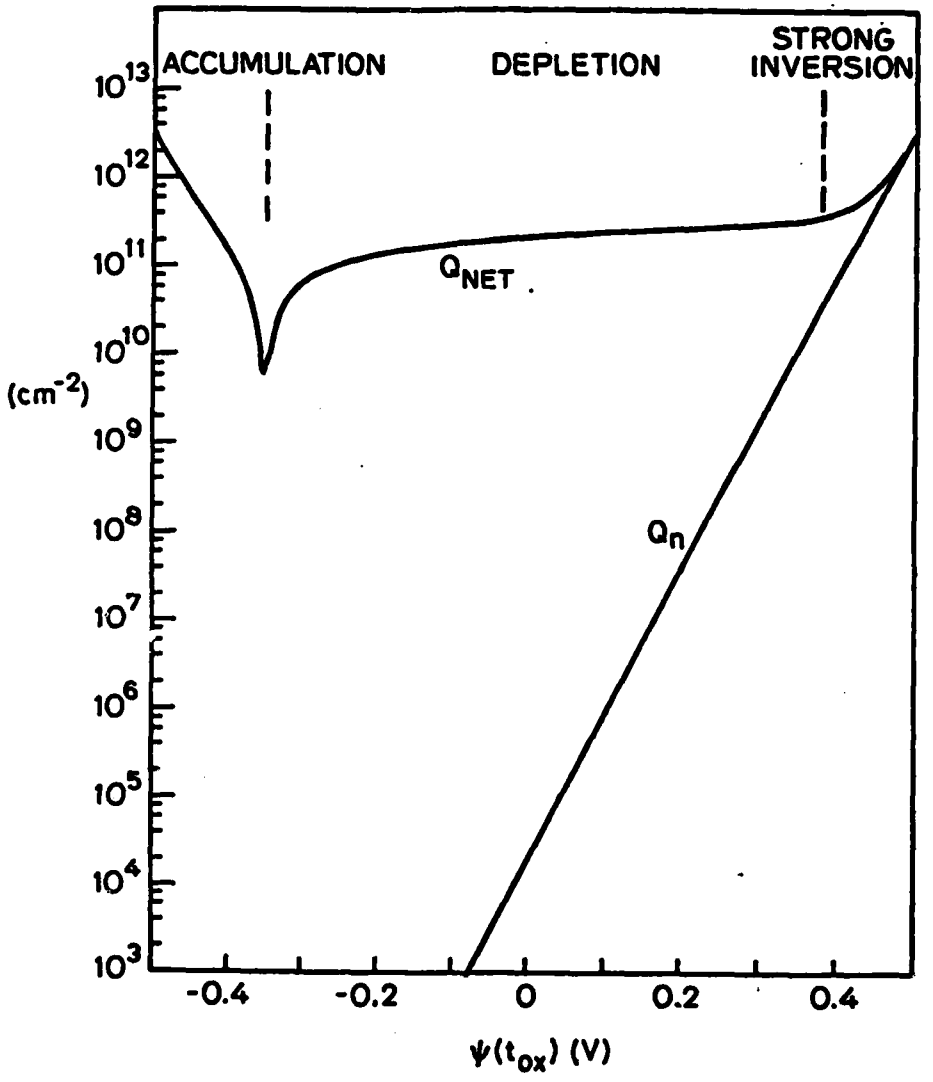



Fig. 4.2.8 Total electron and net charge versus surface potential for the MOS capacitor of Figure 4.7.

TABLE 4.2.1

Normalized Parameters, Normalizing Factors and Physical Constants

| Normalized quantity | | Normalization factor | | |
|---|---|---|---|---|
| Description | Symbol | Symbol | Value | Units |
| Position coordinant | $y$ | $L \equiv \sqrt{V_T \varepsilon_{si}/q\, n_{io}}$ | $3.405 \times 10^{-3}$ | cm |
| Electrostatic potential | $\psi$ | $V_T \equiv kT/q$ | .02586 | volts |
| Quasi-fermi levels | $\phi_n, \phi_p$ | $V_T$ | .02586 | volts |
| Electric field | $E$ | $V_T/L_I$ | 7.595 | V/cm |
| Carrier densities and effective intrinsic electron concentration | $n, p, n_{iE}$ | $n_{io}$ | $1.45 \times 10^{10}$ | $cm^{-3}$ |
| Net, donor and acceptor densities | $N, N_A, N_D$ | $n_{io}$ | $1.45 \times 10^{10}$ | $cm^{-3}$ |
| Carrier mobilities | $\gamma_n^{-1}, \gamma_p^{-1}$ | $\mu o$ | 1.0 | $cm^2/V\text{-sec}$ |
| Recombination, electron and hole current densities | $J_r, J_n, J_p$ | $J_o \equiv q\mu_o n_{io} V_T/L_I$ | $1.762 \times 10^{-8}$ | $\frac{coul}{sec\text{-}cm^2}$ |
| Recombination rate | $U$ | $U_o \equiv \mu_o V_T n_{io}/L_I^2$ | $3.234 \times 10^{13}$ | $1/sec\text{-}cm^3$ |
| Physical Constants | | | | |
| Boltzmann's constant | $k$ | | $8.62 \times 10^{-5}$ | ev/°K |
| Temperature | $T$ | | 300 | °K |
| Electronic charge | $q$ | | $1.6 \times 10^{-19}$ | coul. |
| Si permittivity | $\varepsilon_{si}$ | | $1.04 \times 10^{-12}$ | Farad/cm |
| $SiO_2$ permittivity | $\varepsilon_{ox}$ | | $3.3 \times 10^{-12}$ | Farad/cm |

TABLE 4.2.2
Mobility Parameters

| | $\mu_{max}$ $cm^2/V\text{-}sec$ | $\mu_{min}$ $cm^2/V\text{-}sec$ | $N^*$ $cm^{-3}$ | k - | $E_c$ V/cm |
|---|---|---|---|---|---|
| n | 1350 | 130 | $5 \times 10^{17}$ | 0.72 | 7,396 |
| p | 475 | 90 | $3 \times 10^{16}$ | 0.76 | 20,000 |

TABLE 4.2.3
Recombination Parameters

| | $C_{Auger}$ $cm^6\ sec^{-1}$ | $N_{SRH}$ $cm^{-3}$ | $\tau_{o\ SRH}$ sec |
|---|---|---|---|
| n | $2.8 \times 10^{-31}$ | $5 \times 10^{16}$ | $.5 \times 10^{-6}$ |
| p | $9.9 \times 10^{-32}$ | $5 \times 10^{16}$ | $.5 \times 10^{-6}$ |

## APPENDIX 4.2.1

Program Description

The SEDAN data structure is divided into an integer, a real simple precision and a real double precision working arrays. Furthermore, several real double precision arrays contain the input profile, the dependent variables values and all the informations derived from the dependent variables. Description of the content of some of these arrays follows:

| | |
|---|---|
| PSI | potential values for each grid points. |
| CN | net impurity concentration (positive n-type, negative p-type) |
| DY | grid spacing between two grid points. |
| NC | electron concentration |
| PC | hole concentration |
| TNP & TNN | concentration dependent lifetimes for Shockley-Read-Hall recombination |
| CNIE | effective intrinsic carrier concentration |
| MUN & MUP | mobility values for electrons and holes |

Besides these double precision arrays the are six real single precision arrays containing bias dependent results:

| | |
|---|---|
| JBA | base terminal current density |
| JCA | collector terminal current density |
| VBEA | base emitter bias voltages |
| VCEA | collector emitter bias voltages |
| VCBA | collector-base bias voltages |
| BCJC | base emitter or base collector junction capacitance. |

SEDAN is composed of the main program and 30 subroutines:

| | |
|---|---|
| SEDAN | main |
| BEGAN | effective intrinsic carrier concentration (band gap narrowing) calculation |
| BIAS P | input bias processor |
| CALD | main matrix solution. Simultaneous solution for $\psi$, n, p. |
| CJUN | junction capacitance calculation |
| COMST | input comments processor |
| DERFC | error function calculation |
| DEVIP | input device cord processor |
| EQUID | Poisson equilibrium solution |
| ERSET | syntax errors processor |
| FGRID | input grid processor |
| GETLN | reads a line of input |
| GETRL | reads numbers from input cards |
| GUMM | calculates base and emitter Gummel numbers |
| IMPZ | calculates the input profile when specified analytically |
| ISCOM | compares two strings of characters |
| LGTOP | gets logical values from input cards |
| MODEP | input model card processor |
| NORML | calculates the tail area of the standardized normal curve |
| PARSE | reads the input line and determines the values of parameters |
| PRINP | print card processor |
| RSUPZ | reads the file saved in the program SUPREM and interpolates the impurity concentration according to the new grid specification |

SDATS controls the output

SEINP controls the input

SESYN checks the syntax

SHEE calculates sheet resistivities

SINIT initializes and sets boundary conditions

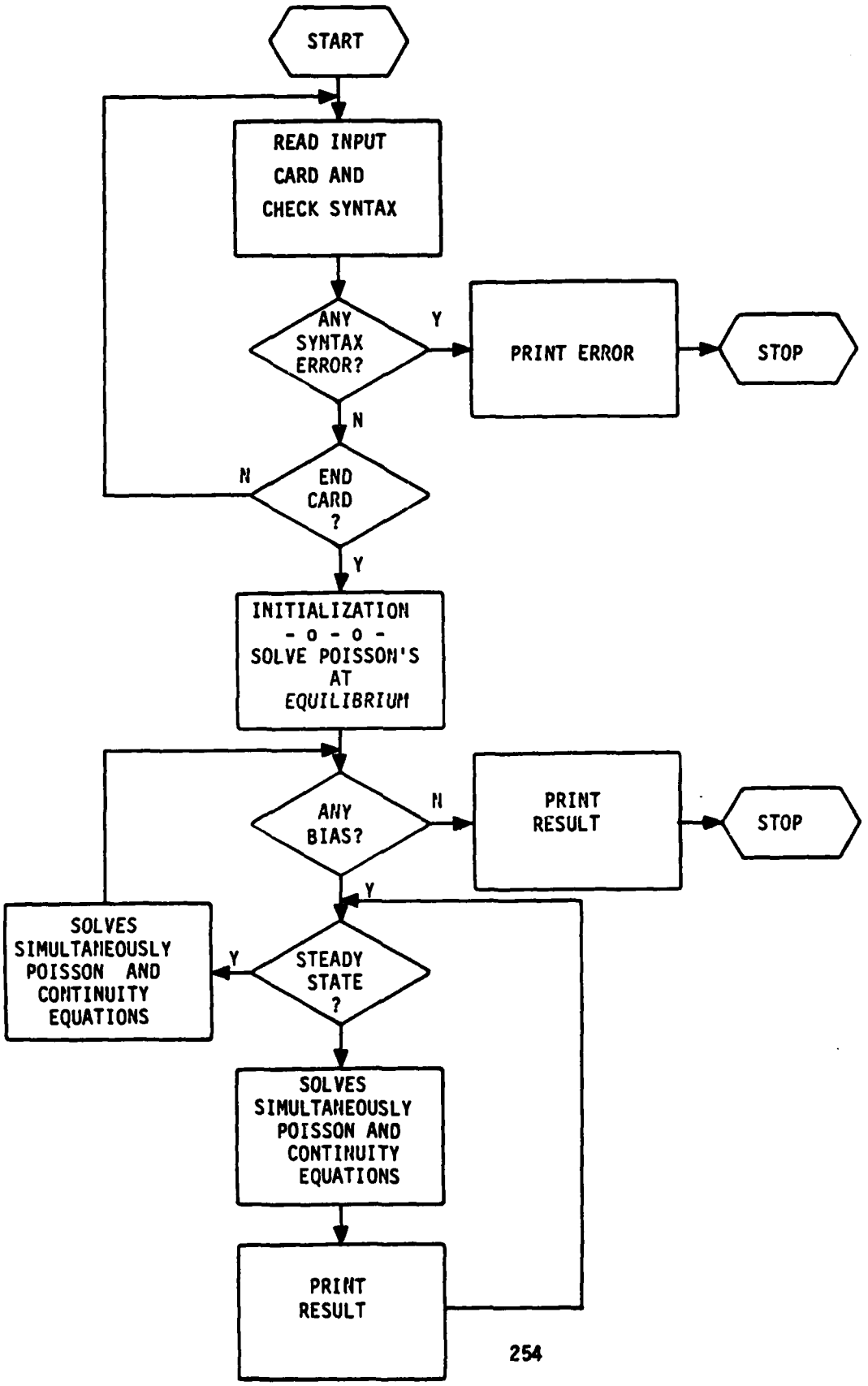
TGSIN input title processor

TYPEP . type parameters processor

After completing the input and initialization the program calls the subroutine EQUID which solves Poisson's equation at equilibrium. If no bias is specified (see flowchart) the program outputs the result and stops; if a bias is specified, SEDAN calls CALD and solves simultaneously for potential $\psi$ and electron and hole concentrations n and p, with the equilibrium solution as the initial guess. One of the input parameters is the time increment for the transient response. A modification of this parameter permits to have a steady state solution or a transient solution. There is no automatic increment of the time step inside the program, so, if one wants to see the evolution of the solution in time (See Figs. 4.2.4a and 4b), he must give increasing values to this parameter and perform the solution for each value.

If one specifies a short timestep, say 10 picoseconds he must also specify only one bias point and have the transient solution to that bias situation. If any new bias has been specified, incrementing one or both of the input voltages, but in this case asking for the steady-state solution, the previous solution of CALD is used as an initial guess for the next set of bias conditions. After each bias point solution and at the end of the analysis, SEDAN prints out the values of all the requested parameters.

SEDAN FLOW CHART



# S E D A N

## S E M I C O N D U C T O R  D E V I C E  A N A L Y S I S

DON D'AVANZO
MASSIMO VANZI
ROBERT W. DUTTON

INTEGRATED CIRCUIT LABORATORY
STANFORD UNIVERSITY
STANFORD, CA. 94305

MAY 15 1979

# A P P E N D I X 4.2.2

USERS MANUAL

FOR THE VERSION OPERATING ON AN IBM 370/168 USING THE FORTRAN-4 LEVEL H COMPILER

## GENERAL DESCRIPTION OF SEDAN

SEDAN (SEMICONDUCTOR DEVICE ANALYSIS) IS A COMPUTER PROGRAM WHICH PERFORMS THE ONE-DIMENSIONAL ANALYSIS OF A SEMICONDUCTOR STARTING FROM THE IMPURITY PROFILE (ANALYTICAL OR OUTPUT OF SUPREM ) AND SOLVES FOR THE ELECTRICAL PROPERTIES OF THE DEVICE.
THESE ELECTRICAL PROPERTIES CAN BE COMPLETELY DESCRIBED BY FIVE PHYSICAL RELATIONSHIPS: - POISSON EQUATION, ELECTRON AND HOLE TRANSPORT EQUATIONS, ELECTRON AND HOLE CONTINUITY EQUATIONS.
WITH THE APPROPIATE BOUNDARY CONDITIONS THE COUPLED EQUATIONS CAN BE SOLVED FOR CARRIER CONCENTRATION AND ELECTROSTATIC POTENTIAL (DEPENDENT VARIABLES).
THE RECOMBINATION RATE INCLUDES SHOCKLEY READ HALL AND AUGER RECOMBINATION MECHANISMS.
THE TRANSPORT EQUATIONS HAVE BEEN WRITTEN ASSUMING THE APPLICABILITY OF THE EINSTEIN RELATION FOR A NON-DEGENERATE SEMICONDUCTOR.

A HIGH IMPURITY CONCENTRATION EFFECT, THE SO CALLED "BAND GAP NARROWING" IS INCLUDED. IT CONSISTS OF AN EFFECTIVE FORBIDDEN GAP NARROWING DUE TO OVERLAPPING OF THE ELECTRON WAVE FUNCTIONS ASSOCIATED WITH THE IMPURITY ENERGY LEVELS AND THE INTRINSIC CONDUCTION AND VALENCE BAND EDGES.

THE ONE-DIMENSIONAL GRID IS DIVIDED INTO A NUMBER OF REGIONS, NOT EXCEEDING 10. A UNIQUE GRID SPACING IS SPECIFIED FOR EACH REGION.
AN IMPROVED FINITE DIFFERENCE APPROXIMATION IS OBTAINED CONSIDERING THE CURRENT DENSITY EXPRESSIONS AS DIFFERENTIAL EQUATIONS AND INTEGRATING THEM BETWEEN TWO ADJACENT GRID POINTS.
THE INTEGRATION IS CARRIED OUT BY ASSUMING THE CARRIER MOBILITY, THE ELECTRIC FIELD INTENSITY, AND THE CURRENT DENSITY TO BE CONSTANT BETWEEN ADJACENT GRID POINTS.

THE TIME-DEPENDENT NATURE OF THE EQUATIONS IS ACCOUNTED FOR BY A TIME ADVANCEMENT METHOD CONSISTING OF A QUASI-LINEARIZATION TECHNIQUE IN WHICH THE TWO CONTINUITY EQUATIONS ARE EXPRESSED IN A TRUNCATED TAYLOR SERIES.
THE MATRIX EQUATION IS READILY INVERTED BY "LU" DECOMPOSITION AND FORWARD AND BACK SUBSTITUTION ARE USED TO SOLVE SIMULTANEOUSLY FOR THE POTENTIAL AND THE ELECTRON AND HOLE CONCENTRATIONS.

ALL THIS FORMS THE NUMERICAL BASIS OF SEDAN, WHICH IS USEFUL FOR ANALYZING BOTH TWO AND THREE TERMINAL, ONE DIMENSIONAL DEVICES.

THE INFORMATION OBTAINED FROM THE DEVICE ANALYSIS INCLUDES VALUES OF THE DEPENDENT VARIABLES (P, N, PSI), AS WELL AS ANY PARAMETERS WHICH CAN BE DERIVED FROM THE DEPENDENT VARIABLES SUCH AS QUASI-FERMI LEVELS, TERMINAL CURRENTS, JUNCTION CAPACITANCES, INTEGRATED CHARGE, ETC.
IT IS POSSIBLE THEN TO DERIVE AN APPROPIATE MODEL FOR CIRCUIT SIMULATION PROGRAMS LIKE MSINC™ OR SPICE™ AND TO STUDY HOW PHISYCAL PARAMETERS SUCH AS DOPING PROFILE, LIFETIMES AND CARRIER MOBILITIES, ARE RELATED TO THE DEVICE ELECTRICAL BEHAVIOUR AND TO THE CIRCUIT PERFORMANCE.

## SEDAN INPUT FORMAT

THE INPUT PARAMETERS FOR SEDAN ARE SPECIFIED BY A SEQUENCE OF LINES IN A DATA FILE. EACH OF THESE LINES BELONGS TO ONE OF THE FOLLOWING TYPES.

| INITIALIZATION/ANALYSIS | INPUT/OUTPUT | MODEL |
|---|---|---|
| TITLE | PRINT | MODEL |
| COMMENT | PLOT | |
| GRID | LOAD | |
| DEVICE | | |
| PROFILE | | |
| BIAS | | |
| END | | |

EACH CARD CONSISTS OF ONE OF THE ABOVE CARD TYPE IDENTIFIERS FOLLOWED BY EITHER A PARAMETER LIST OR A CHARACTER STRING (OF WHICH ONLY THE FIRST FOUR CHARACTERS ARE SIGNIFICANT). THE PARAMETER LIST OR CHARACTER STRING IS SEPARATED FROM THE CARD TYPE IDENTIFIER BY EITHER A COMMA OR ONE OR MORE BLANKS. ONLY THE FIRST 72 COLUMNS OF A LINE ARE READ BY SEDAN'S INPUT PROCESSOR. IF NOT ALL OF THE PARAMETERS OF A CARD'S PARAMETER LIST WILL FIT ONE ONE LINE, THEY MAY BE PLACED ON THE FOLLOWING LINE IF A PLUS (+) IS USED AS THE FIRST NON-BLANK CHARACTER OF THAT LINE. CARD TYPES THAT USE A CHARACTER STRING INSTEAD OF A PARAMETER LIST MAY NOT HAVE A CONTINUATION LINE. ALL BLANK LINES ARE IGNORED.

THE CARD TYPES THAT HAVE A CHARACTER STRING INSTEAD OF A PARAMETER LIST ARE TITLE, COMMENT, AND END. THE STRINGS IN ANY TITLE OR COMMENT CARDS ARE USED BY THE OUTPUT ROUTINES, WHILE TEXT IN STOP OR END CARDS IS IGNORED. THE PARAMETERS IN THOSE CARDS WHICH USE A PARAMETER LIST ARE SPECIFIED BY A PARAMETER NAME TO WHICH A VALUE IS EQUATED (I.E. <NAME>=<VALUE>). THE PARAMETER/VALUE PAIRS MAY OCCUR IN ANY ORDER WITHIN THE LIST, SEPARATED FROM EACH OTHER BY COMMAS. ANY BLANKS OR SPACES WITHIN THE LIST ARE IGNORED.

THE VALUES ASSIGNED TO THE VARIOUS PARAMETERS MAY BE OF SEVERAL DIFFERENT TYPES (I.E. NUMERICAL, LOGICAL, ETC.). IN THE CARD TYPE DESCRIPTIONS ON THE FOLLOWING PAGES, THE TYPE OF VALUE THAT IS ASSIGNED TO A PARAMETER IS INDICATED BY ONE OF THE FOLLOWING SYMBOLS.

| SYMBOL | TYPE | EXAMPLE OR (DESCRIPTION) |
|---|---|---|
| <N> | NUMERICAL | 10., .2, 1.2E-14, 350.236 |
| <L> | LOGICAL | T, TRUE, Y, YES, F, FALSE, N, NO |
| <D> | DEVICE TYPE | NPNT, MOSC, PNDI |
| <P> | PROFILE TYPE | SUPR, ANAL |
| <F> | LAYER TYPE | CNST, EXPO, GDIF, ERFC, GIMP |

FOLLOWING IS A DESCRIPTION OF EACH CARD TYPE AND ITS ASSOCIATED PARAMETERS. IN THIS DESCRIPTION OPTIONAL PARAMETERS OR PARAMETER GROUPS ARE ENCLOSED BY ROUND BRACKETS.

## A. INITIALIZATION/ANALYSIS CARDS

### TITLE CARD

THE TITLE CARD SPECIFIES THE CHARACTER STRING USED BY THE PRINT AND PLOT ROUTINES TO LABEL THE OUTPUT. THE TITLE CARD ALSO CAUSES THE INITIALIZATION OF MANY OF THE PROGRAMS DATA BUFFERS AND SO IT MUST BE THE FIRST CARD IN ANY PROCESSING SEQUENCE. IF SEVERAL PROCESSES ARE TO BE RUN, THEY MAY BE PLACED IN THE SAME INPUT FILE, SEPARATED BY THEIR TITLE CARDS.

TITLE <CHARACTER STRING>

### COMMENT CARD

COMMENT CARDS CAN BE PLACED AT ANY POINT IN A PROCESSING SEQUENCE AFTER THE INITIAL TITLE, GRID AND SUBSTRATE CARDS. THE CHARACTER STRING SPECIFIED IN THE LAST COMMENT CARD BEFORE A STEP CARD IS USED BY THE PRINT AND PLOT ROUTINES TO LABEL THE OUTPUT.

COMMENT <CHARACTER STRING>

### END CARD

THE END CARD IS USED TO TERMINATE THE PROGRAM AND SO MUST BE THE LAST CARD IN THE INPUT FILE OR CARD DECK.

END <CHARACTER STRING>

## GRID CARDS

THE ONE DIMENSIONAL SIMULATION SPACE IN SEDAN IS REPRESENTED BY UP TO 400 GRID POINTS. THERE CAN BE SEVERAL REGIONS, EACH WITH ITS UNIFORM GRID SPACING (DISTANCE BETWEEN TWO ADJACENT GRID POINTS). IN MOST CASES IT WILL NOT BE NECESSARY TO USE ALL THE 400 GRID POINTS. A SAVING IN GRID POINTS REFLECTS IN A SAVING IN CPU TIME. FOR VERY STEEP PROFILES OR FOR VERY NARROW BASES (THERE MUST BE A SUFFICIENT NUMBER OF GRID POINTS IN THE BASE), THE SOLUTION ACCURACY CAN DEPEND ON THE SELECTED GRID. IN THESE VERY PARTICULAR CASES CAN BE USEFUL TO TAKE ADVANTAGE OF THE POSSIBILITY OF HAVING SEVERAL REGIONS.

TO DEFINE THE GRID THE USER HAS TO SPECIFY A GRID CARD FOR EACH REGION.

THE FORMAT FOR EACH GRID CARD IS :

GRID NREG=<N>, NSTP=<N>, STSZ=<N>

WHERE

NREG REFERS TO THE NUMBER OF THE REGION, STARTING FROM THE BEGINNING OF THE SIMULATION SPACE (OXIDE-SILICON INTERFACE FOR NPN TRANSISTORS AND PN DIODES OR GATE CONTACT FOR MOS CAPACITORS).

NSTP REFERS TO THE NUMBER OF STEPS IN THE REGION

STSZ REFERS TO THE GRID SPACING IN MICRONS FOR THAT REGION.

THESE PARAMETERS MUST BE SPECIFIED FOR EACH REGION, SO THERE WILL BE AS MANY GRID CARDS AS MANY REGIONS THERE WILL BE IN THE SIMULATION SPACE. THE SUM OF THE NSTEPS IN ALL REGIONS CANNOT BE GREATER THAN 400.

IN THE CASE THAT THE INPUT IMPURITY PROFILE FOR SEDAN IS DIRECTLY TAKEN FROM THE OUTPUT OF SUPREM (STANFORD UNIVERSITY PROCESS ENGINEERING PROGRAM), THE USER HAS TO KNOW THE DEPTH OF THE SIMULATION SPACE IN SUPREM (YMAX). IF HE SPECIFIES A TOTAL DEPTH IN SEDAN SIMULATION SPACE GREATER THAN THE ONE IN SUPREM, THE IMPURITY CONCENTRATION VALUE OF THE LAST GRID POINT IN SUPREM WILL BE EXTENDED TO THE ACTUAL NEW TOTAL DEPTH.

TO GIVE AN IDEA OF THE POSSIBLE GRID SPACINGS TO USE, .1 MICRON CAN BE SUFFICIENT FOR CONSTANT OR LOW VARYING CONCENTRATION PROFILES, WHILE .01 OR .008 MICRONS SHOULD BE SUFFICIENT FOR VERY STEEP PROFILES.

EXAMPLE

```
GRID NREG=1, STSZ=.02, NSTP=50
GRID NREG=2, STSZ=.01, NSTP=100
GRID NREG=3, STSZ=.05, NSTP=80
```

## A SIMPLER WAY TO SPECIFY THE GRID

CONSIDERING THAT, IN MOST CASES, A UNIFORM GRID SPACING WILL BE SUFFICIENT THROUGHOUT THE DEVICE, A SIMPLER ALTERNATIVE WAY TO SPECIFY THE GRID MAY BE USED.
IT DOES NOT REQUIRE THE SPECIFICATION OF ANY GRID CARD AT ALL.
IT IS SUFFICIENT TO SPECIFY THE TOTAL LENGTH OF THE DEVICE USING THE PARAMETER -LGTH- IN THE PROFILE CARD (SEE LATER).
THE PROGRAM AUTOMATICALLY WILL TAKE CARE OF THE BEST CHOICE TO HAVE GOOD SAVING IN COMPUTATION TIME TOGETHER WITH SOLUTION ACCURACY.

## DEVICE CARD

THE DEVICE CARD IS USED TO SPECIFY THE TYPE OF DEVICE TO BE ANALYZED.

THE FORMAT IS :

DEVI TYPE=<D>, (BACO=<N>), ( OXTH=<N> )

THE TYPE CAN BE:

NPNT FOR BIPOLAR NPN TRANSISTORS
MOSC FOR MOS CAPACITORS
PNDI FOR PN DIODES

BACO (BASE CONTACT) REFERS TO THE DEPTH AT WHICH THE PROGRAM MUST CONSIDER THE BASE CONTACT TO BE WHEN TYPE=NPNT (MICRONS)

OXTH (OXIDE THICKNESS) REFERS TO THE THICKNESS OF THE OXIDE LAYER IN THE CASE WHEN TYPE=MOSC (MICRONS).

EXAMPLE

DEVI TYPE=NPNT, BACO=1.5

## PROFILE CARD

THE PROFILE CARD CONTAINS THE IMPURITY PROFILE INFORMATION.

THE FORMAT IS:

```
PROF   TYPE=<P>, (NLAY=<K>), (TLAY=<F>), (PAR1=<N>), (PAR2=<N>), (PAR3=<N>),
+      (LGTH=<N>)
```

WHERE TYPE CAN BE:

ANAL  IF THE INPUT PROFILE IS SPECIFIED ANALITICALLY. FOR THE ANALYTICAL SPECIFICATION THE PROFILE IS CONSTRUCTED OF SEVERAL 'LAYERS' (UP TO 10), AND A PROFILE CARD IS REQUIRED FOR EACH 'LAYER'.

SUPR  IF THE OUTPUT OF SUPREM IS READ INTO SEDAN, IN THIS CASE A LOAD CARD REFERRING TO A FILE PREVIOUSLY SAVED IN ASCII FORMAT FROM SUPREM, MUST BE PRESENT AMONG INPUT CARDS.

NLAY  IS THE NUMBER OF THE LAYER

LGTH  IS THE TOTAL LENGTH OF THE DEVICE IF NO GRID CARD IS SPECIFIED

TLAY  (TYPE OF LAYER) CAN BE:

| | |
|---|---|
| CNST | FOR CONSTANT IMPURITY DISTRIBUTION |
| EXPO | FOR EXPONENTIAL |
| GDIF | FOR GAUSSIAN DIFFUSION |
| ERFC | FOR ERROR FUNCTION DISTRIBUTION |
| GIMP | FOR GAUSSIAN IMPLANTATION |

PAR1, PAR2, PAR3,  ARE THREE PARAMETERS THAT HAVE DIFFERENT MEANING DEPENDING ON THE TYPE OF LAYER THEY REFER TO.

I)  IF TLAY=CNST
    PAR1 IS THE BEGINNING OF THE CONSTANT LAYER IN MICRONS
    PAR2 IS THE END OF THE CONSTANT LAYER IN MICRONS
    PAR3 IS THE VALUE OF CONCENTRATION, NEGATIVE FOR P-TYPE AND POSITIVE FOR N-TY (CM-3)

II) IF TLAY=GDIF OR EXPO OR ERFC
    PAR1 IS THE DIFFUSION SURFACE (0 FROM X=0, 1 FROM X=LAST X)
    PAR2 IS THE DIFFUSION LENGTH IN MICRONS
    PAR3 IS THE SURFACE CONCENTRATION (CM-2)

THE EQUATIONS ARE:

A) FOR THE GAUSSIAN DIFFUSION

C(X) = PAR3*EXP-(1L*PAR1-X1/PAR2)**2

B) FOR THE ERROR FUNCTION

C(X) = PAR3*ERFC(1L*PAR1-X1/PAR2)

III)IF TLAY=GIMP
PAR1 IS THE RANGE OF THE GAUSSIAN (MICRONS)
PAR2 IS THE STANDARD DEVIATION (MICRONS)
PAR3 IS THE PEAK VALUE OF CONCENTRATION (CM-3)

THE EQUATION IS:

C(Y) = PAR3*EXP-((X-PAR1)/PAR2)**2

EXAMPLE

```
PROF  TYPE=PEAL
PROF  NLAY=1, TLAY=GDIF, PAR1=0, PAR2=.5, PAR3=1E19
PROF  NLAY=2, TLAY=CNST, PAR1=.5, PAR2=2.5, PAR3=1E15
PROF  NLAY=3, TLAY=GDIF, PAR1=1, PAR2=.5, PAR3=1E19
```

OR

```
PROF TYPE=SUPR
```

## BIAS CARD

THE BIAS CARD SPECIFIES THE VOLTAGES AT WHICH ONE WANTS THE ANALYSIS TO BE PERFORMED.

DEPENDING ON THE TYPE OF DEVICE TO ANALYZE THE VOLTAGES CHANGE NAME.

IF NPNT :

VBEF = FIRST BASE-EMITTER VOLTAGE
VBEL = LAST BASE-EMITTER VOLTAGE
VBES = STEP IN BASE-EMITTER VOLTAGE, THE BASE-EMITTER VOLTAGE IS INCREMENTED FROM VBEF TO VBEL AT STEPS OF VBES

VCEF = FIRST COLLECTOR-EMITTER VOLTAGE
VCEL = LAST COLLECTOR-EMITTER VOLTAGE
VCES = STEP IN COLLECTOR-EMITTER VOLTAGE

IF NOSC :

VGSF = FIRST GATE-SOURCE VOLTAGE
VGSL = LAST GATE-SOURCE VOLTAGE
VGSS = STEP IN GATE-SOURCE VOLTAGE
VSBF = FIRST SOURCE-BULK VOLTAGE
VSBL = LAST SOURCE-BULK VOLTAGE
VSBS = STEP IN SOURCE-BULK VOLTAGE
VFBA = FLAT BAND VOLTAGE

IF PNDI :

VDDF = FIRST DIODE VOLTAGE
VDDL = LAST DIODE VOLTAGE
VDDS = STEP IN DIODE VOLTAGE

## B. INPUT/OUTPUT CARDS

THE INPUT/OUTPUT CARDS ARE THE PRINT, PLOT AND LOAD CARDS. THE FIRST TWO OF THESE, THE PRINT AND PLOT CARDS, ARE USED FOR CONTROLLING THE HARDCOPY OUTPUT OF THE PROGRAM.
THE OTHER INPUT CARD, THE LOAD CARD, IS USED TO READ AS INPUT TO SEDAN A FILE OF THE SAVE TYPE OF THE PROGRAM SUPREM. THE FORMAT IS THE SAME AS IN PROGRAM SUPREM, THE ONLY CONSIDERATION IS THAT THE SUPREM SAVE FILE HAS TO BE SAVED IN ASCII FORMAT.

### PRINT CARD

THE PRINT CARDS CONTROL SEDAN'S PRINTED NUMERICAL OUTPUT, WHICH IS OF TWO BASIC TYPES. THE FIRST TYPE, CALLED THE HEADING INFORMATION, IS CONTROLLED BY THE PRINT CARD'S HEAD PARAMETER. THE HEADING CONTAINS INFORMATION ABOUT THE VALUES OF THE TERMINAL (EMITTER, BASE AND COLLECTOR) CURRENTS, FOR EACH BIAS POINT, THE SHEET RESISTANCES FOR EACH LAYER OF DIFFERENT TYPE, THE VALUES OF THE DC CURRENT GAIN BETA, THE EMITTER AND BASE GUMMEL NUMBERS AND OTHER UTILITY INFORMATION.

THE SECOND TYPE OF PRINTED NUMERICAL OUTPUT IS DIVIDED INTO THREE GROUPS OF DATA, WE COULD CALL THEM PRIMARY INFORMATIONS AND SECONDARY ONES.
THESE LOGICAL PARAMETERS CONTROL THESE THREE SETS OF PRINTED OUTPUT.
THE INFORMATIONS COMING OUT WHEN EACH OF THE THREE PARAMETERS IS SET TRUE ARE DESCRIBED IN THE CARD FORMAT DESCRIPTION.
THE PRINT CARD PARAMETERS ARE LOGICAL FLAGS AND SO THEY MUST BE SET TRUE OR FALSE IN THE INPUT DECK OF CARDS.

THE PRINT CARD FORMAT IS :

PRINT (HEAD=<L>), (PRT1=<L>), (PRT2=<L>), (PRT3=<L>)

HEAD — REFERS TO THE HEADING PAGE OF GENERAL INFORMATIONS AND UTILITY DATA. THE DEFAULT VALUE IS TRUE.

PRT1 — CONTROLS THE PRINTED OUTPUT OF THE GRID NUMBER, THE DEPTH IN MICRONS CORRESPONDING TO EACH GRID NUMBER, THE ELECTROSTATIC POTENTIAL, THE ELECTRON AND HOLE CONCENTRATIONS, THE ELECTRON AND HOLE CURRENT DENSITY, THE RECOMBINATION RATES.
THE DEFAULT VALUE IS FALSE.

PRT2 — CONTROLS THE PRINTED OUTPUT OF THE GRID NUMBER, THE DEPTH IN MICRONS CORRESPONDING TO EACH GRID NUMBER, THE ELECTRIC FIELD, THE IMPURITY CONCENTRATION, THE ELECTRON AND HOLE QUASI-FERMI POTENTIALS.
THE DEFAULT VALUE IS FALSE.

PPT3 CONTROLS THE PRINTED OUTPUT OF ALL THE BIAS DEPENDENT VALUES LIKE BASE-EMITTER OR BASE-COLLECTOR JUNCTION CAPACITANCE, BETA, THE COLLECTOR AND BASE CURRENT DENSITIES. THE DEFAULT VALUE IS FALSE.

## LOAD CARD

THE LOAD CARD IS USED TO READ THE SAVED FILE OF THE PROGRAM SUPREM.

THE LOAD CARD HAS ONLY ONE PARAMETER. IT SPECIFIES THE FORTRAN LOGICAL UNIT NUMBER THAT IS ASSIGNED TO THE DATA FILE IN THE JOB CONTROL LANGUAGE DECK. THE TYPE OF DATA TRANSFER MUST BE ASCII AND SO MUST BE THE DATA STORED IN THE SUPREM SAVE FILE.

THE FORMAT OF THE LOAD CARD IS:

LOAD LUNM=<N>

LUNM : A FORTRAN I/O LOGICAL UNIT NUMBER WHICH HAS BEEN ASSIGNED TO A DATA FILE IN THE J.C.L DECK.

## PLOT CARD

THE PLOT CARDS CONTROL SEDAN'S PLOTTED OUTPUT (LINE PRINTER PLOT IN THE IBM VERSION). ALL THE PARAMETERS HAVE LOGIC VALUES AND ARE DEFAULTED FALSE (NO). EACH PARAMETER SET TRUE(YES) CAUSES A PAGE PLOT. THERE ARE TWO DIFFERENT SETS OF PARAMETERS, ONE DEPTH DEPENDENT WILL BE PLOTTED AFTER EACH BIAS POINT, ONE BIAS DEPENDENT WILL BE PLOTTED AFTER THE LAST BIAS POINT IN A SEQUENCE OF BIAS POINTS.

THE FORMAT IS:

```
PLOT (EFIF=<L>   (PHIN=<L>), (PHIP=<L>), (PSIP=<L>), (NCON=<L>),
+    (HCON=<L>), (ECON=<L>), (PETA=<L>), (PCDE=<L>), (RCDE=<L>),
+    (PFJC=<L>), (RFJC=<L>)
```

THE FIRST SEVEN PARAMETERS ARE DEPTH DEPENDENT AND THEN DEPTH WILL BE THE X AXIS. THE LAST FIVE PARAMETRS ARE BIAS DEPENDENT AND THE APPROPIATE VOLTAGE WILL BE THE X AXIS.

## C. MODEL CARD

THE MODEL CARD CONTAINS THE PHISICAL MODELS SPECIFICATION, THE VALUES OF THE PHYSICAL CONSTANTS, THE REQUEST FOR THE CALCULATION OF THE JUNCTION CAPACITANCES AND OTHER UTILITY PARAMETERS.
NOTICE THAT IF THE CALCULATION OF A JUNCTION CAPACITANCE IS ASKED (ONLY ONE AT A TIME CAN BE ASKED), THE BASE AND COLLECTOR (DIODE) BIAS MUST BE SET IN THE APPROPIATE WAY TO ASSURE GOOD CALCULATIONS BOTH IN THE CASE OF FORWARD OR REVERSE BIAS OF THE JUNCTION.
FOR EACH SPECIFIED BIAS POINT THE PROGRAM ACTUALLY CALCULATES TWO BIAS SITUATIONS SEPARATED BY 20 MILLIVOLTS, INTEGRATES THE CHARGE AT THE TWO CONDITIONS AND CALCULATES THE CAPACITANCE AS THE RATIO BETWEEN THE DIFFERENCE IN CHARGE DEVIDED BY THE DIFFERENCE IN VOLTAGE.

THE FORMAT OF THE MODEL CARD IS :

```
MODEL (SRHR=<L>), (AUGE=<L>), (BGNR=<L>), (TINC=<N>), (CNAU=<N>),
+     (CPAU=<N>), (ELFT=<N>), (HLFT=<N>), (MOB0=<N>), (MOB1=<N>),
+     (MOB2=<N>), (MOB3=<N>), (MOB4=<N>), (MOB5=<N>), (MOB6=<N>),
+     (MOB7=<N>), (MOB8=<N>), (MOB9=<N>), (NSRH=<N>), (BCJC=<L>),
+     (BEJC=<L>), (BBJC=<L>)
```

WHERE:

SRHR — IS A LOGICAL FLAG WHICH CONTROLS THE SHOCKLEY READ HALL RECOMBINATION MECHANISM. IF IT IS FALSE OR NO (F OR N), THE SRH RECOMBINATION IS INSERTED IN THE MODEL BUT WITH CONSTANT ELECTRON AND HOLES LIFETIME (EQUAL TO ELFT AND HLFT). IF SRHR IS TRUE, CONCENTRATION DEPENDENT LIFETIMES ARE INSERTED IN THE MODEL (SEE 'PHYSICAL MODELS').

AUGE — IS A LOGICAL FLAG WHICH CONTROLS THE AUGER RECOMBINATION MECHANISM (SEE 'PHYSICAL MODELS'). IF IT IS TRUE, AUGER RECOMBINATION IS INSERTED IN THE MODEL, OTHERWISE IT IS NOT.

BGNR — IS A LOGICAL FLAG WHICH CONTROLS THE SO CALLED 'BAND GAP NARROWING PHISICAL MECHANISM (SEE 'PHYSICAL MODELS'). IF IT IS TRUE, THE BAND GAP NARROWING IS CONSIDERED, OTHERWISE IT IS NOT.

TINC — (TIME INCREMENT) REFERS TO THE TRANSIENT ANALYSIS. IT INDICATES THE TIME (SECONDS) AFTER WHICH THE SOLUTION IS DESIRED.

CNAU, CPAU — ARE TWO CONSTANTS IN AUGER RECOMBINATION FORMULATION (SEE EQUATION).

$$U_{AUGER} = C_{NAU}(P N^2 - N N_{IE}^2) + C_{PAU}(N P^2 - P N_{IE}^2)$$

FLFT — REFERS TO THE VALUE OF ELECTRON LIFETIME IN THE SHOCKLEY READ HALL RECOMBINATION MECHANISM. THE LIFETIME FOR ELECTRON IS EQUAL TO FLFT EVERYWHERE IF SPHF=N, ONLY AT LOW CONCENTRATIONS IF SRHS=Y (SEE EQUATION).

HLFT — REFERS TO THE VALUE OF HOLE LIFETIME IN THE SHOCKLEY READ HALL RECOMBINATION MECHANISM (SEE EQUATION).

USRH = (N*P-NIF**2)/(TNN*(N+N1)+TPP*(P+P1))

WHERE TNN = FLFT/(1+(NT(X)/NSRH))

AND TPP = HLFT/(1+(NT(X)/NSRH))

MOB0 TO MOB9 — REFER TO THE VALUES OF TEN CONSTANTS IN THE MOBILITY AS A FUNCTION OF ELECTRIC FIELD AND IMPURITY CONCENTRATION SPECIFICATION. HERE FOLLOWS THE RELATION WHERE MOB0 TO MOB4 REFER TO THE EXPRESSION FOR ELECTRONS AND MOB5 TO MOB9 TO THE SIMILAR EXPRESSION FOR HOLES

MOBILITY = (MOB0+(MOB1-MOB0)/(1+(N(X)/MOB2)*MOB3))*(1+E(X)/MOB4)**-1

NSRH — REFERS TO THE VALUE OF THE CONSTANT NSRH IN TABLE III AND IN RELATIONSHIPS 24 AND 25 OF THE REPORT, IN THE SHOCKLEY-READ-HALL RECOMBINATION (SEE BEFORE).

BCJC — LOGIC FLAG FOR THE CALCULATION OF THE BASE-COLLECTOR JUNCTION CAPACITANCE.

BEJC — LOGIC FLAG FOR THE CALCULATION OF THE BASE-EMITTER JUNCTION CAPACITANCE. CAN BE REVERSE OF FORWARD DEPENDING ON THE SPECIFIED BIAS.

DDJC — LOGIC FLAG FOR THE CALCULATION OF THE DIODE JUNCTION CAPACITANCE.

DEFAULT VALUES :

SRHR = Y

AUGE = N

BGNW = Y

TINC = 10 (STEADY STATE)

CNAU = 2.8E-31

CPAU = 9.9E-32

ELFT = .5E-6

HLFT = .5E-6

MOB0 = 1407.3

MOB1 = 71.1

MOB2 = 1.1E17

MOB3 = .720

MOB4 = 7.396E3

MOB5 = 467.7

MOB6 = 60.7

MOB7 = 1.4E17

MOB8 = 0.700

MOB9 = 2.054

NSRH = 7.1E15

BCJC = N

BFJC = N

BDJC = N

## 4.3 TWO DIMENSIONAL MODELING OF NONPLANAR DEVICES ON MINICOMPUTERS

J. A. Greenfield

### 4.3.1 Introduction

At the present time the design of integrated circuits using metal-insulator-semiconductor (MIS) technologies relies heavily upon empirical data for the optimization of device performance. There is also some use of theoretical results based on various analytic and numerical models of the device physics. The process of optimization involves the determination of device parameters which yield some form of optimal device performance.

When using empirical data to perform the optimization, each device parameter variation requires a modification of the processing schedule and the initiation of a new process run. This technique can be very expensive as well as time consuming. Once a modified device has been fabricated the device evaluation is accomplished through measurements of device performance. These measurements are necessarily subject to uncertainty introduced by process variations and measurement inaccuracies. The evaluation of a device through measured device performance also provides little physical insight into the factors governing the device operation.

The use of computer and analytic prediction in the optimization avoids the disadvantages associated with the use of empirical data and provides the device designer with good physical insight into device performance. However, most predictive techniques currently in use are based on one dimensional approximations to the true device structure.

Accurate results may be expected from these techniques only when the assumed approximations are valid. Even when the approximations fail the predictive techniques may still be useful for anticipating trends in the variation of device performance with changes in device parameters. However, the techniques no longer provide accurate quantitative information and lose their value as an optimization tool. More accurate results can be provided by two dimensional modeling capabilities.

The present state of the art in device technology allows the construction of many devices which may no longer satisfy the one dimensional approximations. This problem is currently of importance in the design of integrated circuits which have high packing densities or have the capability to operate at high applied voltages.

The attainment of high packing density requires that the device dimensions be reduced to the point where the devices are highly two dimensional in nature. The device operation then depends on the two dimensional characteristics of the device. Particular examples in this area include random access memories and serial memories utilizing charge coupled devices [4.8, 4.9].

The remainde of this section deals with various considerations involved in the two dimensional modeling of semiconductor devices. Section 4.3.2 describes the advantages and restrictions associated with the use of a minicomputer to simulate general MIS device structures. Section 4.3.3 describes several of the many applications for a computer program which solves the electrostatic Poisson equation. Section 4.3.4 discusses the device structure characteristics which should be allowed by the computer program in order that general MIS devices may be treated.

Section 4.3.5 treats the discretization of Poisson's equation using a five point finite difference approximation. Section 4.3.6 describes the numerical techniques and iterative algorithms which are used to obtain the solution to Poisson's equation.

### 4.3.2 Two Dimensional Modeling on Minicomputers

In many instances, the use of a minicomputer is preferable to time sharing a large computer system. The minicomputer can offer the advantages of lower operating expense, greater convenience and less total execution time. The major disadvantages of a minicomputer are typically the limited available memory and the slower execution of arithmetic operations. Minicomputers operate with interactive systems which introduce minimal system overhead into the total time necessary to accomplish a task. A time sharing system services many users and must necessarily introduce significant system overhead. Thus, although the minicomputer executes arithmetic operations more slowly than the large computer, the total time required to perform a task is usually smaller on the minicomputer because of the lower system overhead and the lack of a large number of simultaneous users.

The implementation of two dimensional modeling capabilities on an interactive minicomputer system provides the device designer with a tool which is well suited to the process of device optimization. During the optimization process, each change to a device structure can be determined based on the device performance predicted by simulations

of the previous device structures. The interactive nature of the minicomputer system greatly facilitates this iterative procedure.

The use of a minicomputer does impose some restrictions upon the type of numerical calculations which may reasonably be performed. The speed of arithmetic operations on a minicomputer is usually slow enough that simulations requiring excessive amounts of computation should be avoided. Of course, even on a time sharing system, the requirement for large amounts of computation is undesirable because of the expense involved. This point will be addressed later in this section. The storage requirements for a simulation may easily exceed the available memory of a minicomputer. However, the use of iterative relaxation techniques in performing the simulations allows most of the storage to be provided by a mass storage device such as a disk or a tape. Minicomputers with limited memory generally have mass storage devices available. The relaxation techniques will be described in Section 4.3.6.

In a fairly general case, the two dimensional analysis of a MIS device involves the simultaneous solution of Poisson's equation and the continuity equations for electrons and holes within the entire device cross section. This presents a formidable task requiring a great deal of computer storage and execution time. Obtaining these solutions on a minicomputer would generally be unreasonable. Several authors have developed computer programs which implement various techniques for performing these solutions [4.10-11]. However, the complexity of the problem generally leads to the treatment of only a very limited set of device structures. For example, in the simulation of MIS transistors, the device is often taken to have rectangular boundaries with portions of the boundary used

to approximate the source and drain diffusions [4.12]. The excessive computational requirements severely restricts the number of device examples which can be investigated, limiting the usefulness of the solutions for device optimization.

If the treatment of charge transport may be eliminated, then it is only necessary to solve the electrostatic Poisson equation. In many cases these solutions yield sufficient information to aid in device optimization. This will be elaborated on further in Section 4.3.3. The amount of computer storage and execution time required for the solution of the electrostatic Poisson equation is much smaller than for the solution of the entire charge transport problem. Reducing the amount of computation makes it feasible to obtain the solutions using minicomputers. It also becomes more reasonable to increase the generality of the allowed device structures. By doing this the simulation capabilities may be applied to a more complex and diverse set of devices.

### 4.3.3 Applications for the Simulations

There are several types of device design problems where solutions to the electrostatic Poisson equation provide useful information for device structure optimization. While applications for CCD's and high voltage devices have been considered [4.13], this work will focus on the applications pertinent to VLSI device structures.

The design of short channel MIS transistors is aided by the ability to determine the electrostatic potential distribution. When a transistor is not conducting significant current the phenomena of punch through can be investigated. Two dimensional simulations are used to determine when the potential barrier between the source and drain is

lowered sufficiently to allow the flow of a subsurface current, Figure 4.3.9a shows the cross section of a MIS transistor in which the drain-source bias is insufficient to cause punch through to occur. The equipotential lines shown indicate that the depletion regions associated with the source and drain diffusions have not merged. Figure 4.3.9b shows the same device cross section with an increased drain source bias. In this case the equipotential lines indicate that the depletion regions associated with the source and drain diffusions have merged below the channel region. The potential has the form of a saddle in this region. It is evident that the potential barrier between the source and the drain is smaller in the region of the saddle potential than it is along the channel. Figure 4.3.9c shows the potential as a function of the distance along a line between the source and the saddle region for the two drain-source biases illustrated. This figure illustrates the lowering of the potential barrier caused by the onset of punch through. The threshold voltage of complicated device structures can also be determined by finding the gate potential which causes the formation of a conducting channel. The current versus voltage characteristics in the subthreshold region may be obtained by using two dimensional simulations along with some assumptions concerning the form of the channel charge distributions [4.14].

Another application of the calculation of the electrostatic potential is in the determination of capacitance in highly two dimensional devices. The capacitance may be determined by calculating the change in the charge induced on an electrode for a small change in the electrode potential. The induced electrode charge is calculated by integrating

the normal component of the electric field along the electrode. The electric field is obtained from the electrostatic potential distribution provided by a two dimensional simulation. Figure 4.3.10 shows the cross sections of the two short channel MOS transistors fabricated using metal gate technology. The devices differ only in the oxidation conditions used during the growth of the gate oxide. The figures show the equipotential lines within the device. In Figure 4.3.10a the oxidation conditions are such that the oxide growth rate is nearly the same over the lightly doped channel region and the heavily doped source and drain regions. The gate-source and gate-drain overlap capacitances are substantial and may be calculated easily and accurately from knowledge of the final device structure. In Figure 4.3.10b the oxidation conditions have been modified to enhance the oxide growth rate over the heavily doped source and drain regions. This results in a significant reduction in the overlap capacitances. However, the two dimensional nature of the device now necessitates the use of two dimensional simulation to calculate these capacitances.

### 4.3.4 Required Device Structure Characteristics

The solution of the electrostatic Poisson equation for general device structures imposes certain requirements on any computer program which performs the solutions. The program must allow for the specification of parameters which describe the physical characteristics of these device structures. This specification should be done in a manner which is consistent with standard terminology used in describing the structure and physics of devices. Of course, some terminology conventions may be necessarily imposed by the implementation of a computer program to perform

the simulations. The remainder of this section describes the physical device characteristics which should be considered when simulating general device structures.

In order to simulate a two dimensional cross section of a device it must be assumed that the device structure is invariant in the orthogonal third dimension. The coordinates X and Y will be used to represent the two dimensional cross section of the device. The coordinate Z will represent the orthogonal third dimension. We will assume that the X coordinate lies parallel to the surface of the material in which the device is fabricated. For a device with a nonuniform surface, the X coordinate would typically lie parallel to the surface of the starting material. The Y coordinate is taken orthogonal to X and oriented such that Y increases from the outer device surface into the substrate. The X and Y coordinates generally represent two coordinates in a rectangular coordinate system. However, it is sometimes desirable to let X and Y represent the radial and axial coordinates, respectively, in a circular cylindrical coordinate system. In this situation the device is invariant in the angular coordinate which is represented by Z. The relations between the coordinate axes and typical device structures for both rectangular and circular cylindrical coordinates are shown in Figure 4.3.11.

MIS devices are fabricated in a semiconductor material which is covered by insulators and high conductivity electrodes. The semiconductor usually extends in the Y direction for a distance which is large compared to the vertical extent of the active device region. However, some devices are fabricated in a relatively thin piece of semiconductor which is supported by an underlying insulating substrate.

One example of this situation is the SOS technology in which a thin silicon layer is grown on an insulating sapphire substrate. Simulations of devices must account for the fact that the dielectric permittivities of the semiconductor, the top insulator and the supporting insulator may all have different values. There should also be allowance for the presence of fixed charge associated with the interfaces between the semiconductor and the insulator regions.

The active device region of most MIS devices consists of a semiconductor with a highly nonuniform impurity distribution. One technique used to introduce impurities is through ion implantation. The resulting impurity profile is characterized by the location and concentration associated with the peak of the profile and by parameters which describe the profile variation with distance. A second technique for introducing impurities is through thermal diffusion from an impurity source in contact with the surface of the semiconductor. For a diffusion which forms a metallurgical junction the impurity profile is often characterized by the impurity concentration at the semiconductor surface and the location of the junction. The use of ion implantation or thermal diffusion may result in two dimensional impurity profiles. A third technique for introducing impurities which results in a one dimensional impurity profile is the use of epitaxial deposition. The impurity profile must be specified through the variation of impurity concentration with vertical distance into the substrate. The specification of the semiconductor impurity distribution should be done in terms of parameters which are conveniently available to the device designer.

Many MIS devices which show two dimensional characteristics have nonplanar physical boundaries. One important instance of such a boundary is the interface between the semiconductor and the top insulating region. A second instance is the upper surface of the top insulating region. These boundaries can become nonplanar due to such effects as local oxidation and local etching. Requiring that the Y locations of these boundaries be single valued functions of X is not overly restrictive. Such a restriction greatly simplifies the numerical simulation of the device. It is also reasonable to restrict the boundaries at the left and right sides to be planar.

The solution of the electrostatic Poisson equation within a region of a device requires that some conditions be imposed on the potential at the boundaries of the region being considered. The boundary conditions along the sides of the region depend on the type of device which is being simulated. There often occur vertical symmetry lines at the center of a device or between identical sections of a repeated device structure. The use of a symmetry line as one side of the solution region requires that reflection symmetry be used as the boundary condition on the potential at the appropriate side. For a region where a device approaches the limit of a one dimensional device the appropriate boundary condition is a one dimensional solution to Poisson's equation. The use of this type of boundary condition along a side of the solution region requires that the one dimensional solution be performed in the Y direction. The simulation of cylindrically symmetric devices requires that a circular cylindrical coordinate system be used with an appropriately modified Poisson equation. Boundary conditions along the sides of the solution

region could utilize one dimensional solutions or reflection symmetry.

The boundary conditions on the top surface of the device are dependent on the type of electrode structure covering the surface. The simulation of general MIS devices requires that this electrode structure be arbitrary. The potential is determined at any point on the surface where an electrode is present. At points on the surface where no electrode is present, the boundary condition consists of the specification of the normal gradient of the potential. The value of the normal gradient is proportional to the density of charge on the surface.

The boundary condition on the bottom surface of the solution region depends on whether an insulating substrate is supporting the semiconductor which contains the active device. If there is a suporting insulator, then there is typically a single electrode along the bottom surface of this insulator. The potential of this electrode establishes the potential which is used as the boundary condition at every point along the electrode. If there is no supporting insulator, then the semiconductor typically extends much further in the Y direction than the extent of the active device region. At some point below the active device region the potential usually begins to exhibit one dimensional characteristics. The bottom of the solution region may be taken at this point if the boundary condition along this line is taken to be a one dimensional solution to Poisson's equation. The one dimensional solution in this case is performed in the X direction.

In many devices the active region contains mobile charge which must be properly determined. The most accurate method of treating this charge is to represent the charge as a nonlinear function of the electrostatic potential and the quasi-Fermi potential. If both minority and

majority carriers are handled in this manner, then the locations of the depletion edges within the device do not need to be treated as separate boundaries. They are automatically included due to the nonlinear relation between charge and potential. The functional dependence of the mobile charge on the potential should properly account for degeneracy effects as the quasi-Fermi potentials approach the appropriate energy band edge potentials. This may be handled by using the Fermi-Dirac integral of one half order which reduces to the standard exponential function in non-degenerate situations.

The desire to simulate devices with reverse biased p-n junctions requires that it be possible to specify the electron and hole quasi-Fermi potentials at each point within the substrate. The bias applied to a given region may be specified through the majority carrier quasi-Fermi potential for the region. Since an electrostatic situation is assumed during the solution, a quasi-Fermi potential must necessarily be constant wherever the mobile charge associated with that potential is a significant portion of the total charge. However, a quasi-Fermi potential may have any behavior in regions where the associated mobile charge is negligible.

### 4.3.5 Discretization of Poisson's Equation

The electrostatic Poisson equation which must be solved when analyzing the characteristics of two dimensional MIS device structures is a second order nonlinear elliptic partial differential equation in two space variables. Several approaches exist for obtaining solutions to such equations. These include the use of Green's functions, Fourier series, conformal transformations and discrete numerical techniques. Only the last of these is applicable to the nonlinear equation within an

arbitrary device boundary.

The general discrete numerical approach involves the approximation of the continuous electrostatic potential by the potential values at a discrete grid of points located within the device boundary. At each grid point a discrete approximation to the continuous Poisson equation is used to relate the potential value at this point to the values at neighboring points. The result is a system of nonlinear equations which must be solved to obtain tne potential values at the grid points.

Several types of grid structures have been used extensively for the numerical solution of partial differential equations. The most common is a rectangular grid which in its most general form may have nonuniform spacing between the points in both space directions. The nonuniform grid spacing is necessary to provide the flexibility required for accurate and efficient representation of device features having various sizes. Its major disadvantage is that the use of a large number of grid points to represent a localized feature may introduce an unnecessary increase in the number of grid points elsewhere in the device. This phenomena is illustrated in Figure 4.3.12 which shows the cross section of a MIS transistor. A large grid point density is desired near the source and drain junctions. However, the horizontal grid density is also increased in the substrate where the additional accuracy is not necessary. Other grid schemes which do not suffer from this disadvantage have attracted increasing interest. Thes include nonrectangular orthogonal grids and general finite element grids. Their main disadvantage arises form the increased complexity introduced by the arbitray location of the grid points. For this work the nonuniform rectangular grid was chosen because it provides good flexibility and accuracy while requiring

only minimal organizational complexity.

The discrete approximation to Poisson's equation may relate the potential values at various numbers of adjacent grid points. As this number increases, there is a corresponding increase in the complexity of the nonlinear system of equations which results from the discretization. For a rectangular grid the simplest discretization scheme uses five points consisting of a center point and the four nearest neighbor points. This is generally referred to as a five point finite difference scheme on a rectangular grid, as described below.

The finite difference approximation at each point is derived by integrating Poisson's equation over the rectangular region bounded by the perpendicular bisections of the lines joining the point to its four nearest neighbor points. Figure 4.3.13 shows the five grid points involved in the discretization. Each grid point is identified by a pair of indices representing its location in the X and Y directions. The line segments comprising the boundary of the rectangular region are designated as $C_i$. This figure also shows the triangular areas $S_i$ associated with the line segments and the potential values $V_i$ associated with the five grid points. The continuous form of Poisson's equation is

$$\bar{\nabla}\cdot[K\bar{\nabla}V] = -Q(V) \quad Q(V) = q[N+p(V) - n(V)]$$
$$p(V) = F_{\frac{1}{2}}(q(\phi_p - V)/kT) \qquad (4.3.33)$$
$$n(V) = F_{\frac{1}{2}}(q(V-\phi_n)/kT)$$

Integrating this equation over the rectangular region S bounded by C gives

$$\iint_S \bar{\nabla}\cdot[K\bar{\nabla}V]\,dA = -\iint_S Q(V)\,dA \qquad (4.3.34)$$

The application of Green's theorem allows the integral on the left to be converted from a surface integral to a line integral along the boundary of the rectangle.

$$\oint_C K\bar{\nabla}V \cdot \hat{n}\, d\ell = -\iint_S Q(V)\, dA$$

$$\sum_{m=1}^{4} \int_{C_m} K[\bar{\nabla}V \cdot \hat{n}]\, d\ell = -\sum_{m=1}^{4} \iint_{S_m} Q(V)\, dA \qquad (4.3.35)$$

This last equation must now be approximated in terms of the potential values at the five grid points. The most reasonable and simple approximations for the partial derivatives are

$$\bar{\nabla}V \cdot \hat{n}_m = \frac{V_m - V_o}{h_m} \qquad (4.3.36)$$

We assume that these terms are constant over the integration paths. A simple approximation for the surface integral in (4.3.35) is to assume that Q(V) within the rectangle has a constant value determined by the impurity concentration and potential at the center point. Using these assumptions, (4.3.35) can be reduced to

$$K_s \sum_{m=1}^{4} L_m[\bar{\nabla}V \cdot \hat{n}_m] = -Q(V_o) \sum_{m=1}^{4} A_m$$

$$K_s \sum_{m=1}^{4} \frac{L_m}{h_m}[V_m - V_o] = -Q(V_o) \sum_{m=1}^{4} A_m \qquad (4.3.37)$$

$$L_m = \int_{C_m} \frac{K}{K_s}\, d\ell$$

Here $A_m$ represents the area in Si over which Q(V) is nonzero. This becomes important when nonzero charge only occurs in a portion of the

rectangular region. Such a situation arises when the interface between the semiconductor and the insulator passes through the center grid point.

*For a grid point within the semiconductor the discrete approximation* to Poisson's equation becomes

$$\sum_{m=1}^{4} A_m = \frac{[h_1 + h_3]\,[h_2 + h_4]}{4} = L_1 L_2$$

$$L_1 = L_3 = \frac{h_2 + h_4}{2} \qquad (4.3.38)$$

$$L_2 = L_4 = \frac{h_1 + h_3}{2}$$

$$\frac{2K_s}{h_1(h_1 + h_3)}[V_o - V_1] + \frac{2K_s}{h_2(h_2 + h_4)}[V_o - V_2] + \frac{2K_s}{h_3(h_1 + h_3)}[V_o - V_3]$$

$$+ \frac{2K_s}{h_4(h_2 + h_4)}[V_o - V_4] = Q(V_o) \qquad (4.3.39)$$

Figure 4.3.14 shows a situation where the interface between the semiconductor and the insulator passes through the grid point at which the discretization is being performed. For this case the discrete approximation to Poisson's equation becomes

$$\sum_{m=1}^{4} A_m = \frac{h_2(h_1 + h_3)}{4}$$

$$L_1 = L_3 = \frac{h_2 + h_4\, K_I/K_s}{2} \qquad (4.3.40)$$

$$L_2 = \frac{h_1 + h_5}{2}$$

$$L_4 = \frac{h_1 + h_3}{2}\,\frac{K_I}{K_s}$$

$$\frac{2(K_s + h_4\ K_I/h_2)}{h_1(h_1 + h_3)}[V_o - V_1] + \frac{2K_s}{h_2^2}[V_o - V_2] + \frac{2(K_s + h_4\ K_I/h_2)}{h_3(h_1 + h_3)}[V_o - V_3]$$

$$+ \frac{2K_I}{h_2 h_4}[V_o - V_4] = Q(V_o)$$

By using the form for the discretized equation given in(4.3.37) other situations such as nonuniform interfaces can easily be treated.

4.3.6 Numerical Techniques and Iterative Solution Algorithm

The discretization of Poisson's equation using a rectangular grid of points generates a system of nonlinear equations which relate the values of potential at neighboring grid points. This system must be solved to obtain the grid potentials which approximate the solution to the continuous problem. Figure 4.3.15 illustrates the outside boundary of a simple device structure having a nonuniform surface. The discretization grid is shown as well as the X and Y indices used to label each grid point. The Y index of the top point within the device boundary along the vertical line at X index is denoted by $j_s(i)$. Identification of the grid potentials by the appropriate grid indices yields a system of equations of the form

$$C_1(i,j)\ V(i,j) + C_2(i,j)\ V(i-1,j) + C_3(i,j)\ V(i+1,j)$$

$$+ C_4(i,j)\ V(i,j-1) + C_5(i,j)\ V(i,j+1) = Q(V(i;j));$$

$$i=1, \ldots, N_x;\ j=j_s(i),\ldots,M_y \qquad (4.3.41)$$

The coefficients multiplying the grid potentials are obtained by applying (4.3.37) for the various grid points within the device.

In order to efficiently solve (4,3.41) for the grid potentials the grid points within the device boundary should be assigned some sequential order. A vector V may then be constructed by placing the grid potentials in V according to the order of the associated grid points. A vector Q may also be constructed by placing the grid charge densities from (4.3.41) in Q in the same order as the elements of V. Although the orderings of V and Q do not need to be identical, such a convention is standard and will be used in this work. A sparse coefficient matrix C may be constructed from the coefficients defined by (4.3.41). The solution of (4.3.41) can now be obtained by solving a nonlinear matrix equation of the form

$$CV = Q(V) \qquad (4.3.42)$$

The structure of C depends upon the ordering of the elements within V and Q. Proper choice of the matrix structure can greatly facilitate the solution of the matrix equation.

There are two reasonable choices for the grid point order which generate appropriate structures for the matrix C. The first of these takes consecutive grid points along successive horizontal lines of grid points. The second takes consecutive grid points along successive vertical lines of grid points. In this work, the simulation of devices with nonuniform surfaces dictates that the second method of ordering the points by vertical lines be used for maximum efficiency. This ordering is illustrated in Figure 4.3.16 for the device shown in Figure 4.3.15. Figure 4.3.16 also shows the structure of the coefficient matrix C for this case. Matrix C can easily be partitioned so that it is block tridiagonal. This is accomplished by treating each vertical line

of grid points as a unit. The vector V is divided into individual vectors corresponding to these units. The diagonal blocks of C are tridiagonal matrices with orders equal to the number of points in the corresponding vertical line of grid points. The off-diagonal blocks of C have at most one nonzero element in each row. The block tridiagonal structure is convenient in the implementation of relaxation solution techniques discussed later.

A system of nonlinear equations requires the use of an iterative technique to obtain a consistent solution. One standard and reliable approach is the use of the matrix equivalent of the Newton-Raphson technique. Thisscheme is formulated by linearizing the matrix equation about the current approximation to the solution to obtain

$$
\begin{aligned}
& CV^{n+1} \cong Q(V^n) + J(V^n)\,(V^{n+1} - V^n) \qquad ; J_{p,q}(V^n) = \frac{\partial Q_p}{\partial V_q^n} \\
& [C - J(V^n)]\,V^{n+1} = Q(V^n) - J(V^n)\,V^n \qquad (4.3.43) \\
& C'V^{n+1} = Q'(V^n)
\end{aligned}
$$

The superscripts on V denote the Newton-Raphson iteration numbers. At each iteration we evaluate Q and J based on the most recent approximation to V. The linear matrix equation in (4.3.43) must then be solved to obtain the next approximation to V.

The discretization of Poisson's equation which we have used places only one nonzero element in each row of the matrix J. In addition, since V and Q have identical ordering of their elements, J is a diagonal matrix. Thus, the matrix C' differs from C only in its diagonal elements. The properties of the matrix C' allow several techniques to be applied to the solution of (4.3.43).

By proper scaling of the individual equations comprising (4.3.43) the matrix C' can be made symmetric, Since C' is also a very sparse matrix, an algorithm for solving sparse symmetric matrix equations could be applied to (4.3.43). The number of storage locations required for such a solution is on the order of Nx*Ny*Ny. The solution algorithm requires that the stored data be accessed in a relatively random manner. It is often necessary to allow Nx and Ny to be as large as 100 for a relatively complicated device. For a device requiring a large number of grid points this solution technique is not consistent with implementation on a minicomputer having limited available memory. Other direct solution algorithms exist which also require large amounts of randomly accessible information.

The structures of V and C' are well suited to the use of an iterative relaxation technique for solving (4.3.43). The number of storage locations required is on the order of Nx*Ny because it is only necessary to store V,Q' and the nonzero elements of C'. The solution algorithm allows the stored data to be accessed in a serial manner. This solution technique is consistent with implementation on a minicomputer where limited available memory requires most of the storage to be provided by a mass storage device. During each relaxation iteration the required mass storage accesses are limited to those necessary to perform one complete read and one complete write of the entire data set.

The relaxation techniques which are commonly used include successive overrelaxation (SOR) and successive line overrelaxation (SLOR) [4.15 ]. The SLOR algorithm typically exhibits convergence characteristics which are superior to those of the SOR algorithm. However, both techniques provide only first order error reduction versus iteration count. During each iteration the SOR algorithm separately improves

each element of the solution vector, In contrast, the SLOR algorithm groups the elements of the solution vector into blocks and separately improves each block as a unit. In this work the SLOR algorithm has been chosen for its improved convergence properties and its applicability to a system such as (4.3.43) having a block tridiagonal coefficient matrix.

The partitioning of C' which results in its block tridiagonal structure may be used to reformulate (4.3.43) as the following simultaneous matrix equations.

$$C'(i,i-1)\,V(i-1) + C'(i,i)\,V(i) + C'(i,i+1)\,V(i+1) = Q'(i);\ i=1, \ldots, N_x \qquad (4.3.44)$$

The unknown vectors V(i) in (4.3.44) represent the components of $V^{n+1}$ resulting from the partitioning of C'. The matrices C'(i,j) represent the blocks into which C' is partitioned. The application of one iteration of the SLOR algorithm to (4.3.44) is accomplished as follows

$$V^*(i) = [C'(i,i)]^{-1}\,[Q'(i) - C'(i,i-1)\,V^{m+1}(i-1) - C'(i,i+1)\,V^{m}(i+1)] \qquad (4.3.45)$$

$$V^{m+1}(i) = V^{m}(i) + \omega[V^*(i) - V^{m}(i)];\ i=1, \ldots, N_x$$

The subscripts m and m+1 are the SLOR iteration numbers. The relaxation factor is denoted by $\omega$ and must lie between 0 and 2. Values of $\omega$ less than unity correspond to underrelaxation while values greater than unity correspond to overrelaxation. It has been found in our work that relaxation factors in the range of 1.4 to 1.7 typically yield satisfactory convergence rates. Each relaxation iteration modifies the V(i) in the order of increasing X index. At each step the most recent approximations to V(i-1) and V(i+1) are used. It is possible to alternate the order

in which the X indices are taken and thereby improve the convergence of the iteration. However, such a scheme is inconsistent with the use of a serially accessible mass storage device which must be utilized when the algorithm is implemented on a minicomputer.

The maximum number of SLOR iterations which are required at each Newton-Raphson iteration step depends upon the degree of convergence which is desired in the solution of (4.3.44). Although the Newton-Raphson iteration should provide second order reduction, only first order global error reduction is achieved because of the use of the SLOR iteration to solve (4.3.44). As the number of SLOR iterations performed during each Newton-Raphson iteration is reduced the effects of the nonlinearities in (4.3.42) have a larger effect on the convergence rate. The limiting case is the use of a single SLOR iteration for each Newton-Raphson iteration step. This is the approach which has been used in our work. It has the advantage of reducing the storage requirements since it is no longer necessary to precalculate and store Q' and the diagonal elements of C'.

The number of Newton-Raphson iterations which are necessary to achieve convergence depends on several factors. As the number of SLOR iterations per Newton-Raphson iteration decreases the total number of required Newton-Rapohson iterations increases. This is a result of the dominance of the error reduction characteristics by the SLOR iterations. In order to stop the Newton-Raphson iterations a convergence criteria is necessary. A standard method is to require that the percentage change in potential fall below some upper bound. The imposition of such a restriction at every grid point within the device boundary is far too

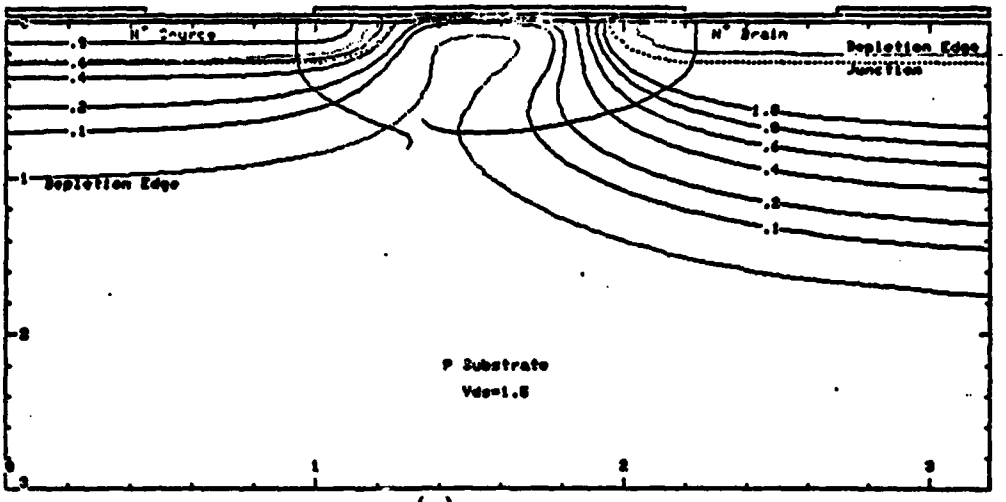
severe. The potential in the region of most concern to the device designer usually converges long before the convergence criteria is satisfied in the remainder of the device. Thus, a more reasonable approach is to choose only a small set of grid points where the convergence is monitored. For example, this set of points might be taken along the signal channel in a charge coupled device or along the metallurgical junction in a reverse biased p-n junction. We have found that such an approach works quite satisfactorily.

### 4.3.7 Conclusion

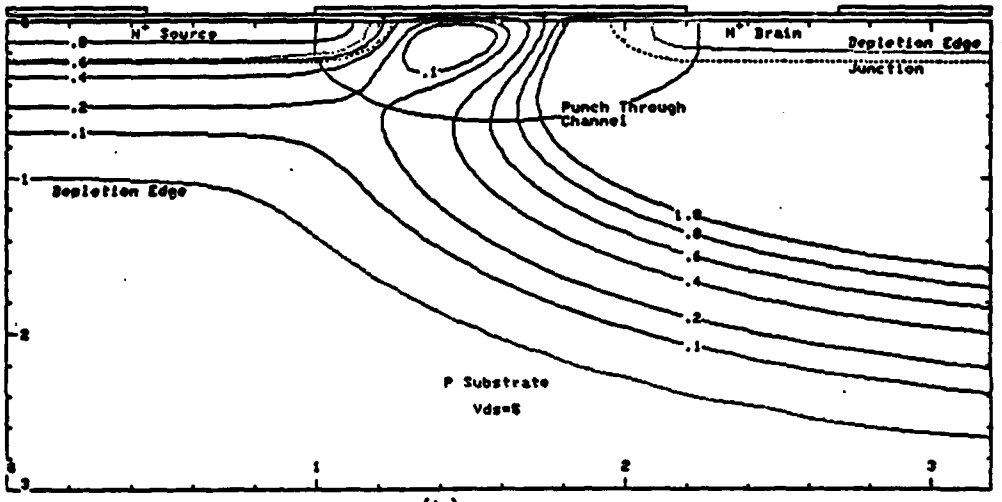
This section has described factors which must be considered when developing computer programs to solve Poisson's equation in MIS devices for VLSI applications. In particular, the treatment of two dimensional impurity distributions and highly nonuniform device boundaries is valuable in this context. It is important to provide a unified analysis capability ranging from process modeling to device characterization. To accomplish this goal the output from SUPREM can be used as input to a computer program which solves Poisson's equation. The SUPREM outputs determine the impurity distributions along key lines through the device substrate. The impurity distribution in the entire substrate is then calculated by using appropriate transformations.

Several applications for solutions to Poisson's equation have been mentioned. Those which are the most important to VLSI technology include the following,

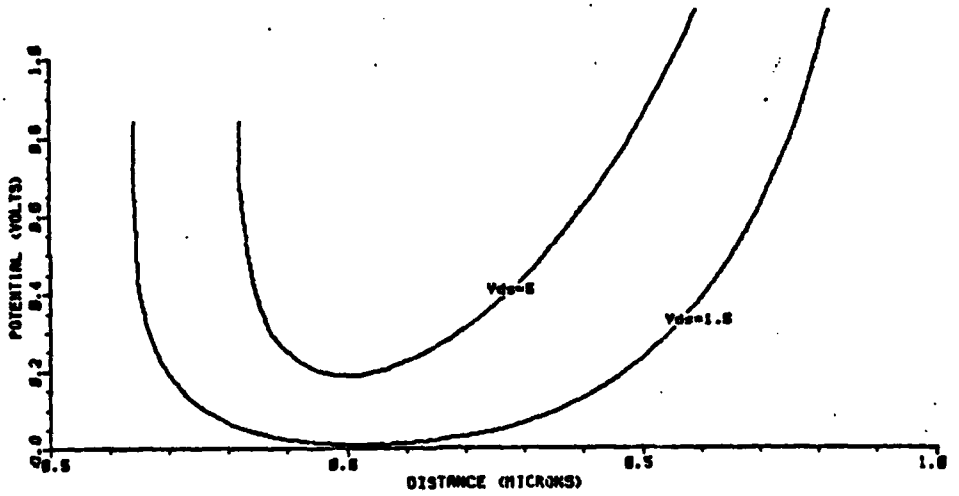
1) Threshold voltage calculation
2) Investigation of the onset of punch through
3) Calculation of device characteristics near and below threshold
4) Calculation of device capacitances.




(a)




(b)




(c)

Figure 4.3.9a Equipotential contours before the onset of punch through. The approximate location of the developing bulk channel is shown.

4.3.9b Equipotential contours after the onset of punch through. The location of the bulk channel is shown.

4.3.9c Potential along the bulk channel between the source and drain for the situations illustrated in a and b.

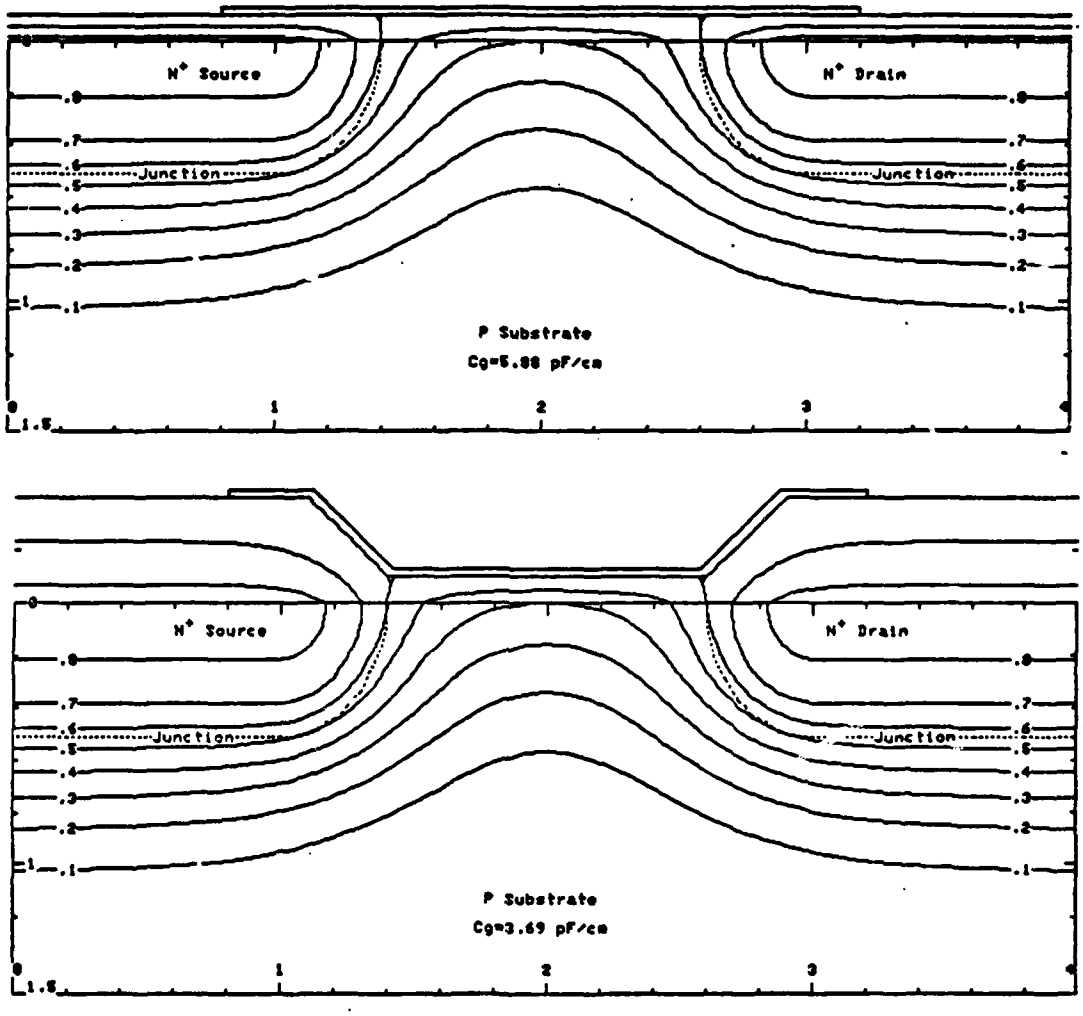
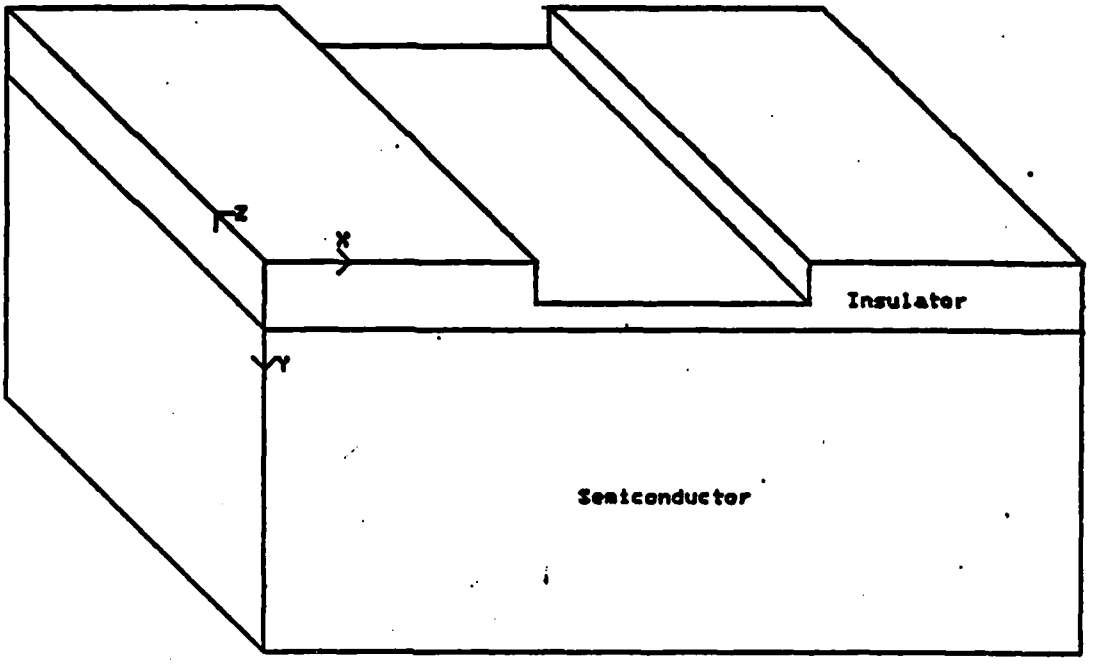
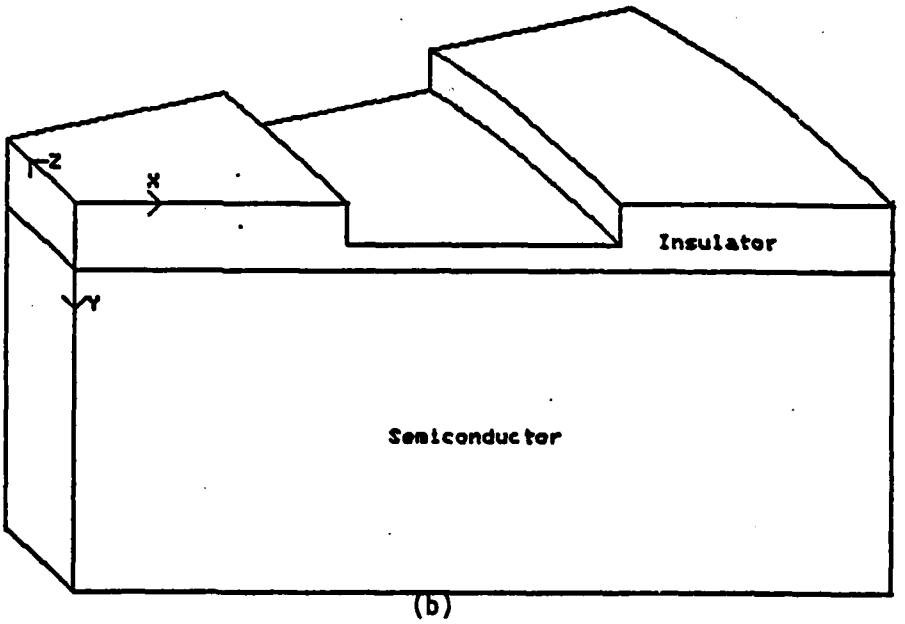



Figure 4.3.10 Equipotential contours for a MOSFET with two different oxide thicknesses over the source and drain regions. The gate electrode capacitances are indicated for each case.
(a) thin oxide
(b) thick oxide




(a)




(b)

Figure 4.3.11a Coordinate axis orientations for a rectangular coordinate system.

4.3.11b Coordinate axis orientations for a circular cylindrical coordinate system.

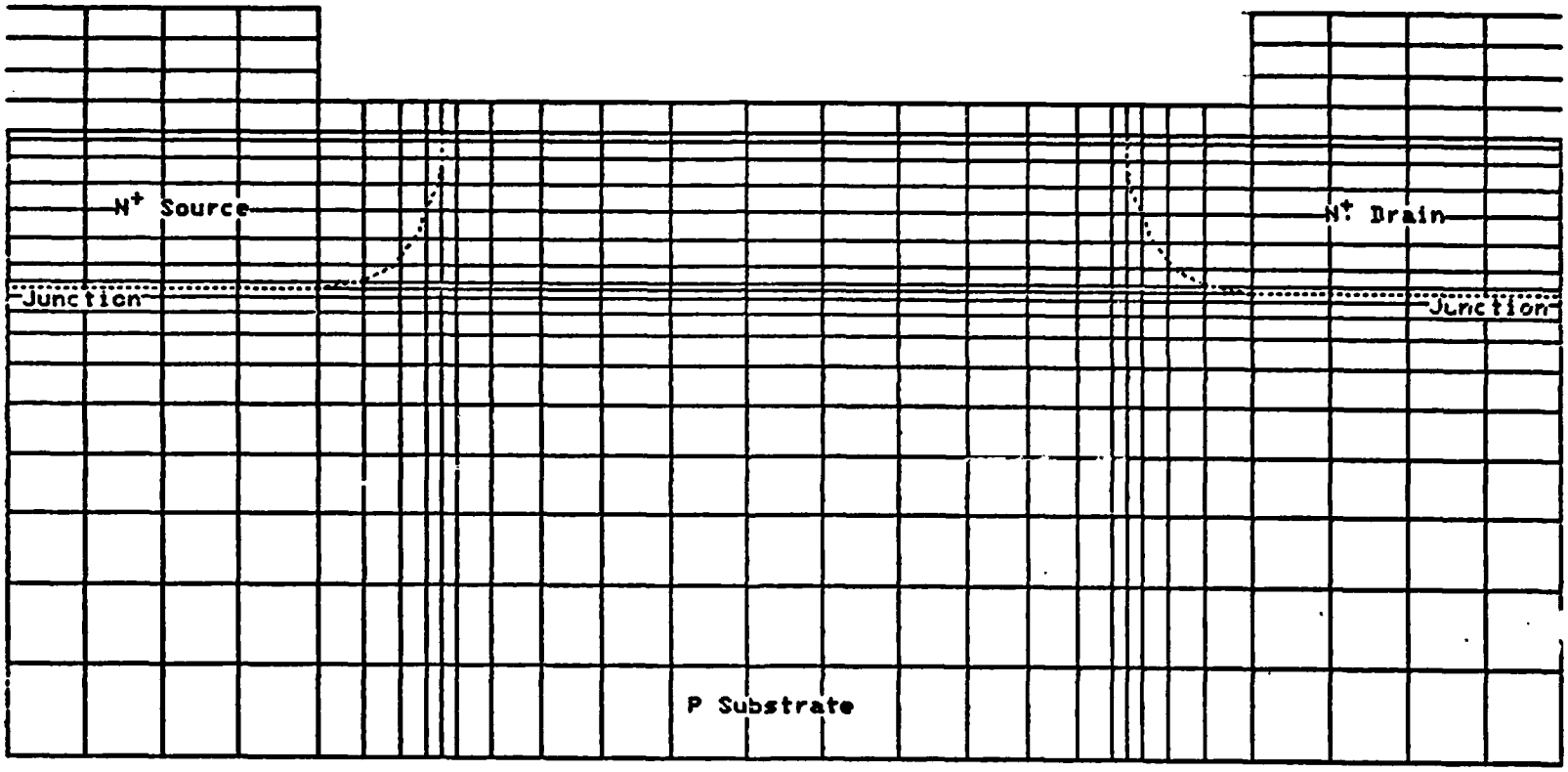



Figure 4.3.12 Typical nonuniform finite difference discretization grid for a MOSFET.

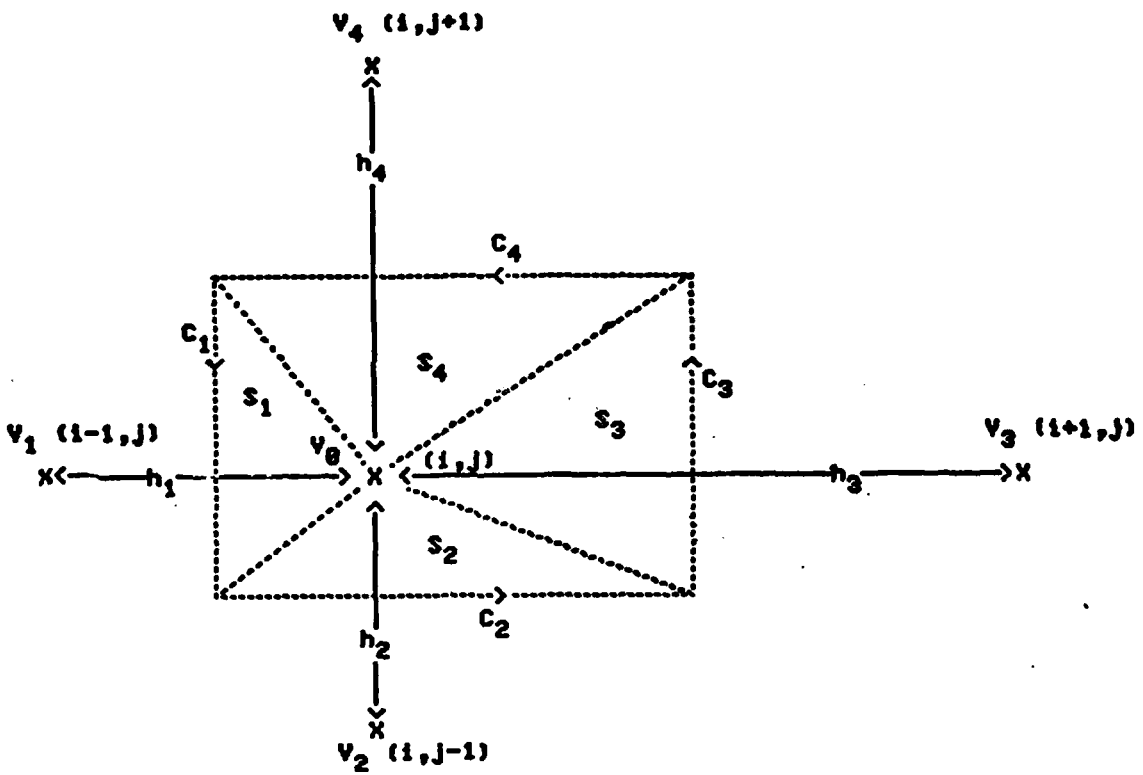



Figure 4.3.13 The five grid points and the associated terminology involved in the discretization of Poisson's equation at a general grid point.

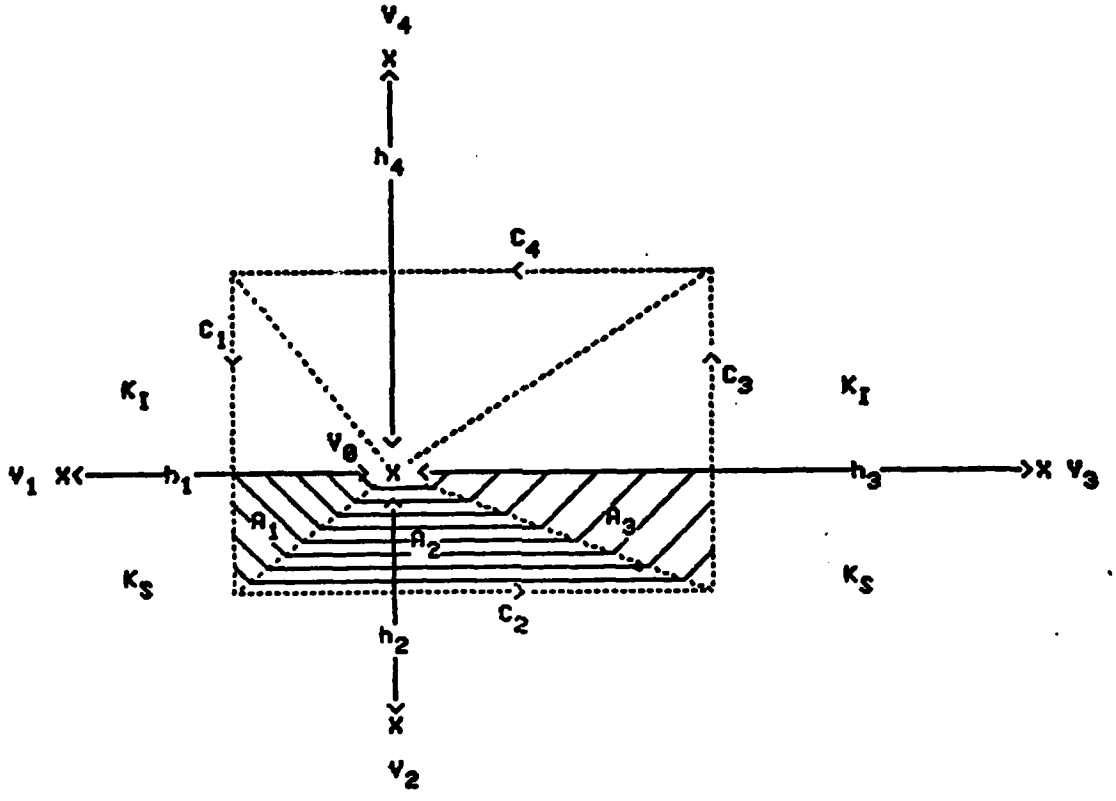



Figure 4.3.14 The five grid points and the associated terminology involved in the discretization of Poisson's equation at a grid point lying along the insulator - semiconductor interface.

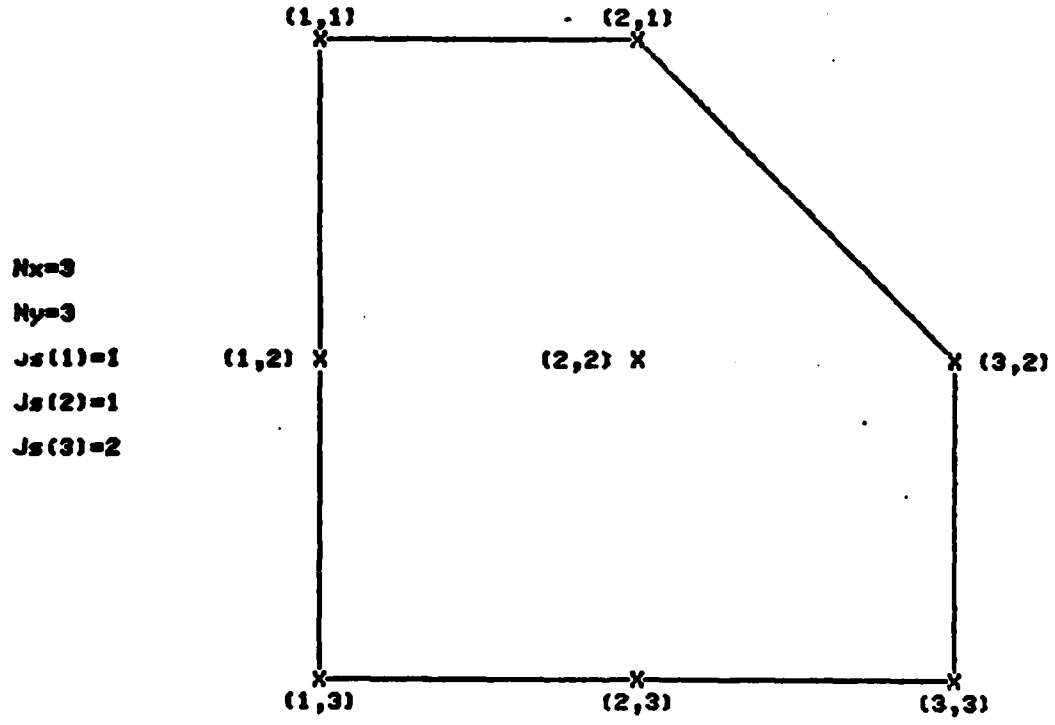



Figure 4.3.15 Grid index labels for a simple device structure.

$$\Psi = \begin{bmatrix} V(1,1) \\ V(1,2) \\ V(1,3) \\ \hline V(2,1) \\ V(2,2) \\ V(2,3) \\ \hline V(3,2) \\ V(3,3) \end{bmatrix} \qquad Q = \begin{bmatrix} Q[V(1,1)] \\ Q[V(1,2)] \\ Q[V(1,3)] \\ \hline Q[V(2,1)] \\ Q[V(2,2)] \\ Q[V(2,3)] \\ \hline Q[V(3,2)] \\ Q[V(3,3)] \end{bmatrix}$$

$$C = \left[\begin{array}{ccc:ccc:cc} C_1(1,1) & C_5(1,1) & 0 & C_3(1,1) & 0 & 0 & 0 & 0 \\ C_4(1,2) & C_1(1,2) & C_5(1,2) & 0 & C_3(1,2) & 0 & 0 & 0 \\ 0 & C_4(1,3) & C_1(1,3) & 0 & 0 & C_3(1,3) & 0 & 0 \\ \hline C_2(2,1) & 0 & 0 & C_1(2,1) & C_5(2,1) & 0 & 0 & 0 \\ 0 & C_2(2,2) & 0 & C_4(2,2) & C_1(2,2) & C_5(2,2) & C_3(2,2) & 0 \\ 0 & 0 & C_2(2,3) & 0 & C_4(2,3) & C_1(2,3) & 0 & C_3(2,3) \\ \hline 0 & 0 & 0 & 0 & C_2(3,2) & 0 & C_1(3,2) & C_5(3,2) \\ 0 & 0 & 0 & 0 & 0 & C_2(3,3) & C_4(3,3) & C_1(3,3) \end{array}\right]$$

Figure 4.3.16 Grid point ordering and matrix partitioning for the grid in Figure 4.3.15. The structure of the discretization coefficient matrix C is shown.

## 4.4 A SIMPLIFIED TWO DIMENSIONAL NUMERICAL ANALYSIS OF MOS DEVICE - DC CASE

Soo Young Oh

### 4.4.1 Introduction

In order to analyze the transient phenomena like the charge pumping in MOS devices, two dimensional transient analysis is essential, since the transient in MOS devices is basically two dimensional. Two dimensional MOS device problems in steady state have been analyzed by the finite difference method [4. 16 ] [4.17 ] or finite element method [4. 18 ], [4. 19 ] as reported last year. However the transient analysis with no dielectric relaxation time limitation [4. 22 ] can be done only by the simultaneous solution of Poisson's equation and the hole and electron continuity equation. The simultaneous solution method increases the number of variables three times more than that of the steady state solution by the iteration method. Hence it increases the CPU time and memory allocation tremendously. For example, 400-1000 node points are necessary to solve a typical two dimensional transient MOS problem by the finite element program, SEDAN, which has been developed by Stanford University and uses the simultaneous solution method. Due to the large analysis grid and the required simultaneous solution method, excessive CPU time and memory allocations result. Typically, 800 k - 2 mega byte memory and the CPU time for one bias point is of the order of 1-2 minutes by SEDAN using an IBM 370-168.

Due to the CPU time and memory requirements, two dimensional analysis using the finite difference and finite element method is of marginal engineering utility for DC analysis and totally impractical

for transient analysis. Some simplifications should be made to reduce the CPU time and memory allocations for the transient analysis. When the topology and transport effect for MOS devices are investigated carefully, the finite difference or finite element method mentioned above are found to overspecify the problem. First, current flows mainly along the channel and is carried by the majority carriers of the source and drain. Hence, the current continuity equation can be solved only for the majority carriers. Nonetheless Poisson's equation needs to be solved over the entire region. In the work described here, Poisson's equation in MOS devices is reduced to the initial solution and the boundary value problems for the incremental potentials with the relevant incremental boundary conditions. In addition, the carrier transport is solved along the channel with appropriate field coupling to the boundary conditions. In this method, nodes are allocated only along the boundary and channel. Therefore, the number of the nodes are much fewer than the finite difference and finite element method. This new method is an order of magnitude faster than the result quoted above, due to the substantial reduction in nodes allocated for analysis.

In Section 4.4.2 , the basic theory of the new method will be discussed. In Section 4.4.3, its practical implementation will be discussed. In Section 4.4.4, very short channel MOSFET devices are simulated by the new method and compared with the simulation by CADDET. The applications and prospects of the new method will be discussed in Section 4.4.5.

### 4.4.2 Semiconductor Device Equations and Their Simplification for MOS Devices

#### A. Semiconductor Device Equations

Semiconductor devices can be completely described by Poisson's equation, the hole continuity equation and the electron continuity equation [4.28]. These equations are as follow:

$$\bar{\nabla} \cdot (\varepsilon \bar{\nabla} \psi) = -q(p - n + N_d) \tag{4.4.46}$$

$$\frac{\partial p}{\partial t} + \bar{\nabla} \cdot \left(\frac{\bar{J}_p}{q}\right) = - R \tag{4.4.47}$$

$$\frac{\partial n}{\partial t} + \bar{\nabla} \cdot \left(\frac{\bar{J}_n}{-q}\right) = - R \tag{4.4.48}$$

where

| | | |
|---|---|---|
| $\psi$ | : | potential distribution |
| $P$, $n$ | : | hole and electron density |
| $N_d$ | : | ionized net doping density |
| $R$ | : | net electron-hole pair recombination rate |
| $J_n$, $J_p$ | : | hole and electron densities |
| $-q$ | : | electron charge |
| $t$ | : | time |

In order to calculate the distributions of $\psi$, p and n, these three equations should be solved simultaneously over the entire region with the given boundary conditions. However, when one considers MOS devices, several simplifications can be made without significantly affecting the accuracy.

For practical reasons, a two dimensional device geometry will be considered so that uniformity is assumed in the third (Z) direction. In this paper, only n-channel MOS devices will be treated. A p-channel MOS device can be simulated with an appropriate transformations.

B. Simplifications for MOS Devices.

Three assumptions allow the MOSFET analysis problem to be reduced to a rather simplified form. First, it is assumed that the main current flow is between the source and drain and is a result of electron current. Hole current is negligible in most cases so that the hole continuity equation can be omitted and the hole quasi fermi level can be held constant with the same value as the fermi level of the substrate throughout the region.

Second, it is assumed that for most cases the electron current flows only along the channel* in the device except in the saturation and punch-through regions. Hence, there is no reason to solve the electron continuity equation in the bulk but rather it can be solved only in one dimension along the channel.

Third, the effect of the mobile carriers in the depletion region is assumed negligible compared to the fixed ionized doping. The depletion region is not the region of action, however the depletion region consumes many nodes points in the finite difference and finite

*The channel or current path can be at the surface or in the bulk [4.29]. It is required that the current path should be sufficiently thin so that it can be treated as one dimensional path.

element methods since the potential changes rapidly in this region. In order to overcome this wasteful use of grid in computing potential, Poisson's equation is reduced to a boundary value problem for the incremental potential. Hence, the problem is reduced from a bulk solution on a two-dimensional grid to a boundary value problem to be solved on a one dimensional type grid located on the boundary only. Since this simplification of Poisson's equation is the key means to reduce the number of nodes and yields a more efficient solution, the method will be discussed in some detail in the next section.

C. Simplification of the Potential Calculation

For studying the distribution of the potential and charges of MOS devices with bias, one can change the bias voltage by a small increment $\Delta V$ and study the resulting changes of the potential and charges. Figure 4.4.17 shows the change of the depletion region and the induced charges on the gate and in the channel. When the depletion approximation is assumed, no mobile carriers are in the depletion region except the channel. Hence, the incremental charges are induced only around the boundary of the depletion region, on the gate and in the channel. These incremental charges are the only sources of the incremental potential. For these increments, the problem is equivalent to a boundary value problem whose boundary includes the depletion region in the silicon as well as theoxide region and source terms are the induced surface charges located on the boundary and in the channel.

The above concept can be expressed mathematically as follows. The potential in one point is the sum of the potential contributions from the charges in the whole space. Hence, potential can be expressed as follows:

$$\psi(\bar{x}) = \int_S \rho(\bar{x})\, G_N(\bar{x}, \bar{x}')\, d^2x' \qquad (4.4.49)$$

where

S = the whole space

$G_N(\bar{x}, \bar{x}')$ = Green's function that satisfies the following conditions in two dimension.

$$\nabla^2 G_N(\bar{x},\bar{x}') = \frac{1}{\epsilon}\, \delta(\bar{x}-\bar{x}')$$

$$\lim_{|\bar{x}-\bar{x}'| = \infty} G_N(\bar{x},\bar{x}') = 0$$

If the incremental form of equation (4.4.49) is taken,

$$\Delta\psi(\bar{x}) = \int_S \Delta\rho(\bar{x}')\, G_N(\bar{x}, \bar{x}')\, d^2x' \qquad (4.4.50)$$

If $\Delta\psi$ is small, the changes of the depletion regions are small and the incremental charges can be treated as a sheet of surface charge.

$$\Delta\psi(\bar{x}) = \int_\Gamma \Delta\sigma(\bar{x}')\, G_N(\bar{x}, \bar{x}')\, G_N(\bar{x}, \bar{x}')\, d^2x' \qquad (4.4.51)$$

where

$\Gamma$ = the depletion boundary, the channel and gate

$\Delta\sigma$ = surface changes on $\Gamma$

Eq. (4.4.51) expresses the incremental potentials only in terms of the induced incremental surface charges at the channel and the boundary. The fixed charges in the bulk do not contribute to the incremental potential. Thus the allocation of the nodes inside the bulk can be eliminated. In this way, the potential problem can be reduced from a bulk solution on a two dimensional grid to a boundary value problem to be solved on a one dimensional grid located on the boundary and in the channel. This reduction of the node allocations makes it possible to decrease the CPU time and memory allocation by orders of magnitude.

In order to implement this method for MOS devices, the potential must be solved explicitly in two dimensions for an initial bias voltage. The solution for the other bias voltages can be solved by the boundary value problem with the corresponding small incremental bias voltages formulated above. The specific procedure is as follows:

1) The initial solution is calculated using a two dimensional grid and either a finite difference or finite element solution method.
2) Using the initial solution, the boundaries of the depletion regions are calculated.
3) Next the boundary value problem with a specific small incremental bias is solved together with the one dimensional electron continuity equation along the channel for the incremental potentials and electron densities.
4) Terminal currents are calculated. The incremental channel and boundary surface charges are calculated and the boundaries of the depletion regions are updated.

5) The same procedure is repeated to the final bias voltage.

In essence, the bias is stepped to obtain the new solution in a manner very much like curve tracer measurements.

### 4.4.3 Formulation and Solution of the Matrix Equation for MOS Devices

As outlined above, the semiconductor device equations can be simplified to a boundary value problem and the one dimensional electron continuity equation along the channel. The boundary value problem is equivalent to Poisson's equation result. The one dimensional electron continuity equation along the channel is used to obtain current flow. In this section, the practical implementation of the new method will be discussed in detail.

A. Simplification of the Geometry for the Practical Implementation

In MOS devices, the significant amount of the incremental surface charge is induced only along the boundaries confined by the source, gate drain and substrate. In this case, it is enough to consider the simple rectangular region as illustrated in Figure 4.4.18. The effects of the incremental charges outside the region can be represented by the proper boundary conditions along the rectangular boundary using the equivalent principle in the electromagnetic theory [ 4.20 ]. This simplification method is illustrated in Appendix 4.4.3.

B. Discretization of the Boundary Value Problem - Poisson's Equivalent

When the geometry of the device is simplified and fixed, the boundary value problem can be completely specified if the relationships between the incremental potential and surface charge on the

boundary and in the channel are determined. In order to find these relations, the boundary is divided into $N_d$ small cells and the channel is divided into $N_c$ small cells as illustrated in Figure 4.4.19. If the incremental potential in each cell is assumed constant over the cell, the small cells can be treated as sheets of conductors. Then the capacitance matrix* can be formulated to give the relation between the incremental potential and surface charge in the cells:

$$\sum_{j=1}^{N} C_{ij} \Delta\psi_j = \Delta Q_i \qquad i = 1, N \qquad (4.4.52)$$

where

$\Delta\psi_j$ = incremental potential in cell j

$\Delta Q_i$ = incremental surface charge in cell i per unit width

$C_{ij}$ = capacitance matrix

$N_d$ = number of cells on the boundary

$N_c$ = number of cells on the channel

$N = N_c + N_d$

Each element of the capacitance matrix, $C_{ij}$ is the induced incremental charge in cell i due to the incremental voltage $\Delta\psi_j$. The incremental potential $\Delta\psi_i$ on the boundary are given by the boundary whereas the $\Delta\psi_i$ on the channel are unknown. However there are additional relations on the channel inversion layer, given by

$$\Delta Q_i = q \cdot \ell_i \cdot \Delta X_{ch} \cdot n_i \left[ e^{(\psi_i + \Delta\psi_i - \phi_{fn}, i)} - e^{(\psi_i - \phi_{fn}, i)} \right] \qquad (4.4.53)$$

where

$\ell_i$ = length of the ith cell

*the detailed derivation of the capacit nce matrix in the finite region will be given in Appendix 4.4.4.

$\Delta X_{ch}$ = channel thickness

$\phi_{fn}$, i = electron quasi fermi potential on the ith cell

$n_i$ = intrinsic carrier concentration

Hence, if equation (4.4.43) is used in (4.4.42), there are N unknown and N equations. Therefore, $\Delta\psi_i$ along the channel and $\Delta Q_i$ on the boundary can be evaluated using equations (4.4.42) and (4.4.43) as well as the boundary conditions.

C. Discretization of Electron Continuity Equation

As explained in Section 4.4.2, the hole continuity equation can be neglected for n-channel MOS devices and only the electron continuity equation is solved along the channel. As a result, a one dimensional electron continuity equation is solved along the channel:

$$\frac{dn}{dt} = \frac{1}{q}\frac{dJ_n}{dq} - R \qquad (4.4.54)$$

The Gummel-Scharfetter expression [4.21] is used for $J_n$

$$J_n(i+\tfrac{1}{2}) \; q \cdot \mu_n(i+\tfrac{1}{2}) \cdot E(i+\tfrac{1}{2}) \; \frac{n(i+1)}{1 - e^{-\beta \cdot E(i+\frac{1}{2}) \cdot \Delta y_i}}$$

$$+ \frac{n(i)}{1 - e^{\beta \cdot E(i+\frac{1}{2}) \cdot \Delta y_i}} \qquad (4.4.55)$$

where

$\beta$ : q/kT

$\mu_n$ : electron mobility

Equation (4.4.54) can be discretized as follows

$$\frac{n(i)^{t+1} - n(i)^{t}}{\Delta t} = \frac{1}{q} \; \frac{J_n^{t+1}(i+\frac{1}{2}) - J_n^{t+1}(i-\frac{1}{2})}{(\Delta y(i+1) - \Delta y(i))/2} \qquad (4.4.56)$$

Where the appropriate index change using equation (4.4.55) is implied. Infinite recombination velocity is assumed at the boundary of the source and drain contacts. Hence, electron densities are the same as the equilibrium value at the source and drain contact. An implicit time scheme [4.22] is used to discretize the time dependent term. The numerical solution of equation (4.4.56) results in the terminal currents required at each bias point. The overall solution implementation will be discussed next.

D. Flow Chart of the Program

The method described above has been implemented on the computer; the flow chart is illustrated in Figure 4.4.20. First, input data are read and the initial Poisson solution is generated using finite difference solution method. Next the depletion boundary is determined and divided into small cells. The typical number of the cells on the boundary and channel ranges from 40 to 60. After these initial steps, bias voltages are increased incrementally. When the bias voltages are given, the capacitance matrix equation (4.4.49) and the one dimensional electron continuity equation (4.4.56) are solved simultaneously for $\Delta\psi_i$ and n(i). Equation (4.4.52) and (4.4.56) are nonlinear and Newton's method is used at each iteration. When the convergent solution is obtained, the terminal current is calculated as well as $\Delta Q_i$ on the boundary. The change of the depletion region boundary and updated boundary are calculated using $\Delta Q_i$ and the specified doping density. The above procedure is repeated until the final bias potential is achieved.

Due to the small number of cells, the implementation of the simultanous solution requires only a small CPU time and memory allocation.

Furthermore, the simultaneous solution using Newton's method guarantees the convergence of the solution at any bias condition. The average CPU time for one bias point is of the order of 1 second using the IBM 370-168. This is an order of magnitude faster than results of the finite difference method program, CADDET.

Moreover, the incremental method with the simultaneous solution method shows minimal effects at higher bias voltages, where this is not true for CADDET because it uses the iteration between Poisson's equation and electron continuity equation which fails in the strong inversion cases. Table 4.4.4 summarizes the iteration and total CPU time/bias point differences between the two methods.

4.4.4 Discussion of the Accuracy of the New Method Compared with CADDET

In the practical implementation of the new boundary value method, several approximations are used to simplify and reduce the computations. The validity of these approximations should be checked. All the approximations used in the new method are summarized and discussed in Appendix C. In order to check the accuracy of the method and to identify the sources of inaccuracy, the simulation results of BV method have been compared with the results of CADDET. These results are not compared with the measurements of the real devices, since additional errors can be introduced such as the effective channel length and doping profile specifications. A short channel MOS device with the 2 μm effective channel length is used in comparison. The geometry of the device is illustrated in Figure 4.4.21. The source and drain region are approximated as a uniformly doped rectangular region. The doping concentrations are 1.E20/cm**3 for each region. The substrate has uniform doping of 1.E16/cm**3 .

The device parameters used for both programs are shown in Table 4.4.5. The initial solution for the new method is taken from CADDET at the bias of $V_{gs}$ = 2.0 v and $V_{ds}$ = 0 V, which is in the subthreshold region.

Two primary sources of the inaccuracy are identified. The most significant is the approximated boundary conditions on the simplified rectangular region, especially the boundary condition between the drain and bulk for high $V_{ds}$ bias. Two different approximation methods are used to evaluate the boundary conditions between the source and the bulk, and drain and bulk. The first one is the one dimensional planar p-n junction approximation as illustrated in Figure 4.4.22. The second one is the one dimensional cylindrical p-n junction approximation as illustrated in Figure 4.4.23. The results using both approximations are compared with CADDET in Figure 4.4.23. The approximations of these boundaries have significant effects on the output conductance of $I_{DS}$ vs. $V_{GS}$ characteristics in the saturation region. The cylindrical approximation gives more reasonable results as expected. However, it still gives a slightly larger output conductance. More efforts will be needed to accurately approximate these boundary conditions.

The next major source of the inaccuracy is the approximation of the finite thickness incremental charges on the depletion boundary as the infinitesimal thickness of the surface charge in the boundary value problem. Due to the above approximation, the bias voltage step is limited to less than 0.05 V in the subthreshold region to minimize the inaccuracy due to the above approximation and hold it to within 1% of the CADDET result. However, much larger bias steps can be taken above threshold voltage due to the screening effects of the channel.

As illustrated in Figure 4.4.25, the bias step $\Delta V = 0.5$ V can be taken and still maintain 3% error -- steps of $V = 1$ V result in only 5% inaccuracies.

Using the one-dimensional cylindrical p-n junction approximation on the boundary, full characteristics of the n channel MOS devices are simulated and compared with the results of CADDET in Figures 4.4.26-28. In Fig.4.4.10, $I_{DS}$ vs $V_{GS}$ characteristics in the linear region is plotted. The drain voltage is set to 0.1 V. The gate voltage is varied from 2.0 V to 4.0 V. The agreement is very good in the linear region except the slight difference in the slope. The same data is plotted in log scale to compare the currents in the subthreshold region as illustrated in Figure 4.4.27. $I_{DS}$ vs. $V_{DS}$ characteristics in the linear and saturation region are plotted and compared in Figure 4.4.28. The output conductance in the saturation region of the new method is slightly larger than that of CADDET as explained above. The overall agreement between the new method and CADDET is very good in spite of the approximations mentioned above. The discrepancy is less than 5% in all cases. The potential distribution along the channel in both cases are plotted and compared also in Figure 4.4.29. The gate voltage is set to 3.0 V and the drain voltage is varying from 0.2 V to 1.8 V with 0.4 V step. The agreements are good except the discrepancy near the drain. That is due to the difference in the drain boundary condition mentioned in Appendix 4.3.5.

4.3.5 Conclusion

A new method has been developed for fast and simplified two dimensional simulations of MOS devices. Two dimensional Poisson's equation

is reduced to the initial solution and the boundary value problem. The current transport is solved along the channel with appropriate field coupling to the boundary conditions. These simplifications lead to the substantial reductions in nodes allocated for analysis. In spite of these simplifications, the simulations of the short channel MOSFET by the new method are very good agreements with the results of CADDET.

The new program is very suitable for the mini-computer implementation and the two dimensional transient analysis, due to its small memory requirement and fast speed. The small number of node points make the speed of the new program faster than that of the equivalent finite difference and finite element method programs. Each dc point takes about 1-2 seconds on IBM 370-168 which is an order of magnitude faster than CADDET.

## APPENDIX 4.4.3

## EQUIVALENCE PRINCIPLE AND UNIQUENESS THEOREM

Equivalence Principle [4.20]

Many sources distributed outside a region can produce the same field inside the region. Two sources which produce the same field within a region of space are said to be equivalent within that region. For consideration of the field in a given region, it is not always necessary to know the actual sources, since the equivalent sources will serve as well. Using the uniqueness theorem, many electrostatic field problems can be simplified when the actual source is replaced by the corresponding boundary conditions or equivalent sources. The uniqueness theorem is summarized as below:

Uniqueness theorem

In a closed boundary region, the potential distribution is uniquely specified if

1) The potential is given over the boundary (Dirichlet B. C.)
2) Normal derivative of the potential is given along the boundary (Neumann B. C.)
3) The potential is given over a part of the boundary and its normal derivatives over the remaining part (mixed B. C.)

Using the uniqueness theorem and the equivalence principle, the potential generated by the incremental charges in Figure 4.4.17 can be equivalently generated within the rectangular region by the proper boundary conditions as illustrated in Figure 4.4.18.

In the boundary value problem of Figure 4.4.18, the incremental boundary potentials on the source, gate, drain and substrate are already given by the incremental bias. The incremental boundary potentials between the 4 terminals should be evaluated. Some approximated evaluation methods are discussed in Section 4.4.4.

## APPENDIX 4.4.4

## DERIVATION OF THE CAPACITANCE MATRIX IN THE FINITE REGION AND ITS PRACTICAL EVALUATION

1) Derivation of the Capacitance Matrix in the Finite Region

In order to derive the relation between the incremental potentials and surface charges in the finite region, equation (4.4.46) is modified as follows [4.23].

$$\psi(\bar{x}) = \langle\psi\rangle_\Gamma + \int_s \rho(\bar{x}')\, G_N(\bar{x},\bar{x}')\, d^2x' + \int_\Gamma \epsilon \frac{\partial\psi}{\partial n'} \cdot G_N\, dx' \qquad \text{(B-1)}$$

where

$s$ = the two dimensional finite region to be considered

$\Gamma$ = the boundary of s

$G_N(\bar{x},\bar{x}')$ = two dimensional Green's function that satisfy the following conditions

$$\nabla^2 G_N(\bar{x},\bar{x}') = \begin{cases} -\frac{1}{\epsilon}\ \delta(\bar{x}-\bar{x}') & \text{inside s} \\ 0 & \text{outside s} \end{cases}$$

$$\frac{\partial G_N}{\partial n'}(\bar{x},\bar{x}') = -\frac{1}{L_\Gamma}$$

$L_\Gamma$ = the total length of

$\langle\psi\rangle_\Gamma$ = average of $\psi$ on $\Gamma$

If small increments are taken, there are not significant changes in the boundary of the depletion region. Hence $G_N$ is constant and (B-1) becomes:

$$\Delta\psi(\bar{x}) = \langle\Delta\psi\rangle_\Gamma + \int_s \Delta\rho(x')\, G_N(\bar{x}, \bar{x}')\, d^2x' + \int_\Gamma \varepsilon \frac{\partial\Delta\psi}{\partial n} \cdot G_N dx' \quad \text{(B-2)}$$

$\varepsilon \frac{\partial\Delta\psi}{\partial n'}$ in the third term of the right hand side is identified as incremental surface charges along the boundaries using Gauss theorem. If the depletion approximation is used for MOS devices, there are only incremental channel charges in s. Hence, the surface integral of the second term can be reduced to the line integral as follows.

$$\int_s \Delta\rho(\bar{x}')\, G_N(\bar{x}, \bar{x}')\, d^2x' = \int_{\text{channel}} \Delta\sigma(\bar{x}')\, G_N(\bar{x}, \bar{x}')\, dx' \quad \text{(B-3)}$$

where

$$\Delta\sigma(\bar{x}') = \Delta\rho(\bar{x}') \cdot \Delta X_{ch}$$

$\Delta X_{ch}$ = channel depth

Then

$$\Delta\psi(\bar{x}) = \langle\Delta\psi\rangle_\Gamma + \int_{\text{channel}} \Delta\sigma(\bar{x}')\, G_N(\bar{x},\bar{x}')\, dx' + \int_{\text{boundary}} \nabla\sigma(\bar{x}')\, G_N(\bar{x},\bar{x}')\, dx' \quad \text{(B-4)}$$

If the boundary and channel are divided into N small cells and the mean value theorem [10] is used,

$$\Delta\psi(\bar{x}) = \langle\Delta\psi\rangle_\Gamma + \sum_{j=1}^{N} \int_j \Delta\sigma(\bar{x}') \cdot G_N(\bar{x},\bar{x}')\, dx' \quad \text{(B-5)}$$

$$= \langle\Delta\psi\rangle_\Gamma + \sum_{j=1}^{N} \Delta\sigma(\bar{\xi}_j) \left[\int_j G_N(\bar{x},\bar{x}')\, dx'\right]$$

where $\bar{\xi}_j$ is the point in the jth cell.

Using Garlerkin's method [4.25], both sides of (B-5) are multiplied by the shape function $\phi_i$, where $\phi_i$ is one in the ith cell and zero otherwise. Then it is integrated in the whole region. Then

$$\Delta\psi_i = \frac{1}{L_r} \sum_{j=1}^{N} \ell_j \Delta\psi \quad \sum_{j=1} \Delta\sigma(\bar{\xi}_j) \cdot \frac{1}{\ell i} \int_i \int_j G_N(\bar{x}_i, \bar{x}_j) dx_i \, dx_j \qquad (B\text{-}6)$$

$i = 1, N$

where

$\ell i$ : length of the ith cell

$$\Delta\psi_i = \frac{1}{\ell i} \int_i \Delta\psi(\bar{x}) \, dx$$

After some mathematical manipulation, (B-6) becomes

$$\Delta\psi_i = \sum_{j=1}^{N} P_{ij}' \, \Delta Q_j \qquad (B\text{-}7)$$

where

$$\Delta Q_j = \ell i \cdot \Delta\sigma(\bar{\xi}_j)$$

$$P_{ij} = \sum_{k=1}^{N} D_{ik}^{-1} P_{kj}$$

$$P_{ij}' = \frac{1}{\ell i \ell j} \int_i \int_j G_N(\bar{x}_i, \bar{x}_j) \, dx_i \, dx_j$$

$$D_{ij} = \delta_{ij} - \ell i / L_r$$

If (B-7) is inverted,

$$\Delta Q_i = \sum_{j=1}^{N} C_{ij} \Delta\psi_j \qquad i=1,N \qquad (B\text{-}8)$$

where $C_{ij} = P'^{-1}_{ij}$

The capacitance matrix can be obtained using the above definition. However in practical cases, conformal mapping method is used as illustrated in the next section.

2. Subregion Method

In order to treat MOS devices, multiple dielectric regions should be considerd. It is very complicated to derive the relation between $\Delta\sigma_i$ and $\Delta\psi_i$ in this case. A different approach has been taken to solve this

problem where an imaginary boundary wall is located along the channel as illustrated in Figure (B-1). Using the equivalence principle $\Delta\sigma_{ox} = \epsilon_{ox} \cdot \Delta\bar{E}_{ox} \cdot \bar{n}_{hx}$ is put on the side of the oxide region at the channel and $\Delta\sigma_{sf} = \epsilon_{sj} \cdot \Delta\bar{E}_{si} \cdot \bar{n}_{si}$ on the side of the silicon region. The boundary surface charges in the oxide region and $\Delta\sigma_{ox}$ generate the same incremental field as before in the oxide region but give zero field outside. Using the same argument, the boundary surface charges in the silicon region and $\Delta\sigma_i$ generates the same incremental fields in the silicon region and zero field outside. Therefore, the oxide region and silicon region can be separated using these equivalent sources $\Delta\sigma_{ox}$ and $\Delta\sigma_{si}$. After the separation, the relation between the incremental potentials and surface charges in each region can be formulated independently.

The evaluation of the capacitance matrix becomes much easier in this homogeneous dielectric case. After evaluating the capacitance matrix for each region, they can be combined to obtain the total capacitance matrix for the entire region using the relations as follows.

$$\Delta\psi_{ox} = \Delta\psi_{si} \tag{B-9}$$

$$\Delta\sigma_{channel} = \Delta\sigma_{ox} + \Delta\sigma_{si} \tag{B-10}$$

at the channel

The rules for combining $C_{ij}$ is:

1) $C_{ij} = C_{ij}^{ox}$ or $C_{ij}^{si}$ if i c channel or j c channel

2) $C_{ij} = C_{ij}^{ox} + C_{ij}^{si}$ i, i c channel and jc channel.

In subsequent discussion, only the homogeneous case will be considered.

3) The Conformal Mapping Method [4.26]

The potential and field distribution of a complicated boundary can be easily obtained from that of the simple boundary when conformal mapping and the theorem [4.26] presented hereafter are used

Theorem

If $\nabla \psi(u,v) = 0$ in W-plane, then, $\nabla^2 \psi[u(x,y), v(x,y)] = \nabla^2\phi(x,y) = 0$ in Z-plane when $Z = (w)$ is analytical.

where $\phi(x,y) = \psi[u(x,y), v(x,y)]$

According to the theorem above, the potential at one point in Z-plane is the same as the potential at the corresponding point in W-plane if the boundary conditions are properly mapped. The parallel plates in Figure (B-2a) can be mapped into the region in Figure (B-2b) such that the upper plate is mapped into the ith cell and the lower plate into the other boundary, using conformal mapping. Then the potential and field at the Z-plane are the same as those of the corresponding point in W-plane.

Hence, the incremental potential and field distribution are easily obtained for the given geometry and boundary conditions in Figure (B-2b). Once the incremental potential and field distribution are known, the capacitance matrix $C_{ij}$ can be easily obtained, because $C_{ij}$ is, by definition, the induced charge in the ith cell when $\Delta\psi_j$ is 1 V and others are zero.

4. Surface Charge Approximation of the Finite Thickness Incremental Charges at the Depletion Boundary

When the potential problem is reduced to the boundary value problem for the incremental potential, the finite thickness incremental charges on the depletion boundary are approximated as infinitesimal thickness surface charge assuming the incremental potentials are small so that the thickness of the incremental charges are thin enough. Due to the above approximation, the bias voltage step is limited 0.05 V below the threshold voltage to keep the accuracy within 1%. However, it turns out that the bias step can be taken up to 0.5 V within the accuracy of 3% above the threshold voltages.

5. The Approximated Boundary Conditions on the Simple Rectangular Region

In simplifying the geometry of the boundary value problem the effects of the incremental charges outside the region are equivalently replaced by the boundary conditions as illustrated in Figure4.4.18. There are no errors introduced in the above replacement. However, significant errors can be introduced when the approximated boundary conditions are used. The detailed discussion about it is in Section 4.4.4.

## APPENDIX 4.4.5

## DISCUSSION OF THE APPROXIMATIONS USED IN THE NEW METHOD

All the approximations used in the new method are summarized and their validities are discussed.

1. One Carrier Approximation

In n-channel MOS devices, the hole current contribution to the channel current is very small and can be neglected. Hence, only the electron continuity equation is solved. This is usually used in two dimensional MOS device simulations [4.27, 4.29].

2. One Dimensional Approximation of the Channel Current Transport

In usual bias cases of MOS devices, the current flows only along the surface except the punch-through and saturation region. In the saturation region, the effect of the channel spreading is small. Hence, the approximation can be used in the saturation region without significant errors. However, the new method cannot be used in the punch-through region due to the failure of this approximation.

3. Abrupt Junction Approximation at the Drain and Source Boundary

The source and drain boundaries for the one dimensional electron continuity equation are located at the metallurgical junction using the abrupt junction approximation. Due to above approximation, the potential and electron concentration near the drain and source of the new method is slightly different with the results of CADDET, illustrated in Figure 4.4.29. However, this inaccuracy of the boundary point is not significant considering the small width of the depletion region in the high concentration region.

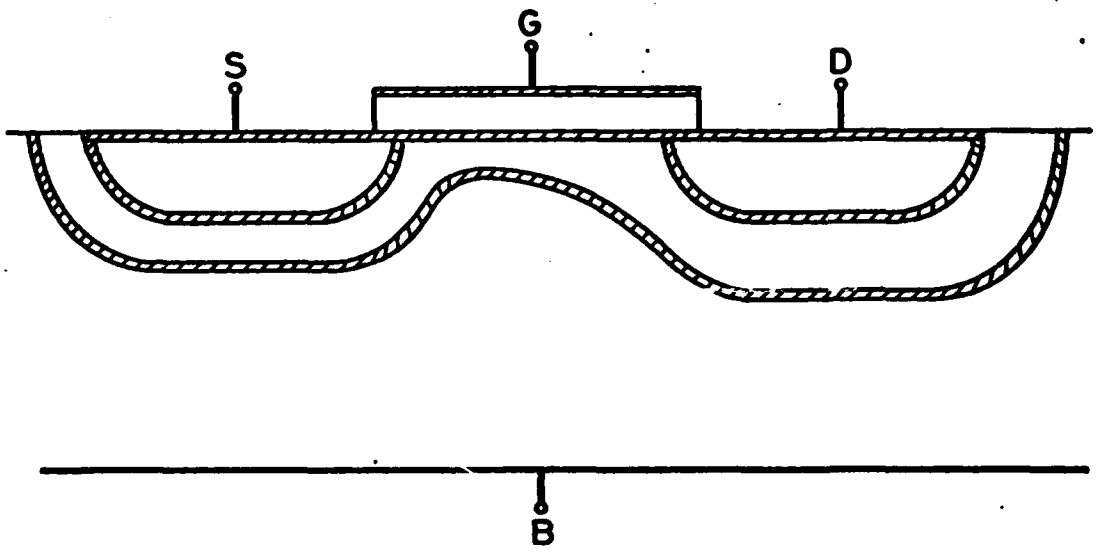



Figure 4.4.17 Induced incremental charges around the depletion region of the source, drain and bulk, and on the gate and channel in the MOS device.

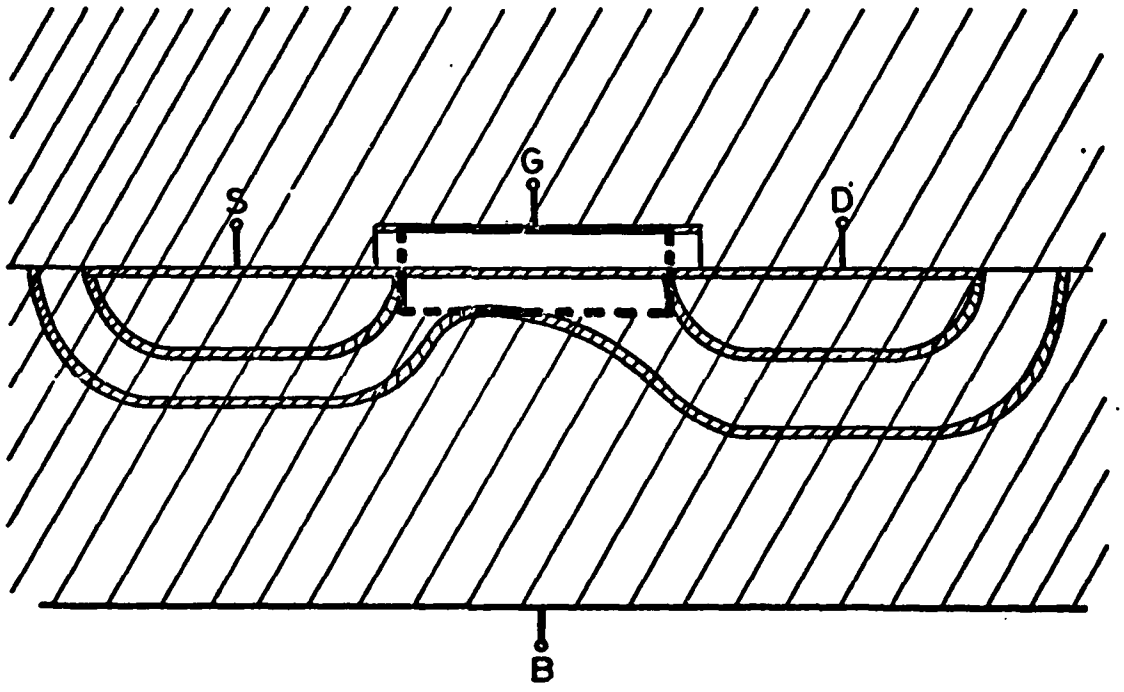



Figure 4.4.18 Simplification of the boundary value problem on a whole space with the actual source to the equivalent boundary value problem on a rectangular region with the equivalent boundary conditions.

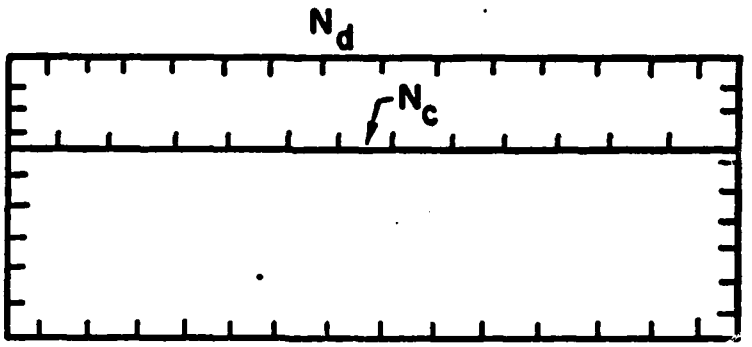



Figure 4.4.19 Discretization of the boundary and channel into small cells.

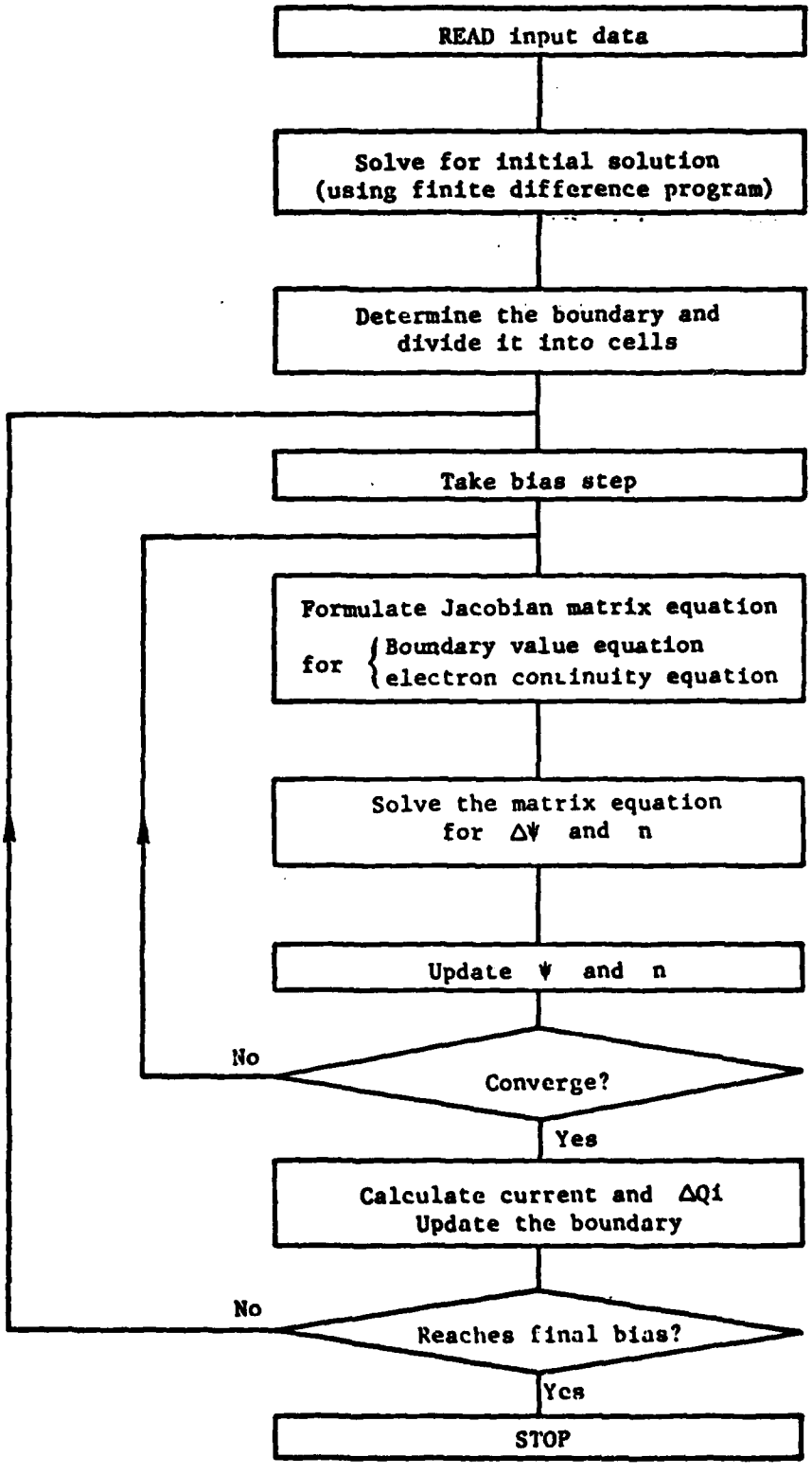



Figure 4.4.20 Flow chart of the new boundary value method program.

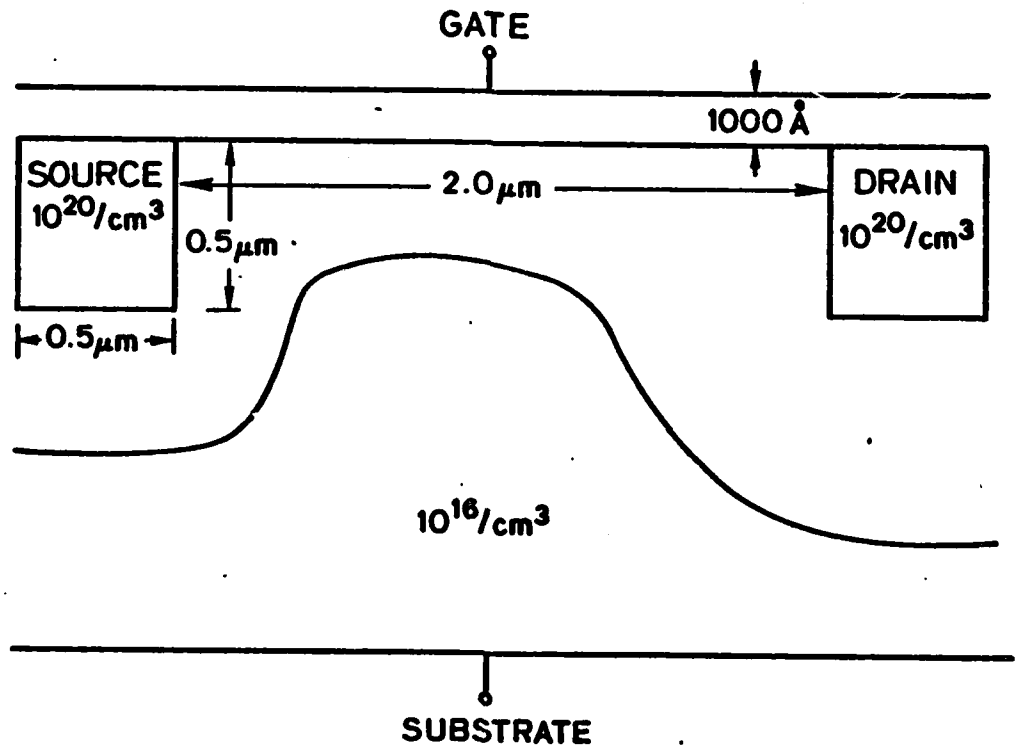



Figure 4.4.21 Topology and device parameters of the simulated short channel MOS device.

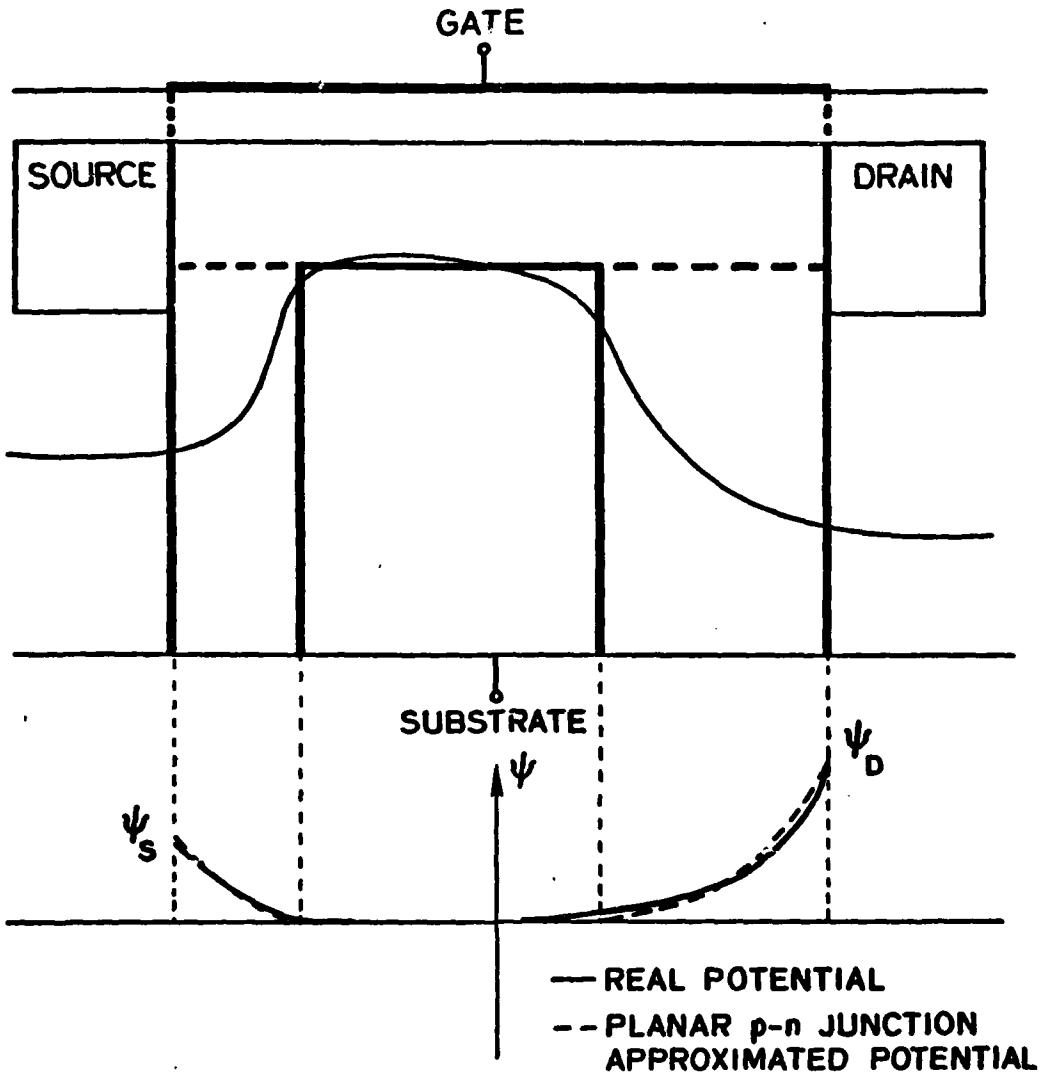



Figure 4.4.22 One dimensional planar p-n junction approximation of the boundary potential between source-bulk and drain-bulk.

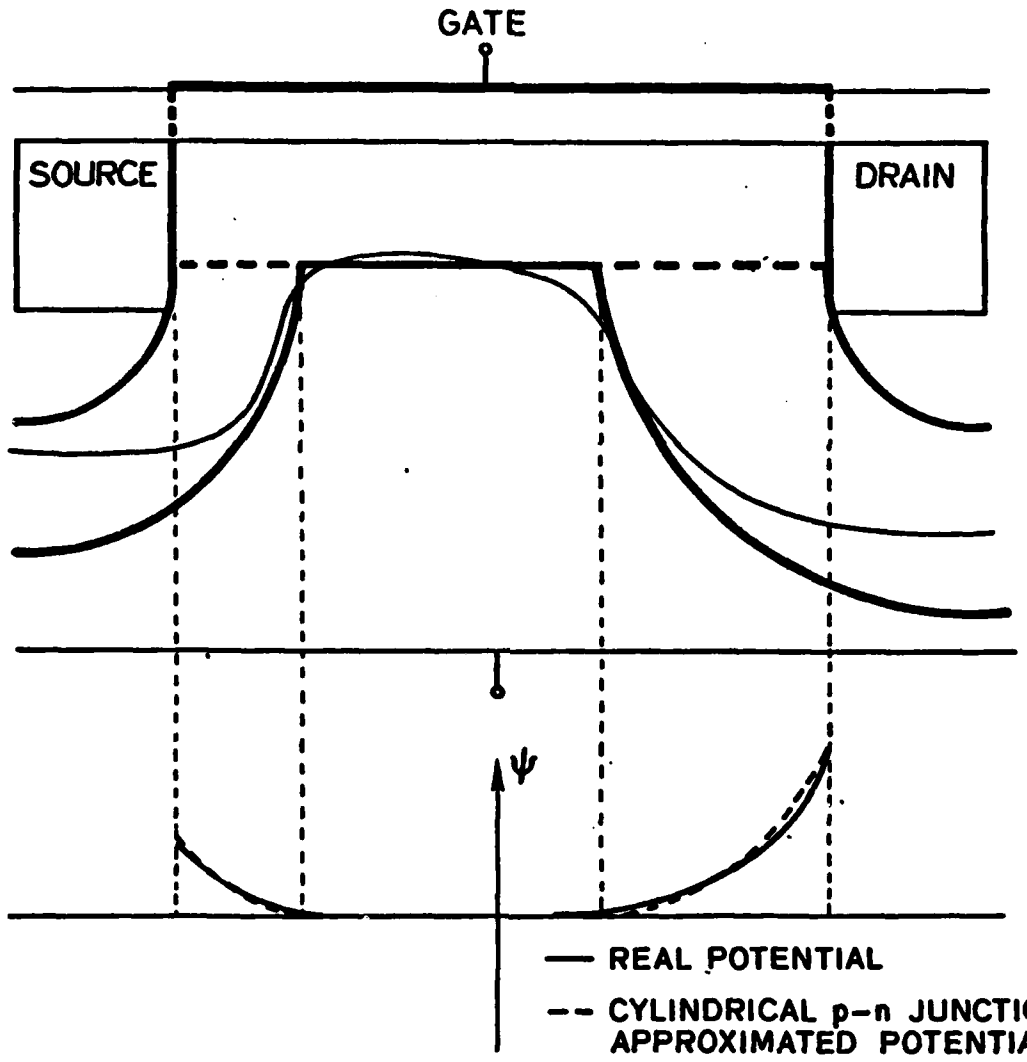



Figure 4.4.23 One dimensional cylindrical p-n junction approximation of the boundary potential between source-bulk, and drain-bulk.

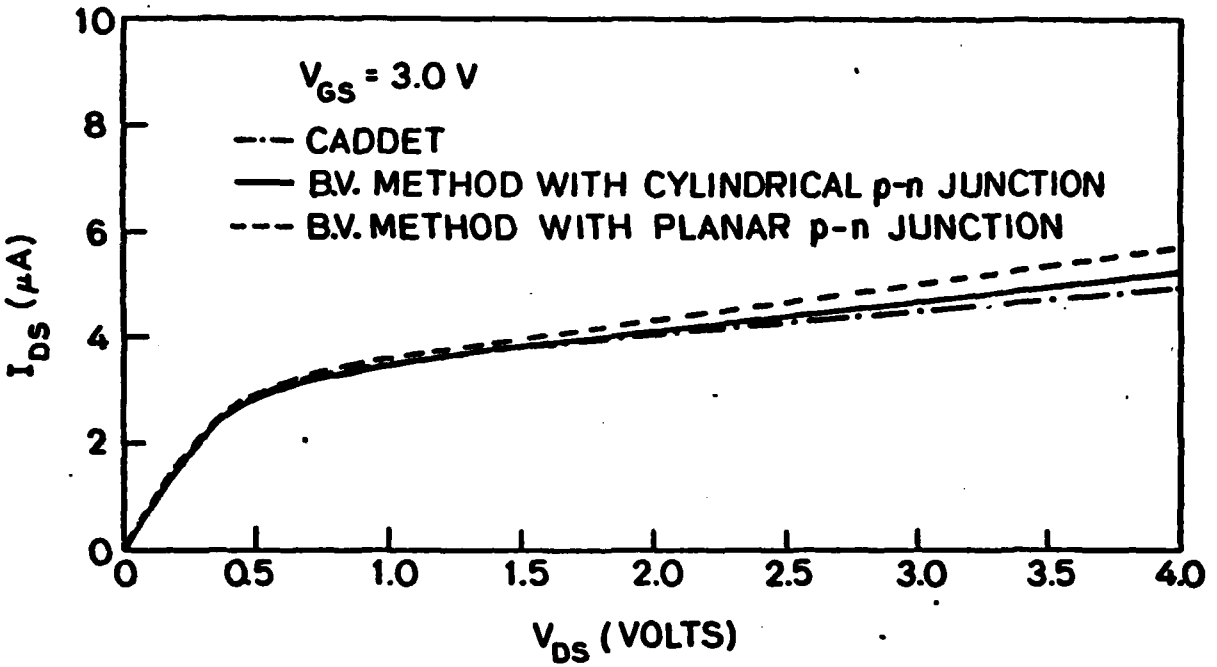



Figure 4.4.24 Comparison of the simulation results of the one-dimensional planar and cylindrical p-n junction approximation with those of CADDET.

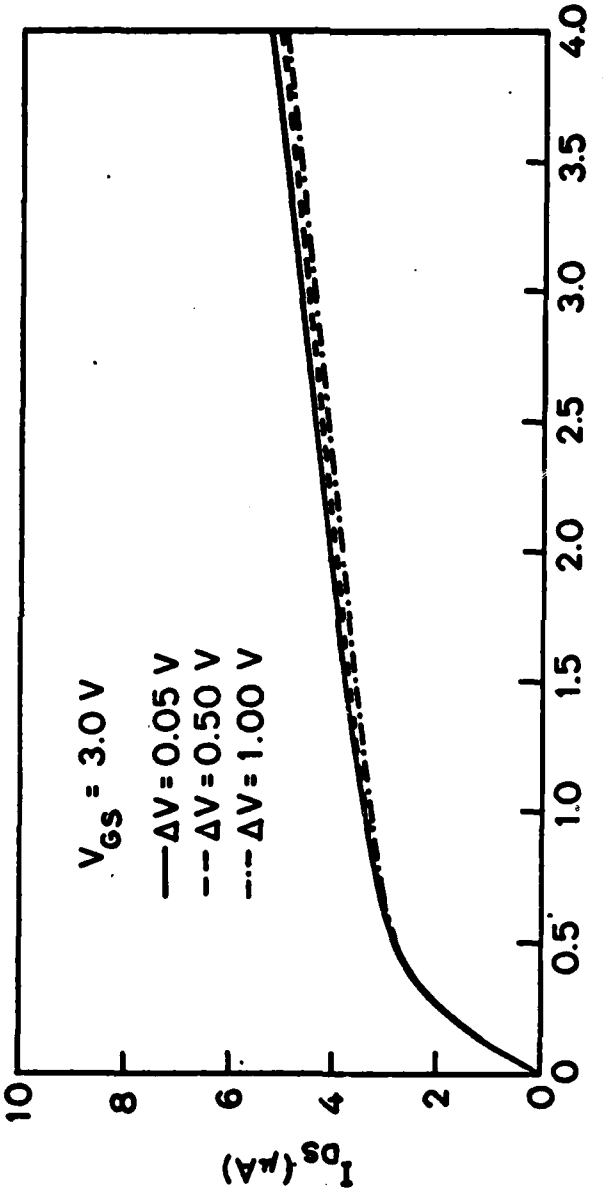



Figure 4.4.25 Effects of the bias voltages step on the drain current.

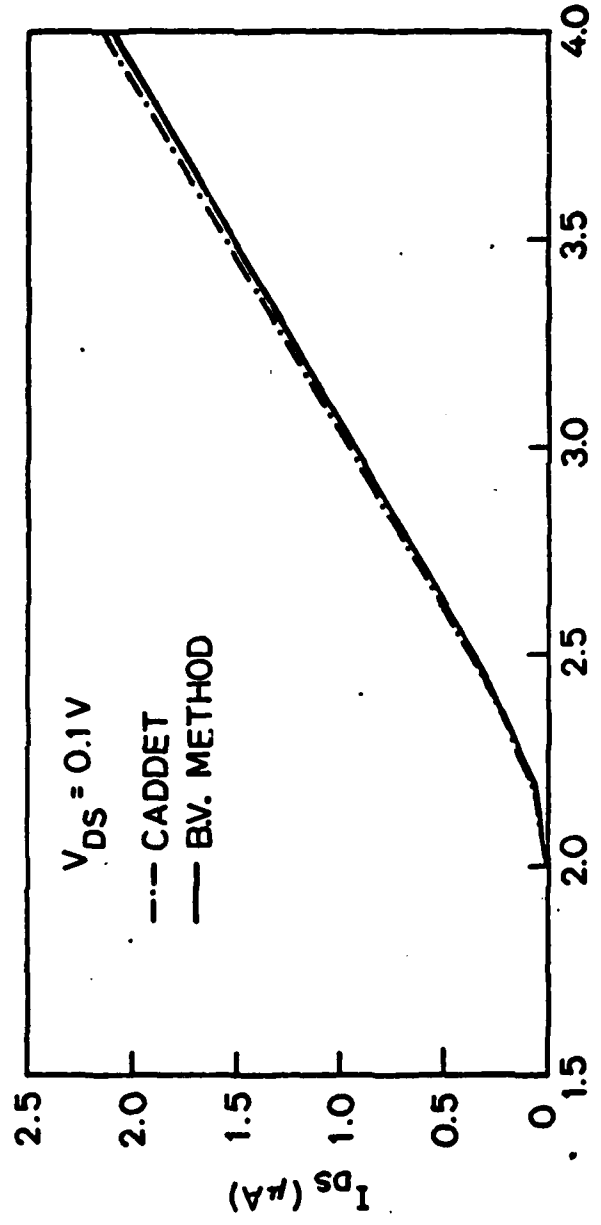



Figure 4.4.26 Comparison of the B.V. method with CADDET in $I_{DS}$ vs. $V_{GS}$ characteristics in the linear region.

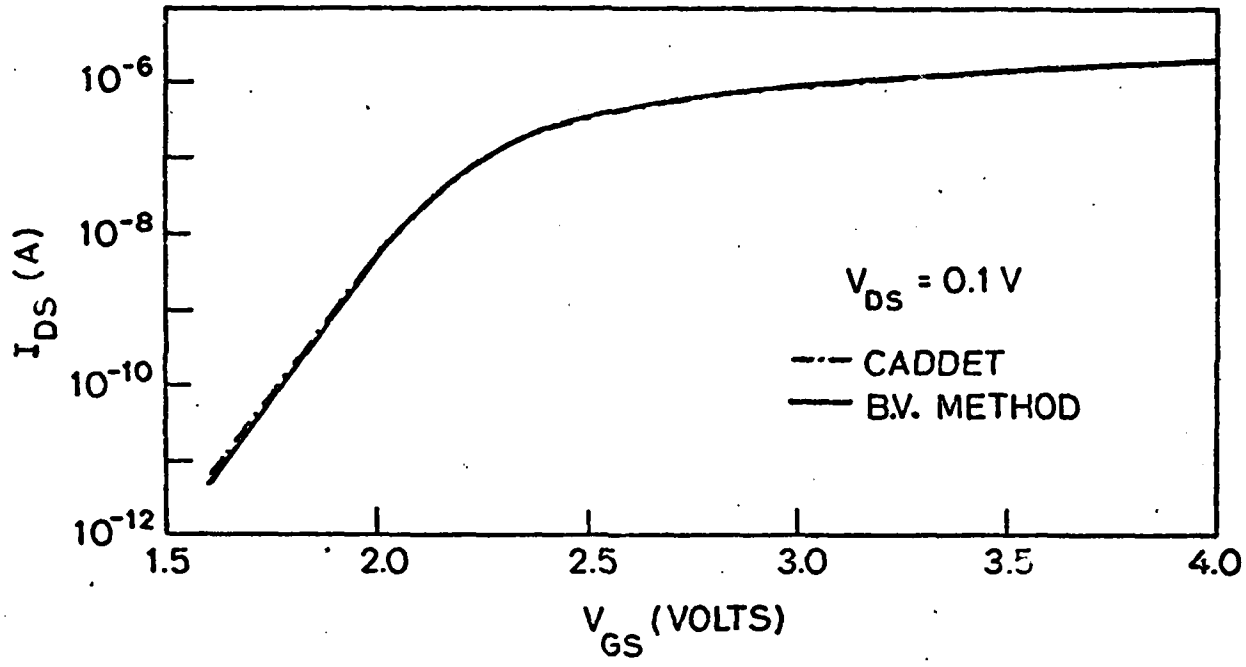



Figure 4.4.27 Comparison of the B.V. method with CADDET on $I_{DS}$ vs. $V_{GS}$ characteristics in the subthreshold region.

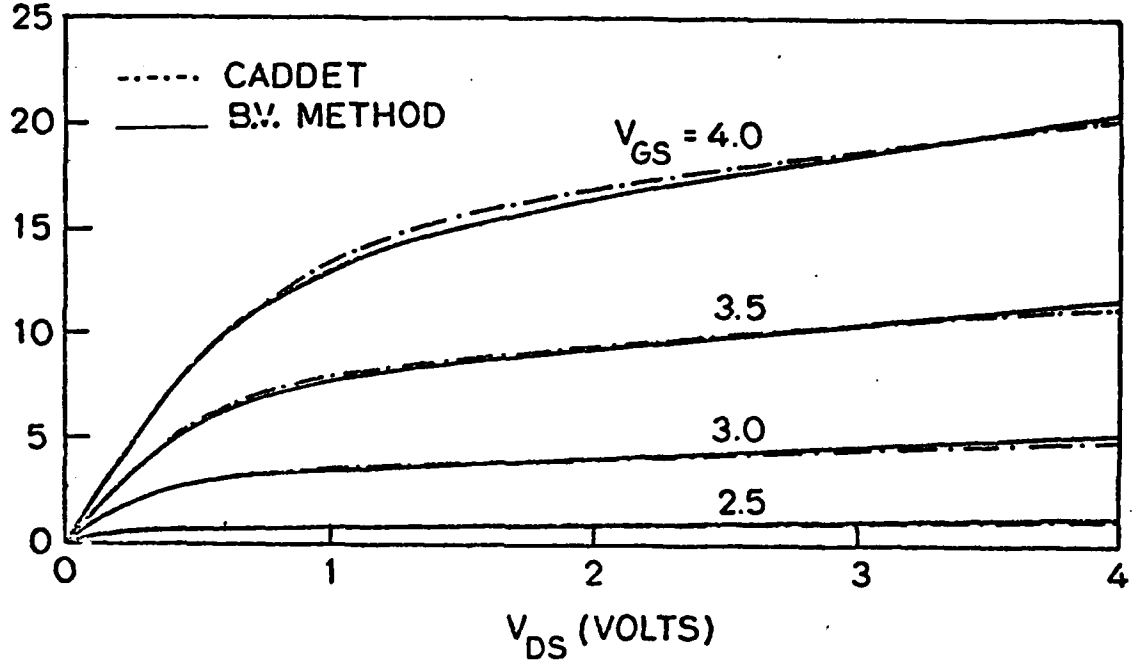



Figure 4.4.28 Comparison of the B.V. method with CADDET on $I_{DS}$ vs. $V_{DS}$ charactetistics in the linear and saturation region.

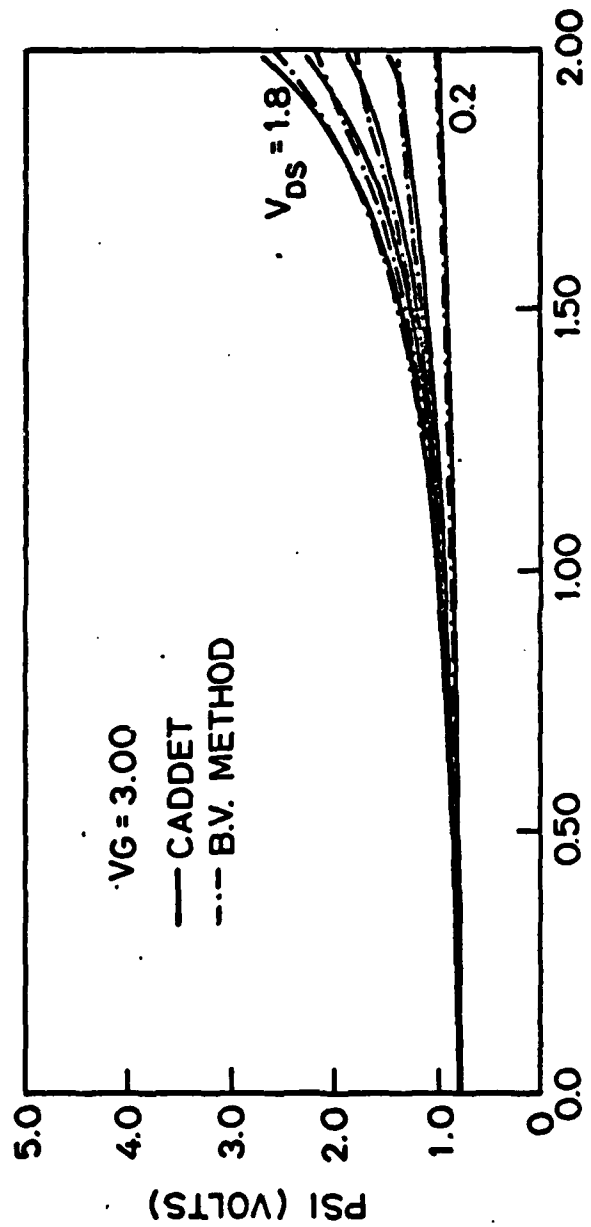



Figure 4.4.29 Comparison of the B.V method with CADDET on the potential distribution along the channel with the various drain voltages.

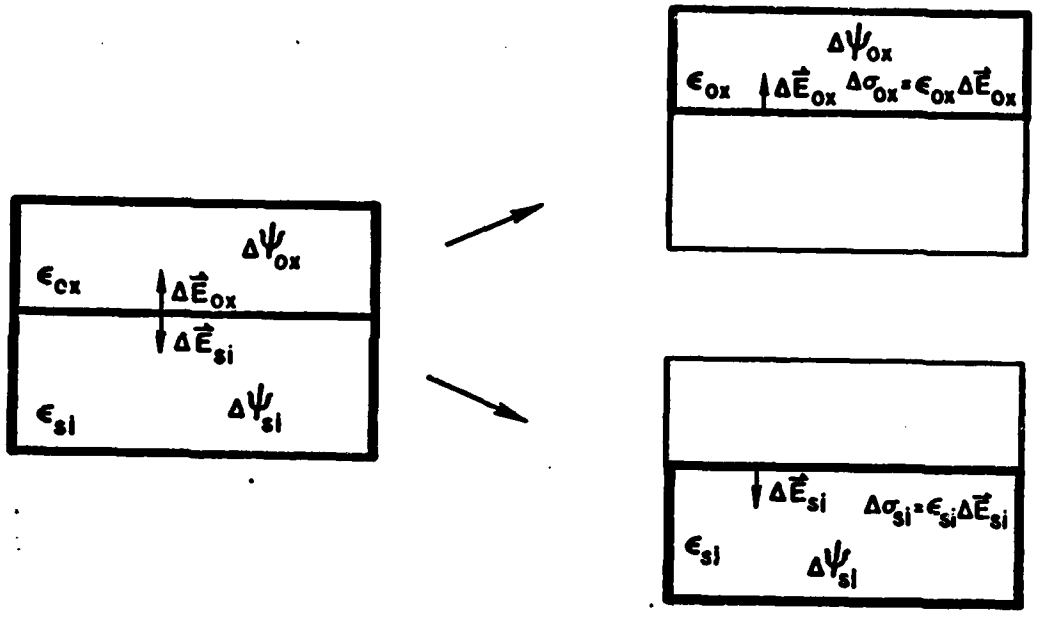



Figure B-1 Separation of the inhomogeneous region into the equivalent two homogeneous region.

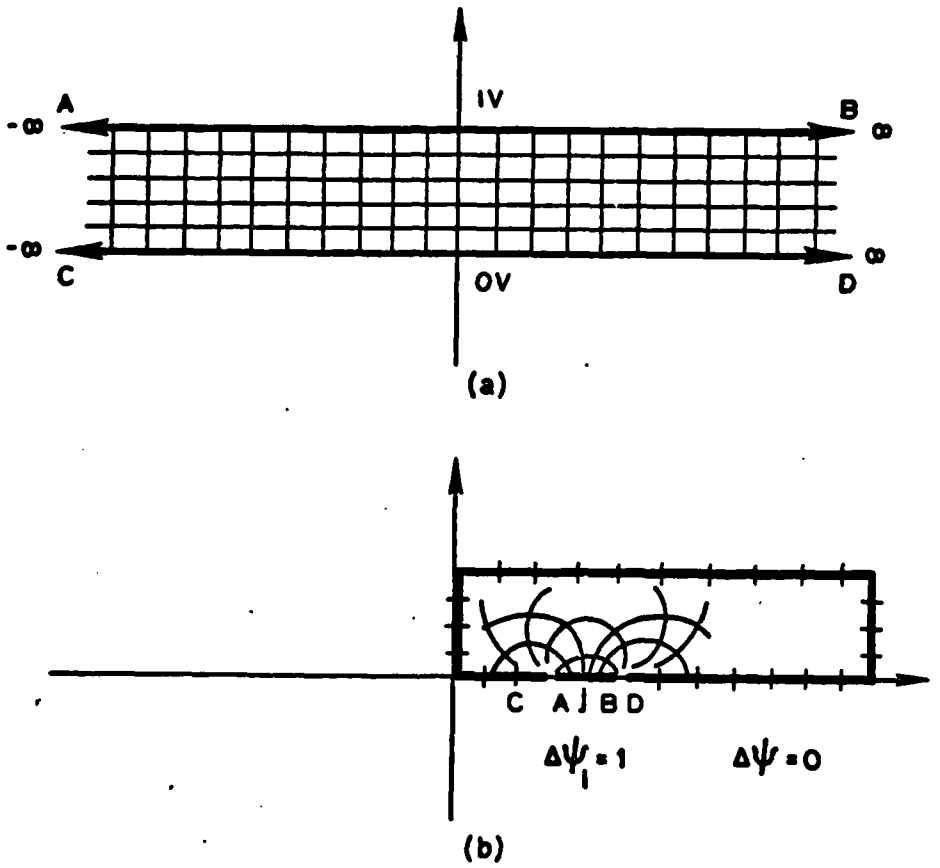



Figure B-2 The evaluation of the capacitance matrix using the conformal mapping method.

TABLE 4.4.4

CPU TIME COMPARISON*BETWEEN CADDET AND THE BOUNDARY VALUE PROBLEM (BV) METHOD

| $V_{GS}$ | PROGRAM | ITERATION** | CPU/BIAS*** | $I_{DS}$ |
|---|---|---|---|---|
| 0.0 V | CADDET | 12 | 4.3 sec | 0.0 A |
| | BV-METHOD | 5 | 1.3 sec | 0.0 A |
| 2.0 V | CADDET | 25 | 8.5 sec | 1.51E-8 A |
| | BV-METHOD | 5 | 1.3 sec | 1.69E-8 A |
| 3.0 V | CADDET | 26 | 10.8 sec | 4.95E-6 A |
| | BV-METHOD | 5 | 1.3 sec | 5.25E-6 A |
| 4.0 V | CADDET | 47 | 18.8 sec | 2.03E-5 A |
| | BV-METHOD | 4 | 1.1 sec | 2.13E-5 A |
| 5.0 V | CADDET | 72 | 30.4 sec | 4.50E-5 A |
| | BV-METHOD | 4 | 1.1 sec | 4.58E-5 A |
| 6.0 V | CADDET | 96 | 46.9 sec | 7.77E-5 A |
| | BV-METHOD | 4 | 1.1 sec | 7.80E-5 A |
| 7.0 V | CADDET | 100 | 47.9 sec | 1.16E-4 A |
| | BV-METHOD | 4 | 1.1 sec | 1.16E-4 A |

* $V_{DS}$ = 4 V CADDET Node = 1840 ΔV tolerance = 1 mV
B.V. method node = 46 ΔV tolerance = 1 μV

** CADDET iteration is the one between the Poisson's equation and N continuity equation. The iteration in the new method is Newton's iteration.

*** CPU time/bias point is measured at IBM 370-168.

TABLE 4.4.5

| | |
|---|---|
| $L_{eff}$ | : 2 μm |
| W | : 1 μm |
| $T_{ox}$ | : 0.1 μm |
| $V_{fb}$ | : 0 V |
| Electron life time | : 1 μsec |
| Electron mobility | : 684 cm**2/V sec |
| Substrate doping density | : 1E16 /cm**3 |
| Drain doping density | : 1E20 /cm**3 |
| Source doping density | : 1E20 /cm**3 |

## 4.5 TWO DIMENSIONAL GRID GENERATION FOR SEMICONDUCTOR DEVICE SIMULATION

C. H. Price

### 4.5.1 Introduction

The execution time and storage requirements of two dimensional semiconductor device simulation programs are directly dependent on the number of grid points (nodes) in the discretization grid. The number of equations to be solved is generally linearly related to the number of nodes and the number of arithmetic operations (principally multiplier) required for the solution is proportional to $n^{\alpha}$ where n is the number of nodes and $\alpha$ is somewhere between 1.5 and 2. Reducing the number of nodes in a simulation is, therefore, a matter of great importance.

Reduction of the number of nodes and the ease of matching non-rectangular structures are two of the principal factors in popularity of the finite element method for solving the equations of semiconductor device simulation. Some of these advantages are lost, however, if no simple means for grid generation is provided to the user for optimizing the grid to match the device topology. Most often, the user has only two choices, he may tediously locate every grid point (or some large subset thereof) in the x, y coordinate plane or else he may specify a much smaller subset of the grid points (such as boundary points only) and allow the computer to interpolate the remainder of the grid from the specified points. Unless a highly "intelligent" grid placement program is used, the computer generated grid will be sub-optimal for most problems of interest.

If limited to a specific device the degree of program "intelligence" needed is reduced and nearly optimal grid placement may be obtained. Although many of the interesting problems involve novel modifications to the standard device which would render the grid sub-optimal, it is a logical first step to initially limit the scope of the investigation into grid generation to a specific device. We have chosen the short channel MOS transistor as a representative problem of interest.

This work is just beginning and few results have been obtained yet since the initial efforts have been focused on developing an efficient (in terms of execution time and storage requirements) simulation program which uses the finite element grid [4.1 ]. Thus, the emphasis in this report is on the description of the grid generation problem and the directions proposed in pursuit of the solution.

4.5.2 Grid Density Criteria

It can be shown [4.30] that since abrupt changes in mobile carrier density are smeared out with a characteristic length of $L_d$ (the Debye length), accurate discretization of the continuous problem must have grid spacings of less than $L_D$ in such regions. The Debye length is given by:

$$L_D = \left[\frac{KkT}{4\pi q^2 n_i}\right]^{1/2} \qquad (4.5.56)$$

where kT/q is the thermal voltage, $\varepsilon$ is the permittivity, q is the electron charge and n is the mobile carrier concentration. Typical values for these parameters yield an $L_D$ of .002 microns for $n=10^{19}cm^{-3}$. Thus, even when modeling one micron dimension devices, the required grid spacing is excessive in terms of number of grid points needed. The question, then, is what is lost by exceeding the "required" grid

spacing maximum? Furthermore, what is the required grid spacing in other areas of the device away from abrupt mobile charge variations such as the depletion and neutral regions?

Another criteria is the requirement that the potential be accurately represented by linear interpolation between grid points. Thus, curvature of the potential requires a greater density of grid points which, as a consequence of Poisson's equation, places the densest grid where the net charge is greatest, i.e.

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\psi = -\rho/\varepsilon \qquad (4.5.57)$$

where $\psi$ is the electrical potential, $\rho$ is the net charge density and $\varepsilon$ is the permittivity of the region. Although this criterion places a linear relation between grid density and net charge density, such relations are seldom, if ever, followed. Ignoring neutral regions, net charge often varies over more than five orders of magnitude whereas grid density seldom varies over more than two orders of magnitude.

Another attribute which the grid must possess is flexibility over a range of bias conditions without need for modification. This is not an absolute requirement as adaptive solution methods do exist; however, the overhead involved in recomputing the geometrical connectivity information of the grid (and in redetermining the matrix decomposition steps for direct solution methods) is considered excessive for device simulation. Ideally, each region of the grid should have a grid point density equal to the densest grid needed for that region under any of the range of bias conditions. Practically, however, it may be preferable to utilize different grids for different families of bias conditions. For example,

the dense grid needed in the channel region of an MOS transistor above threshold would be largely wasted in the study of sub-threshold effects.

The principal techniques used to determine the effects of grid will be global and local grid density doubling combined with Richardson extrapolation. The grid density will be increased until further increases have no significant effect on the solution, then density will be selectively decreased in various regions of the model device to determine the minimum grid density necessary for each region. If possible, an adaptive technique for finite element problems developed by Bank [4.31] at the University of Texas at Austin may also be used.

### 4.5.3 Types of Grids

For the purpose of this report, grids are classified into two groups -- finite difference and finite element types. The distinction is not as great as the naming convention might imply, however, since it is possible to implement finite difference analysis on a non-rectangular "finite element" grid. Examples of several members of each type of grid are shown in Figure 4.5.30. Grids a,b,c and d in Figure 4.5.30 are finite difference grids and grids e and f are finite element grids -- the dominant difference is the non-rectangular nature of the latter.

The finite difference equations for the grid becomes increasingly more complicated progressing from Figure 4.5.30a to Figure 4.5.30d with an associated increase in overhead for execution time and storage. However the total number of grid points decreases over this range. Allowing the boundary surfaces to be non-rectangular further complicates the equations. The finite element grids e and f are seen to much more easily conform to the shape of internal and external boundaries but it is equally clear that specifying the grid is much more difficult.

The efficiency of each of the grids of Fig. 4.5.30 in discretizing a model MOS transistor is shown in Fig. 4.5.31. In this highly simplified example, it is assumed that the vertical grid density required in box labeled "v" and the horizontal grid density in the boxes labeled "h" is $\alpha$ and the density everywhere else is $\beta$ either horizontally or vertically. It is further assumed that $\beta = 1$ and $\alpha = 10*\beta = 10$, an arbitrary choice of grid density ratios. Using these values, the relative number of grid points needed for each grid point to satisfactorily cover the device can be calculated and used as a figure of merit for each grid. These results clearly show the desirability of the finite element grids; however, the most common grid, even for finite element problems, is grid type c because only the boundary node locations have to be specified by the user. This grid is seen to be only moderately efficient in use of grid.

A finite difference grid which is analogous to type f could be made by starting with a coarse regular rectangular grid and locally subdividing the rectangles in regions requiring high grid point density. It is not clear, however, how one should handle the differential equation discretization at nodes which do not have four neighbors. A special case of this type of grid appears as grid type d in which the added restrictions have been imposed that all horizontal lines must span the device and that along any horizontal line: the number of nodes must be a power of two, the node spacing must be constant and the number of nodes on adjacent horizontal lines may not differ by more than two. This grid allows use of Fourier transforms in the horizontal direction which can be used in a method devised by Hockney [4.32] for the solution of Poisson's equation. The benefits of this technique may outweigh the disadvantages posed by the excess grid required.

Primary emphasis for the grid generation studies discussed here will be on grid types e and f.

### 4.5.4 Generation Methods for Finite Element Grids

There are three methods for grid generation under consideration: subdivision of a coarse grid based on desired grid point density, relaxation techniques on a regular rectangular grid, and a refinement which maps and follows equipotential and electric field lines.

The subdivision scheme begins with a very coarse user-specified triangular grid which serves principally to fix the device shape. Based on the desired grid density, the triangular elements are subdivided recursively until the desired grid density is achieved in each region of the device. Two kinds of triangular element subdivision are allowed as shown in Figure 4.5.32. This scheme is taken directly from the adaptive method of Bank in [4.31].

One problem with this method is that it is nearly impossible to prevent the generation of obtuse triangles which are undesirable in finite element methods. Although there are at least two schemes for accomodating obtuse triangles, neither is very appealing in representing the physics of the device. The only alternative is to develop an algorithm for seeking out obtuse triangles and reconnecting the offending nodes into acute triangles. Whether such an algorithm is feasible will be determined.

The method involving relaxation of a rectangular grid also appears attractive but suffers from the same obtuse triangle problems as the previous method. A triangular grid is generated from the quadrilateral grid by connecting diagonal nodes. The relaxation method consists of moving each node by a weighted average of the vectors pointing to its

four neighbors. The weighting is the object of study but it is expected to be a function of the relative locations of the nodes and the grid density criteria for the region. Variations of this method are commonly used in structural finite element problems for smoothing the grid [4.33]. It is hoped that non-uniform weighting can similarly un-smooth the grid for our purposes. The grid of Figure 4.5.30e was generated using this procedure.

The third grid generation scheme involves mapping the equipotential and electric field lines within the device. Placing nodes at the intersections of these lines results in elements which are nearly rectangular due to the orthogonality of the equipotential and electric field lines. Connecting one diagonal node pair within the element, divides the nearly rectangular element into two triangular elements which are nearly right triangles. This method has been applied to the modeling of BJT's by Chan [4.34 ].

### 4.5.5 Source of Grid Density Values

Once the proper grid density criteria is determined, this criteria must be applied to each region of the device to generate a grid density value for that region. Because this criteria will probably include potentials and charge concentrations - values which are not known beforehand - a rough estimate of the solution is required in order to generate the grid. Various solutions to this dilemna are proposed:

1) Solve the problem on a relatively coarse regular grid.
2) Fit the impurity profile data to analytic functions (e.g. Gaussian, error function, etc.) and obtain a "textbook" solution. Use the potentials and analytic impurity profiles for grid generation.

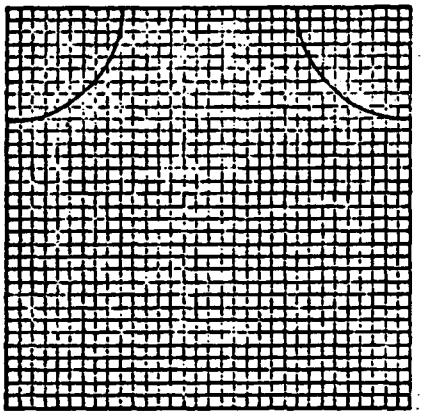
3) Same as 2) except apply the depletion edge and inversion layer locations obtained to the actual impunity profile to determine the net charge density.

4) Use simple one-dimensional abrupt or graded junction approximations and the complete depletion approximation to calculate the location of depletion edge and inversion layers, then proceed as in 3).

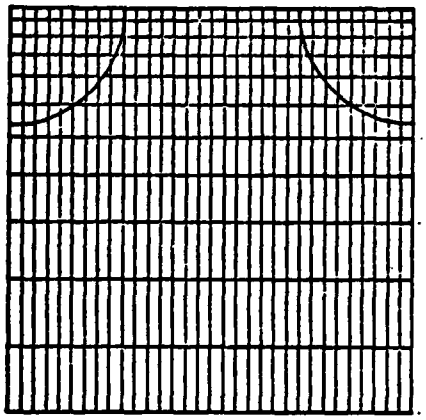
These methods should (to varying degrees) provide results which are sufficiently good for grid generation. The advantages of an adaptive method where grid and soluti.. ; are improved iteratively is apparent.

4.5.6 Conclusion

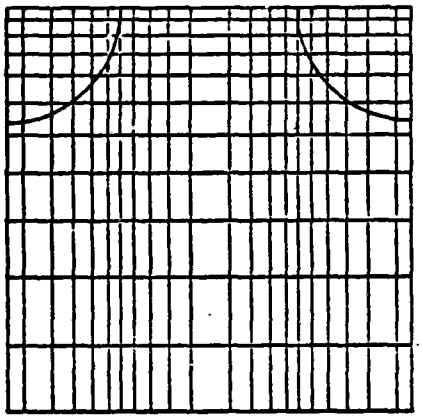
The need for improved methods of generating finite element grids in order to take advantage of the grid efficiency of the finite element method has been presented. Work has just begun in this area and has started with the development of a suitable device simulation program for use in comparing the various grids. One of the principal objectives will be the determination of a grid density criteria for optimal placement of grid points. The second objective will be the investigation of various methods for generating grid according to the criteria determined. The three principal methods to be compared are: subdivision of elements, grid relaxation, and following equipotential and electric field lines.



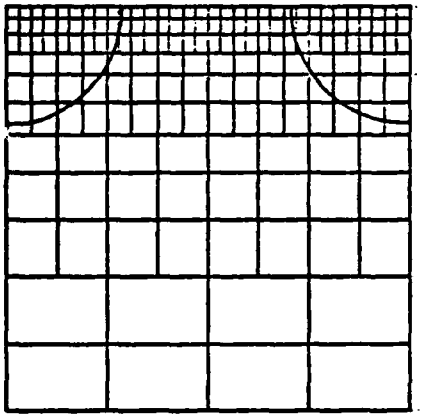
(a) Regular Rectangular



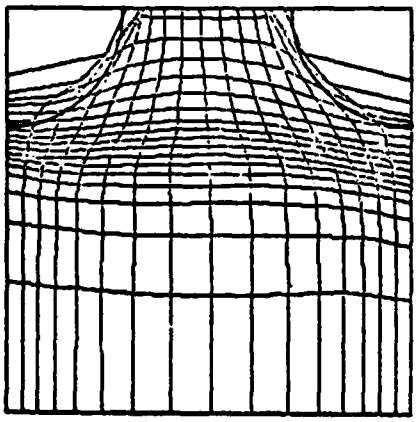
(a) Semi-regular Rectangular



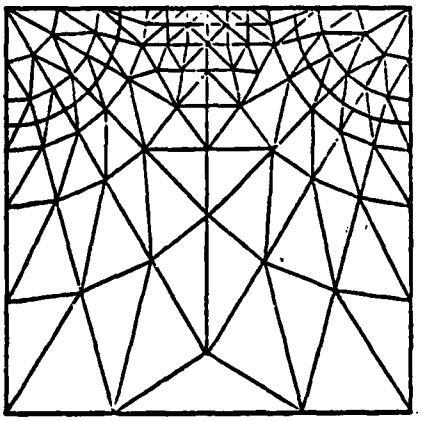
(c) Rectangular



(d) Truncated Semi-regular Rectangular

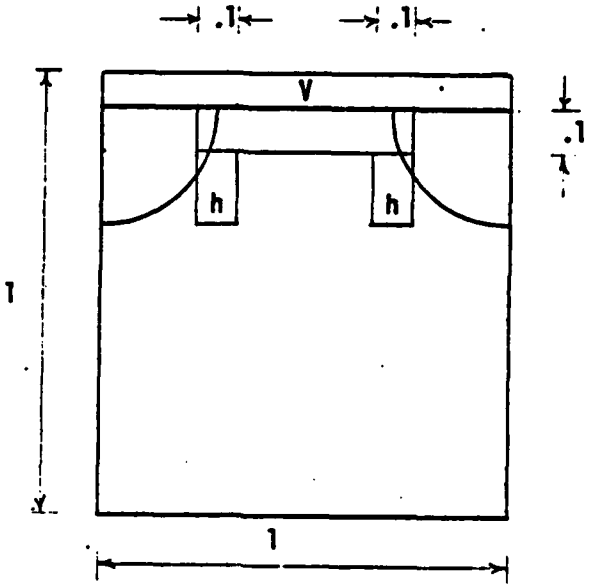


(e) Relaxed Rectangular



(f) Triangular

Figure 4.5.30 Grid types




| Grid | Relative Number of Nodes Required |
|---|---|
| a) $\alpha^2$ = | 100 (all fine) |
| b) $.1\ \alpha^2 + .9\ \alpha\ \beta$ = | 19 |
| c) $.02\ \alpha^2 + .26\ \alpha\beta\ .72\ \beta^2$ = | 5.3 |
| d) $.3\ \alpha\beta + .7\ \beta^2$ = | 3.7 |
| e) $.1\ \alpha\beta + .9\ \beta^2$ = | 1.9 |
| f) $.04\ \alpha\beta + .91\beta^2$ = | 1.8 |
| | 1 (all coarse |

Figure 4.5.31 Relative efficiencies of the grids of Figure 4.5.30.

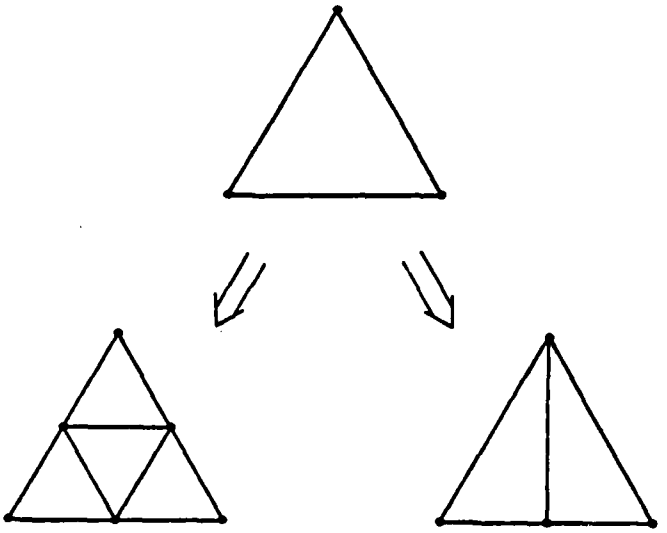


Figure 4.5.32 Sub-division of triangular elements

## 4.6 TEST STRUCTURES FOR TWO-DIMENSIONAL PROFILE MEASUREMENTS

H. G. Lee

Earlier work with a CMOS test-chip [4.35] successfully demonstrated a method to calculate surface impurity profiles of laterally diffused regions using p-well and channel stop implants. Subsequently, two questions have been posed which must be answered before this technique can be established as a useful design aid for the realization of VLSI devices.

1) The present measurement method involves simplifying assumptions concerning the transport behavior of inverted MOS channel region. What will be the ultimate application limits of this technique?
2) How easily can this test structure approach be adapted for a process other than a CMOS process?

This section will cover results of work that has evolved from efforts to answer these questions. The two subsections below correspond to each of the points above. In the first subsection, several theoretical calculations including a two-dimensional device simulation program (TANDEM/Poisson) is used to examine the accuracy of the algorithm and the sensitivity with the parameters involved. In the next subsection, fabrication and measurement results on a DMOS test-chip will be discussed.

### 4.6.1 Accuracy of the Algorithm to Calculate Surface Impurity Profiles of Laterally Diffused Regions

The measurement method, which will be briefly reviewed in the following, converts the measured $I_D$ vs. $V_G$ relationship for an MOS transistor

into the local threshold voltage along the surface, and hence the surface impurity profile. It is important to understand the validity of the method. Since no other measurement technique is available with which to compare, the following approach has been taken.

An MOS structure with known impurity profiles, as shown in Figure 4.6.34 is used first. When a gate potential is applied the inversion charge in the channel region can be calculated, and hence the channel current, as a function of the gate and substrate potential is determined. Using the calculated $I_D$ vs. $V_G$ data at two values of substrate bias, the surface impurity profile is reconstructed using the method. The profile should be identical to the original profile provided the calculation method is valid. This test is important because it not only evaluates the accuracy of the method, but also facilitates the study of the sensitivity of the result on the assumptions and parameters involved in the calculation. One of the assumptions made in the algorithm is that electron mobility in the channel does not depend on the channel doping. We will estimate the error this assumption can introduce in the final profile. The result indeed shows that the assumption results in a nonnegligible error. An improved computational method will be suggested to handle this problem.

1. Algorithm for Profile Calculation Based on Measurement

The surface impurity profile in the width direction of n-channel transistors with non-uniform channel regions can be calculated from the device transconductance. The transistor can be considered as a parallel combination of many devices having threshold voltage $V_T$ variations in the x direction only, as shown in Figure 4.6.34. At a

small drain voltage $V_T$, elements of width dx whose $V_T = V_T(x)$ contribute

$$dI_D = 2 \frac{dx}{L} \mu(V_G, V_B, N_A) C_o [V_G - V_T(x)] V_D \qquad (4.6.58)$$

where

$L$ = length of the channel

$C_o$ = oxide capacitance per unit area

$V_G$ = gate potential

$V_D$ = drain potential

$N_A$ = surface impurity concentration

$\mu(V_G, V_B, N_A)$ = electron mobility in the channel

and the factor of two results from equal contributions from the two sides of the device.

The analysis is simplified using an approximate analytic form for mobility. Since lateral mobility variations of the test devices are due to the nonuniform channel doping, and this variation is smaller than variations due to $V_G$ or $V_B$, we assume that $\mu$ is not a function of doping (and therefore independent of x). Then the total current at a given gate voltage $V_G$ will be

$$I_D = \frac{2\mu(V_G, V_B)}{L} C_o V_D \int_{x(V_G)}^{W} \{V_G - V_T(x)\} \, dx \qquad (4.6.59)$$

where $V_G = V_T$ at $X(V_G)$ and $X(V_G)$ is the maximum lateral extent of inversion at $V_G$. Differentiating (4.6.59) with respect to $V_G$, the transconductance is

$$g_m = \left.\frac{\partial I_D}{\partial V_G}\right|_{V_D=\text{const}} = \frac{I_D}{\mu(V_G,V_B)} \cdot \frac{\partial \mu(V_G,V_B)}{\partial V_G}$$

$$+ \frac{2\mu(V_G,V_B)}{L} C_o V_D \{W-X(V_G)\} \qquad (4.6.60)$$

Measured mobilities are fitted to an empirical equation

$$\mu(V_G,V_B) = \frac{\mu_o(V_B)}{1+\theta(V_B)V_G} \qquad (4.6.61)$$

at each substrate bias. The transconductance $g_m$ is calculated from the measured $I_D$ vs. $V_G$ data, and we obtain $X(V_G)$ from (4.6.60) and (4.6.61). Given $X(V_G)$, we know $V_T(x)$ since $V_T=V_G$ at $X(V_G)$.

Using measured threshold voltage at two substrate biases, we obtain

$$V_{T1}(x) - V_{T2}(x) = \frac{1}{C_o}\sqrt{2q\epsilon_s N_A(x)} \cdot \{\sqrt{V_{SB1} + 2\phi_F(x)}$$

$$- \sqrt{V_{SB2}+2\phi_F(x)}\ \} \qquad (4.6.62)$$

where

$$\phi_F(x) = \frac{kT}{q} \ell n \ \{N_A(x)/n_i\} \qquad (4.6.63)$$

The lateral dependence of surface states or oxide charge on position has been eliminated. $N_A(x)$ is obtained from this equation by iteratively solving for $\phi_F(x)$ from $N_A(x)$.

Two assumptions made in the method are apparent.

1) Mobility dependence on channel doping is negligible.
2) All the current measured from the device come from surface regions which are strongly inverted. This is not exact, since a small

but finite portion of the measured current is the contribution of weakly inverted regions. Also two-dimensional effects due to nonuniform channel doping are assumed to be negligible.

## 2. $I_D$ vs. $V_G$ Characteristics Calculation for a Given MOS Structure

The structure used to calculate the $I_D$ vs. $V_G$ characteristics is shown in Figure 4.6.35. The surface impurity concentration is given by

$$N(x) = \frac{N_{peak} - N_{sub}}{2 \text{ erf } \frac{W}{4\sqrt{Dt}}} \left[ \text{erf}\left(\frac{x + \frac{W}{2}}{2\sqrt{Dt}}\right) - \text{erf}\left(\frac{x - \frac{W}{2}}{2\sqrt{Dt}}\right) \right] + N_{sub} \qquad (4.6.64)$$

and the vertical profile of the p-diffusion is a Gaussian with $\exp(-y^2/4Dt)$ dependence. We calculate the inversion charge $Q_{in}(x)$ for the given gate potential $V_G$. For a very small drain bias, the gradual channel approximation is valid. Therefore,

$$I_D = \int_{X(V_G)}^{W} \mu \, Q_{inv}(x) \cdot \frac{V_D}{L} \, dx \qquad (4.6.65)$$

The current voltage relationship is calculated using three methods as shown below, with the order of increasing complexity.

(Case A) Strong inversion assumption with mobility dependence on $V_G$ and $V_B$

$$I_D = \int_{X(V_G)}^{W} \mu(V_G, V_B) \, C_o \, \{V_G - V_T(x)\} \cdot \frac{V_D}{L} \, dx \qquad (4.6.66)$$

where $X_{max}$ is the point where the gate voltage $V_G$ equals the threshold voltage $V_T(X_{max})$. Mobility is assumed to have only the $V_G$ and $V_B$

dependence as in (4.4.61)

The drain current calculated is shown in Figure 4.6.36. Since the assumptions made in calculating the current in this case are exactly the same as the ones for the lateral profile calculation method, we should expect to reconstruct the surface impurity profile without any error. Therefore any error will be due to the numerical technique used in the method e.g., round-off errors, etc. Using this idea, we actually obtained the criterion for the minimum number of grid points in discretizing the space for the current calculation, and the number of points in $V_G$ at which the drain current $I_D$ is calculated to reconstruct the lateral surface profile. It is found that dividing the structure into 100 sections for $I_D$ calculation, with more allocation of points where the profile changes rapidly, and 200 data points of 10 millivolt step for $0<V_G<2$ volts will give a satisfactory reconstruction of the lateral profile as shown in Figure 4.6.37(a).

(Case B) Strong Inversion Assumption with Mobility Dependence on $V_G$, $V_B$ and $N_A$

In the previous $I_D$ calculation, we assumed that the channel mobility does not vary with channel doping. We can be more exact by adding the dependence on channel doping with the following mobility model

$$\mu = \mu_1(N_A) \cdot \mu_2(V_G, V_B)$$

where

$$\mu_1(N_A) = \mu_{min} + \frac{(\mu_{max} - \mu_{min})}{1 + (N_A/N_{ref})^{\beta}} \qquad (4.6.67)$$

$$\mu_2(V_G, V_B) = \frac{1}{1 + \theta(V_B) \cdot V_G}$$

The equation for $\mu_1(N_A)$ is frequently used for the bulk mobility estimation [4. 36]. We adjust $\mu_{min}$ and $\mu_{max}$ so that it fits the measured channel mobility. Then

$$I_D = \int_{X(V_G)}^{W} \mu_1(N_A)\, \mu_2(V_G,V_B) C_o \{V_G - V_T(x)\} \cdot \frac{V_D}{L} \cdot dx \qquad (4.6.68)$$

The calculated current is shown in Figure 4.6.36. Even though $I_D$ is calculated with the mobility dependence on channel doping taken into account, it is not possible to include this dependence in the existing method for lateral profile calculation. Therefore, the reconstructed lateral profile will now be different from the original profile. As shown in Figure 4.6.37(b), the reconstructed lateral profile has a steeper slope in the center region of the structure where the surface concentration changes from substrate concentration to peak concentration of the p-diffusion. However, the peak concentration value is still predicted accurately.

(Case C) Two-Dimensional Poisson Simulation

In cases A and B, simple current calculations were used to examine the cases where the assumption of strong inversion is valid. Using a numerical two-dimensional device simulation program, we can calculate the total current which, in addition to the mobility dependence of case B, takes i) weak inversion current; ii) two-dimensional effect; iii) nonuniform channel doping in the vertical direction, all into account. The inversion charge distribution as a function of depth near the surface is calculated and shown in Figure 4.6.38.

One disadvantage of using a numerical two-dimensional analysis is the substantial calculation time. As shown in case A, we need 200 $I_D$ vs. $V_G$ data points for each substrate bias, and it is not practical to calculate all these data points. However, we know that the $I_D$-$V_G$ characteristics of the lateral profile device are very well behaved. Therefore, we will evaluate the current value at a few bias points and use spline fit interpolation to obtain the other data points (Figure 4.6.36). In this study, 15 to 20 points are taken for each substrate bias, with relatively more points allocated in the bias range where the slope of $I_D$ vs. $V_G$ characteristics changes rapidly. The current data calculated using this method were used to reconstruct the lateral surface profile using the method and the result is shown in Figure 4.37(c). Since the spline interpolation is a polynomial fit which matches the given data points exactly, the data points obtained by interpolation will contain small but finite oscillations compared to an ideal smooth curve fit. Due to the data differentiation step in the method , the effect of this oscillation is amplified and is apparent in the reconstructed profile. However, the general feature of the lateral profile is quite similar to case B. Errors due to neglecting the weak inversion current are not observed because the use of equation (4.4.62) compensates this assumption to a certain degree by using the difference in threshold voltages instead of using the absolute values of threshold voltages. Also we do not see an appreciable difference due to any two-dimensional effect in the center region of the device, because the concentration gradient in this region is not steep enough to result in a significant two-dimensional effect.

From study of the above three cases, we conclude that accurate modeling of mobility dependence on channel doping is the single most important area where the present lateral profile calculation method needs additional refinements. It is found that by using an iterative approach, we can significantly improve the algorithm.

3 An Iterative Method to Improve the Lateral Profile Calculation

We use equation (4.6.69) which is rewritten here.

$$I_D = 2\,\mu_2(V_G,V_B) \int_{X(V_G)}^{W} \mu_1(N_A)\, C_o\, \{V_G - V_T(x)\}\, \frac{V_D}{L}\, dx \qquad (4.6.69)$$

where the factor of two has been discussed above concerning equation (4.6.58). Since $N_A$ is a function of x, $\mu_1(N_A)$ is also a function of x. $\mu_1(x)$ will be the channel mobility along the device width at zero bias. Then

$$I_D = 2\,\mu_2(V_G,V_B)\, C_o\, \frac{V_D}{L} \left[ V_G \int_{X(V_G)}^{W} \mu_1(x)dx - \int_{X(V_G)}^{W} \mu_1(x)\, V_T(x)dx \right] \qquad (4.6.70)$$

The transconductance $g_m$ is

$$g_m = \frac{I_D}{\mu_2(V_G,V_B)} \cdot \frac{\partial \mu_2(V_G,V_B)}{\partial V_G} + \frac{2\mu_2(V_G,V_B)}{L}\, C_o V_D \int_{X(V_G)}^{W} \mu_1(x)\, dx \qquad (4.6.71)$$

Equation (4.6.71) reduces to (4.6.60) when $\mu_1(x)$ is constant.

Since the lateral profile is unknown, we cannot directly calculate the integration to obtain $X(V_g)$. We first assume $\mu_1(x)$ to be a constant and obtain a lateral profile with the method used in the previous discussion. Then we obtain a second approximation for $\mu_1(x)$ from the calculated lateral profile using equation (4.6.68).

This second approximation for $\mu_1(x)$ is used in eq. (4.6.70) to obtain a new $X(V_G)$. It can be shown from Figure 4.6.39 that this process actually improves the result as the iterations are repeated. In Figure 4.6.39, curve A is the constant mobility used in the first approximation and curve B is what is obtained from the calculated lateral profile.. For a given $V_G$ and $g_m$ in equation (4.6.71), $X(V_G)$ obtained using curve B will be closer to the peak of p-diffusion (x=0) than $X(V_G)$ obtained using curve A. This process of iteratively improving the lateral profile will converge to the correct lateral profile with a few iterations.

### 4.6.2 Fabrication and Measurement of DMOS/Profile Test Chip

The double-diffused or DMOS transistors have been proposed as potential technologies for VLSI high speed digital applications. Successive p-type and n-type diffusions made through the same oxide opening results in the characteristic short diffused channel lengths, offering an improved performance over conventional NMOS at moderate masked channel lengths.

Due to the graded impurity profile of the channel region, the electrical characteristics of these devices are considerably different from NMOS transistors. Efforts to fit the measured DMOS characteristics to conventional NMOS transistor model equations were not satisfactory, requiring the use of an effective channel doping and threshold voltage. These parameters are typically obtained by adjusting the NMOS device parameters with empirically determined factors, which can vary with the change of fabrication procedures.

Earlier work involving DMOS modeling [4.37 ] has indicated that a more comprehensive short channel device model is needed which takes the actual impurity distributions into account. The simple two-

dimensional diffusion theory [4.38 ] has been utilized to model the device characteristics [4.39 ], but the actual profiles have never been measured, and the inherent two-dimensional effects of the DMOS structure remains poorly understood.

The lateral surface impurity profile measurement technique can be a powerful tool for the determination of the channel impurity distributions of the DMOS devices. The technique is capable of resolving concentration variations within one micron lateral dimension for impurity concentrations in the range of $10^{16}cm^{-3}$. Some physical/chemical techniques might have comparable resolution capabilities in the lateral dimension, but these techniques are confined only for very high impurity concentrations (typically above $10^{19}cm^{-3}$).

Test structures have been developed for the measurement of DMOS channel profiles. On the same test chip, test devices for two other related areas of research have been also included. These areas will not be further discussed here, except for the brief objectives of the studies as given below:

(1) Determination of the smallest device dimensions for which the lateral profile measurement technique will be useful;

(2) Application of similar test structures for the study of several other two-dimensional diffusion/oxidation effects. One example is the determination of impurity distributions near the locally oxidized areas.

Two Lateral Impurity Profile (LIP) measurement structures are designed with several modifications of device geometries from earlier designs in order to accomodate for different fabrication procedures. The first LIP device (name it as LIP-1) is very similar

to the original structure of the CMOS counterpart. Photoresist stripes define the area for p-type implantation and a drive-in nitrogen ambient will result in lateral as well as vertical diffusion of implanted atoms. A gate oxide is then grown followed by n+ source and drain diffusion. Aluminum gates are used for these devices. Due to the great similarity in design with the previous devices, this structure is virtually fail-safe. However, the laterally diffused regions do not exactly correspond to the DMOS channel region, since the p-implant for DMOS is self-aligned to the poly-silicon gate.

A second LIP device (LIP-2) is designed so that the laterally diffused profile experience similar processing conditions with DMOS channel profiles as much as possible. The cross-section of a basic LIP-2 device as well as that of an n-channel poly-silicon gate DMOS device is shown in Figure 4.6.40. To prevent stray conduction at the edges of the LIP-2 device, a channel stop implant is performed which is also self-aligned to the poly-silicon gate edge. Compared to LIP-1 device, LIP-2 has the advantage of having a laterally diffused region which follows as close as possible to the real DMOS channel profile. This advantage comes at the cost of an extra processing step (channel stop implantation). It is expected, however, that under proper fabrication conditions both LIP-1 and LIP-2 should provide similar lateral profile information. For both LIP devices, the actual layout utilizes the parallel connection of 32 identical elements in order to minimize the measurement fluctuations and edge effects. The layout of LIP-2 device is shown in Figure 4.6.41.

### 4.6.3 Fabrication

LIP devices have been fabricated with a modified n-channel poly-silicon gate process using boron ion-implanted p-region and phosphorus diffused source and drain. The p-region is implanted through an oxide of 1000 Å with an energy of 60 kev and dose of 4 to 8 x $10^{12}$/ $cm^2$ which is self-aligned to the poly-silicon gate. A drive-in cycle in a nitrogen ambient is made for 95 min at 1200 °C. Phosphorus source/ drain diffusion and boron channel stop implantation is followed. Since the p+ channel stop defines the width of the LIP-2 devices which should not perturb the laterally diffused p-region, the channel stop implantation is performed after all the major heat cycles. It is expected that the lateral encroachment of the channel stop implants during any subsequent heat cycles should be on the order of the lateral diffusion of n+ source diffusion for the DMOS devices, so that no useful information will be lost. Annealing of the channel stop implantation and diffusion of the source/drain is accomplished in a single low temperature oxidation step. The gate oxide of the LIP-1 device is also grown in this step.

### 4.6.4 Measurement

A desktop calculator measurement systems [4.40] has been used to obtain the electrical characteristics of the test devices. Surface profiles measured from LIP devices are shown in Figure 4.6.42. The apparently steeper measured profile of the LIP-1 device compared to the theoretical profile of equation (4.6.64) is due to the mobility effect described in the previous subsection. This measurement again confirms the validity of the two-dimensional diffusion model [4.35, 4.38]. Measured profile from LIP-2 device shown in the second figure agrees

closely to the theoretical model as well as the corresponding regions of the LIP-1 device. It is interesting to note that the channel profile for DMOS devices can be well approximated by an exponential function (which will be a straight line in Figure 4.6.42(b)). This approximation allows the derivation of simple but accurate analytic expressions for threshold voltage and drain current which for the first time incorporate nonuniform profile and two-dimensional field effects [4.41]. The analytic results compare favorably with two-dimensional simulations.

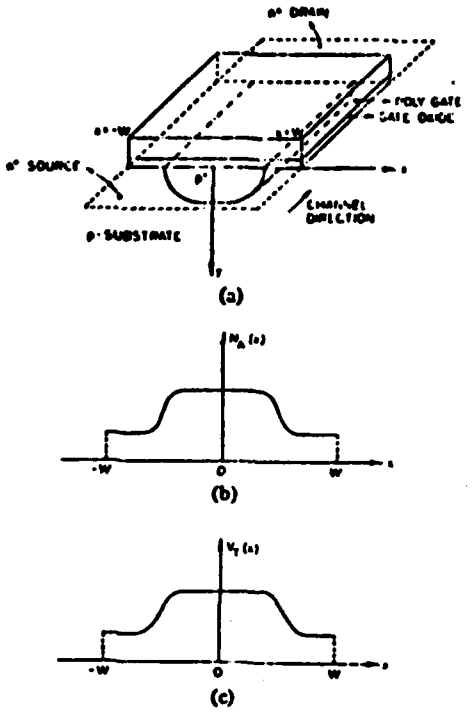



Fig.4.6.34. (a)n-channel MOSFET with nonuniform doping across the width of the channel(x-direction). (b)surface impurity concentration in x-direction. (c)Threshold $V_T$ in x-direction.

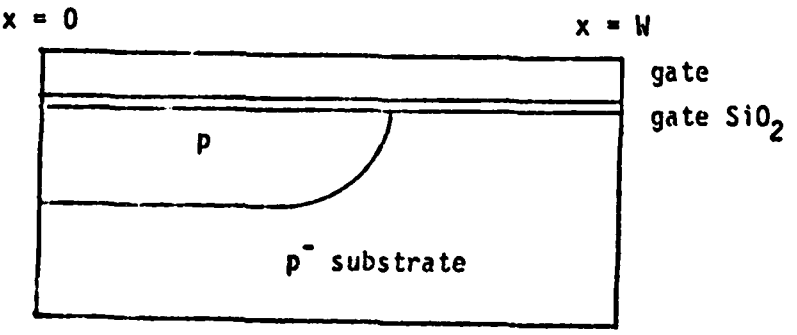



Fig. 4.6.35. MOS structure for current calculations for cases A, B and C.

$N_{sub} = 8 \times 10^{14}$ Cm-3

$t_{ox} = 670$ Å

$W = 20$ μm

$N_{peak} = 2 \times 10^{16}$ Cm-3

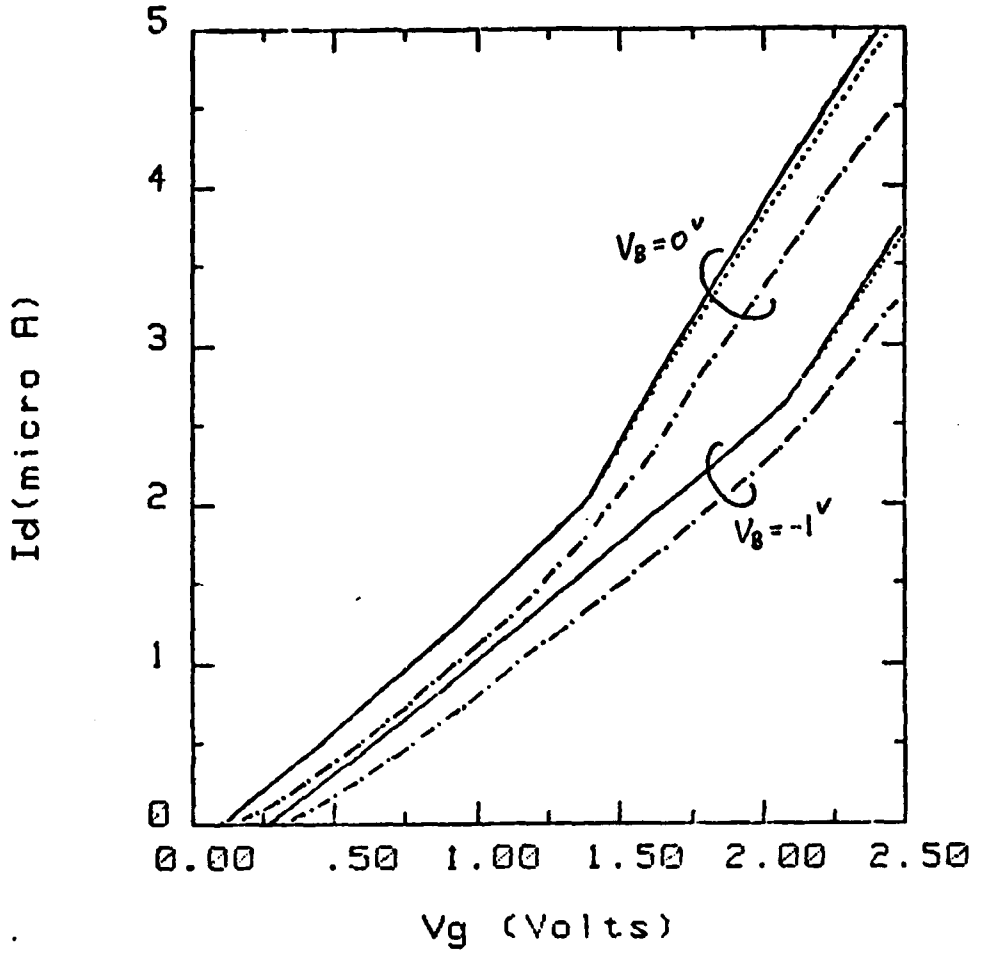
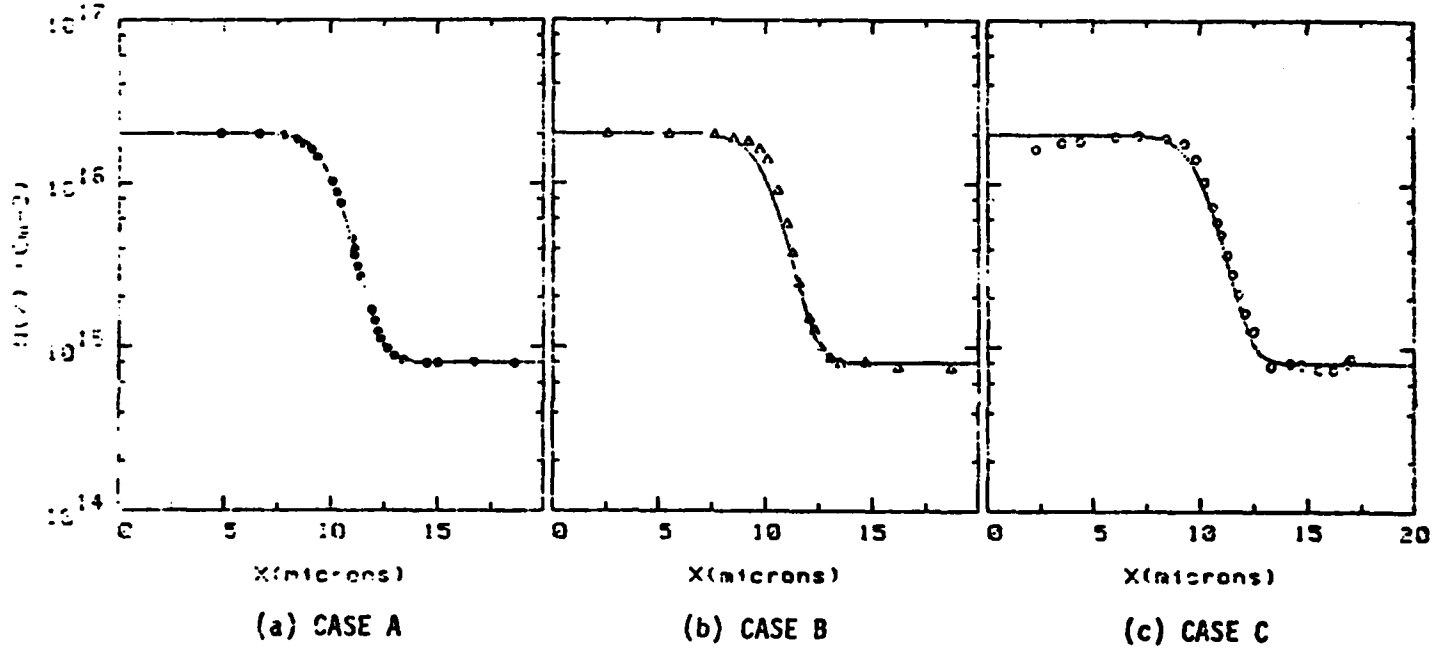



Fig.4.6.36. Calculated $I_D$ - $V_G$ characteristics for the device of Figure 2 using three methods.

CASE A ——————

CASE B .....

CASE C —·—·—·—·—




(a) CASE A (b) CASE B (c) CASE C

Fig.4.6.37 The surface Impurity Profile of the Device in Figure 2.

solid plot : actual profile

symbol plot : calculated profile

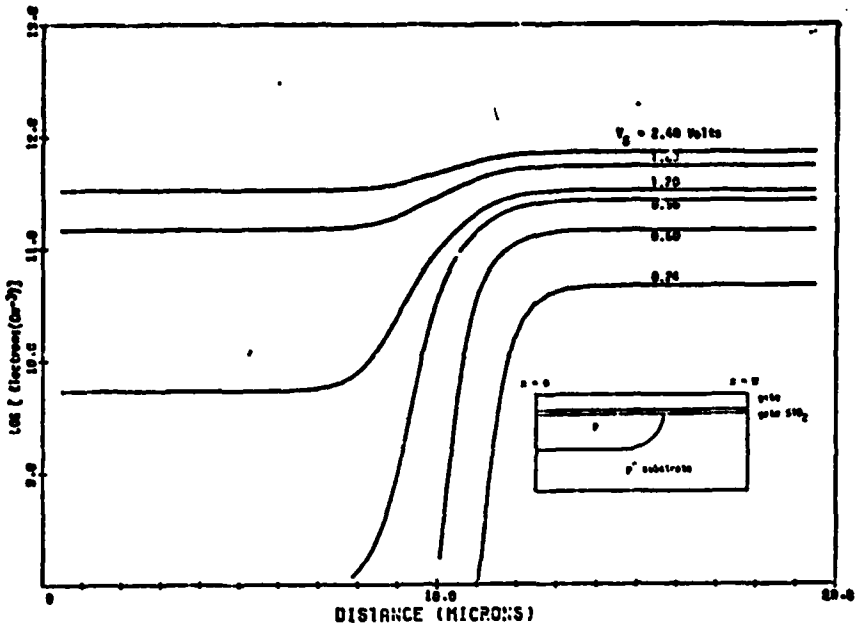



Fig.4.6.38 Inversion Carrier concentration at several gate potentials calculated by a numerical two-dimensional device simulation program.

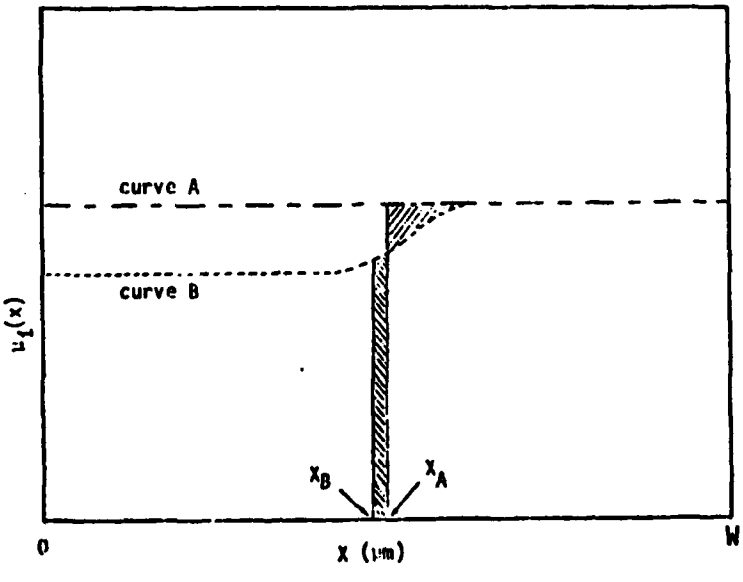



Fig. 4.6.39 Two successive approximations for the channel mobility across the device width, $\mu_1(x)$.

$$\int_{X_A}^{W} \mu_1(x)\,dx = \int_{X_B}^{W} \mu_1(x)\,dx$$

curve A curve B

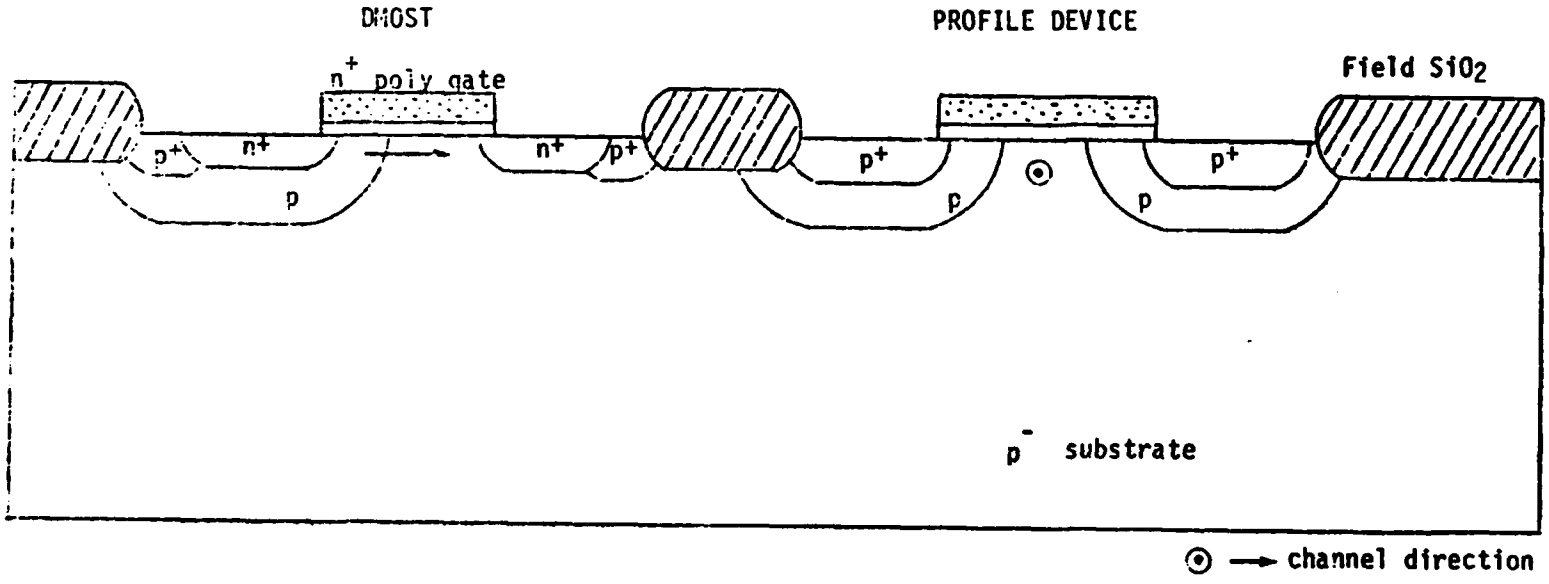
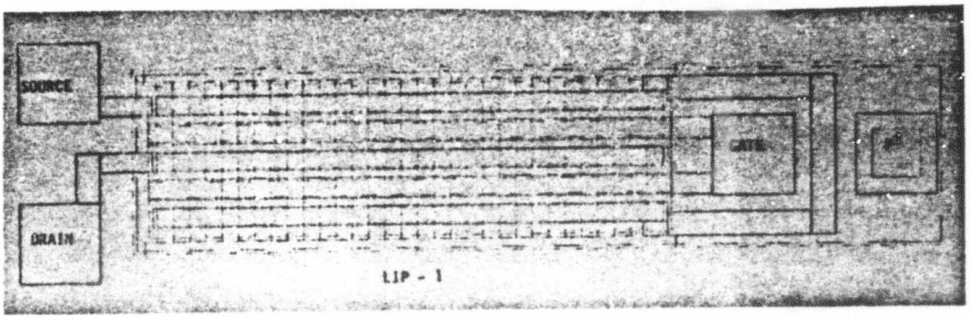
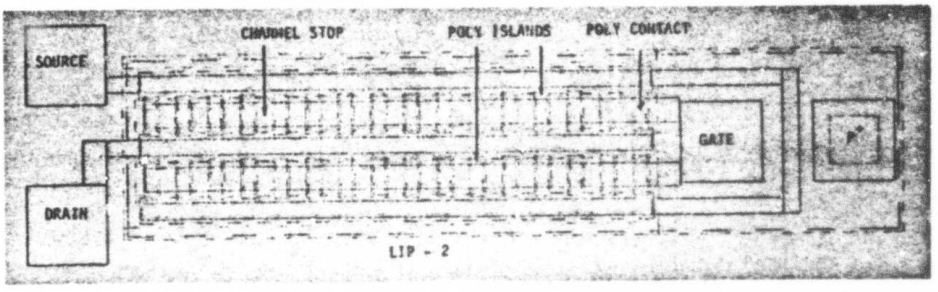



Fig.4.6.40 The cross-sectional view of the LIP-2 Device and the corresponding DMOS transistor. The electrical channel direction(source to drain) is indicated by arrows.



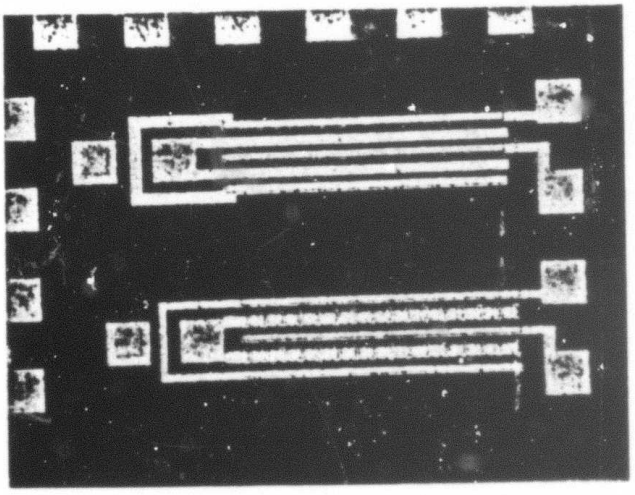
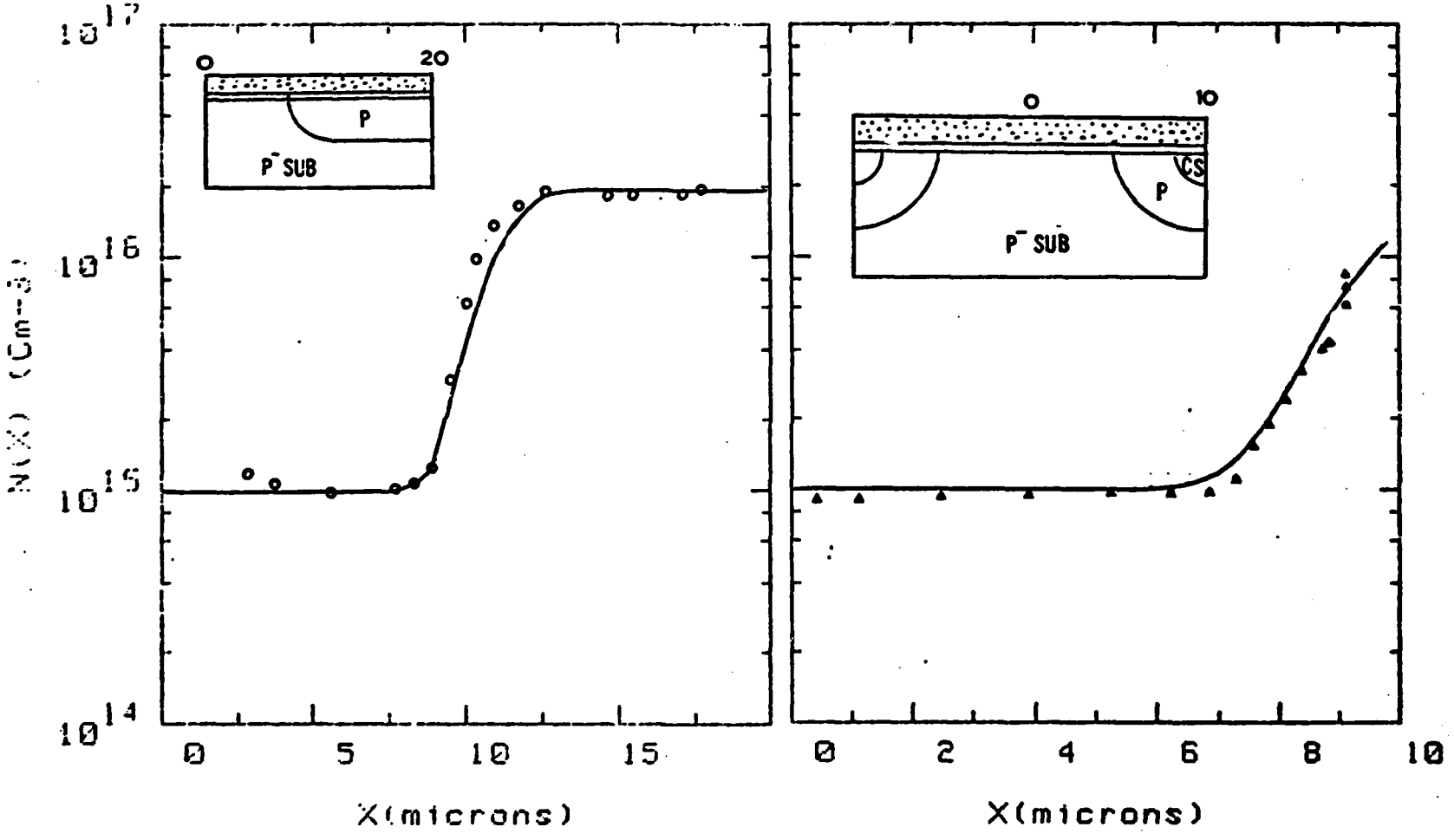



Fig.4.6.41 Layout of LIP devices and actual test devices fabricated.




(a) LIP - 1

(b) LIP - 2

Fig. 4.6.42 Measured DMOS Channel profiles from LIP Devices. Symbol plots are measured data and solid plots are theoretical profiles.

## 4.7 LATCH-UP IN CMOS INTEGRATED CIRCUITS

D. B. Estreich

In a previous report [4.1 ] the physics of latch-up in CMOS integrated circuits was described. Recent results of this CMOS latch-up research is described below.

A new test mask set has been designed and fabricated to give additional insight into latch-up physics. This mask set has been completed in conjunction with Sandia Laboratories, Albuquerque, New Mexico [4.42]. The LURIC test mask - Latch-up and Radiation Integrated Circuit - was designed to investigate CMOS latching and radiation effects. More specifically, LURIC was intended to (a) provide information on the physics of CMOS latch-up, (b) study the layout dependence of such parasitic pnpn paths, and (c) provide controlled latch-up test structures for the development and verification of latch-up models. Many of the devices and test patterns on LURIC are also well suited for radiation effects studies.

The LURIC mask set contains eighty-six devices and related test structures. A twelve layer mask set allows both metal gate and silicon gates CMOS test cells to be fabricated. Six categories of test devices and related test structures are included. There are (a) the RCA CD 4007A metal gate dual complementary pair plus inverter IC with auxiliary test cells (often used for latch-up study), (b) the Sandia Laboratories extended linear array CMOS high-density layout cells, (c) field-aided lateral pnp test transistors, (d) p-well sheet resistance and substrate spreading resistance test patterns, (e) symmetrical latch-up path test

test structures (for study of dimensional influence on latch-up thresholds), and (f) support test patterns (e.g., MOS capacitors, p+n leakage diodes, MOS monitor transistors, van der Pauw and Kelvin contact resistance test patterns, etc.). A standard probe pattern array has been used on all twenty-four subchips for testing convenience.

Fig. 4.7.43 shows all twenty-four subchips of the LURIC mask set. Ten field-aided lateral pnp test transistors are included on LURIC. As previously discussed [4.1], field-aiding can greatly increase the current gain of the parasitic lateral pnp in a CMOS latch-up path. For this reason, it is important to have a theory to predict the current gain enhancement from field-aiding when modeling a pnpn latch-up path. Fig. 4.7.44 shows the surface layout of the lateral pnp transistors as included on LURIC. Note that separate substrate bars are used to control the aiding field from emitter-to-collector.

Fig.4.7.45 shows a cross-section of lateral pnp as it often occurs in CMOS layouts. To model the field-aided lateral pnp transistor, two major current components are shown: a lateral current component $J_\ell$ and vertical (downward) current component $J_V$. The component $J_\ell$ is in the proper direction to experience field-aiding or drift enhanced transport. The electric field E therefore improves the lateral transport efficiency of the lateral current $J_\ell$. In addition, the presence of the electric field means that there is lateral debiasing of the emitter junction such as to concentrate the minority carrier injection into the base at the emitter edge closest to the collector of the pnp. This reduces the current component $J_V$, which is a loss current component because it completely recombines in the base region. Therefore, the existence

of the electric field will improve or increase the current gain of the parasitic pnp. Unfortunately, for latch-up considerations this is undesirable because it means that the IC will be more susceptible to latch-up due to the higher lateral pnp current gain.

A theory has been developed to predict the pnp current gain enhancement with aiding field strength. In addition to the two current components listed above, both the emitter-base depletion layer recombination and surface recombination currents are included in the theory. Fig.4.7.46 shows the behavior of the lateral pnp current gain as a function of the base width to diffusion length ratio ($W_B/L_B$) and lateral electric field strength (E). In Fig. 4.7.46 the solid lines are without depletion layer recombination, whereas, the dashed lines shows the characteristics with depletion layer recombination included. Two features are worthy of note: (a) the lateral pnp has low current gain for $W_B/L_B > 0.2$, and (b) small electric field strengths produce dramatic increases in the current gain. The field strength range given in Fig.4.7.46 is representative of the fields encountered in latch-up conditions for CMOS integrated circuits. This deoendence can be more effectively interpreted in the plot shown in Fig. 4.7.47. Here the lateral pnp current gain is normalized to the zero field current gain and plotted as a function of electric field strength for several values of $W_B/L_B$ (Fig.4.4.47 includes depletion layer recombination). The gain enhancement is greater for longer base devices.

Theory is compared to experimental data in Fig. 4.4.48. Two devices where $W_B/L_B$ is equal to 1/2 and 1 are represented by the experimental points. Excellent agreement is attained for lower field strengths.

Neglect of the charge between the vertical and lateral (horizontal) currents or charge distributions leads to the discrepancy at higher fields.

An additional aspect of this work has produced a novel way to estimate the minority carrier diffusion length $L_B$ (and, hence, lifetime). The current gains at E=0 are measured on four lateral pnp transistors of differing base widths ($W_B$). The theoretical gain expression is then used to fit these gains with $L_B$ varied. Fig. 4.7.49 shows typical results where base widths in the ratio of 1:2:4:6 have been used.

Neutron irradiation is sometimes used to control CMOS latch-up [4.43]. The effect of neutron irradiation (Energy > 10 KeV) on lateral pnp current gain has been modeled. The lifetime ($\tau$) of minority carriers varies with neutron fluence $\Phi$ (neutrons/cm$^2$) as

$$\frac{1}{\tau(\Phi)} = \frac{1}{\tau_{prerad}} + K\,\Phi\ , \qquad (4.7.72)$$

where $\tau_{prerad}$ is the minority carrier lifetime before irradiation and K is the empirically determined damage constant. The recombination rate is expected to be proportional to $\Phi$ because the number of induced defects is proportional to $\Phi$ while the lifetime is known to be inversely proportional to the number of recombination defects. With this relationship between $\tau$ and $\Phi$, the diffusion length ($L_B$) in the base of the pnp depends upon $\Phi$ as

$$\frac{1}{L_B} = \sqrt{\frac{1}{L_{prerad}}^2 + \frac{K}{D_p}\Phi} \qquad (4.7.73)$$

where $D_p$ is the diffusion coefficient and $L_{prerad}$ is the diffusion length before neutron irradiation. The depletion layer recombination current,

denoted $I_{RG}$, is proportional to

$$I_{RG} \alpha \frac{W_T}{\tau_{sc}} (1 + K_{sc} \Phi) \quad . \tag{4.7.74}$$

Width $W_T$ is the depletion layer width, $\tau_{sc}$ is the lifetime in the depletion layer, and $K_{sc}$ is the effective damage constant in the depletion layer; again, empirically determined. Taking data from Sokol and Adams [4.44], the lateral pnp model has been used to predict the decrease in pnp current gain as a function of neutron fluence. Typical results are shown in Fig. 4.7.50. In Fig. 4.7.50 the pnp current gain has been normalized to the preradiation level and the aiding field is near zero. Generally, good agreement is obtained for predicting radiation decreased current gain.

In summary, recent work has focused on modeling the wide-base, field-aided lateral pnp transistor which forms one of the parasitic bipolar transistors in CMOS latch-up paths. It has been shown that small electric fields greatly enhance the pnp current gain. Both depletion layer and base surface recombination are included in the model. Agreement between the model and experimental results from the LURIC test chip have been demonstrated. Finally, the lateral pnp model has been developed to include neutron irradiation gain degradation effects.

The help of Dr. C. W. Gwyn, and Dr. A. Ochoa, Jr. at Sandia Laboratories is gratefully acknowledged.

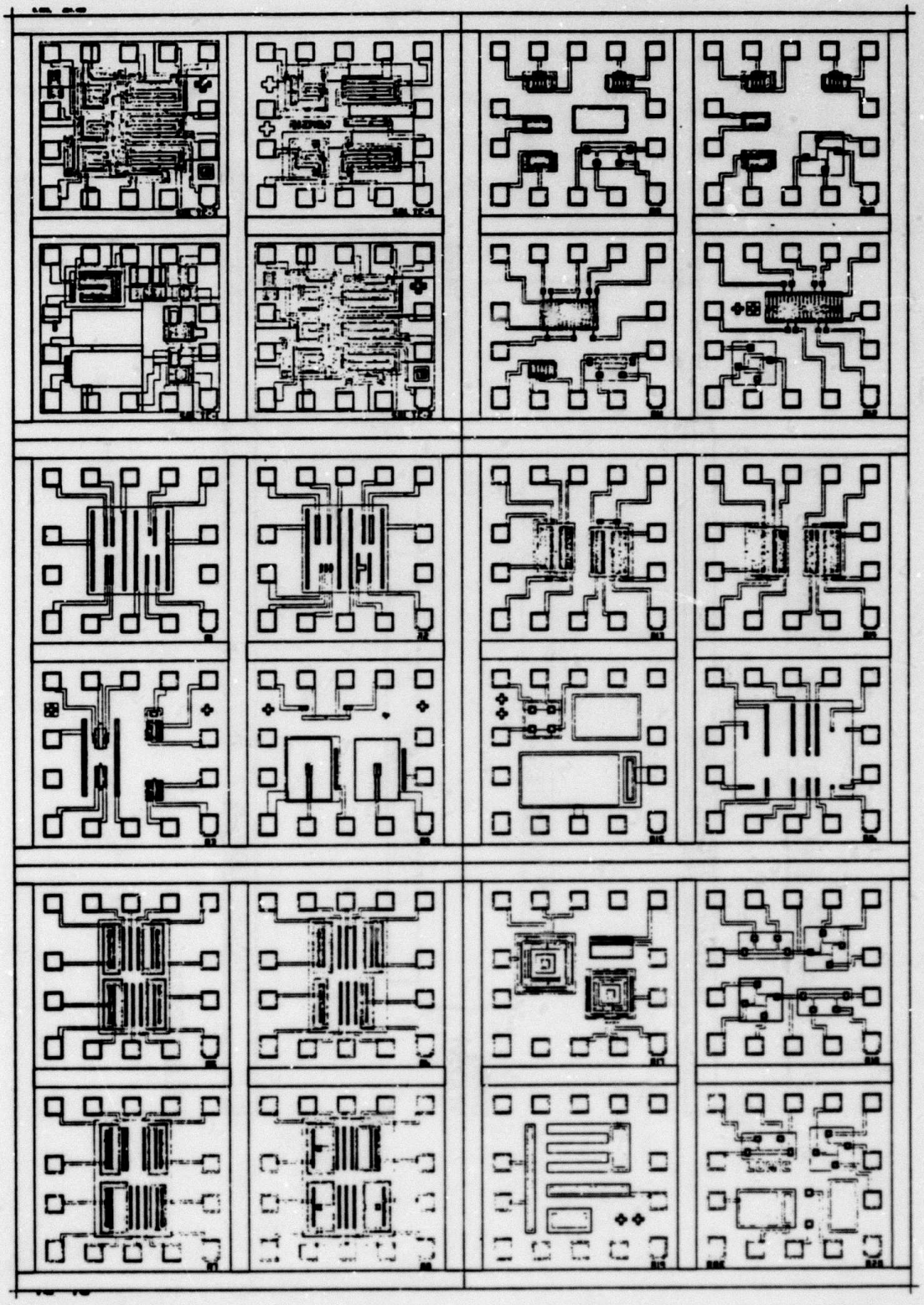


Fig. 4.7.43. Twenty-four subchips of the LURIC test mask set [4.42].

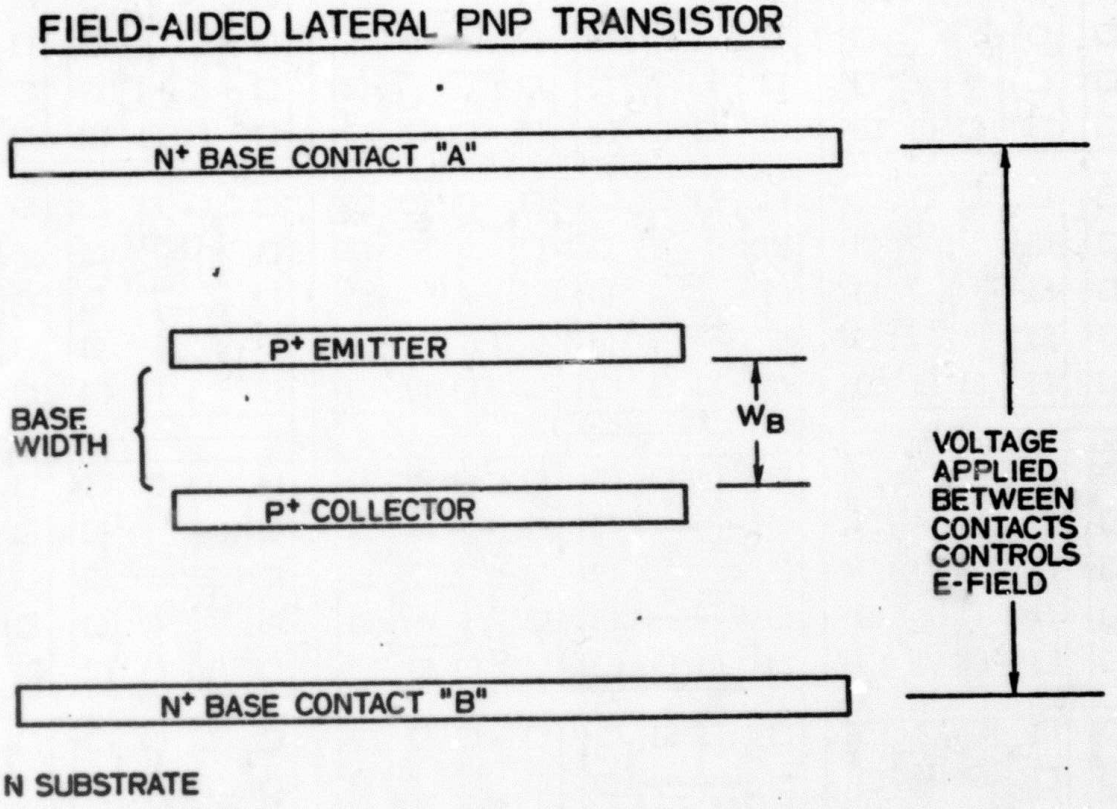



Fig. 4.7.44. Surface layout of field-aided lateral pnp.

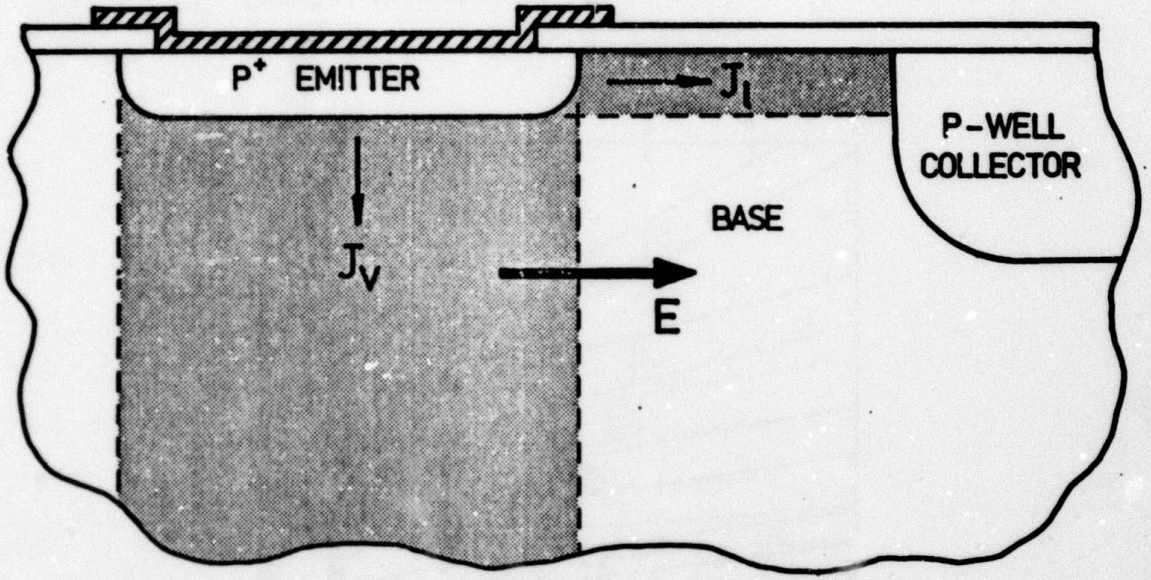



Fig. 4.7.45. Cross section of lateral pnp with both vertical and lateral current paths.

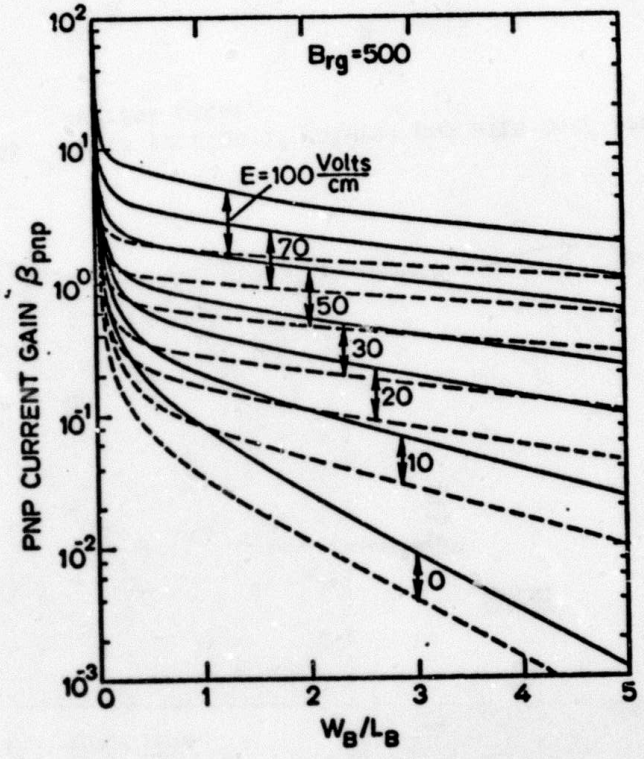



Fig. 4.7.46. Computed current gain versus normalized base width ($W_B/L_B$) with electric field as a parameter. The dashed curves show the result of including depletion region combination in the computations.

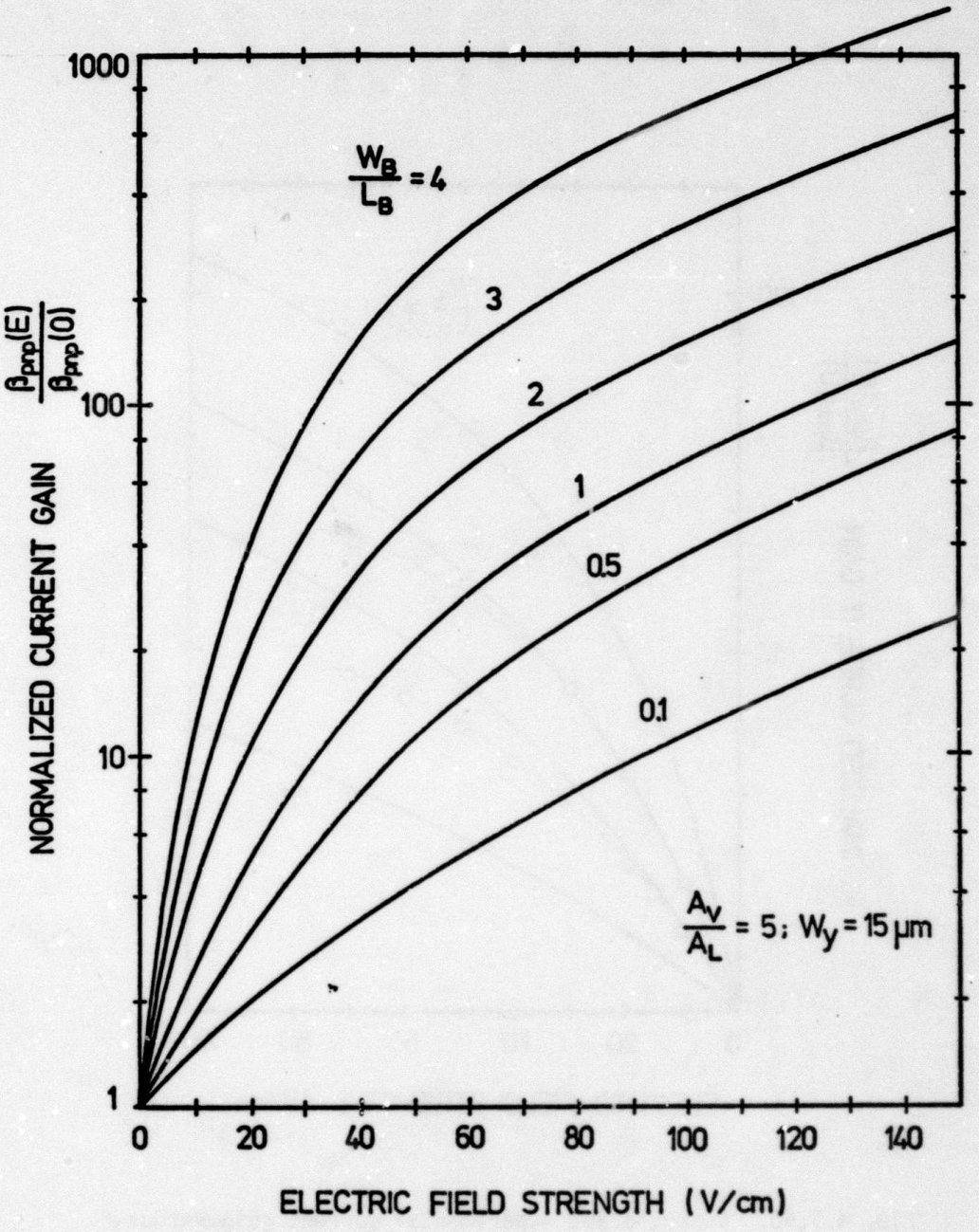



Fig. 4.7.47. Computed normalized current gain [β(E)/β(0)] versus electric field strength with normalized base width ($W_B/L_B$) as a parameter.

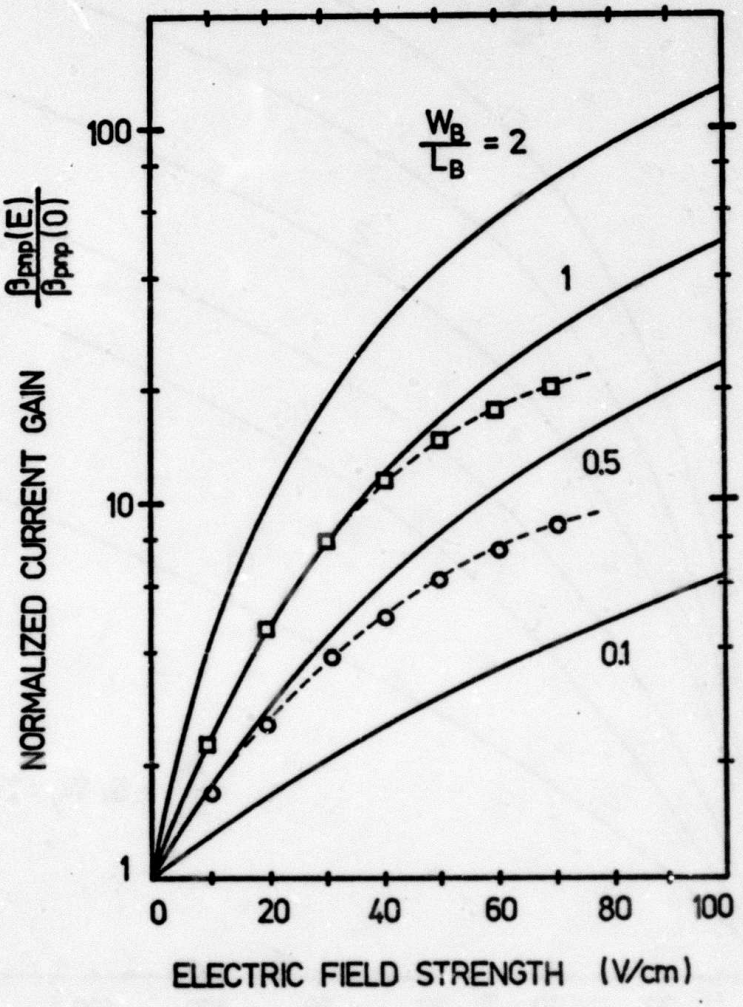



Fig. 4.7.48. Computed and experimental current gain versus electric field strength.

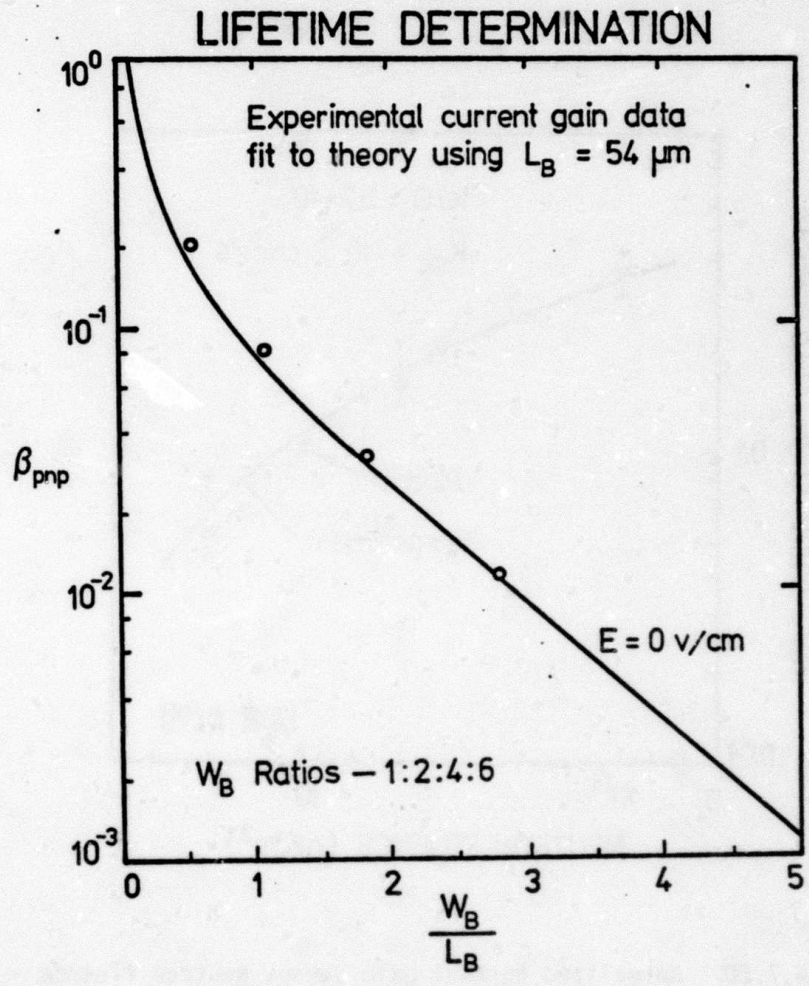



Fig. 4.7.49. Life-time determination based on fit of current gain versus normalized base width.

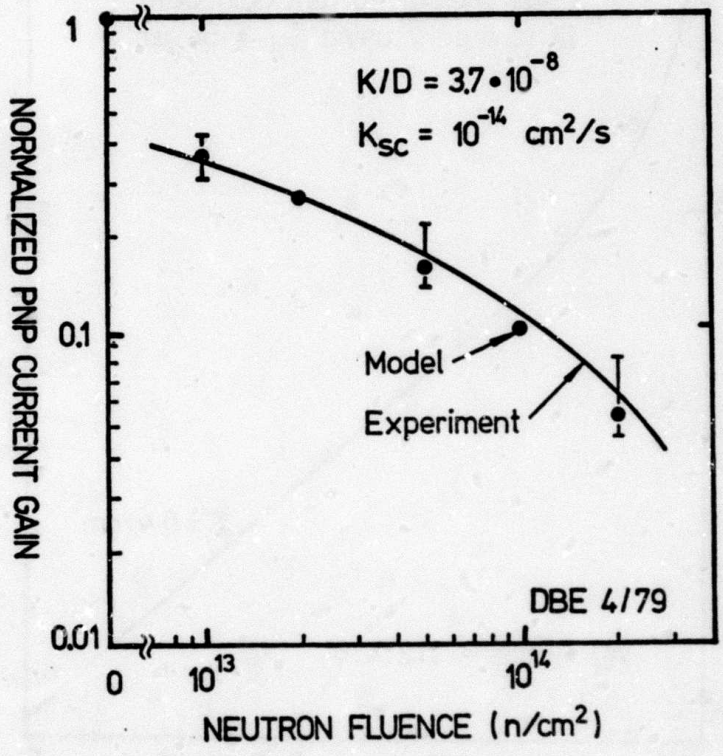



Fig. 4.7.50. Normalized current gain versus neutron fluence -- both experimental and model results.

## 4.8 A CHARGE ORIENTED MOS CAPACITANCE MODEL

D. Ward

The principles of a new MOS transistor capacitance model have been set forth in reference [4.45 ]. The modeling approach will be briefly outlined here and compared with the conventional approach. Several unique consequences of the model are discussed.

Both the charge-oriented and conventional models are based on the equation

$$i_G = \frac{dQ_G}{dt} \tag{4.8.75}$$

where $i_G$ is the transient gate current and $Q_G$ is the charge on the gate electrode. Although gate capacitances are discussed here, the equations (with suitably modified subscripts) and discussion apply to all four terminals of the transistor. Both models invoke the quasi-static approximation by assuming that the charge $Q_G$ is a function of the instantaneous values of the terminal voltages. The conventional model rewrites the right-hand side of (4.8.75) as

$$\frac{dQ_G}{dt} = C_{GD}\,\frac{d(V_G-V_D)}{dt} + C_{GS}\,\frac{d(V_G-V_S)}{dt} + C_{GB}\,\frac{d(V_G-V_B)}{dt} \tag{4.8.76}$$

where

$$C_{GD} \equiv -\frac{\partial Q_G}{\partial V_B}, \quad C_{GS} \equiv -\frac{\partial Q_G}{\partial V_S}, \quad C_{GB} \equiv -\frac{\partial Q_G}{\partial V_B}$$

This suggests that the gate current is the sum of the currents through three capacitors. Each component of current is given by

$$i = C\,\frac{dV}{dt} \tag{4.8.77}$$

A circuit simulation program integrates this expression over one time-step, to give

$$\int_{to}^{t_1} i\,dt = \bar{C}\,(V_1 - V_o) \qquad (4.8.78)$$

where $\bar{C}$ is the capacitance averaged between the voltages $V_o$ and $V_1$ at times $t_o$ and $t_1$ respectively. $\bar{C}$ is unknown and must be approximated. This approximation can lead to numerical errors (non-conservation of charge), slow convergence to solution, and numerical instability. In addition, all other capacitive currents are usually obtained from the three capacitors derived from the gate charage. This implies a reciprocity which does not exist e.g. that $i_G \Big/ \frac{dV_B}{dt} = i_D \Big/ \frac{dV_G}{dt}$ and omits capacitive effects between bulk and drain, bulk and source, and drain and source.

The charge-oriented model works directly with equation (4.8.75). This is integrated by the circuit simulation to give

$$\int_{to}^{t_1} i_G dt = Q_{G1} - Q_{Go} \qquad (4.8.79)$$

The right-hand side is evaluated by linearizing $Q_{G1}$ about the operating point obtained from the most recent iteration. Thus at convergence the right-hand-side of (4.8.79) is given exactly while the right-hand-side of (4.8.78) contains the unknown $\bar{C}$. Equations analogous to (4.8.75) and (4.8.79) may be written for each of the other terminals. The calculation of charge directly from the terminal voltages insures that charge is conserved, while the complete linearization of (4.8.79) provides for numerical speed and stability.

This model displays the other features not found in the conventional model. Firstly, capacitive effects are defined between every pair of terminals. Thus bulk-drain, bulk-source, and drain-source effects are modeled. Secondly, the relationship between each pair of terminals may be non-reciprocal. Thus, for example,

$$C_{GD} \equiv \frac{\partial Q_G}{\partial V_D} \neq \frac{\partial Q_D}{\partial V_G} \equiv \quad DG \tag{4.8.80}$$

implies that the effect of the drain voltage on the gate current is not the same as the effect of the gate voltage on the drain current. This is illustrated in Figure 4.8.51. ($C_{DG}$ and $C_{GD}$ may be viewed a capacitive transconductance and reverse transconductance, respectively, or $C_{GD}$ may be viewed as a conventional two-terminal capacitance and $C_{DG}$- $C_{GD}$ as a capacitive transconductance.)

The charge-oriented model requires that a charge obeying (4.8.75) be associated with each electrode and that these charges are determined by the instantaneous terminal voltages. The second requirement is satisfied if the quasi-static approximation is assumed. The first requirement is easily met for the gate and bulk, as the charges associated with these electrodes are electrically isolated from the other terminals. The drain and source charges, however, join to form the channel and thus are not readily identified physically. These charges may be found by calculating the transient drain and source currents and integrating. The accuracy of the model depends on the assumptions made in finding the charges and in the validity of the quasi-static approximation. For implementation into a circuit simulator, the following aproximations were made:

(1) gradual channel (one dimensional) approximation

(2) depletion approximation

(3) strong inversion approximation

(4) conduction by drift only

(5) saturation occurs by pinch-off of the channel; no channel-length modulation effects included

Unless otherwise noted, model results presented here are based on these approximations.

4.8.1 Capacitance Measurements

Measurements have been made to verify features of the capacitance model, particularly the predictions of non-reciprocity. Measurement equipment and techniques will be briefly described. Measured data will then be presented and compared with model predictions.

Capacitance measurements were made with a Hewlett-Packard 4271B LCR meter. This instrument applies a 1 MHz signal at 20 mV rms to the "high" test leads and in effect measures the magnitude and phase of the 1 MHz current at the "low" test leads. Thus, to measure $C_{GD}$ the "high" test leads are connected to the drain of the device while the "low" test leads are connected to the gate. The test leads are reversed to measure $C_{DG}$. Transistor bias is provided by three HP-6131C digital voltage sources. All instruments are under programmed control of an HP-9845 desktop computer.

Measurements were made on a large (300 x 300 μm) device made with a silicon-gate NMOS process. A threshold-tailoring implant resulted in a non-uniform bulk doping. An average bulk doping value (extracted from measured I-V data) was used in the model.

Measured plots of $C_{GD}$ and $C_{DG}$ vs. $V_G$ are compared with theoretical curves in Figure 4.8.51. Note that whereas $C_{GD}$ is zero in saturation as predicted by conventional theory, $C_{DG}$ is not. It is seen that the charge-oriented model for capacitance is in good agreement for larger gate voltages. The rounding of the measured characteristics at low gate voltages is due to failure of the quasi-static approximation for the long device at the 1 MHz measurement frequencies.

### 4.8.2 Device Analysis Using CADDET

Two-dimensional numerical device analysis has become an invaluable tool for model development. Such an analysis tool can be used to probe the physics of device operation; to isolate the effects of various physical phenomena (e.g. velocity saturation); and to evaluate and/or model new devices and processes.

We have used two-dimensional numerical device analysis to obtain the charges required by the capacitance model. The device analysis program (CADDET) will be briefly described and some of the results will be discussed. The charge distributions in a short-channel device will then be considered.

The CADDET (for Computer-Aided Device DEsign in Two dimensions) program, developed at Hitachi, performs a two-dimensional, steady-state, numerical device analysis using the stream function method described by Mock [4.46]. The program can be used to find values of gate, bulk, and channel charges which are not subject to the approximations listed above. The principle theoretical limitations are the use of Boltzmann statistics and restriction to steady-state conditions, while the principle practical limitations are the numerical accuracy and required

computation time. The method for deriving source and drain charges from two-dimensional steady-state analysis has not yet been developed.

Although the model presented here uses the charges in a device, its consequences are more easily understood by considering the associated capacitances. In addition, the capacitances are a more sensitive indicator of model accuracy than are the charges (see Figure 4.8.52). The computation of capacitances by CADDET, however, is done by numerical differencing, and thus requires greater solution accuracy.

Comparisons were made between CADDET calculations and the analytic model applied to a hypothetical device having lengths of 7-9 microns, deep junctions, a substrate doping of $5 \times 10^{15} cm^{-3}$, and an oxide thickness of 1000 Å. Typical DC characteristics are shown in Figure 4.8.53. The main differences between the CADDET and analytic models is the curvature near threshold and the difference ($\sim 5 \frac{kT}{q}$) in threshold voltages. These are to be expected from the depletion and strong-inversion approximations used in the analytic model. Typical capacitance characteristics are shown in Figure 4.8.54. Again the numerical solutions show smoother transitions between cutoff and saturation and between saturation and linear regions.

Also shown in Figures 4.8.53 and 4.8.54 are results obtained using a simplified Pao and Sah model which does not use the depletion, strong inversion, or drift approximations at the source end of the device. It is seen that this model is in much better agreement with CADDET for gate voltages near threshold.

It was noted that for gate voltages below threshold, the values of $C_{GB}$ predicted by CADDET are consistently lower than those derived from the analytic models. The difference between the two models arises

from the two-dimensional distribution of charge in the corner regions of the device, at the source and drain junctions near the surface. Analysis of these regions using CADDET has led to the development of a capacitance model for short devices.

For long devices the charge in the device is taken as the sum of the net charges from the three idealized case of Figure 4.8.55. Thus the gate and oxide on depleted bulk material, p-n junction, and gate and oxide on accumulated source/drain are treated as non-interacting one-dimensional structures. This approximation is valid for long devices, in which the charge in the corner regions is negligible compared to the charge in the rest of the device. For short devices the corner charge must be considered. The superposition of these three cases results in a charge distribution which is adequate for modeling long devices but is in considerable error in the corner regions. Figure 4.8.56 illustrates the actual distribution in these regions as calculated using CADDET. Four places where the long-channel approximation is in error are indicated. The largest error is due to the overlap of the channel and junction depletion regions. (①in Figure 4.8.56). In the example shown, the charge given by the long-channel approximation differs from the computed charge by -1.2 pico Coulombs. Most of this is compensated by an increased accumulation of electrons at the surface near the junction (③ in figure), an error of 0.9 pc. The other two error terms are indicated by ② and ④in the figure. The term ② consists of the rounding of the depletion region near the junction and is small compared to term ① ($\Delta Q_2$ = 0.2 pC). Term ④ consists of an added gate charge. While it is small for the case considered here ($\Delta Q_4$ = .1 pC), it is larger

for smaller gate voltages.

Incorporation of these terms into the charge equations produces a capacitance model for short-channel devices. Terms ① and ② are added as corrections to the bulk charge while terms ③ and ④ modify the source/drain and gate charges, respectively. Approximations to these charges have been developed which account nicely for the error in $C_{GB}$ below threshold although only term one has been accurately modeled.

### 4.8.3 Discussion and Conclusions

It has been shown that an approach based directly on the charge distributions within a device can provide more complete modeling of capacitive effects with improved numerical accuracy. Measurements have confirmed the predictions of the model.

An important new tool is the two-dimensional numerical device analysis program. This tool allows the model developer to study the physics of a two-dimensional structure and to gain insight unobtainable by analytical or experimental means. We have shown how such a program (CADDET) has been used to evaluate on analytical model and to develop a model for the charge distributions in the corner regions of a short-channel device.

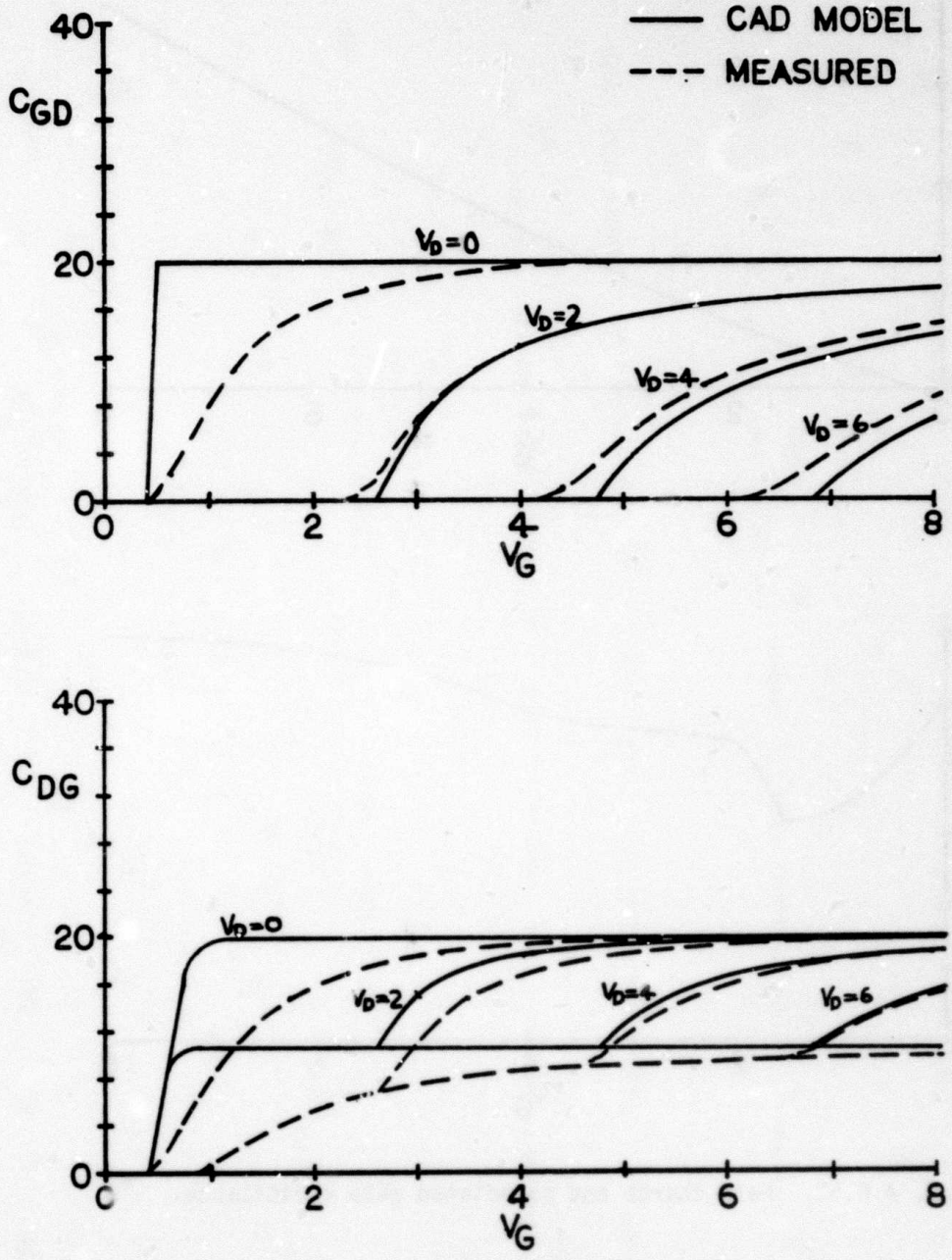



Fig. 4.8.51. Theoretical and measured gate-drain capacitances.

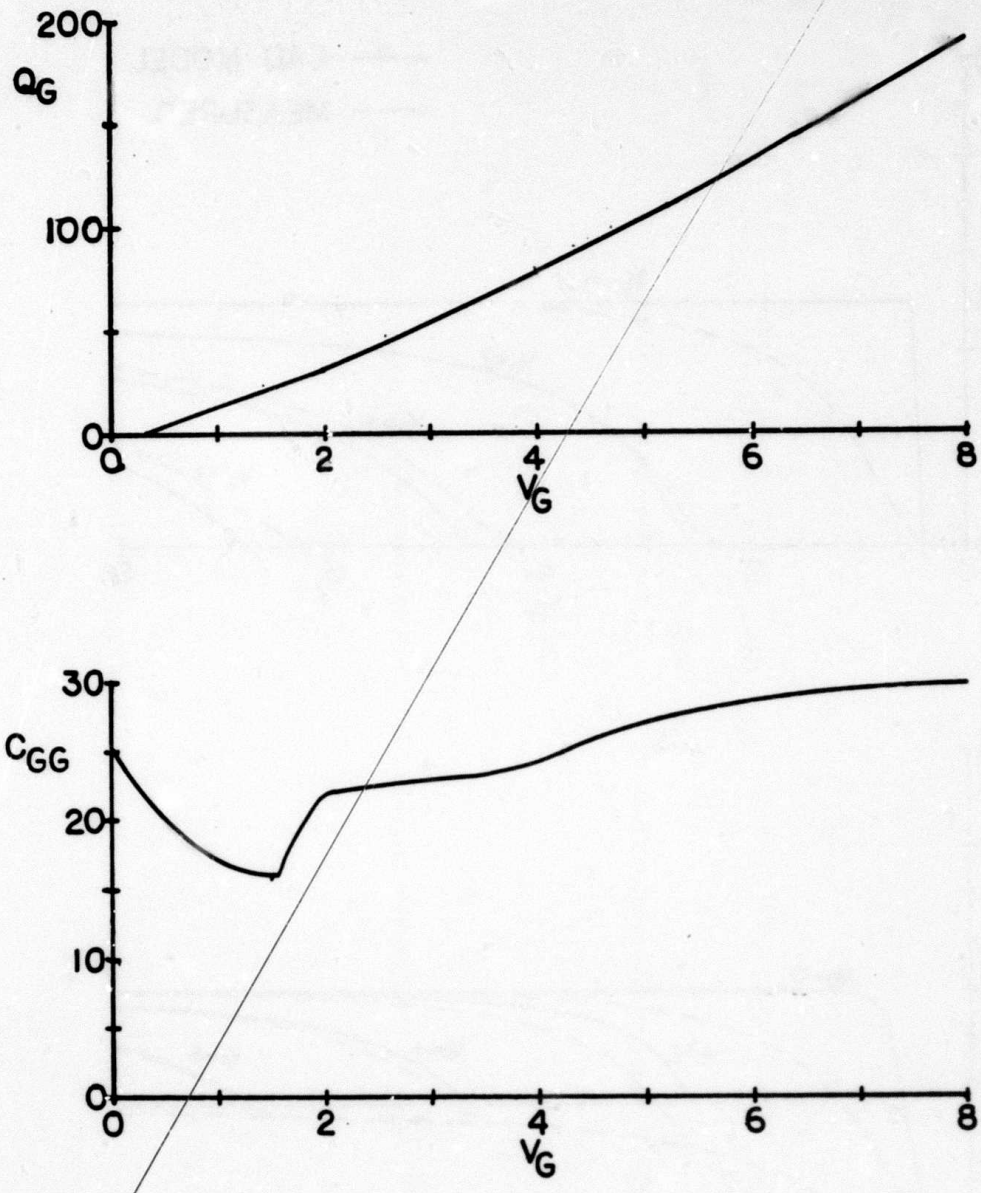



Fig. 4.8.52. Gate charge and associated gate capacitance.

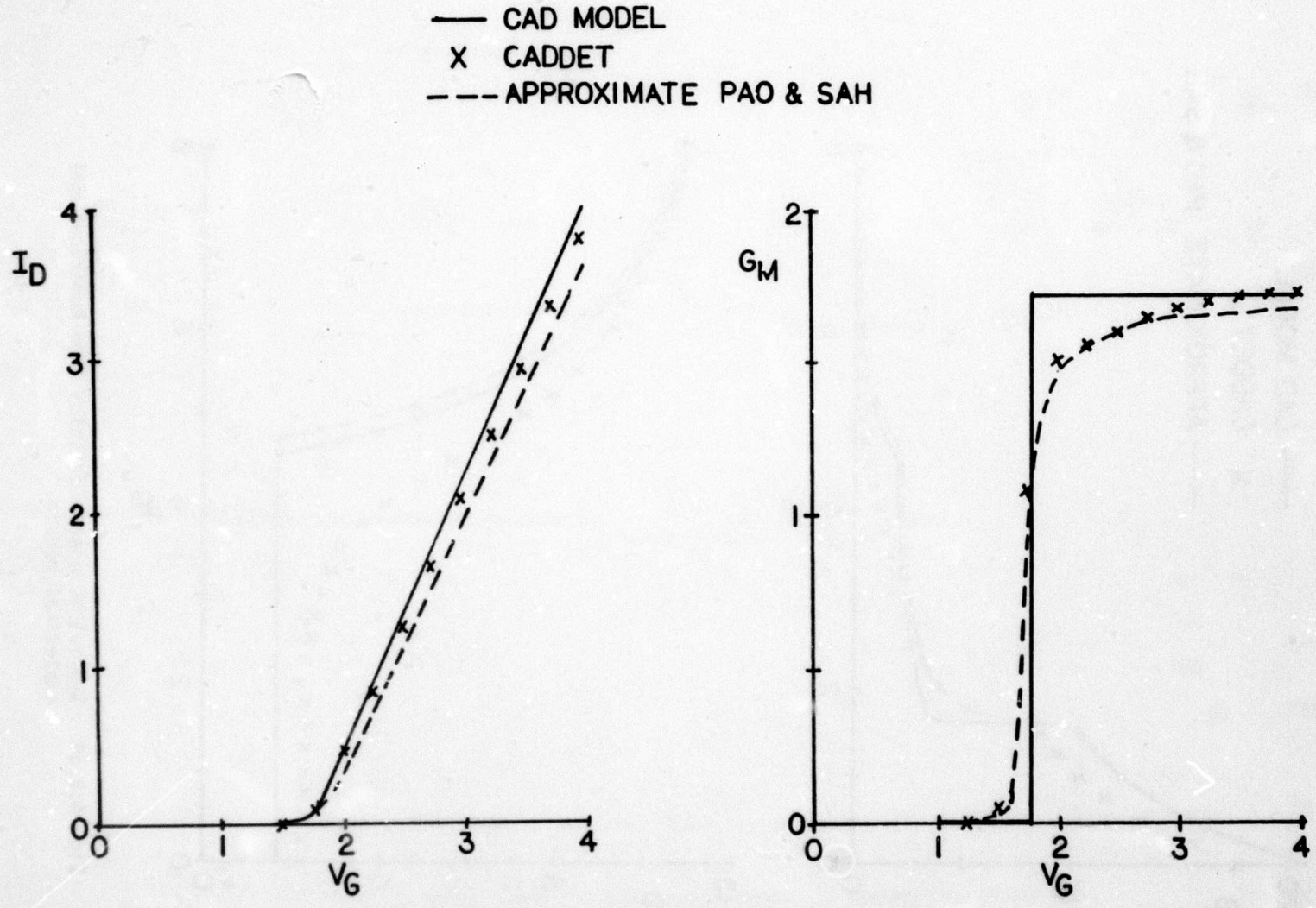



Fig. 4.8.53. Typical DC characteristics from numerical and analytical models.

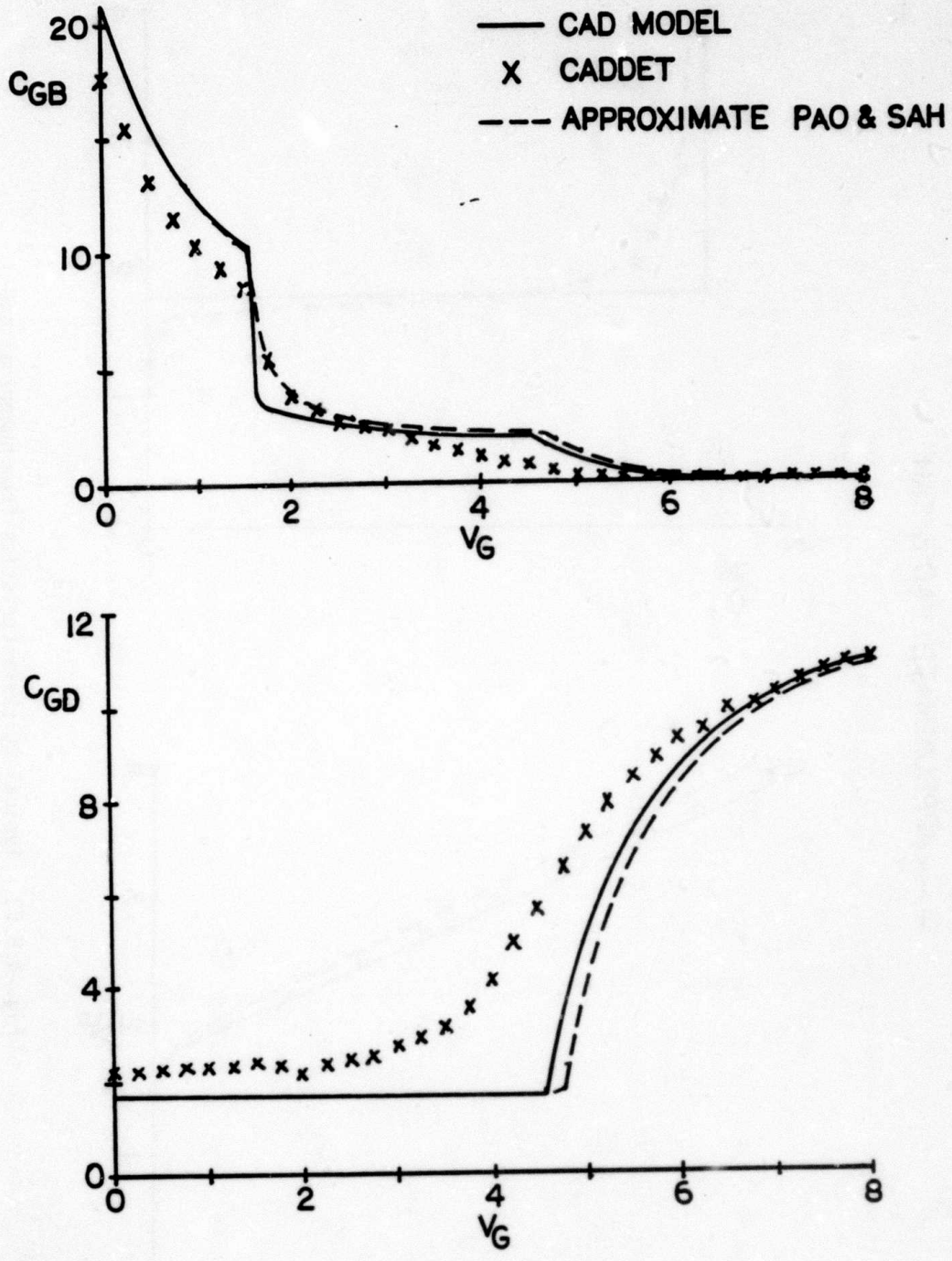



Fig. 4.8.54. Typical AC characteristics from numerical and analytical models.

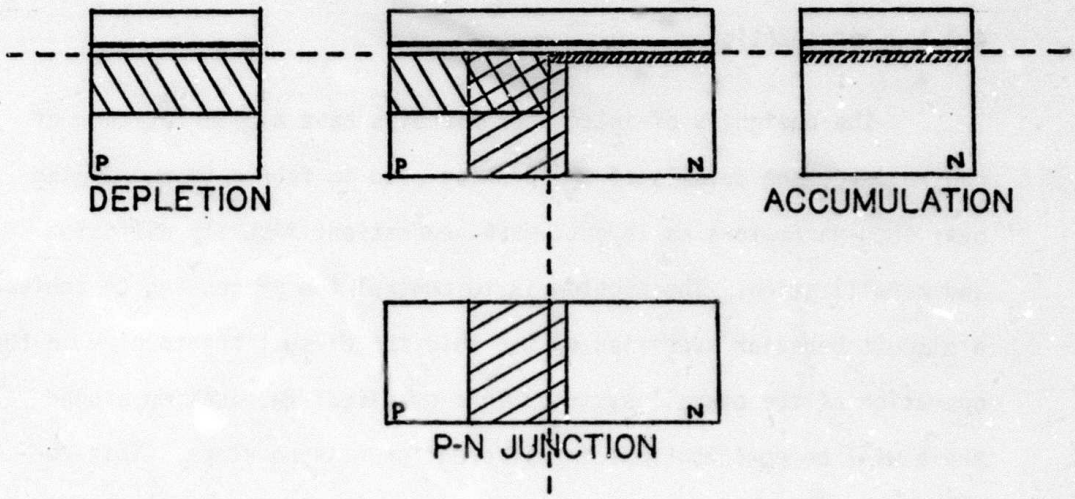



Fig. 4.8.55. Approximation of charge in corner region.

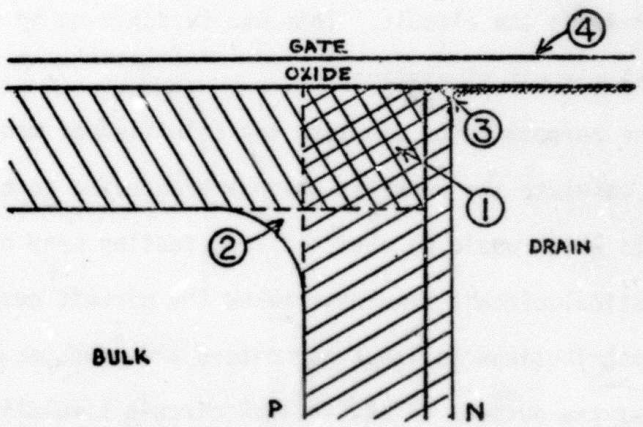



Fig. 4.8.56. Charge errors associated with corner region approximation.

## 4.9 DC STATISTICAL CIRCUIT ANALYSIS FOR BIPOLAR IC'S USING PARAMETER CORRELATIONS -- AN EXPERIMENTAL EXAMPLE

D. Divekar, W. J. McCalla

### 4.9.1 Introduction

The designers of integrated circuits have a wide latitude of control over the details of the process used in fabrication, ranging over such parameters as layout, mask-generation, impurity diffusion and metallization. The problem is to control the processing to achieve a circuit behavior specified by the role the circuit has to play in the operation of the overall system. In a practical fabrication procedure, there will be perturbations in different processing steps. This results in variations in terminal behavior of the devices which in turn is reflected as a spread in the circuit performance. To avoid the enormous cost of trial-and-error, the spread in circuit behavior must be calculable from the perturbations in the terminal behavior of devices comprising the circuit. This can be achieved by the use of statistical circuit simulation.

The purpose is to simulate the selection of many sample circuits and tabulate the results, which are expected to correspond to the results which would be obtained from testing many physical samples. The statistical circuit simulator takes the circuit description and element distributions as input parameters and produces circuit performance as the output. A statistical circuit simulation program has been developed to predict the effect of component variations on network performance. The program was designed for operation on

the Hewlett-Packard 2100 series mini-computers with 32K words of memory although the program can be used on other computing facilities with little difficulty. The program is written in HP Fortran IV. Some of the important features of the program are described here. More details are discussed elsewhere [4.47, 4.48].

Terminal behavior of devices is described to the simulation program in terms of model parameters. The problem to be faced is how to simulate the variations in model parameters taking into account their interrelationships. To study this dependence, a sample of devices is first characterized in terms of their model parameters. A statistical data reduction technique is then used to identify a few key independent parameters. The results of this model reduction are verified using the statistical simulation program. The performance of an emitter-coupled logic gate is simulated. These simulated results are compared with the measurements made on a sample of test die for this circuit. It is shown that favorable agreement can be obtained between measured and simulated results if model parameter correlations are taken into account using techniques to be described later [4.49 ].

### 4.9.2 Monte Carlo Technique

The simulation program uses the Monte Carlo Method [4.50] for statistical analysis. This method computes the probability distribution of circuit performance quantities from known probability distributions of input parameters. Circuit analysis provides the simulated functional relationship between the input element values and output circuit performance. The procedure is to analyze the circuit many times, each time letting

the circuit elements assume random values according to their prescribed distributions. Figure 4.9.57 shows the flow chart of this method.

An advantage of this technique is that detailed distribution of the circuit performance quantities can be obtained. Another advantage is that the Monte Carlo method makes no assumption concerning the distribution shapes themselves or the linearity of the input-output functional relations. It does, however, assume that all input parameters are statistically independent. Correlations which exist between circuit elements, resulting from the manufacturing process, can be explicitly accounted for by entering these correlations or tracking relationships [4.51] into the random selection process, as explained later. The Monte Carlo technique is also easy to implement in a circuit simulation program.

The method involves two basic steps, simulating the circuit performance and generating random element values.

Simulating Circuit Performance - One of the steps in statistical analysis is to simulate the circuit performance. This provides the functional relationship between the input parameters and output performance for the statistical analysis. The circuit simulator mini-MSINC [4.52,4.53] is used for this purpose. Mini-MSINC is a program capable of performing non-linear dc and time-response simulations of electronic circuits. The program uses a linked list structure for efficient allocation of memory. Nodal analysis [4.51] is used to compute circuit performance. The non-linear equations are solved with a modified Newton-Raphson method and sparse matrix techniques are employed to minimize storage requirements [4.51 ].

The program mini-MSINC can handle linear resistors, capacitors and inductors as well as diodes and MOS transistors. The modified Gummel-Poon model [4. 54] is implemented in the program to enable analysis of circuits containing bipolar junction transistors. Thus statistical analysis can be performed for circuits consisting of both MOS and bipolar transistors using the built-in models.

Random Number Generation - Random numbers are used to simulate stochastic phenomena and are produced using the arithmetic operations of a computer. Sequences generated in a deterministic way such as this are usually called pseudo-random or quasi-random sequences [4.55,4.56]. These sequences do possess the essential characteristitcs of a truly random sequence in the context of their particular application.

The main aim is to obtain sequences which behave as if they are random. Theory of statistics provides some quantitative measure of randomness. These tests can be divided into two groups: empirical tests, for which the computer manipulates groups of numbers from the sequence and evaluates certain statistics; and theoretical tests, for which the characteristics of the sequence are established by using number theoretical methods [4. 55 ]. Since the random number sequence is the heart of element generation, it is important to test its randomness. The random number generator used in the program is tested by applying over fifteen tests of randomness [4.55,4.56 ].

Circuit Description for Statistical Analysis - In order to provide a complete statistical description of electronic circuit elements used in integrated as well as discrete circuit designs, three fundamental capabilities must exist within a program [4.51 ]. First is the ability to describe

probability distribution for statistically varying input quantities. Second is the ability to generate variables which track each other. Finally the program should be able to override the randomness of variables in such a way as to simulate their interrelationships. The distinction is made here between tracking of the same parameter in different devices and correlations among different parameters within the same device.

The input language is designed in such a way that if a circuit description for analysis using mini-MSINC exists, then it can be used without any changes for performing statistical simulation of that circuit. The statistical description can simply be appended to the original description. On the other hand, if an input is being prepared for the first time, the statistical description can be incorporated very easily.

Shapes of statistical distributions of circuit elements or model parameters can be described in two ways. One of the standard distributions provided by the program can be used and its shape can be modified by specifying appropriate scale factors. The second way is to enter the distribution in the form of a table. Both discrete and continuous tables can be entered specifying either the probability density function or cumulative distribution function.

Tracking, in the present context, means the degree of matching between elements or model parameters within different circuits and devices. For example, the resistors on a single die of an integrated circuit will track each other. The tracking is specified in terms of a pivot element which is used as the independent variable and a tracking

coefficient. In integrated circuit technology more than one level of tracking can exist. For instance, the devices on a single die track each other. Also devices on different dice on a wafer track each other. Therefore, three levels of tracking are provided for in the program.

Correlation represents the interrelationships between various model parameters of the same device. These relationships can be entered using the equation specification. Equations can be specified in the form of FORTRAN-like expressions and once specified, they can be used repeatedly by passing desired parameters. In addition, certain parameters can be defined which are not a part of the circuit. These parameters can be used to simulate complex interrelationships between various device parameters.

Statistical Outputs - The program provides statistical information about the circuit performance. Envelope plots are provided in terms of minimum and maximum values as a function of time with the mean value running in between. The envelope plots can be sampled at the time points of interest to get more detailed information in the form of histograms. The histogram is a plot of the number of statistical cases which fall into each class interval versus the range of values of the output variable. Histograms of circuit elements as well as model parameters can also be plotted. A scattergram is a parametric plot which plots one output variable against another with the statistical cases as the parameter of the plot. These can provide information about the relationships between various output quantities and circuit elements or model parameters.

### 4.9.3 Circuit Example

The simulation program is used in conjunction with a simple circuit to illustrate the procedure of statistical circuit analysis. The circuit used is a low power version of an emitter-coupled logic gate with a single input [4. 57]. Figure 4.9.58 shows the circuit diagram alongwith the component and power supply values. The circuit is designed to operate between the logic levels of 0 V and -200 mV at an operating current of 2 μA. The output voltage and the input current when the input is 0 V are the performance quantities of interest. The transistors and resistors available on the KITCHIP are used for this study. The KITCHIP is a monolithic array of bipolar transistors, resistors and other components which can be interconnected to form a desired circuit by using a single metal mask [4. 58]. Fig. 4.9.60 shows the photograph of a KITCHIP die with the metal mask used for this study. Fifty dice are selected from a KITCHIP wafer and two npn transistors and three epitaxial resistors from each die are used to form the emitter-coupled gate. The output voltage and input current are measured. The results of measurements on the sample of 50 circuits are summarized below.

| | Min. | Max. | Mean | Std. Dev. |
|---|---|---|---|---|
| $V_{out}$ (mV) | -264.4 | -192.2 | -237 | 17.164 |
| $I_{in}$ (nA) | 8.8 | 90.14 | 24.6 | 18.12 |

Model Reduction -- To analyze the circuit using the simulation program, it is necessary to describe the terminal behavior of active devices in terms of their model parameters. In statistical simulation, device behavior over a wide range of operating currents and voltages is to be predicted. The simplified Gummel-Poon model [4. 54] achieves this purpose

without using an excessively large number of parameters. For statistical analysis, this model needs to be reduced to a few independent parameters, for the following reason. In statistical analysis, a circuit is analyzed many times, each time letting the varying elements assume random values according to their prescribed distributions. Parameters of a device usually change together in a certain manner, since they represent interrelated physical effects. Choosing such parameters randomly from their distributions can give rise to unrealistic combinations of model parameters, since this procedure implicitly assumes that no interrelationships exist. Such combinations will not represent physical effects and circuits analyzed using these model parameters may give misleading results, as will be shown later. Therefore it is necessary to find a set of independent parameters from which other parameters are calculated.

The transistors used in the test circuit are characterized in terms of their Gummel-Poon model parameters. The model parameters of interest are defined in Fig.4.9.61which shows the plot of collector and base currents as a function of base-emitter voltage on a semi-log scale. A study of the Gummel-Poon model parameters using the statistical data reduction technique of factor analysis has shown that the saturation current $I_s$ controls a large proportion of variance of the measured data [4.59 ]. Therefore, model parameters of importance to the circuit under consideration, are expressed in terms of $I_s$ using linear regression. These equations are shown below.

$$\beta_{FM} = 27.004\ I_s + 22.46$$

$$n_e = -0.063\ I_s + 1.598$$

$$\sigma_2 = -38.055\ I_s + 320.959$$

$$\beta_{fm} = 0.448\ I_s + 1.954$$

Statistical Circuit Simulation -- It is possible to use the circuit simulation program to generate element values for statistical analysis in three possible ways. Fig . 4.9.62 shows the flow chart of statistical analysis, highlighting these possibilities. Fig. 4.9.63(a) and (b) show this parameter selection process pictorially, where a simple case of generating values of two parameters is considered. Figure 4.9.63(a) shows the measured scatter plot of forward beta versus the saturation current $I_s$. Three different ways of accounting for this parameter distribution are illustrated in Fig.4.9.63b. In worst case analysis (WC), parameter values are shown at the minimum or maximum end points of their respective ranges so as to obtain extreme values for the circuit performance. For Monte Carlo simulation without correlation (MC 1), values of parameters are selected from their distributions. For Monte Carlo simulation with correlation (MC 2), parameter values are computed from the independent parameters. For the simple case shown in Fig .4.9.63b PARM 1 is considered independent parameter, for example, and PARM 2 is calculated from the values of PARM 1.

## 4.9.4 Results and Discussion

The emitter-coupled logic gate is simulated for the three cases shown in Fig.4.9.63. The first step is to analyze the circuit under worst case conditions to check if the performance remains within desired limits. For the circuit under consideration, when the input is 0 V, the output voltage will have limiting values under following conditions. The minimum

| | $I_s$ | $\beta_{Fm}$ | $R_e$ | $C_2$ | $\beta_{Rm}$ | RC1 | RC2 | REE | $V_{out}$ (mV) | $I_{in}$ (mA) |
|---|---|---|---|---|---|---|---|---|---|---|
| WC 1 | max | min | min | max | min | min | min | max | -1 | 1008 |
| WC 2 | min | max | max | min | max | max | max | min | -620 | 612 |

and maximum values of the parameters are obtained from their measured distributions and the circuit is analyzed. Resulting output voltage and input current are also shown above. The circuit behavior in either case is far from satisfactory, the expected nominal value of the output voltage being -200 mV. The performance values are unreasonable and it may be concluded that the circuit does not behave as desired.

One drawback of performing the worst case analysis is that it only gives limits of performance but it does not give any idea for how many circuits on a wafer will these performance limits be observed. Therefore this gives an overly pessimistic picture of the circuit. In reality, the probability of occurrence of these worst case conditions may be very small.

To get an idea of the distributions of the output quantities, statistical analysis of the circuit needs to be performed. Such an analysis without taking into account parameter correlations (MC 1) is performed and the results are tabulated in Table 4.9.6. Although the spread in values is less than that obtained from worst case analysis, they still do not compare well with the measured results. During this analysis, all device parameters are chosen independently from their respective distributions.

In practice, device parameters will not be completely independent. Hence their interdependence must be taken into account. The reduced

model obtained earlier provides an easy method of accounting for the interaction between model parameters. Here the parameters are expressed in terms of the saturation current $I_s$ using linear regression equations. The value of $I_s$ is chosen from its distribution and remaining parameters are computed using these equations. The correlation between the three epitaxial results is also taken into account in a similar manner. Resistor $R_{EE}$ is chosen independently from its distribution and RC 1 and RC 2 are computed using linear regression equations. The results of this simulation are also summarized in Table 4.9.6. The spreads in the two performance quantities are less than those obtained without correlations (MC 1) and they are much closer to the corresponding measured values.

Thus, following important observations can be made from the results of this simulation. Circuit simulation using nominal or typical values does not provide information about the variation in circuit performance that will be observed when a large number of such circuits are fabricated. Worst case analysis is a simple way of estimating these variations. But the results of this type of analysis are often too pessimistic and may lead to unreasonable conclusions about the circuit behavior. Therefore, statistical analysis is necessary to obtain information about the distribution of circuit performance. The importance of the interdependence of device parameters is illustrated by performing the statistical simulation first without and then with model parameter correlations. It is illustrated that by accounting for this interdependence using simple linear regression, simulation results can be obtained which compare well with the measured circuit performance distribution.

### 4.9.5 Summary

The features necessary in a program for performing statistical circuit simulation are described. A computer program is developed based on these considerations. The program runs on mini-computer and can simulate MOS and bipolar integrated circuits. The program serves the need for statistical simulation to determine the distributions of circuit outputs due to input parameter variations such as component tolerances, aging and temperature effects.

The performance of an example circuit is simulated using this program. The worst case analysis is shown to give large and unrealistic variations in circuit performance. Simulations making use of parameter distributions give substantially better results. By properly considering parameter correlations, a realistic simulation results which agrees with the measured circuit performance. This indicates the usefulness of statistical simulation and the improvements achieved by accounting for the physical correlations between various device parameters.

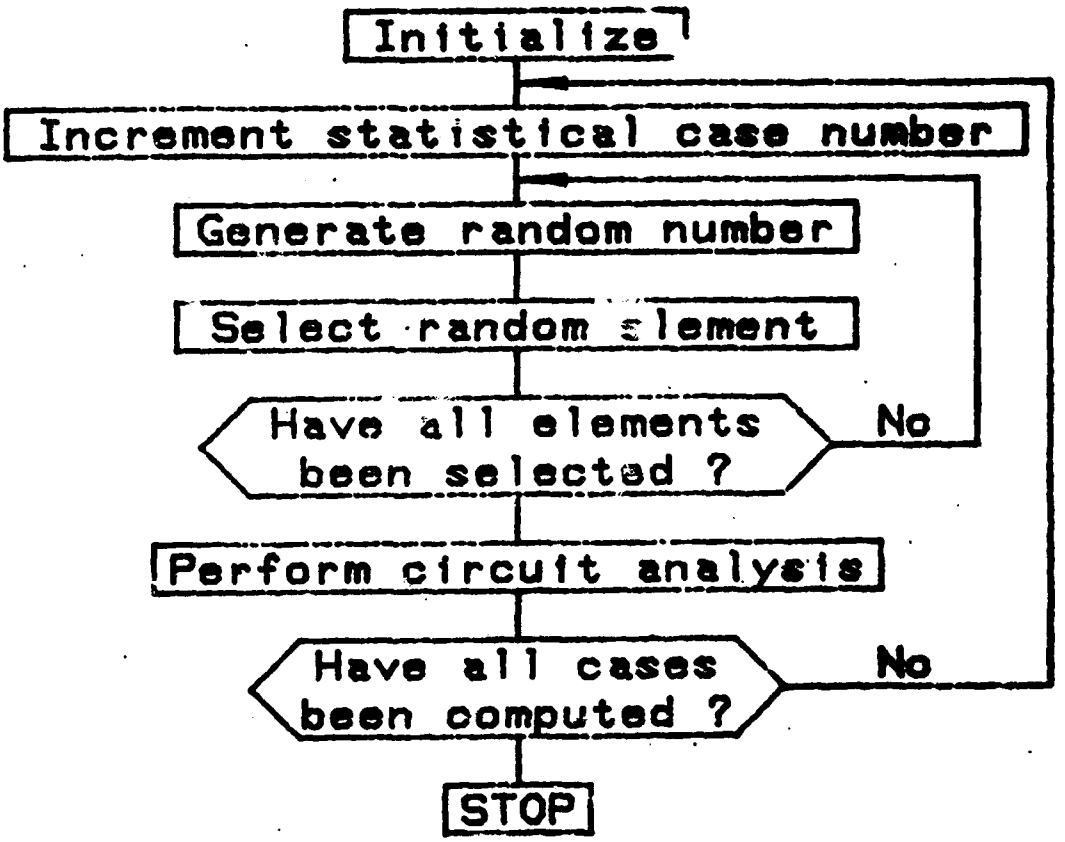



Figure 4.9.57 Flow chart of the Monte Carlo Method

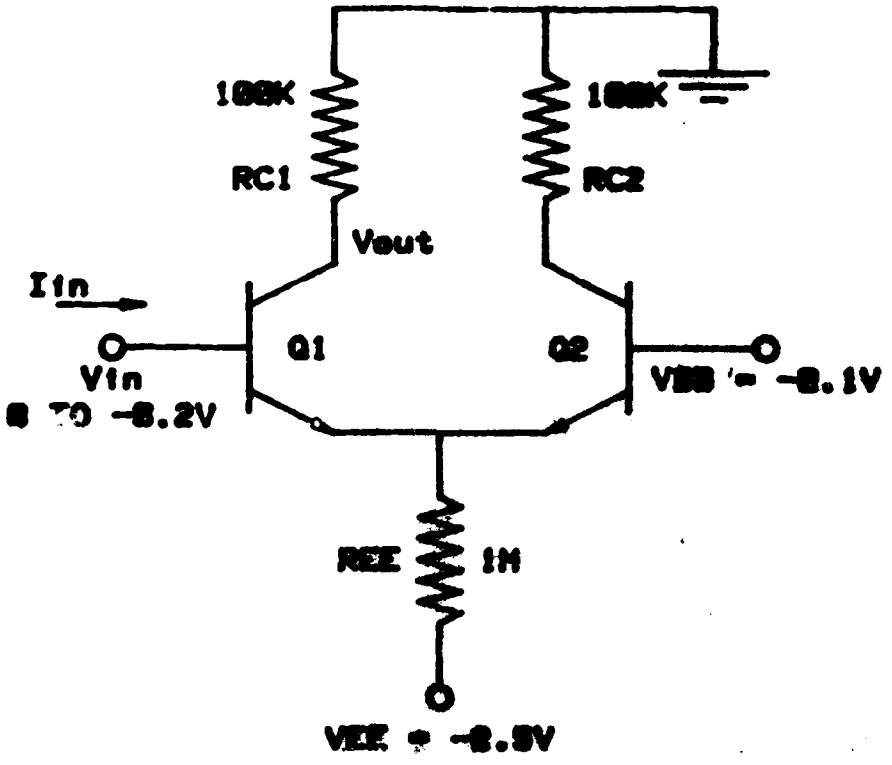



Figure 4.9.38 Single input ECL gate

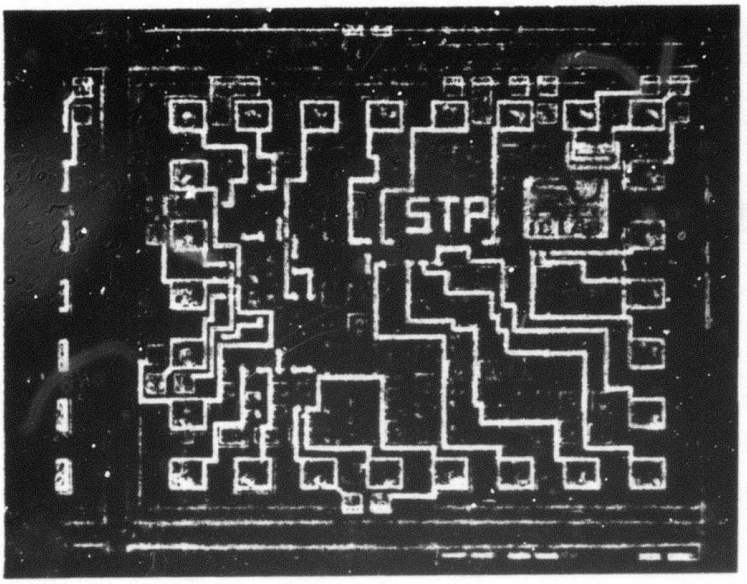



Figure 4.9.59 Photograph of a KITCHIP die.

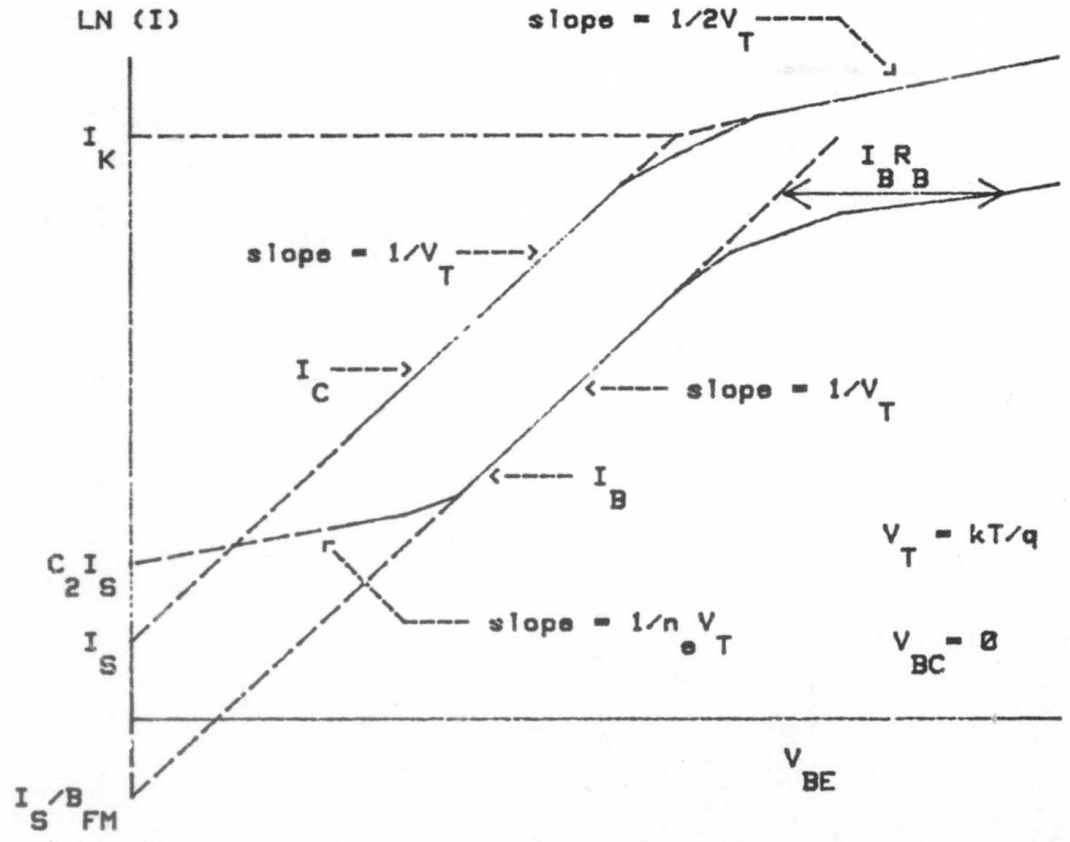



Figure 4.9.61 Gummel-Poon model parameter definitions in terms of collector and base currents as a function of base-emitter voltage.

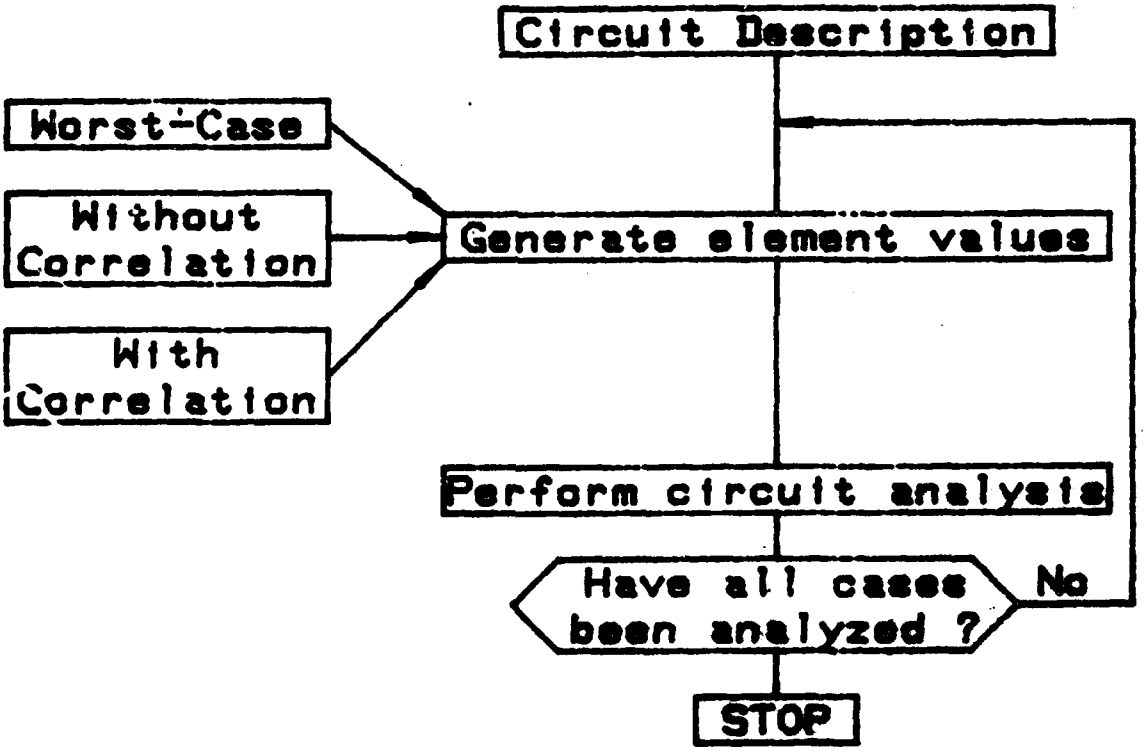
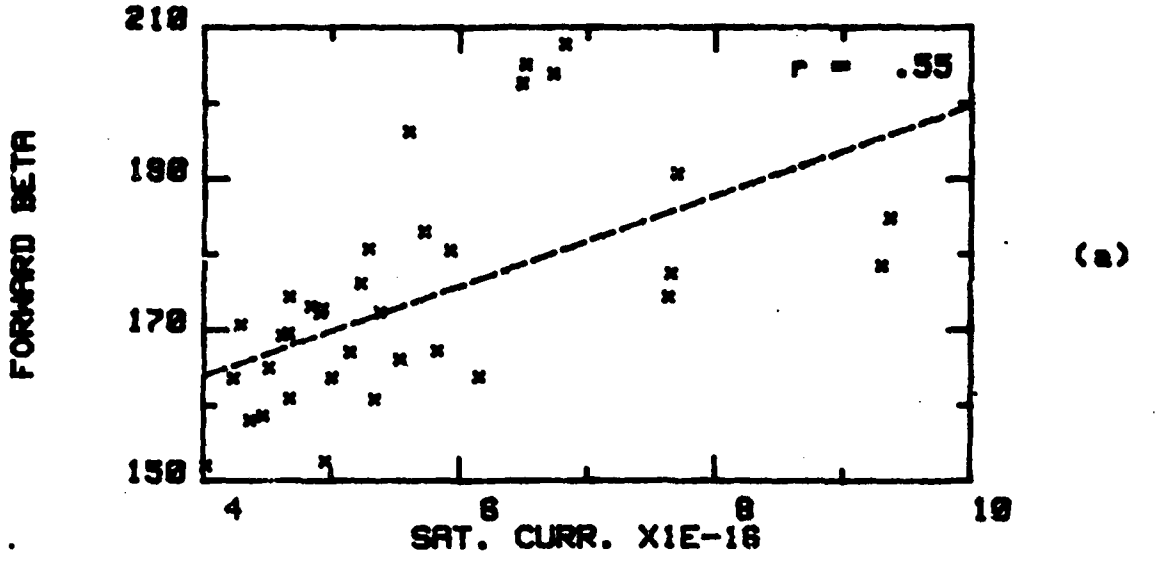



Figure 4.9.62 Flow chart of statistical analysis



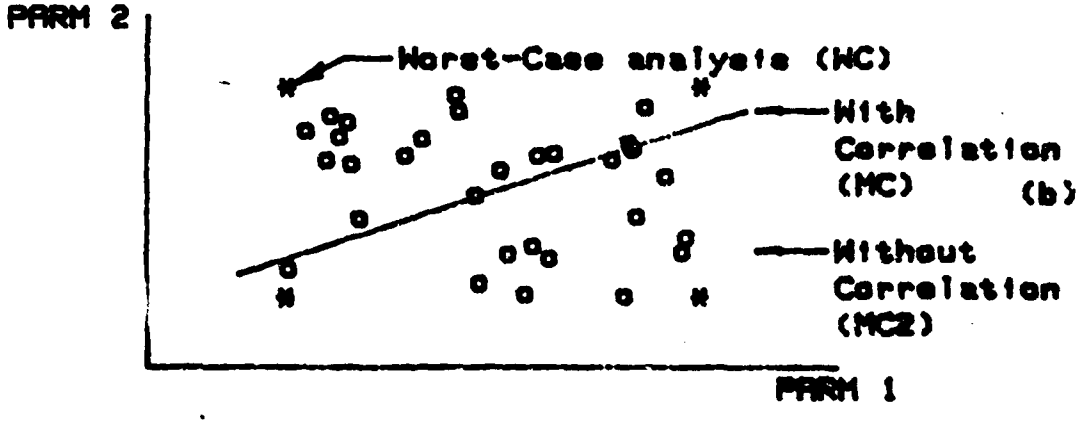




Figure 4.9.63(a) Measured scatter plot of $\beta$ vs. $I_s$
(b) Parameter selection for statistical analysis

**Statistical Analysis With and Without Correlations**

| | $V_{out}$ (mV) | | |
|---|---|---|---|
| | Measured | No Corr. (mc2) | With Corr. (mc) |
| Minimum | -264.4 | -489.72 | -265.84 |
| Maximum | -192.2 | - 80.66 | -167.18 |
| Mean | -236.959 | -214.76 | -218.18 |
| Std.Dev. | 17.164 | 97.22 | 21.017 |

| | $I_{in}$ (nA) | | |
|---|---|---|---|
| | Measured | No Corr. (mc2) | With Corr. (mc) |
| Minimum | 8.793 | 10.0 | 5.822 |
| Maximum | 90.14 | 1036.9 | 122.886 |
| Mean | 24.603 | 287.7 | 68.753 |
| Std.Dev. | 18.121 | 271.8 | 40.6 |

Table 4.9.6 Results of statistical simulation with and without model parameter correlations.

REFERENCES


[4.1] J. D. Meindl, et al, "Interim Report on Computer-Aided Engineering of Semiconductor Integrated Circuits," Stanford Electronics Laboratories, TR No. 4969-3, Feb. 1978.

[4.2] D. A. Antoniadis, S. E. Hansen, and R. W. Dutton, "SUPREM II - A Program for IC Process Modeling and Simulation", Stanford Electronics Laboratory Technical Report, SEL 78-020, June 1978.

[4.3] D. A. Divekar, R. W. Dutton and W. J. McCalla, "DC Statistical Circuit Analysis for Bipolar IC's Using Parameter Correlations -- An Experimental Example, Submitted to IEEE JSSC.

[4.4] J. W. Slotboom, "The pn-product in Silicon", Solid State Electronics, 20, 1977, pp. 279-283.

[4.5] R. J. vanOverstraeten, H. J. De Man and R. P. Mertens, "Transport Equations in Heavy Doped Silicon," IEEE Transactions on Electron Devices, ED-20, March 1978, pp. 290-298.

[4.6] J. W. Slotboom, Analysis of Bipolar Transistors, Ph.D. Dissertation, Eindhoven University, 1977, p. 15.

[4.7] D. B. Estreich, private communication.

[4.8] P. K. Chatterjee, G. W. Taylor, R. L. Easley, H-S Fu, A. F. Tasch, Jr., "A Survey of High Density Dynamic RAM Cell Concepts", IEEE Transactions on Electron Devices, vol. ED-26, pp. 827-839, June 1979.

[4.9] P. K. Chatterjee, G. W. Taylor, A. F. Tasch, Jr., "Charge-Coupled Device Structures for VLSI Memories", IEEE Transactions on Electron Devices, vol. ED-26, pp. 871-881, June 1979.

[4.10] M. S. Mock, "A Two-Dimensional Mathematical Model of the Insulated-Gate Field-Effect Transistor", Solid State Electronics, Vol. 16, pp. 601-609, 1973.

[4.11] J. J. Barnes, R. J. Lomax, "Finite Element Methods in Semiconductor Device Simulation", IEEE Transactions on Electron Devices, Vol. ED-24, pp. 1082-1089, August 1977.

[4.12] A. D. Sutherland, "A Two-Dimensional Model for MOSFET Operation," ECOM 75-1344-F, pp. 216-256, August 31, 1977.

[4.13] J. A. Greenfield, "Analysis and Limitations of Charge-Coupled Devices for VLSI", Thesis - to be published.

[4.14] E. Demoulin, J. J. Barnes, R. W. Dutton, "Bulk Punchthrough Characterization of Submicron Transistor Using Poisson's Solution", submitted to 1979 European Solid State Device Research Conference.

[4.15] R. S. Varga, Matrix Iterative Analysis, Prentice Hall, 1962.

[4.16] M. B. Barron, "Computer Aided Analysis of IGFET Transistors", Stanford Electronics Laboratory, Technical Report No. 5501-1, November 1969.

[4.17] D. Vandorpe, J. Borel, "An Accurate Two-Dimensional Numerical Analysis of the MOS Transistor," Solid State Electronics, Vol. 15, pp. 547-557, 1972.

[4.18] P. E. Cottrell and E. M. Buturla, "Steady State Analysis of Field Effect Transistors via the Finite Element Method", IEDM Technical Digest, pp. 51-54, December 1975.

[4.19] J. J. Barnes and R. J. Lomax, "Two-Dimensional Finite Element Simulation of Semiconductor Devices," Electron Lett., Vol. 10, pp. 341-343, August 8, 1974.

[4.20] R. F. Harrington, "Time-Harmonic Electromagnetic Fields", McGraw-Hill, New York, 1961, p. 106.

[4.21] D. L. Scharfetter and H. K. Gummel, "Large Signal Analysis of a Silicon Read Diode Oscillator," IEEE Trans. Electron Devices, Vol. ED-16, pp. 64-67, January 1969.

[4.22] M. Reiser, "On the Stability of Finite-Difference Schemes in Transient Semiconductor Problems," Comp. Meth. Appl. Mech. Eng., Vol. 2, pp. 65-68, January 1973.

[4.23] J. D. Jackson, "Classical Electrodynamics", John Wiley & Sons, Inc., NY, 1962, pp. 14-20.

[4.24] T. M. Apostol, "Mathematical Analysis", Addison-Wesley, 1957.

[4.25] K. H. Huebner, "The Finite Element Method for Engineers" John Wiley & Sons, Inc. NY 1975, p. 108.

[4.26] G. Artken, "Mathematical Methods of Physicists", Academic Press, New York 1970, pp. 330-344.

[4.27] T. Yoyabe, K. Yamaguchi, S. Asai, and M. S. Mock, "A Two-Dimensional Avalanche Breakdown Model of Submicron MOSFET," IEDM '77 Technical Digest, pp. 432-435.

[4.28] W. Van Roosbroeck, "Theory of Flow of Electrons and Holes in Germanium and other Semiconductors," Bell Sys. Tech. J. 29 (1950), pp. 560-607.

[4.29] D. P. Kennedy, "Computer Aided Engineering of Semiconductor Integrated Circuits", Research and Development Technical Report ECOM-75-1344, August 1977.

[4.30] S. M. Sze, Physics of Semiconductor Devices, John Wiley and Sons, Inc. New York, pp. 430-431, 1969.

[4.31] R. E. Bank and A. H. Sherman, "Algorithmic Aspects of the Multi-level Solution of Finite Element Equations," Report No. CNA-144, Center for Numerical Analysis, The University of Texas at Austin, October 1978.

[4.32] R. W. Hockney, "The Potential Calculation and Some Applications", Methods in Computational Physics, Vol. 9, Plasma Physics, Academic Press, pp. 135-211, 1970.

[4.33] R. E. Jones, "QMESH: A Self-Organizing Mesh Generation Programs", Report No. SLA-73-1088, Applied Mathematics Division, Sandia Laboratories, Albuquerque, New Mexico, July 1974.

[4.34] N. Chan, Ph.D thesis, to be published.

[4.35] H. G. Lee, J. D. Sansbury, R. W. Dutton and J. L. Moll, "Modeling and Measurement of Surface Impurity Profiles of Laterally Diffused Regions," IEEE JSSC Vol. SC-13, 4, August 1978.

[4.36] D. M. Caughey and R. E. Thomas, "Carrier Mobilities in Silicon Empirically Related to Doping and Field," Proc. IEEE Vol. 55, 1967.

[4.37] D. C. D'Avanzo, S. R. Combs and R. W. Dutton, "Effects of the Diffused Impurity Profile on the DC Characteristics of VMOS and DMOS Devices," IEEE JSSC Vol. SC-12, 4, August 1977.

[4.38] D. P. Kennedy and R. R. O'Brien, "Analysis of the Impurity Atom Distribution Near the Diffusion Mask for a Planar p-n Junction," IBM J., Vol. 9, May 1965.

[4.39] D. H. Harper and R. E. Thomas, "Diffusion Current Effects in DMOS Transistors," IEDM Tech. Digest, p. 34, December 1978.

[4.40] E. Khalily, "T.E.C.A.P. - An Automated Characterization System", Technical Report No. 5015-1, Stanford IC Lab., March 1979.

[4.41] D. C. D'Avanzo, Ph.D. Thesis, to be published.

[4.42] D. B. Estreich, Latch-Up and Radiation Integrated Circuit - LURIC: A Test Chip for CMOS Latch-Up Investigation, Sandia Laboratories Report SAND 78-1540, November 1978.

[4.43] C. E. Barnes, et. al, Latch-Up Prevention in CMOS, Sandia Laboratories Report SAND 76-0048, March 1976.

[4.44] R. Sokol, and J. Adams, preprint of paper to be presented at 1979 Nuclear and Space Radiation Effects Conference, Santa Cruz, CA, July 1979.

[4.45] D. E. Ward, R. W. Dutton, "A Charge-Oriented Model for MOS Transistor Capacitances, IEEE JSSC, Vol. SC-13, No.5, October 1978, pp. 703-708.

[4.46] T. Toyabe, K. Yamaguchi, S. Asai and M. S. Mock, "A Numerical Model of Avalanche Breakdown in MOSFET's, IEEE Transactions on Electron Devices, ED-25, July 1978, p. 825.

[4.47] Divekar D. A, McCalla W. J. and Dutton R. W., "Statistical Circuit Circuit Simulation in Computer Aided Design", Asilomar Conference on Circuits, Systems and Computers, Asilomar, California, November 1978.

[4.48] Divekar D. A., "Device Modeling for Statistical Circuit Simulation," Stanford Electronics Laboratories, TR No. 5021-3, July 1978.

[4.49] Divekar D. A. and Dutton R. W., "Model Parameter Correlations in Statistical Circuit Simulation", 21st Midwest Symposium on Circuits and Systems, Iowa State University, Ames, Iowa, September 1978.

[4.50] Calahan D. A., Computer Aided Network Design, McGraw-Hill Publishing Co., 1972.

[4.51] McCalla W. J., Computer Aided Circuit Simulation Techniques, Class Notes EE 392B, Spring 1977, Electrical Engineering Department, Stanford University.

[4.52] Young T. K and Dutton R. W., Mini-MSINC - A Minicomputer Simulator for MOS Circuits with Modular Built-in Model, Stanford Electronics Laboratories, Technical Report No. 5013-1, March 1976.

[4.53] Young T. K. and Dutton R. W., "Mini-MSINC - A Minicomputer Simulator for MOS Circuits with Modular Built-in Model", IEEE J. of Solid State Circuits, Vol. SC-11, No. 5, October 1976, pp. 730-732.

[4.54] Nagel L. W. and Pederson D. O., "SPICE - Simulation Program with Integrated Circuit Emphasis", 19th Midwest Symposium on Circuit Theory, Waterloo, Ontario, April 12, 1973.

[4.55] Knuth D. E., The Art of Computer Programming - Vol. 2, Addison-Wesley Publishing Company, 1969.

[4.56] Ralston A. and Wilf H. S. (Editors), Mathematical Methods for Digital Computer - Vol. II, John Wiley and Sons Inc.

[4.57] Meindl J. D., Micropower Circuits, John Wiley and Sons, Inc, 1969.

[4.58] Combs S. R. and Meindl J. D., "KITCHIP - A Generalized IC for Biomedical System Design", Biotelemetry III (T. B. Fryer, H. A. Miller and H. Sandler, editors), Academic Press, New York, 1976, pp. 91-95.

[4.59] Divekar, D. A., Dutton R. W., and McCalla W. J., "Experimental Study of Gummel-Poon Model Parameter Correlations for Bipolar Junction Transistors", IEEE J. of Solid State Circuits, Vol. SC-12, No. 5, October 1977, pp. 552-559.

## APPENDIX I

## SUPREM USERS LIST

Chris Davis
RCA, Solid State Division

Norman Goldsmith
RCA, David Sarnoff Research Center

Paul Langer
Bell Telephone Labs, Allentown

Wolfgang Fichtner
Bell Telephone Labs, Murray Hill

Alexander Voshchenkov
Bell Telephone Labs, Holmdel

N.J. Schneier
TRW Defense and Space Systems Group

Mike Payne
Prime Computer Inc.

Randy Reitmeyer
ECOM

Tom Leedy
National Bureau of Standards

Dr. William Daughton
Texas Instruments

Michael D. Jack
Hughes Aircraft Co., Carlsbad, CA

Doug Culmer
Hughes Aircraft Co., Newport Beach, CA

Mr. Bill Bandy
National Security Agency

L.A. Doyal or Dr. Charles Gwyn
Sandia Laboratories

Tak Young
Signetics

Jennifer Robinson
Bell Northern Research

Walter Jobke
Control Data Corporation

Brian Biehl
Harry Diamond Labs

Dr. Ali Bahraman
Northrop Corporation

Mr. H. Wong
National C.S.F. Inc.

George Walker
National Semiconductor

Young Park
Intel Corporation

Prof. D. Pederson
University of California at Berkeley
Dept. of Electrical Engineering

Keming W. Yeh
Xerox Corp., El Segundo, CA

Robert Tremain
Xerox Corp., Palo Alto Research Center

Prof. C.A.T. Salama
University of Toronto
Dept. of Electrical Engineering

Hassan Nosrati
Philips Laboratories

Jeff Steinwedel
Solid State Scientific

Ken Harmon
Lawrence Livermore Labs
University of California

John Foggiats
Amdahl Corporation

A. Bhattacharyga
IBM, Essex Junction

Philip Mubley
IBM, East Fishkill

Mike Kump
American Microsystems, Inc.

Dr. Rajinder P. Khosla
Eastman Kodak Co.

Dr. V.A. Twaddell
General Dynamics, Pomona, CA

S.V. Pabbisetty
Motorola Inc.

Harald Skarelven
Central Institute for Industrial
Research (Norway)

David A. Wayne
Dionics Inc.

Daryl Delano
Fairchild Semiconductor
Portland, Maine

Dr. Reda Razouk
Fairchild Semiconductor
Palo Alto, CA

John R. Debolt
General Electric Co.
Corporate Research and Development

James Topich
Case Western Reserve University

Prof. Jeffrey Frey
Cornel University
School of Electrical Engineering

Dr. A. Brunnschweiler
Southampton University
Department of Electronics
(England)

Jim Hayes
Synertek

Wes Lukaszek
Zilog

Ed Drummond
United Computing Systems, Inc.

Prof. P. Antognetti
University of Genoa (Italy)

Prof. M. Ilegems
Ecole Polytechnique Federale
de Lausanne
Institut Interdepartemental de
Microelectronique
(Switzerland)

Monsieur Lacroix
Thomson-C.S.F. (France)

Charles Thompson
Burroughs Corp.
Tech Center, San Diego, CA

W. Fitchner
Institut fur Physikalische
Elektronick
Techniche Universitat Wien
(Austria)

Ken Weiner
Information System Design

Gliceria Roflox
Advanced Micro Design

Dilip Pyne
Kooper Instruments Inc.

Prof. Joel Pereira de Souza
Univ. de Sao Paulo (Brazil)

Prof. Davis
Arizona State University
Dept. of Electrical Engineering

Dr. D.D. Meyer
Martin Marietta

Mohammed I. Elmasry
University of Waterloo
Dept. of Electrical Engineering

Rattan Dhar
Digital Equipment Corp.

Mr. Camus
Centre National D'Etudes des
Telecommunications (France)

Joseph Walas
Teletype Corporation

Gilbert Declerck
Catholic University of Louvain
(Belgium)

The Hailey Company

Prof. A. Ganesan
S.U.N.Y. at Stonybrook
Dept. of Electrical Engineering

R.E. Hayes
University of Colorado at
Boulder

G.Y. Robinson
University of Minnesota
Dept. of Electrical Engineering

P.D. Scovell
I.T.T. Standard Telecommunications
Laboratories (England)

Christopher W. Kapral
G.T.E. Laboratories

Brian Craig
Mitel Corp.

Prof. Francois C. Anceau
Ecole Nationale Superieure D'Informatique
et de Mathematiques Appliquees
(France)

Dr. H.C. Lin
University of Maryland
Electrical Engineering

T.N. Casselman
Honeywell Technology Centre

I. Brown
Standard Telephones and
Cables Limited
(England)

Paul Pellegrini
USAF/RADC

Randy Stout
EMM Semi Inc.

J.P. Spoto
Harris Semiconductor

J.J. Dupcavitch
Naval Research Laboratory

Geoffrey Batchelder
NCR Corporation, Microelectronics
Division

Prof. Ron J. Lomax
Electron Physics Laboratory
Dept. of Electrical and
Computer Eng.
University of Michigan

Grant G. Meyers
Dept. of Electrical Engineering
University of Nebraska-Lincoln

A.J. Giovinazzo
Massachusetts Institute of
Technology
Lincoln Laboratory

Surinder S. Rai
Interdata, Inc.

Ian Getreu
Tektronix

Eve Shen
Rockwell International

Marvin White
Westin Elect. Corp.

Scott W. Owen
Univac

Klocker-Moeller
Elektrizitats (West Germany)

J.J. Luferski
Brown University
Division of Engineering

Dr. Richard Vaughn
School of Electrical Engineering
University of NSW (Australia)

Gerald M. Oleszek
College of Engineering and Computer
Science
University of Colorado

A. Pilipowski
(Australia)

Luke W. Dru
Western Digital Corp.

ELECTRONICS TECHNOLOGY AND DEVICES LABORATORY

MANDATORY CONTRACT DISTRIBUTION LIST

101 Defense Technical Information Center
ATTN: DTIC-TCA
Cameron Station (Bldg 5)
012 Alexandria, VA 22314

203 GIDEP Engineering & Support Dept
TE Section
PO Box 398
001 NORCO, CA 91760

205 Director
Naval Research Laboratory
ATTN: CODE 2627
001 Washington, DC 20375

301 Rome Air Development Center
ATTN: Documents Library (TILD)
001 Griffiss AFB, NY 13441

437 Deputy for Science & Technology
Office, Asst Sec Army (R&D)
001 Washington, DC 20310

438 HQDA (DAMA-ARZ-D/Dr. F. D. Verderame)
001 Washington, DC 20310

482 Director
US Army Materiel Systems Analysis Actv
ATTN: DRXSY-T
001 Aberdeen Proving Ground, MD 21005

563 Commander, DARCOM
ATTN: DRCDE
5001 Eisenhower Avenue
001 Alexandria, VA 22333

564 Cdr, US Army Signals Warfare Lab
ATTN: DELSW-OS
Vint Hill Farms Station
001 Warrenton, VA 22186

705 Advisory Group on Electron Device
201 Varick Street, 9th Floor
002 New York, NY 10014

579 Cdr, PM Concept Analysis Centers
ATTN: DRCPM-CAC
Arlington Hall Station
001 Arlington, VA 22212

602 Cdr, Night Vision & Electro-Optics
ERADCOM
ATTN: DELNV-D
001 Fort Belvoir, VA 22060

603 Cdr, Atmospheric Sciences Lab
ERADCOM
ATTN: DELAS-SY-S
001 White Sands Missile Range, NM 8800

607 Cdr, Harry Diamond Laboratories
ATTN: DELHD-CO, TD (In Turn)
2800 Powder Mill Road
001 Adelphi, MD 20783

609 Cdr, ERADCOM
ATTN: DRDEL-CG, CD, CS (In Turn)
2800 Powder Mill Road
001 Adelphi, MD 20783

612 Cdr, ERADCOM
ATTN: DRDEL-CT
2800 Powder Mill Road
001 Adelphi, MD 20783

680 Commander
US Army Electronics R&D Command
000 Fort Monmouth, NJ 07703

1 DELEW-D
1 DELET-DD
1 DELSD-L (Tech Library)
2 DELSD-L-S (STINFO)
* Originating Office

681 Commander
US Army Communications R&D Command
ATTN: USMC-LNO
001 Fort Monmouth, NJ 07703

ELECTRONICS TECHNOLOGY AND DEVICES LABORATORY

SUPPLEMENTAL CONTRACT DISTRIBUTION LIST

(ELECTIVE)

103 Code R123, Tech Library
DCA Defense Comm Engrg Ctr
1800 Wiehle Ave
001 Reston, VA 22090

104 Defense Communications Agency
Technical Library Center
Code 205 (P. A. Tolovi)
001 Washington, DC 20305

206 Commander
Naval Electronics Laboratory Center
ATTN: Library
001 San Diego, CA 92152

403 Cdr, MICOM
Redstone Scientific Info Center
ATTN: Chief, Document Section
001 Redstone Arsenal, AL 35809

407 Director, Ballistic Missile Defense
Advanced Technology Center
ATTN: ATC-R, PO Box 1500
001 Huntsville, AL 35807

517 Commander
US Army Satellite Communications Agcy
ATTN: DRCPM-SC-3
001 Fort Monmouth, NJ 07703

518 TRI-TAC Office
ATTN: TT-SE
001 Fort Monmouth, NJ 07703

519 Cdr, US Army Avionics Lab
AVRADCOM
ATTN: DAVAA-D
001 Fort Monmouth, NJ 07703

*For Millimeter & Microwave Devices Only

475 Cdr, Harry Diamond Laboratories
ATTN: Library
2800 Powder Mill Road
001 Adelphi, MD 20783

477 Director
US Army Ballistic Research Labs
ATTN: DRXBR-LB
001 Aberdeen Proving Ground, MD 21005

482 Director
US Army Materiel Systems Analysis Actv
ATTN: DRXSY-T, MP (In Turn)
001 Aberdeen Proving Ground, MD 21005

507 Cdr, AVRADCOM
ATTN: DRSAV-E
PO Box 209
001 St Louis, MO 63166

511 Commander, Picatinny Arsenal
ATTN: SARPA-FR-5, -ND-A-4,
-TS-S (In Turn)
001 Dover, NJ 07801

515 Project Manager, REMBASS
ATTN: DRCPM-RBS
001 Fort Monmouth, NJ 07703

520 Project Manager, FIREFINDER
ATTN: DRCPM-FF
C01 Fort Monmouth, NJ 07703

521 Commander
Project Manager, SOTAS
ATTN: DRCPM-STA
001 Fort Monmouth, NJ 07703

531* Cdr, US Army Research Office
ATTN: DRXRO-PH (Dr. Lontz)
DRXRO-IP (In Turn)
PO Box 12211
001 Research Triangle Park, NC 27709

604 Chief
Ofc of Missile Electronic Warfare
Electronic Warfare Lab, ERADCOM
001 White Sands Missile Range, NM 88002

606 Chief
Intel Materiel Dev & Support Ofc
Electronic Warfare Lab, ERADCOM
001 Fort Meade, MD 20755

608 Commander
ARRADCOM
DRDAR-TSB-S
001 Aberdeen Proving Ground, MD 21005

614 Cdr, ERADCOM
ATTN: DRDEL-LL, -SB, -AP (In Turn)
2800 Powder Mill Road
001 Adelphi, MD 20783

617 Cdr, ERADCOM
ATTN: DRDEL-AQ
2800 Powder Mill Road
001 Adelphi, MD 20783

619 Cdr, ERADCOM
ATTN: DRDEL-PA, - ILS, - ED (In Turn)
2800 Powder Mill Road
001 Adelphi, MD 20783

701 MIT - Lincoln Laboratory
ATTN: Library (RM A-082)
PO Box 73
002 Lexington, MA 02173

703 NASA Scientific & Tech Info Facility
Baltimore/Washington Intl Airport
001 PO Box 8757, MD 21240

704 National Bureau of Standards
Bldg 225, Rm A-331
ATTN: Mr. Leedy
001 Washington, DC 20231

707 TACTEC
Batelle Memorial Institute
505 King Avenue
001 Columbus, OH 43201