AD A091969

LEVEL II

INTEGRATED CIRCUITS LABORATORY

STANFORD ELECTRONICS LABORATORIES
DEPARTMENT OF ELECTRICAL ENGINEERING
STANFORD UNIVERSITY · STANFORD, CA 94305

LELAND STANFORD JUNIOR UNIVERSITY
ORGANIZED 1891

TR DXG501 ICL 17-78

COMPUTER-AIDED ENGINEERING OF SEMICONDUCTOR INTEGRATED CIRCUITS

prepared by

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

INTEGRATED CIRCUITS LABORATORY
STANFORD UNIVERSITY
Stanford, Ca. 94305

DTIC
NOV 7 1980
C

July 1980

Final Report for Period 1 January 1979 . 15 February 1980

Distribution Statement
Approved for public release;
distribution unlimited

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

U S Army Electronics Technology & Devices Laboratory

DDC FILE COPY

8011 03 209

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: DELET-TR-79-0257-1

2. GOVT ACCESSION NO.: AD-A092 969

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE *(and Subtitle)*: Models of Fabrication Processes, Devices, Circuits for Computer Aided Design of VLSI (Computer-Aided Engineering of Semiconductor Integrated Circuits)

5. TYPE OF REPORT & PERIOD COVERED: Final Report 1 January 1979-15 February 1980

6. PERFORMING ORG. REPORT NUMBER: 2-DXG501-80

7. AUTHOR(s): R.W. Barton, G. Bronner, L.A. Christel, B.E. Deal, R.W. Dutton, J.F. Gibbons, J.A. Greenfield, C.P. Ho, T.I. Kamins, H.G. Lee, L. Lie, M.M. Mandurah, H.Z. Massoud, J.D. Meindl, F. Mohammadi, S. Mylroie,

8. CONTRACT OR GRANT NUMBER(s): MDA903-79-C-0257

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Integrated Circuits Laboratory
Stanford University
Stanford, CA 94305

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

12. REPORT DATE: June 1980

13. NUMBER OF PAGES: 396

14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*:

15. SECURITY CLASS. *(of this report)*: Unclassified

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 MDA903-79-C-0257, issued by Defense Supply Service - Washington. 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 aided circuit design, computer aided device design, computer aided process design, 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 impeded by inadequate process, device and circuit models that handicap low cost computer aided design. The principal 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 circuits. The basic areas under investigation are (1) thermal oxidation, (2) ion implantation and diffusion,

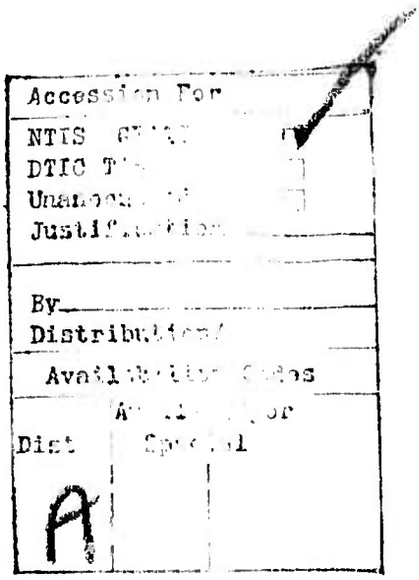
DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

7. Continued

S.Y. Oh, J.D. Plummer, C. Price, R.R. Razouk. J. Rouse,
K.C. Saraswat, S.A. Schwarz, W.A. Tiller, D.E. Ward,
C.R. Helms



20. (Continued)

(3) chemical vapor deposition of silicon and refractory metal silicides, (4) device simulation and analytic measurements. This report discusses the fourth year of the program.

TR DXG501

ICL 17-78

(9) FINAL REPORT. 1 Jan 79-15 Feb 80.

ON

(6) COMPUTER-AIDED ENGINEERING

OF

SEMICONDUCTOR INTEGRATED CIRCUITS.

1 JANUARY 1979 -- 15 FEBRUARY 1980

(12) 40 p.

(10) J.D./Meindl R.W./Dutton J.F./Gibbons
C.R./Helms
J.D./Plummer

This work was supported by
a contract from the

(11) Jul 80

DEFENSE ADVANCED RESEARCH
PROJECTS AGENCY

(15) MDA903-79-C-0257, ARPA Order-3709

ARPA Order No. 3709

(14) ICL-17-78, TR-DXG501-80

(18) DELET

Approved for public release

Distribution unlimited

(19) TR-79-0257-2

James D. Meindl
Stanford Electronics Laboratories
Stanford University Stanford, California

409117

TABLE OF CONTENTS

	Page
Introduction	1
1. THERMAL OXIDATION	11
1.1 Introduction	11
1.2 Summary of Principal Accomplishments to Date	14
1.3 Detailed Discussion of Results	31
1.3A High Pressure Oxidation	31
1.3A1 Experimental Procedure	32
1.3A2 Results and Discussion	34
1.3A3 Conclusions	38
1.3B Thin Oxide Kinetics and Charge Densities	42
1.3B1 Processing Conditions and Equipment for the Characterization of Thin SiO_2 Layers	42
1.3B2 Growth Kinetics	47
1.3B3 Charge Densities in Thin Oxides	50
1.3C Physical Modeling of the Thermal Oxidation Process	53
1.3C1 Species Redistribution	56
1.3C2 Structural Model of Transformation	58
1.3C3 Point Defect Thermodynamics	61
1.3C4 Interstitial Partition Coefficient	62
1.3C5 Summary	65
1.4 Future Work/Conclusions	66
References	69
2. ION IMPLANTATION	113
2.1 Introduction	113
2.2 Major Accomplishments to Date	116
2.3 Detailed Discussion of Accomplishments	119
2.3.1 Extension of the TE Modeling: The Scattering Cross Section	119
2.3.2 Accuracy of TE Modeling Calculations of Primary Ion, Recoil and Damage Distributions	120
2.3.3 Simulation of Implanted Ion Channeling	125
2.3.4 Implant Damage Effects on Thin Oxides	125

Page

2.3.5 Shallow Implanted Junctions 127
2.3.6 High Temperature-Resistant Electrical Contact/Interconnect Metallurgy 129
2.4 Conclusions/Future Plans 134
2.4.1 TE Implementation in SUPREM and Channeling 135
2.4.2 Shallow Implanted Layers and Implant Damage Interactions 135
References 137

3. POLYCRYSTALLINE SILICON 156
3.1 Introduction 156
3.2 Major Accomplishments to Date 157
3.3 Resistivity of Doped LPCVD Poly-Silicon 159
3.3.1 Resistivity of Boron-Doped LPCVD Poly-Silicon 159
3.3.2 Resistivity of Arsenic-Doped LPCVD Poly-Silicon 163
3.3.3 Summary 165
3.4 Arsenic Segregation in Polycrystalline Silicon 165
3.4.1 Theory 167
3.4.2 Sample Preparation 173
3.4.3 Measurement and Interpretation of Hall Constant 174
3.4.4 Results and Discussion 176
3.4.5 Summary 181
3.5 Conclusions 182
References 183

4. REFRACTORY METAL SILICIDES 200
4.1 Introduction 200
4.2 Summary of Major Accomplishments to Date 202
4.3 Silicide Formation Techniques 204
4.3.1 Laser Irradiation of W and Mo Films on Silicon to Form Silicides 205
4.4 Formation of WSi_2 by CVD of W on Si and Subsequent Heating . . 206
4.5 Thermal Oxidation of WSi_2 209
4.5.1 Oxidation Kinetics 209
4.5.2 Mechanism of Oxidation 212
4.6 WSi_2 Gate MOS Devices 218
4.7 Conclusions 221
References 223

Page

5. MATERIALS ANALYSIS/INTERFACE PHYSICS 246

5.1 Introduction 246

5.2 Principal Accomplishments 248

5.2.1 Introduction 248

5.2.2 Summary 249

5.3 Detailed Discussion of Results 250

5.3.1 Sensitivity and Spatial Resolution 251

5.3.2 An Ion Knock-On Mixing Model 253

5.3.3 Redistribution of Phosphorus During the Thermal Oxidation of Silicon 260

5.3.4 Auger Sputter Profiling Studies of SiO_2 Grown in O_2/HCl Mixtures 268

References 273

6. COMPLETE PROCESS AND DEVICE SIMULATION 297

6.1 Introduction 297

6.1A Device Modeling 298

6.1B Analysis Tools 301

6.1C Test Structures 304

6.2 Device Modeling 306

6.2.1 Device Modeling Using the Two-Dimensional Solution of Poisson's Equation 306

6.2.2 A Charge-Oriented Capacitance Model for MOS Transistors 315

6.3 Analysis Tools 325

6.3.1 Two-Dimensional Simulation of One Carrier Steady State Current Transport in MOS Devices 326

6.3.2 A Simplified Two-Dimensional Numerical Analysis of MOS Devices Including Transient Phenomena 333

6.4 Test Structures 344

6.4.1 Test Structures for Extracting Two-Dimensional Impurity Profiles 344

References 353

MODELS OF FABRICATION PROCESSES, DEVICES, AND CIRCUITS FOR COMPUTER AIDED DESIGN OF VLSI

Introduction

In military systems in particular, although to a growing extent in commercial applications as well, there is often a necessity to customize the design of an integrated circuit to fulfill the critical needs of a specific system or class of systems. A major barrier which prevents the economic production of small quantities of high performance custom integrated circuits is the cost of design. The source of this difficulty lies in empirical, and costly, engineering techniques used currently both in integrated circuit design and production. Optimal and efficient system design requires accurate computer models for silicon fabrication processes, elemental device structures, monolithic circuit configurations and system architectures. Current capability to predict changes in system performance resulting from silicon fabrication process modifications is impeded by deficiencies in process, device, circuit and system models.

The overall objectives of this program are:

(1) To develop process models based upon fundamental physical phenomena which accurately predict device structures which would result from a given process sequence.

(2) To integrate these process models in a comprehensive, user-oriented computer simulator (SUPREM) which is generally useful for silicon structures.

(3) To develop methods to utilize geometrical and process inforamtion to specify simple device models including two-dimmensional effects.

(4) To develop circuit analysis capabilities which reflect statistical model effects including parameter correlations.

(5) To disseminate the results to industry.

The results described in this report build upon an established program which has resulted in the release of a second generation process modeling computer program--SUPREM II--which is generally reagarded as the pre-eminent tool of its type available today. It is used by more than 100 industrial firms and university research groups for the simulation and design of state-of-the-art MOS and bipolar fabrication processes. The development of accurate physical models for SUPREM has been a principal objective of this program for the past four years. The work described here extends these models and their implementation in SUPREM to the small geometry device structures which are expected to be used in VLSI circuits in the next several years.

The historical trend in silicon integrated circuit technology has been towards increased complexity and smaller active device dimensions. This has resulted today in lateral device dimensions on the order of 2-3 μm and vertical dimensions well below 1 μm in many commercially available circuits. There are no basic physical mechanisms which prevent a reduction in each of these dimensions by an order of magnitude; such device structures are widely projected for the 1985-1990 time period. Such projections are predicated, however, on our ability to understand physically and accurately model the fabrication techniques which will be used in the construction of such devices.

For devices with relatively large geometries (> 5 μ) and loose processing tolerances, relatively simple models suffice for prediction of vertical impurity profiles resulting from a given fabrication sequence. As device dimensions shrink, however, it becomes essential to employ more robust process models and to consider the interaction both laterally and vertically of various processing steps, if accurate simulation of structures is to be obtained. This is important even with today's 2-3 μm device geometries; it will become essential for smaller devices.

Large geometry devices can be successfully modeled as one-dimensional structures. This is true for both process models and electrical models. Devices with lateral dimensions below a few microns, however, require two-dimensional models for accurate simulation. This need has stimulated a large body of work in recent years on two-dimensional electrical models of device current-voltage characteristics. This work has resulted in remarkable advances in our understanding of small geometry device physics. Recent work under this program in this area is described in Section 6 of this report.

Progress has not been as rapid, however, in two-dimensional process modeling. This is a direct result of our need for improved physical models of oxidation, ion implantation, diffusion, and CVD. It is quite clear that these processes are not one-dimensional. Recent experimental evidence has clearly indicated that oxidation or impurity diffusion in a localized region of a silicon substrate can substantially affect oxidation or diffusion rates in laterally or vertically adjacent regions of the substrate. There is no clear agreement at the present time on the basic physical mechanisms responsible for such results. It is clear, however, that we must quantitatively understand such phenomena if we are to accurately model small device structures.

A specific goal of this program is to understand and model these two-dimensional effects. We have made substantial progress in this regard in the past year. It appears now that the basic physical phenomena underlying these interactions are the roles of point defects--silicon vacancies and interstitials--in impurity diffusion, thermal oxidation, and other processes. The generation and consumption of these point defects during high temperature fabrication steps appear to be the unifying physical effects which can explain many of the phenomena which have been regarded as anomalous to date. We have used such models to quantitatively understand a variety of process phenomena and have incorporated some of these models in SUPREM II. The unifying role of these mechanisms is a cornerstone of the work described in this report. We regard such models as absolutely essential to accurate modeling of two-dimensional effects in small devices.

The most widely used model of the thermal oxidation kinetics of silicon is that provided by Deal and Grove [1], who suggest an initial reaction rate limited linear relationship between oxide thickness and time followed by a transport limited parabolic form. Although this model is useful, it does not include the effects of substrate orientation, multiple species in the oxidation ambient,and substrate doping levels. It does not explain the enhanced kinetics of thin SiO_2 layers, has no predictive capability for oxide charge densities and is one dimensional. A major part of our activity in this area has been aimed at examining the interface reaction which converts silicon to SiO_2. This work has pointed out the important roles of silicon point defects in this reaction [2,3], and has led to a generalized model for thermal oxidation which has quantitatively modeled for the first time the effect of substrate

doping levels on oxidation kinetics [4]. This model has been incorporated in SUPREM II. The role of these point defects in oxidation is fundamental to the work which is described in Section 1 of this report. Such phenomena are the means of coupling oxidation to diffusion and oxide charge densities, and are expected to play important roles in our studies of thin oxide kinetics. In addition, generation of these point defects during oxidation is regarded as the key mechanism in understanding two-dimensional effects of thermal oxidation.

The most widely used model for ion implantation is based upon Linhard, Scharff and Schiott (LSS) range theory [5]. Although the LSS model is indeed useful, the approach fails to account for two important features of ion implantation in integrated circuits: the effects of implanting in multilayer targets as encountered in threshold shifting of MOS transistors and the effects of upper layer "knock-on" ions on device characteristics. Also not modeled is the experimentally observed dopant profile's deep channeling tail that results from ions scattered into relatively unobstructed channels in the crystalline silicon substrate. To overcome these shortcomings, a new ion implantation model based on the Boltzmann transport equation which predicts the flux of incident and "knock-on" ions at a series of closely spaced distances within a multilayer target has been developed [6]. This model has also recently been applied to the prediction of channeling tails on implanted profiles.

Many of the physical phenomena associated with ion implantation and other processes are transient in nature (damage annealing and gettering for example). As a result, physical modeling has been difficult due to the need for obtaining a large number of experimental data at closely spaced time intervals. In the past this has implied a large number of experimental samples. A major effort has been initiated in the past year

to develop a temperature tolerant metallurgy for electrical contact to bulk regions and for MOS structures. Such a metallurgy would make possible in-situ or time dependent monitoring of a single sample and greatly improve our understanding of these transient processes. This work is regarded as vital for developing improved models for many of the processes described in this report. Recent work in ion implantation modeling and in the development of this temperature tolerant metallurgy is described in Section 2 of this report.

The most widely used model for thermal diffusion is based upon Fick's second law or the so-called diffusion equation assuming a constant diffusion coefficient. It is quite clear that this model is inadequate at high dopant levels, or during thermal oxidation, or in the presence of multiple diffusing species. A new multistream diffusion model which describes the interaction of diffusing ions with silicon point defects has been developed and demonstrated substantial improvements in modeling capability [7]. Models involving point defects have also been developed which explain the enhancement of diffusion coefficients during thermal oxidation [8]. The large amount of work in this program related to point defect phenomena is expected to further improve impurity diffusion models.

Current models of dopant inclusion in silicon epitaxial layers are based upon simple steady-state boundary layer concepts [9]. Such approaches cannot account for gas flow transients and their impact on doping, especially in very thin epitaxial layers. A new physiochemical model for dopant inclusion has been developed, based upon a transfer function approach. It achieves accurate prediction of arbitrary epitaxial layer doping profiles through convolution of the reactor transfer characteristics with appropriate gas flow functions [10,11]. Modeling of this type has recently been extended to thin epitaxial layers ($< 2\ \mu$) which will be

required for VLSI devices. Because polysilicon is so widely used in MOS devices, it has become the subject of a significant portion of our CVD activities. The physical structure and thermal stability as well as electrical conductivity of doped films have been studied in layers deposited both at atmospheric and low pressures. Refractory metal silicides hold great promise for reduced interconnect resistances in VLSI structures. For this reason, a detailed study of their physical, electrical and chemical properties is underway with the aim of evaluating their potential for small geometry devices. Recent work in the polysilicon area is described in Section 3, activity in the Refractory Metal Silicides is described in Section 4.

In cooperation with the physical modeling of fabrication processes described above we have also developed advanced materials analysis techniques based upon the use of Auger electron spectroscopy (AES) and secondary ion mass spectrometry (SIMS). These techniques are routinely used to analyze samples and provide experimental inputs to the oxidation, ion implantation and CVD activities. In addition, both AES and SIMS have been used extensively in careful studies of the segregation of impurities in the Si-SiO_2 structure, the atomic nature of the Si-SiO_2 interface and the grain boundary chemistry in polysilicon [12,13]. This work is described in Section 5 of this report.

The new process models for oxidation, ion implantation, diffusion, and CVD outlined above and described in detail in the remainder of this report, are specifically being developed for incorporation in a new version of SUPREM [14]. SUPREM III which is projected for release during the current calander year, will for the first time enable simulation and accurate modeling of many of the technologies associated with small geometry devices. In order to model shallower junctions and thinner

oxides and epitaxial layers, transient kinetics must be included in the process simulator. In addition, most VLSI device structures employ multilayer structures--for example polysilicon layers for diffusion contacts and SiO_2-Si_3N_4 layers both for active devices and for local oxidation masking. SUPREM III will have to be substantially altered to accommodate these capabilities. In addition to these fundamental changes in simulation capability all of the models in SUPREM will be upgraded to incorporate results generated by the oxidation, ion implantation, diffusion, and CVD groups. One final change which will be made in the program is to alter its output format and structure so that the process variables being used at any time by the simulator are accessible. This improvement is expected to be highly useful to the groups developing improved physical models for SUPREM and should allow for closer coordination of basic physical work and SUPREM implementation.

The viability of process modeling for IC design hinges on prediction of device performance. The final section 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) are described. Moreover, the applications of process and device analysis to test structures and circuit simulator model development are presented.

REFERENCES

1. B.E. Deal and A.S. Grove, "General Relationship for the Thermal Oxidation of Silicon," J. Appl. Phys., 36, 3770 (1965).

2. J.D. Meindl et al., "Final Report on Computer-Aided Semiconductor Process Modeling," Stanford Electronics Labs, TR DXG569, July 1979.

3. W.A. Tiller, "On the Kinetics of the Thermal Oxidation of Silicon. Part 1," J. Electrochem. Soc., 127, 625, 1980.

4. 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., 126, 1516-1522, Sept. 1979.

5. J.L. Lindhard, M. Scharff and H.E. Schiott, Mat. Fys. Medd. Dan. Vid. Selsk, 33, No. 14, 1963.

6. D. Smith and J. Gibbons, "Application of the Boltzmann Transport Equation to the Calculation of Range Profiles and Recoild Implantation in Multilayer Media," Proc. of Ion Implantation Conf., Boulder, CO, Aug. 1976, p. 333.

7. A. Chu and J. Gibbons, "A Theoretical Approach to the Calculation of Impurity Profiles for Annealed Ion Implanted B in Si," Ion Implantation Conf., Boulder, CO, Aug. 1976.

8. D.A. Antoniadis, A.G. Gonzalez and R.W. Dutton, "Boron in Near Intrinsic (100) and (111) Silicon Under Inert and Oxidizing Ambients--Diffusion and Segregation," J. Electrochem. Soc., 125, 813 (1978).

9. T.J. Rodgers, "Advanced Integrated Circuit Technology for Micropower ICs," SEL-75-034, Stanford Electronics Labs, Stanford, Calif., pp. 58-98, Aug. 1975.

10. R.Reif, T.I. Kamins and K.C. Saraswat, "A Model for Dopant Incorporation into Growing Si Epitaxial Films: I. Theory," J. Electrochem. Soc., 126, No. 4, April 1979, pp. 644-652.

11. R. Reif, T.I. Kamins and K.C. Saraswat, "A Model for Dopant Incorporation into Growing Si Epitaxial Films. II. Comparison of Theory and Experiments," J. Electrochem. Soc., 126, No. 4, April 1979, pp. 653-660.

12. J.S. Johannessen, W.E. Spicer and Y.E. Strausser, "A Study of the Chemical Composition of MOS and MNOS Structures by Auger Electron Spectroscopy," Thin Solid Films, 32:311 (1976).

13. S.A. Schwarz, R.W. Barton, C.P. Ho and C.R. Helms, "Studies of Phosphorus Pile-up at the Si-SiO_2 Interface Using Auger Sputter Profiling," submitted for publication.

14. D.A. Antoniadis and R.W. Dutton, "Models for Computer Simulation of Complete IC Fabrication Processes," IEEE Trans. Electron Devices, ED-26, No. 4, pp. 490-500 (1979).

Part 1

THERMAL OXIDATION

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

1.1 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 20 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, oxide properties, and interface charge densities is the

---

*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.

continuing decrease in active device 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 10 to 15 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, and presence of a chlorine species during the oxidation, the use of multiple species (such as O_2/H_2O mixtures) during oxidation, ambient pressures greater than or less than one atmosphere, and thin oxides (<500 Å) in which kinetics are often dramatically changed.

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 quanitatively 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 reviewed 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 example of the importance of these goals and the gaps which exist in our present understanding is illustrated in Fig. 1.1. The device shown here is an oxide isolated NMOS silicon gate transistor, similar in configuration to many presently being fabricated in the semiconductor industry. One can identify a number of problem areas, in the context of this structure, in thermal oxidation and related processes.

1. Heavily doped regions (N^+ source, drain and polysilicon gate layer) show experimentally much larger oxidation rates than lightly doped regions.

2. The growth kinetics of thin oxide layers (gate dielectric) are dramatically enhanced over thicker layers due to mechanisms which are not well understood. Si/SiO_2 interface charge properties associated with these thin layers also appear to be quite different from thicker layers.

(4)

3. Orientation effects on oxide growth kinetics and charge densities are very important at the edges of locally oxidized regions. Two dimensional effects dominate these transition regions.

4. Dopant segregation and redistribution effects at the Si/SiO_2 interface largely determine field oxide inversion voltages and can affect active device thresholds under the thin gate dielectric.

5. Enhanced impurity diffusion coefficients under and adjacent to oxidizing regions can dramatically affect dopant profiles and therefore device characteristics. These effects are intimately connected with atomistic processes occurring during oxidation at the Si/SiO_2 interface.

6. Fixed oxide charges (N_f) and interface state charges (D_{it}) 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 process phenomena will permit optimization of present device structures and will minimize the amount of empiricism needed to develop new structures.

1.2 SUMMARY OF PRINCIPLE ACCOMPLISHMENTS TO DATE

Historically, the kinetics of the thermal oxidation of silicon have been described by a linear-parabolic growth law originally derived by Deal and Grove [1.1]. This model recognized the three steps important in silicon oxidations (Fig. 1.2) -- transport of O_2 or H_2O from the gas phase to the SiO_2 surface, diffusion of the O_2 or H_2O through the SiO_2 layer and reaction of the oxidizing species with silicon at the Si/SiO_2

interface. By equating the fluxes involved in these three steps, a growth law given by

$$\frac{X_o^2}{B} + \frac{X_o}{B/A} = t + \tau \qquad (1.1)$$

where X_o = oxide thickness and t = oxidation time results. A, B, and τ 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.

As formulated above, the Deal-Grove kinetic model lumps all physical processes occurring at the Si/SiO_2 interface into an effective rate constant k_s (or B/A). While these parameters may be experimentally measured under various process conditions, the underlying mechanisms responsible for observed values of k_s and B/A are not easily revealed

by this approach. Basic understanding and quantitative prediction of parameters associated with the oxidation process would appear to be achieved more easily by a more detailed physical model. Such a model may be regarded as an extension of the basic linear parabolic growth law with particular attention paid to the atomistic processes occurring at the Si/SiO_2 interface. This has been a major thrust of our activities in this program. We believe that such detailed understanding is essential for accurate prediction of kinetics, charges and other parameters such as enhanced diffusion coefficients associated with the oxidation process.

This "working model" is under continuing development under this program. Its essence was described in last year's annual report [1.2]. Recent work is described in the main body of this report; to provide background for this later discussion, we review here the major points of the model. It should be emphasized that this model is not complete. It is qualitative in many aspects although some specific aspects have been quantified with excellent agreement with experiment. 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 the model is that it shows promise of allowing quantitative prediction of all these phenomena and their inter-relationships. The basis for the model is shown in Fig. 1.3.

O_2 from the gas phase is incorporated into the surface of the SiO_2 layer and is known to diffuse down to the SiO_2 interface where the reaction converting silicon to SiO_2 takes place. The work of Doremus [1.3], Jorgensen [1.4], Rayleigh [1.5], and Collins and Nakayama [1.6] presented

conflicting evidence concerning the presence or absence of an important charged oxygen ion in the diffusion process. The work of Mills and Kroger [1.7] 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 [1.1] seems to indicate that the dominant diffusion species is molecular oxygen. Blanc [1.8] recently proposed a model for the enhanced growth kinetics of thin SiO_2 layers which postulated O_2 as the diffusion species, but with the dissociation of the molecular oxygen into 2O at the SiO_2 interface as an important part of the overall kinetics. 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 are currently using for kinetic experiments 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, dissocation 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 [1.9,1.10]. 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. 1.3. 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.

Two alternative means of providing the necessary free volume at the interface are shown in the middle and bottom of Fig. 1.3 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.

The bottom mechanism of Fig. 1.3 has been postulated by Dobson [1.11, 1.12] and the excess interstitial flows produced during thermal oxidation may be the mechanism of oxidation enhanced dopant diffusion effects [1.11,

1.12,1.13]. In addition, the interstitials produced during the reaction have been linked to oxidation-induced stacking faults (OISF) [1.13]. Both of these oxidation related phenomena are likely due to very small quantities of excess interstitials. It has been estimated by Hu [1.13] that perhaps only 1 in 1000 silicon atoms at the interface ends up as an interstitial in the substrate. This is supported by the fact that within experimental measurement, the oxidation process appears to go to completion, i.e. essentially all silicon atoms at the interface are converted to SiO_2. This implies that under normal conditions most Si_I flow into the SiO_2 layer and react with O atoms to form SiO_2. The relative few that do go into the silicon substrate, however, have profound effects on diffusion coefficients and OISF growth rates. Finally, it might be conjectured at this point that some of the excess interstitials which move into the SiO_2 layer are closely related to the observed oxide charge N_f that has been discussed as incompletely oxidized silicon atoms [1.14].

A substantial amount of work has been reported in the literature which tends to support the general framework in Fig. 1.3. Much of this work is related to growth rates of OISF and to enhanced impurity diffusion coefficients in an oxidizing environment. Stacking faults may be nucleated at the surface of the substrate where there is a lattice mismatch with the growing SiO_2 layer. They also are generally observed to nucleate and grow in the bulk of the substrate. Nuclei for these bulk stacking faults appear to be present in the wafer prior to oxidation and may be due to the crystal growth defects, oxygen precipitates, impurity centers or mechanical or ion implantation damage. Once nucleated, all stacking faults are observed to grow at the same rate, at least up to distances of several tens

of microns away from the oxidizing interface. This growth rate is directly linked to the velocity of oxidation.

Many experiments have been reported in which OISF growth rates have been measured. All of these results are consistent with the role of point defects (Si_I and Si_V) in controlling such growth. If an oxidation is taking place at the Si/SiO_2 interface, with a velocity V = dx/dt, then excess Si_I are generated at a rate which is directly related to this dx/dt. As described above, a small fraction of these interstitials become mobile in the substrate. Such defects are extremely mobile and would be expected to diffuse to large distances away from the surface. Once in the bulk, these Si_I may recombine with available Si_V or they may be absorbed by stacking faults which are present. During their diffusion away from the interface, they would be expected to enhance diffusion coefficients of any impurities present via the interstialcy mechanism.

Various attempts at modeling quantitatively the growth rates of OISF based upon the above ideas have been presented in the literature [1.13, 1.15,1.16] with generally good agreement with experiment. In general, these models demonstrate the following observations.

1. At any given temperature, higher oxidation velocity implies higher OISF growth rates. Low partial pressure oxidations produce lower dx/dt and therefore slower growth. High pressure oxidation produces higher dx/dt and therefore faster growth. It appears that the OISF growth rate depends less than linearly on dx/dt so that a given oxide thickness can be grown with smaller stacking faults if dx/dt is maximized.

2. At higher temperatures, dx/dt is larger because both B and B/A depend exponentially on T. However, it is observed, that as T increases, OISF growth rates first increase, reach a maximum, and then actually decrease, causing the OISF to shrink. This phenomenon is likely related to the fact that at higher temperatures, concentrations of vacancies in the substrate increase. This is illustrated in Fig. 1.4. The activation energy of vacancy formation is not known precisely but is ≈3-4 eV. Since the activation energy of B/A and hence

the rate of excess Si_I generation is ≈2 eV, the general behavior illustrated in the figure would be expected. Thus as T increases, vacancies in the substrate increasingly provide recombination sites for the excess Si_I and finally at high enough temperatures, can be absorbed by the OISF, allowing them to shrink in size.

3. HCl additions to the oxidizing ambient are observed to decrease and in some case stop the growth of OSIF [1.17-1.20]. The mechanism in this case is believed to be bonding of the Cl with the excess Si_I at the interface, preventing their release into the substrate.

4. OISF growth rates are observed to be dramatically higher on oxidizing (100) surfaces than on oxidizing (111) surfaces. This has led to the proposal that some of the excess Si_I are captured by active surface sites (kinks or surface steps) and take part in surface regrowth. The density of such kinks is higher on (111) surfaces [1.13].

5. In heavily doped regions in which Si_I concentrations are much higher (see later discussion) OISF are observed to shrink [1.21,1.22]. This is consistent with the vacancies being absorbed on the stacking faults and therefore causing them to shrink. In experiments in which oxidation of heavily doped regions is taking place, the situation is much more complicated because now the oxidizing interface acts as a sink for vacancies. This may lead to growth or retrogrowth of the OISF depending upon doping level, T, etc.

A large body of data also exists with respect to the influence of oxidation on impurity diffusion coefficients. It has been well established that boron [1.23,1.24], phosphorus [1.25,1.26], and arsenic [1.25] all exhibit enhanced diffusion coefficients under oxidizing conditions. The magnitude of the effect depends upon the impurity, on temperature, and on crystal orientation. The orientation and temperature effects are similar to those observed for OISF growth, leading to the hypothesis that the underlying mechanism is the same in both cases. In general, enhancement of diffusion coefficients is more pronounced at lower temperatures and is larger on (100) oriented silicon.

Si_I generated during oxidation have been used to explain these effects, exactly as in OISF growth. Excess interstitials increase impurity diffusion via the interstialcy or Watkins replacement mechanism [1.27]. This

process consists of a Si_I removing a substitutional impurity to an interstitial site, followed by the impurity creating a new Si_I by jumping on to a neighboring lattice site. The temperature dependence of the enhancement factor has been explained by the same mechanism illustrated in Fig. 1.4. That is, at higher temperatures, the equilibrium concentration of vacancies increases relative to the excess Si_I concentration. This results in more recombination sites for the excess Si_I and also an increase in "normal" diffusion processes which are vacancy related. The net result is that oxidation enhancement effects are less important at the higher temperatures. The orientation effect [(100) > (111)] is likely due to the same phenomenon described above in connection with OISF, i.e. the increased density of surface kinks on (111) silicon surfaces which act as sinks for the Si_I. It is also interesting that HCl oxidations produce smaller enhancements of impurity diffusion coefficients [1.28], again likely because the Cl at the Si/SiO_2 interface chemically bonds with the Si_I, preventing their diffusion into the substrate.

Some interesting work has also been recently reported [1.29] on lateral effects of oxidation enhanced diffusion. In these experiments, locally oxidized surface regions were shown to dramatically enhance boron diffusion coefficients both vertically into the substrate and laterally under Si_3N_4 masked regions. The lateral extent of these effects was small compared to the distances over which vertical effects have been seen. Again assuming that excess Si_I are responsible, effective diffusion lengths of these defects of >25 μ vertically and 2-3 μ laterally were observed. Two interesting conclusions arise from these experiments. The first relates to the difference in lateral and vertical effects. This may be due to

sinks for the excess Si_I at the Si/SiO_2 interface (perhaps the kinks postulated above or perhaps strained regions under the Si_3N_4 layer). The second and more general conclusion is that in small geometry device structures, process parameters (diffusion coefficients, etc.) may be geometry sensitive. That is, lateral diffusion lengths for the Si_I of 2-3 μ are comparable with modern device dimensions, implying that process models for such devices will of necessity be two dimensional.

In addition to the above results, which are largely taken from the literature, substantial work under this program has provided direct experimental and theoretical verification of the ideas expressed in Fig. 1.3. The most complete example of this relates to enhanced oxidation rates over heavily doped regions [1.30,1.31]. Experimental data such as that shown in Fig. 1.5 gathered early in this program demonstrated that such enhanced oxidation rates could be as much as five times over lightly doped rates. The experiments further demonstrated that the enhancement was related to the interface reaction rate k_s (B/A) predominantly. These observations led to the mechanism shown in the middle of Fig. 1.3 (Si_V) being postulated as the cause.

Under the assumption that the vacancy related mechanism acts in parallel with others which likely dominate under lightly doped conditions, we may write

$$\frac{B}{A} = R_1 + k_o C_{VT} \tag{1.2}$$

where R_1 is the mechanism associated with lightly doped oxidation (likely the other two suggested in Fig. 1.3) and C_{VT} is the total vacancy concentration which is a function of temperature and doping level. k_o is an empirically determined constant which effectively expresses the efficiency

of the vacancies in providing reaction sites for the oxidation process.

Since C_{VT} can be explicitly calculated as a function of doping level and T, the expression in Eq. (1.2) can be quantitatively tested with experiment. A large body of data generated under this program and taken from the literature has shown excellent agreement with the predictions of Eq. (1.2). Typical of the results are the data and calculations shown in Fig. 1.6 which show effective values of B/A as a function of T and N^+ doping level. The model and experiment are in excellent agreement. Additional implications of this work are the following.

1. All N type dopants should behave identically since the relevant parameter is the substrate vacancy concentration which depends only on the Fermi level position and T. This has been experimentally verified [1.30,1.31].

2. The enhancement of B/A should be the same for dry O_2 and H_2O oxidations since it is a substrate parameter which is relevant. This has also been experimentally verified [1.30,1.31].

3. The enhancement of B/A in P^+ silicon should be smaller than in N^+ silicon because a smaller total vacancy concentration exists at any given doping level in P^+ silicon. This has also been verified [1.30, 1.31].

4. The enhancement factors for (111) and (100) silicon should be comparable. This also appears to be true [1.30,1.31].

5. To the extent that an electrically active doping concentration and hence Fermi level position can be defined in polysilicon, the enhancement mechanism should be similar. This also appears to be true although growth rates are generally not as enhanced over doped polysilicon layers because of dopant segregation and carrier trapping effects at grain boundaries. These are discussed in more detail in the polycrystalline section of this report.

Based upon the ideas represented in Fig. 1.3, we can also make some interesting observations about the origin of the various oxide charges (Q_f and N_{it}). It seems likely that these are related to the excess Si_I which move into the oxide and as a result represent incompletely oxidized

silicon. As we have seen, the vast majority (>99%) of the Si_I generated during oxidation appear to go into the oxide, while only a small percentage become free and move into the bulk silicon. We have also seen that some of the excess Si_I which go into the bulk may be trapped at surface kinks where they become involved in interface reconstruction. (This mechanism was used to partially explain the difference in (111) and (100) surfaces in oxidation enhanced diffusion and OISF growth.) In either case, the density of such Si_I should be directly related to dx/dt, the oxidation velocity. Based upon this model and referring again to Fig. 1.4, we can begin to understand some of the experimental data which exists on oxide charge densities vs. process parameters. The well known Q_f triangle [1.33], for example, is shown in Fig. 1.7.

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 [1.32]. 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 [1.1]. Van Vechten and Thurmond [1.34] conclude from bulk quenching data that $[Si_V] \alpha e^{-2.4/kT}$, while Seidel and MacRae [1.35] propose that $[Si_V] \alpha e^{-3.4/kT}$. In either case, behavior similar to that shown in Fig. 1.4 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_f 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 as we have seen to account for the growth rate of oxidation induced stacking faults

[1.15]. An intriguing relationship between Q_f 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 [1.15]. This is also the approximate temperature at which the dry O_2 Q_f curve in Fig. 1.7 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 ≃0 because sufficient Si_V are present to eliminate them.

In the case of H_2O, shown in Fig. 1.7, the interface reaction rate is faster at all temperatures (although the interface reaction rate activation energy is still ≃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 ≃0 should be higher. In the case of oxidation-induced stacking faults, this temperature is ≃1350°C [1.15]. In the case of the Q_f triangle, the high temperature behavior of Q_f in H_2O is not well known but the point of intersection of the H_2O curve in Fig. 1.7 with the "minimum" Q_f 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_f curve in Fig. 1.7 has a smaller slope vs. temperature than does the dry O_2 curve. This is in spite of the face that more Si_I are generated per unit time in the H_2O case. One might, therefore, expect a higher Q_f value of H_2O at all temperatures; this is not observed as Fig. 1.7 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_f.

One can predict several other interesting features of Q_f 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 [1.33], the value of Q_f should be lower. Measurement difficulties make this difficult to verify although the fact that OISF growth rates are lower and oxidation enhanced diffusion effects are smaller tend to support the general argument that Q_f should be smaller also. 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_f should be lower. A recent publication confirms this prediction [1.36]. In fact, the inert gas curve in Fig. 1.7 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_f drops to its "minimum" value. (It is well known [1.15] that the growth rate of oxidation-induced stacking faults decreases in low partial pressures of O_2 or H_2O as do oxidation enhanced diffusion effects, 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_f. We are not aware of any measurements of Q_f 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_f) before they are removed from the furnace. Some discussion of charges associated with high pressure oxidation is included later in this report.

One final interesting observation follows from the discussion above. The "minimum" value of Q_f shown in Fig. 1.7 is generally shown as a constant value, independent of temperature [1.32]. 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 of Q_f 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_f caused by trace amounts of O_2 or H_2O in the furnace gases. In fact, there have been a few reports [1.37] of very low Q_f 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_f 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.

In summary, we have reviewed 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 has been placed on the roles of point defects (Si_I and Si_V) 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 at the Si/SiO_2 interface. The attraction to this model is 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. Figure 1.8 summarizes many of the results and phenomena described in this section and illustrates the intimate relationships between the various process phenomena.

The implication of this model for future process models is that point defect concentrations in the substrate will play essential roles in many process parameters. It will thus likely be necessary to calculate the distribution and concentrations of these defects throughout the device structure. These concentrations and distributions will depend upon doping levels, temperature, oxidation velocities and likely device geometry as well. This implies that accurate process modeling will of necessity become two dimensional.

We complete this section with a listing of the principal accomplishments of this program in the past several years. Detailed discussion of the results of the past year are contained in the next section.

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 our first annual report [1.38].

2. Complete experimental characterization of O_2/HCl oxidation kientics for 0-10% HCl and T = 900-1100°C has been completed. These results were also described in our first annual report [1.38] and are the subject of journal publication [1.39].

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 our second annual report [1.40] and are the subject of a journal article [1.41].

4. The kinetics of H_2O and H_2O/HCl mixtures using a pyrogenic system have been studied, and the results were described in our second annual report [1.40] and in a journal publication [1.42].

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 f (phosphorus chemical concentration) have been derived. These results were described in our second annual report [1.40] and are the subject of a journal article [1.43].

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 our second annual report [1.40] and in last year's annual report [1.2] and is the subject of two journal articles [1.30,1.31].

7. The dependence of the oxidation-induced interface state charge density (N_{it}) on process variables has been determined. Process variables investigated included oxidation temperature (800°-1200°C), ambient (O_2, H_2O, O_2/HCl), and annealing/cooling conditions (O_2, N_2, Ar), as well as silicon type and orientation. A direct relationship between interface states and fixed oxide charge (N_f) has been established. The results are described in last year's annual report [1.2] and a journal publication [1.44].

8. The effect of low temperature hydrogen anneal parameters on interface state densities has been investigated. The results indicate that the best annealing condition for premetallization anneal is at 450°C for 10 min or at lower temperatures for longer times in a 100% hydrogen ambient. In the case of post-metallization anneals, the best parameters are 10 to 15 min, 10 to 25% hydrogen in nitrogen at 400°C. Results were presented in last year's annual report [1.2], and a journal paper is being prepared.

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

10. 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.

11. 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.

12. Experimental measurements of the physical and optical properties of thin, thermally grown SiO_2 layers have been completed. These were described in detail in last year's annual report [1.2].

13. Initial experimental measurements of the growth kinetics of thin (0-500 Å) oxide layers were described in last year's annual report [1.2]. More detailed measurements are described later in this report.

14. 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 was described in detail in last year's annual report [1.2] and was reviewed earlier in this report.

1.3 DETAILED DISCUSSION OF RESULTS

A. High Pressure Oxidation

The thermal oxidation of silicon in high pressure dry O_2 or steam may have a significant impact on IC device fabrication in the near future. Although high pressure oxidation of silicon was carried out almost 20 years ago by Ligenza [1.45], the availability of commercial high pressure oxidation units such as Gasonics' "Hipox" [1.46] has only recently permitted the use of high pressure oxidation in high volume integrated circuit fabrication.

One of the keys to VLSI circuit development seems to be low temperature processing. This is the result of constraints being imposed by device scaling in the continuing effort to achieve higher packing densities and better performance. Dealing with some of these constraints requires reducing dopant redistribution and oxidation induced defects. High pressure thermal oxidation of silicon allows the growth of films of the order of 1 μm on (100) substrates at 800°C in times less than 2.5 hours. The low oxidation temperature possible with high pressure leads to substantial reductions in dopant redistribution as well as reductions in oxidation induced stacking faults [1.47,1.48]. Some disadvantages may also

result, such as increased lateral oxidation in oxide-isolated structures fabricated by selective oxidation [1.49].

One of the major goals of the present program is the development of an advanced process modeling capability. This requires the investigation of the latest technological advances and their incorporation in the present process modeling program. For high pressure oxidation, the kinetics of oxide growth in dry O_2 and pyrogenic steam ambients need characterization for both (100) and (111) oriented silicon. The variables under investigation are oxidation temperature, ambient, time, and pressure. Also of importance is the density of oxide charges resulting from the oxidation and differences, if any, in charge density between oxidation at 1 atm and at high pressure.

During the past year, a high pressure oxidation system was purchased from Gasonics and installed at Fairchild. The system was modified by including booster pumps in order to increase the pressure of the needed gases (originating from liquid sources) to the 1000-1200 psi required for system operation. Mass flow meters were installed to monitor gas flow and other modifications carried out to allow more complete system monitoring and ease of operation. Kinetics of oxide growth at pressures of 5, 10, 15, and 20 atm in pyrogenic steam were characterized for temperatures in the range of 800° to 1000°C. Fixed oxide charge, interface state, and mobile ionic charge densities were measured for the various temperatures and pressures and compared to typical charge densities in 1 atm oxides.

A1. Experimental Procedure

P-type (100) and (111) silicon wafers of resistivity 5-9 Ω-cm and 1-3 Ω-cm respectively were used in these experiments. The wafers

were cleaned in hot sulfuric, aqua regia, and 10:1 HF/DI with appropriate DI rinses. After cleaning, the wafers were spin dried in a "Corotek" spin/dryer and loaded into the furnace in nitrogen. A cross section of the high pressure oxidation system is shown in Fig. 1.9 and shows the various components of the system. An injector baffle is placed between the wafers and the injector and helps in maintaining uniformity of oxide thickness between the front and back of the quartz boat. A second baffle is placed in front of the stainless steel cap and helps reduce the temperature at the cap in order to prevent deterioration of the sealing O-ring.

A typical pressurization cycle is shown in Fig. 1.10. Pressure in the tube increases almost linearly with time and fluctuations in pressure are about 3-6 psi. The oxidation procedure used in this work is summarized below.

1. Wafers are loaded in nitrogen at temperature.
2. Loading door is closed and system activated.
3. System is purged for 5 min in dry O_2.
4. System is purged for 5 min in steam.
5. Pressurization start with the wafers in the steam ambient.
6. Wafers are oxidized for the preselected time.
7. Depressurization starts. The steam flow terminates and is replaced by nitrogen flow. Depressurization is complete in 3-4 minutes.
8. The wafers are purged in nitrogen for an additional 7 min and pulled from the furnace in N_2 (pull time approximately 2 min).

Oxide thickness was measured by a Gaertner L-116 Automatic Ellipsometer and for thick oxides the thickness range was verified by an additional measurement on a Nanometrics "NanoSpec/AFT" microspectrophotometer.

Wafers to be used for oxide charge measurements were metallized following oxidation with Al-4% Cu-2% Si by cold flash evaporation and annealed in a 10% H_2 in N_2 ambient at 400°C for 10 minutes. High frequency and quasistatic C-V plots were used to calculate N_f and D_{it} and bias stress measurements performed to obtain mobile ionic charge density N_m.

A2. Results and Discussion

The work carried out in this part of the program can be divided into three parts:

1. Characterization of the system and oxidation growth for both (100) and (111) silicon at 800°, 850°, 900°, 950°, and 1000°C in pyrogenic steam at pressures of 5, 10, 15, and 20 atm.

2. Determination of the rate constants for high pressure steam oxidation and the relationship between B, B/A and pressure. This will have a direct bearing on the incorporation of high pressure oxidation into Stanford's process modeling program SUPREM.

3. Determination of the effect of high pressure oxidation on oxide charges.

(a) System Characterization and Oxide Growth Data. P-type (100) and (111) silicon was used in obtaining the oxide growth data summarized in Figs. 1.11-1.13 for oxidations at 5, 10, and 20 atm. The oxidation time plotted in the figures includes the ramp-up time which is shorter at 5 atm and longer at 20 atm. The effect of ramp-up time will be discussed later. The faster oxidation of (111) oriented silicon compared to (100) is evident and maintained at all pressures tested. As temperature and pressure increase the parabolic rate constant dominates and the orientation differences are less pronounced.

(b) Rate Constant Determination. The general oxidation relationship [1.1] for the case of no initial oxide can be written as:

$$x_o - B\left(\frac{t}{x_o}\right) = -A \tag{1.3}$$

and a plot of x_o vs. t/x_o can then be used to obtain the rate constants B and B/A. In obtaining the rate constants from the high pressure oxidation data, it is necessary to account for the initial pressurization time. During that time pressure is increasing from 1 atm to the desired pressure resulting in a continuous change in the oxidation rate constants. This problem is avoided by carrying out an "initial run" which is long enough to allow the system to pressurize and stabilize at pressure. The oxide thickness obtained from the "initial run" will be the initial oxide thickness (x_i) and the general relationship can then be expressed as:

$$x_o^2 + Ax_o = B(t + \tau) = Bt + x_i^2 + Ax_i \tag{1.4}$$

which can be written as

$$x_o^2 - x_i^2 + A(x_o - x_i) = Bt \tag{1.5}$$

or

$$(x_o + x_i) + A = B\left(\frac{t}{x_o - x_i}\right) \tag{1.6}$$

A plot of $(x_o + x_i)$ vs. $(t/x_o - x_i)$ is shown in Fig. 1.14 and both the parabolic rate constant (B) and the linear rate constant B/A can be extracted.

It should be noted that this technique results in values of B and B/A based upon relatively thick oxides. This is due to a limit in the minimum time allowed for the system to pressurize and stabilize coupled with the fast oxidation rates in steam ambients. Rate constants obtained

by this technique indicated that both B and B/A are directly proportional to pressure with constant activation energies. The change in activation energy for temperatures below 900°C previously observed (1.42) was not evident.

One of the important goals of this program is the incorporation of high pressure into the process modeling program SUPREM. Towards that end equations were developed based upon the general oxidation relationship to allow for the calculation of oxide thickness given oxidation temperature, pressure, and an oxidation time which include a ramp-up period. The ramp rate chosen was that determined experimentally and is constant at all pressures and linear with time. The oxidation cycle was divided in five intervals as shown in Fig. 1.15 with the oxide thickness at the end of each interval identified (x_1 - x_f). The equations describing each interval are:

$$x_1 \simeq 0 \tag{1.7}$$

$$x_2 = \frac{1}{2}\left[-A + \sqrt{[A^2 + 4B'(t_{x_2} - t_{x_1})]}\right] \tag{1.8}$$

$$x_3 = \frac{1}{2}\left[-A + \sqrt{A^2 + 2B'(t_{x_3} - t_{x_2})(P + 1) + 4x_2^2 + 4Ax_2}\right] \tag{1.9}$$

$$x_4 = \frac{1}{2}\left[-A + \sqrt{A^2 + 4B'P(t_{x_4} - t_{x_3}) + 4x_3^2 + 4Ax_3}\right] \tag{1.10}$$

$$x_f = \frac{1}{2}\left[-A + \sqrt{A^2 + 2B'(t_{x_f} - t_{x_4})(P + 1) + 4x_4^2 + 4Ax_4}\right] \tag{1.11}$$

where the parabolic rate constant B = B'P and B' = rate constant at 1 atm. Results obtained by use of Eqs. (1.7)-(1.11) are in very good agreement with experimental values for oxidation temperatures greater than 900°C. For lower oxidation temperature an over-estimation of oxide thickness occurs if a constant activation energy is assumed for B and B/A. Equations (1.7)-(1.11) were therefore used to calculate the rate constant that would yield the best estimate of oxide thickness over the entire temperature range. The calculation was done by obtaining the locus of B/P and B/AP points that for a given oxidation will yield the experimentally measured oxide thickness. The intercept of the loci for various oxide thicknesses yields the values of B/P and B/AP that would, when used with this model, result in an oxide thickness estimate within $\pm$5% of the experimentally obtained value. An example of this method is shown in Fig. 1.16 where for each oxide thickness a band is shown representing variations of $\pm$2%.

Rate constants obtained are summarized in Figs. 1.17 and 1.18 and include B/P and B/AP values obtained from data at 1, 5, and 10 atm. These results show that the rate constants are proportional to pressure and indicate the presence of a second activation energy at 900°C and below for both rate constants. For B the activation energy increases from 18 Kcal/mole above 900°C to approximately 27 Kcal/mole while for B/A the activation energy decreases from 47 Kcal/mole to about 38 Kcal/mole. The reasons for this change are being investigated at this time. The difference between these results and those obtained by $(x_o + x_i)$ vs. $(t/x_o - x_i)$ plots is the incorporation of the short oxidation time in the calculation therefore extending the data to thinner oxides. This has a significant effect particularly at 800° and 850°C.

(c) Oxide Charges. High frequency and quasistatic C-V measurements were carried out on P-type (100) and (111) silicon MOS structures and indicated that both fixed oxide charge (N_f) and interface state densities (N_{it}) are similar in high pressure and 1 atm oxides grown at a given temperature and annealed in situ in N_2 following oxidation. Bias stress measurements indicated that mobile ionic charge densities (N_m) are compatible with standard MOS processes.

(d) Other Results and Technical Accomplishments. Unless otherwise specified, most P-type (111) material available commercially has a slice misalignment of 4° from (111) toward (110) parallel to the flat. The effect of this misalignment on oxide charges was investigated and a 2-13% variation in fixed oxide charge and interface state density observed with the (111) +4° silicon yielding lower oxide charges as expected.

Other activities included the determination of metal-semiconductor work function differences for the incorporation in C-V measurement programs and the effect of cooling procedures on oxide charges.

A3. Conclusions

1. A high pressure oxidation system purchased from Gasonics was installed, tested, and activated. System modifications including the installation of booster pumps and mass flowmeters were completed.

2. Kinetics of oxide growth at pressures up to 20 atm in pyrogenic steam were characterized for temperature in the range of 800° to 1000°C. The pressure dependence of the rate constants was determined and the general relationship used to calculate oxide thickness for a given wafer orientation, oxidation temperature, oxidation pressure, and oxidation time (including pressure ramp-up time).

3. Oxide charges in pyrogenic steam SiO_2 were measured and found to be compatible with present technological requirements.

B. Thin Oxide Kinetics and Charge Densities

Very thin (<500 Å) layers of thermally grown SiO_2 are certain to

become essential components of VLSI circuits. In fact some commercially available, high performance MOS memory circuits currently use gate oxides as thin as 400 Å. There is a wide body of experimental evidence which indicates that such thin layers of SiO_2 are substantially different from thicker layers. These differences include growth kinetics, optical properties, breakdown characteristics, and physical structure. Present versions of SUPREM model such layers by employing approximate empirical correction factors to the general relationship expressed in Eq. (1.1). It is thus of great importance to study these layers and their properties in order to improve the predictive capabilities of SUPREM 3.

The atomic structure and optical properties of thin oxides were studied experimentally during the past contract year [1.2], using transmission electron microscopy, X-ray diffraction, and incremental ellipsometry. TEM showed that the interface is at least partially crystallized in an α-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. These studies are described in detail in our last annual report [1.2].

Some experimental data does exist in the literature on the growth kinetics of thin SiO_2 layers [1.50-1.57]. One of the first detailed studies was reported by Van der Meulen and Ghez [1.50,1.51]. Their work

involved oxidations between 700°C and 1000°C in dry O_2 and in O_2/N_2 mixtures with the O_2 partial pressure as low as 0.01 atm. They found that both (111) and (100) orientations exhibited essentially linear growth at least down to ≈50 Å in thickness. The pressure dependence of the growth rate, however, varied from close to $p^{0.5}$ at 700°C up to close to $p^{1.0}$ at temperatures of 1000°C and higher. This result is in contradiction to the predictions of the linear-parabolic growth law [Eq. (1.1)] which predicts a $p^{1.0}$ pressure dependence under all conditions.

In an effort to explain their experimental results, Van der Meulen and Ghez proposed that the interface reaction is actually more complex than that proposed by Deal and Grove. They suggested that three reactions occur, given by

$$\text{Si-Si} + O_2 \rightarrow \text{Si-O-Si} + O \tag{1.12}$$

$$\text{Si-Si} + O \rightarrow \text{Si-O-Si} \tag{1.13}$$

$$O_2 \rightleftarrows 2O \tag{1.14}$$

While the first two reaction are not stoichiometric in SiO_2 as written they are intended simply to indicate that both molecular and atomic oxygen are assumed capable of reacting with the silicon to form SiO_2. Based upon their kinetic data, Van der Meulen and Ghez determined rate constants for the first reaction (k_1) and for the combination of reactions (1.13) and (1.14) (k_3). The corresponding activation energies were found to be 1.91 eV for k_1 and 0.58 eV for k_3. Because of this difference in activation energies, k_1 would be expected to dominate at higher temperatures and

therefore a $p^{1.0}$ pressure dependence and the normal Deal-Grove relationship should apply at these temperatures. At lower temperatures, k_3 should dominate, leading to a $p^{0.5}$ pressure dependence and a growth law with different effective rate constants. Thus they concluded that oxidation by molecular oxygen was dominant at high T while the atomic oxygen reaction becomes increasingly important as T decreases. The decrease in activation energy for the interface reaction at lower temperatures is consistent with our recent experimental data as well (see Fig. 1.18).

Recently Blanc [1.57] has proposed a kinetic model for thin oxides which in essence uses only Eqs. (1.13) and (1.14). That is, he assumes that the reaction involving molecular oxygen [Eq. (1.12)] does not occur at all. Based upon this, he derives a growth law given by:

$$\frac{1}{2} u + \frac{1}{4} (\exp 2u - 1) = bt \qquad (1.15)$$

where $ax = \sinh u$ and a and b are rate constants. For thin oxides, this predicts a linear parabolic growth law similar to Eq. (1.1), but with different effective rate constants. Blanc has compared his theoretical predictions with the data of Irene [1.58] at temperatures of 780°-980°C, with apparently good agreement. As Blanc points out, however, his model predicts a $p^{0.5}$ pressure dependence for the interface reaction rate; this would seem to be in disagreement with the data of Van der Meulen and Ghez.

Other models have been proposed to account for the enhanced growth rates of thin oxide layers. Grove [1.59], for example, proposed that molecular oxygen is ionized at the outer SiO_2 surface, forming $O_2^=$. Coupled diffusion between the $O_2^=$ and holes then effectively increases the $O_2^=$ diffusion rate until an oxide thickness on the order of the extrinsic

Debeye Length is reached (≈150 Å for dry O_2 and ≈5 Å for H_2O). This model might also be expected to be at least partially successful in explaining the experimental results of Jorgensen [1.4] in which externally applied electric fields were found to affect oxidation rates. An extension of this model was proposed in last year's annual report [1.2] in which it was pointed out that the concept of $O_2^=$ diffusion in thin oxides is not inconsistent with the role of point defects (Si_V) in the oxidation process.

Accurate modeling of the thin oxide regime is most hindered today by the lack of a comprehensive body of experimental data. If such data did exist, it would then likely be possible to differentiate between the various proposed models or to define a new model. A major effort in this program has, therefore, been aimed at obtaining such a body of data. This is a continuing effort which is likely to show its most significant outputs in the next several months. Much of the basic work, however, has been completed and is described here, along with some preliminary results on oxidation kinetics and charge densities in thin SiO_2 layers.

B1. Processing Conditions and Equipment for the Characterization of Thin SiO_2 Layers

Historically one of the principle problems in gathering reproducible data on thin oxide kinetics has been the fact that growth rates have been shown to be dependent on surface preparation techniques. The most dramatic example of this is the data of Schwettman et al. [1.60] which showed that oxide growth rates could be substantially altered simply by changing the chemical cleaning procedure used prior to oxidation.

In an effort to avoid these problems, we have installed an oxidation furnace with several special capabilities. A schematic

of the furnace arrangement is shown in Fig. 1.19. Basically a double-walled quartz tube is used in the furnace with the provision for Ar, O_2 and HCl ambients in the region between two tubes. This permits constant or periodic cleaning and gettering of the environment around the inner quartz tube in which the wafers are placed. The inner tube has provisions for Ar, O_2, low partial pressure O_2, N_2 and HCl ambients, in any combination. This will permit cleaning of the inner tube, growth of SiO_2 layers in O_2, O_2/HCl and O_2/Ar mixtures as well as inert gas annealing in N_2 or Ar. An additional feature is the ability to in-situ clean the silicon wafers in an Ar/HCl ambient prior to thermal oxidation. Such a provision will be shown below to remove any traces of pre-oxidation cleaning treatments, since it etches off 40-50 Å of silicon from the surface, without causing any defects.

Initial experiments designed to characterize the effectiveness of this new furnace were carried out as follows. Phosphorus-doped (100)-oriented (1-4 Ω-cm) wafers were cleaned, dried in isopropyl alcohol vapor, and oxidized at 1000°C according to the following sequence:

1. 15 min pre-heating.
2. Pre-oxidation etch 1% HCl in Ar, 20 min.
3. Pre-oxidation argon purge (0-180 min).
4. Oxidation in dry O_2 for 45 min.
5. Fast pull.

Steps 2 and/or 3 were omitted for some of the wafers. Measured oxide thicknesses obtained from high frequency capacitance-voltage plots are shown in Fig. 1.20. The horizontal axis in this figure is the Ar purge time (step 3) following the in-situ HCl/Ar etch.

From this data, it is seen that pre-oxidation purges of 15 min or longer result in the same oxide thickness. Therefore a 15 min period of argon purging is required to flush from the furnace traces of the HCl used in the pre-oxidation etch. These traces of HCl in the ambient enhance the oxidation rate and result in the higher oxide thickness obtained on the sample that received no argon purge.

Wafers which did not receive any pre-oxidation etch ranged in thickness from 421 to 439 Å and it is believed that variations in the cleaning procedure resulted in this scatter in the data. It can also be noticed that the thicknesses obtained in this case were smaller than those obtained with the pre-oxidation etch, suggesting that the cleaning leaves a surface film that retards the oxidation reaction slightly.

Oxide fixed charge density N_f measured on these samples was $1.05 \pm 0.05 \times 10^{11}$ cm^{-2} for the pre-oxidation etched samples while the wafers which were not pre-oxidation etches ranged from 1.05×10^{11} to 1.60×10^{11} cm^{-2}. These results demonstrate the importance of using proper in-situ cleaning and subsequent purging to obtain reproducible and representative oxide growth kinetic data as well as charge characteristics. We believe that this furnace configuration will allow, for the first time, accurate characterization of thin oxide kinetics.

Another important question with regard to these experiments is the choice of N_2 or Ar as the "inert" gas to be used along with the O_2 oxidant. In many experiments, such an inert gas is required for the following operations:

1. Pre-oxidation annealing or warming up.

2. During the oxidation in oxygen partial pressure.

3. Post-oxidation annealing of oxide fixed charge.

Argon and nitrogen are widely used for these purposes. A body of experimental results support the following observations [1.61-1.63]:

1. Pre-oxidation annealing of freshly-etched silicon wafers in nitrogen for 15 min at 900°C results in a film of ≈20 Å thick (as determined by ellipsometry). It has been suggested that a nitridation process occurs which results in a barrier to subsequent oxidation [1.61].

2. No such oxidation barrier is found in the case of pre-oxidation annealing in argon where the wafers are found to oxidize identically to freshly-etched wafers [1.61].

3. Nitrogen diffuses to the oxide-silicon interface where it chemically reacts with Si at elevated T. ESCA, etch rate, and oxidation-barrier data suggest that nitrogen is chemically-bound in the dielectric layer near the interface either as oxynitride or as a nitride mixed with oxide [1.61].

4. Exposure of wafers pre- or post-oxidation to nitrogen has also been shown to adversely affect the dielectric strength of thermal oxides [1.62]. Substituting argon for nitrogen prevented oxide deterioration.

5. Post-oxidation annealing of oxide fixed charge (N_f) in nitrogen is characterized by an initial charge reduction asymptotically to the minimum detectable level (1×10^{10} cm^{-2}), followed by a dramatic increase. This increase occurs at an annealing time that is a function of the oxide thickness (the time at which this charge increase starts is monotonically proportional to the thickness). No such oxide fixed charge increase is observed when the oxide is post-oxidation annealed in argon. The same trends were also observed for interface trap density N_{it} [1.63].

These observations are based both upon experiments performed under this program at Stanford and upon the literature references cited. Taken together, they imply that the use of N_2 should be avoided and Ar substituted, if the growth kinetics and charge densities of thin oxides are to be studied successfully. This is the procedure we have adopted.

In investigating the growth kinetics and interface charges of thin SiO_2 layers, second order effects become important. These include the variation of oxide thickness and interface charge density with the following parameters.

1. Transient Effects: The time required for the wafers to reach the oxidation temperature becomes comparable to the time needed to oxidize the silicon to ultrathin thicknesses. This time is affected by whether the wafers are pushed in argon (and left until they reach the oxidation temperature and then switched to oxygen) or pushed in oxygen. It is also affected by all considerations that affect the thermal time constant of the wafers (e.g. wafers flat or standing up on the boat). Simple comparative experiments designed to show basic differences indicated that the wafers should be pushed, standing up, in argon. After a warm-up period of about 15 min, O_2 or the oxidizing ambient could be introduced. To investigate the intrinsic oxide fixed charge density, the wafers should be pulled as fast as possible from the furnace and removed immediately from the oxidation quartz boat. This results in the lowest measured values of N_f as compared to cases where the wafers are pulled to the tube entrance and then left there to cool or to the case where the wafers are pulled fast but were left to cool on the boat.

2. Metallization-Alloying Effects: Simultaneously oxidized (100)-oriented phosphorus-doped (1-4 Ω-cm) silicon wafers were metallized in an E-beam metallization system and a flash evaporation system. The flash evaporation system was used to avoid the radiation damage that occurs in the E-beam evaporation system. The wafers were post-metallization-lithography annealed at 420°C for 40 min in nitrogen. The oxide thicknesses obtained from high frequency C-V measurements were 1005 Å and 1025 Å for the E-beam and the flash systems respectively. This indicates that the alloying reaction of SiO_2 with E-beam aluminum consumes 20 Å more than the flash evaporation case. Since we are interested in studying ultrathin layers of SiO_2, the flash system wil be adopted as the metallization technique. Additional attractive features of this system include the absence of radiation damage and the lower value of interface trap density when compared with E-beam-metallized samples.

3. Wafer Drying: No noticeable differences in oxide thickness, oxide fixed charge density or midgap interface trap density between wafers that were blown dry with a nitrogen gun or dried used boiling isopropyl alcohol vapors were found. The isopropyl alcohol technique was subsequently adopted for its additional attractive feature which is that the wafers do not come in contact with filter paper during the drying process.

B2. Growth Kinetics

Phosphorus- and boron-doped, (100)- and (111)-oriented silicon wafers with resistivities between 1 and 10 Ω-cm were cleaned using the following procedure.

1. Soap, ultrasonic agitation - rinse in DI H_2O.
2. H_2SO_4:H_2O_2 (1:1) - 65°C 15 min - rinse in DI H_2O.
3. NH_4OH:H_2O_2:H_2O (1:1:5) - 65°C 10 min - rinse in DI H_2O.
4. HCl:H_2O_2:H_2O (1:1:5) - 65°C 10 min - rinse in DI H_2O.
5. H_2O:HF (50:1) - 30 sec.
6. Rinse in DI H_2O.
7. Isopropyl alcohol dry 3 min.

The wafers were subjected to the following high-temperature treatment (1000°C):

1. Twenty min heating in argon.
2. Twenty min 1% HCl:argon pre-oxidation etch.
3. Twenty min pre-oxidation purge in argon.
4. Oxidation in dry O_2 to thicknesses of 0-1750 Å.
5. Fast pull in O_2 and immediate quench.

The wafers were then metallized in a cold flash evaporation system on the front side. The back-side oxide was then removed using buffered oxide etch followed by metallization on the back side for good contact during the high frequency C-V measurement. After the capacitor dot lithography, a post-metallization low-temperature anneal was done at 420°C for 40 min in N_2.

No differences were observed in the growth kinetics on N-type and P-type samples whether (100)- or (111)-oriented. Figures 1.21 (100) and 1.22 (111) show the oxidation rate obtained experimentally. Also shown is the oxidation rate dx/dt as calculated from the Deal-Grove general oxidation relationship at 1000°C. The enhanced growth rate for thicknesses less than about 500 Å is apparent in the figures.

Table 1.1 shows the values of oxidation rate dx/dt on (100) and (111) samples, the predicted Deal-Grove corresponding rates, the ratio of enhancement over the Deal-Grove rate for each orientation and the ratio of the rates on (111) and (100). It is apparent that both orientations show substantial enhancement factors. It is also interesting that the enhancement factor is larger for (111) than it is for (100). This is in agreement with other data obtained for the two orientations [1.38]. In fact, the limiting value for the ratio of the two growth rates at x_{ox} = 200 Å (1.718) is very close to reported ratios of B/A for the two orientations (1.7) [1.38]. The decrease in the growth rate ratio for thicker oxides is a reflection of the increasing onset of the parabolic regime of oxidation and is to be expected since B is independent of substrate orientation.

The body of data which exists at the present is not sufficient to draw conclusions about the physical mechanisms involved in the growth kinetics of thin SiO_2 layers. Data at other temperatures (700-1000°C) and at oxidant pressures less than 1 atm is presently being gathered and will be used during the present year to arrive at definite conclusions regarding these kinetics.

TABLE 1.1

	(100)			(111)			
x_{ox} (Å)	$\dot{x}_{ox}$ (Å/min)	D-G $\dot{x}_{ox}$ (Å/min)	$\frac{\dot{x}_{ox}}{\dot{x}_{ox}^{(D-G)}}$	$\dot{x}_{ox}$ (Å/min)	D-G $\dot{x}_{ox}$ (Å/min)	$\frac{\dot{x}_{ox}}{\dot{x}_{ox}^{(D-G)}}$	$\frac{\dot{x}_{ox}\ (111)}{\dot{x}_{ox}\ (100)}$
2000	2.8	2.80	1.000	3.30	3.3	1.000	1.178
1000	4.1	4.10	1.000	5.40	5.3	1.018	1.317
500	5.9	5.25	1.123	9.00	7.4	1.216	1.525
400	6.8	5.60	1.214	10.80	8.1	1.333	1.588
300	8.0	6.00	1.333	13.25	8.9	1.488	1.656
200	9.6	5.60	1.416	16.50	9.8	1.683	1.718

B3. Charge Densities in Thin Oxides

Charge characterization at the Si/SiO_2 interface is accomplished by measuring both the high-frequency and quasi-static capacitance vs. voltage of MOS capacitors as shown in Fig. 1.23. The expression for the flat-band voltage of such a structure can be written as:

$$V_{fb} = \phi_{MS} - \frac{qN_f}{C_{ox}} - \frac{1}{C_{ox}} \int_0^{x_{ox}} \frac{x}{x_{ox}} \rho(x)dx + \frac{q}{C_{ox}} \int_{\Psi_{s(v)}}^{\Psi_{s(fb)}} \cdot N_{it}^a (\Psi_s)d\Psi_s - \frac{q}{C_{ox}} \int_{\Psi_{s(fb)}}^{\Psi_{s(c)}} N_{it}^d (\Psi_s)d\Psi_s \quad (1.16)$$

which includes the contributions of metal-semiconductor work function differences (ϕ_{MS}), the oxide fixed charge (N_f), any charge distribution $\rho(x)$ in the bulk of the oxide and the charge on acceptor (N_{it}^a) and donor (N_{it}^d) interface traps.

Under the assumptions of:

1. No contamination in the cleaning, oxidation or metallization procedures [$\rho(x) = 0$].

2. Proper post-metallization annealing so that the contribution of interface traps in negligible, Eq. (1.15) reduces to:

$$V_{fb} = \phi_{MS} - \frac{qN_f}{C_{ox}}$$

$$= \phi_{MS} - \left(\frac{qN_f}{\varepsilon_{ox}}\right) x_{ox} \quad (1.17)$$

which results in a straight line plot of V_{fb} vs. x_{ox} in which the intercept is ϕ_{MS} and the slope is proportional to N_f.

From Eq. (1.17) it is seen that whereas an error of 50 mV in measuring V_{fb} or in the value of ϕ_{MS} results in an error of 1.08×10^{10} in the value of N_f for a 1000 Å oxide, it results in an error of 5.4×10^{10} in the value of N_f if the oxide is 200 Å. Therefore, if the maximum allowable error in N_f for a 200 Å oxide is set to $1\sim2 \times 10^{10}$ cm^{-2}, the flat-band voltage V_{fb} and the value of ϕ_{MS} must be accurate to within 15 to 25 mV.

The same matrix of samples used for the 1000°C growth kinetics characterization was used in determination of ϕ_{MS}. Figures 1.24-1.27 show V_{fb} vs. x_{ox} for (100), (111), N-type and P-type samples. Values of ϕ_{MS} in the (100) N-type case agree with a previous matrix of samples oxidized at 900°C in our laboratory, and either pulled fast in oxygen or annealed at 1200°C in argon for 30 min with the purpose of reducing N_f below measurement resolution. This is shown in Fig. 1.28.

The values of ϕ_{MS} obtained in the different cases are shown in Fig. 1.29, where ϕ_{MS} values given by Deal, Snow and Mead [1.64] and by Werner [1.65] are also plotted for comparison. Our data agrees well with the published results of [1.64] for (111) samples but disagrees with the published (100) data of [1.65]. It should be pointed out that from theoretical considerations, the difference between ϕ_{MS} for N and P type samples of a given orientation and doping should be $2kT/q[\ln(N/n_i)]$. Our data and that of [1.64] follow this prediction, while the data of [1.65] does not.

It should also be pointed out that an ongoing part of our research is aimed at explaining the measured differences between (100) and (111) ϕ_{MS} values. Theoretical predictions for this difference are

not quantitative and there appears to be some indication in our recent work that this difference may in fact be closely connected with the different oxide charge densities on the two orientations. Resolution of this question must be achieved before accurate statements about N_f values on thin oxides can be made, because, as we have seen, errors of only 25-50 mV in the value of ϕ_{MS} can have significant impact on inferred values of N_f.

Nevertheless, to the extent that the above experimental values of ϕ_{MS} are correct, we have extracted values of N_f for the 1000°C oxides described above. Typical results are shown in Figs. 1.30 and 1.31. In the (100) samples there appears to be a distinct increase in N_f as oxide thickness decreases. Such a trend is not apparent in the (111) samples; in fact the opposite may be true, but this conclusion must be regarded with caution at this time because of uncertainty in the exact ϕ_{MS} value. It should also be re-emphasized that these N_f values are for as-grown oxides which did not receive any high temperature Ar anneal.

The data of Fig. 1.30 is in agreement with the predictions of the "working model" described earlier in this report. In the discussion of that model, it was proposed that oxide charge densities should be related directly to dx/dt and hence to the generation rate of excess Si_I. The apparent relationship seen in Fig. 1.30 is replotted in Fig. 1.32 in which it is seen that the higher dx/dt of the thinner oxides is apparently directly related to N_f densities. The data at this point is too sparse and too many questions about ϕ_{MS} remain to draw any definite conclusions. In particular, the very different behavior of (111) oriented samples in Fig. 1.31 is unexplained at this time. Many of these questions should be

resolved during the present year as a larger body of data becomes available. The ultimate goal of this work is, of course, to develop accurate physical models for thin oxide kinetics and charges and to implement these models in SUPREM.

C. Physical Modeling of the Thermal Oxidation Process

In addition to the experimental work in connection with high pressure oxides and thin oxide kinetics and charges described above, an ongoing portion of our work has been aimed at theoretically modeling the thermal oxidation process. This work has provided the framework for the working model described earlier and on a continuing basis provides a perspective on physically probable processes. Three main foci of attention that have benefited from our recent efforts in this area are:

1. The analysis of point defect and solute (P) redistribution in the silicon as a result of oxidation at velocity V = constant in the presence of an interface field, $\Delta G^0(x)$.

2. The development of a structural model of the transformation of crystalline Si to vitreous SiO_2 via the intermediate state of an α-cristobalite SiO_2 layer containing Si and SiO_x interstitials.

3. The initial development of a corona discharge technique for altering both the electric field in SiO_2 and the state of the oxygen species in the SiO_2 environment during the thermal oxidation of Si wafers.

The first two items are an essential part of our atomistic level modeling program. It leads to a detailed understanding of the orientation and temperature dependences of oxidation-enhanced diffusion (OED), oxidation-induced stacking faults (OISF), fixed oxide charge (Q_f), interface trap density (N_{it}), interface stoichiometry (SiO_x), enhanced interface density as well as the actual oxidation velocity (V). The third term arose as a prediction from theory concerning the electrical nature

of the oxidizing species and it has led to a strongly enhanced oxidation rate as expected.

An atomistic level modeling on the kinetics of the thermal oxidation of silicon for the last few years has been extremely fruitful [1.9, 1.10,1.66]. This work has provided a perspective overview of the partitioning of the total driving force for the oxidation process into its component key parallel subprocesses, illustrated in Fig. 1.33, and has delineated the boundary value problems needed to quantitatively connect these components [1.9]. An overall frame work has been given for viewing all aspects of this process and, as such, constitutes a new and generalized description of oxidation with specific attention to the oxidation of Si. The free volume supply condition required to sustain such a transformation was explicitly given and shown to be perhaps the key condition determining the mechanistic pathways involved [1.9]. Because of this free volume condition, attention was focused on vacancy and interstitial formation in the substrate as well as on viscous flow in the oxide.

The first success of this new postulate was the qualitative explanation of the enhanced oxidation rate as a result of heavy doping in the oxide [1.9]. The movement of the Fermi energy away from mid-band at oxidizing temperatures because of the heavy doping led to the creation of a high population of ionized vacancies such that the total vacancy concentration increased strongly with doping level. The higher vacancy concentration produced a readily accessible source of free volume and thus a higher oxidation rate. A one-to-one correspondence was observed between the enhanced vacancy production due to doping and the enhanced oxidation rate as described earlier in this report [1.30,1.31].

By making a budget balance of the total thermodynamic driving force for the oxidation reaction, it was possible to show that the key features limiting growth of the oxide was not the transport of neutral oxygen species through the oxide, as most workers had previously assumed, but rather the excess free energy storage in the system at the interface due to strain [1.10]. It was recognized that ionized oxygen species as well as neutral oxygen species are involved in the oxidation process and that a crossed grid of extra half planes of Si terminated at the silicon side of the Si/SiO_2 interface. Propagation of the interface into the Si-phase required continuous removal of these extra atoms via one of several mechanisms. Based upon this new theoretical understanding, several procedures for enhancing the oxidation rate were proposed, one of which was the corona charging method.

The next step was to postulate a unique structural transformation between the Si and the SiO_2 at the interface [1.66]. Rather than a direct transformation to vitreous silica, which would have produced a large strain energy, it was proposed that the silicon transforms to α-crystobalite plus four interstitial Si defects per unit cell of the silicon transformed [1.66]. This was predicted to be the lowest energy path for the transformation. These interstitial Si then partitioned themselves between the two phases Si and SiO_2 for thermodynamic and kinetic reasons; i.e., a fraction of them leaked back into the Si. Those that flowed back into the Si were found to qualitatively account for the OED via a vacancy and interstitial "wind" effect and for the OISF via a Si interstitial supersaturation effect. Those that flowed into the SiO_2 qualitatively accounted for the Q_f and the N_{it}. These mechanisms were discussed in detail earlier in this report.

Because most of the Si_I defects flowed into the SiO_2, this accounted for the observed interface chemical stiochiometry of SiO_x where x is slightly larger than unity. Because the Si_I defects in the SiO_2 become progressively oxidized and place swelling stresses on the SiO_2, some enhanced density effects will arise as a result of incomplete stress relaxation. In addition, some residual charge (Q_f) remains because of incomplete oxidation within the interface region. Finally, this internal oxidation and the resulting swelling stresses quickly distorts the structure away from the α-cristobalite form to the vitreous form [1.66].

C1. Species Redistribution

Both qualitative and quantitative analyses of an interface field, $\Delta G^o(x)$, on solute redistribution during a phase transformation at constant velocity, V, have been conducted. The interface field may be an electric field, a stress field, a structural disorder field, etc., and it may be either short range or of long range. A rigorous mathematical analysis of the steady state distribution for a two-phase system transforming at constant velocity, V, and with an exponentially varying interface field, $\Delta G^o(x) = \beta e^{-\alpha x}$ (x refers to distance away from the interface in each phase) has been made [1.67]. For dilute binary solutions, the interface distribution coefficient for the solute, k_i, is given by:

$$k_i = k_o \exp[\Delta(\delta G^o(0))/kT] \tag{1.18a}$$

$$\Delta[\delta G^o(0)] = [(\Delta G_S^{01} - \Delta G_L^{01}) - \frac{v^{01}}{v^{02}} (\Delta G_S^{02} - \Delta G_L^{02})] \tag{1.18b}$$

In Eq. (1.18), k_o is the phase diagram partition coefficient for the solute between the two phases S and L, the superscripts 1 and 2 refer to solvent

and solute respectively and v^o is the specific volume in the standard state.

In order to consider the case of actual phase transformation at some velocity, V, a net interface distribution coefficient, k_i^*, is defined as

$$k_i^* = C_S(-5\alpha_S^{-1})/C_L(5\alpha_L^{-1}) \tag{1.19}$$

where $5\alpha_S^{-1}$ is the distance to the edge of the interface field in the new phase, S, and $5\alpha_L^{-1}$ is the distance to the edge of the interface field in the old phase, L. For a variety of numerical cases, k_i^* has been evaluated as a function of k_i and V. One expects k_i^* to be orientation dependent since it depends upon the magnitude and extent of the interface field (orientation dependent) and also upon the ratio of solute diffusion coefficients D_S and D_L to the interface velocity, V. In Fig. 1.34, a plot of $k_i^*/k_o'(k_c' \equiv k_o)$ for an ideal solution, as a function of $V/\alpha_S D_S$ is given for several values of β_S/k' under the condition where $\beta_L = 0$ and $\alpha_S = 10^4\alpha_L = 10^6\ cm^{-1}$. This illustrates the remarkable effect of interface adsorption ($\beta_S < 0$) and interface desorption ($\beta_S < 0$) upon the ultimate incorporation of solute in the new phase (the S-phase).

Extensive attempts to obtain a rigorous initial transient distribution for the solute redistribution process in the presence of an exponential interface field have not yet been completely successful. This is essential for the application of such concepts to the thermal oxidation of Si since the constant velocity regime is short-lived in that case. However, an approximate theoretical model for phosphorus redistribution during thermal oxidation for an interface field, which is a square well

potential of some width d and length U at the interface (and is zero at larger distances), has been developed. Theory predicts that the interface region will be 90% saturated with P after 90 minutes of oxidation at 900°C and after 420 minutes at 800°C. At earlier times (10 minutes at 900°C), the theory predicts a smaller pile-up and, just away from the interface proper on the Si side, a depletion of P. All of these results are in accordance with the interface field (or interface segregation) concept and were confirmed by Auger concentration profile measurements. These results are discussed in detail in Section 5 of this report.

Within the limits set by the Auger technique, which tends to broaden the peak of the concentration profile, estimates of the P pile-up were 25 A° < d < 47 A°. The surface segregation of phosphorus, m = exp(Q/kT), was experimentally found to be defined by the limits 0.22 eV < Q < 0.52 eV and 9 < m < 170 while the limits on U were 0.2 eV < U < 5.2 eV.

C2. Structural Model of Transformation

The SiO_2 molecular volume increases from β-quartz (38.8 $A°^3$) to α-cristobalite (43.6 $A°^3$), β-cristobalite (45.3 $A°^3$) and then to tridymite (52.0 $A°^3$) compared to the atomic volume of Si(20.0 $A°^3$) and vitreous silica (44.5 $A°^3$). Transformation of the Si phase to the SiO_2 phases requires massive reconstruction except for the cristobalites and the vitreous phase. Direct intrusion of O atoms between each Si-Si bond in the Si lattice creates the β-cristobalite lattice under a compressive strain of 31%. Similar direct formation of vitreous SiO_2 from the Si lattice results in an amorphous film under a 30% compressive strain. However, the direct formation of α-cristobalite seems to offer a

potentially much lower value of ΔG_E.

Consider the (100) face of the Si unit cell and the (100) face of the tetragonal α-cristobalite unit cell whose projects are shown in Fig. 1.35. In the Si unit cell we have eight atoms. By completely removing the 4-Si atoms at the 1/4 and 3/4 positions and slightly displacing the other 4-Si atoms followed by oxidation of these four remaining Si atoms, we obtain the α-cristobalite unit cell. This produces a biaxial tension in the film parallel to the interface of ~9% and no strain in the perpendicular direction because the unit cell can expand freely in this direction. This would yield a strain energy <9% of that generated by forming the β-cristobalite directly. However, we still have the four removed Si atoms per unit cell to place back into the system in interstitial positions and this will move the system towards compressive strain and will increase the excess free energy. Placing these interstitials into the Si and the SiO_2 at the interface will stretch the respective unit cells to varying degrees depending on the degree of partitioning of these four extra Si atoms between the two phases. Thus, if the majority of these extra Si atoms enter the SiO_2, we can expect relaxation to occur at the interface such that some extra half-planes on the Si-side will terminate at the interface (see Fig. 1.36). The number of these planes per unit area of interface will depend upon the partitioning ratio of the Si interstitials between the two phases. Since the dangling bonds at the interface ends of these half-planes constitute interface electronic states, this is probably the source of N_{it}. We thus see that the magnitude of N_{it} depends critically upon the interstitial partitioning ratio at the interface. In addition, we note in Fig. 1.35 that, at 45° to the

β-cristobalite unit cell, the structure is pseudo diamond-cubic with a lattice spacing slightly smaller than that of β-cristobalite. Thus, the proposed transformation leads to a relatively well matched diamond-cubic phase that is then stressed in compression by the interstitials placed therein. This leads to an increased density of the interfacial silicate and to an interface silicate stochiometry of SiO_x where $x \tilde{>} 1$. If the β-cristobalite phase or the vitreous phase had formed directly from the Si, it would not have been possible to obtain this interface stochiometry or excess interfacial density.

The key point of this concept is that the SiO_2 form a suitable structure with extra atoms being placed in interstitial positions rather than it form a structure without interstitials. The former choice is the energetically more favorable one by ~100 Kcal/mole. The second point concerning this concept relates to the actual structure that the interface SiO_2 takes. There is only a slight preference for α-cristobalite over β-cristobalite since both produce extra half-planes of Si at the interface that must end up in interstitial positions. There is less reason to choose vitreous SiO_2 because this requires some additional disordering mechanism and there is no rationale for choosing which Si atoms are destined to become interstitial species. With this model a mechanism exists for the rapid transition of the crystalline SiO_2 phase to the vitreous SiO_2 phase within a few molecular distances of the interface. The crystalline SiO_2 phase thus exists only as a buffer layer between the Si and the vitreous SiO_2 but it plays a crucial role in the overall transformation process.

C3. Point Defect Thermodynamics

Neglecting any electric field or strain field effects, the chemical potential of defect species i, i = vacancy (V) or interstitial (I), in each phase j = Si or SiO_2 to be

$$\mu_{ij} = kT \ln \left[\frac{n_{ij}(n_{sij} - n^*_{ij})}{n^*_{ij}(n_{sij} - n_{ij})} \right] \quad \begin{matrix} i = V,I \\ J = Si,SiO_2 \end{matrix} \tag{1.20}$$

Here, n_i is the number of i-defects per unit cell, n_{si} is the number of possible i-defect sites per unit cell and n^*_j is the equilibrium number of i defects given by

$$\frac{n^*_i}{n_{si} - n^*_i} = e^{\Delta S^V_i/k} \; e^{-\Delta H_i/kT} \tag{1.21}$$

where ΔS^V_i is the vibrational entropy contribution and ΔH_i is the enthalpy of formation of the i defect. In general, we expect $n^*_i << n_s$ for each phase. For Si, $n_{sV} = 8$, $n_{sI} = 5$, $\Delta H_V = 2.4$ eV while $\Delta H_I \sim 8$ eV [1.68]. For α-cristobalite, $n_{SV} = 4$ for Si, O or SiO_2 vacancies, while $n_{sI} \sim 15$ (one I-defect adjacent to each O site in a ring), and no information is available on the ΔH values; however, it is well known that many different kinds of interstitial ions enter the interstitial spaces in natural β-cristobalite so it is expected that ΔH_I will be small for this defect in SiO_2 [1.69]. Certainly, it is expected that $\Delta H_I^{SiO_2} << \Delta H_I^{Si}$. For the vacancy species as well as the interstitial species in SiO_2, one must consider Si, O, I_2, SiO, SiO_2, SiO_3, SiO_4 and $(O\text{-}Si\text{-}O)_n$ chains as candidates

For each of these defect species, the free energy change for the phase, ΔG_{ij}, has the form given by Fig. 1.37 where the minimum gives the location of the equilibrium i-defect content, $n_i^*(\mu_i^* = 0)$. For our interest here, we will generally be interested in i-defect contents far removed from n_i^* where ΔG_{ij} is no longer negative but may have a large positive value.

C4. Interstitial Partition Coefficient

The equilibrium partition coefficient, k_{Io}, between the concentration of interstitial species in SiO_2 at the interface to the concentration in Si at the interface will be determined by the equality of the electrochemical potential, η_I, of this defect species, Si_I, in each phase at the interface

$$\eta_I^{Si} = \eta_I^{SiO_2} \qquad \text{at } Si/SiO_2 \tag{1.22}$$

Neglecting any electric field or strain field effects, Eq. (1.20) is inserted in Eq. (1.22) with $\eta_I = \mu_I$.

We are actually interested in the nonequilibrium partition coefficient, k_I, for the I-defect at the interface. The initial I-defects can be thought of as being created in the α-cristobalite lattice by the breaking of Si-Si bonds (thermally activated process) so that the corner Si atoms are free to become SiO_2 while the four inner Si atoms become I-defects. The entrance of the I defects into the Si side of the interface then involves an activation step because the enthalpy of formation of the I-defect in Si is expected to be much higher in SiO_2. This is conceptually illustrated in Fig. 1.38.

For the two-step process, a chemical potential difference, $\Delta\mu_I(V,\theta)$, developed between the I-defects on the SiO_2 side of the interface

compared to the I-defects on the Si side of the interface in order to supply I-defects to the Si at a rate determined by the oxidation rate, V. This can be thought of in analogy with water trickling through a leaky dam; the higher is the water level, the greater is the leak rate. This leak rate is expected to be a function of orientation, θ, because the activation barrier existing right at the interface in Fig. 1.38 is expected to be a function of orientation. It is also expected to be a function of interface velocity, V, since the flux, J_{II}, of I defects into the Si depends upon V. Using classical reaction rate theory for the two-step process we expect to find

$$J_{II} \approx A\, e^{-Q(\theta)/kT}\, e^{+\Delta\theta/kT} \tag{1.23a}$$

where A is a constant with a slight orientation dependence for a given number of I-defects in the SiO_2 and

$$Q(\theta) = \Delta H_I^{Si} + Q^*(\theta) - \Delta H_I^{SiO_2} \tag{1.23b}$$

(see Fig. 1.38). Thus, instead of Eq. (1.22), for the zero interface field case we must use

$$\Delta\mu(V,\theta) + \mu_I^{Si} = \mu_I^{SiO_2} \tag{1.24}$$

to determine the nonequilibrium interface partition coefficient, $k_I(V,\theta)$.

Utilizing Eqs. (1.24, (1.22), and (1.20), k_I is given by

$$\beta = \left(\frac{1 - n^*_{I1}/n_{s1}}{1 - n^*_{I2}/n_{s2}} \right) \frac{n^*_{I2}}{n^*_{I1}} \exp\left[\Delta\mu(V,\theta)/kT\right] \qquad (1.25b)$$

In our case, $n^*_{I1}/n_{s1} << 1$ and probably $n^*_{I2}/n_{s2} << 1$ as well so that the first term on the right of Eq. (1.25b) is ~1. From Eq. (1.23a), we also note that $\beta \alpha J_{I1}$. Although we do not have any information on $\Delta\mu(V,\theta)$ or on n^*_{I2}/n^*_{I1}, we can evaluate k_I in terms of β. In Fig. 1.39, plots of k_I and $n_{I2}(0)$ are given as a function of $n_{I1}(0)/n_{s2}$ for a range of β-values. We note that k_I decreases from its saturation value of ~∞ at very small values of $n_{I1}(0)$ to an inverse dependence on $n_{I1}(0)$ at large values of $n_{I1}(0)$. We note also that $n_{I2}(0) = k_I n_{I1}(0)$ increases towards its ultimate value of four per unit cell of Si transformed as $n_{I1}(0)$ increases and at a rate less than linear.

If we make the approximation that $n_{I2}(0) \approx 4$ in the derivation leading to Eq. (1.25a), then we are led to the result that

$$k_I \approx \beta \left(1 - \frac{4}{n_{s2}}\right) + \frac{4}{n_{s1}} \qquad (1.26)$$

Thus, in this limit, from Eqs. (1.23a) and (1.25a), we see that $k_I \alpha J_{I1}$. For this same approximation, we shall find the excess free energy stored in I-defects, ΔG_E, to be given by

$$\Delta G_E = kT \left\{ 4 \ln \left[\frac{4(n_{s2} - n^*_{I2})}{n_{s2}\, n^*_{I2}} \right] + (n_{s2} - 4) \ln \left(\frac{n_{s2} - 4}{n_{s2}} \right) \right\} \qquad (1.27a)$$

$$\approx kT \left\{ 4 \ln(4/n^*_{I2}) + 11 \ln(11/15) \right\} \qquad (1.27b)$$

assuming that $n_{s2} = 15$ for α-cristobalite. For $n^*_{I2} = 10^{-2}$ and 10^{-4}, this amounts to values of $\Delta G_E/kT \approx 6.5$ and 12.4 Kcals/mole, respectively. Thus, by type of structural transition involving interstitials proposed here rather than a direct transition with no interstitials, the value of ΔG_E is reduced from ~150 Kcals/mole to less than 20 Kcals/mole.

C5. Summary

A qualitative model has been proposed based upon the initial transformation of silicon to α-cristobalite plus almost four interstitial Si ions per unit cell of Si transformed, with the remainder interstitial Si ions rejected into the bulk silicon phase. Subsequent oxidation of the interstitial Si ions in the SiO_2 matrix produces rapid distortion and transformation of the structure to the vitreous form. The following are the qualitative predictions of the model.

1. OED, OISF, N_it and Q_f can be accounted for.
2. The orientation dependence gives, in accordance with experiment,

$$k_I(100) < k_I(111)$$
$$OED_{(100)} > OED_{(111)}$$
$$OISF_{(100)} > OISF_{(111)}$$
$$Q_f(100) < Q_f(111)$$
$$N_{it(100)} < N_{it(111)}$$
$$V_{(111)} > V_{(100)}$$

3. The temperature dependence gives k_I increasing as T increases above ~ 950°C so that all the orientation dependences tends to decrease.
4. At the lower growth temperatures, a layer of increased density SiO_2 will exist at the Si interface which will yield an increased index of refraction for the layer in accordance with experiment. The chemical stoichiometry of the layer will range from almost SiO (slightly Si deficient) at the interface with Si to almost SiO_2

(slightly O deficient) at the transition with bulk density vitreous SiO_2 which occurs within ~3-5 unit cell distances of the interface.

The next major step is to quantitatively account for all of these phenomena and, for this, a special class of computations is needed. Interstitial formation energy, ΔE_F, for a wide range of species Si, Si^{n+}, O_2, O, O^{m-}, SiO, SiO_2, SiO_4, H_2O, etc., placed at various locations in the SiO_2 ring structure will be made. The following information will be sought: (a) the number of interstitial sites per ring, (b) the magnitude of the energy well at these sites, (c) the charge redistribution occurring in the ring and the outward force on the ring. From these results, it is expected that the enthalpies of solution, ΔH_j, for such interstitial j species, their activation energies for diffusion, Q_j, swelling stresses on the SiO_2 ring, σ_r, and molar free energy of reaction for constrained oxidation may be obtained. With this information plus additional information on the flow characteristics of SiO_2, a first principles calculation of oxidation velocity becomes possible.

1.4 FUTURE WORK/CONCLUSIONS

In general the ongoing work under this program and its future directions follow directly from the objectives discussed above. With respect to high pressure oxidation, the characterization of oxide charges associated with steam oxidation in the temperature range 800°-1000°C will be completed. In addition dry O_2 high pressure oxidation process will be characterized and oxide charge properties determined. As with steam, the pressure range will be 1-25 atmos, while the temperature range will be slightly greater, 700°-1000°C. Cooperative work with C.R. Helms and co-workers will continue in characterizing chlorine-containing oxides from the standpoint of

chlorine concentration profiles and electrical properties, as described in Section 5 of this report.

With respect to thin oxides, a major effort is currently underway to gather a large body of experimental data on kinetics and charge properties using the $Ar/HCl/O_2$ furnace described in the body of this report. This data should help to resolve many of the questions which remain about physical mechanisms in thin oxide growth. Our work on physical modeling of the oxidation process will emphasize continued development of the "working model" described above with particular emphasis on quantifying this model in areas in which it is at present largely qualitative.

In addition to these ongoing activities several new efforts have begun or will begin shortly. Several of these are associated with multilayer structures which are of particular interest to VLSI devices. A specific example of interest is polysilicon on SiO_2 on silicon. This structure is widely used in NMOS circuits today and at present it cannot be modeled in SUPREM. Parameters of interest include the thermal oxidation of polysilicon layer (especially when it is heavily doped) and the resistivity of the poly layer as a function of thermal processing and oxidation. Close cooperation with the group currently working on polysilicon (see Section 3) is underway to address these problems. In addition, the numerical techniques needed for SUPREM to handle such multilayer structures are being developed (see Section 6). A second example is Si_3N_4 on SiO_2 silicon which is a structure widely used in local oxidation processes. In this case the masking properties of the Si_3N_4 with respect to oxidation and dopant diffusion are of interest.

A new effort is also currently underway to experimentally measure the annealing properties of oxide charge densities in inert (Ar) ambients. Work in this area has been only sparsely reported in the literature [1.63] but should be enormously useful in clarifying the physical mechanisms responsible for oxide charges. The same experimental set-up and techniques currently being used for thin oxide kinetic studies are also being used in these annealing studies.

The overall goal of all this work, of course, is the development of generally useful process models for SUPREM. As devices become smaller and processes more precisely controlled, such models must be increasingly accurate to realistically predict device structures. We believe this implies that process models must be increasingly based upon fundamental physical understanding, rather than on empirically determined parameters. This concept is the underlying motivation for the work described in this report.

REFERENCES

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

[1.2] J.D. Meindl et al., "Final Report on Computer-Aided Engineering of Semiconductor Integrated Circuits," Stanford Electronics Laboratories, TRDXG569, July 1979.

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

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

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

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

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

[1.8] J. Blanc, "A Revised Model for the Oxidation of Si by Oxygen," Appl. Phys. Letters, 33(5), Sept. 1, 1978, 424.

[1.9] W.A. Tiller, "On the Kinetics of the Thermal Oxidation of Silicon: I. A Theoretical Perspective," J. Electrochem. Soc., 127, 1980, 619-624.

[1.10] W.A. Tiller, "On Kinetics of the Thermal Oxidation of Silicon: II. Some Theoretical Evaluations," J. Electrochem. Soc., 127, 1980, 625-632.

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

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

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

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

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

[1.16] A.M. Lin, D.A. Antoniadis, R.W. Dutton, W.A. Tiller, "The Rate Control Model of Oxidation Stacking Fault Growth in Silicon," Electrochem. Soc. Extended Abstracts, Los Angeles, 1979.

[1.17] H. Shiraki, "Stacking Fault Generation Suppression and Grown-In Defect Elimination in Dislocation Free Silicon Wafers by HCl Oxidation," Japan J. Appl. Phys., 15, 1, 1976.

[1.18] H. Shiraki, "Elimination of Stacking Faults in Silicon Wafers by HCl Added Dry O_2 Oxidation," Japan J. Appl. Phys., 14, 1975, 747.

[1.19] H. Shiraki, "Suppression of Stacking Fault Generation in Silicon Wafer by HCl Added to Dry O_2 Oxidations," Japan J. Appl. Phys., 15, 1976, 83.

[1.20] T. Hattori, "Elimination of Stacking Faults in Silicon by Trichloroethylene Oxidation," J. Electrochem. Soc., 123, 1976, 945.

[1.21] G.D. Watkins and J.W. Corbett, "Defects in Irradiated Silicon: Electron Paramagnetic Resonance and Electron-Nuclear Double Resonance of the Si-E Center," Phys. Rev. A, 134 1964, 1359.

[1.22] Y. Enomoto, T. Ishioka, S. Matsumoto, T. Niimi, unpublished.

[1.23] D.A. Antoniadias, A.G. Gonzalez, R.W. Dutton, "Boron in Near Intrinsic (100) and (111) Silicon Under Inert and Oxidizing Ambients - Diffusion and Segregation," J. Electrochem. Soc., 125, 1978, 813.

[1.24] W.G. Allen and K.V. Anand, "Orientation Dependence of the Diffusion of Boron in Silicon," Solid State Elec., 14, 1971, 397.

[1.25] D.A. Antoniadis, A.M. Lin, R.W. Dutton, "Oxidation-Enhanced Diffusion of Arsenic and Phosphorus in Near-Intrinsic (100) Silicon," Appl. Phys. Letters, 33(12), Dec. 1978, 15.

[1.26] G. Masetti, S. Solmi, G. Soncini, "On Phosphorus Diffusion in Silicon Under Oxidizing Atmospheres," Solid State Elec., 16, 1973, 1419.

[1.27] G.D. Watkins, in Radiation Damage in Semiconductors, Dunod, Paris, 1964, 97.

[1.28] Y. Nabeta, T. Uno, S. Kubo, H. Tsukamoto, "Restrained Diffusion of Boron and Phosphorus in Silicon under HCl-Added Oxygen Atmosphere," J. Electrochem. Soc., 123, 1976, 1416.

[1.29] A.M. Lin, R.W. Dutton, D.A. Antoniadis, "The Lateral Effect of Oxidation on Boron Diffusion in (100) Silicon," Appl. Phys. Letters, 35(10), Nov. 15, 1979.

[1.30] 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., 126, Sept. 1979, 1516-1522.

[1.31] C.P. Ho and J.D. Plummer, "Si/SiO_2 Interface Oxidation Kinetics: A Physical Model for the Influence of High Substrate Doping Levels. II. Comparison with Experiment and Discussion," J. Electrochem. Soc., 126, Sept. 1979, 1523-1530.

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

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

[1.34] 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, 1976, 3551.

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

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

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

[1.38] J.D. Meindl et al., "Final Report on Computer-Aided Semiconductor Process Modeling," Stanford Electronics Laboratories, TR-4969-73-F, Oct. 1977.

[1.39] D.W. Hess and B.E. Deal, "Kinetics of the Thermal Oxidation of Silicon in O_2HCl Mixtures," J. Electrochem. Soc., 124, May 1977, 735-739.

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

[1.41] B.E. Deal, D.W. Hess, J.D. Plummer, 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, Feb. 1978, 339-346.

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

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

[1.44] R.R. Razouk and B.E. Deal, "Dependence of Interface State Density on Silicon Thermal Oxidation Process Variables," J. Electrochem. Soc., 126, Sept. 1979, 1573-1581.

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

[1.46] R. Champagne and M. Toole, "High-Pressure Pyrogenic Oxidation in the Production Environment," Solid State Tech., 20, Dec. 1977, 61-63.

[1.47] N. Tsubouchi, H. Miyoshi, H. Abe, "Suppression of Oxidation-Induced Stacking Fault Formation in Silicon by High Pressure Steam Oxidation," Japan J. Appl. Phys., 17, Suppl. 17-1, 1978, 223-228.

[1.48] L.E. Katz and L.C. Kimerling, "Defect Formation During High Pressure, Low temperature Steam Oxidation of Silicon," J. Electrochem. Soc., 125, Oct. 1978, 1680.

[1.49] H. Miyoshi, N. Tsubouchi, A. Nishimoto, "Selective Oxidation of Silicon in High Pressure Steam," J. Electrochem. Soc., 125, Nov. 1978, 1824-1829.

[1.50] Y.J. Van der Meulen, "Kinetics of Thermal Growth of Ultra-Thin Layers of SiO_2 on Silicon. Part 1. Experiment," J. Electrochem. Soc., 119, 1972, 530.

[1.51] R. Ghez and Y.J. Van der Meulen, "Kinetics of Thermal Growth of Ultra-Thin Layers of SiO_2 on Silicon. Part 2. Theory," J. Electrochem. Soc., 119, 1972, 1100.

[1.52] F.P. Fehlner, "Formation of Ultra-Thin Oxide Films on Silicon," J. Electrochem. Soc., 119, 1972, 1723.

[1.53] J.A. Aboaf, "Formation of 20-25 Å Thermal Oxide Films on Silicon at 950-1140°C," J. Electrochem. Soc., 118, 1971, 1370.

[1.54] E. Taft and L. Cordes, "Optical Evidence for a Silicon-Silicon Oxide Interlayer," J. Electrochem. Soc., 126(1), 1979, 131.

[1.55] K. Hamano, "Breakdown Characteristics in Thin SiO_2 Films," Japan J. Appl. Phys., 13, 1974, 1085.

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

[1.57] J. Blanc, "A Revised Model for the Oxidation of Si by Oxygen," Appl. Phys. Letters, 33(5), 1978, 424.

[1.58] A.E. Irene, "Silicon Oxidation Studies: Some Aspects of the Initial Oxidation Regime," J. Electrochem. Soc., 125, 1978, 1708.

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

[1.60] F.N. Schwettman, K.L. Chiang, W.A. Brown, "Variation of Silicon Dioxide Growth Rate with Pre-Oxidation Clean," Electrochem. Soc. Extended Abstracts, Vol. 78-1, Abs. 276, Seattle, Washington, May 21-26, 1978.

[1.61] S.I. Raider, R.A. Gdula, J.R. Petrak, "Nitrogen Reaction at Silicon-Silicon Dioxide Interface," Appl. Phys. Letters, 27, Aug. 1975, 150.

[1.62] B.H. Vromen, "Nitrogen-Silicon Reaction and Its Influence on the Dielectric Strength of Thermal Silicon Dioxide," Appl. Phys. Letters, 27, Aug. 1975, 152.

[1.63] M. Revitz, S.I. Raider, R.A. Gdula, "Effect of High-Temperature, Post-Oxidation Annealing on the Electrical Properties of the Si-SiO_2 Interface," J. Vac. Sci. Tech., 16(2), Mar./Apr. 1979, 345.

[1.64] B.E. Deal, E.H. Snow, C.A. Mead, "Barrier Energies in Metal-Silicon Dioxide-Silicon Structures," J. Phys. Chem. Solids, 27, 1966, 1873.

[1.65] W.W. Werner, "The Work Function Difference of the MOS-System with Aluminum Field Plates and Polycrystalline Silicon Field Plates," Solid State Elec., 17, 1974, 769.

[1.66] W.A. Tiller, "On the Kinetics of the Thermal Oxidation of Silicon: III. Coupling with Other Key Phenomena," submitted to J. Electrochem. Soc.

[1.67] W.A. Tiller and K.S. Ahn, "Interface Field Effects on Solute Redistribution During Crystallization," J. Crystal Growth, 49, 1980, 483-501.

[1.68] J.A. Van Vechten, Phys. Rev., 517, 1978, 3197.

[1.69] W.A. Deer, R.A. Howie, J. Zusman, An Introduction to the Rock Forming Minerals, John Wiley and Sons, New York, 1966.

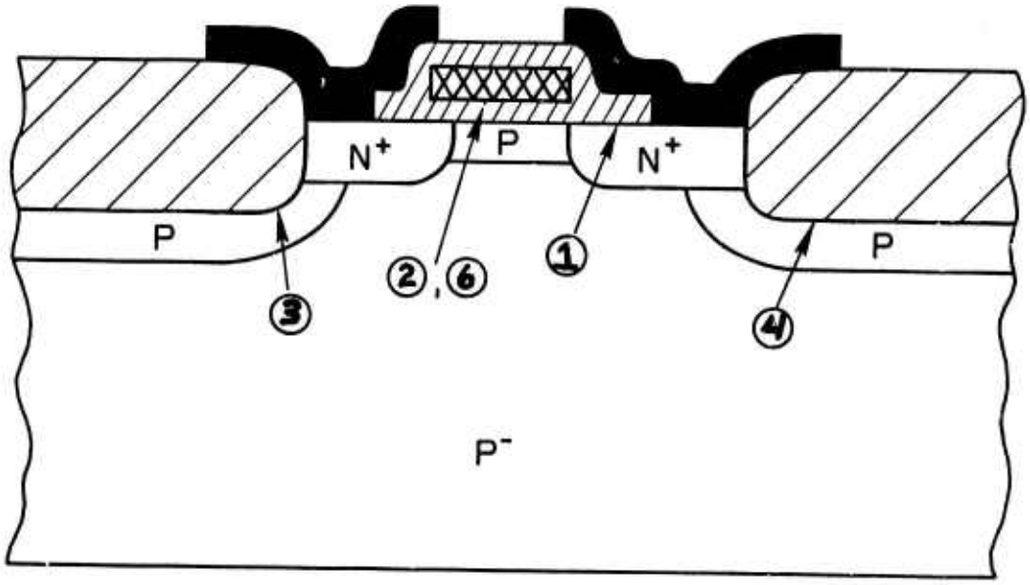


Fig. 1.1. Cross section of silicon gate NMOS device with current problem areas in oxidation kinetics indicated.

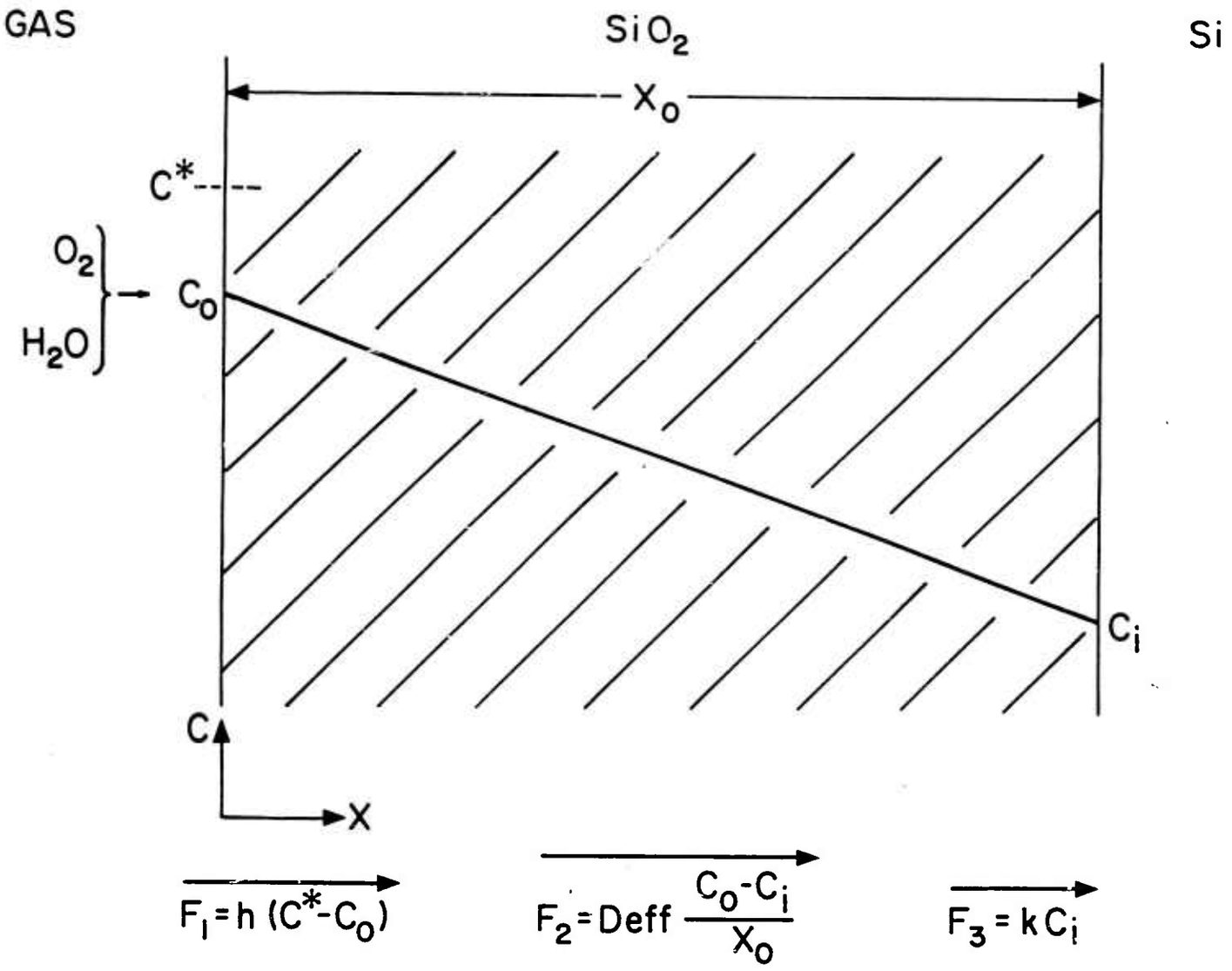


Fig. 1.2. Fluxes involved in the thermal oxidation process.

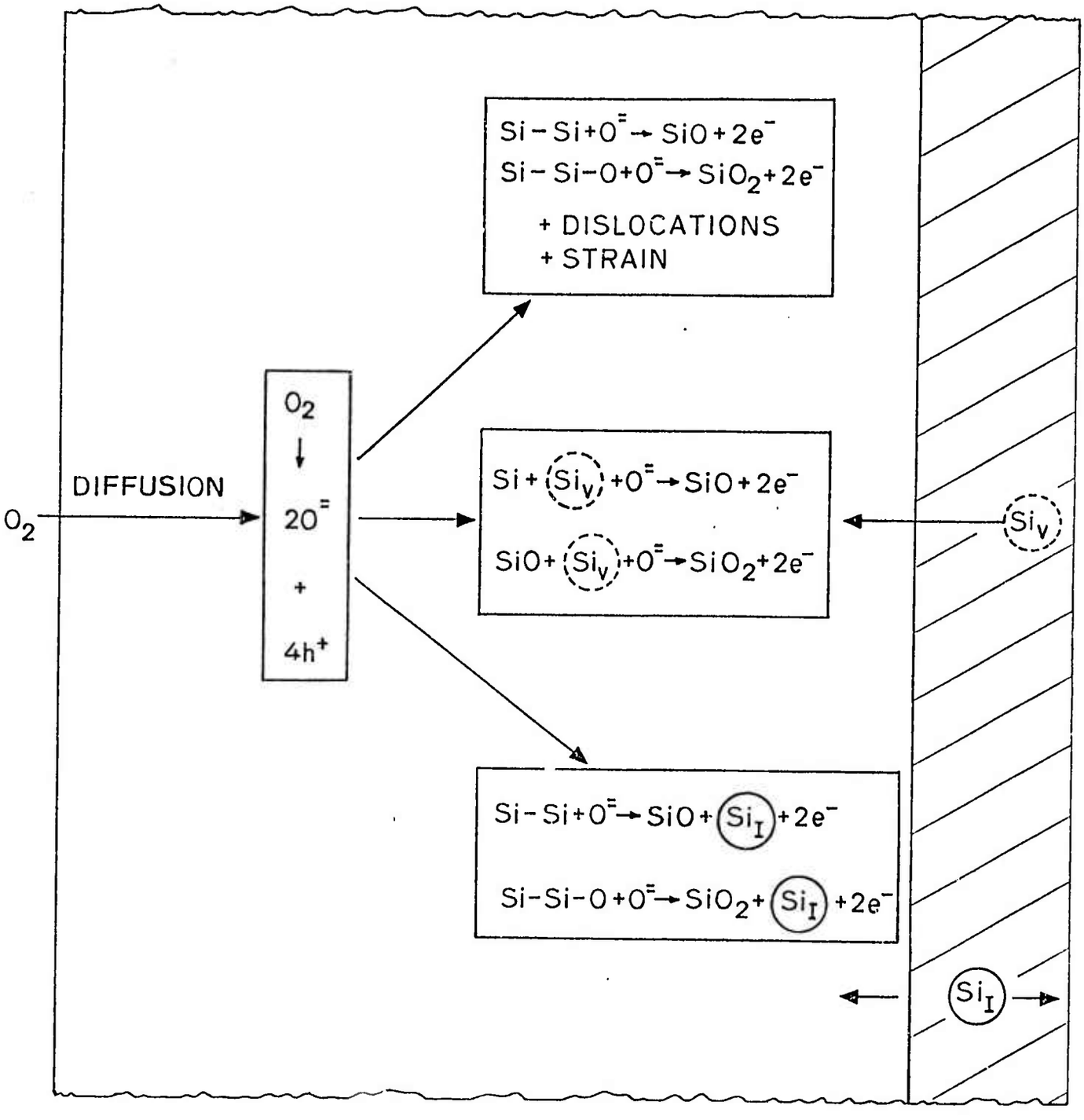


Fig. 1.3. Detailed model of the interface oxidation reaction including possible sources for the required "free volume" for the reaction Si + O_2 → SiO_2 to proceed.

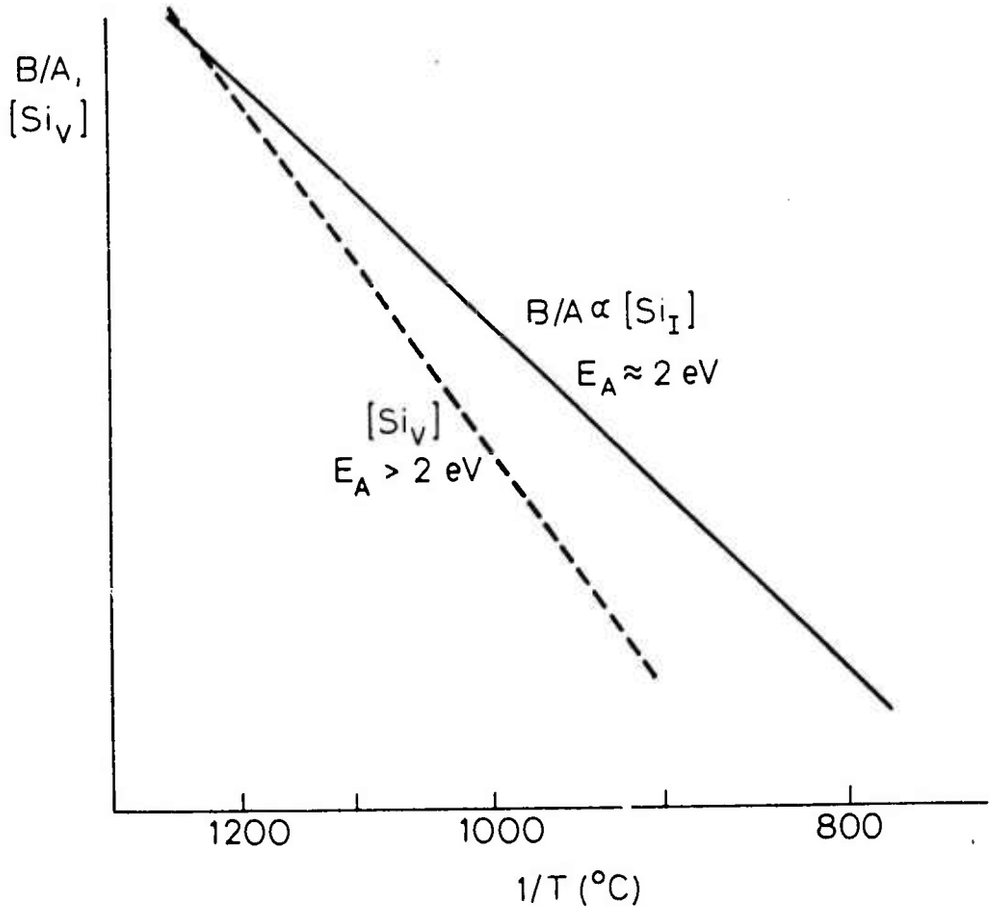


Fig. 1.4. Relative activation energies of interface reaction rate B/A and generation rate of Si_V.

SILICON OXIDE THICKNESS vs OXIDATION TIME
FOR VARYING PHOSPHORUS CONCENTRATIONS [(111) Si]

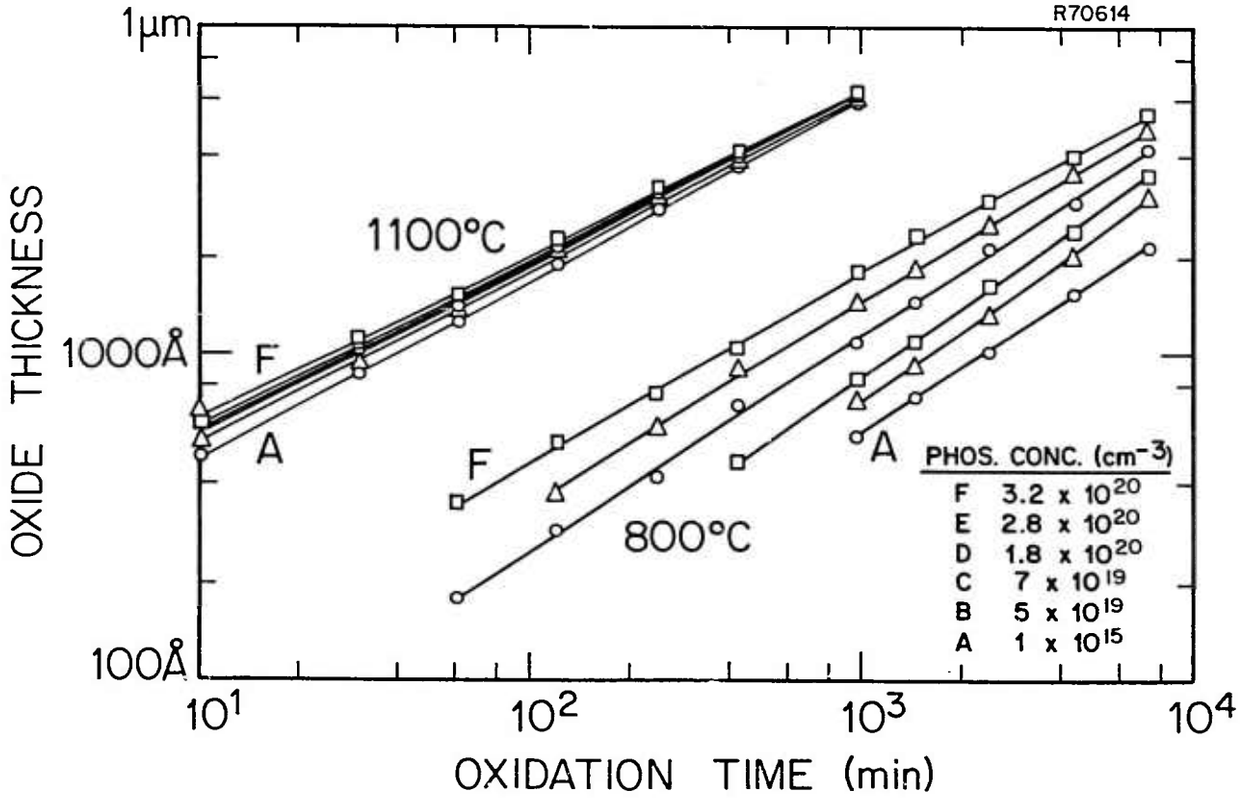


Fig. 1.5. Oxide thickness vs oxidation time for varying phosphorus concentrations.

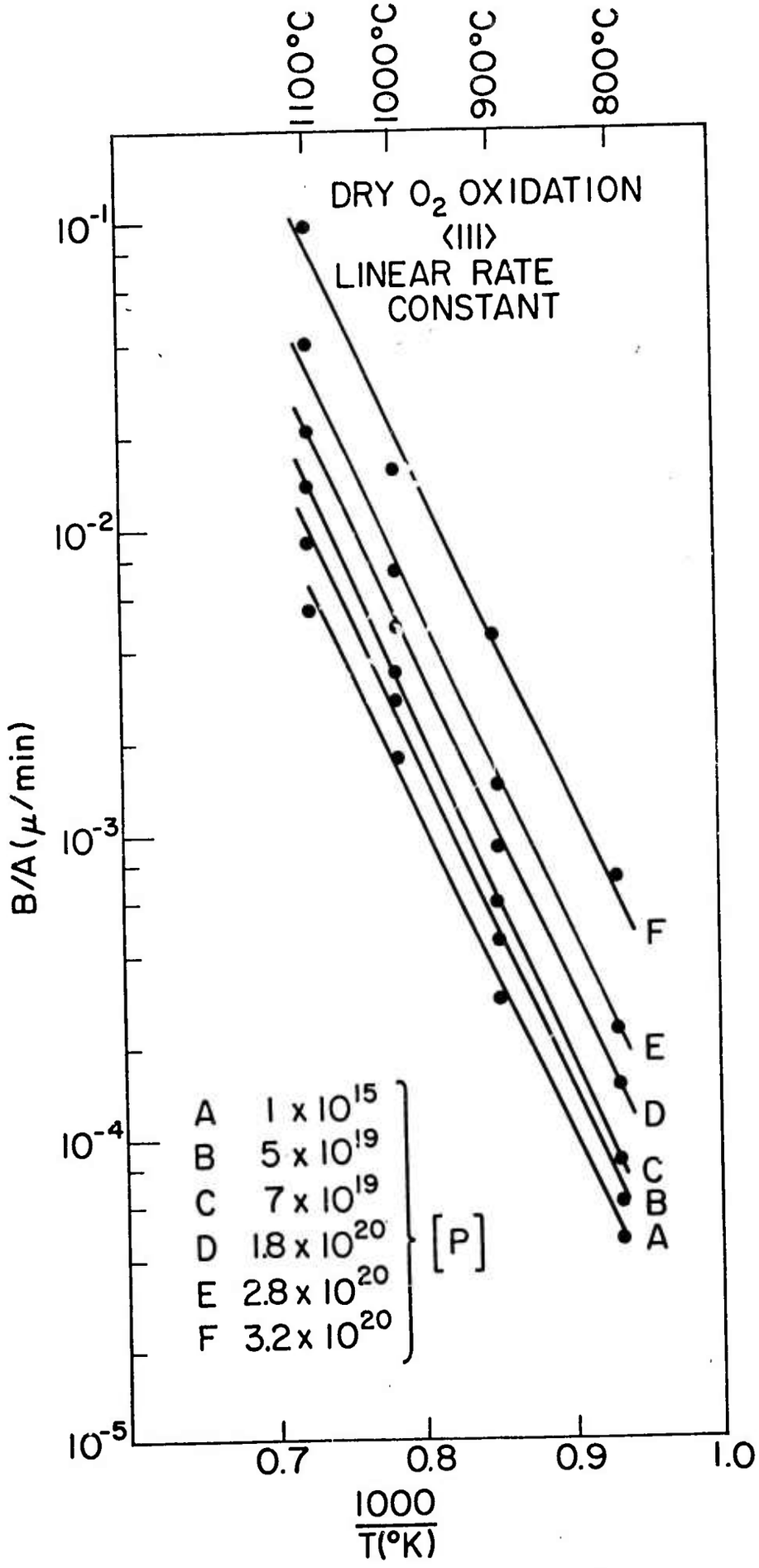


Fig. 1.6. B/A vs 1/T for the samples in Fig. 1.5.

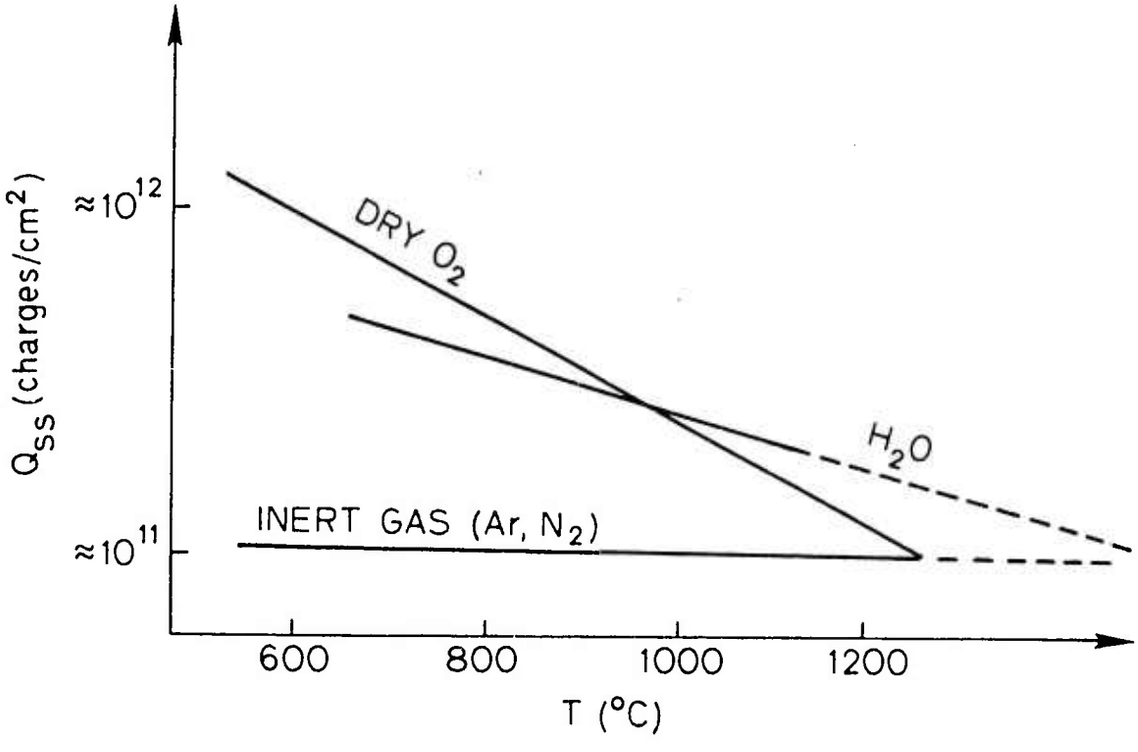


Fig. 1.7. "Q_f triangle" after Deal [1.1].

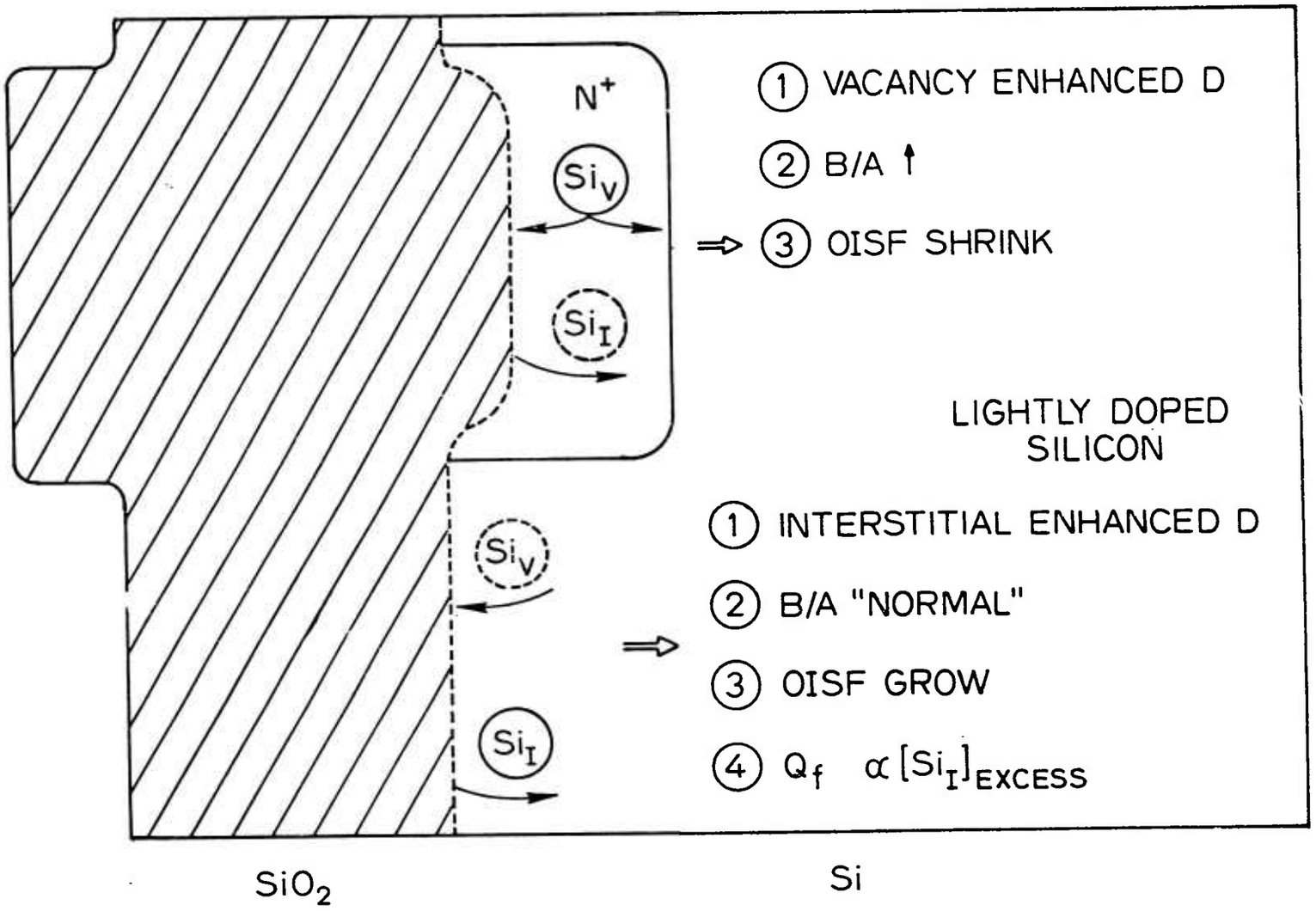


Fig. 1.8. Qualitative effects of interstitial generation and vacancy consumption during thermal oxidation.

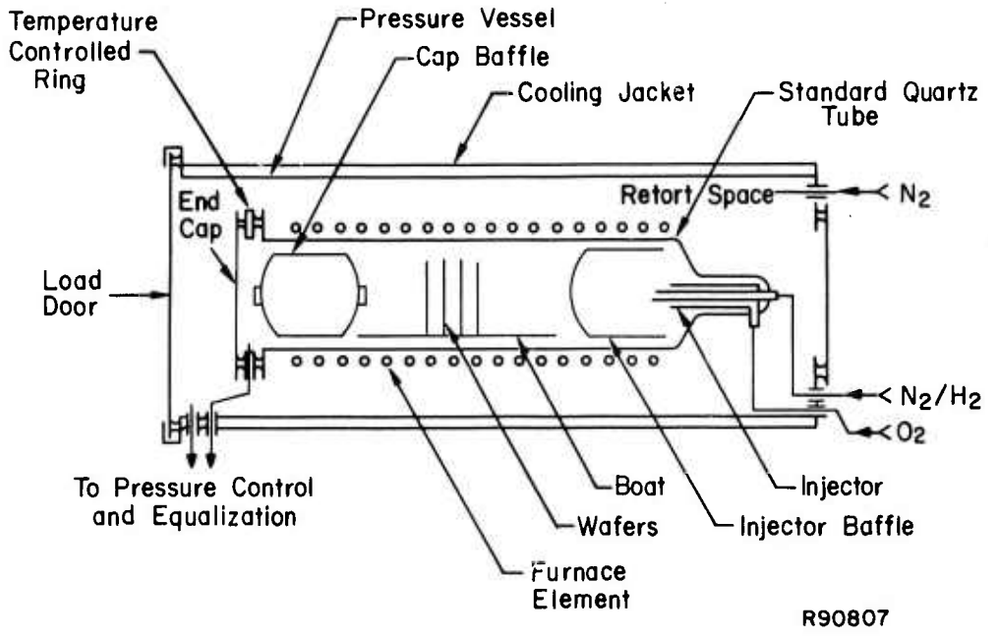


Fig. 1.9. Cross section of the high pressure oxidation system showing the inner quartz tube and the stainless steel outer shell. A standard furnace element is used.

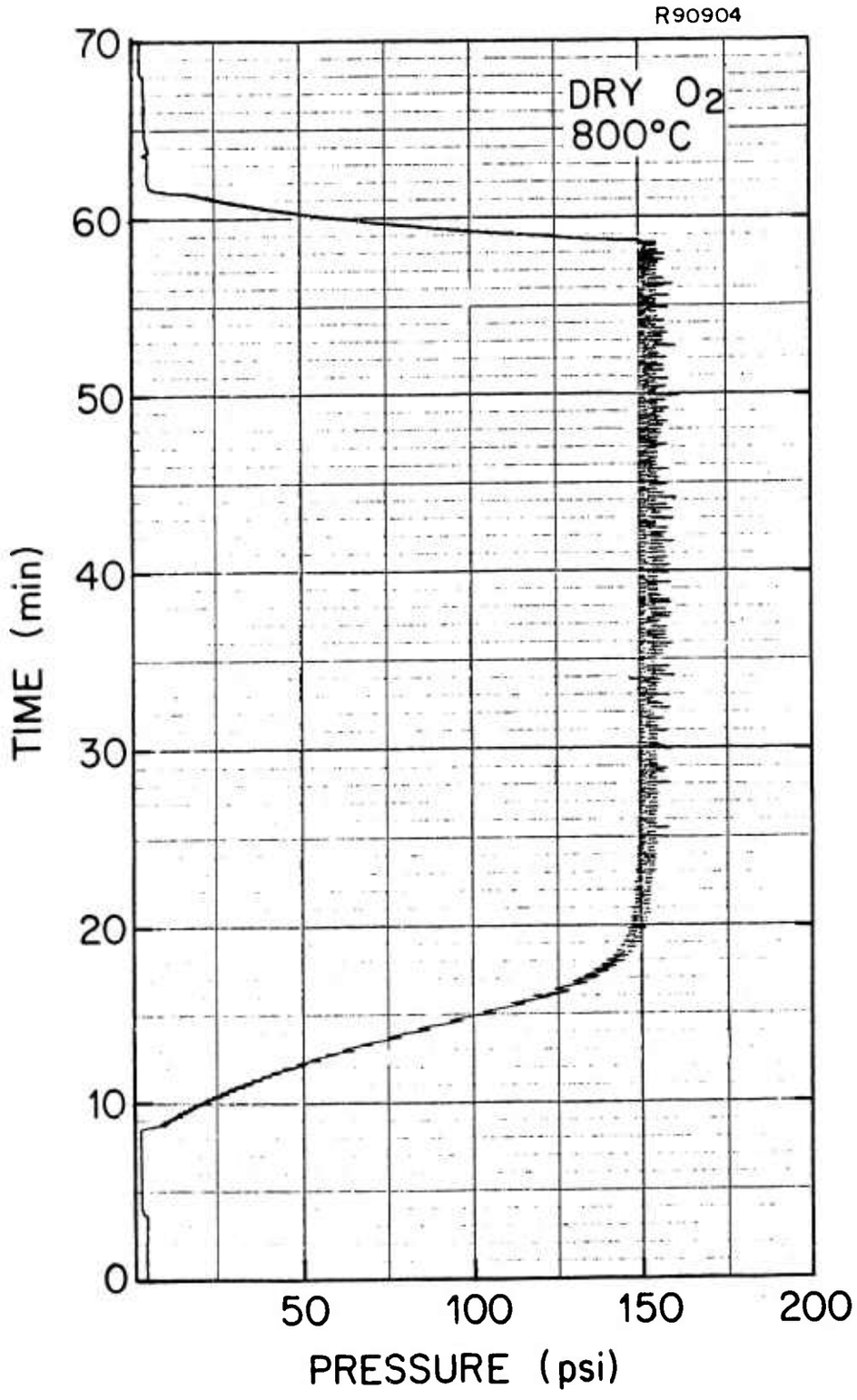


Fig. 1.10. A typical high pressure oxidation cycle showing pressurization to 10 atm in dry O_2. Pressure fluctuations of about 5 psi are typical at pressure.

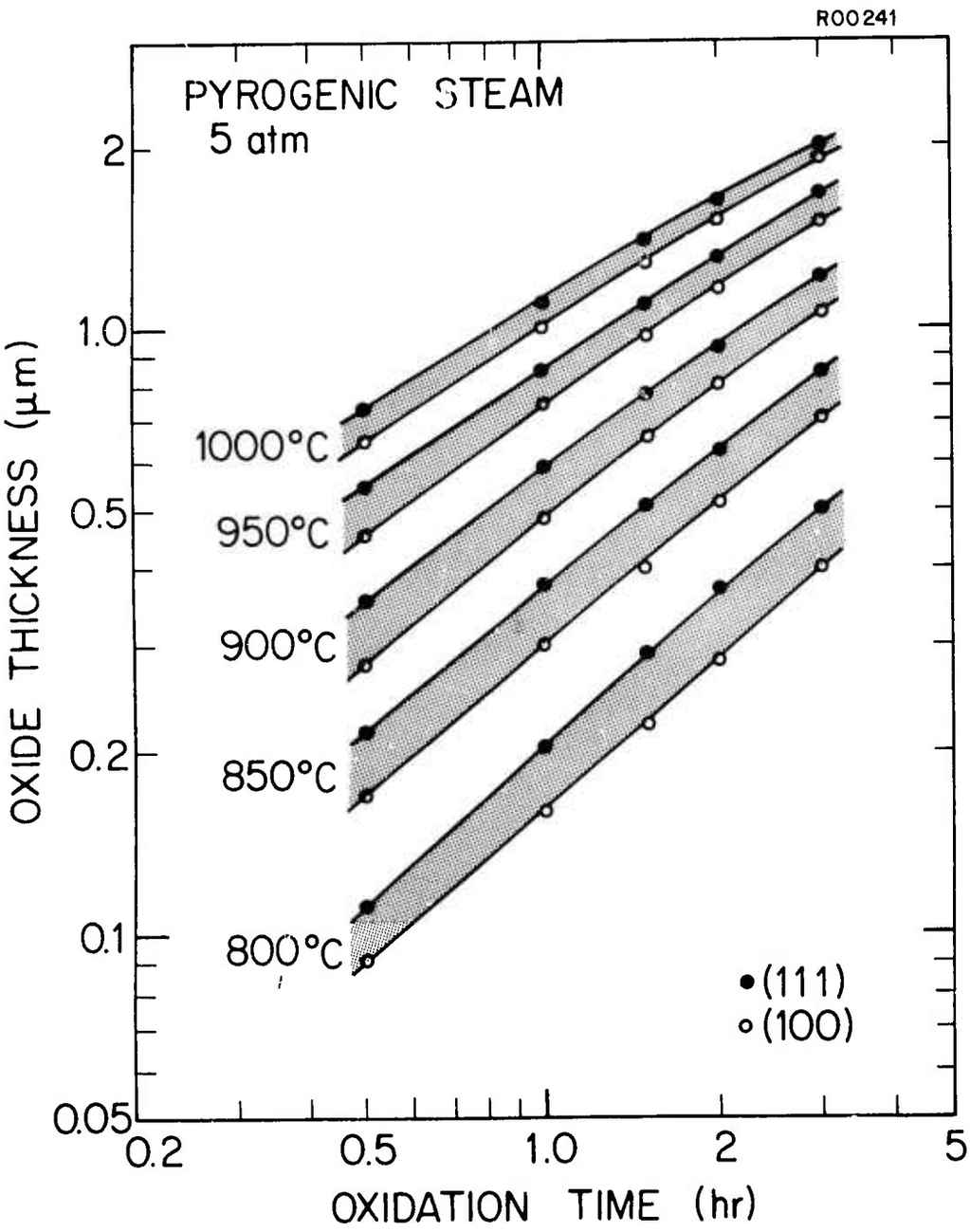


Fig. 1.11. Oxide thickness versus oxidation time for (100) and (111) silicon oxidation in pyrogenic steam at 5 atm.

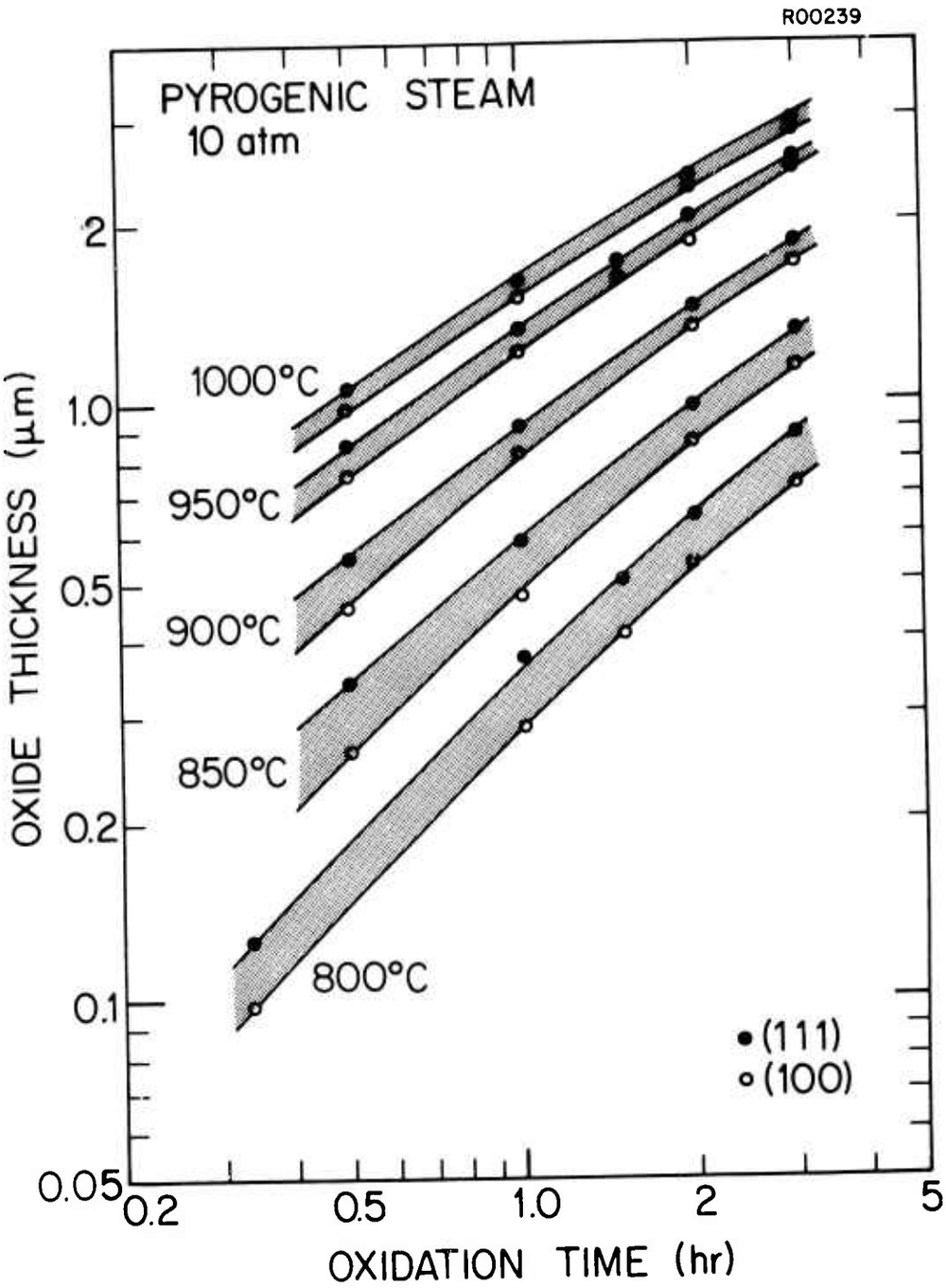


Fig. 1.12. Oxide thickness versus oxidation time for (100) and (111) silicon oxidation in pyrogenic steam at 10 atm.

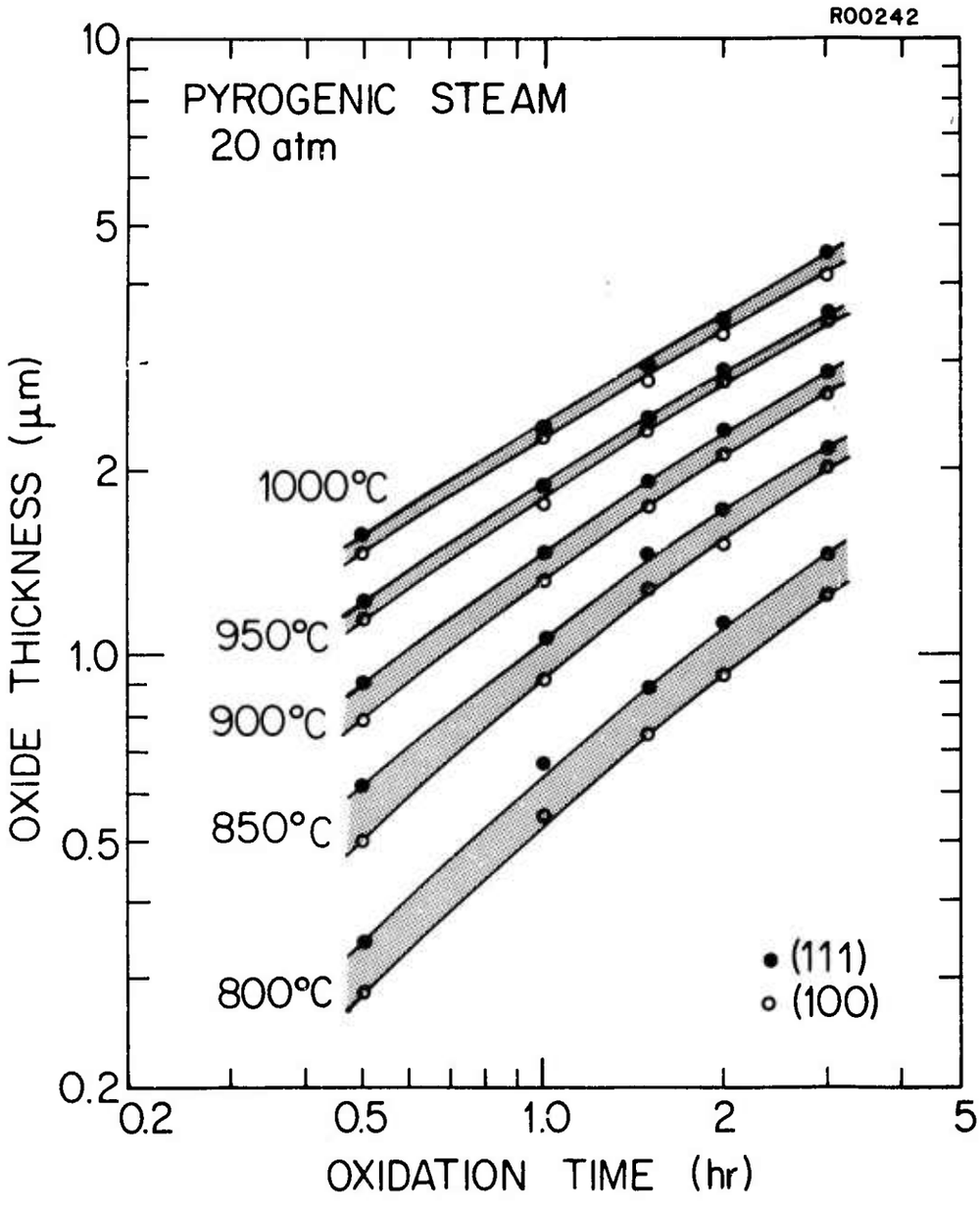


Fig. 1.13. Oxide thickness versus oxidation time for (100) and (111) silicon oxidation in pyrogenic steam at 20 atm.

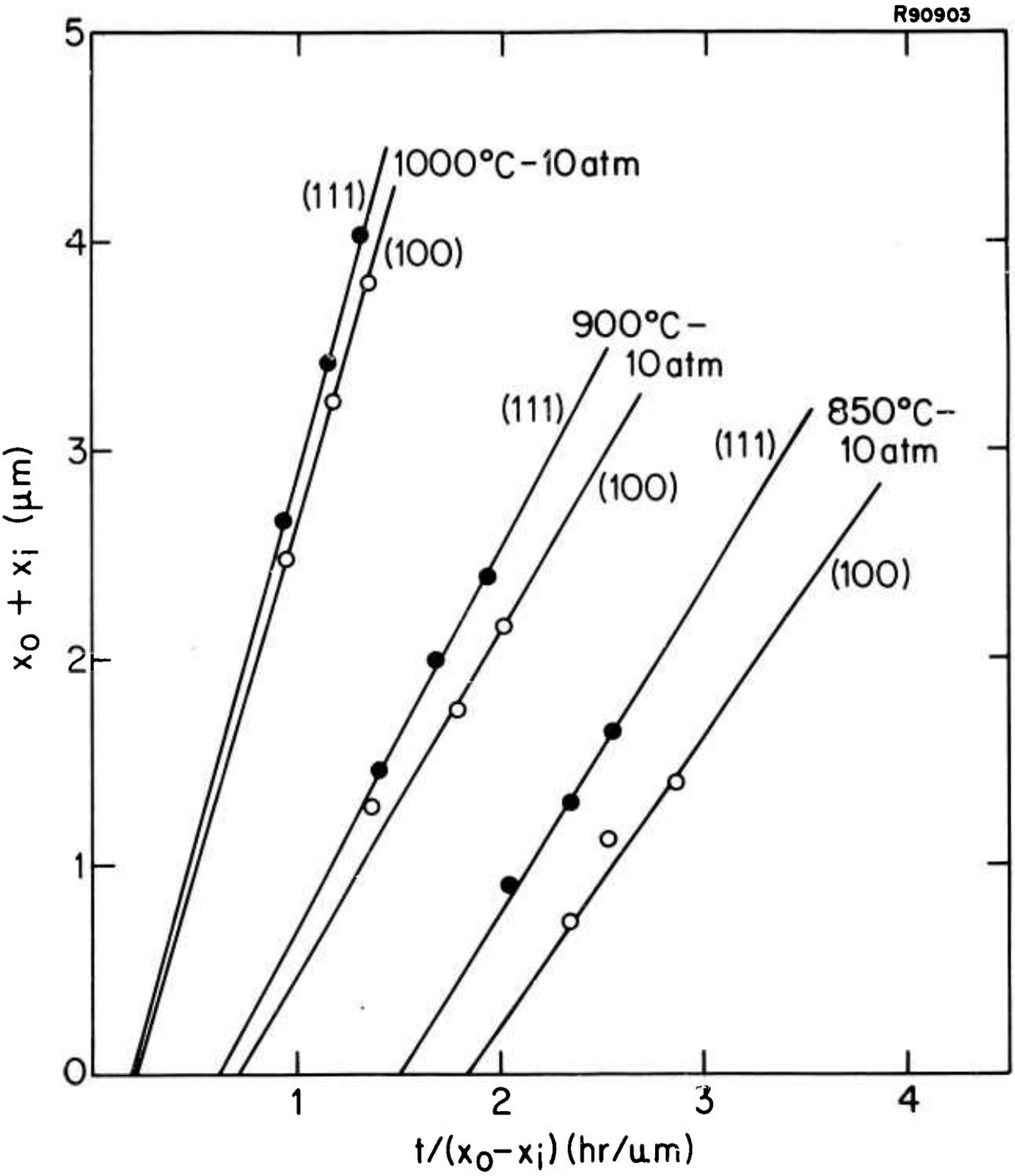


Fig. 1.14. $(x_o + x_i)$ versus $(t/x_o - x_i)$ for (100) and (111) silicon oxidized at 850°, 900°, and 1000°C in pyrogenic steam at 10 atm.

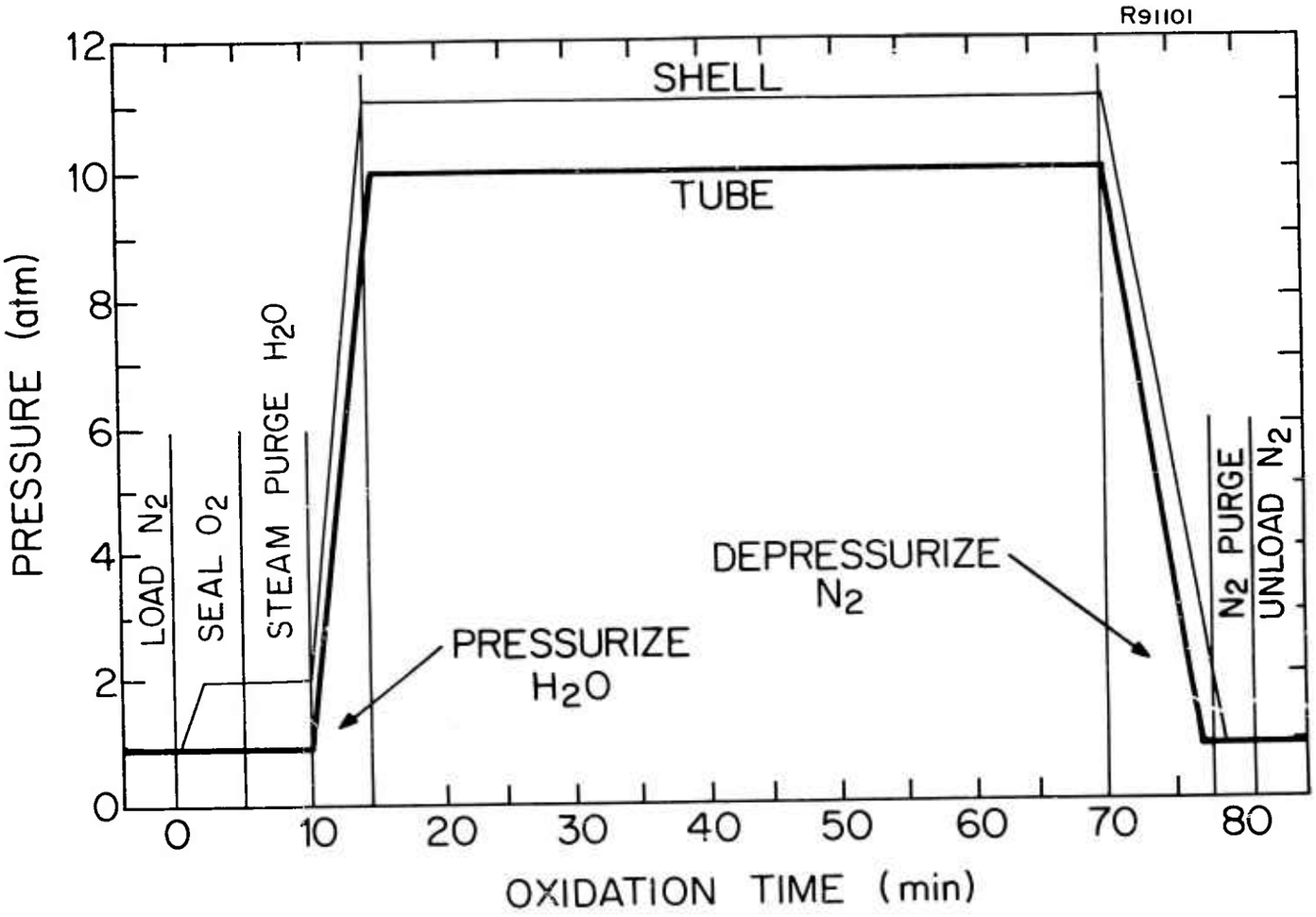


Fig. 1.15. Typical oxidation cycle showing the various regions considered in modeling the oxidation.

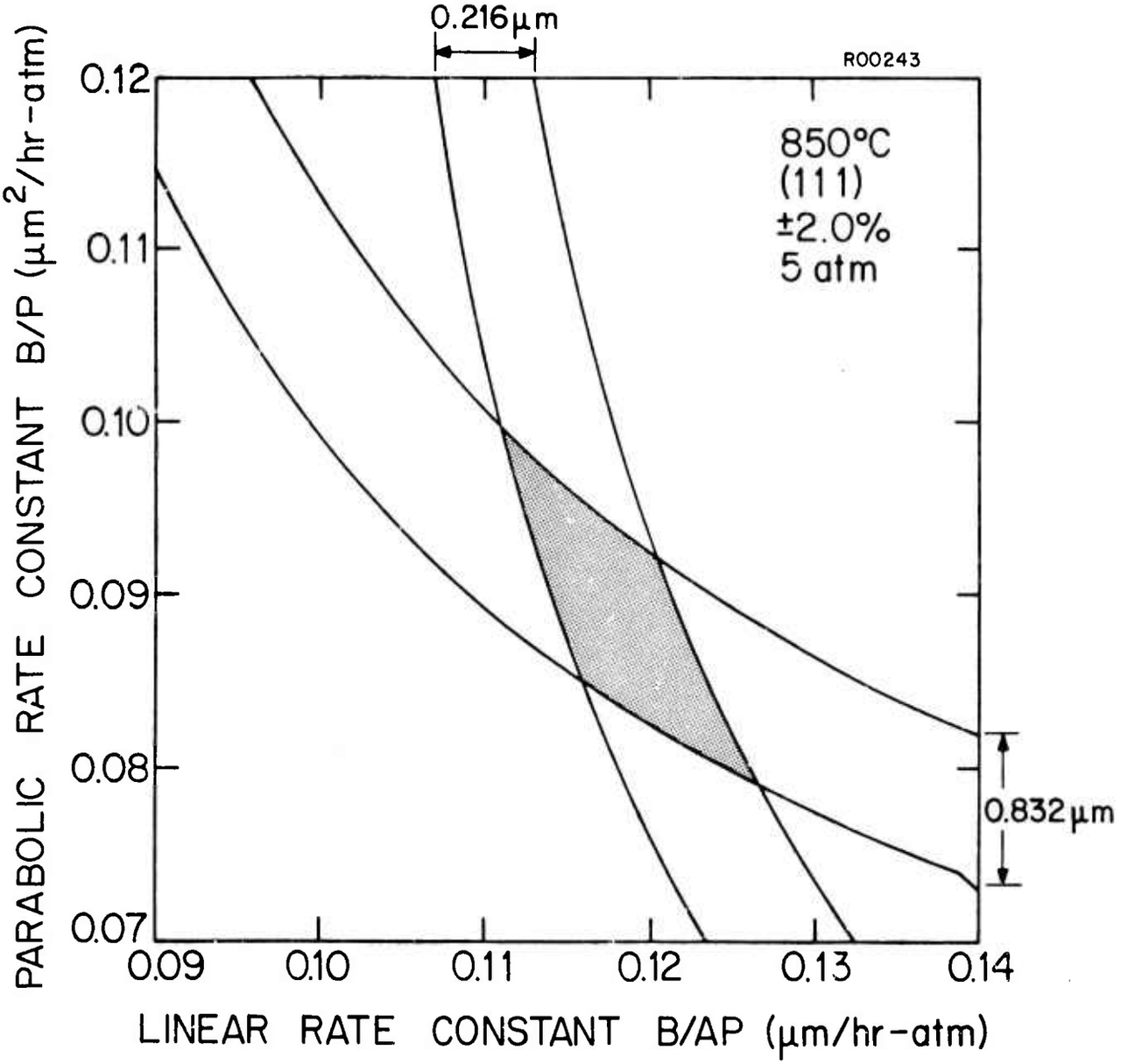


Fig. 1.16. Parabolic rate constant (B/P) versus linear rate constant (B/AP). The shaded area represents values of B/P and B/AP which when used with equation [1.7-1.11] will yield oxide thickness values within ±2% of the experimental results.

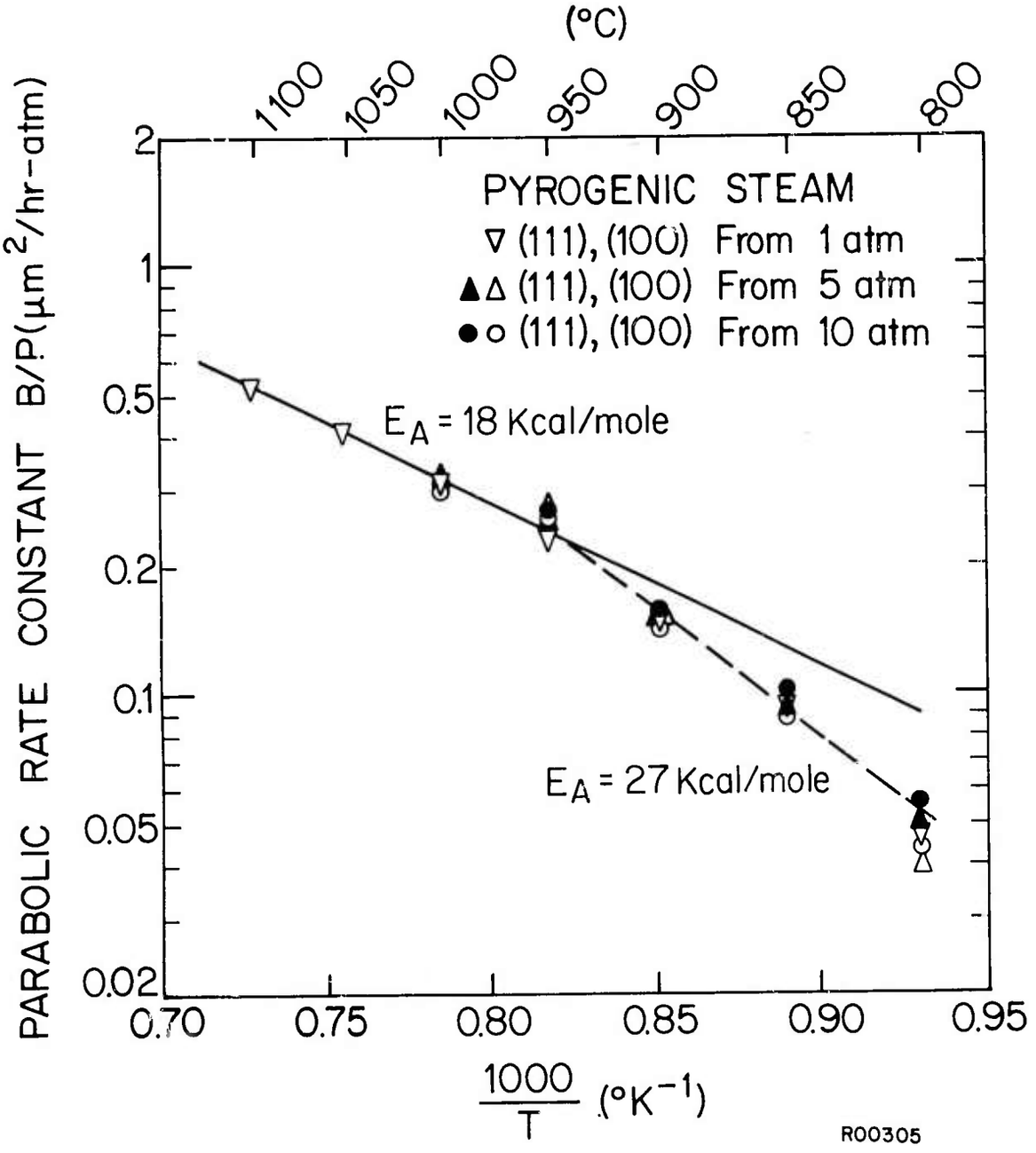


Fig. 1.17. Parabolic rate constant B/P versus 1000/T for (100) and (111) silicon wafers oxidized in pyrogenic steam as obtained from 1, 5, and 10 atm data.

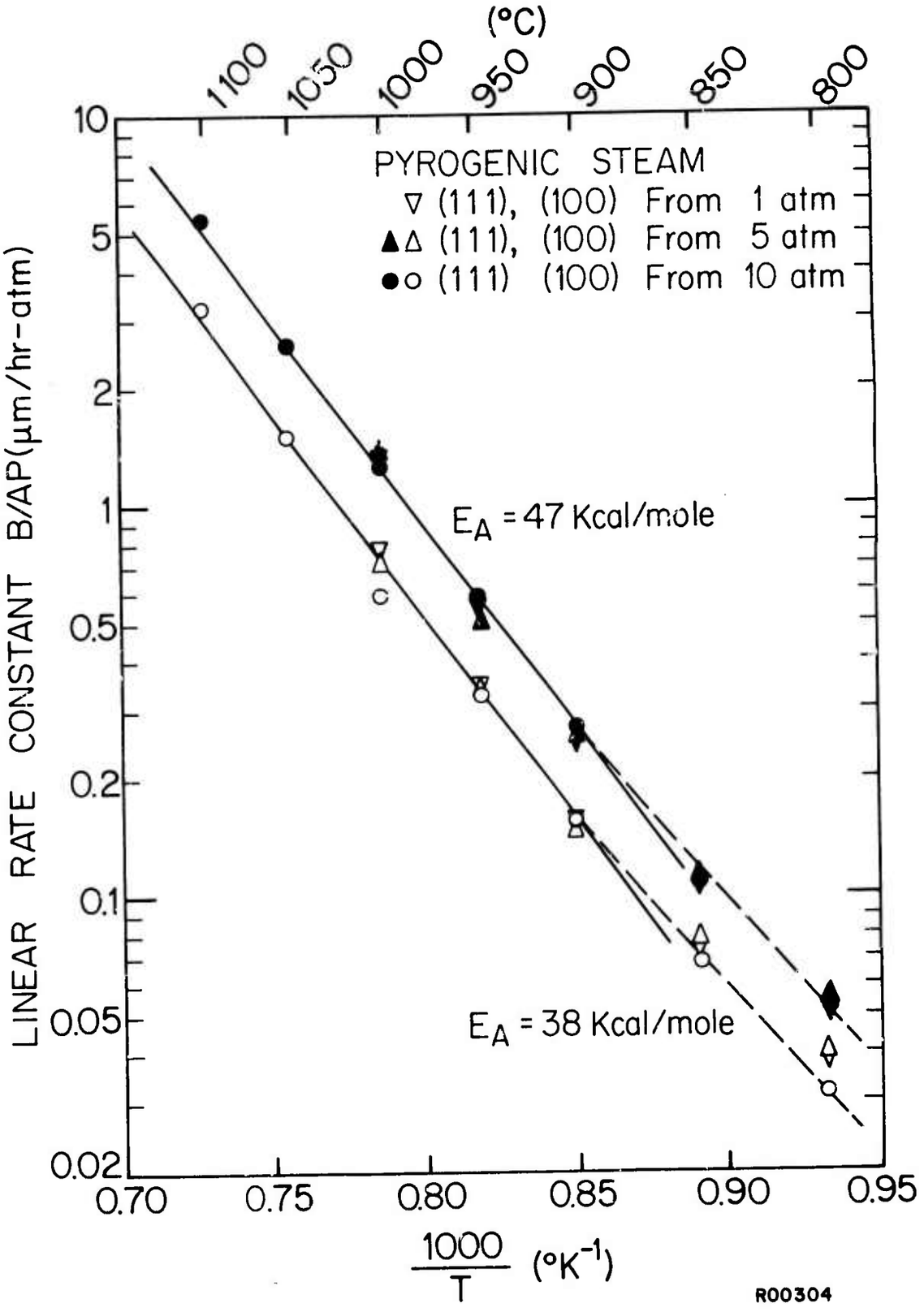


Fig. 1.18. Linear rate constant B/AP versus 1000/T for (100) and (111) silicon wafers oxidized in pyrogenic steam as obtained from 1, 5, and 10 atm data.

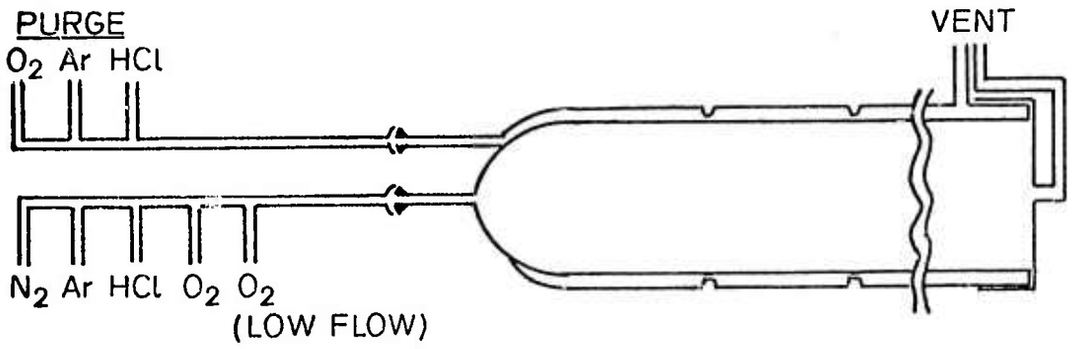


Fig. 1.19. Experimental furnace setup for the study of thin oxide kinetics and charges.

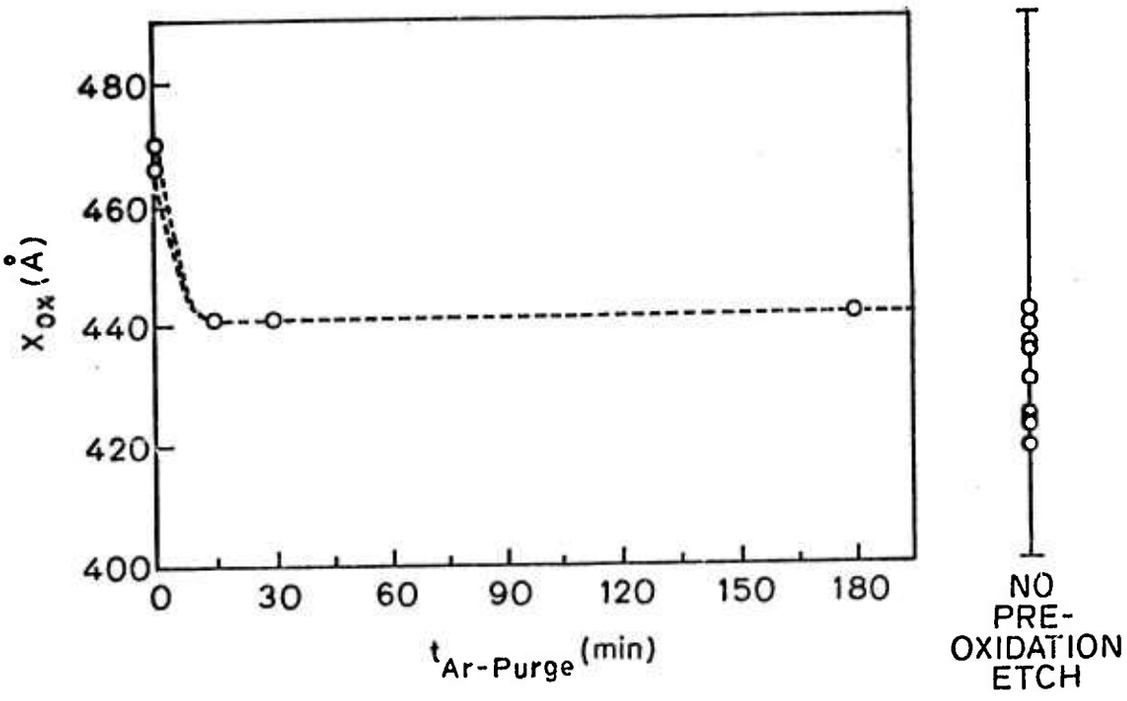


Fig. 1.20. Effect of preoxidation Ar/HCl etch and Ar purge on the reproducibility of oxides grown at 1000°C for 45 minutes in dry O_2, (100) orientation.

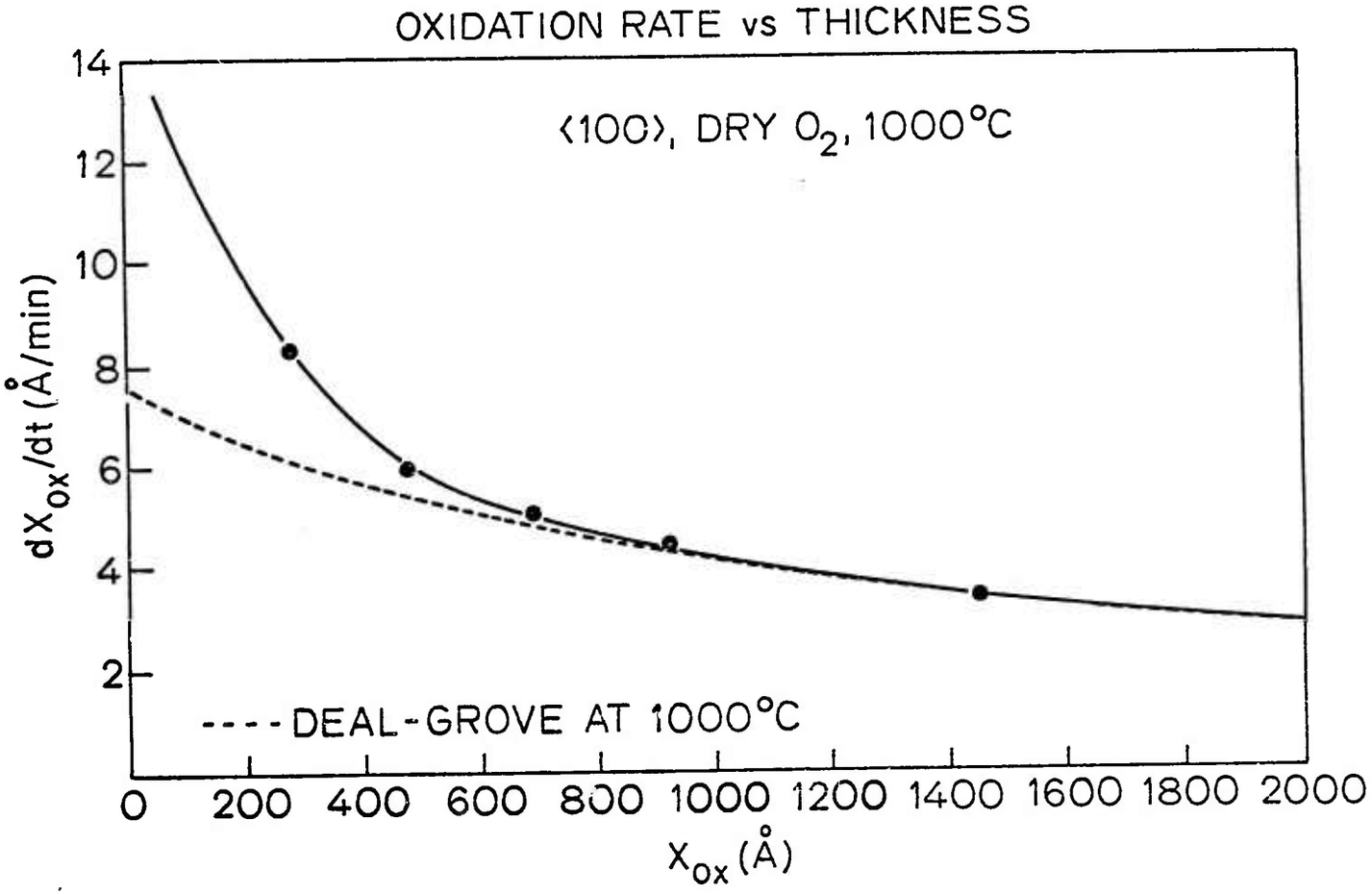


Fig. 1.21. Oxidation rate dx/dt vs oxide thickness for (100) silicon at 1000°C in dry O_2.

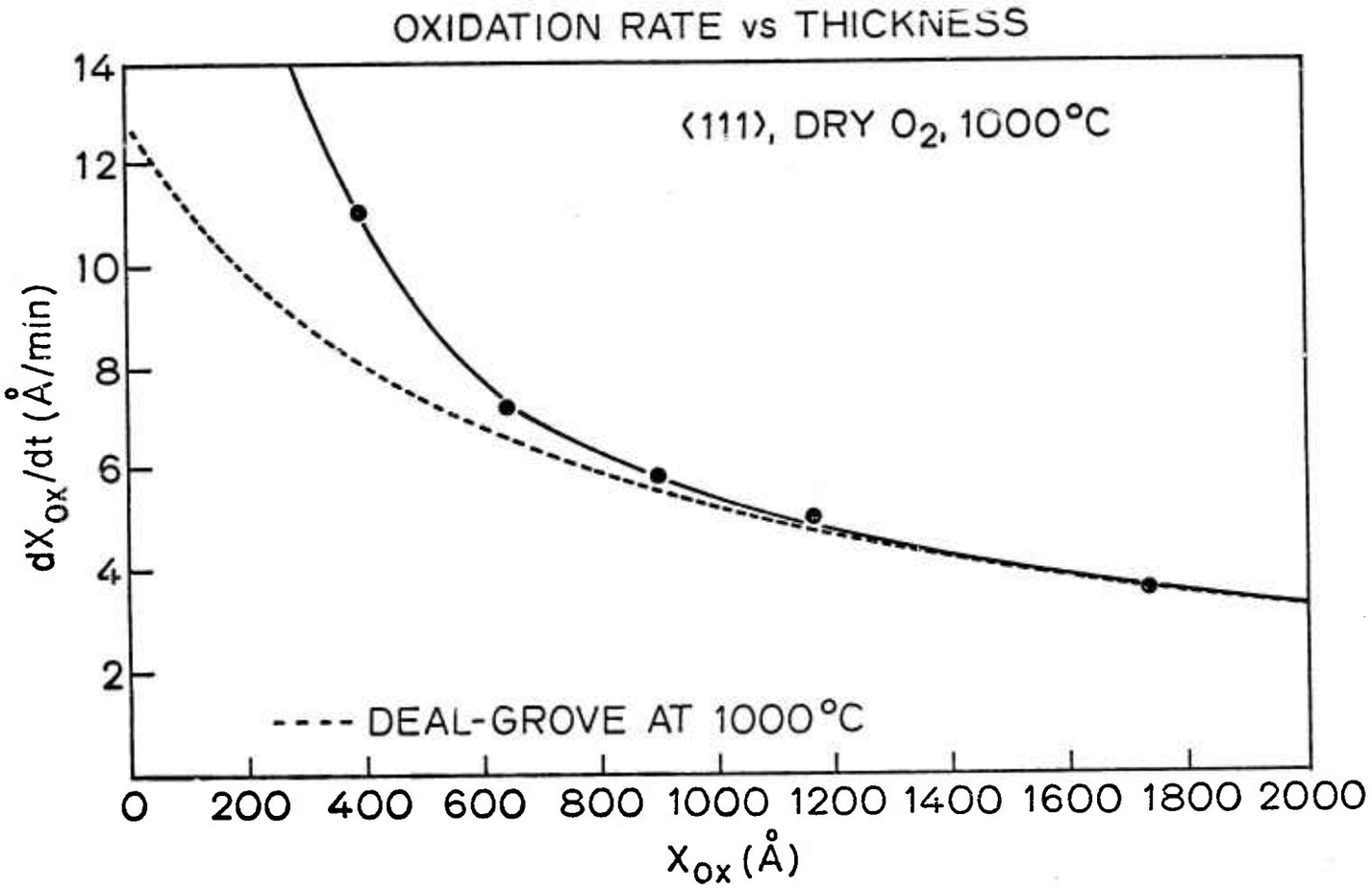


Fig. 1.22. Oxidation rate dx/dt vs oxide thickness for (111) silicon at 1000°C in dry O_2.

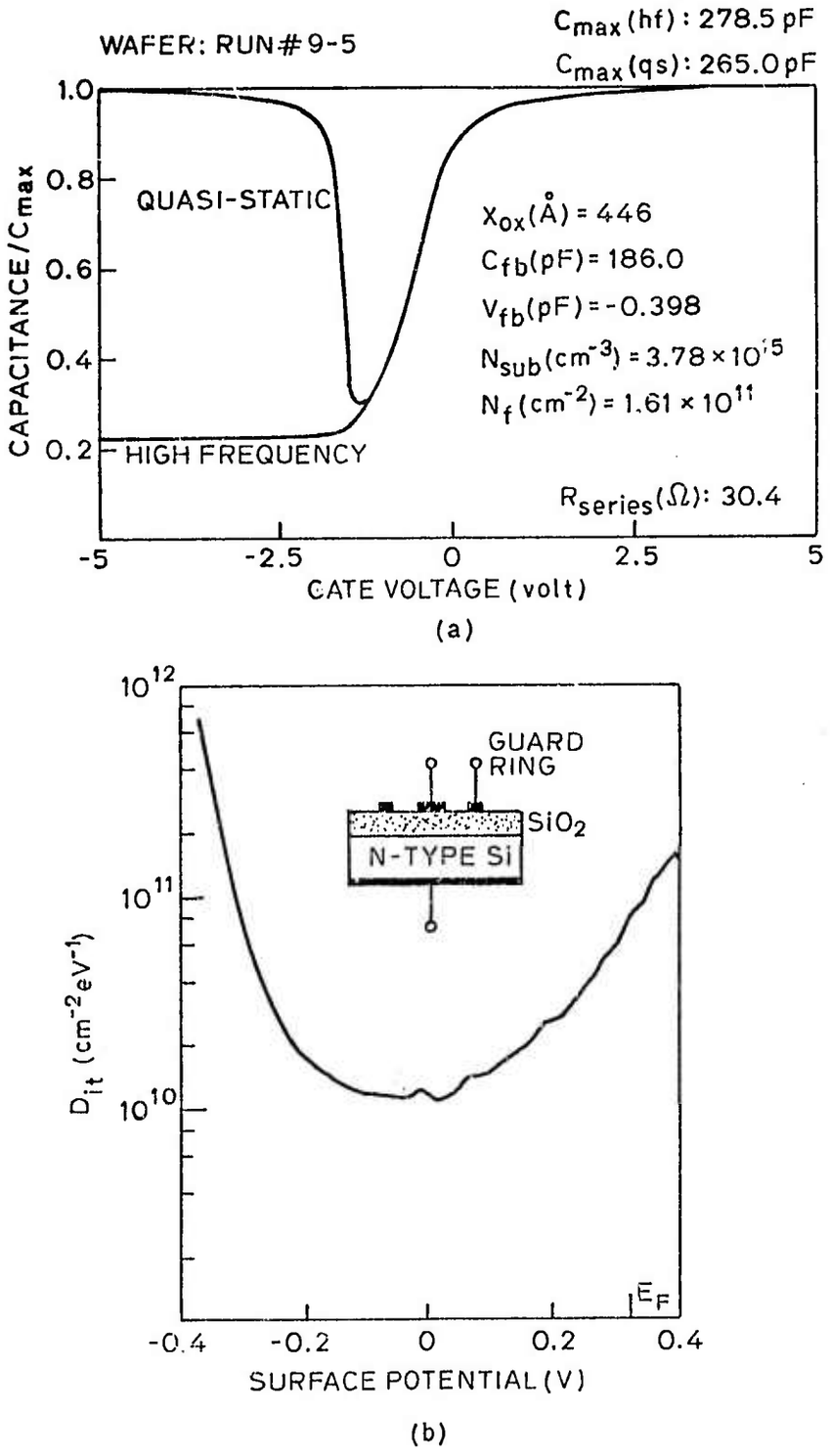


Fig. 1.23. Examples of (a) high frequency and quasi-static CV plots and (b) interface state density distribution.

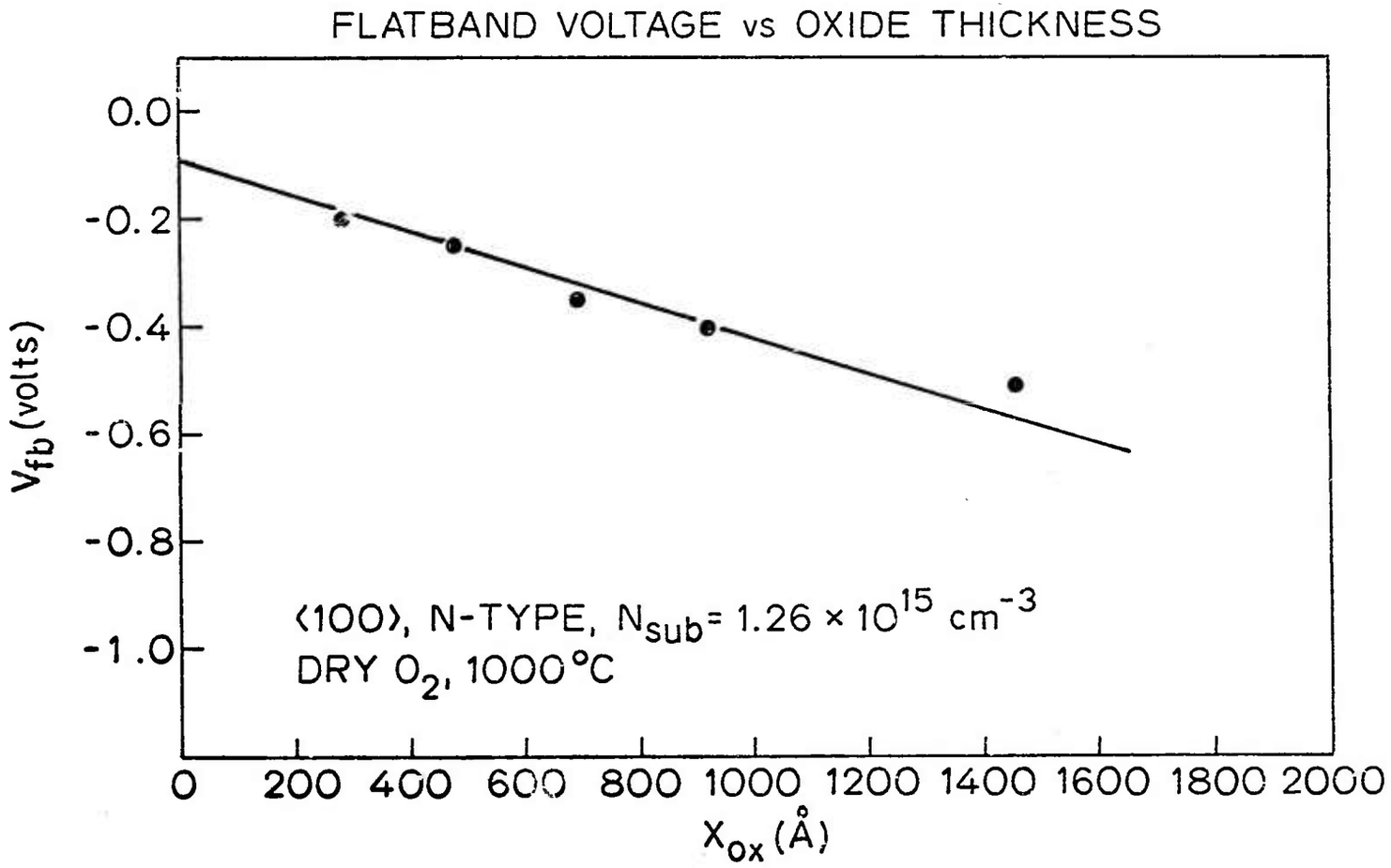


Fig. 1.24. Flatband voltage vs oxide thickness for (100) N-type wafers oxidized at 1000°C and fast pulled from O_2 ambient.

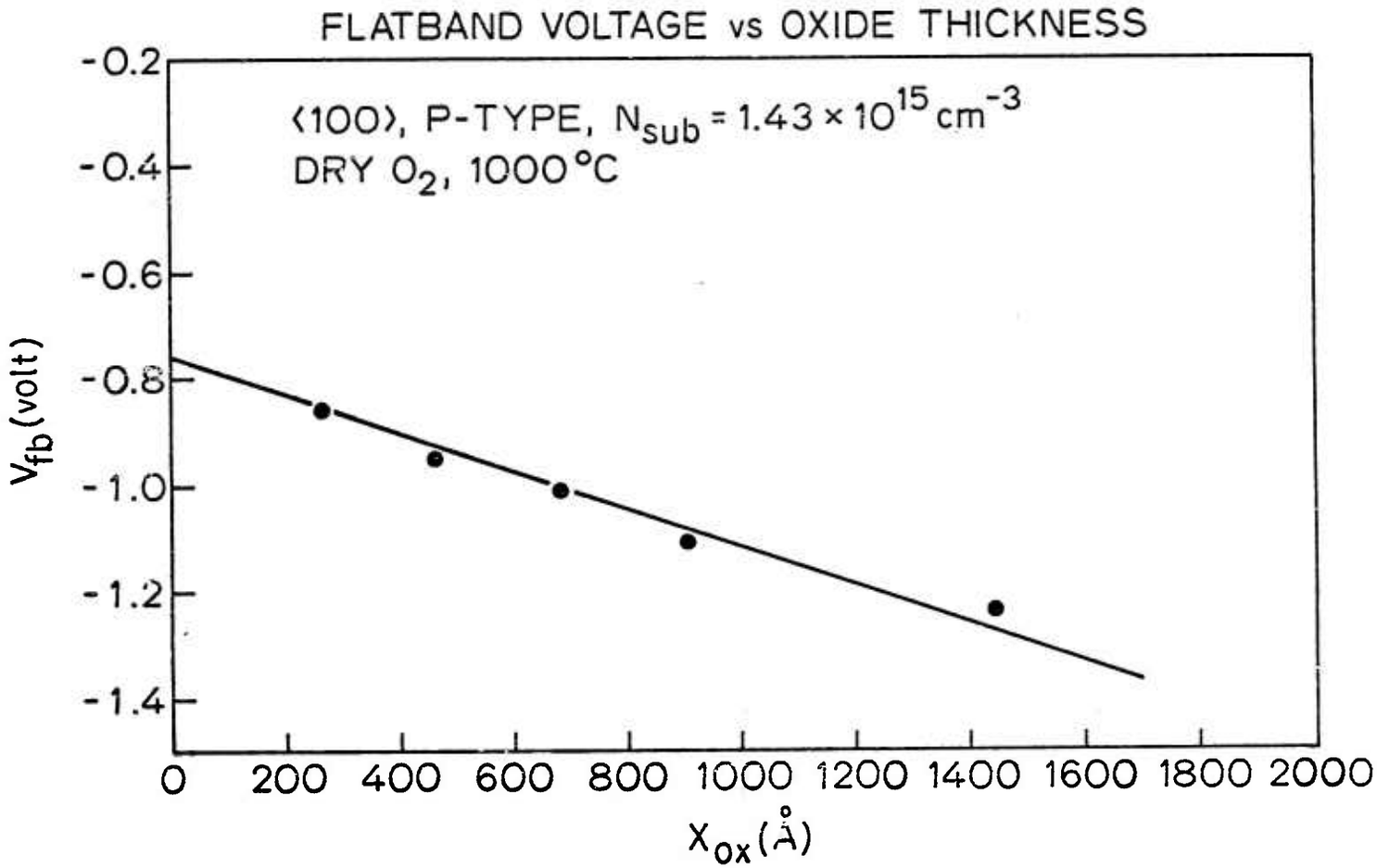


Fig. 1.25. Flatband voltage vs oxide thickness for (100) P-type wafers oxidized at 1000°C and fast pulled from O_2 ambient.

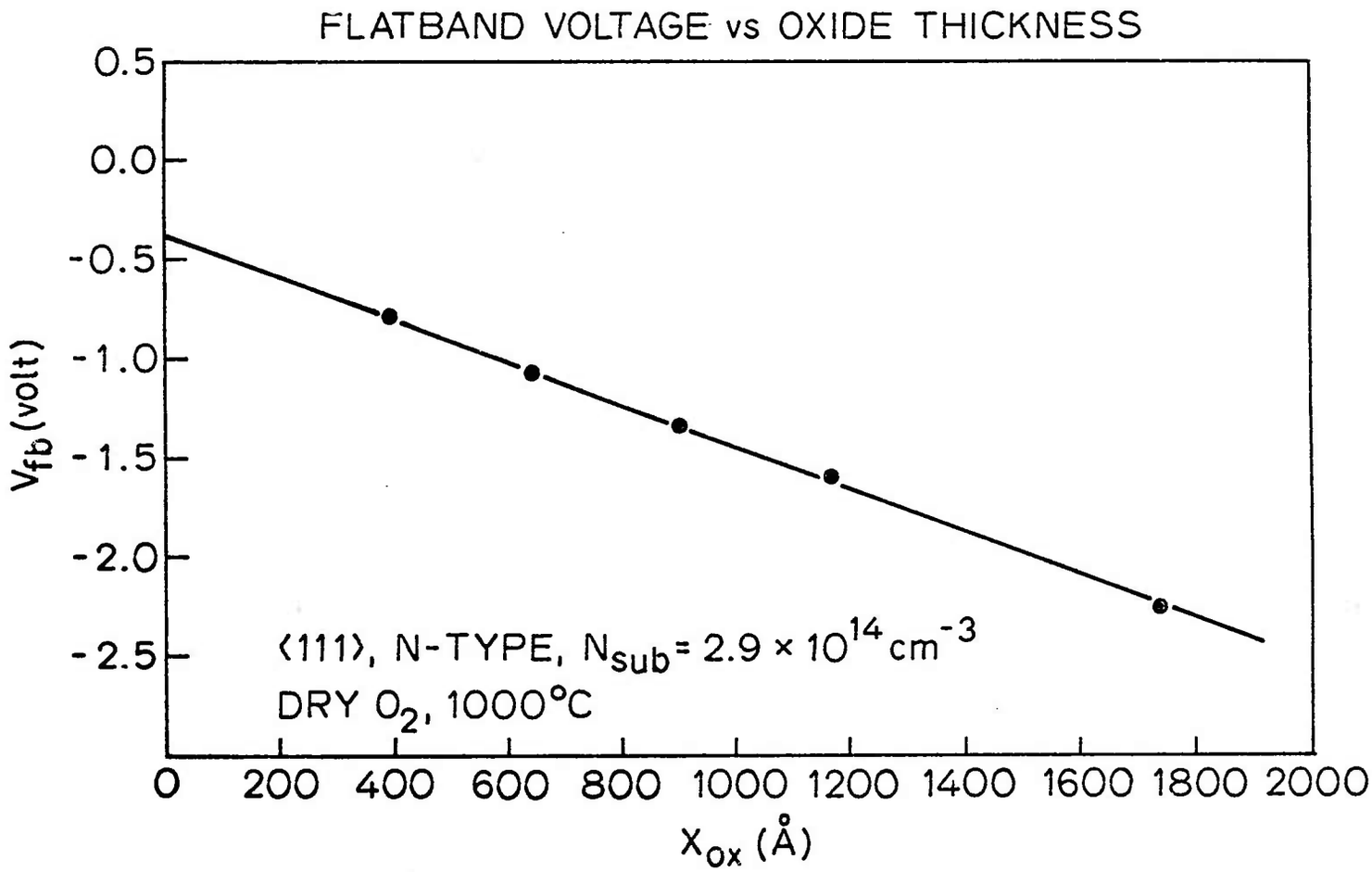


Fig. 1.26. Flatband voltage vs oxide thickness for (111) N-type wafers oxidized at 1000°C and fast pulled from O_2 ambient.

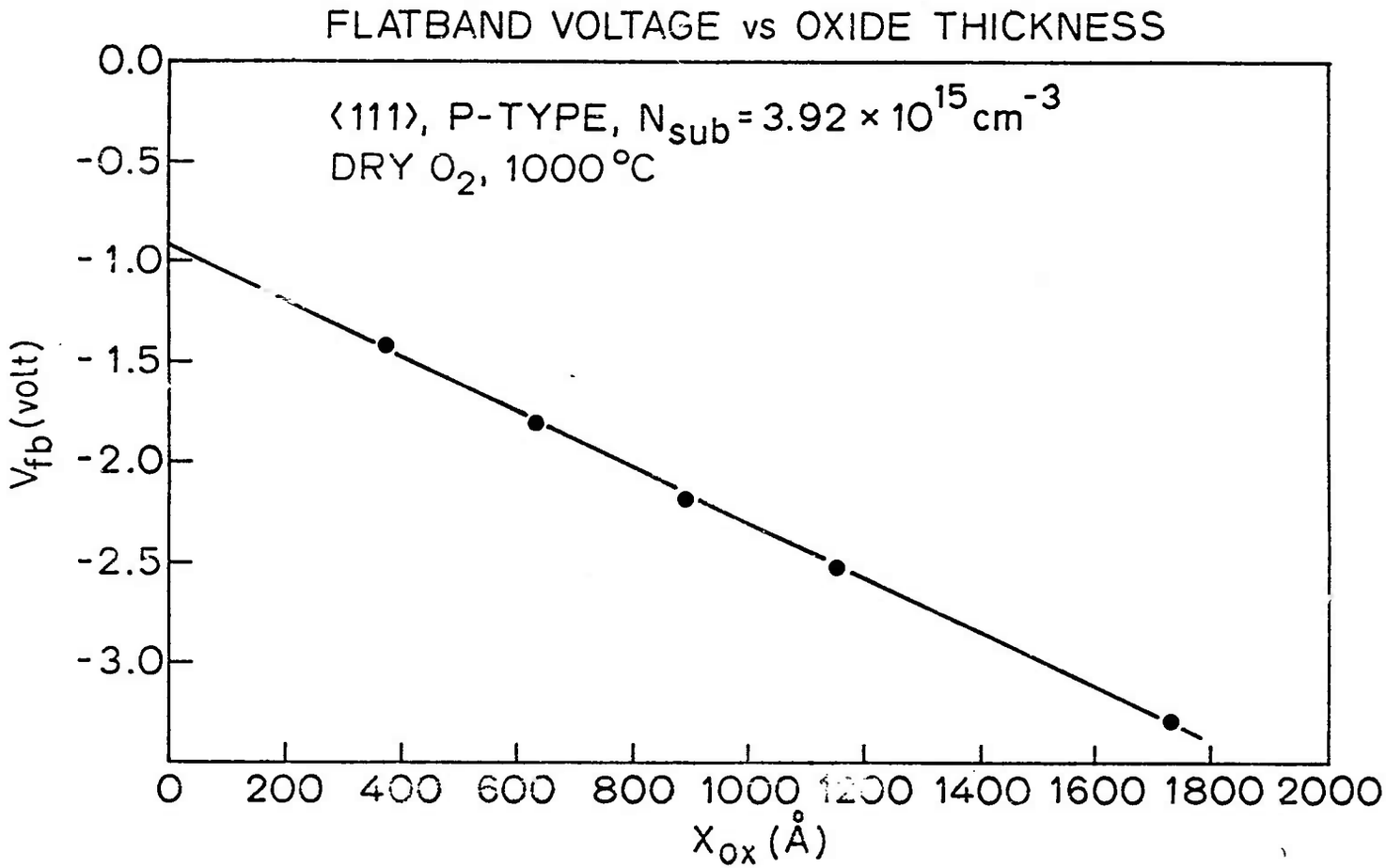


Fig. 1.27. Flatband voltage vs oxide thickness for (111) P-type wafers oxidized at 1000°C and fast pulled from O_2 ambient.

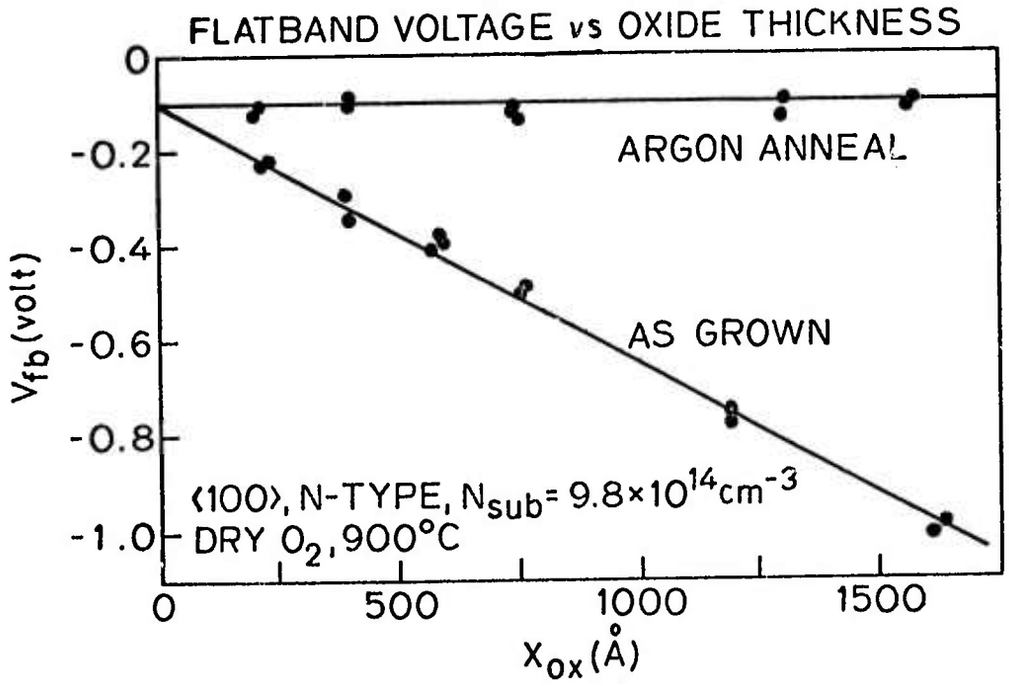


Fig. 1.28. Flatband voltage vs oxide thickness for (100) N-type wafers oxidized at 900 C and either fast pulled or Ar annealed.

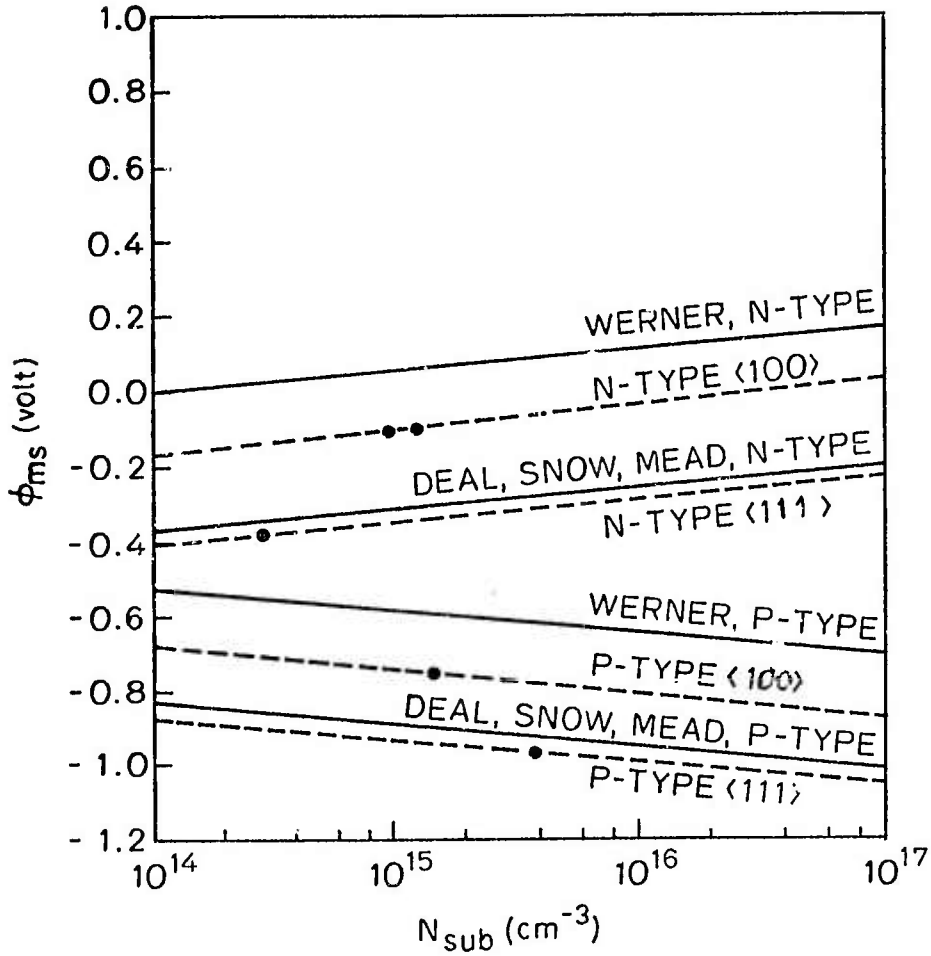


Fig. 1.29. ϕ_{ms} vs substrate doping for (100) and (111) substrates. The dashed lines correspond to this study.

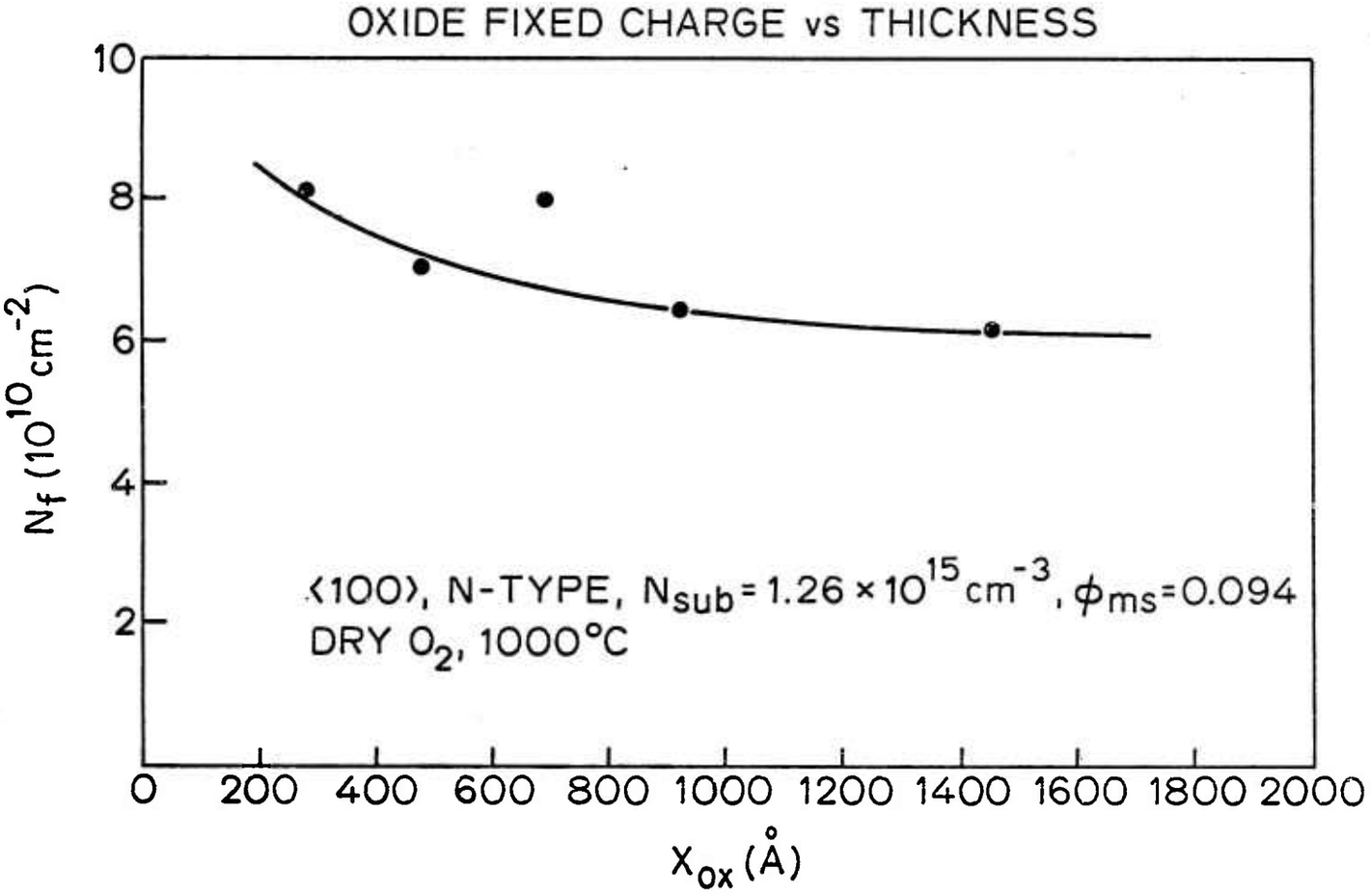


Fig. 1.30. Fixed oxide charge N_f vs oxide thickness for (100) N-type wafers oxidized in dry O_2 at 1000°C and fast pulled with no Ar anneal.

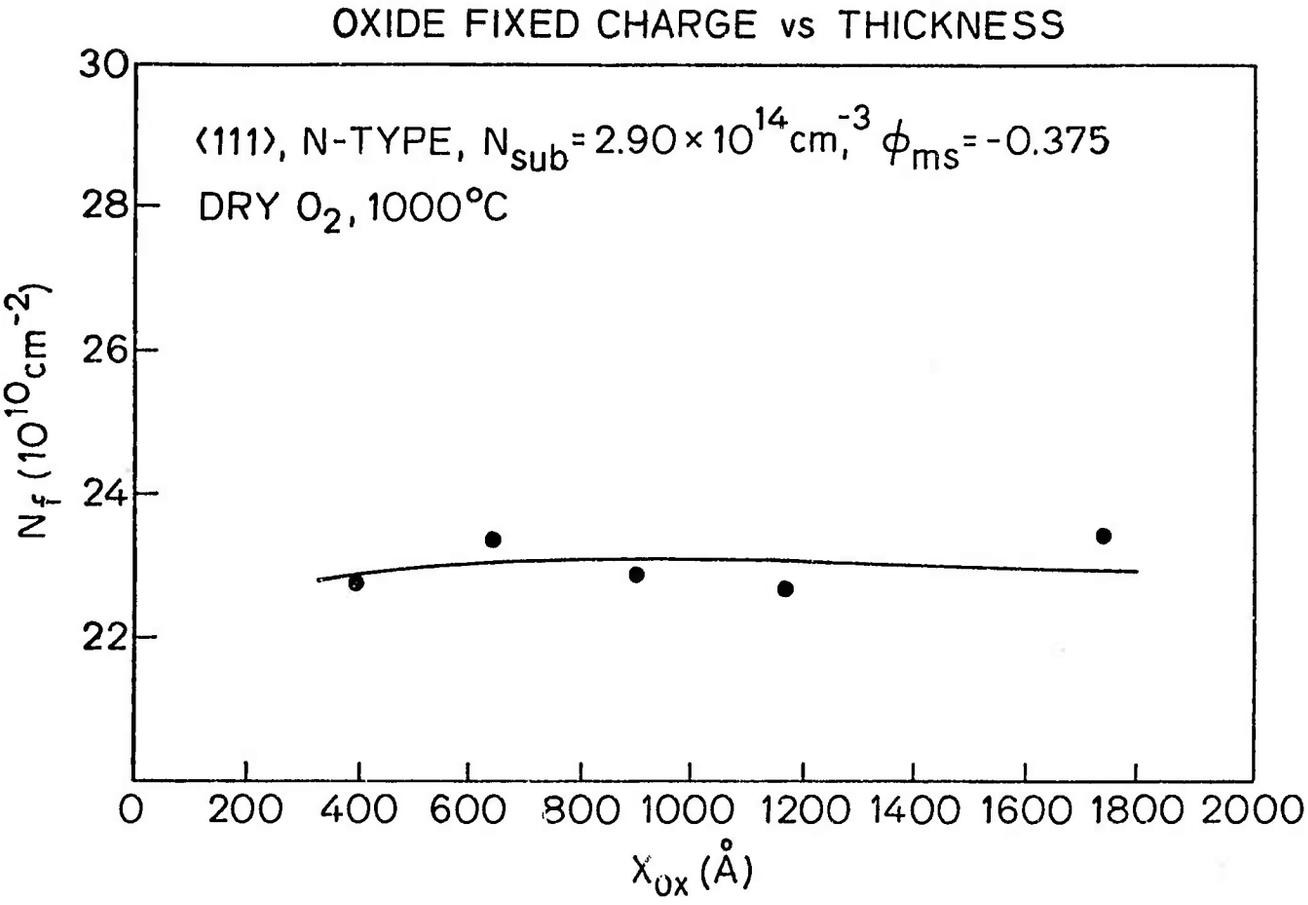


Fig. 1.31. Fixed oxide charge N_f vs oxide thickness for (111) N-type wafers oxidized in dry O_2 at 1000°C and fast pulled with no Ar anneal.

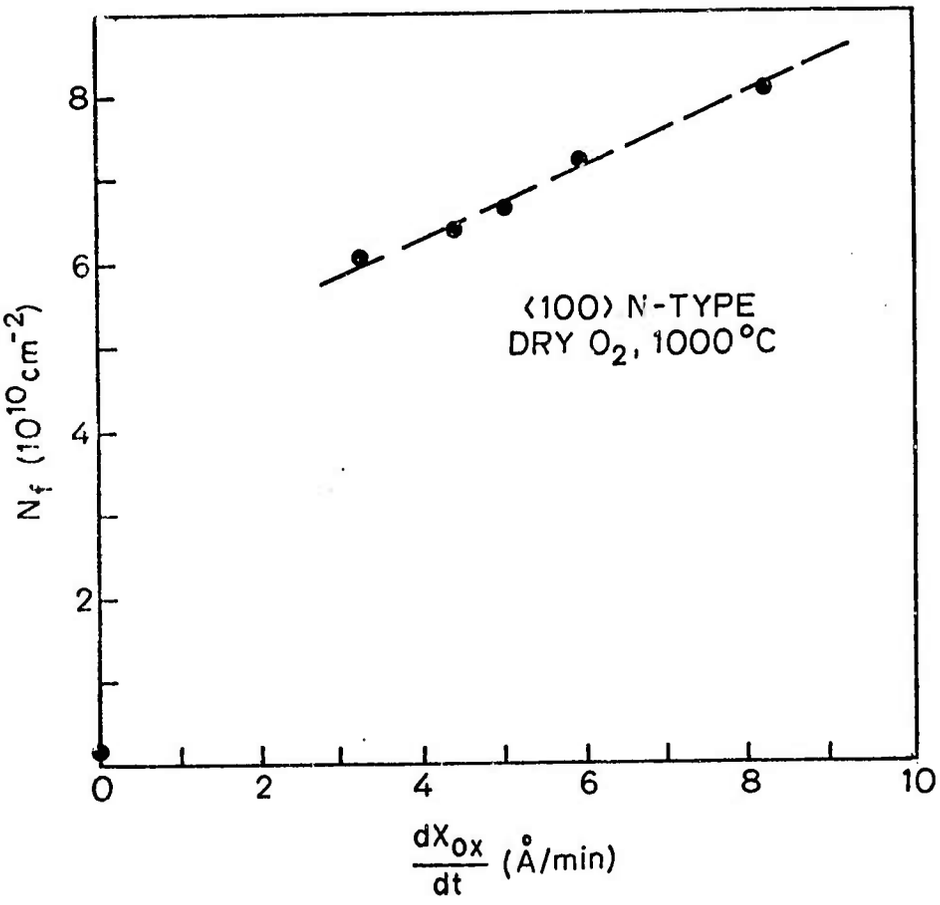


Fig. 1.32. Fixed oxide charge N_f vs oxide growth rate dx/dt for oxides grown in dry O_2 at 1000°C and fast pulled with no Ar anneal.

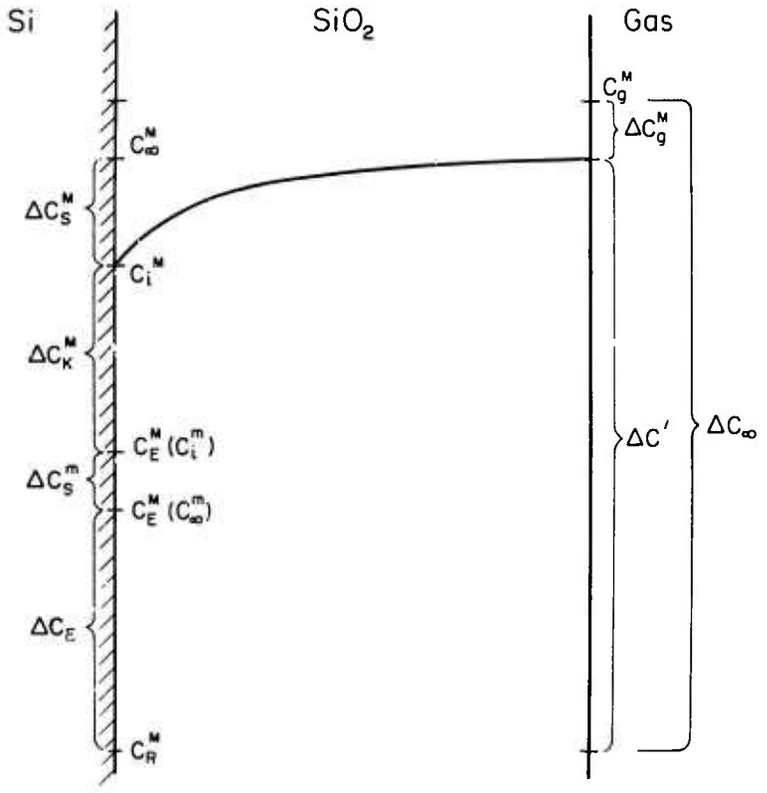


Fig. 1.33. Schematic representation of supersaturation coupling equation involved in the thermal oxidation of silicon.

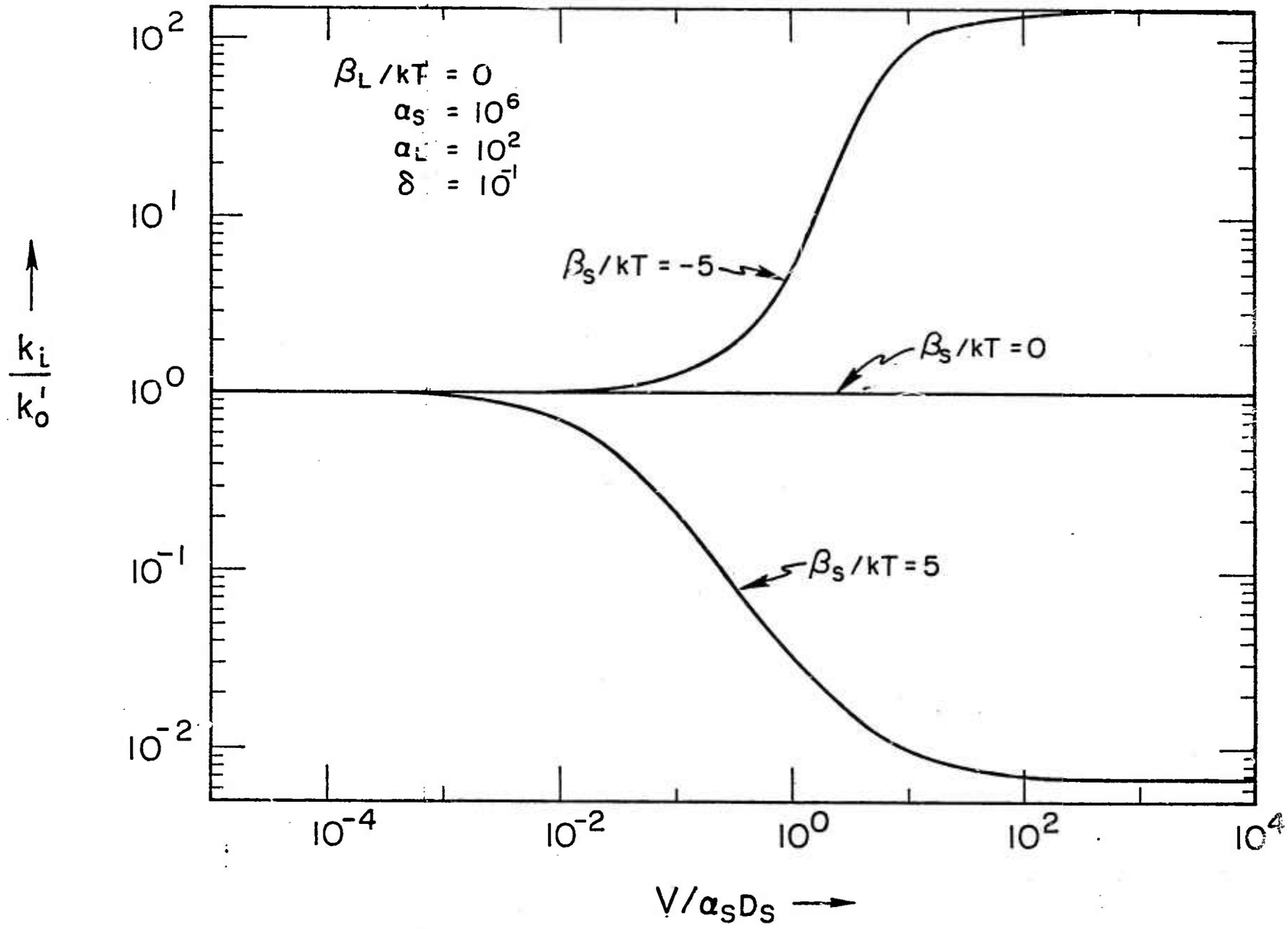


Fig. 1.34. Calculated plots of k_i/k_o' vs $V/\alpha_s D_s$ illustrating the importance of β_s/kT for the value of α_s and α_L ($D_s = 10^{-8}$ cm/sec, $D_L = 10^{-5}$ cm^2/sec).

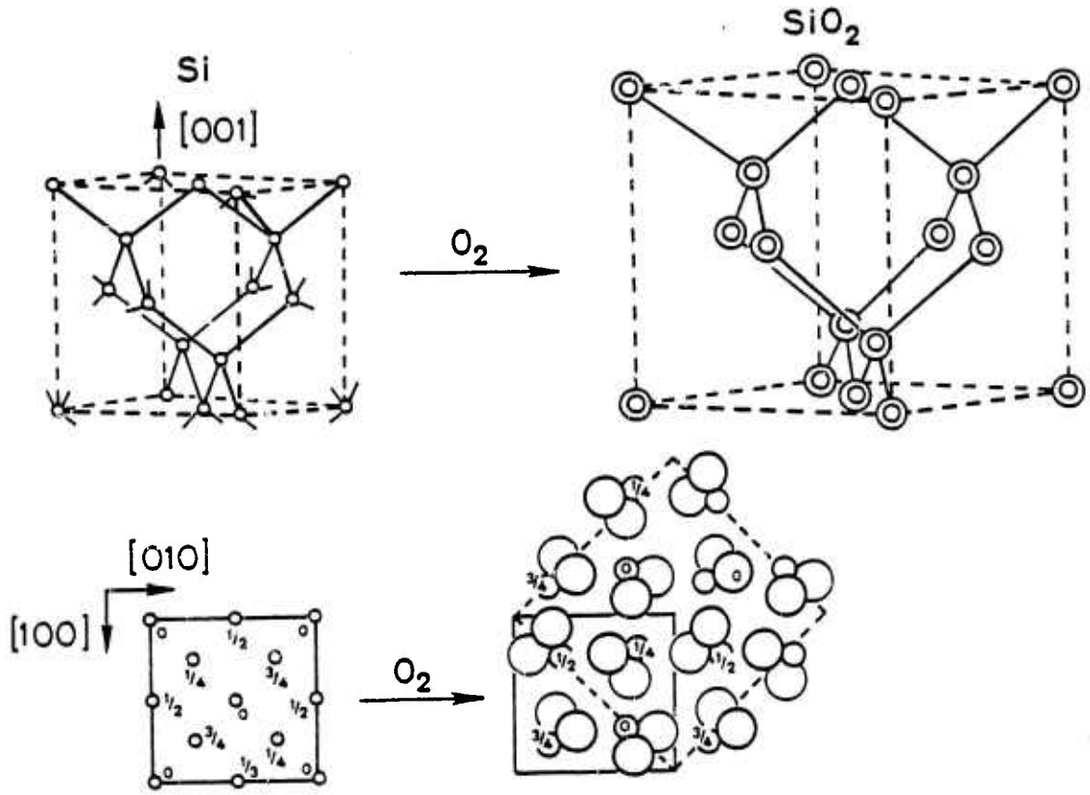


Fig. 1.35. Projection of the Si and the α-cristobalite unit cells onto their (100) planes. The dashed square represents the pseudo diamond cubic unit cell of α-cristobalite.

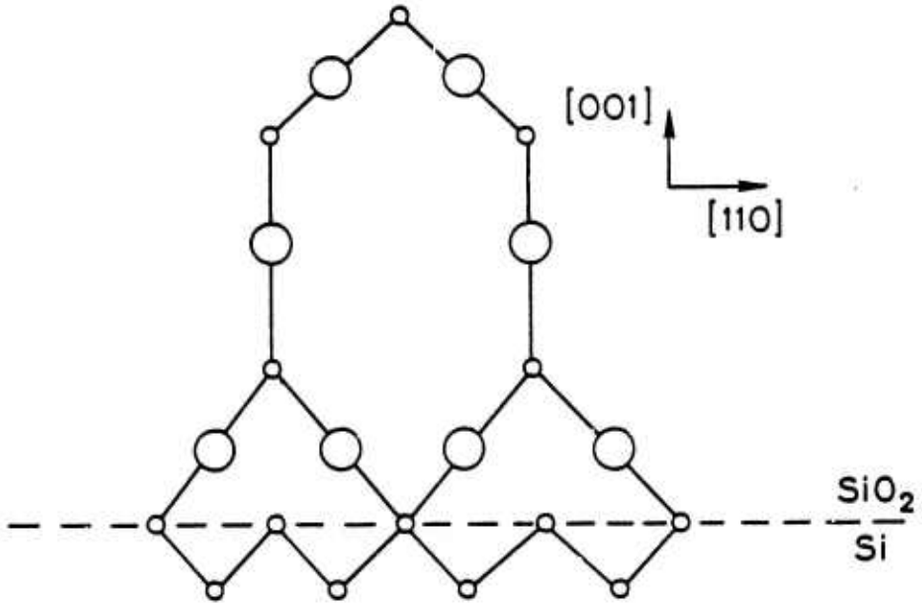


Fig. 1.36. Projection of Si/SiO_2 interface cross section illustrating the half-planes of Si that terminate at the interface with α-cristobalite.

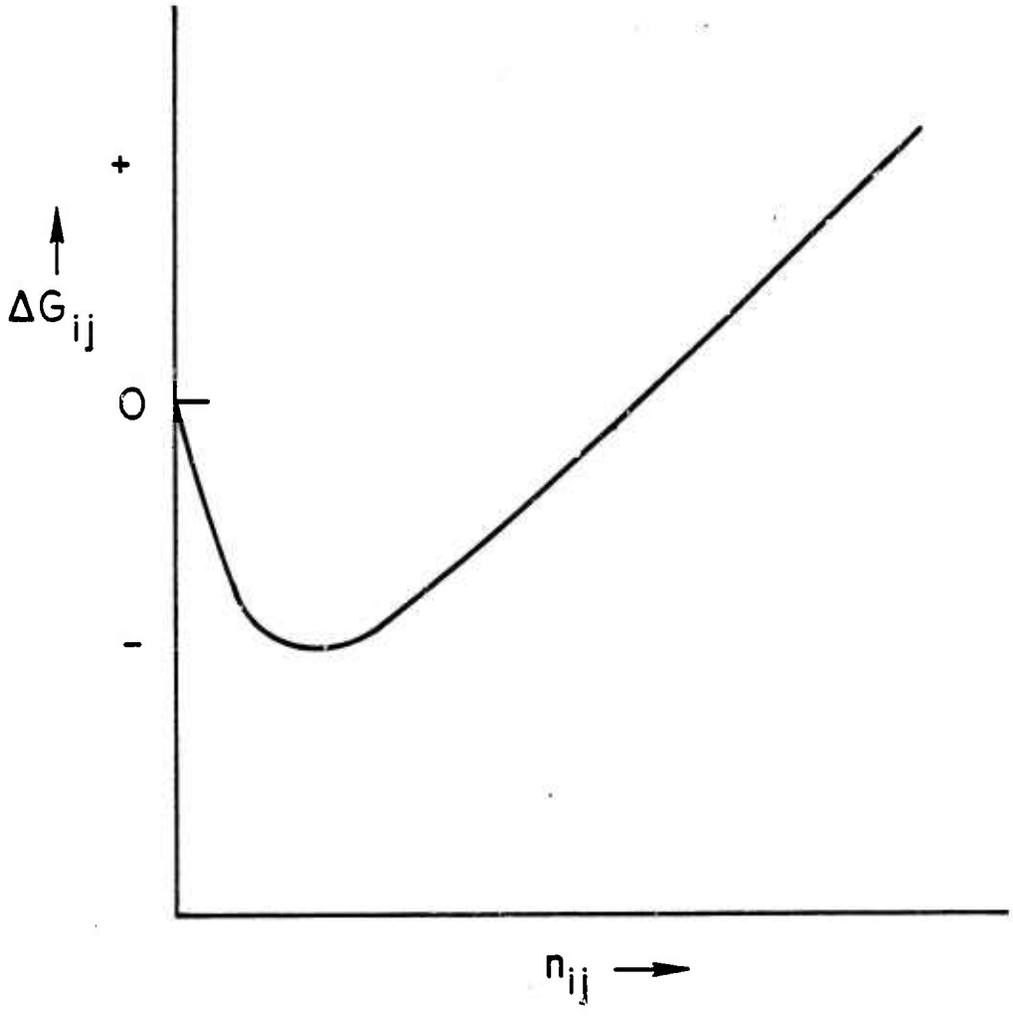


Fig. 1.37. Excess free energy of formation, ΔG_{ij}, of i-type defects in phase j as a function of number of defects per unit cell, n_{ij}.

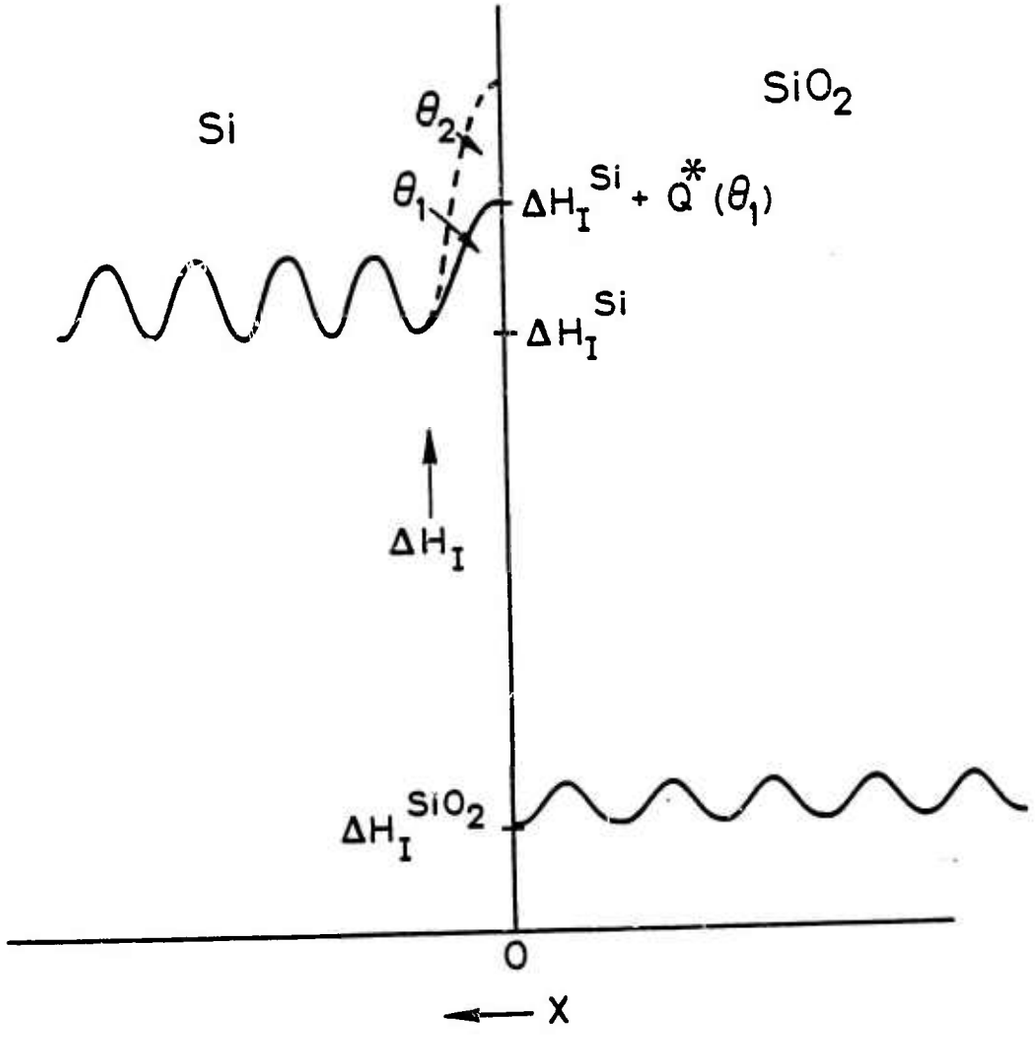


Fig. 1.38. Schematic illustration of the enthalpy of formation, ΔH_I, for I-defects as a function of distance from the interface. Also included is the interface barrier for two different Si wafer orientations θ_1 and θ_2.

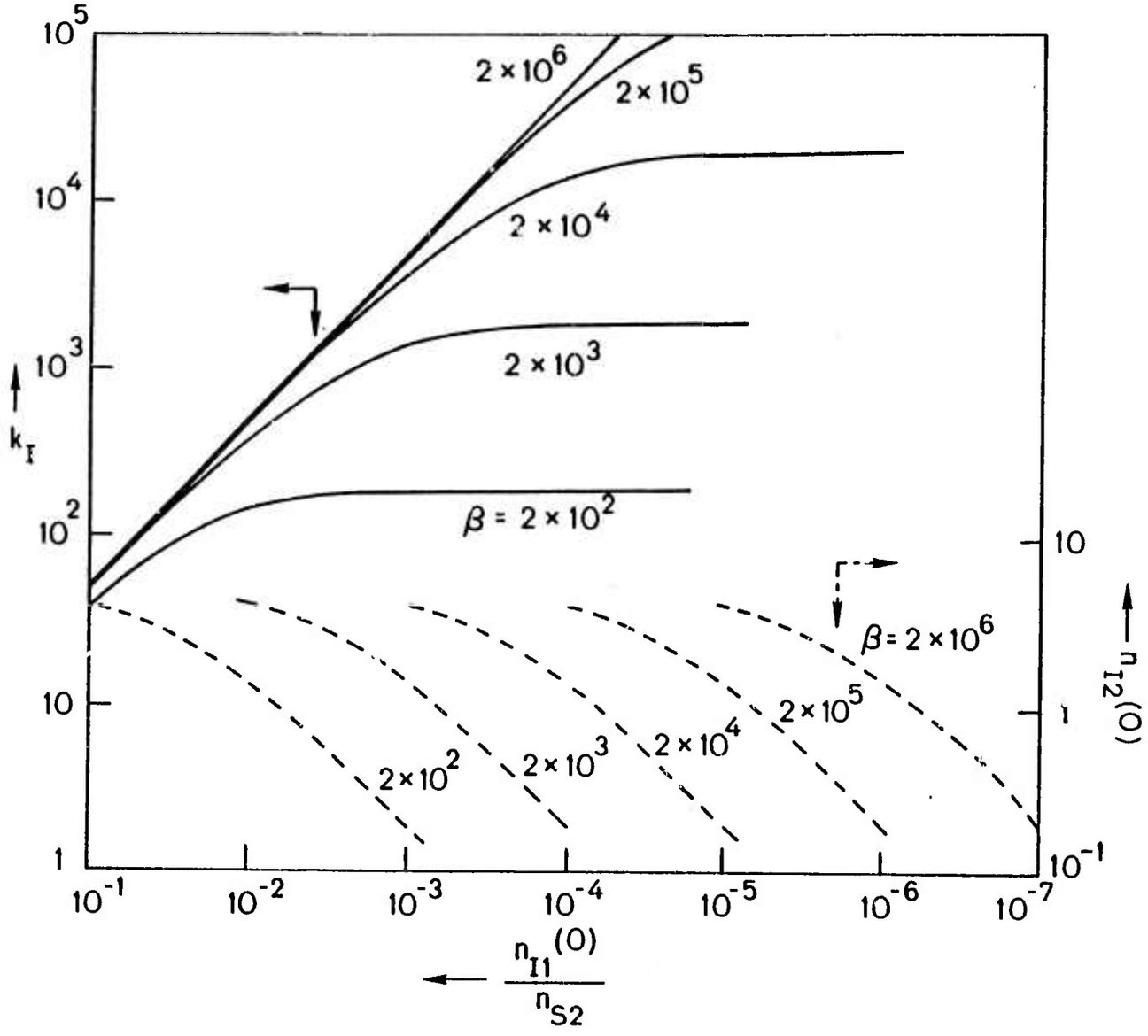


Fig. 1.39. Plot of k_I and $n_{I2}(0) = k_I n_{I1}(0)$ as a function of $n_{I1}(0)/n_{s2}$ for several values of the parameter β.

2. ION IMPLANTATION

L. Christel, G. Bronner, C. Ho, J. Plummer, J. Gibbons, S. Mylroie

2.1 Introduction

Ion implantation has found extensive use in introducing precisely controlled quantities of impurity-dopant atoms at desired depths into the target substrate in silicon integrated circuit technology. The reproducibility and uniformity across the wafer, and from wafer to wafer, are essential for such applications as boron implants for threshold control in the silicon-gate n-channel MOS transistor illustrated in Fig. 2.1. Further, as such devices are scaled to very small dimensions with gate oxide thicknesses approaching 200 Å or less, the requisite shallow drain-source junctions are achievable with and perhaps practical only with ion implantation.

As implantation applications grow in number and significance, detailed understanding and accurate modeling of the process and its applications correspondingly become more essential. Such modeling also can no longer be considered in a one-dimensional "isolated" sense. Rather, as device dimensions shrink both laterally and vertically, interactions of implantation with other processes involved in IC technology may dominate such very small device structures. Two-dimensional modeling of such interactions becomes vital and additionally will contribute enormously to understanding of the basic physical mechanisms of those other processes themselves.

The overall goals of the implantation modeling work discussed here address these needs:

- Process models for direct calculation of range and energy distributions of implanted species in multilayer targets, including implanted ion profiles, recoil implants ("knock-ons"), residual channeling, and damage distributions.

- Process models for annealing behavior of implanted species in silicon, including damage annealing, electrical activation, and dopant profiles resulting from interactions of dopants with silicon point defects.

- Process models for the roles of implant damage/silicon point defects in other IC processes and device behavior, including residual damage after annealing, traps produced by knock-ons, and interactions with oxidation kinetics, oxide charge densities, dopant diffusion, and epitaxy on ion implanted silicon.

- Incorporation of these models in a user-oriented process modeling program (SUPREM) to allow simulation of very small geometry device structures.

Following the work of Lindhard, Scharff and Schiott (LSS) [2.1] a great deal of progress was made in the characterization of the implantation process. The LSS theory is formulated in terms of a probability density $p(r,E)$ for ions of energy E to travel a range r. Conservation of probability implies an integro-differential equation for p. In order to eliminate spatial variables, an infinite homogeneous target is assumed and spatial moments of the equation for p are taken. The resulting equations may be solved for the first few moments (e.g., mean and standard deviation) of the range distribution. Approximate concentration profiles may then be constructed from these moments. Calculations for numerous ion-target combinations have been tabulated [2.2, 2.3] and the resulting profiles agree well with appropriate experiments in most cases.

Unfortunately the LSS approach suffers from several distinct drawbacks. Specific areas of the implantation process itself as well as its implications and interactions have not been well understood or adequately modeled.

As in the threshold control application already noted in Fig. 2.1, practical device processing very frequently involves implantation into target structures composed of a substrate on which one or more thin films of varying composition have been grown or deposited. LSS theory assumes an isotropic or amorphous semi-infinite target. The effect on ion range of implantation into such multilayer target is not considered. Also not modeled is the experimentally

observed dopant profile's deep channeling tail that results from ions scattered into the relatively unobstructed open "channeling" directions of the crystalline silicon substrate. Further, the secondary implantation of "knock-on" atoms from a masking surface layer into the substrate as recoils from collisions with the primary ion beam critically affect electrical properties and device performance. This "recoil implantation" process also cannot be analyzed by LSS theory.

The substrate itself can be severely damaged by implantation. An initially crystalline substrate may be driven amorphous or large densities of vacancy-interstitial point defect pairs created. Such damage distributions are integrally involved in implant annealing and activation, but have not been adequately modeled. An approach similar to that of LSS has been used by Winterbon et al. [2.4] to calculate energy deposition (damage) profiles. In this case however the asymmetry of such profiles necessitates the use of many moments for an accurate description. In addition, non-nuclear effects including electronic stopping must be neglected. Brice [2.5] has circumvented these problems to some extent by using a variation of the LSS moments approach to determine the more symmetric intermediate range distributions and then calculating energy deposition profiles directly from these. Unfortunately such approaches again require the assumption of a homogeneous target and therefore fail to address the problem of multilayered media.

Finally, as suggested by the developing view of the central role of point defects in most high temperature processes involved in device fabrication (for example, see the Oxidation section for related work), such implant damage may interact with and substantially influence those other processes both in the implanted region and in adjacent areas of the device wafer. Yet, the nature and magnitude of these interactions have not been characterized.

As device geometries diminish and new structures develop, understanding and modeling of all these aspects of implantation, as in SUPREM, will aid immeasurably in optimizing processes and minimizing costly and time-consuming empiricism.

2.2 Summary of Major Accomplishments to Date

During the past several years, emphasis in this project had been oriented very strongly toward modeling of the detailed mechanisms of the implantation process itself. A very promising new approach for direct calculation of primary ion and recoil range distributions that is ideally suited to multilayer target implants was one major development. By numerical integration of the linearized Boltzmann transport equation, the implant range of relatively massive ions such as As into SiO_2/Si target was successfully modeled [2.6].

In the past 18 months elaboration of the transport equation model and implementation in a computer program have remained a major thrust of the work. Extension of the Boltzmann formalism to enable light ion implant (wide angle nuclear scattering) as well as knock-on, channeling, and damage distribution calculations have been described in last year's annual report [2.7] and in a recent paper submitted for publication [2.8].

Effort of this past year has involved continued program development and refinement through detailed comparison of calculations with experimental results. A second major direction concerning the applications and interactions of implantation and implant damage has recently also been initiated. Activity of this year has been highlighted by the following:

- Extension of the Boltzmann transport equation (TE) modeling program, including incorporation of an improved hybrid nuclear scattering cross section, wide angle nuclear scattering, and electronic scattering [2.8].

- Detailed comparisons of TE calculations of primary ion range, recoil range, and damage density distributions with experimental data and other calculations (where applicable). The TE approach is demonstrated to be at least as accurate as other methods in those cases where comparisons can be made, and considerably more powerful and varied in its application to problems of practical interest [2.8].

- Initial simulations of implanted ion channeling, with some encouraging correlation to experimental observations.

- Initial demonstration and efforts to define more concretely the scope and magnitude of implant damage effects specifically on silicon oxidation kinetics and oxide charges in the thin oxide regime.

- Fabrication of shallow implanted junctions to attempt characterization of annealing behavior, minimum realizable sheet resistances, and implant damage gettering effects.

- Development of a high temperature-resistant electrical contact/interconnect technology to facilitate transient characterization of implant dopant activation, damage annealing and interactions with other high temperature transient processes.

These recent activities will be discussed in detail in the next section. For reference, the Boltzmann TE modeling approach will first be reviewed briefly below. (Additional detail may be found in [2.7, 2.8].)

In a transport equation approach quantities of interest are determined by calculating in a step-wise fashion the momentum distributions of both primary and recoil particles as a function of distance z from the target surface. It is assumed that the target is amorphous or aligned in a random direction and possesses translational symmetry in all directions parallel to its surface. Thus only two components of momentum, or equivalently an energy and direction cosine, are necessary to define a particle's state of motion. Let $F(p,z)dp$ be the number of particles (integrated over time) with momentum in the region dp about p and depth z. There will be one such distribution function for each projectile type. The spatial evolution of each distribution is governed by the Boltzmann transport equation:

$$\frac{\partial F(p)}{\partial z} = N \int \left[\frac{F(\vec{p}')d\sigma(\vec{p}'\rightarrow\vec{p})}{\cos\theta_{\vec{p}'}} - \frac{F(\vec{p})d\sigma(\vec{p}\rightarrow\vec{p}')}{\cos\theta_{\vec{p}}} \right] + Q(\vec{p}) \qquad (2.1)$$

Here for simplicity we assume a single element target with number density N. The total differential scattering cross section $d\sigma(\vec{p}\rightarrow\vec{p}')$ represents a differential area presented to an incident ion by each target atom. Any ion which enters this area with momentum p is scattered to momentum $\vec{p}'$. A particle which scatters to a momentum (or energy) below some fixed limit is considered stopped and is removed from the momentum distribution. In this way the range profiles of particles are generated. The factors of $\cos(\theta)$ result from the fact that a particle moving at an angle θ with respect to the z axis travels a distance $dz/\cos(\theta)$ as the entire distribution moves the distance dz. The quantity Q is a generation term which (for recoil distributions) allows particles to be created from rest.

The distribution functions are assumed known at the surface plane $z = 0$. Recoil distributions are identically zero there and the momentum distribution of the primary ion is taken to be a delta function:

$$F(\vec{p},0) = \Phi\delta\,(\vec{p} - \vec{p}_o) \qquad (2.2)$$

where Φ is the total dose (per cm^2) and $\vec{p}_o$ is the (unique) momentum of ions in the incident beam. With these initial conditions the transport equation is integrated numerically to determine the distributions at all depths $z > 0$. Crossing a material interface during the integration, as occurs in the case of multilayered target, necessitates changing only the number density and cross section to quantities appropriate to the new material. No changes in the momentum distributions occur because such distributions are continuous functions of z for any target composition. Such an approach is thus easily applied to targets of practical interest.

To apply the transport equation (TE) effectively to problems in ion implantation some additional assumptions must be made. These assumptions are not inherent in the TE approach but arise from our limited knowledge of the scattering process itself and are typically made in all methods. In particular we assume (1) the scattering cross section $d\sigma$ is the sum of an elastic nuclear part $d\sigma_n$ and an inelastic electronic part $d\sigma_e$, (2) scattering consists of a series of independent binary events so that contributions from elements in a multielement target may be simply summed, (3) electronic stopping acts only as an energy dependent drag force which removes energy continuously from the distribution, (4) only nuclear events generate recoils, produce lattice disorder, or cause significant angular deflections, (5) classical two body mechanics may be used to describe the angular deflections of both incident and recoil particles, and (6) for each nuclear event involving energy transfer T (assumed much greater than the binding energies of lattice atoms), a particle of energy T is created in the appropriate recoil distribution consistent with assumption 5.

2.3 Detailed Discussion of Accomplishments During the Period March 1979-February 1980

2.3.1 Extension of the TE Modeling: the Scattering Cross Section

The results of such TE calculations have been found to be quite sensitive to the differential nuclear cross section used. The scattering function $f(t^{1/2})$ used in $d\sigma_n(t)$ is determined by parameter constants λ, m, and q [2.9, 2.10]. The constants $\lambda = 1.309$, $m = 0.333$ and $q = 0.667$ were originally fit to an interatomic potential based on the Thomas-Fermi model of the atom and have been used in most work to date. However, Kalbitzer [2.11] has proposed that the parameters $\lambda = 2.54$, $m = 0.25$ and $q = 0.475$ will provide a better fit to a large collection of experimental data taken under

conditions where nuclear stopping dominates (i.e., when the LSS-normalized energy $\varepsilon < 1$). The two scattering functions differ most in the low energy region and are compared in Figure 2.2. We believe that the Kalbitzer cross section is better at low energy transfers but that near $\varepsilon = 1$ values closer to the Thomas-Fermi cross section should be used. We have therefore constructed a hybrid section which agrees with that of Kalbitzer at low energy transfers but evolves continuously into the Thomas-Fermi cross section at higher transfers. The mark shown in Fig. 2.2 gives the approximate point where the hybrid crosses over from Kalbitzer to Thomas-Fermi. A cross section which is very similar to this hybrid but which has a more fundamental basis was recently proposed by Wilson, et al. [2.12]. Calculations using the various cross sections are discussed below.

2.3.2 Accuracy of TE Modeling Calculations of Primary Ion, Recoil, and Damage Distributions

In this section we present comparisons of TE calculations with both LSS-based calculations and experimental results. Several examples of varying complexity have been chosen in an attempt to study the overall accuracy and applicability of the TE approach. Broadly speaking, the results obtained are found to be at least as accurate as those obtained using other methods in those cases where comparisons can be made. In addition the TE approach is considerably more powerful and varied in its application to problems of practical interest. Unfortunately, the results cannot be compactly codified as is the case for the LSS approaches where tables of moments have been prepared.

As our first example we consider the range distributions of three common dopants (arsenic, phosphorus and boron) in semi-infinite silicon targets. These three ions represent a fairly wide range of mass ratios and will provide a reasonable test of the consistency of calculations.

In Fig. 2.3 TE calculations using both the Kalbitzer and hybrid cross sections are compared with a Pearson type IV distribution generated from LSS moments for a dose of 10^{16} arsenic per cm^2 at 355 keV. Also shown is experimental data from Hirao, et al. [2.13] of arsenic implanted into single crystal silicon substrates which were misaligned by 8° from the (111) direction. For direct comparison the Kalbitzer cross section was used in the LSS calculation. Under these conditions ($\varepsilon \approx 0.5$) we expect nuclear scattering to be the dominant stopping mechanism and hence sensitivity to the nuclear cross section is most effectively examined. The figure shows that the TE and LSS calculations using the Kalbitzer cross section agree well near the peak of the distribution. However the LSS profile is more skewed, and both the TE and LSS calculations yield profiles that are somewhat deeper than the experimental data would suggest. By comparison, the TE calculation using the hybrid cross section is seen to agree quite well with experiment. The disagreement which occurs in the deepest part of the distribution is most likely due to channeling effects.

Figures 2.4 and 2.5 show similar results for phosphorus at 160 keV and boron at 100 keV. The experimental results for phosphorus are also from Hirao [2.13] and data for boron in amorphous silicon is from Hofker et al. [2.14]. For these lighter ions, electronic stopping is more significant and since the electronic cross section is not as well known, disparity between theory and experiment is expected to increase. In general, however, it can be seen that the hybrid cross section gives better agreement than the pure Kalbitzer cross section and that any error in the TE results is comparable to that of the LSS calculations.

As our second example we consider the damage density distributions resulting from implantations of antimony and boron into silicon. Only energy

deposited into nuclear scattering events is included in these profiles since it is believed that only such events contribute to the production of lattice disorder. Because recoil atoms transport some deposited energy to greater depths, we must distinguish between the as-deposited energy deposition profile and the final damage distribution. In the TE calculation the as-deposited profile is calculated by summing all nuclear energy transfers which occur at each depth. The final damage density profile is determined by assuming that the number of displaced substrate atoms which come to rest at any depth is proportional to the final damage density at that depth. Thus the recoil range distribution is a measure of this final damage density profile.

TE calculations of the as-deposited and final damage density profiles for 100 keV antimony implanted into silicon are shown in Fig. 2.6. The as-deposited damage (solid line) is a maximum at the sample surface and decreases smoothly to zero in a distance equal to about twice the projected range of the antimony. This is the expected behavior for nuclear energy loss when $\varepsilon \lesssim 1$ since under these conditions the nuclear stopping power decreases monotonically with decreasing energy.

The final damage density (proportional to the recoil range distribution) is also shown in Fig. 2.6 together with a calculation by Brice [2.15] and an experimental measurement of lattice disorder by Bøgh [2.16]. The TE recoil distribution is normalized to the peak of the experimental data. In the calculation by Brice it is assumed that recoil atoms deposit their energy in a distributed manner along their path while the TE calculations associate a damage with the final position of the recoil atom. The TE calculations are thus expected to yield somewhat deeper final damage distributions and this is verified by the figure. It may be the case that other quantities which are easily extracted from the TE calculations,

such as the concentration of vacancies, are more appropriate in characterizing the final damage and this possibility is being investigated.

Similar calculations of the as-deposited energy deposition profile for 100 keV boron are compared in Fig. 2.7. Also shown in this figure is the effect of neglecting those nuclear events which transfer less than 30 eV of energy. Such a procedure may be desirable if one wishes to account for the fact that a minimum amount of energy is required to remove an atom from its lattice position. The Brice calculation includes a correction for recoil energy loss due to electronic processes and is shown only to display the similarity between the two calculations. We note however that the rapid decay of the damage profile at large depths which occurs in the TE result is more consistent with the high degree of skewing and associated sharp decay which is observed in boron range distributions.

In the boron case ($\varepsilon = 11$) the calculations of Brice indicate that recoil transport of as-deposited energy is significant only near the sample surface. However, if one associates damage density with the recoil range distribution as in the TE calculations, significant transport of the as-deposited energy to greater depths is observed even in the boron case. In Fig. 2.8 we compare the two calculations. The Brice calculation has been normalized so that the total integrated damage agrees with the TE result. One can see that the TE calculation yields a significantly deeper final damage distribution.

The previous examples have shown that TE calculations give results which are consistent with more conventional calculations. We now turn our attention to several examples of situations which cannot be directly addressed at all by other methods, namely distributions in multilayered targets.

We begin by considering an implantation of arsenic into a target consisting of a silicon substrate coated with silicon dioxide (SiO_2). During the implantation, oxygen atoms will be recoil-implanted from the oxide into the silicon. Since oxygen is a deep level in silicon it is of interest to determine the total number of oxygen atoms (per cm^2) which ultimately reside in the silicon substrate. For a given initial ion energy, this number will be a function of the oxide thickness. Goezberger, et al. [2.17] have used surface state techniques to measure the dependence of total oxygen recoil flux on oxide thickness for an implant of 5×10^{13} cm^{-2} arsenic at 100 keV. We have calculated the same quantity using the Kalbitzer cross section with the results given in Fig. 2.9. As can be seen from the figure the TE results are in good agreement within the experimental error. Such a calculation can be very valuable in assessing the merits of coventional implantation processes.

The most recent calculations have concentrated on the oxygen recoil range distributions resulting from boron, phosphorus and arsenic implants into SiO_2/Si structures. In Fig. 2.10 an example for 180 keV As into 550 Å SiO_2 on Si is compared with the experimental SIMS results of Hirao et al. [2.18]. It can be seen that agreement is quite good. Fig. 2.11 shows some typical results for boron into 750 Å SiO_2 on Si.

As a final illustration of the versatility of the TE method we have calculated the range distributions of silicon and nitrogen which have been recoil-implanted into gallium arsenide from a 1000 Å thick coating of silicon nitride during an implantation of 2.5×10^{12} cm^{-2} selenium at 400 keV. The result is shown in Fig. 2.12 where we see that the magnitudes of the Si and N concentrations are comparable to those of the primary selenium. In particular, the silicon tail at large depths is larger than the selenium concentrations and could easily be mistaken for a channeling tail of selenium.

Since silicon may act as an n-type dopant in gallium arsenide, the presence of recoiled silicon may have a significant effect on electrical properties of ion implanted GaAs devices. These results could also explain the observation of greater than 100% electrical activity under such conditions. It is clear that such recoil distributions represent a significant effect and must be considered when designing modern implantation processes. At present the TE method is the only technique which can be used to accurately calculate such distributions.

2.3.3 Simulation of Implanted Ion Channeling

Initial attempts were made at simulating the channeling of ions along directions of high symmetry in crystalline solids. The (110) direction of silicon was modeled by a configuation of six string potentials and the trajectories of individual ions at a given energy were studied as a function of position and angle of entry into the channel. These calculations were unfortunately only moderately successful in illuminating the mechanisms by which the exponential tails observed in concentration profiles arise. Some results showed a slope which was near the experimentally observed value, but the total number of particles contained in the calculated tail was an order of magnitude too small. While these results are nevertheless encouraging, further work is needed.

2.3.4 Implant Damage Effects on Thin Oxides

Silicon (100), 11-20 Ω-cm, n-type wafers were heavily damaged to the point of driving the sample surface amorphous by implanting Si^+ sequentially at the following energies and doses.

50 keV	4×10^{14} cm^{-2}
100 keV	8×10^{14} cm^{-2}
150 keV	1×10^{15} cm^{-2}

These samples were then oxidized with controls in both dry and wet O_2 ambients for various times and temperatures, and the resulting oxide thicknesses measured.

Results included those shown in Table 2.1:

Table 2.1

Oxidation Conditions	Oxide Thickness	
	Control	Si^+ Implant Damaged
O_2: 60 min @ 900°C	189 Å	69 Å
Wet O_2: 60 min @ 800°C	373 Å	390 Å

Apparently, the damage created in the silicon substrate by the implantation process itself, independent of any implanted impurity atom effects, can substantially alter oxide growth kinetics. Preliminary work to add impurity effects to the damage, via implantation of both relatively inert impurities such as Ar, F, and Cl and electrically active dopants such as B, P, and As, has also begun. Both retardation and enhancement of oxidation rates have been observed, in agreement with the few reports found in the literature [2.19, 2.20].

As was also observed in the literature, reproducibility has not been sufficient for quantitative modeling, emphasizing the transient nature of the multiple mechanisms likely participating. In view of the growing realization of the unifying role of silicon point defects in most high temperature processes (see for example the Thermal Oxidation section for related work), the net oxidation rate likely results from the intrinsic interface oxidation mechanism and the damage annealing/solid-phase epitaxial regrowth process [2.21, 2.22] competing for the point defects produced in the

implant damage. Further confusing the results is the influence of the electrical activation of the implanted impurities that in turn may then substantially alter the point defect statistics and therefore influence both oxidation and regrowth rates. Such implantation process parameters as beam current and substrate temperature, to the extent that they influence possible during-implant annealing and therfore the damage present at the onset of high temperature processing, would also seem to be crucial. All of these probabilities will need to be considered in attempting to clarify the effects of implant damage on and the role of point defects in thin oxide kinetics.

2.3.5 Shallow Implanted Junctions

As device geometries shrink, implantation is seeing increasing use for the fabrication of shallow junctions in, for example, the MOS drain/source regions illustrated in Fig. 2.1. The achievable quality of such implanted junctions in comparison to the deeper thermally diffused junctions they are replacing is of great practical interest. Particularly relevant are the sheet resistance and junction leakage of such layers.

Shallow phosphorus implanted layers were fabricated in (100), 11-20 Ω-cm p-type wafers and, after thermal annealing, compared electrically to deeper thermally deposited and diffused layers of similar total charge. While shallow implanted phosphorus layers will have limited utility compared to implanted arsenic layers, this initial study would suggest the promising directions and possible problems to be encountered in subsequent arsenic characterization.

The diffused junction control samples received a predeposition of 90 minutes at 950°C and drive-in of 100 minutes at 1000°C for an estimated junction depth of ~1.6 μ and ~5 x 10^{15} cm^{-2} total charge. Both a single implant

and tailored multiple implants [2.23, 2.24] were used. The multiple implants were intended to approximate a square profile (as shown in Fig. 2.13) for improved sheet resistance. Implant schedules and estimated as-implanted junction depths were:

a. single implant

100 keV	5×10^{15} cm^{-2}	$x_j \sim 0.35\ \mu$

b. multiple implants

35 keV	4.74×10^{14} cm^{-2}	$x_j \sim 0.62\ \mu$
90 keV	1.05×10^{15} cm^{-2}	
145 keV	1.53×10^{15} cm^{-2}	
200 keV	1.94×10^{15} cm^{-2}	

The implanted and diffused samples were annealed together for 4 hours at 800°C. Resulting sheet resistances of the various structures are indicated below in Table 2.2.

Table 2.2

	X_j	Ω/□
Diffused	1.6 μ	12
Multiple Implant	0.6 μ	16.7
Single Implant	0.35 μ	19

It can be seen that the implanted junctions compare reasonably well with the different junctions, especially in view of the rather poor control of the total charge in the diffused junctions. The estimated diffused charge could easily be in error by 10% or 20%, making the sheet resistance of the implanted structures possibly even more comparable to that of the diffused junctions. Junction leakage currents of the implanted

layers were higher than the diffused layer, but were found to be improved considerably by backside argon implant damage gettering [2.25, 2.26].

These results appear promising, and point to a number of crucial physical issues. Residual implant damage may have lowered the carrier mobility in the implanted layers, increasing the sheet resistance. Perhaps, there was incomplete electrical activation of the implanted species. Clearly, understanding of the transient annealing mechanisms of shallow implanted layers is needed. Similarly, the mechanism and time dependence of implant damage gettering should also contribute to a unified understanding and modeling of the role of point defects in most high temperature processes such as oxidation and diffusion as well as implant annealing. Data and models arising from work addressing these physical issues should be immediately useful in SUPREM.

2.3.6 High Temperature-Resistant Electrical Contact/Interconnect Metallurgy

A common theme has become apparent in many of the processes being characterized and modeled in this contract. The critical high-temperature mechanisms most challenging to modeling are transient and time-dependent in nature. This is certainly true of the thermal activation and damage annealing of shallow implanted layers and also of implant damage gettering issues addressed above. It is also characteristic of other processes such as oxide charge annealing and diffusion studies addressed in other sections of this report or closely linked with this work.

In most instances, experimental characterization or monitoring, often electrical in nature, has involved multiple samples corresponding to different points in the time course of these transient processes. Multiple samples, however, are time consuming, difficult to handle, and of questionable accuracy. Undesirable variation between samples other than the different

heat treatments received may be an unknown factor which is difficult to control. Thus, often very few samples or time points are used in the data base. Modeling then necessarily employs a time-averaging to find parameter values for SUPREM. Such values may have limited predictive value in SUPREM when new process variations are to be simulated.

Ideally, a single sample should be monitored vs process time and temperature "in situ," as, for example, in the use of automated ellipsometry to monitor oxidation kinetics [2.27, 2.28]. Alternatively, a single sample should be used and repeatedly heat-cycled. After each heat treatment, its characteristics could be measured and from these the progress of the transient mechanism inferred. In electrical techniques requiring electrical contact or interconnect, however, present aluminum metallization is not compatible with the high temperatures of most processes being characterized. An electrical contact technology capable of withstanding high process temperatures, and conventional chemicals used in sample preparation, would be extremely useful in this work. Additionally, this contact metallurgy should not interfere with the transient high temperature processes that will be monitored--its deposition must be relatively low temperature and the metallurgy should not introduce impurities, particularly electrically active ones, into the sample.

Heavily doped polysilicon, widely used as a gate material for self-aligned MOS device structures, is one possibility. However, it appears to have serious limitations. While low pressure deposition (LPCVD) at ~620°C is relatively low temperature, a high temperature doping process is still necessary to reduce resistivity to a usable range of 10^{-3}-10^{-2} Ω-cm [2.7]. Also, while doped polysilicon is stable on oxide for MOS structures, it has been used as an excellent diffusion source for doping single crystal silicon

substrates [2.29]. This suggests that it is unsuitable for directly contacting the substrate. Dopants introduced from the poly into the substrate at such contacts would seriously confuse characterization of electrical activation, diffusion, and other dopant and point defect mechanisms.

Refractory metals, another candidate that has been used for MOS gate electrodes [2.30-2.32], may also have limitations. Deposition can be low temperature, and resistivities such as 1.5×10^{-5} Ω-cm for Mo and 2.5×10^{-5} Ω-cm for W [2.33] are certainly more than adequate. They have high melting points, their thermal expansion coefficients match well with that of silicon, and they mask common silicon dopants. However, they are not necessarily stable at high temperatures. Studies of W films show that above 800°C WSi_2 starts to form. This formation often leads to cracking or lifting of the W film. $MoSi_2$ also forms by the heating of Mo films on Si. There are indications that this may occur at temperatures as low as 500°C [2.34]. Many people have reported the cracking, lifting, and poor adhesion of Mo and W to Si and SiO_2 [2.31, 2.33-2.35]. Finally W and Mo have both been reported to introduce trap levels into the silicon bandgap that could potentially interfere with electrical charcterization. Molybdenum has reported levels in silicon at [2.36]:

$$E_c - 0.33 \text{ eV}$$

$$E_v + 0.34 \text{ eV}$$

$$E_v + 0.3 \text{ eV}$$

For tungsten the following levels are reported [2.37]:

$$E_c - 0.25 \text{ eV}$$

$$E_c - 0.28 \text{ eV}$$

(Note that these trap levels do lie at a reasonable level from mid-gap. The generation-recombination currents due to these levels should not be too severe.)

Silicides of the refractory metals, on the other hand, appear quite promising (see the Silicide section of this report) for this proposed technology. They are stable at process temperatures, not tending to react with the underlying silicon. Their deposition, by CVD, coevaporation, or sputtering, is low temperature. While a high temperature anneal has been found desirable to reduce their resistivities to very low levels (~10^{-4} Ω-cm for $MoSi_2$ [2.38] and ~5 x 10^{-4} Ω-cm for WSi_2 [2.7, 2.39]) desirable for use in very small device structures of VLSI, their as-deposited resistivities (~7 x 10^{-4} Ω-cm for $MoSi_2$ [2.38] and ~6 x 10^{-4} Ω-cm for WSi_2 [2.39]) are clearly superior to heavily doped polysilicon and more than adequate for the intended use of process characterization and modeling.

(A potential problem with the silicides might arise from the presence of unreacted metal in metal-rich, non-stoichimetric "silicide" films. This metal might diffuse through junction contacts into the bulk silicon and adversely affect the junction properties. Still, as noted above, the metal-linked bulk trap levels should not create too great a G-R current problem. Also, use of stoichiometric or silicon-rich films may circumvent this hazard.)

So, development of an electrical contact/interconnect technology employing silicides was undertaken, clearly complementary to the work described in the CVD section of this project. It was necessary to check that silicides could actually be used without degrading device performance. Starting with [100], 0.1-1 Ω-cm, p-type silicon, both n^+-p junctions and MOS capacitors were fabricated. Tungsten silicide (WSi_2) was rf diode sputtered from a hot-pressed target made of the silicide. The capacitors and diodes were characterized after the silicide electrodes/contacts were deposited, and after each heat cycle in sequential one hour nitrogen anneals at 700°C, 900°C, and 1100°C.

Table 2.3 summarizes the MOS capacitor characteristics:

Table 2.3

	N_f (cm^{-2})	D_{it} ($cm^{-2}ev^{-1}$)
Al control	1.0×10^{11}	2×10^{10}
WSi_2 as-deposited	2.0×10^{11}	1×10^{11}
700°C	1.6×10^{11}	7×10^{10}
900°C	1.8×10^{11}	3×10^{10}
1100°C	1.2×10^{11}	2×10^{10}

We see that the quality of the Si/SiO_2 interface is not appreciably affected by the WSi_2. Over the course of the temperature cycling both N_f and D_{it} improve slightly, but even in the case of the as-deposited sample, their values are acceptable. It has been reported that sputtering may damage the oxide and Si/SiO_2 interface [2.40]; so, these affects could be due to the sputtering process and not inherent to the WSi_2.

One unexpected result of the MOS capacitor measurements was evidence that the as-deposited WSi_2 films were not clean. Bias temperature stress (BTS) measurements were performed by raising the substrate temperature to 250°C and then applying a positive voltage to the gate of the capacitor for five minutes. Comparison of the HF-CV curves before and after BTS revealed flat band voltage shifts in excess of 10 V. Thus, we see that the as-deposited WSi_2 has very large amounts of contamination. It is interesting to note that this contamination disappears after the 900°C anneal, although it is not clear what happens to it.

Because of this contamination, the results of the measurements on the n^+-p diodes were inconclusive. After annealing the diodes at 700°C

the reverse leakage currents did increase appreciably. See Fig. 2.14. However, the behavior of these leakage currents is not what one would expect for W contamination [2.37]. One expects leakage current proportional to $V^{1/2}$ due to the generation-recombination currents followed by breakdown. In this curve, however, we see the leakage current rising at a much faster rate. It is possible that this leakage may be due to the contamination seen in the films; however, that is not clear.

To resolve this question a new set of experiments were run, similar to the first except for the deposition system for WSi_2. For this run, the film was cosputtered in a different system from separate W and Si targets. These films were made slightly silicon rich to discourage the W from moving into the bulk silicon. These experiments are not completed, but preliminary results are encouraging. Specifically, the contamination which manifested itself in the form of shifts in capacitor flat band voltage under BTS seems to have disappeared. This indicates that such contamination is not inherent in the silicide films. Measurements on the p-n diodes have yet to be completed.

2.4 Conclusions/Future Plans

The following activities are anticipated in the current year and beyond as desired continuations of the very promising results in the two major directions of implantation modeling work reported above:

- Elaboration of the TE modeling approach and consideration of the steps necessary to
    a. implement the Boltzmann transport equation modeling of multilayer target implantation in an alternative form suitable for SUPREM
    b. analyze channeling effects more accurately
- Completion of the contact/interconnect metallurgy development and subsequent use of the technology for characterization of

very shallow implanted layers, electrical activation/damage annealing and gettering, and the interactions of implant damage in very thin oxide kinetics and charge properties.

2.4.1 TE Implementation in SUPREM and Channeling

Fairly accurate calculations of primary and recoil range distributions as well as damage distributions in multilayered targets are now possible. Such a calculation requires between 5 and 50 minutes of CPU time on the Stanford SCORE DECsystem20, depending on complexity.

The TE calculations in their present form are too complex to be incorporated within the SUPREM framework. An attempt is being made to produce a streamlined version of the program which will calculate only primary ion range distributions in the hopes that this can be incorporated directly as a subroutine in SUPREM III. Alternatively, calculations for common conditions will be made and then a means of interpolating between these will have to be found.

The channeling problem will require a lot of future work before these effects can be incorporated simply in a model. However, a strong first step has been made and this activity will be continued in the coming year.

2.4.2 Shallow Implanted Layers and Implant Damage Interactions

Development of the silicide contact/interconnect technology, in conjunction with the CVD portion of this project, will continue. Specifically, WSi_2 contacts to bulk silicon must be characterized more fully. Also, stability and feasibility of silicide electrodes over very thin oxides and substrate contacts to very shallow implanted junctions will be examined.

As this work is completed, experimental characterization of the electrical activation time-dependence, damage annealing behavior, and minimum realizable sheet resistance of shallow implanted layers will be facilitated. Complementing this work should be characterization of the time-dependence and

physical mechanisms of backside implant damage gettering presumably to getter undesired (metallic) impurities from the wafer frontside in which active devices would be fabricated. This will be done in conjunction with the very thin oxide MOS structures and the shallow implanted junctions on the front side to characterize effects on their electrical properties. Active devices may be incorporated in later stages of this study.

Understanding of these damage annealing and gettering mechanisms will in turn complement a study of the roles of implant damage/silicon point defects on other high temperature processes such as oxidation and diffusion. Initially, characterization of the influence of implant damage, both directly in the region to be oxidized and indirectly in laterally or vertically removed sections of the silicon, on very thin oxide kinetics and charge properties will be continued. Close cooperation here with the Thermal Oxidation portion of this work may also aid in clarifying understanding of the role of point defects in the "intrinsic" interface silicon oxidation mechanism, particularly for very thin oxide layers.

It should be expected that first practical data bases suitable for SUPREM implementation will be produced in the coming year. Understanding and modeling of the roles of implant damage and point defects in, and interactions with, all high temperature silicon processes should then be facilitated.

REFERENCES

[2.1] J. Lindhard, M. Scharff, H. E. Schiott, Mat. Fys. Medd. Dan. Vid. Selsk 33, No. 14 (1963).

[2.2] J. Gibbons, W. S. Johnson, S. Mylroie, Projected Range Statistics in Semiconductors, 2nd edition (distributed by John Wiley & Sons) 1975.

[2.3] Bernard Smith (Pergamon Press).

[2.4] K. B. Winterbon, Rad. Eff. 13, 215 (1972).

[2.5] D. K. Brice, Rad. Eff. 6, 77 (1970).

[2.6] D. H. Smith, J. F. Gibbons, Ion Implantation in Semiconductors, 1976, (Plenum Publishing Corp. 1977), p. 333.

[2.7] J. D. Meindl, et al., "Computer Aided Engineering of Semiconductor Integrated Circuits," Stanford Electronics Laboratories TR-77-26874-3, July, 1979.

[2.8] L. A. Christel, J. F. Gibbons, and S. Mylroie, "Application of the Boltzmann Transport Equation to Ion Range and Damage Distributions in Multilayered Targets," submitted to J. Appl. Phys.

[2.9] J. Lindhard, V. Nielsen, M. Scharff, Mat. Fys. Medd. Dan. Vid. Selsk. 36, No. 14 (1970).

[2.10] K. B. Winterbon, P. Sigmund, and J. B. Sanders, Mat. Fys. Medd. Dan. Vid. Selsk. 37, No. 14 (1970).

[2.11] S. Kalbitzer and H. Oetzmann, Proc. of Intern. Conf. on Ion Beam Modification of Materials (1978).

[2.12] W. D. Wilson, L. G. Haggmark, J. P. Biersack, Phys. Rev. B. 15, No. 5, 2458 (1977).

[2.13] T. Hirao, K. Inoue, S. Takayanagi, Y. Yaegashi, J. Appl. Phys. 50, 1, (1979).

[2.14] W. K. Hofker, D. P. Oosthoek, N. J. Koeman, H. A. M. DeGrefte, Rad. Eff. 24, 223 (1975).

[2.15] D. K. Brice, J. of Appl. Phys. 46, 3385, 8 (1975).

[2.16] E. Bøgh, P. Hogild, I. Stensgard, Proc. of the Symposium on Radiation Damage in Reactor Materials, Vienna, 1969, Vol. 77 (unpublished). Also, Proceedings of the First International Conference on Ion Implantation om Semiconductors, edited by F. H. Eisen and L. T. Chadderton (Gordon and Breach, London, 1970), p. 431.

[2.17] A. Goetzberger, D. J. Bartelink, J. P. McVittie, and J. F. Gibbons, Appl. Phys. Lett. 29, 259, 8 (1976).

[2.18] Hirao et al., J. of Appl. Phys. 50, 8, 5251 (1979).

[2.19] G. Mezey et al., Proc. of the Intl. Conf. on Ion Implantation in Semiconductors, Boulder, CO, 1976, p. 49.

[2.20] K. Nomura et al., Proc. of the Intl. Conf. on Ion Implantation in Semiconductors, Osaka, Japan, 1974, p. 681.

[2.21] L. Csepregi et al., J. Appl. Phys., 48, 4234 (1977).

[2.22] Y. Tamminga and W. J. M. J. Josquin, Appl. Phys. Lett., 32, 13 (1978).

[2.23] G. T. Marcyk and B. G. Streetman, "Boron Impurity Profile Tailoring in Silicon by Ion Implantation and Measurement by Glow Discharge Optical Spectroscopy," J. Electrochem. Soc., 123, 1388 (1976).

[2.24] A. J. Zarenba, et al., "Optional Summation of Gaussians for Ion Implantation Profile Control," IEEE Trans. Electron Devices, ED-24, 163 (1977).

[2.25] T. E. Seidel, et al., "Direct Comparison of Ion-Damage Gettering and Phosphorus Diffusion Gettering of Au in Si," J. Appl. Phys., 46, 600 (1975).

[2.26] A. G. Nassibian, et al., "Generation Lifetime Investigation of Ion-Damage Gettered Silicon Using MOS Structure," J. Appl Phys., 47, 992 (1976).

[2.27] M. A. Hopper, et al., "Thermal Oxidation of Silicon - In Situ Measurement of the Growth Rate Using Ellipsometry," J. Electrochem. Soc., 112, 1216 (1975).

[2.28] E. A. Irene and R. Ghez, "Silicon Oxidation Studies: The Role of H_2O," J. Electrochem. Soc., 124, 1757 (1977).

[2.29] J. Middelhoek and A. Kooy, "A Polysilicon Source and Drain MOS Transistor (PSD MOST)," IEEE Trans. Electron Devices, ED-23, 523 (1976).

[2.30] D. M. Brown, et al., "Self-Registered Molybdenum-Gate MOSFET," J. Electrochem. Soc., 115, 874 (1968).

[2.31] D. M. Brown, et al., "Retractor Metal Silicon Device Technology," Solid-State Electronics, 11, 1105 (1968).

[2.32] D. Brown, et al., "The P-Channel Retractory Metal Self-Registered MOSFET," IEEE Trans. Electron Devices, ED-18, 931 (1971).

[2.33] L. E. Terry and R. W. Wilson, "Metallization Systems for Silicon Integrated Circuits," Proc. IEEE, 57, 1580.

[2.34] Holmwood and Glang, J. Electrochem. Soc., 112, 827 (1965).

[2.35] V. Kumar, "Fabrication and Thermal Stability of W-Si Ohmic Contacts," J. Electrochem. Soc., 123, 262 (1976).

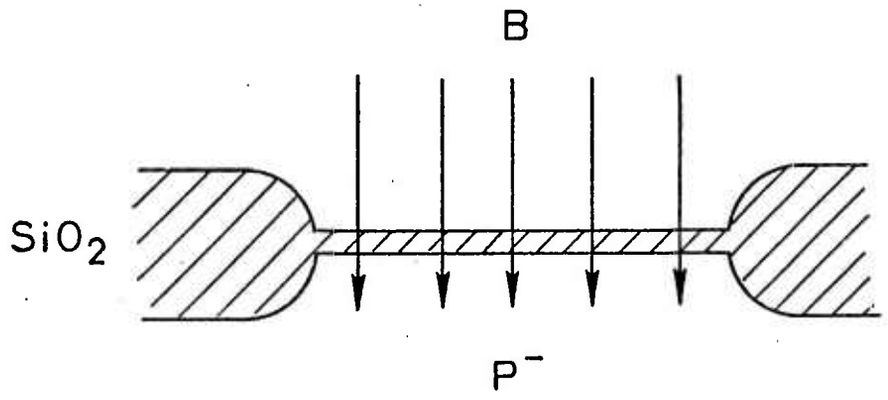
[2.36] A. Zibuts, et al., "Some Properties of Silicon Containing Mercury, Tungsten, Molybdenum and Platinum Impurities," Soviet Physics-Solid State, 5, 2416 (1964).

[2.37] H. H. Busta and H. A. Waggener, "Precipitation-Induced Currents and Generation-Recombination Currents in Intentionally Contaminated Silicon P^+ N Junctions," J. Electrochem. Soc., 124, 1424 (1977).

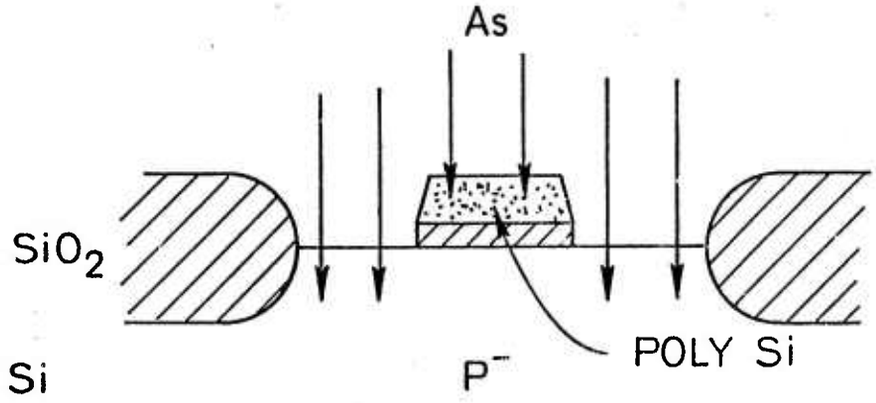
[2.38] T. Mochizutti, K. Shibuta, T. Inoue, and K. Ohuchi, Japan. J. Appl. Phys. Suppl., 17, 37 (1978).

[2.39] F. Mohammadi and K. C. Saraswat, "Properties of Sputtered Tungsten Silicide for MOS Integrated Circuit Applications," J. Electrochem. Soc., 127, 450 (1980).

[2.40] A. Rodriguez and M. Misra, "Fabrication of Short Channel MOSFETS with Refractory Metal Gates Using R_F Sputter Etching," Solid-State Electronics, 19, 17 (1976).



THRESHOLD CONTROL



SHALLOW JUNCTIONS

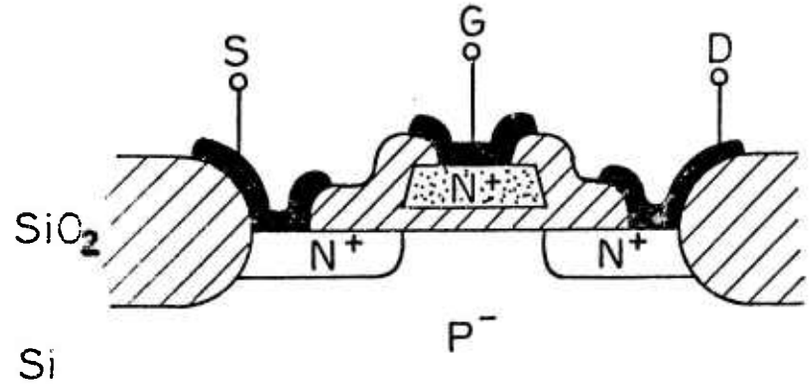


Fig. 2.1. Applications of ion implantation to very small geometry silicon gate NMOS transistor.

PRECEDING PAGE BLANK-NOT FILMED

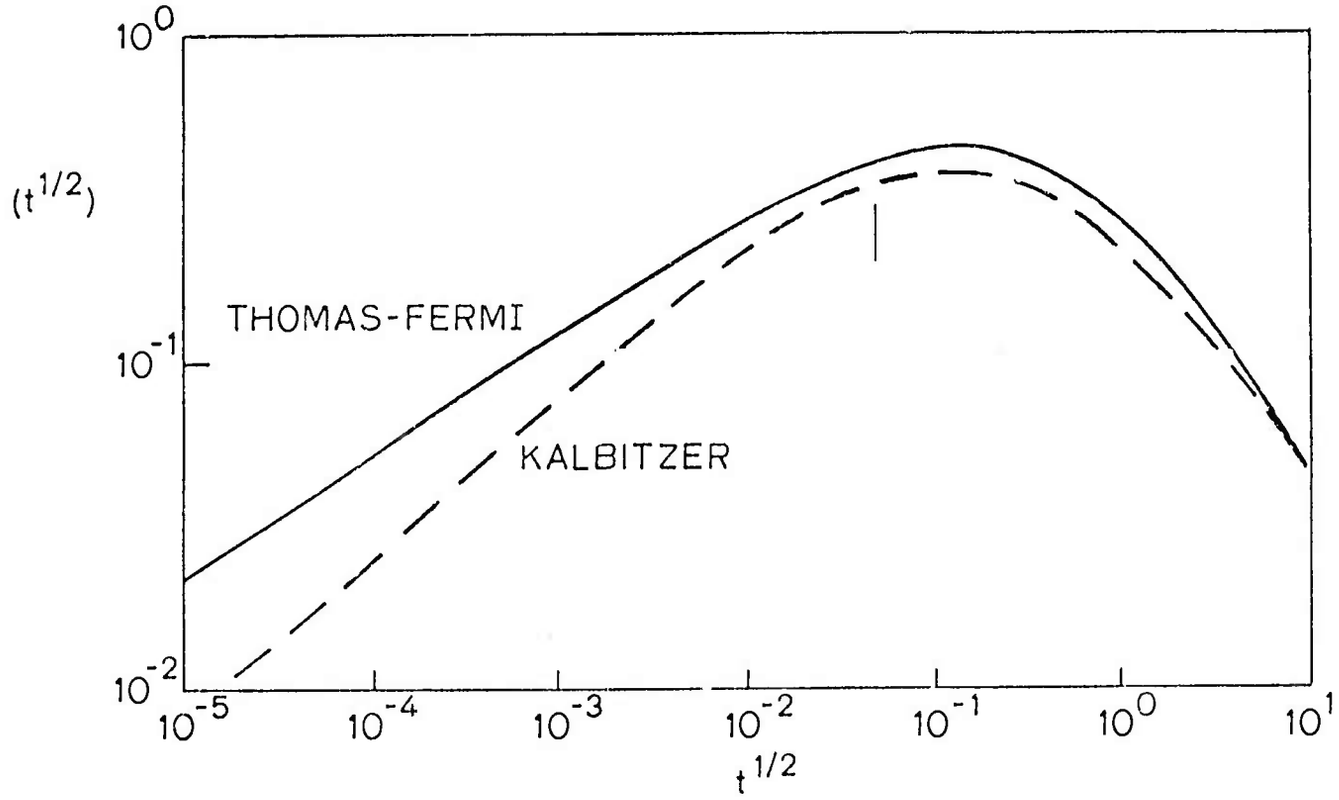


Fig. 2.2. Scattering functions $f(t^{1/2})$ used in the nuclear scattering cross section $d\sigma(t) = 1/2\ \pi a^2 t^{-3/2} f(t^{1/2})$. The marker indicates the point at which the hybrid cross section switches from the Kalbitzer form on the left to the Thomas Fermi form on the right.

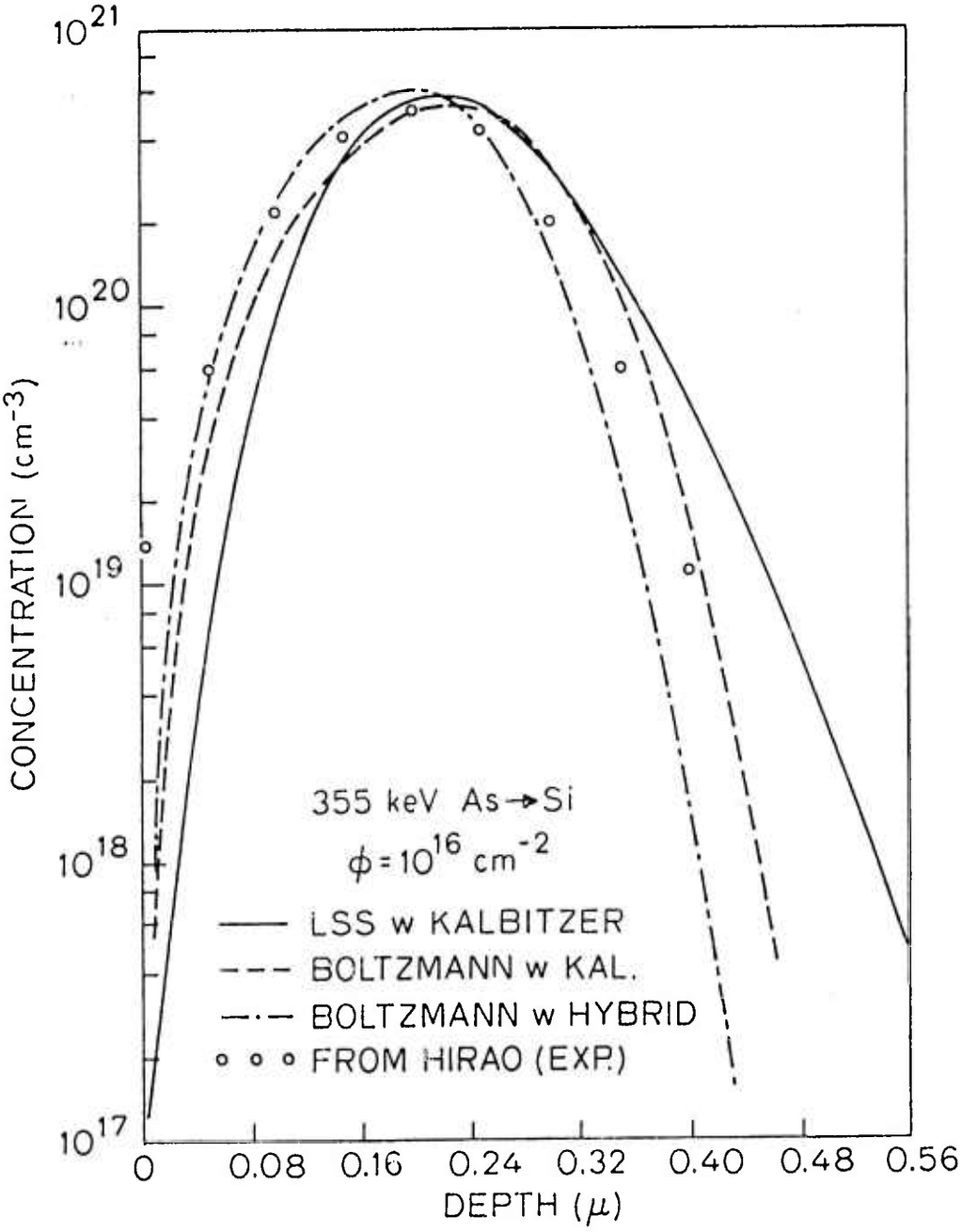


Fig. 2.3. Comparison of LSS and transport equation calculations and experimental results for the range profile of 355 keV arsenic implanted into silicon to a dose of 10^{16} cm^{-2}. The cross section used for each calculation is indicated in the key.

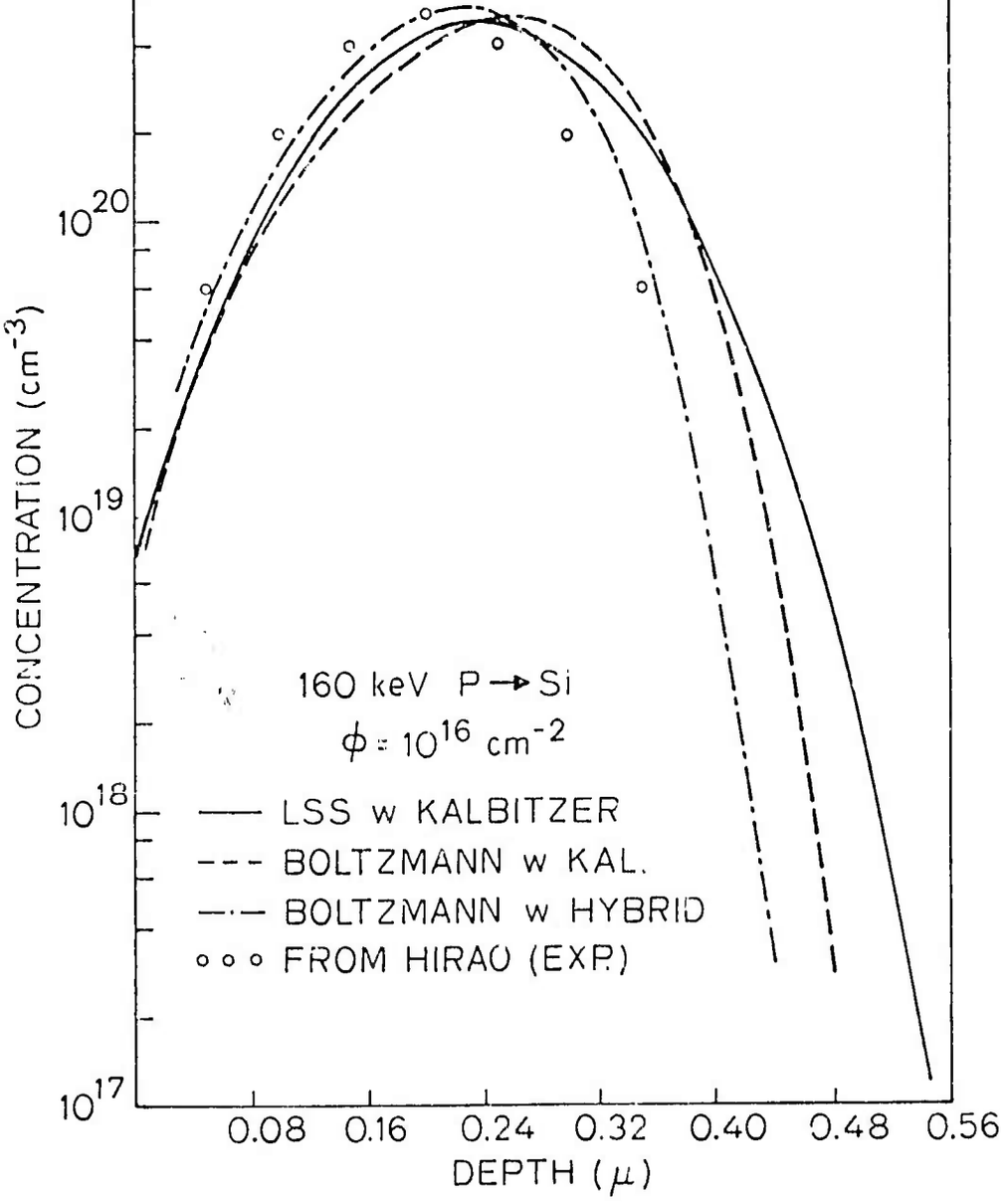


Fig. 2.4. Comparison of LSS and transport equation calculations and experimental results for the range profile of 160 keV phosphorus implanted into silicon to a dose of 10^{16} cm^{-2}. The cross section used for each calculation is indicated in the key.

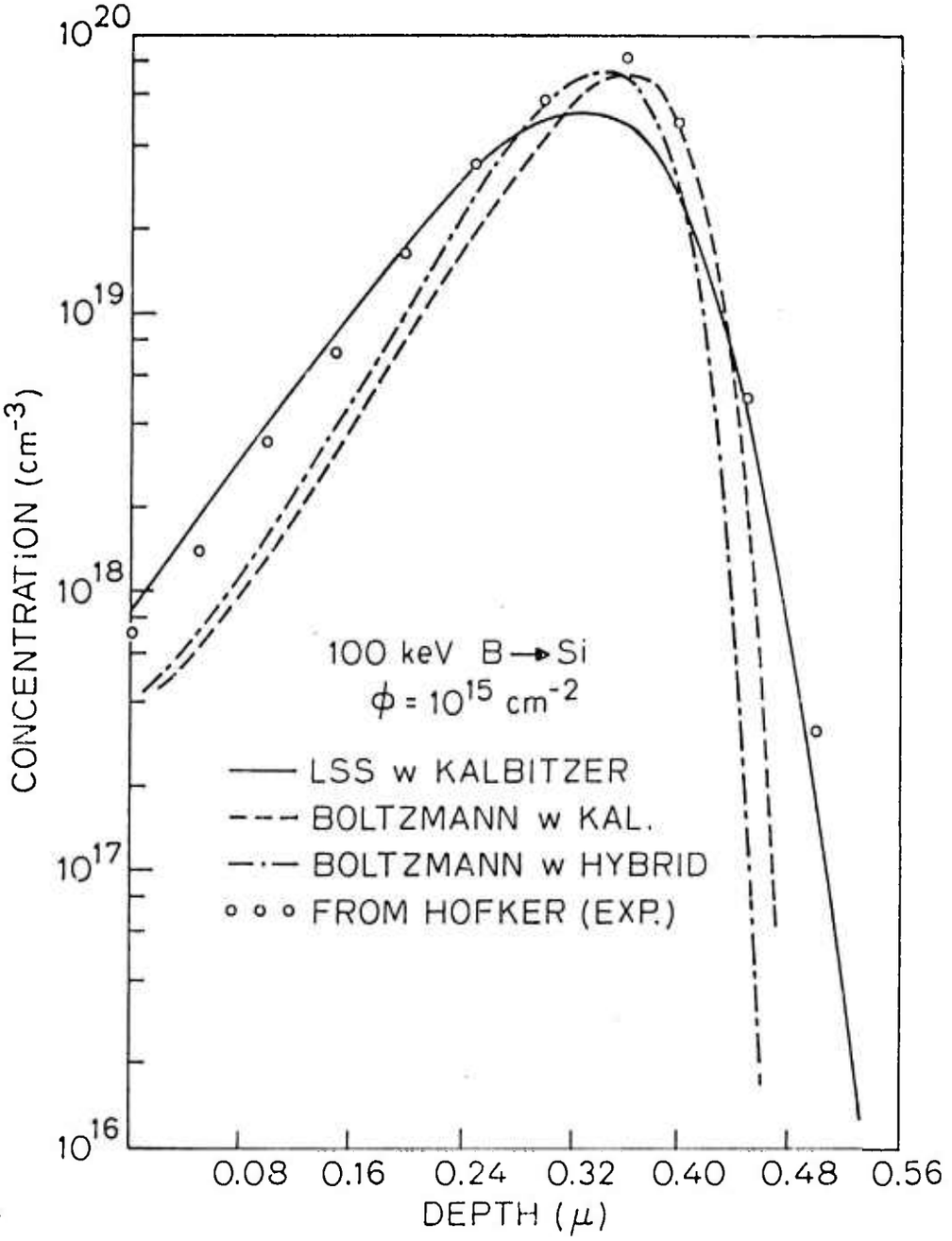


Fig. 2.5. Comparison of LSS and transport equation calculations and experimental results for the range profile of 100 keV boron implanted into silicon to a dose of 10^{15} cm^{-2}. The cross section used for each calculation is indicated in the key.

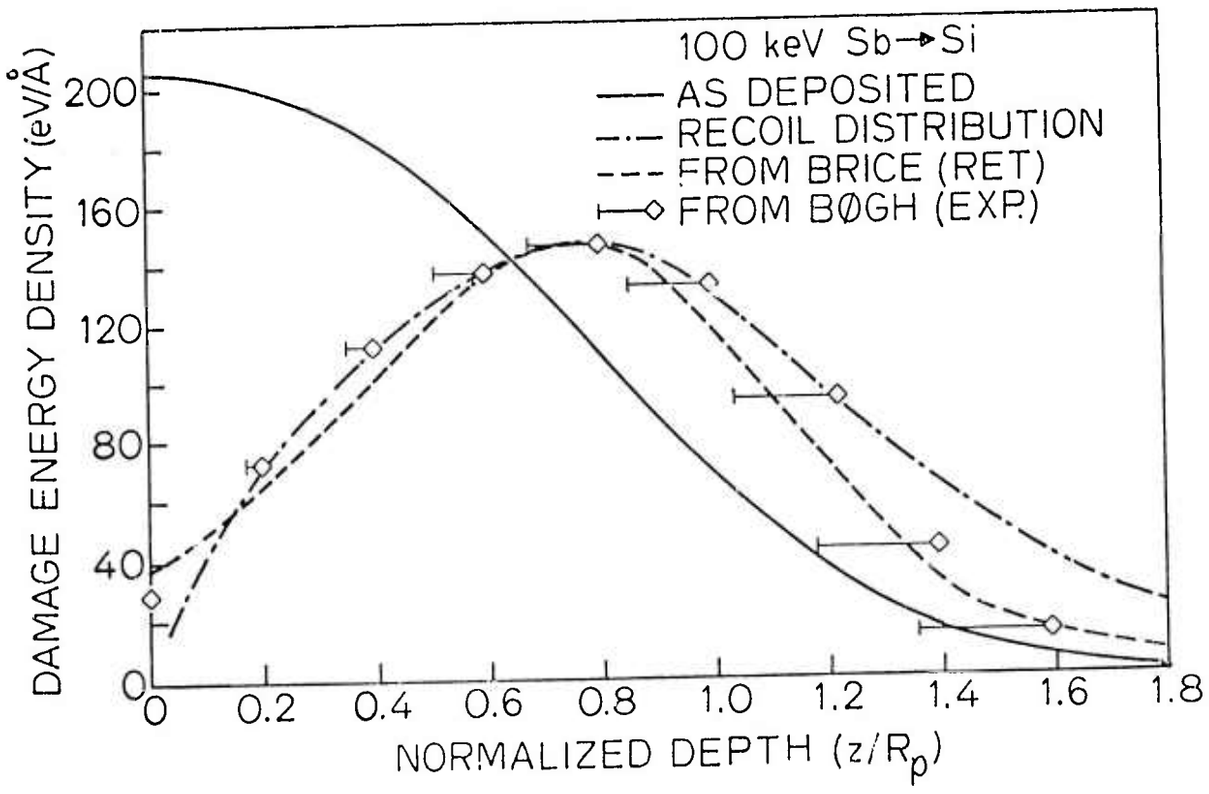


Fig. 2.6. Energy deposition profiles for 100 keV antimony implanted into silicon. The experimental data and recoil range distributions have been normalized to the peak of the Brice calculation. The hybrid cross section was used in the transport equation calculations. The abscissa is normalized to the projected range of the antimony, and the ordinate is energy density per incident particle.

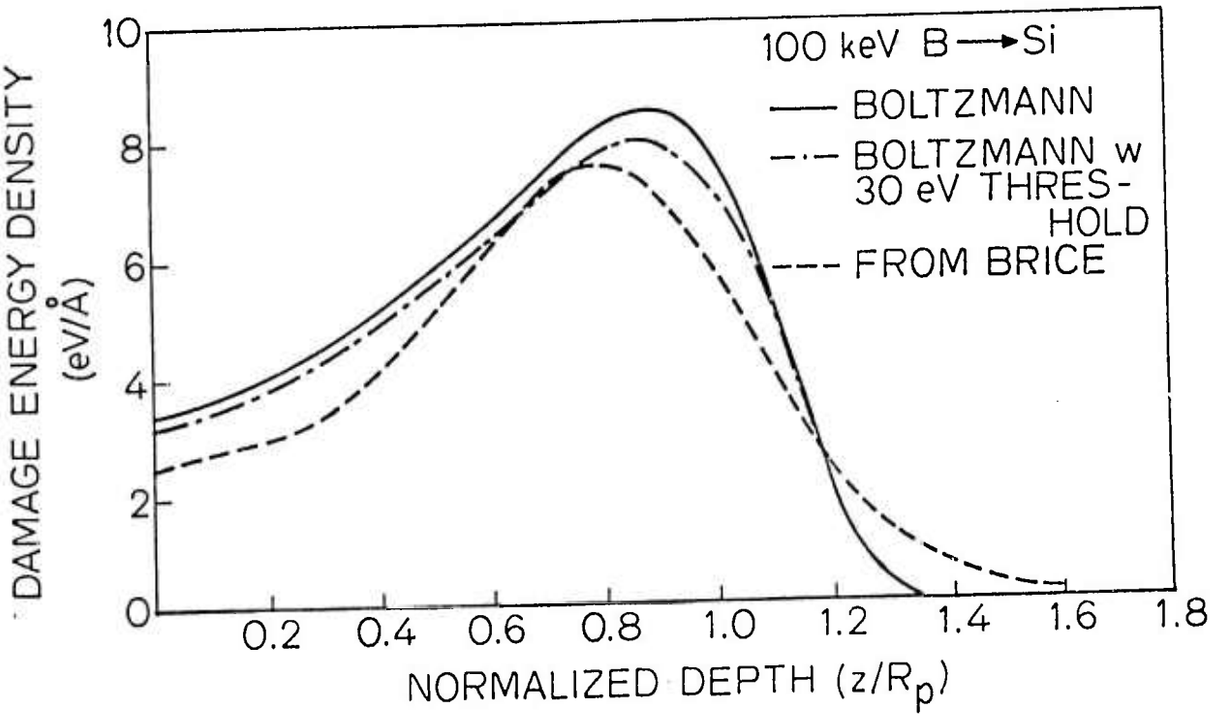


Fig. 2.7. As-deposited energy deposition profiles for 100 keV boron into silicon comparing the Brice and transport equation calculations. The abscissa is normalized to the projected range of the boron and the ordinate is energy density per incident particle.

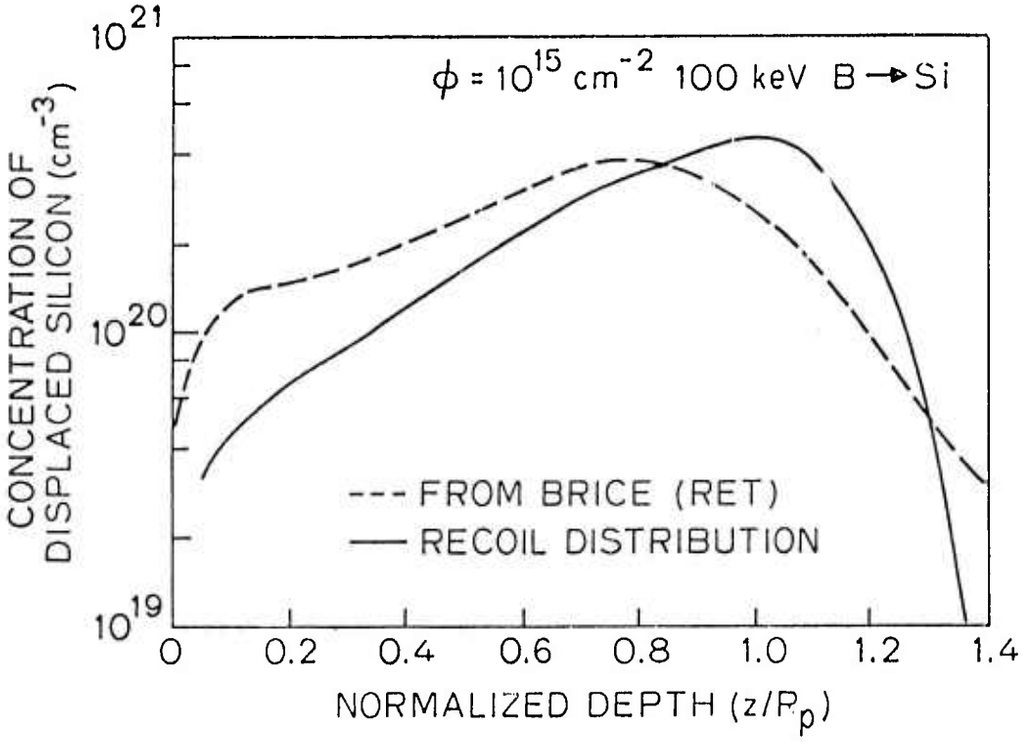


Fig. 2.8. Final damage distributions for 100 keV boron into silicon expressed in terms of the concentration of displaced substrate atoms. The Brice calculation has been normalized so that the integrals of the two curves agree. The abscissa is normalized to the projected range of the boron.

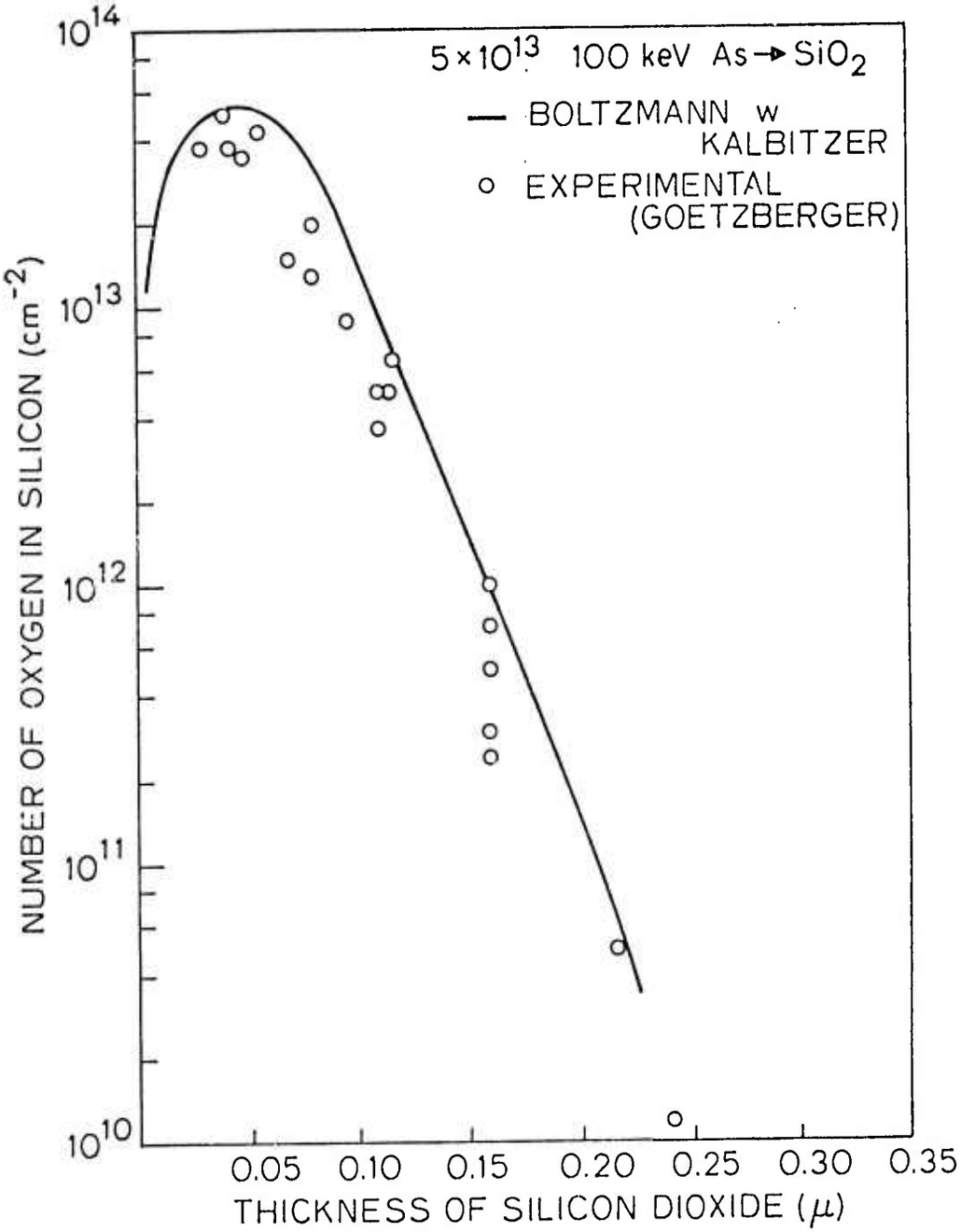


Fig. 2.9. Comparison of a transport equation calculation and experimental results for total oxygen recoil fluxes into silicon resulting from an arsenic implant at 100 keV through various thicknesses of oxide.

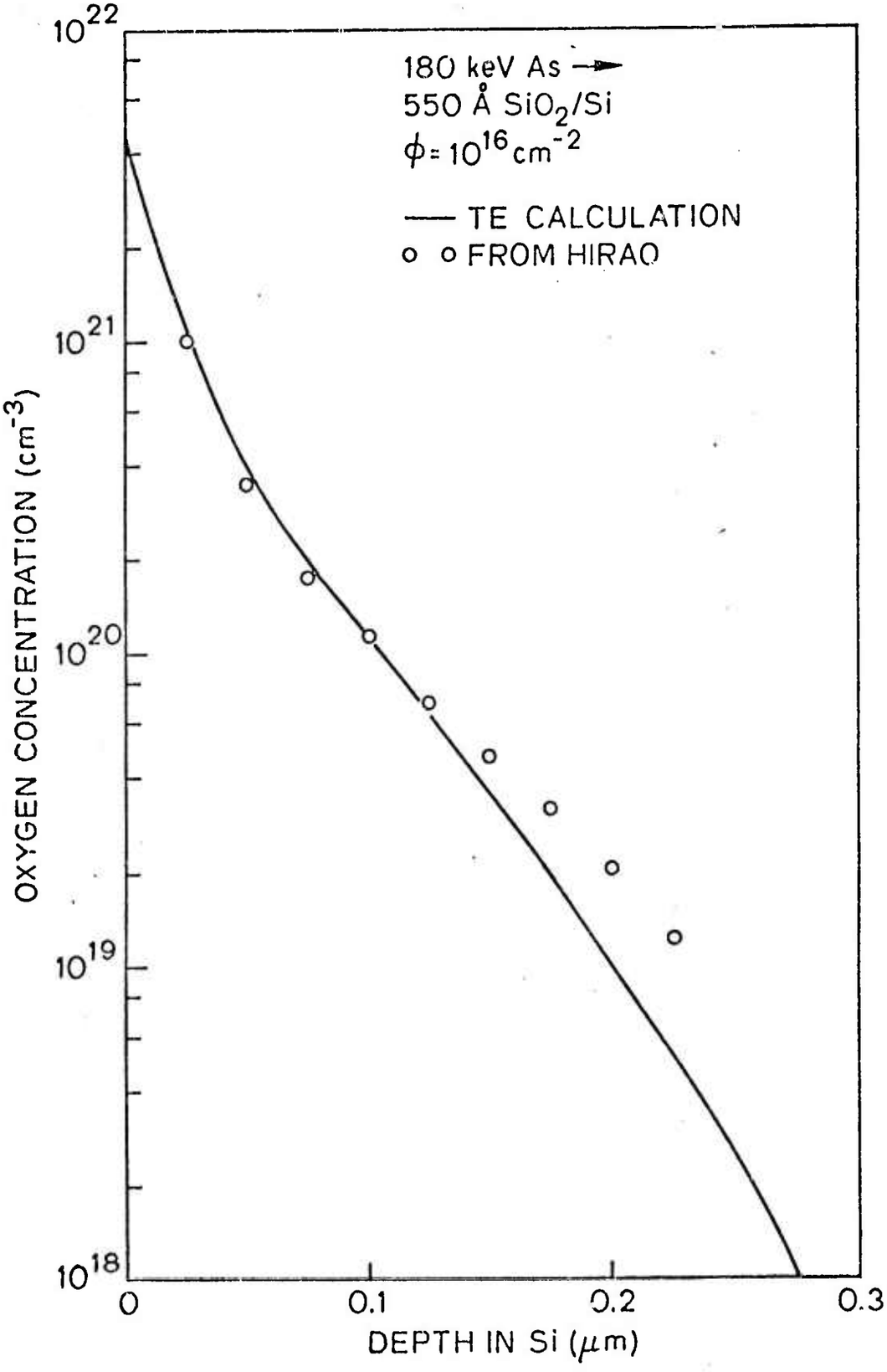


Fig. 2.10. Comparison of transport equation calculation and experimental results for oxygen recoil distribution resulting from an implant of 180 keV As into 550 Å SiO_2 on Si.

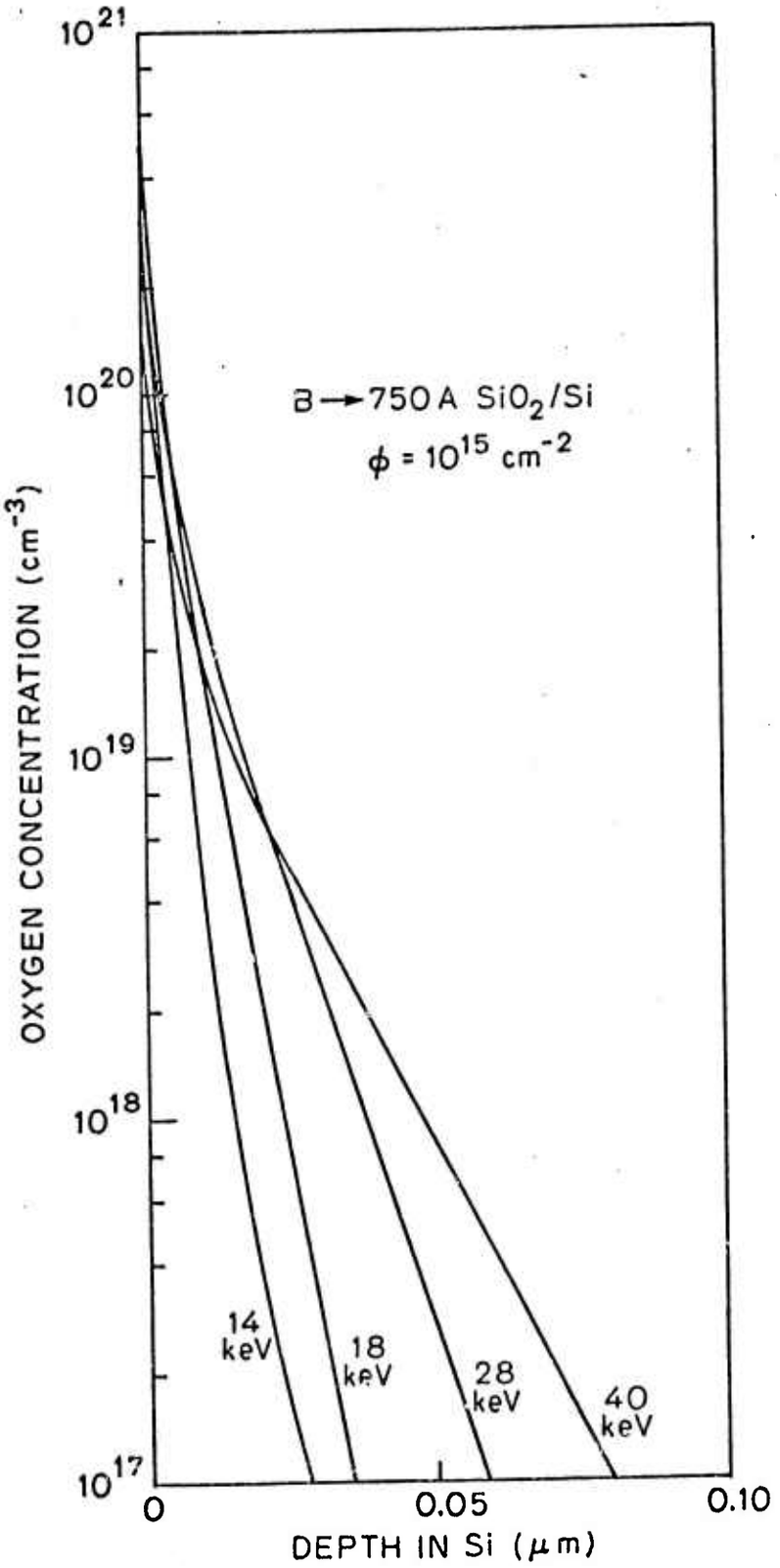


Fig. 2.11. Transport equation calculations of the oxygen recoil distributions resulting from boron implants into silicon coated with 750 Å of SiO_2.

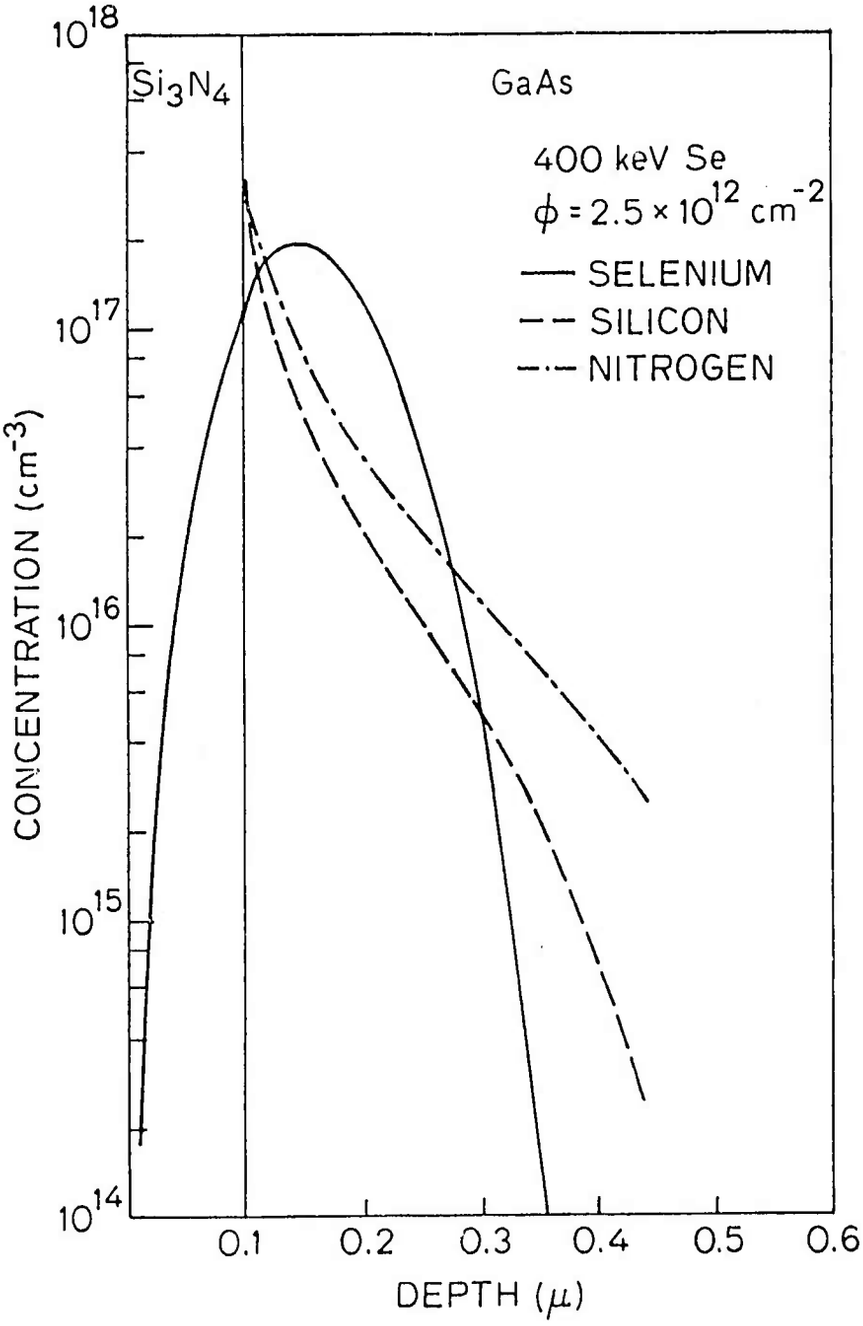


Fig. 2.12. Transport equation calculation of the recoil range distributions of silicon and nitrogen in gallium arsenide as the result of an implantation of 400 keV selenium through 1000 Å of silicon nitride to a dose of 2.5 x 10^{12} cm^{-2}.

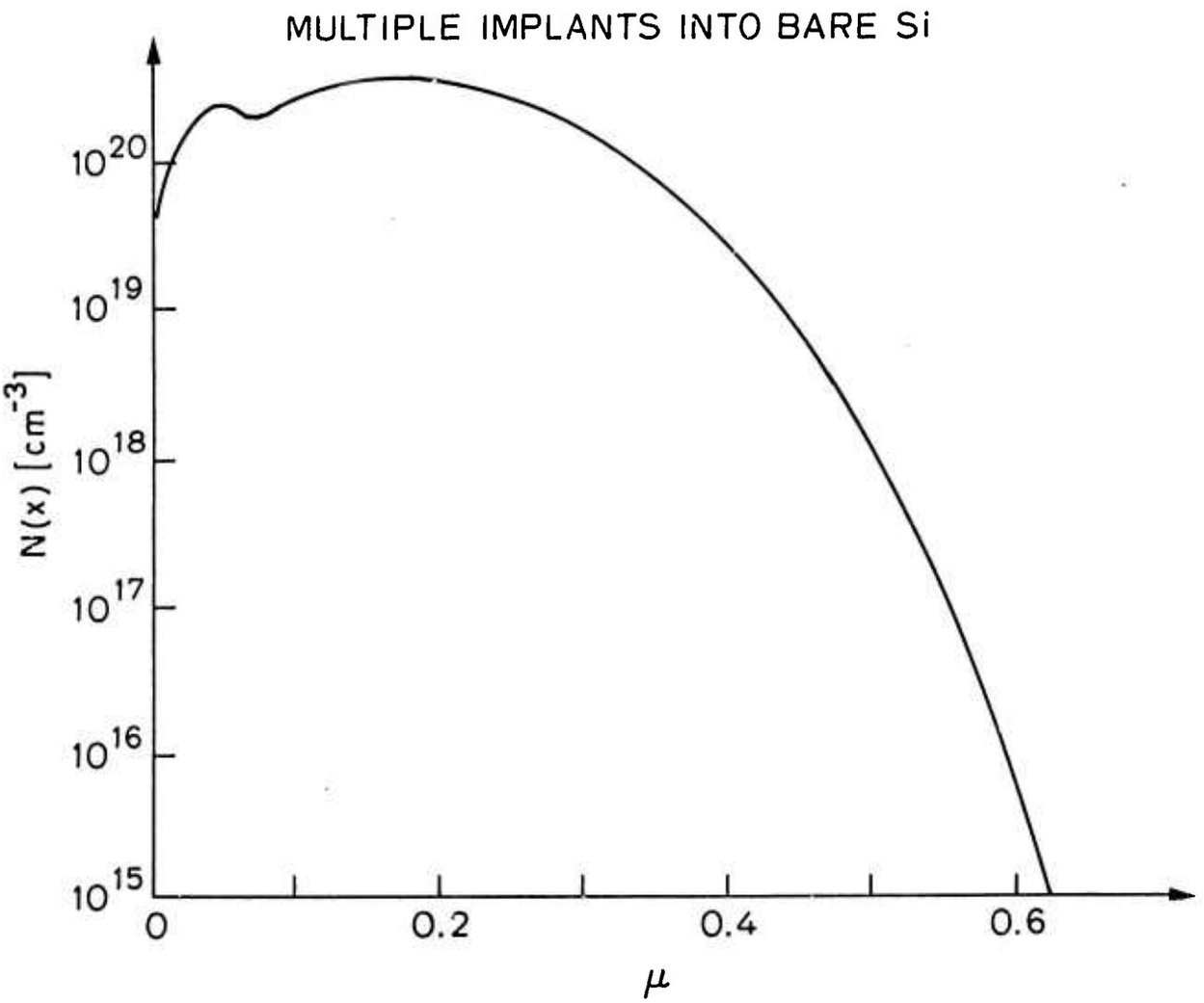


Fig. 2.13. Impurity profile resulting from the use of multiple implants of phosphorous into silicon.

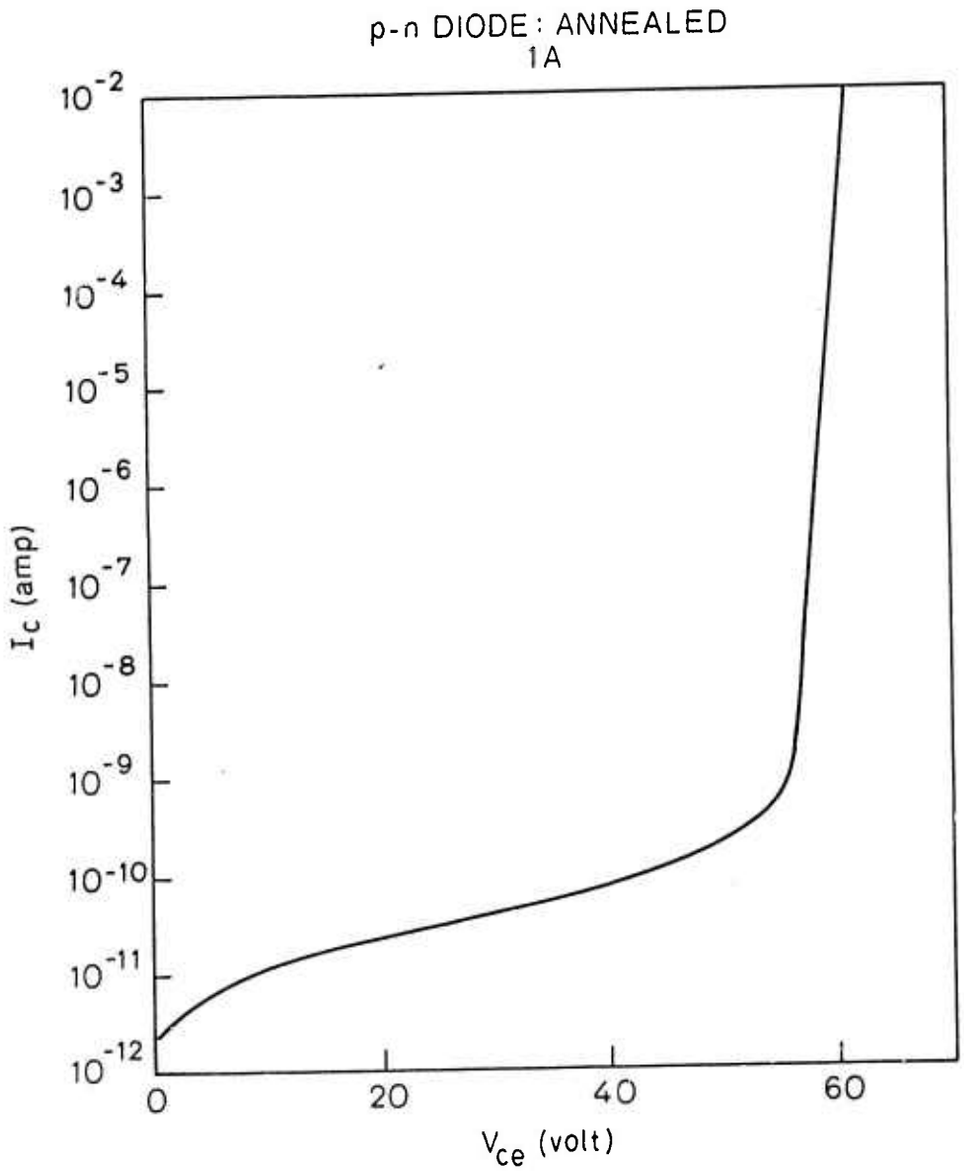


Fig. 2.14. (a) Reverse leakage current of reference p-n diode with Al contacts.

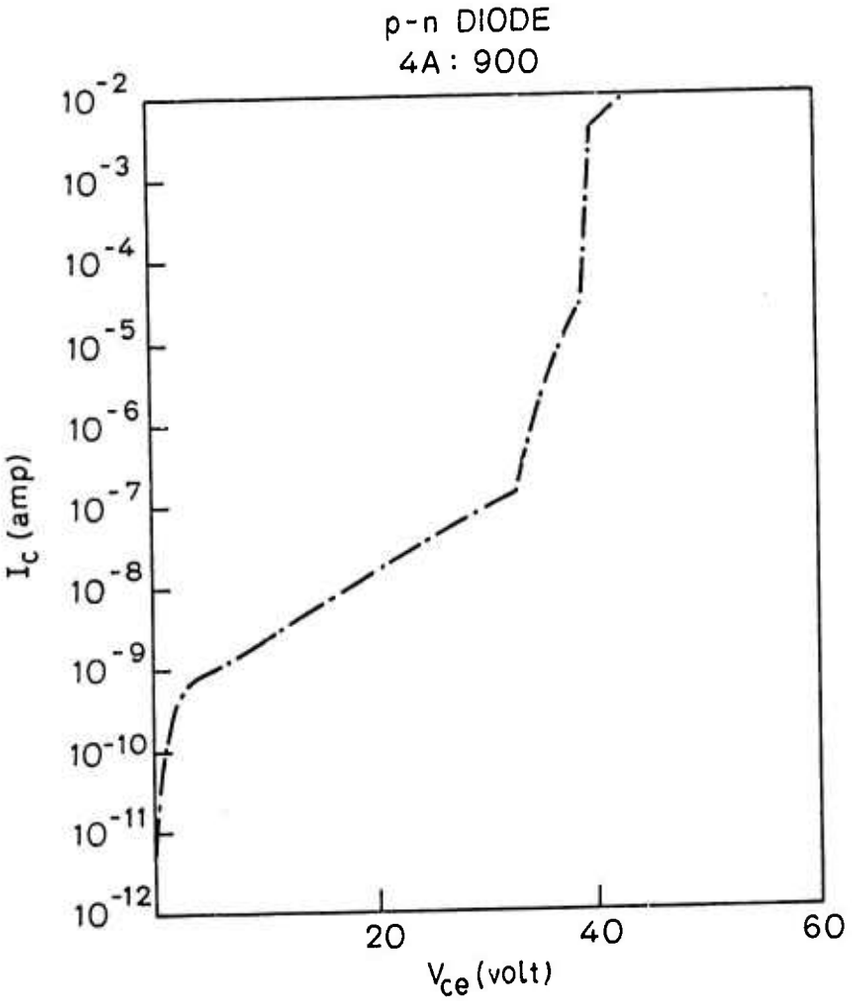


Fig. 2.14. (b) Reverse leakage current of p-n diode with WSi_2 contacts after high temperature anneal.

3. POLYCRYSTALLINE SILICON

K.C. Saraswat, M.M. Mandurah, T.I. Kamins and C.R. Helms

3.1 Introduction

Polycrystalline silicon has many important applications in integrated circuits technology. The success of MOS IC's can be largely attributed to the silicon-gate MOS technology[3.1], where heavily doped polycrystalline silicon is used as the gate and interconnection material. Lightly doped thin films are finding increasing use as high-value resistors in MOS integrated circuits, e.g., static memory circuits [3.2]. Thin films of polycrystalline silicon have been used to form monolithic distributed RC devices [3.3]. Since polycrystalline silicon is used so widely in the semiconductor industry it is very important to understand its properties.

As the technology is advancing the device dimensions continue to decrease, at the same time the overall chip size and complexity continue to increase. Since polycrystalline silicon is used as gate, interconnection and high-value resistor, the reduction in the device size as well as the increase in chip size have important implications. Today we are at the threshold of submicron technology. Thus the grain size in polycrystalline silicon can be of the same dimension as the minimum feature size. Therefore the structure of this material can seriously affect the electrical properties. It is of utmost importance to know the physics of this material and how it could impact the technology and device performance. Since the physical properties of polycrystalline silicon change as a function of some of the steps normally encountered during the fabrication of integrated circuits, electrical properties are also affected accordingly. Therefore the approach in this program is to link the physical properties to the electrical properties.

The overall goal of this research program is to develop models to accurately predict the electrical properties of polycrystalline silicon as a function of deposition, doping and further high temperature processing. The electrical properties of such films have been investigated by many researchers [3.4-3.8], however, much more work needs to be done to achieve the goal we have outlined above. 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. Within the last several years, low pressure, chemical vapor deposition (LPCVD) systems have been developed [3.9,3.10]. The high capacity and low cost of these systems led to their rapid acceptance in integrated circuit manufacturing facilities. Therefore in this work more emphasis has been put on LPCVD films.

3.2 Summary of Major Accomplishments to Date

Since the deposition rate, temperature, and pressure in these reactors are markedly different from those conventionally employed, a study of the properties of the material deposited in LPCVD reactors was needed. Therefore in the first phase of the study being conducted at Stanford, to investigate the properties of polycrystalline silicon deposited at low pressure, we have reported the structure of low-pressure, polycrystalline silicon films and its stability during further heat treatment [3.11]. The electrical properties of phosphorus-implanted, low-pressure polycrystalline silicon were then investigated and compared to those of films deposited at atmospheric pressure [3.12].

In the present report, the electrical properties of boron-doped films and arsenic-doped films are considered. First the resistivity of boron and arsenic doped films as a function of dopant concentration over a wide range

was studied. Films were doped with boron and arsenic by ion implantation over a wide range of doses corresponding to average dopant concentration of 2×10^{15} to 2×10^{20} cm^{-3}. For either dopant, at the lowest dopant concentration the polysilicon films had a resistivity approximately five orders of magnitude higher than that of correspondingly doped single crystal silicon. The resistivity decreased rapidly as the dopant concentration increased, finally falling to within half an order of magnitude of the single crystal resistivity at the highest dopant concentration. Throughout the whole range it has been found that films deposited at 580°C always had lower resistivity than did those deposited at 620°C.

These results are similar in nature to those obtained earlier for phosphorus [3.12]. A closer observation showed that arsenic doped films always had higher resistivity compared to similarly doped phosphorus films. This discrepancy led to our work on dopant segregation at grain boundaries in polycrystalline silicon. Arsenic was ion implanted into low-pressure, chemically-vapor-deposited polycrystalline silicon films. All films were then annealed at 1000°C for one hour, and some were subsequently further annealed at 800, 850 or 900°C for 64, 24 or 12 hours, respectively. The room-temperature resistivity of the films was found to be higher after annealing at lower temperatures. By successively annealing the same sample at lower and higher temperatures, the resistivity would repeatedly increase and decrease, indicating dopant segregation. Hall measurements were used to estimate the number of active dopant atoms within the grains and the number of atoms segregated at the grain boundaries as a function of annealing temperature. A theory of segregation in systems of small particles was developed. Using this theory, the heat of segregation of arsenic in polycrystalline silicon was calculated.

3.3 Resistivity of Doped LPCVD Poly-Silicon

As the increasing applications of polysilicon films begin to require more controlled film properties, further investigation of the behavior of the films and the sensitivity of the properties to the deposition parameters is necessary. In particular, resistivity of polysilicon varies over eight orders of magnitude for practical ranges of dopant addition and, in certain ranges, is a rapidly changing function of dopant concentration. In certain applications, such as high-impedance load elements for static random access memories [3.2], the resistivity must be carefully controlled in this rapidly varying region and, consequently, further characterization of the material was needed.

3.3.1 Resistivity of Boron-Doped LPCVD Polysilicon

a. Sample Preparation

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 commercial LPCVD reactor with no intentionally added dopant. Nominal deposition temperatures of 580° and 620°C were used with 100% silane gas. The thickness of the films was about 0.5 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). The films deposited at 580°C were amorphous immediately after deposition, while those deposited at higher temperatures were initially polycrystalline.

After the film deposition, boron was implanted with doses ranging from 3×10^{11} cm^{-2} to 6.9×10^{15} cm^{-2} at 80-100 KeV, energies which place the peak of the implanted distribution in the center of the film.

(As shown by Hirao, et al. [3.13], the projected range is expected to be the same in polysilicon films as in single-crystal silicon.) The samples were then protected with a low-temperature deposited oxide and annealed at 1100°C for 30 minutes in dry nitrogen atmosphere. The annealing was intended to eliminate most of the damage produced by the implantation and to create a uniform impurity distribution since the dopant diffuses much more rapidly in polysilicon than in single-crystal silicon [3.14]. During this annealing process, the amorphous films deposited at 580°C crystallized into polycrystalline silicon which has been shown to have larger grains than the initially polycrystalline films [3.11].

After annealing, a thin oxide film (less than 100 Å thick) was grown to minimize the surface leakage current during electrical measurements. This oxide was purposely kept very thin to avoid significant loss of the implanted boron from the polysilicon film due to segregation into the growing oxide film. Simulation using the SUPREM process simulation program for a dose of 1×10^{15} cm^{-2} showed that the impurity concentration was constant over most of the film, except for the region near the surface where loss of boron by dopant segregation could be seen. Because of the thinness of the oxide, and the oxidation conditions chosen, however, the simulation shows that only about 6 percent of the total dose was lost, so that the final impurity concentration was not seriously changed.

Subsequently, a 5000 Å thick layer of low-temperature oxide deposited over the thin thermal oxide at 500°C to serve as a mask for the poly-silicon etch which was used to isolate the devices. A ring-dot contact structure (Fig. 3.1) was used for measurement of the resistivity. For low doses additional boron was added to the contact regions in order to improve the contact quality. Aluminum was electron-beam deposited and photolithographically

defined, and the contacts were alloyed at 450°C for 30 minutes in nitrogen to complete the structure.

b. Results and Discussion

Electrical measurements were made on the Stanford calculator (HP 9830) controlled measurement system using the ring-dot structure. The resistivity was calculated from the following expression:

$$\rho = \frac{2\pi d}{\ln\left(\frac{r_2}{r_1}\right)} \left(\frac{V}{I}\right) \qquad (3.1)$$

where

d = Thickness of Polysilicon Film

r_2 = Radius of Outer Ring

r_1 = Radius of Inner Ring

$\frac{V}{I}$ = Resistance of Test Device (ohms)

Figure 3.2 shows the resistivity as a function of the average dopant concentration (implant dose divided by film thickness) for the three types of films used in the present experiment. Also shown are data from an earlier study of boron-implanted films deposited at atmospheric pressure and a temperature of 650°C in an argon carrier gas [3.8].

As for the case of phosphorus-doped films [3.12], thoughout the entire dopant concentration range investigated the initially amorphous films deposited at 580°C had the lowest resistivity, and the rapid decrease in resistivity occurred at a lower dopant concentration than for the other types of films. The atmospheric-pressure films deposited at 960°C had the next lowest resistivity, while the low-pressure films deposited at 620°C had higher resistivity.

These results may be understood on the basis of the carrier trapping model [3.7,3.8]. The high resistivity at low doping concentrations results from the trapping of most of the carriers at grain boundaries. As the doping concentration is increased, most trapping states will be filled; thus, the capacity of grain boundaries that trap implanted carriers will approach saturation. When the implanted dose is further increased, the potential energy barrier will decrease, causing a rapid decrease in resistivity, as observed. Further increases in doping concentration will lead to a lower resistivity which approaches that of single-crystal silicon.

Since the conductivity is limited by the number of carriers trapped at the grain boundaries, the resistivity and the critical dopant concentration should be lower for material which contains fewer grain boundaries, i.e., larger-grained material.

In our previous study [3.11], 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. 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 crystallite 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 and 960°C 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 [3.8], which is dominated by the resistivity of the crystallites themselves; hence the resistivities of the three types of polycrystalline silicon appraoch a common value close to that of single-crystal silicon.

The higher resistivity previously found in the 650°C atmospheric-pressure films deposited in an argon atmosphere [3.8] can similarly be explained by the extremely fine-grained structure formed under these conditions. The average grain size of about 200 Å is much smaller than the final grain size for any of the films investigated in the present study.

Figure 3.3 compares the boron-implanted and the phosphorus-implanted [3.]2] low-pressure films deposited at 580°C as a function of average doping concentration. The resistivity of p-type, single-crystal silicon is also shown for comparison. The experimental data of two types of impurity almost fall on the same curve, except that boron-implanted films have slightly higher resistivity at high dopant concentrations.

3.3.2 Resistivity of Arsenic-Doped LPCVD Polysilicon

Arsenic is widely used as a dopant in silicon because of its high solid solubility, low diffusivity and excellent lattice fit. In silicon, the solid solubility of arsenic is higher than phosphorus, diffusivity is lower and lattice fit is closer. Because of these properties it might be desirable to dope polysilicon with arsenic, to form low resistivity interconnections and high-value resistors. Therefore arsenic doping of LPCVD polysilicon has been investigated and resistivity as a function of average dopant concentration over a wide range has been determined.

a. Sample Preparation

Polysilicon films were deposited to a thickness of 0.5 μm at 620°C onto thermally oxidized silicon wafers. Arsenic ions were implatned with doses ranging from 1×10^{13} to 1×10^{16} ions/cm^2 at an energy of 200 KeV. The samples were coated with SiO_2 and then annealed for one hour in dry N_2 at 1100°C to activate and uniformly distribute the dopant atoms. In order to measure resistivity accurately on the more lightly doped samples (dose $\leq 10^{15}$ cm^{-2}), diffused-contact concentric rings and Van der Pauw structures were fabricated. After the annealing, 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 minutes at 900°C. Aluminum was deposited using an electron gun and the metal pattern was defined photolithographically. A 30 minute anneal at 450°C in N_2 completed the devices.

b. Results

Figure 3.4 shows the resistivity of the arsenic doped films as a function of average dopant concentraion (implanted dose divided by film thickness) (curve I). For comparison the resistivity of phosphorus doped films (curve II), obtained in our earlier work [3.12] and that of n-type single-crystal silicon (curve III) are plotted on the same figure for comparison. At the lowest dopant concentration the polysilicon films had a resistivity approximately five orders of magnitude higher than that of correspondingly doped single-crystal silicon. The resistiivty decreased rapidly as the dopant concentraiton increased, finally falling to within half an order of magnitude of the single-crystal resistivity at the highest dopant concentration. Throughout the whole range the arsenic doped films had higher resistivity than corresponding phosphorus doped films.

This difference between phosphorus and arsenic is intriguing. It cannot be explained by the generally accepted carrier trapping model. As

we will see in the next section dopant segregation to the grain boundaries is responsible for this phoneomenon.

3.3.3 Summary

This phase of the investigation of polycrystalline silicon films has considered the resistivity of low-pressure films deposited under typical production conditions (620°C) and compared ther behavior with that of films typically deposited at atmospheric-pressure (960°C). The initially amorphous films deposited at 580°C were also studied in order to show that their properties differ significantly from those of the initially polycrystalline films deposited at only slightly higher temperatures, indicating the importance of careful temperature control when depositing films close of the critical transition temperature.

The data found for the boron-doped and arsenic doped low-pressure, polycrystalline silicon films complements that found previously for phosphorus-doped films and provides a data base for implementation of the resistivity model for polycrystalline silicon into SUPREM, the Stanford process modeling computer program.

3.4 Arsenic Segregation in Polycrystalline Silicon

Many investigations of the electrical properties of polycrystalline silicon [3.4-3.8,3.12,3.15-3.18] have shown that electrical properties of polycrystalline silicon are markedly different than those of single-crystal silicon. At low dopant concentrations the resistivity is several orders of magnitude higher than that of single-crystal silicon and is not sensitive to changes in the dopant concentration. At medium dopant concentrations, a small increase in dopant concentration causes a sharp drop in resistivity. At high dopant concentrations the resistivity approaches that of single-crystal silicon although it always remains slightly higher.

Two models have been proposed to explain the resistivity variation with dopant concentration in polycrystalline silicon. The first model, the carrier trapping model, postulates that dopant atoms are uniformly distributed throughout the material and that conductivity is limited by carrier trapping at the grain boundaries, where the trapped carriers create potential barriers which impede the transfer of free carriers between the grains [3.7-3.8,3.15-3.16]. The variation of the resistivity with dopant concentration is explained by the change in the barrier height. The other model, the dopant segregation model hypothesizes that conductivity is controlled by the segregation of dopant atoms to the grain boundaries, where the atoms themselves are trapped and become electrically inactive [3.4,3.6]. Here the variation in the resistivity is explained by the hypothesis that for low dopant concentrations most of the atoms segregate to the grain boundaries, leaving few to contribute to conduction. As the dopant concentration is increased, more atoms remain inside the grains and thus the resistivity approaches that of single-crystal silicon. Each model has tried to explain the electronic conduction properties independently while completely ignoring the other model. Although both models were successful in explaining the variation of resistivity with total dopant concentration, the carrier-trapping model has been more generally accepted because the dopant-segregation model cannot explain either the temperature dependence of the film resistivity or the minimum in Hall mobility observed at intermediate dopant concentrations. In our recent work [3.17,3.18,3.19] we have shown that changes in resistivity caused by annealing cannot be explained by the carrier-trapping model and an appreciable fraction of the dopant atoms segregate to grain boundaries in polycrystalline silicon doped with arsenic and that electrical conduction in polycrystalline silicon is, therefore, influenced by both carrier-trapping and dopant atom trapping at the grain boundaries.

In their work Cowher and Sedgwick [3.6] and Fripp [3.4] have hypothesized that dopant atoms segregate in the grain boundaries, however, they have not given any concrete evidence or a physical model of the segregation process. In the present study additional evidence of dopant segregation in polycrystalline silicon is reported. A theoretical model is developed to explain this phenomenon. Based on the model, the heat of segregation of arsenic in polycrystalline silicon is calculated. The effects of different grain sizes (as determined by TEM) and doping levels are also reported.

3.4.1 Theory

The phenomenon of surface or interface segregation in a multicomponent system has been predicted and experimentally observed for a number of years. It can be described using the Gibbs adsorption equation [3.20-3.22]:

$$d\gamma = -S^s dT - \Gamma_A d\mu_A - \Gamma_B d\mu_B \tag{3.2}$$

where γ is the surface energy, S^s is the surface excess entropy, Γ_A and Γ_B are the surface excess concentrations of component A and component B, and μ_A and μ_B are the chemical potentials. In general Eq. (3.2) cannot be solved for real physical systems. One approximation that is often used is to assume that the only entropy term that need be included is the ideal entropy of mixing [3.21]. The problem then reduces to setting the chemical potential of the surface atoms and bulk atoms equal so that

$$E_s + kT \ln \frac{X_s}{1 - X_s} = E_b + kT \ln \frac{X_b}{1 - X_b} \tag{3.3}$$

where X_s and X_b are the surface and bulk atom fractions of component A and E_s and E_b are the enthalpies of component A at the surface and in the bulk. From Eq. (3.3) we obtain

$$\frac{X_s}{1 - X_s} = \frac{X_b}{1 - X_b} e^{\frac{E_b - E_s}{kT}} \qquad (3.4)$$

Due to the difficulty of calculating the terms in Eq. (3.4) there are no accurate first principle calculations of surface segregation. A number of semiempirical methods have, however, been used to predict surface segregation. For simple alloy systems quasichemical, Ising-like models have been used. The general result of these models is that the component with the lowest surface energy will segregate to the surface or interface [3.20-3.22]. For more complex systems where, for example, atom sizes are considerably different, other factors must also be considered. In the case of atoms of differing sizes the solute atom, whether larger or smaller, will segregate; in general it has been found that the heats of segregation for large solute atoms in a matrix of smaller atoms are larger than the opposite case [3.20,3.21]. For more complex systems, which would also be expected to have complex bulk-liquid phase diagrams, the above two methods are not valid. In some of these cases where the bulk phase diagram is known, segregation of the solute occurs if in the bulk phase diagram the addition of the solute lowers the melting point of the solvent-solute system [3.23].

For arsenic in silicon consideration of these factors strongly suggests that arsenic should segregate to the silicon surface. For example, estimates using the Ising-like model mentioned above, where the surface energy is estimated from the heat of sublimation, yields a heat of segregation of arsenic to a silicon surface of approximately 10 Kcal/mole. The atom size effect should not contribute to the segregation effect for this case since silicon and arsenic in a tetrahedral configuration are nearly

the same size. The distribution coefficient between arsenic in the liquid and solid phases in silicon is approximately 0.3 near the melting point. A detailed analysis of the liquids-solidus curves gives a somewhat higher value for the heat of segregation, 40-50 Kcal/mole.*

From the above discussion we would predict segregation of As to a silicon free surface. In polycrystalline silicon, however, we are not dealing with free surfaces, but grain boundaries. There has been considerable work done on segregation of solute atoms to grain boundaries for metal systems [3.23]. In all cases studied where segregation occurs at grain boundaries, it also occurs for the free surface. The major difference between the grain boundary and the free surface theoretically being the magnitude of the heat of segregation, which would be larger for the free surface.

To calculate the amount of segregation we will take a general model for the distribution of dopant atoms either in the bulk or on the grain boundaries as described by Eq. (3.4). This leads to a relationship for the fraction of dopant atoms on the grain boundaries X_{GB} in terms of the fraction of dopant atoms X_G in the bulk of the grain. For an infinite sample with one grain boundary, we can express X_{GB} in terms of X_G, temperature, and a heat of segregation Q as in Eq. (3.4):

---

*In order to be consistent with the value of a distribution coefficient of 0.3 at the melting point the pre-exponential factor A (in Eq. 3.6) is ~3 x 10^{-6} which appears to be quite small making the 40-50 Kcal/mole heat of segregation an upper limit.

$$\frac{X_{GB}}{1 - X_{GB}} = \frac{X_G}{1 - X_G} e^{Q_o/kT} \qquad (3.5)$$

The heat of segregation $Q_o = E_b - E_s$ is directly determined by the difference in enthalpy between an atom on the grain boundary and one in the bulk and can, in general, be a function of X_G and T. If the binding energies of the dopant atom in the bulk and on the grain boundary are the same, then $Q_o = 0$. If, however, the interfacial free energy is reduced by placing the atom on the grain boundary, Q_o will be positive, thus favoring segregation of the dopant atoms to the grain boundary.

Equation (3.3) from which Ew. (3.5) was derived assumes that the only entropy term is the ideal entropy of mixing. In general, Q_o will include an additional entropy term such that $Q_o = Q - ST$, where Q is the enthalpy and S is the additional entropy. This leads to the modification of Eq. (3.5) by an additional pre-exponential factor, A, yielding

$$\frac{X_{GB}}{1 - X_{GB}} = \frac{AX_G}{1 - X_G} e^{Q/kT} \qquad (3.6)$$

where $A = e^{-S/k}$.

The above analysis assumes that the total number of lattice sites present on the grain boundaries is negligible compared to the total number of sites present in the system. For typical polycrystalline silicon used in integrated circuit technology, grain sizes can be small enough so that an appreciable fraction of the atoms present will be on the grain boundaries. In this case, the segregation relationship can be modified using the following analysis [3.22,3,25,3.26].

If we start with the total number of dopant atoms N_d present per unit volume we can partition the atoms into two categories. First, there are those dopant atoms in the bulk of the grains with concentration N_G per unit volume, which we will assume to have the same activation energy for ionization as in single crystal silicon. Then, there are those dopant atoms at the grain boundaries with concentration N_{GB} per unit volume, which will be electrically inactive or have a different activation energy than that of dopant atoms present in the bulk. By conservation of dopant atoms

$$N_d = N_{GB} + N_G \tag{3.7}$$

If we let the density of grain boundary sites in the system be Q_s cm^{-3} and the total number of both "bulk" and surface sites be N_{Si} cm^{-3}, then from Eq. (3.7)

$$X_o N_{Si} = X_{GB} Q_s + X_G (N_{Si} - Q_s) \tag{3.8}$$

where $X_o = \frac{N_d}{N_{Si}}$ is the average dopant occupancy of all sites. Solving for X_G we obtain

$$X_G = \frac{X_o - (Q_s/N_{Si})\, X_{GB}}{1 - Q_s/N_{Si}} \tag{3.9}$$

Inserting this into Eq. (3.6), we obtain

$$\frac{X_{GB}}{1 - X_{GB}} = \frac{A\,[X_o - (Q_s/N_{Si})\, X_{GB})]}{1 - X_o - Q_s/N_{Si}(1 - X_{GB})}\, e^{Q/kT} \tag{3.10}$$

For typical polycrystalline silicon the average bulk dopant occupancy $X_o \ll 1$ and $Q_s/N_{Si} \ll 1$ so the terms in the denominator of

Eq. (3.10) can be replaced by unity. We then obtain

$$\frac{X_{GB}}{1 - X_{GB}} \approx A \left(X_o - \frac{Q_s}{N_{Si}} X_{GB}\right) e^{Q/kT} \qquad (3.11)$$

If the grain boundaries are far from saturation $X_{GB} << 1$ and Eq. (3.11) can be rewritten as

$$X_{GB} = \frac{AX_o \, e^{Q/kT}}{1 + A(Q_s/N_{Si}) \, e^{Q/kT}} \qquad (3.12)$$

Combining Eq. (3.9) and Eq. (3.12), we can obtain X_G as

$$X_G = \frac{X_o}{1 + (AQ_s/N_{Si}) \, e^{Q/kT}} \qquad (3.13)$$

The term Q_s/N_{Si} can be related to the average grain size since if we know the grain size, then the number of bulk lattice sites N_{Si} and grain boundary sites Q_s can be determined. If we assume the grains are cubes, Q_s/N_{Si} for silicon can be given by $1.62 \times 10^{-7}/L$, where L is the dimension of the cube edge in cm.

Plots of $X_G/X_o \approx N_G/N_d$ calculated from Eq. (3.11) without the approximation $X_{GB} << 1$ are shown in Fig. 3.5 as a function of annealing temperature for grain sizes of 500 Å, 1000 Å, and 2000 Å for Q = 10 Kcal/mole and A = 1. For the 1000 Å and 2000 Å grains N_d has been taken as 5×10^{19} cm^{-3}; for the 500 Å grains plots are shown for $N_d = 5 \times 10^{19}$ and 5×10^{20} cm^{-3}. From Fig. 3.5 it is obvious that for these values of the parameters, from 25-75% of the dopant atoms can be segregated to the grain boundaries.

The next question we address is what effect does the segregation of the dopant atoms have on the carrier concentration. For analysis of the

results to follow, we will assume that any dopant atoms present at the grain boundaries either (1) do not contribute carriers to the system or (2) the carriers that are donated are locally trapped at the grain boundary and do not participate in the conduction process. For arsenic the first case is quite reasonable since a grain boundary would provide an ideal site on which an arsenic atom could have all five bonds saturated [3.27]. In this case the ionization energy of the arsenic atom would be too large for significant carrier activation.

3.4.2 Sample Preparation

Polycrystalline silicon films were deposited at low pressure (0.2 torr) to a thickness of approximately 0.5 m by chemical vapor deposition at 625°C onto thermally oxidized silicon wafers. Arsenic atoms with doses in the range from 1×10^{14} to $1 \times 10^{16}/cm^2$ were ion implanted into the films at an energy of 200 KeV. After the ion implantation, the wafers were first coated with low-temperature SiO_2 to prevent the escape of the dopant atoms and to protect the surface of the films from the ambient during the subsequent annealing cycles. All wafers were then annealed at 1000°C for one hour to activate and redistribute the dopant atoms and to stabilize the grain size so that further, lower temperature, annealing would not change the grain size appreciably [3.28]. In this way dopant segregation effects could be separated from those related to changes in grain size. After the initial 1000°C anneal, some of the samples were further annealed at 900, 850 or 800°C for 12, 24 or 64 hours, respectively. These times were found to be adequate for all resistivity changes to reach saturation. Different samples were used for each of these annealing temperatures, and some samples received no further heat treatment after the initial 1000°C anneal. All anneals were performed in a dry nitrogen ambient.

In order to measure resistivity accurately, especially on the more lightly doped samples, diffused-contact, concentric-ring resistor structures were fabricated. After the annealing, photoresist was applied, and the oxide protecting the polycrystalline silicon films was selectively etched to define the concentric rings. The exposed polycrystalline silicon was then doped with phosphorus from a $POCl_3$ source at the same temperature as the final anneal given to the sample. Aluminum was deposited using an electron gun, and the metal pattern was defined photolithographically. A 30 min anneal at 450°C in nitrogen completed the devices.

The samples for transmission electron microscopy were prepared by etching the supporting substrate and the oxide from the back of the polycrystalline film. The film was usually further thinned chemically from the back so that the electron beam could easily penetrate the film; in all cases, therefore, the results correspond to the portion of the film farthest from the substrate. After preparation the samples were examined in a Philips EM400 electron microscope with a 100 KeV electron beam.

3.4.3 Measurement and Interpretation of Hall Constant

Measurement of Hall constant was made only on the highly doped samples. Unpatterned square and rectangular-shaped samples were used for these measurements. Because of the high dopant concentrations, good ohmic contact between the probes and the samples was obtained without additional contact diffusions. Hall constant was measured 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 equations used to calculate Hall constant, R_H, carrier concentration, n, and Hall mobility, μ_H, are

$$R_H = \frac{tV_H}{BI}$$

$$n = \frac{1}{qR_H} \quad \text{and} \quad \mu_H = \frac{R_H}{\rho} \tag{3.14}$$

where t and ρ are the thickness and resistivity of the film, respectively.

Because of the complicated nature of current conduction in inhomogeneous materials, Hall measurements on these materials must be carefully interpreted. The nature of current conduction in polycrystalline silicon is further complicated by carrier trapping and the subsequent formation of energy barriers at the grain boundaries. However, under certain conditions, it has been shown that Hall measurement can give the number of active carriers with reasonable accuracy.

i) In a simple model, in which the inhomogeneous material is viewed as composed of low resistivity regions separated by narrow, high resistivity regions, Volger [3.29] and Heleskivi and Salo [3.30] showed that the Hall constant should accurately give the total active carrier concentration in the low resistivity region.

(ii) Using the same simple model as in (i), Jerhot and Snejdar [3.31] showed that the presence of small energy barriers (< 5 kT) at the grain boundaries has very little effect on the measured Hall constant.

The number of carriers trapped at the grain boundaries in polycrysta-line silicon has been shown to have a maximum vlaue of $N_d^* = \frac{Q_T}{L}$, where Q_T is the density of trapping states at the grain boundaries/cm^2 and L is the grain size [3.8,3.15]. For dopant concentrations $N_d > N_d^*$, the

energy barrier height at the grain boundaries is inversely proportional to the dopant concentration N_d. Thsu, in relatively highly doped samples, the energy barriers at the grain boundaries are small, and the material satisfies the conditions of the simple model above. Consequently, Hall measurement on these samples should accurately predict the active carrier concentration in the grains.

The density of carrier traps at the grain boundaries, Q_T, has been found to be of the order of $3 \times 10^{12}/cm^2$ [3.8,3.15]. The grain size of the polycrystalline silicon films used in this experiment, L, as will be shown later, is in the order of 0.1 μm. Hence, the number of carriers trapped at the grain boundaries has a maximum vlaue N_d^* of approximately $3 \times 10^{17}/cm^3$. Thus, for dopant concentrations much higher than N_d^*, the number of carriers trapped at the grain boundaries is very small compared to the total carrier concentration; hence, assuming complete ionization of the dopant atoms in the grains, the carrier concentration obtained by Hall measurement is equal to the number of active dopant atoms in the grains.

3.4.4 Results and Discussion

Figure 3.6 shows the room-temperature resistivity as a function of the final annealing temperature for three different arsenic doses equivalent to average dopant concentrations of 2×10^{18}, 1×10^{19} and $2 \times 10^{19}/cm^3$. In all cases the resistivity increased from its initial 1000°C value on subsequent annealing, with a larger change caused by annealing at lower temperatures. The change in resistivity was especially strong for the lowest dopant concentration, with an increase in resistivity of approximately two orders of magnitude after annealing at 800°C.

The change in resistivity was reversible upon further annealing. A polycrystalline silicon film implanted with 1×10^{15} ions/cm^2 of arsenic,

equivalent to dopant cocentration of $2 \times 10^{19}/cm^3$, was sequentially annealed at lower and higher temperatures. The resistivity of the film was measured after each annealing cycle and was found to repeatedly increase and decrease as shown in Fig. 3.7.

Figure 3.8 shows the carrier Hall mobility as a function of annealing temperature for polycrystalline silicon films implanted with 1×10^{15} arsenic ions/cm^2, corresponding to the bottom curve of Fig. 3.6. The mobility increased from its 1000°C value on subsequent annealing, with larger change caused by annealing at lower temperatures. The mobility increased by approximately 50% while the resistivity increased by 22% after annealing at 800°C. Since the mobility in polycrystalline silicon is a function of the thickness of the material [3.7], the values reported here are averages over the thickness of the film.

Figure 3.9 shows the carrier concentration, n, in the crystallites obtained from Hall measurements as a function of annealing temperature for the three average dopant concentrations of 2×10^{19}, 6×10^{19} and $2 \times 10^{20}/cm^3$. The carrier concentration decreased from its 1000°C value on subsequent annealing in all three cases, with a larger change caused by annealing at lower temperatures. Since the dopant concentrations under consideration are much higher than the number of trapped carriers, N_d^* ($\sim 3 \times 10^{17}/cm^3$) [3.8,3.15], the carrier concentration, n, is a direct measure of the total number of ionized dopant atoms in the crystallites.

In single crystal silicon, at moderately high doping levels ($> 10^{18}/cm^3$), the shallow impurity levels in the forbidden gap begin to merge with the conduction band so that the dopant atoms are completely ionized [3.32]. Consequently, in the polycrystalline films the dopant

concentration in the crystallites, which can be assumed to have the nature of single crystal silicon, can also be assumed to be completely ionized and can be directly equated to the carrier concentration obtained by the Hall measurements. Therefore, it can be concluded taht the reduction of carrier concentration upon annealing at lower temperature is caused by the reduction of the number of ionized dopant atoms in the grains, that is, by the segregation of some of the dopant atoms to the grain boundaries. This conclusion is further strengthened by

(i) the observed reversibility of the resistivity upon annealing at lower and higher temperature, as shown in Fig. 3.7, which is a definite indication of the movement of dopant atoms between grains and grain boundaries, and by

(ii) the large change in resistivity upon annealing for the lowest dopant concentration used in this study, as shown in Fig. 3.6, which is consistent with the observed strong dependence of the resistivity on dopant concentration in this intermediate dopant concentration range [3.7,3.8,3.12,3.15].

Using Eq. (3.7) and the above assumptions the following relations hold

$$N_G = n \quad \text{and} \quad N_{GB} = N_d - n \tag{3.15}$$

Substituting into Eq. (3.13) and rearranging the terms, we obtain

$$\ln \frac{N_{GB}}{N_G} = \ln \frac{(N_d - n)}{n} = \ln \frac{A \cdot Q_s}{N_{Si}} + \frac{Q}{kT_A} \tag{3.16}$$

Figure 3.10 is a logarithmic plot of N_{GB}/N_G vs the reciprocal of the annealing temperature T_A for the three arsenic dopant concentrations

of 2×10^{19}, 6×10^{19} and $2 \times 10^{20}/cm^3$. A linear dependence on $1/T_A$ is obtained for each curve; the slope of the lines gives the heat of segregation. Table 3.1 shows values for the heat of segregation, Q, and the pre-exponential factor, $A \cdot Q_s/N_{Si}$, for each doping level. The average value for the heat of segregation is equal to 9.8 Kcal/mole.

Figure 3.11 shows transmission electron micrographs of polycrystalline films doped with arsenic with dopant concentrations of 2×10^{19}, 6×10^{19} and $2 \times 10^{20}/cm^3$ and annealed for one hour at 1000°C. The average grain size in these films is approximately 0.1, 0.12 and 0.16 μm, respectively. This increase in grain size is attributed to the enhancement of grain growth by heavy doping, an effect frequently observed in similar heavily doped samples [3.28]. Using these values for the grain sizes, Q_s/N_{Si} can be calculated and, hence, the pre-exponential factor, A, can be determined for each doping level as shown in Table 3.1.

Values for these pre-exponential terms should, however, be considered estimates at best for two reasons: (1) small errors in Q can cause large errors in the calculated values of these pre-exponential factors, and (2) we are assuming that all sites on a grain boundary are active for dopant segregation. This assumption is certainly not established. In any case the fact that the pre-exponential factor, A, is so close to unity (the ideal solution limit) is encouraging.

Another mechanism which might account for the observed resistivity behavior is the clustering of dopant atoms. Arsenic clustering in heavily doped single, crystal silicon has been observed by many researchers [3.33-3.36]. Similar reversible changes in resistivity upon annealing at lower and higher temperatures were observed and several models were developed to explain this phenomenon. All of these models propose

the formation of "vacancy-impurity" complexes as the cause for arsenic clustering in silicon and they differ only in the nature and configuration of these complexes. All of these models also predict a strong dependence of the clustering on arsenic concentration and on annealing temperature, being maximum at high arsenic concentrations and at lower annealing temperatures.

Table 3.2 shows the amount of arsenic clustering as a percentage of the total dopant concentration as predicted by three different clustering models [3.34,3.35,3.37]. As can be seen in Table 3.2, the amount of arsenic clustering as predicted by the above models is negligible for the two samples with lowest dopant concentrations. Therefore, the carrier concentration changes seen in these samples upon annealing cannot be explained by "arsenic-clustering" or "vacancy-impurity complex formation" and can only be the result of dopant segregation to grain boundaries. It is interesting to mention here that although arsenic clustering may be appreciable in the sample with the highest dopant concentration, the heat of segregation in this case was very close to that of the other two cases with lower dopant concentrations as shown in Table 3.1. This may suggest that any clusters of arsenic atoms in polycrystalline silicon will eventually diffuse to the grain boundaries; thus satisfying the conditions of both arsenic clustering and equilibrium dopant segregation.

Using a different approach to study dopant segregation in polycrystalline silicon, Monkowski et al. [3.38], used the capacitance-voltage technique to compare the dopant atom distribution in polycrystalline silicon to that in single crystal silicon grown by chemical vapor deposition under the same conditions. They observed an identical amount and distribution of dopant atoms in the polycrystalline and single-crystal silicon, indicating

no dopant segregation in polycrystalline silicon. However, the films used in their study contained large grains (5 μm) and were prepared at a high temperature (1225°C), where little dopant segregation is expected.

Additional evidence of dopant segregation to the grain boundaries was reported by Swaminathan et al. [3.39]. They measured the distribution of arsenic in polycrystalline silicon films deposited on single crystal silicon and on oxide by ion backscattering after various high temperature annealing steps. Pile up of arsenic was observed at the interfaces of their deposited films which followed the same trends observed in this experiment, being greatest for the samples annealed at the lowest temperatures. This pile up was explained as the result of arsenic segregation to the grain boundaries, considering the interface between polycrystalline silicon and single crystal silicon analogous to the interface between grains in the polycrystalline silicon material.

3.4.5 Summary

This study has shown that an appreciable fraction of the arsenic dopant atoms may segregate to grain boundaries in chemically-vapor-deposited polycrystalline silicon doped with arsenic by ion implantation, and that the electrical conduction in polycrystalline silicon is, therefore, influenced by both carrier trapping and dopant atom trapping at the grain boundaries.

The fraction of dopant atoms segregating to the grain boundaries is proportional to the density of segregating sites at the grain boundaries and hence inversely proportional to the size of the grains. The heat of segregation was found to be almost independent of dopant concentration and approximately equal to 9.8 Kcal/mole.

3.5 Conclusion

The results of this phase of the investigation show that dopant atom segregation, as well as carrier trapping, must be considered in order to relate the resistivity to the dopant concentration in polycrystalline silicon. Thus, implementation of the resistivity model into the Stanford process modeling computer program SUPREM will require consideration of both effects, as schematically illustrated in Fig. 3.11. First, the number of dopant atoms segregating to the grain boundaries and the number of substitutional dopant atoms remaining within the grains will be found. Then an improved carrier trapping model can be used to find the number of electrically active carriers and their mobility base on the number of substitutional dopant atoms remaining within the grains.

REFERENCES

3.1 F. Faggin and T. Klein, Solid-State Electron., 13, 1125 (1970).

3.2 T.R. O'Connel et al., paper WAM 1.2 presented at the IEEE International Solid State Circuits Conf., Philadelphia, Pa., Feb. 16-18, 1977.

3.3 L. Gerzberg, T.R. No. G557-2, Stanford Electronics Laboratories, Stanford Univ., Stanford, CA, May 1979.

3.4 A.L. Fripp and L.H. Slack, J. Electrochem Soc., 120, 145 (1973).

3.5 P. Rai-Choudhury and P.L. Hower, ibid., 120, 1761 (1973).

3.6 M.E. Cowher and T.O. Sedgwick, ibid., 119, 1565 (1972).

3.7 T.I. Kamins, J. Appl. Phys., 42, 4357 (1971).

3.8 J.Y.W. Seto, ibid., 46, 5247 (1975).

3.9 R.S. Rosler, Solid State Technol., 20, No. 4, 63 (1977).

3.10 W.A. Brown and T.I. Kamins, Solid State Technol. 22, no. 7, p. 51, (July 1979).

3.11 T.I. Kamins, M.M. Mandurah, and K.C. Saraswat, J. Electrochem. Soc., 125, 927 (1978).

3.12 M.M. Mandurah, K.C. Saraswat, and T.I. Kamins, J. Electrochem. Soc. 126, 1019 (1979).

3.13 T. Hirao, et al., "Ion Implantation in Semiconductors," Plenum Press, New York, 1974, pp. 599-604.

3.14 T.I. Kamins, J. Manoliu, and R.N. Tucker, J. Appl. Phys. 43, 83 (Jan. 1972).

3.15 G. Baccarani, B. Ricco, and G. Spadini, J. Appl. Phys. 49, 5565 (1978).

3.16 N.C.C. Lu, L. Gerzberg and J.D. Meindl, IEEE Electron Device Letters, EDL-1, (1980).

3.17 M.M. Mandurah, K.C. Saraswat, and T.I. Kamins, Appl. Phys. Lett. 36, 683 (1980).

3.18 M.M. Mandurah, K.C. Saraswat and T.I. Kamins, Extended Abstracts of the 156th Meeting of the Electrochemical Society, Los Angeles, Calif., 79-2, 1433 (1979).

3.19 M.M. Mandurah, K.C. Saraswat, C.R. Helms, and T.I. Kamins, submitted to J. Appl. Phys.

3.20 See for example: P. Wynblatt and R.C. Ku in "Interfacial Segregation," W.C. Johnson and J.M. Blakely, Eds. (ASM, Metals Park, Ohio, 1979).

3.21 See for example: P. Wynblatt and R.C. Ku, Surf. Sec. 65, 511 (1977).

3.22 See for example: J.J. Burton, E. Hyman, D.G. Fedak, J. Catal. 37, 106 (1975).

3.23 J.J. Burton and E.S. Machlin, Phys. Rev. Letts. 37, 1433 (1976).

3.24 R.W. Balluffi loc. cit., ref. 10.

3.25 C.R. Helms loc. cit., ref. 10.

3.26 F.L. Williams and D. Nason, Surf. Sec., 45, 377 (1974).

3.27 W.E. Spear and P.G. LeComber, Phil Mag., 33, 935 (1976).

3.28 Y. Wada and S. Nishimatsu, J. Electrochem. Soc. 125, 1499 (1978).

3.29 J. Volger, Phys. Rev. 79, 1023 (1950).

3.30 J. Heleskivi and T. Salo, J. Appl. Phys. 43, 740 (1972).

3.31 J. Jerhot and V. Snejdar, Thin Solid Films 52, 379 (1978).

3.32 See for example: V.I. Fistul, Heavily Doped Semiconductors (Plenum, New York, 1969).

3.33 R.O. Schwenker, E.S. Pan, and R.F. Lever, J. Appl. Phys. 42, 3195 (1971).

3.34 M.Y. Tsai, F.F. Morehead, and J.E.E. Baglin, J. Appl. Phys., to be published.

3.35 J. Murota and E. Arai, J. Appl. Phys. 50, 804 (1979).

3.36 R.B. Fair and G.R. Weber, J. Appl. Phys. 44, 273 (1973).

3.37 R.B. Fair, to be published.

3.38 J.R. Monkowski, J. Bloem, L.J. Giling, and M.W.M. Graef, Appl. Phys. Lett. 35, 410 (1979).

3.39 B. Swaminathan, E. Demoulin, T.W. Sigmon, R.W. Dutton and R. Reif, Appl. Phys. Lett., to be published.

Table 3.1

Doping Level N_d	Heat of Segregation Q_o	Pre-exponential Factor $A\ Q_s/N_{Si}$	A
2×10^{19} cm^{-3}	10.1 Kcal/mole	.049	3.02
6×10^{19} cm^{-3}	9.58 Kcal/mole	.036	2.67
2×10^{20} cm^{-3}	9.54 Kcal/mole	.022	2.17

Table 3.2

N_d/cm^3	T_A (°C)	n/cm^3	% Clustering		
			Ref (24)	Ref (25)	Ref (27)
2×10^{19}	800	3.10×10^{18}	< .01%	< .01%	---
2×10^{19}	850	3.62×10^{18}	< .01%	< .01%	---
2×10^{19}	900	4.19×10^{18}	< .01%	< .01%	---
2×10^{19}	1000	5.52×10^{18}	< .01%	< .01%	< .01%
6×10^{19}	800	1.42×10^{19}	.12%	.03%	---
6×10^{19}	850	1.667×10^{19}	.09%	< .01%	---
6×10^{19}	900	1.925×10^{19}	< .01%	< .01%	---
6×10^{19}	1000	2.304×10^{19}	< .01%	< .01%	.05%
2×10^{20}	800	6.76×10^{19}	30%	4.4%	---
2×10^{20}	850	7.82×10^{19}	16%	2.7%	---
2×10^{20}	900	8.59×10^{19}	8.3%	1.4%	---
2×10^{20}	1000	1.00×10^{20}	2.8%	< .01%	1.3%

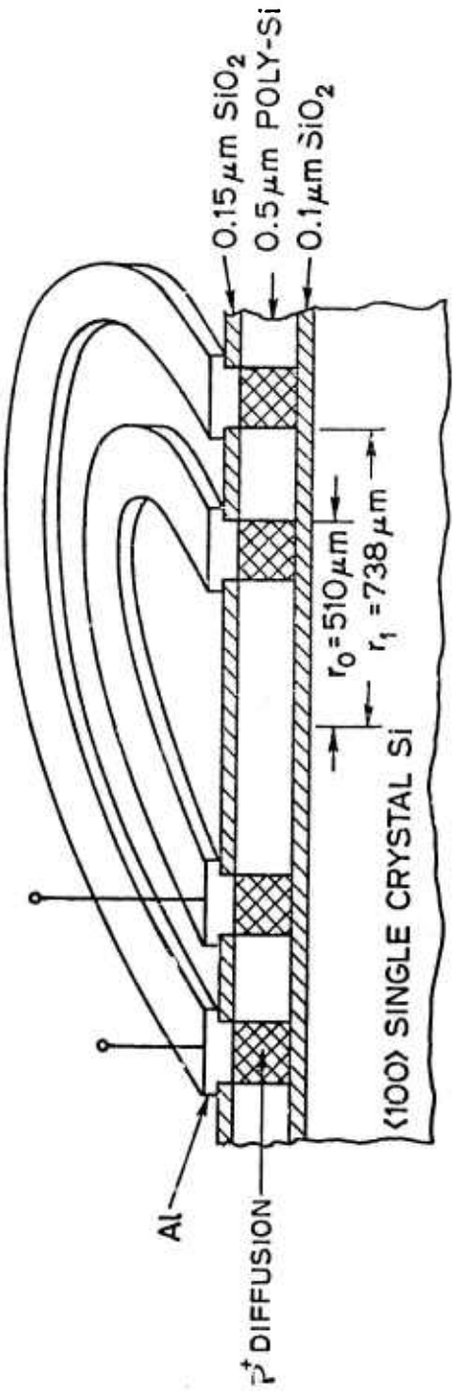


Fig. 3.1. Ring and dot structure for resistivity measurements.

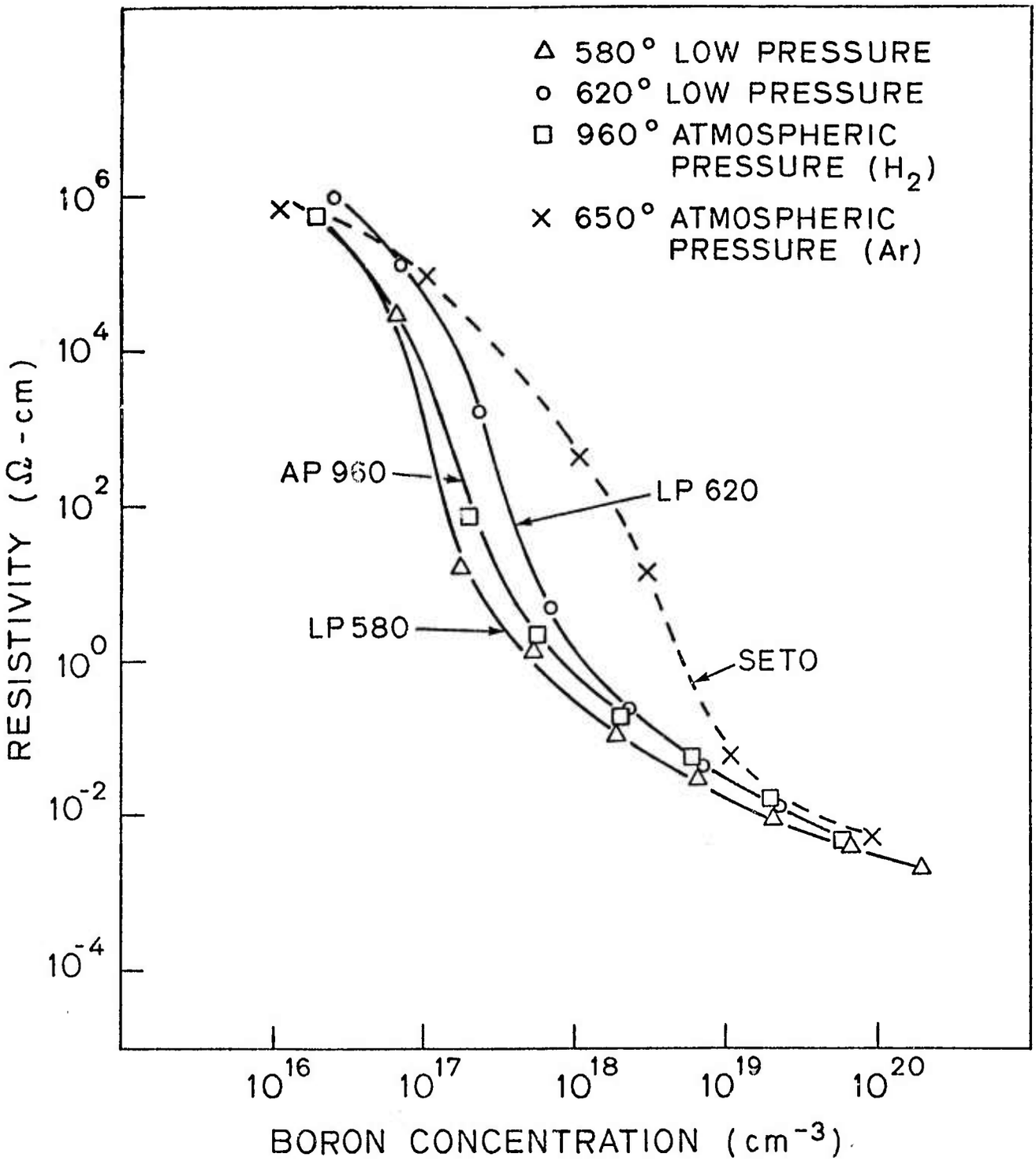


Fig. 3.2. Resistivity of boron-doped poly-silicon as a function of average dopant concentration. (ion-implanted and annealed)

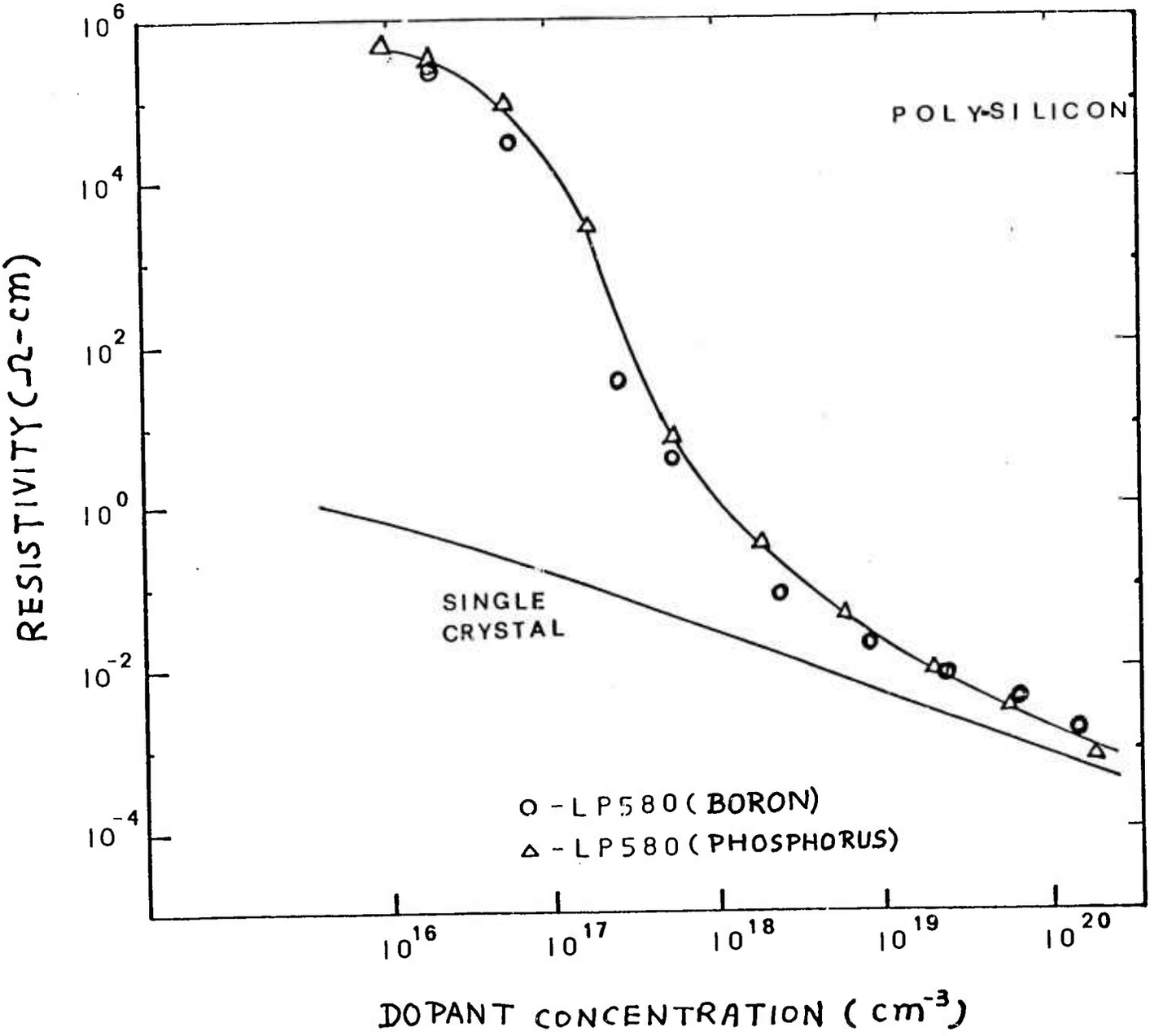


Fig. 3.3. Comparison of boron implanted and phosphorus-implanted low pressure films as functions of average dopant concentration.

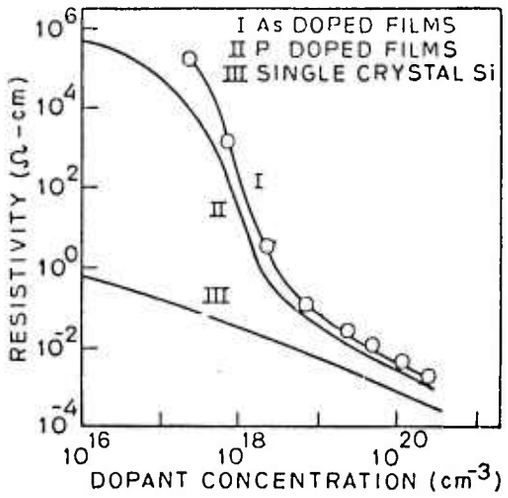


Fig. 3.4. Resistivity as a function of dopant concentration.

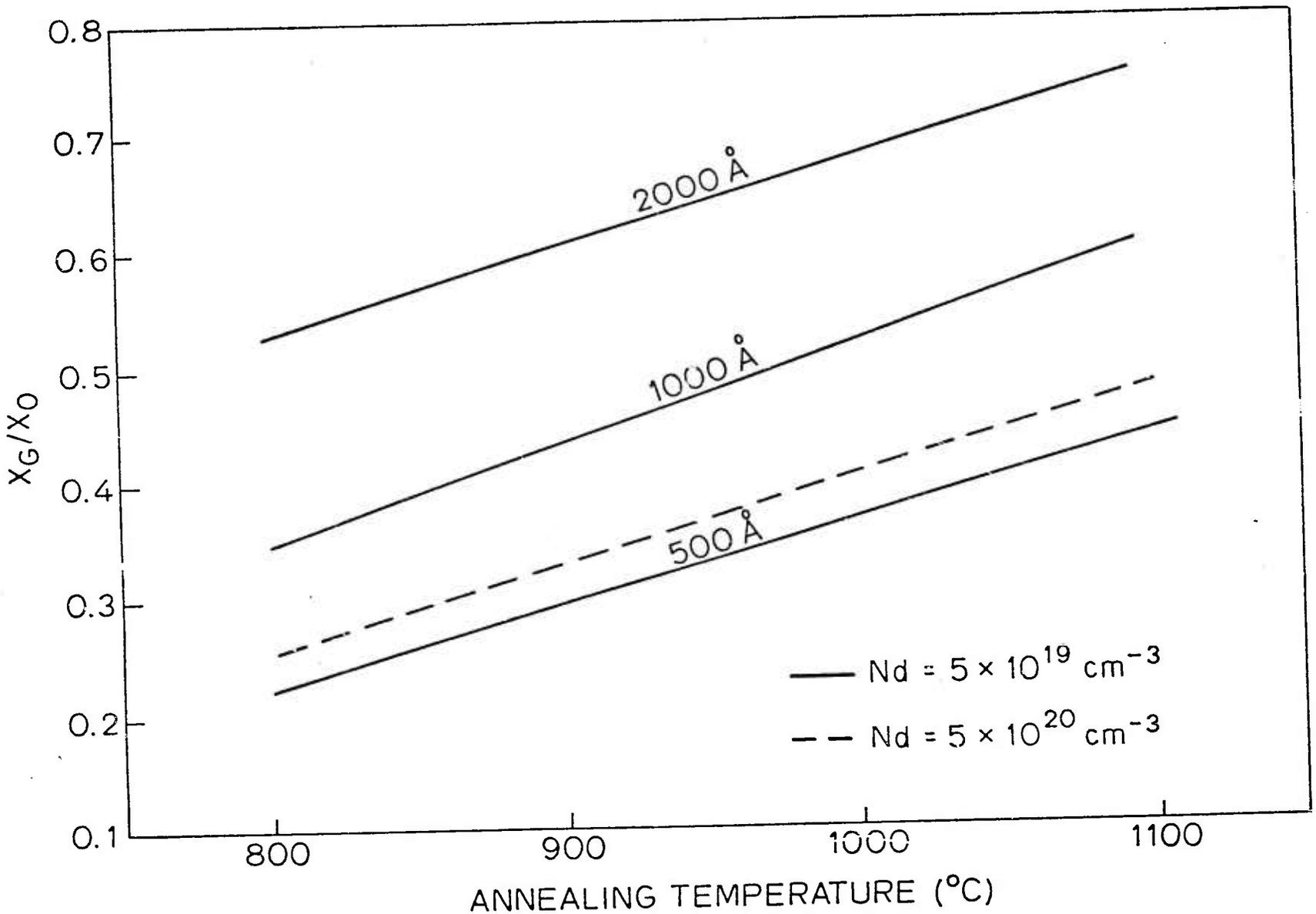


Fig. 3.5. Fraction of dopant atoms in the bulk of the grains calculated as a function of annealing temperature for grain sizes of 2000, 1000, and 500 Å. The solid lines are for $N_d = 5 \times 10^{19}$ while the behavior for $N_d = 5 \times 10^{20}$ is also shown for smallest grain size (dashed line).

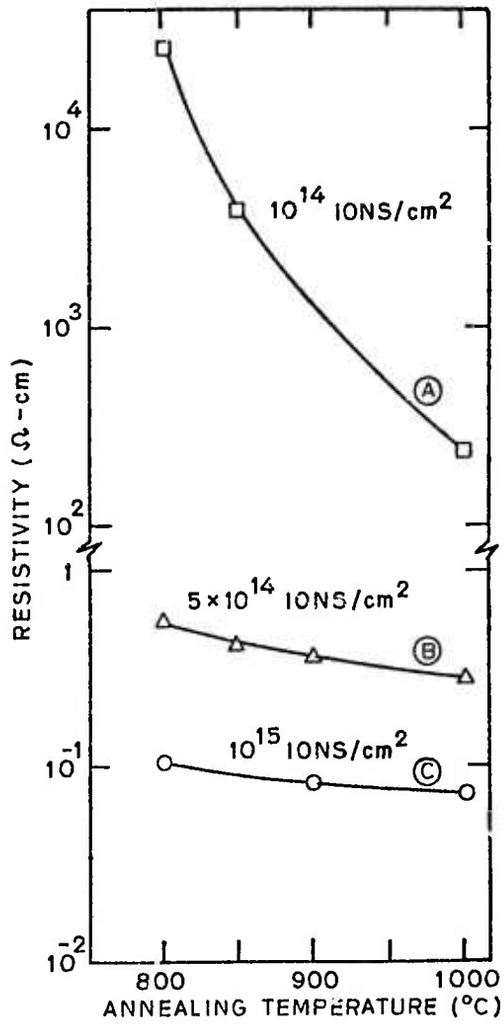


Fig. 3.6. Resistivity as a function of final annealing temperature. The films annealed at 800, 850 and 900°C were first annealed at 1000°C.

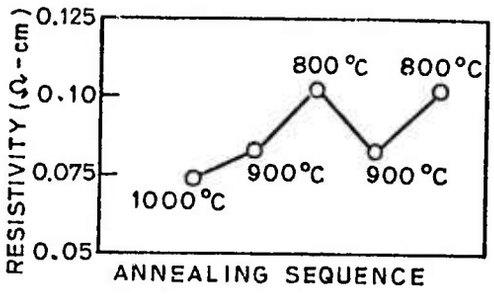


Fig. 3.7. Resistivity of the polycrystalline-silicon film implanted with 10^{15} ions/cm^2 of arsenic, after successive annealing at different temperatures.

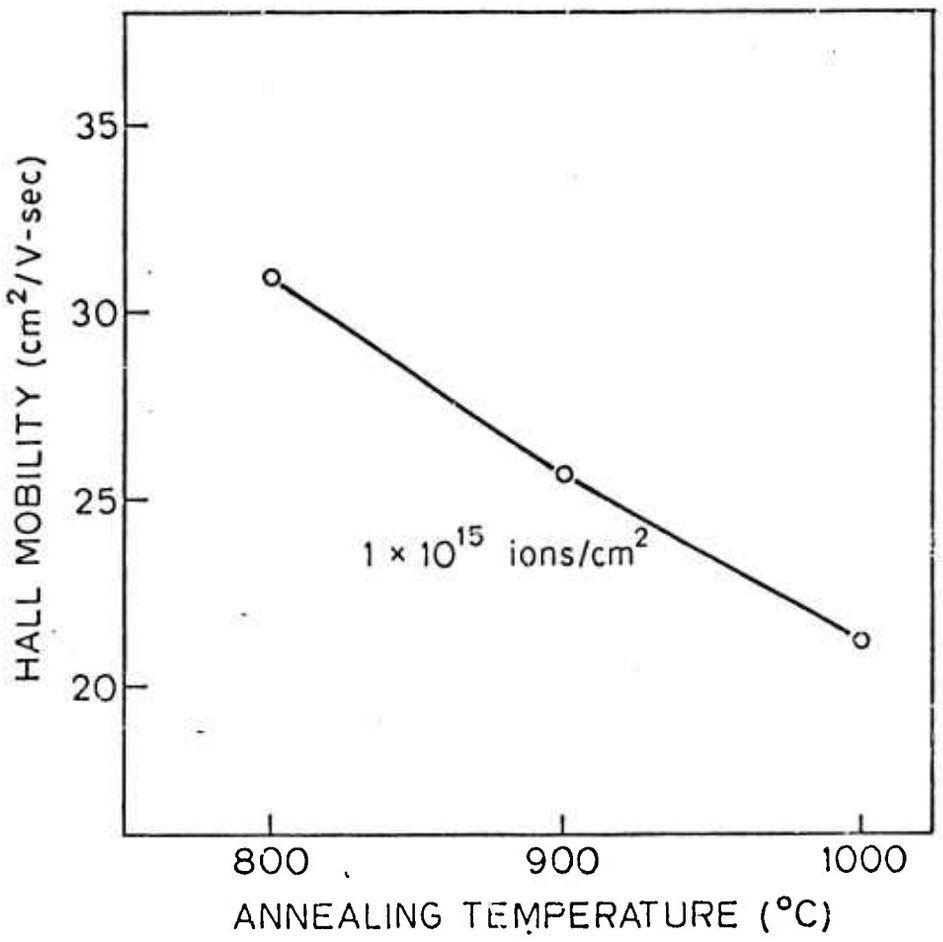


Fig. 3.8. Room-temperature Hall mobility as a function of annealing temperature. The films annealed at 800° or 900°C were first annealed at 1000°C.

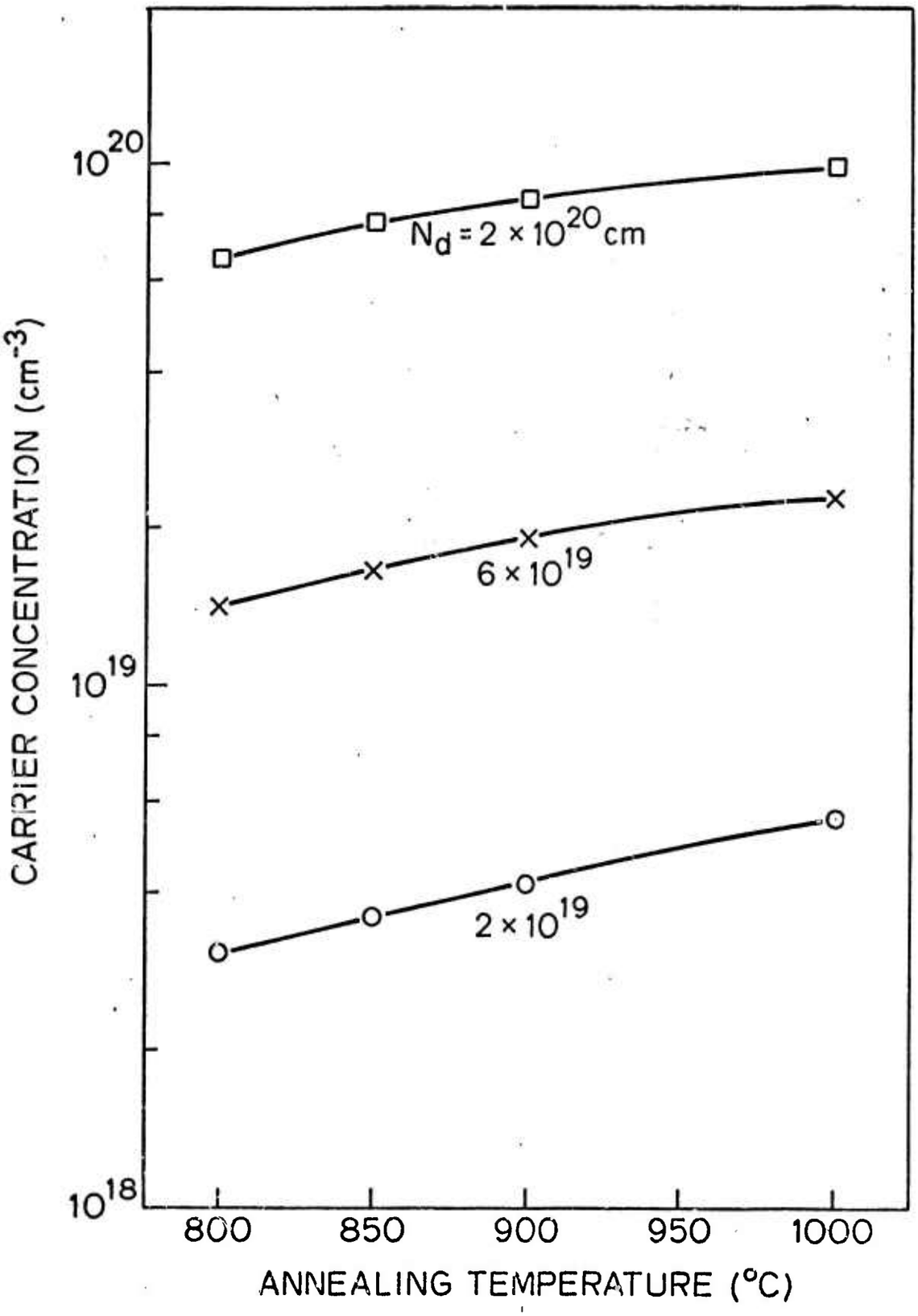


Fig. 3.9. Carrier concentration in the polycrystalline-silicon films as a function of final annealing temperature for average dopant concentrations of 2×10^{20}, 6×10^{19} and 2×10^{19} cm^{-3}.

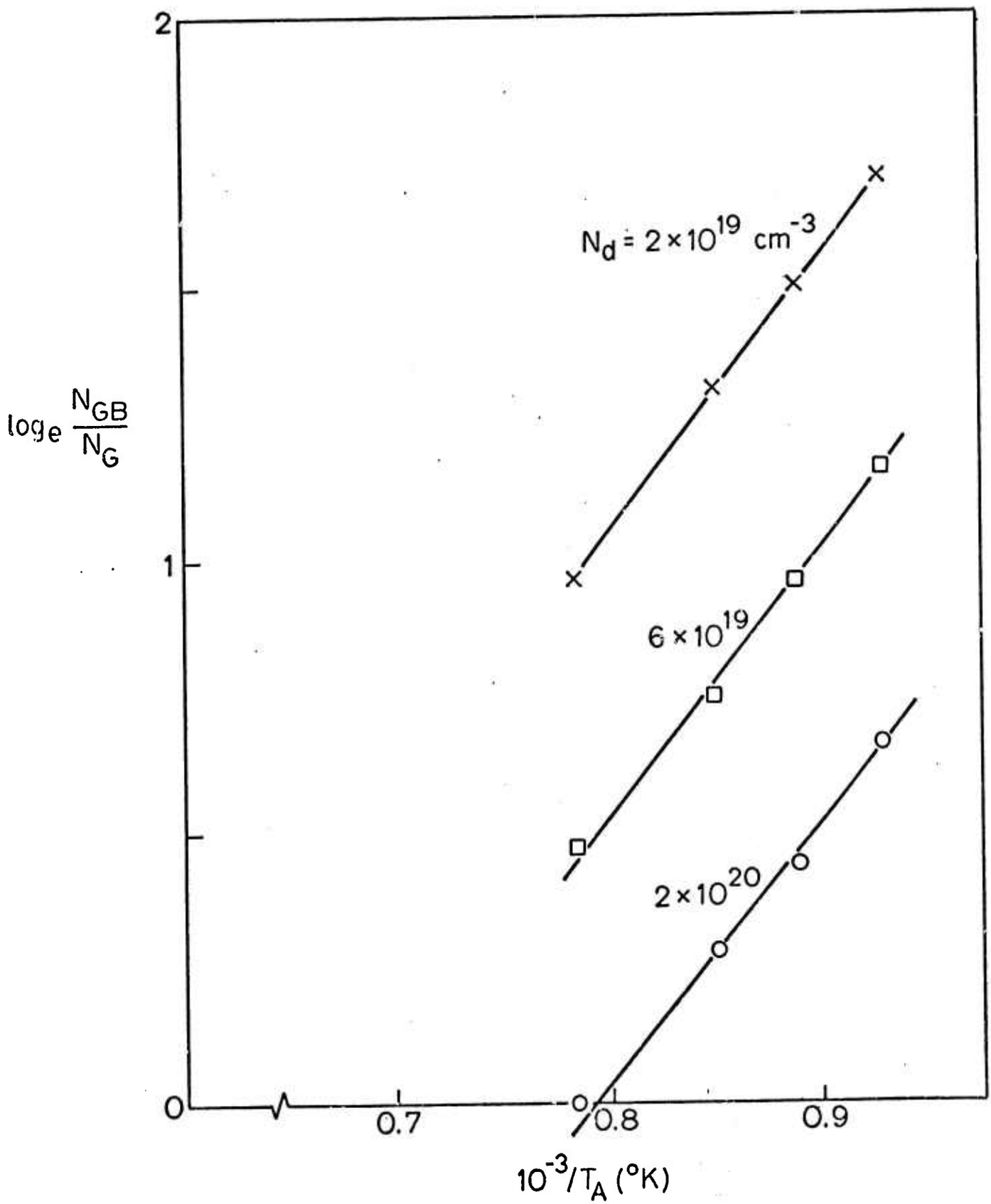


Fig. 3.10. N_{GB}/N_G as a function of reciprocal of final annealing temperature for average dopant concentrations of 2×10^{19}, 6×10^{19} and 2×10^{20} cm^{-3}.

(a)

Fig. 3.11. Transmission electron micrographs of arsenic doped polycrystalline-silicon films with average dopant concentrations of (a) 2×10^{19}, (b) 6×10^{19} and (c) 2×10^{20} cm^{-3} all annealed for one hour at 1000°C.

(b)

(c)

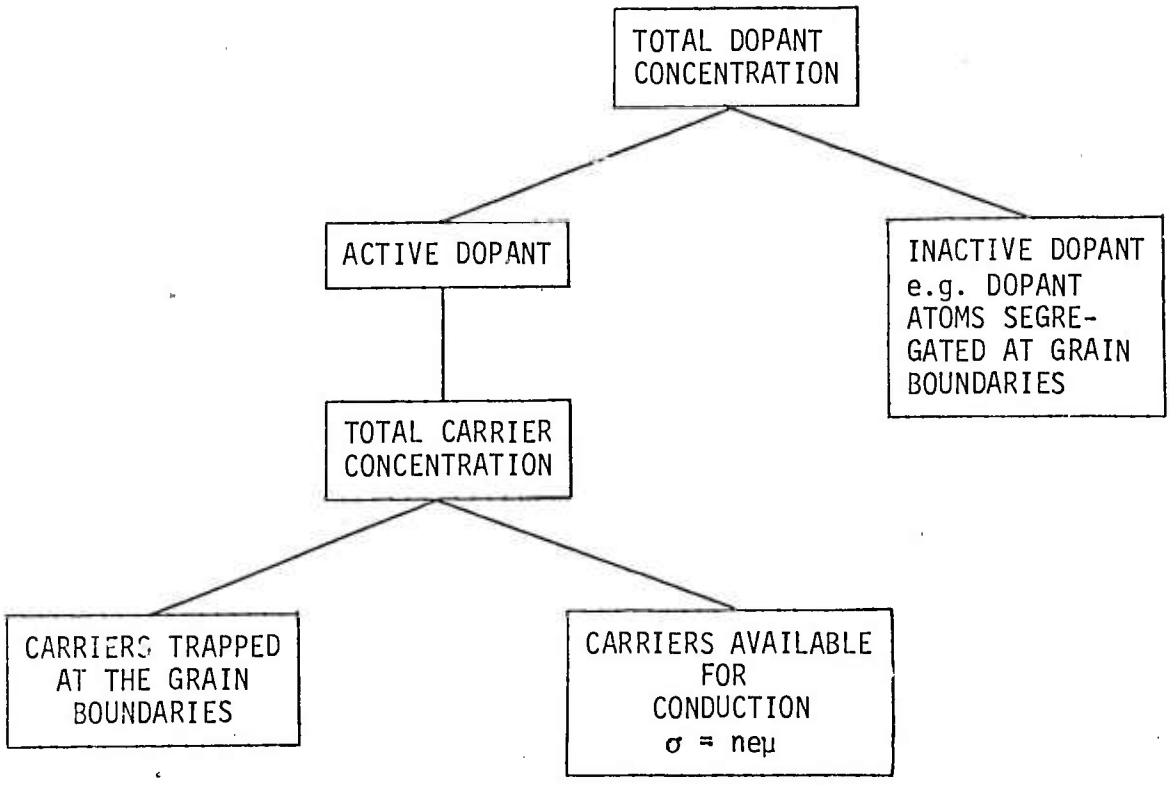


Fig. 3.12. Schematic diagram showing that both segregation of dopant atoms to grain boundaries and carrier trapping must be considered in relating the resistivity to the average dopant atom concentration.

4. REFRACTORY METAL SILICIDES

K.C. Saraswat, F. Mohammadi, J. Rouse and C.R. Helms

4.1 Introduction

With the advances in the integrated circuits technology the device dimensions are continuously decreasing and at the same time the chip area and the complexity is continuously increasing. The samller size of the devices makes them faster which should improve the circuit performance. However because of the larger chip size the interconnection lines are becoming very long. The time delays associated with parasitic capacitance and resistance associated with the interconnection lines can become appreciable. Thus even with very fast devices the overall performance of a large circuit could be seriously affected by the RC delays of the interconnections. As we move towards VLSI the problem becomes even more serious. The entire performance of the circuit could be governed by the RC delays of the interconnections.

The capacitance associated with a line is a problem of geometries used. By reducing the area of the line one could minimize the capacitance. However, the resistance is a problem caused by the material used. In the MOS integrated circuits of today aluminum is used as the last (top) layer of interconnections. Poly-silicon is largely used to form one or more layers of interconnections and gates of the MOS transistors. The sucess of MOSLSI can be largely attributed to the self aligned silicon-gate MOS technology. However, its properties are beginning to limit the performance of VLSI circuits. In general the requirements of the material used to form MOS gates and interconnections are:

1. Low resistivity to reduce RC delay times
2. During fabrication ability to withstand:

a. Chemicals

b. High temperature processing

c. Thermal oxidation

3. Ability to be defined into fine patterns

4. Stable contacts to the final metal layer

5. Good MOS properties

6. Low electromigration

Table 4.1 shows a comparison of the properties of some of the materials which have been either used or proposed for this application. Poly-silicon meets most of the requirements except its high resistivity causes degradation in the speed of the circuits because of R-C delay times. Although the grain size in as-deposited poly-Si can be very small, subsequent high temperature processing increases it markedly [4.1]. This causes a problem in defining very fine lines. Refractory metals such as Mo and W can meet most of the above requirements, but they cannot withstand high temperature oxidizing ambients, because their oxides are generally voaltile, and they cannot withstand chemical reagents commonly encountered during fabrication of integrated circuits. These two requirements can be met by the silicides of some of the refractory metals, such as $MoSi_2$ and WSi_2. It is evident from Table 4.1 that not all of the silicides are suitable to form the gates and interconnections. There are a large number of questions which need to be answered before any of them could be used. We have to establish their compatibility and the advantages.

In the first phase of this program being conducted at Stanford, we studied some properties of sputter deposited tungsten silicide to determine its compatibility with MOS fabrication technology [4.2,4.3]. In the present report additional properties of tungsten silicide have been

studied to establish the feasibility of its use as a material for MOS gate electrodes and low resistivity interconnections. Formation of silicides by CVD [4.4] and laser annealing [4.5] have been investigated. Thermal oxidation of WSi_2 has been studied [4.6-4.9]. WSi_2 gate MOS devices have been fabricated and fully characterized [4.10,4.11].

4.2 A Summary of the Major Accomplishments to Date

During the last two years following major accomplishments have been achieved.

(1) Deposition techniques: Deposition of silicides has been studied by three different techniques; (a) RF diode sputtering from a hot pressed target made of WSi_2 [4.2], (b) laser irradiation of W and Mo films on silicon to form WSi_2 and $MoSi_2$ [4.5], and (c) formation of WSi_2 by CVD of W on silicon and subsequent heating [4.4].

(2) Study of the properties of the silicides: A study of the properties of the films deposited by each of the three techniques has been done to investigate their process compatibility. The films deposited by RF diode sputtering were examined in much more detail [4.2,4.3]. Thin films of WSi_2 were deposited by sputtering on oxidized silicon, with and without a coating of poly-Si. As-deposited films were amorphous, with a resistivity of 6×10^{-4} Ω-cm. Upon annealing above 900°C in N_2, the films became polycrystalline. The resistivity decreased and the average grain size increased as a function of time and temperature. A minimum resistivity of 10^{-4} Ω-cm was obtained. Chemical reagents commonly used for cleaning and etching, when used with WSi_2, showed behavior similar to poly-Si. Al contacts to WSi_2 were stable up to 600°C for $WSi_2/SiO_2/Si$ structure and up to 550°C

for WSi_2/poly-Si/SiO_2/Si structure. These are significantly higher than the temperatures commonly used for sintering Al contacts to Si.

(3) Thermal oxidation of WSi_2: Thermal oxidation of WSi_2 has been studied [4.6-4.9]. Thin films of WSi_2 were deposited by sputtering on (100) silicon and on oxidized silicon, and then oxidized in steam atmospheres at 1000, 1100 and 1200°C for times up to 30 minutes. Steam oxidation gave SiO_2 films of good quality. For WSi_2/Si as well as WSi_2/SiO_2/Si samples, as the oxidation proceeded, WSi_2 reacted with steam to form SiO_2 and W_5Si_3. The rate of oxidation and activation energy were different for the two types of substrates. For WSi_2 films deposited directly on Si, the oxidation process was significantly modified because Si atoms rapidly diffused through the silicides and took part in the oxidation process, and thus affected the conversion of WSi_2 to W_5Si_3. However, with WSi_2 films deposited on oxidized Si, the absence of free Si atoms resulted in the faster transformation of WSi_2 into W_5Si_3.

(4) WSi_2 gate MOS devices: Tungsten silicide gate MOS capacitors have been fabricated on n-type (100) silicon [4.10,4.11]. Gate oxide thickness was varied from 600 to 4000 Å. WSi_2 was deposited by sputtering. High frequency C-V measurements were performed to determine threshold voltage and flat-band voltage as a function of gate oxide thickness. From the relationship between gate oxide thickness and flat-band voltage, the work function of WSi_2 was estimated to be about 4.8 eV and fixed interface charge density, Q_f, was estimated to be about 9×10^{10} cm^{-2}. From the quasi-static C-V measurements, the density of surface states at the mid-band was found to be about 2×10^{10} cm^{-2} eV^{-1}. The details of the accomplishments during the past year of this program are given in the remainder of this chapter.

4.3 Silicide Formation Techniques

Historically several techniques have been used to form the silicides:

(1) Evaporation of metal on silicon and subsequent reaction by thermal heating [4.12].

(2) Deposition by sputtering from a target made of the silicide [4.2,4.13].

(3) Codeposition of metal and silicon from two independently controlled sources and subsequent high temperature annealing to complete reaction [4.14].

(4) Chemical vapor deposition by combining gaseous chemicals of silicon and metal [4.15].

(5) Laser induced reaction between metal deposited on silicon.

The first technique has been tried. The results have been very successful for metals like Pt, Pd and Ni, but for higher temperature refractory metals like W, Mo and Ta the results have not been encouraging. The reaction between the metal and Si in these cases seems to be [4.16] highly dependent upon the interface properties and therefore the presence of even a slight amount of contamination such as a few monolayers of SiO_2 can retard the reaction. Thus this technique has rarely been used to form $MoSi_2$, WSi_2 and $TaSi_2$. The second and third techniques have been widely used to study the properties of silicides and to determine the feasibility of their use in the integrated circuits [4.2,4.13,4.14], however, their use for reliable and low cost fabrication of VLSI circuits has been limited because of a variety of problems, such as cleanliness, reproducibility, step coverage, stress in the films, production cost, etc. During the initial phase of this program we have used the second technique. However other techniques need to be explored in order to overcome the

shortcomings addressed above. In this report we have investigated CVD [4.4] and laser induced reaction [4.5].

4.3.1 Laser Irradiation of W and Mo Films on Silicon to Form Silicides

Films of W and Mo 1000 Å thick were deposited on 5-10 Ω-cm, p-type (100) Si substrate. Tungsten films were deposited by planar magnetron sputtering and Mo films were deposited by RF diode sputtering. Laser irradiation was performed by a high intensity CW Nd:YAG laser (λ = 1.064 μm) operating in TEM_{00} mode. A schematic diagram of the equipment used is shown in Fig. 4.1. The laser beam is held stationary. Scanning is obtained by moving the wafer mechanically. The wafer is held on a vacuum chuck, mounted on an electronically controlled translation stage with X-Y motion provided by a stepping motor. The scanning is obtained by moving the wafer in one direction at a constant speed and then stepping it by a small distance in the other direction. The scanning speed was kept between 5-6 cm/sec in all of the experiments. The minimum spot size of the focused laser beam was about 47 μm. The energy density was varied by changing the spot size up to 200 μm, while the total power was kept constant at 27.5 watts. By overlapping the successive scans continuously reacted layers on substrates as large as 7.5 cm in diameter were obtained. The reacted layers were examined by glancing angle x-ray diffraction technique.

For the wafers with W films, for beam diameter larger than 160 μm, no reaction was observed and only a (110) W diffraction peak was obtained. For beam diameters less than 60 μm, intense reaction between W and Si took place. The reacted areas were easily visible even by the naked eye. X-ray diffraction gave peaks of various orientations of WSi_2 and (111) Si.

Since the substrates were (100) Si, the (111) Si peak was the result of laser irradiation--probably Si melted and diffused through W and re-crystallized. Between beam diameters of 60 μm and 160 μm only WSi_2 diffraction peaks were obtained. It was found that in all the cases no W peak was obtained. Figure 4.2 shows the result of x-ray diffraction of the sample irradiated with 110 μm diameter laser beam. As shown in the figure, diffraction peaks of WSi_2 of orientations (002), (101), (110), (103) and (112) were observed. The (101) and (110) WSi_2 peaks were the strongest. No other phase of the W-Si system or free W was observed.

Similar results were obtained for 1000 Å thick Mo films. The laser power was kept at 27.5 watts and the beam diameter was changed from 60 μm to 160 μm. In all cases the reaction between Mo and Si occurred, but the diffraction peaks were very weak. Figure 4.3 shows the results of x-ray diffraction for the sample annealed with a beam diameter of 110 μm. Diffraction peaks of several orientations of $MoSi_2$ were obtained. All, except two, were ambiguous because peaks of several phases of Mo-Si systems lie close to each other. Since the peaks were very weak it was not possible to identify them properly. However, since all other peaks were of $MoSi_2$ it is improbable that any other phase was present.

4.4 Formation of WSi_2 by CVD of W on Si and Subsequent Heating

Deposition of binary systems such as WSi_2 is not a straightforward technique. Normally the deposited films contain all the silicide phases, along with free silicon and the metal [4.15]. Thus much more work needs to be done to do CVD of silicides for VLSI applications. As a step

towards reaching that goal, one method is to combine the first and the fourth technique, i.e., deposit the metal by CVD on silicon and then subsequently heat them to form the silicide. Since prior to the deposition, the surface of silicon can be cleaned by in situ vapor phase etching the interdiffusion and reaction between the metal and silicon takes place readily.

In our work WSi_2 has been formed by this technique. Tungsten was deposited from WF_6 source in a horizontal, rf-heated reactor operating at atmospheric pressure. A schematic diagram of the reactor is shown in Figure 4.4. Few hundred Å thick layers of tungsten were deposited either in N_2 ambient or by H_2 reduction of WF_6 [4.17] in the temperature range of 500-600°C. Deposition in N_2 ambient took place by the reaction between WF_6 and Si from the substrate

$$2WF_6 + 3Si \rightarrow 2W + 3SiF_6\uparrow \qquad (4.1)$$

Obviously silicon was consumed from the substrate. Deposition was found to be sensitive to the surface cleanliness. Thus if the surface was etched with HCl prior to the deposition of W the uniformity of the deposition was markedly improved. The rate of deposition was found to be a function of time. As the thickness of W increased the deposition rate slowed down, probably because the process was controlled by the diffusion of silicon or WF_6 through the deposited tungsten. Since this method had shortcomings, deposition of W by hydrogen reduction of WF_6 was tried.

$$WF_6 + 3H_2 \rightarrow W + 6HF\uparrow \qquad (4.2)$$

In this case the deposition was not sensitive to the surface cleanliness, deposition rate was a linear function of time and the uniformity was

controlled by deposition parameters rather than the surface quality of the substrate. The as-deposited films of W were very smooth with a resistivity of about 1.8×10^{-5} Ω-cm.

In order to form WSi_2 the samples were annealed immediately after the deposition in the same reactor, at 800°C, in hydrogen ambient. Figure 4.5 shows the scanning electron micrographs of four samples using different techniques to deposit tungsten and form silicide. In general following conclusions can be drawn from this figure:

(a) Deposition in H_2 gives better results than deposition in N_2

(b) HCl etching improves the quality of the films

The optimum conditions for the formation of WSi_2 can be summarized as:

HCl ETCH TEMPERATURE	1180°C
W DEPOSITION TEMPERATURE	600°C
H_2 FLOW	28 l/min
WF_6 FLOW	5 cc/min
WSi_2 FORMATION TEMPERATURE	800-1000°C
WSi_2 GROWTH RATE	~500 Å/min of WF_6 FLOW

Figure 4.6 shows the result of x-ray diffraction of a WSi_2 film formed under the conditions described above. It is evident that neither free tungsten nor any other phase of tungsten silicide besides WSi_2 is present in the film. The surface quality of these films was found to be excellent. After full anneal the resistivity of the films was about 7×10^{-5} Ω-cm. All other properties were found to be similar to the WSi_2 deposited by codeposition [4.14] or sputtering [4.2].

Further work is being continued to optimize this process.

4.5 Thermal Oxidation of WSi_2

Formation of an insulator layer on top of the gate electrode and the interconnection layer is necessary to isolate it from subsequently deposited interconnection layers of metal or silicide. Vapor deposited SiO_2 and Si_3N_4 can be used for this application. However, thermally grown SiO_2 is generally of superior quality. Even if thermal oxidation is not needed to form an insulating layer, it is of extreme importance to know how well WSi_2 can withstand the oxidizing environment generally encountered during the processing. The feasibility of thermal oxidation of WSi_2 was demonstrated by Zirinsky et al. [4.18] by oxidizing WSi_2/Poly-Si/SiO_2/Si structure in steam at 1000°C. In this work we have studied steam oxidation of WSi_2 deposited on single crystal silicon as well as oxidized silicon. In the latter case the substrate is essentially SiO_2, which can be considered inert for all practical purpose.

4.5.1 Oxidation Kinetics

Single crystal Si wafers with (100) orientation were thermally oxidized to grow 1000 Å SiO_2. WSi_2 films were deposited on top of the SiO_2/Si (100) and Si (100) substrates using rf diode sputtering in an argon atmosphere. The substrate temperature was kept below 300°C and the rate of deposition was 360 Å/min with a peak voltage of 1.5 KeV and an rf power of 280 watts. The target was made of hot pressed WSi_2. The last step in the cleaning of the samples was an etch in dilute HF. To further ensure the removal of residual SiO_2, the samples were sputter etched prior to the deposition of WSi_2 to ensure the removal of native oxide on Si. Auger electron spectroscopy and electron microprobe measurements showed that in the as-deposited films the ratio between the number of silicon atoms and the number of tungsten atoms was

approximately 2:1. X-ray diffraction analysis showed no specific peaks of WSi_2 indicating that the unannealed films were amorphous. Annealing of the samples was done in a nitrogen atmosphere at a temperature of 1000°C for 60 min. Annealing resulted in crystallized films of WSi_2 free of having any marked preferred orientation [4.2].

The annealed films were oxidized in steam with N_2 as a carrier gas. The oxidations were carried out at temperatures from 1000°C to 1200°C, for periods from 5 to 30 min. In all experiments WSi_2/Si and $WSi_2/SiO_2/Si$ samples were oxidized along with a single crystal Si (100) wafer as a test sample. In most of the experiments 4000 Å thick WSi_2 films were used. However, to study the effect of WSi_2 thickness on the kinetics of oxidation, films of 1000, 2000 and 8000 Å thickness were also used. After oxidation, strips of SiO_2 were removed by selective etching of SiO_2 in a dilute HF solution, and the thicknesses of grown oxides were measured by Taly Step. Finally the structure of the oxidized films was analyzed by Auger electron spectroscopy and x-ray diffraction. After oxidation at 1000, 1100 and 1200°C in steam, the films appeared to be very smooth and of good surface quality by naked eye and optical microscope observations; and, no WO_3 could be detected. The stoichiometric uniformity of the SiO_2 films was verified by Auger electron spectroscopy.

During the initial experiments, in the range of indicated times of oxidations, the thickness of SiO_2 grown on WSi_2 films of different thicknesses was found to be identical, therefore, all of the subsequent experiments were done only with 4000 Å thick WSi_2 films. Figures 4.7 and 4.8 show the thickness of SiO_2 as a function of time for WSi_2/Si and $WSi_2/SiO_2/Si$ samples, respectively. For comparison

the oxide thickness grown on (100) Si wafers is also plotted. Notice that the WSi_2/Si samples behave more like the single crystal Si than the $WSi_2/SiO_2/Si$ samples. Figures 4.9 and 4.10 show the SiO_2 thickness plotted vs 1/T. From these four figures the kinetics of oxidation can be seen to obey the equation:

$$\bar{X}^{\bar{n}} = K_o \, t \, e^{-E/kT} \qquad (4.3)$$

where $\bar{X}$ is the SiO_2 thickness, t is the time of oxidation, T is the temperature of oxidation, $\bar{n}$ is a constant giving some indication of oxidation mechanisms ($\bar{n}$ = 1 indicates surface reaction rate control and $\bar{n}$ = 2 indicates a reaction limited by diffusion), E is the activation energy for the process controlling the formation of SiO_2, k is the Boltzmann's constant, and K_o is a constant related to several factors including the physical properties of oxidizing species and the structure of the oxide.

From the plot of the thickness of SiO_2 vs oxidation time for WSi_2/Si structure (Figure 4.7), $\bar{n}$ has been calculated to be 1.82. From the plot of the thickness of SiO_2 vs 1/T for the same structure (Figure 4.9), the activation energy has been found, E = 23 Kcal/mole. From Figures 4.9 and 4.10, $\bar{n}$ and E are 1.8 and 8 Kcal/mole, respectively, for $WSi_2/SiO_2/Si$ samples. Note that for the two types of samples the value of $\bar{n}$ is about the same but there is a large difference in the activation energies. In general a value of $\bar{n}$ = 2 implies a diffusion process driven by a constant difference in free energy, where the growing layer of reaction products results in a progressively decreasing chemical potential gradient. Therefore a value of $\bar{n}$ = 1.8 for oxidation of WSi_2 implies that the rate controlling process is diffusion. The large

difference between the activation energies of the two types of samples indicates that the process controlling the formation of SiO_2 is different in the two cases.

4.5.2 Mechanism of Oxidation

In order to understand the mechanism of oxidation, after oxidation the grown SiO_2 was etched off and x-ray diffraction was performed on the samples. For both types of samples, besides WSi_2 diffraction peaks, W_5Si_3 peaks were also observed. This indicates that upon reacting with H_2O, WSi_2 is transformed into W_5Si_3 and SiO_2. This result is contrary to that obtained by Zirinsky et al. [4.18] where they did not see any formation of W_5Si_3 for oxidation of WSi_2 formed by co-evaporation of W and Si. However, for the thermal oxidation of bulk $MoSi_2$, formation of Mo-rich silicide phases has been observed [4.19,4.20]. A comparison was made between the two types of samples, where the thickness of deposited WSi_2 and that of grown SiO_2 were kept the same. X-ray diffraction peaks of WSi_2 were more intense where WSi_2 was deposited on Si, whereas peaks of W_5Si_3 were more intense where WSi_2 was deposited on SiO_2 as shown in Figure 4.11(a). For comparison x-ray diffraction results of an annealed but not oxidized sample of WSi_2 are shown in Figure 4.11(b) showing only WSi_2 peaks. This implies that the transformation of WSi_2 to W_5Si_3 is faster for the samples where WSi_2 was deposited on SiO_2. For the samples where WSi_2 was deposited on Si, at the temperature of oxidation, Si atoms can rapidly diffuse through the layer of silicide and can take part in the chemical reaction occurring at the oxide-silicide interface [4.18]. Since for these samples the values of $\bar{n}$ and E are fairly close to that obtained for oxidation of silicon [4.21], it can be concluded that the basic mechanism of oxidation is also similar and the

rate controlling process is diffusion of steam through the grown SiO_2. In the case of $WSi_2/SiO_2/Si$ samples, since free silicon is not available, the formation of SiO_2 is taking place by reaction between only H_2O and WSi_2, resulting in rapid formation of W_5Si_3. The value of $\bar{n} = 1.8$ still indicates that the rate controlling process is diffusion, however, the value of E = 8 Kcal/mole is much lower than the previous case. Therefore it cannot be concluded whether the oxidation process is controlled by the growth of SiO_2 or W_5Si_3, or both. One must conclude that the oxidation process is complex, probably involving more than one simple mechanism. By necessity the measured temperature dependence must be called an apparent activation energy.

These results are somewhat different compared to those obtained by Zirinsky et al. [4.18] on the oxidation kinetics of WSi_2. Zirinsky et al. [4.18], using dual electron-beam co evaporated WSi_2, found that SiO_2 formation on WSi_2 layers deposited on silicon proceeds completely via diffusion of Si through the WSi_2. The WSi_2 film is left reasonably undisturbed. They report that, for WSi_2 deposited on SiO_2, the oxidation proceeds with the formation of pure tungsten which is lost presumably as a volatile oxide. In our work [4.6,4.7] we have found, however, that sputter-deposited WSi_2 could be oxidized in either case with the formation of W_5Si_3 as a partial reaction product. In order to remove this discrepancy we have done a detailed analysis of the oxidation process of the sputter deposited WSi_2 films using Auger sputter profiling.

Figures 4.12(a) and 4.12(b) show the two structures studied. A has the structure oxide/silicide/silicon and B has the structure oxide/silicide/oxide/silicon with approximate layer thicknesses shown.

These samples were identical to those reported earlier in this section. The WSi_2 layers were formed by sputter-depositing a WSi_2 target onto the substrate material. The samples were then oxidized in steam at 1100°C for 10 min, resulting in specular oxides well bonded to the substrate. The Auger electron beam was operated at 5 A, 4.5 KeV, and rastered over a region 500 m on a side. The sputter crater was made with 1.75 KeV Ar ions (these techniques of Auger sputter profiling are described in detail elsewhere [4.22,4.23]). We profiled each type of sample three times to assure good reproducitiblity.

Shown in Figure 4.13 is a sputter profile of the silicide as-deposited on an SiO_2 substrate with no further processing. We used this sample as a standard to calibrate W and Si levels in the oxidized samples. We monitored the Si 92 eV and W 179 eV transitions. The measured height of the silicon transition in the silicide was lower than the signal from elemental silicon; yet, the actual silicon concentrations of these two regions are nearly equal. This is probably a result of preferential sputtering whereby silicon is removed from the surface more rapidly than the tungsten. This resulted in a steady-state surface condition of lower silicon concentration in the WSi_2 relative to the elemental silicon. Also evident is a slight decrease of the tungsten and silicon levels with increased depth, which may be a result of backscattering effects. These effects become particularly important for layers of widely differing atomic number such as we have in the present case. The substrate oxide is less effective than the tungsten silicide in reflecting the electron beam so that a slight decrease in W and Si signal levels would be expected as sputtering reaches the end of the silicide layer.

Note that the silicon signal shown corresponds to elemental Si (Si^0). The energy of Auger transitions for Si in SiO_2 (Si^{+4}) is shifted by 14 eV from that of elemental Si (Si^0). The Si^{+4} signal was not monitored beyond the SiO_2-WSi_2 interface because of excessive interference from the Si^0 signal of the silicide. It should be noted here that there is no noticeable change in the shape or position of the Si signal as it appears in the W silicides and elemental Si. It is also evident from the figure that the depth scales are not uniform due to the different sputter rates for SiO_2 and the silicides. The depth scales were determined from known thicknesses of each layer, and it was assumed that the sputter rates were constant through each layer.

Figures 4.14(a) and 4.14(b) are profiles of structures A and B, respectively (from Figure 4.12). Examination of structure A reveals an increase in the Si and O signals of the surface oxide upon approaching the interface. This is due to backscattered electrons from the WSi_2 layer, as discussed above. The SiO_2-WSi_2 interface is about 700 Å wide [4.24]. The details of the SiO_2-WSi_2 interface will be discussed in the next section.

Also included in Figures 4.14(a) and 4.14(b) are bars indicating the levels of W and Si in the standard. Again concentrating on structure A, it is evident that the tungsten signal increased and the silicon signal decreased upon oxidation. The oxide growth therefore consumed silicon from the silicide and decreased the thickness of the layer, resulting in higher tungsten and lower silicon concentrations.

The profile for structure B exhibits the same features as that for A, with one exception. The degree of tungsten enrichment and silicon depletion is greater in the case of structure B. Analysis of

the data reveals relative amounts of W and Si of 35% and 65%, respectively, for structure A and 39% and 61% for structure B. This is consistent with a model in which the silicide of both structures contributes silicon atoms to the growing SiO_2, with the substrate serving as an additional silicon source only in the case of a Si substrate.

As described earlier x-ray diffraction revealed that the silicide layer consisted of WSi_2 and W_5Si_3. Calculations from the ASP data yield numbers for the bulk composition of structures A and B of $0.97WSi_2 + 0.03W_5Si_3$ and $0.94WSi_2 + 0.06W_5Si_3$, respectively. These mixtures are nearly uniform throughout the bulk silicide regions, although there is apparently an excess of W_5Si_3 near the SiO_2/silicide interfaces, as will be discussed below. Nucleation sites for the reaction of WSi_2 to form W_5Si_3 are apparently present throughout the silicide region.

In contrast to our findings, Zirinsky et al. [4.18] found no W_5Si_3, but a free tungsten phase. We feel this apparent contradiction may be explained in part by the presence of impurities in our starting silicide material. The major impurity present in these silicides is carbon. A carbon profile for structure B is shown in figure 4.15. As shown in the figure, the carbon is present in the silicide but not in the SiO_2 substrate or grown SiO_2 layer. Using approximate Auger sensitivity factors gives a carbon concentration of $2 \cdot 10^{21}$ cm^{-3}. This carbon may form nucleation sites for the observed W_5Si_3 phase.

Details of the SiO_2-WSi_2 interface after oxidation appear in Figures 4.16(a) and 4.16(b). The excess tungsten curve was obtained by subtracting the WSi_2 component of tungsten from the data of Figure 4.14 using the Si (Si^0) 92 eV signal as a reference. The excess oxygen curve

was obtained by subtracting the SiO_2 component of oxygen in the oxide using the Si (Si^{+4}) 78 eV signal as a reference. The excess tungsten and oxygen plotted, therefore, refer to the excess over what would be accounted for by an interface mixture of SiO_2 and WSi_2.

Two important points emerge from this analysis.

(1) There is a peak in excess W near the interface which indicates that the W_5Si_3 phase, although apparently present throughout the silicide, has a higher concentration near the interface.

(2) There is additional W present at the interface in the form of an oxide which may be in the form of a $W_xSi_yO_z$ compound and may be responsible for the excellent adherence of the SiO_2 film.

Structure B looks much the same as structure A except for the increased amount of the enriched tungsten phase. This illustrates the extent to which structure B is more heavily depleted of Si than is structure A during oxidation.

We have examined two structures of steam-oxidized sputter-deposited WSi_2 layers using ASP to determine the silicide stoichiometry as a function of depth and the interface characteristics. Table 4.2 is a comparison of the composition of the silicide layer in the two structures. This data is consistent with a model in which Si is supplied by the silicide layer of both structures for reaction with oxygen to form SiO_2. Additional Si is supplied by diffusion through the silicide only in the presence of a Si substrate. This oxidation process converts a portion of the WSi_2 to W_5Si_3. The W_5Si_3 is distributed uniformly in a WSi_2 matrix throughout the bulk silicide region. The silicide layer must supply more silicon for oxidation when the substrate is SiO_2. Consequently, the silicide layer contains about 61% Si for an SiO_2 substrate

and 65% Si for a silicon substrate. This bulk silicide region is bound to the oxide by a transition region containing mixed phases of SiO_2, WSi_2, W_5Si_3, and possibly a $W_xSi_yO_z$ compound.

These results contrast with those of steam-oxidized WSi_2 which has been deposited by electron-beam coevaporation. Electron-beam evaporated WSi_2 on a silicon substrate forms an SiO_2 layer only by diffusion of Si through the WSi_2, with no conversion of WSi_2 to any other phase. For e-beam silicide on SiO_2, the oxide layer grows by conversion of the WSi_2 to a free tungsten phase. The results from electron-beam coevaporated silicides and our own sputter-deposited silicides may differ because of different starting compositions, different levels of contaminants, or morphological differences between the two materials due to the different deposition techniques. For example, it is well known that sputter deposited materials can have foreign materials present such as the inert sputtering gas or carbon as shown in this case. These materials may affect film morphology or provide nucleation sites for other phases such as W_5Si_3.

4.6 WSi_2 Gate MOS Devices

In our work so far we have demonstrated the compatibility of WSi_2 with the MOS fabrication technology. In this section fabrication of WSi_2 gate MOS devices, characterization of the properties such as, flat band voltage, threshold voltage, fixed interface charge density, fast surface states density and work function is described.

Single crystal n-type silicon wafers, with (100) orientation and 2-4 Ω-cm resistivity, were used as substrates to fabricate MOS capacitors with gate oxide thickness ranging from 600 to 4000 Å. Different oxidation times were used at 1000°C in steam to obtain different oxide

thicknesses. Following the oxidation all of the wafers were annealed together in N_2 at 1000°C for 30 min to ensure the same value of Q_f. A 2500 Å thick layer of WSi_2 was deposited on all samples using rf diode sputtering in argon atmosphere. The substrate temperature was kept below 300°C and the rate of deposition was 360 Å/min, with a peak voltage of 1.5 KeV and an rf power of 280 watts. The target was made of hot pressed WSi_2. Photolithography was performed to define the gates. A 2% NH_4F + 98% HNO_3 solution was used to etch WSi_2. The as-deposited WSi_2 was amorphous, therefore a 1000°C anneal in argon was performed for one hour to make the films polycrystalline [4.2]. This anneal reduced the resistivity of WSi_2 from 6×10^{-6} to 10^{-4} Ω-cm and also annealed any surface damage encountered during sputtering. A 1.5 μm thick layer of aluminum was deposited on the backside to obtain good ohmic contact. Finally, the wafers were annealed in forming gas (10% H_2 + 90% N_2) at 450°C for one hour.

C-V measurements were performed on the capacitors at a frequency of 1 MHz. From these measurements the gate oxide thickness, W_{OX}, flat-band voltage, V_{FB}, and threshold voltage, V_T, were estimated. W_{OX} was also independently measured by ellipsometer prior to WSi_2 deposition. Figure 4.17 shows a plot of V_{FB} and V_T as a function of W_{OX}. From these curves the value of work functions ϕ_m and Q_f can be determined.

Under thermal equilibrium, the flat-band voltage of a MOS structure fabricated on an n-type substrate can be expressed as follows [4.25]:

$$V_{FB} = \phi_{ms} + \frac{Q_{OX} W_{OX}}{\varepsilon_{OX}} \qquad (4.4)$$

Where Q_{OX} is the total oxide charge density, ε_{OX} is the permittivity of

SiO_2 (3.4×10^{-13} F/cm), and ϕ_{ms} is the metal-to-silicon work function. Figure 4.18 shows the band diagram of an MOS capacitor. From this figure ϕ_{ms} is given by

$$\phi_{ms} = \phi_m - \left(X + \frac{E_g}{2} - \Psi_n \right) \tag{4.5}$$

where ϕ_m is the metal work function, X is the semiconductor electron affinity, E_g is the band gap, and Ψ_n is the magnitude of the bulk potential of Si.

From Eq. (4.4) it is evident that the slope of V_{FB} vs W_{OX} curve will give the total oxide charge density and intercept at zero oxide thickness will give the value of ϕ_{ms}. Using Eq. (4.5), ϕ_m can be calculated since the value of X and E_g are known and Ψ_n can be calculated if the substrate doping density is known.

As shown in Figure 4.17 excellent linearity was obtained in the plot of V_{FB} vs W_{OX}, indicating the validity of the C-V method to obtain ϕ_m and Q_{OX}. Bias-stress C-V measurements showed presence of negligible mobile charge contamination and thus it can be assumed that $Q_{OX} = Q_f$ (the density of fixed interface charge). From Fig. 4.17 the value of Q_f is thus 8.7×10^{10} charges/cm^2. From the intercept of V_{FB} vs W_{OX} curve the value of ϕ_{ms} is 0.5 eV. From C-V plots the value of doping density was found to be about 2.5×10^{15} corresponding to Ψ_n = 0.31 eV. With X and E_g taken 4.05 eV and 1.12 eV, respectively [4.25], the value of the metal work function of WSi_2 was estimated to be 4.8 eV.

A value of 4.8 eV for the work function seems to be reasonable. Table 4.3 shows work function, ϕ_m, and Schottky barrier height, ϕ_B, of PtSi, Pd_2Si, WSi_2, and $NiSi_2$. The Schottky barrier heights have been taken

from the work of Andrews [4.26] and the work functions (except WSi_2) have been measured by Tsang using the C-V technique [4.27]. ϕ_B of WSi_2 falls between that of RhSi and $NiSi_2$. It is well known that ϕ_m bears a relationship to ϕ_B [4.25]. Therefore, it is reasonable to assume that ϕ_m of WSi_2 should fall in between that of RhSi and $NiSi_2$, which it does indeed according to our measurements.

High frequency and quasi-static C-V measurements [4.28] were performed on the capacitors with 1100 Å gate oxide thickness and surface state density distribution (N_{it}) in the band gap was calculated. Figure 4.19 shows the results of 1 MHz and quasi-static C-V measurements. Figure 4.20 plots N_{it} in the band gap calculated from the two C-V measurements. The value of N_{it} near the mid-band was about 2×10^{10} cm^{-2} eV^{-1}.

The values of Q_f and N_{it} are quite comparable to that of Al-gate or Si-gate MOS structures. The work function of WSi_2 is slightly higher than Al or Si. With accurate knowledge of the work function, the fabrication process can be designed accordingly to obtain proper threshold voltages of the MOS devices.

4.7 Conclusions

Use of tungsten silicide to form low resistivity interconnections and MOS gate electrode has been investigated. Various techniques to deposit thin films of tungsten silicide have been studied. Properties of these films have been studied to establish their compatibility with the integrated circuit technology. Thermal oxidation of WSi_2 has been studied in great detail and a model has been developed to explain the observed behavior. WSi_2 gate MOS capacitors have been fabricated which show excellent MOS properties. From this work we can conclude that

tungsten silicide can be used to form low resistivity interconnections and MOS gate electrodes.

REFERENCES

4.1 T.I. Kamins, M.M. Mandurah, and K.C. Saraswat, J. Electrochem. Soc., 125, 927 (1978).

4.2 F. Mohammadi and K.C. Saraswat, J. Electrochem. Soc., 127, 450 (1980).

4.3 J.D. Meindl et al., "Computer Aided Engineering of Semiconductor Integrated Circuits," TR No. DELET-TR DXG 569 Integrated Circuits Laboratory, Stanford University, Stanford, CA, July 1979.

4.4 K.C. Saraswat and F. Mohammadi, Extended Abstracts of the 157th Meeting of the Electrochem. Soc., Vol 80-2, 419 (1980).

4.5 F. Mohammadi, K.C. Saraswat, J. Beaudouin and J.D. Meindl, Extended Abstracts of the Spring 1979 Meeting, Electrochem. Soc., Boston, 79-1, (1979).

4.6 F. Mohammadi, K.C. Saraswat, and J.D. Meindl, Appl. Phys. Lett. 35, 529 (1979).

4.7 F. Mohammadi, K.C. Saraswat and J.D. Meindl. Proceedings of the Symp. on Thin Film Interfaces and Interactions, The Electrochem. Soc., Vol 80-2, 506 (1980).

4.8 J. Rouse, F. Mohammadi, C.R. Helms and K.C. Saraswat, Appl. Phys. Lett. (Accepted for Publication).

4.9 F. Mohammadi, T.W. Sigmon and K.C. Saraswat, Extended Abstracts of the 157th Meeting of the Electrochem. Soc., St. Louis, Vol 80-1, 422 (1980).

4.10 K.C. Saraswat, F. Mohammadi and J.D. Meindl, Technical Digest of the 1979 Int. Electron Device Meet., Washington D.C., pp. 462-464, 1979.

4.11 K.C. Saraswat and F. Mohammadi, IEEE Electron Device Letters, EDL-1, 18 (1980).

4.12 L.D. Locker and C.D. Capio, J. App. Phys., 44, 4366 (1973).

4.13 T. Mochizuki, K. Shibata, T. Inoue and K. Ohuchi, Japan J. Appl. Phys, 17, Suppl. 17-1, 37 (1977).

4.14 B.L. Crowder and S. Zirinsky, IEEE Trans. Electron Devices, ED-26, 369 (1979).

4.15 Dietrich E. R. Kher, Proc. 6th Intl. Conf. on CVD, Electrochem. Soc., 77-5, 511 (1977).

4.16 S. Peterson, J. Baglin, F. d'Heurle, J. Harper and C. Serrano, Extended Abstracts of the Fall 1979 Meeting, Electrochem. Soc., Los angeles, 79-2, (1979).

4.17 J.W. Shaw and J.A. Amick, RCA Review, 31, 306 (1970).

4.18 S. Zirinsky, W. Hammer, F. d'Heurle, and J. Baglin, Appl. Phys. Lett., 33, 76 (1978).

4.19 C.D. Wirkus and D.R. Wilder, J. Am. Ceram. Soc., 49, 173 (1966).

4.20 J.B. Berkowitz-Mattuck and R.K. Dils, J. Electrochem. Soc., 112, 583.

4.21 B.E. Deal and A.S. Grove, J. Appl. Phys., 36, 3770 (1965).

4.22 J.S. Johannessen, C.R. Helms, W.E. Spicer, and Y.E. Strausser, IEEE Trans. Electron Dev. ED-24, No. 5, May 1977, p. 547.

4.23 J.S. Johannessen, W.E. Spicer, and Y.E. Struasser, JAP 47, No. 7, July 1976, p. 3028.

4.24 The measured SiO-WSi_2 interface width for both structures is 725 Å. By profiling samples with varying oxide thicknesses and assuming that the actual interface widths were unchanged, we obtain a sputter-induced broadening of 0.3 D, where D is the oxide thickness. The actual interface width W_A and the measured interface width W_M (= 725 Å) are then related by

$$W_M = [W_A^2 + (0.3D)^2]^{1/2}$$

This gives an actual interface width of 407 Å for both structures.

4.25 S.M. Sze, Physics of Semiconductor Devices, Wiley-Interscience, 1969.

4.26 J.M. Andrews, J. Vac. Sci. Technol., Vol. 11, pp. 972-984, 1974.

4.27 P.J. Tsang, "Measuring the Work Functions of Metal Silicides by the C-V Method," Extended Abstracts of the Spring 1978 Meeting of the Electrochem. Soc., Seattle, Vol. 78-1, pp. 706-708, 1978.

4.28 M. Kuhn, Solid-State Electronics, Vol. 13, pp. 873-885, 1970.

Table 4.1

PROPERTIES OF THE INTERCONNECTION MATERIALS

MATERIALS	THIN FILM RESISTIVITY $\mu\ \Omega$ CM	EUTECTIC TEMP (°C)	CHEMICAL COMPATIBILITY	THERMAL OXIDATION
Poly-Si	500	1410	yes	yes
Al	2.5	577	no	no
W	5.5	1390	no	no
Mo	5	1410	no	no
WSi_2	70	1390	yes	yes
$MoSi_2$	100	1410	yes	yes
PtSi	30	830	yes	?
$TiSi_2$	25	1330	no	yes

Table 4.2

COMPARISON OF THE COMPOSITIONS OF THE SILICIDE LAYERS IN STRUCTURES A, B, AND THE WSi_2 STANDARD. The numbers indicate that Si depletion of the oxidized structure is more severe for an SiO_2 substrate than for a Si substrate.

	W	Si	% WSi_2	% W_5Si_3
Standard	0.33	0.67	100	0
A	0.35	0.65	97	3
B	0.39	0.61	94	6

Table 4.3

WORK FUNCTION ϕ_m AND BARRIER HEIGHT ϕ_B FOR SILICIDES

Silicide	ϕ_m eV	ϕ_B eV
PtSi	5.15	0.84
Pd_2Si	5.2	0.73
RhSi	4.96	0.68
WSi_2	4.8	0.67
$NiSi_2$	4.5	0.65

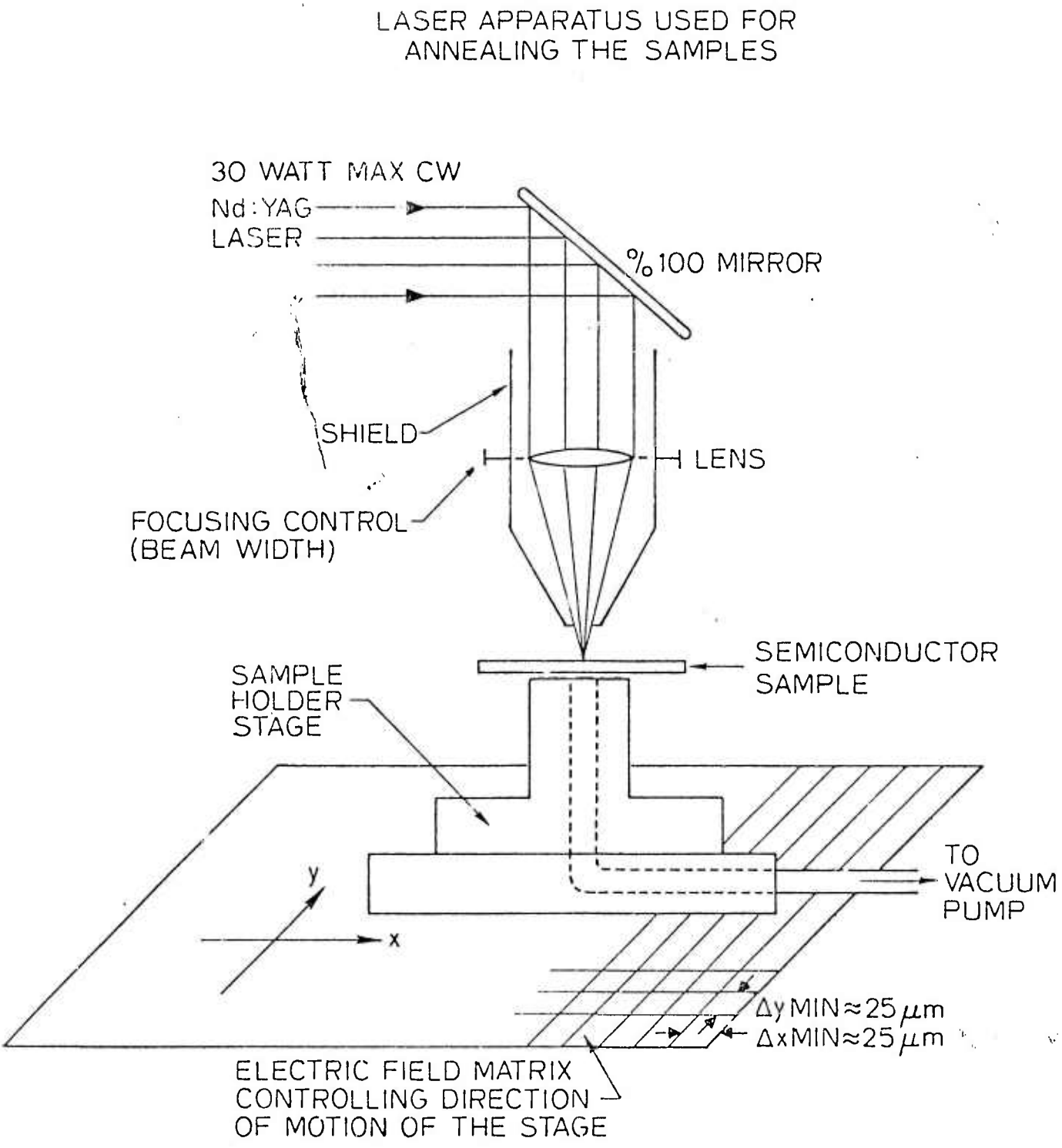


Fig. 4.1. Schematic of the equipment used for laser annealing.

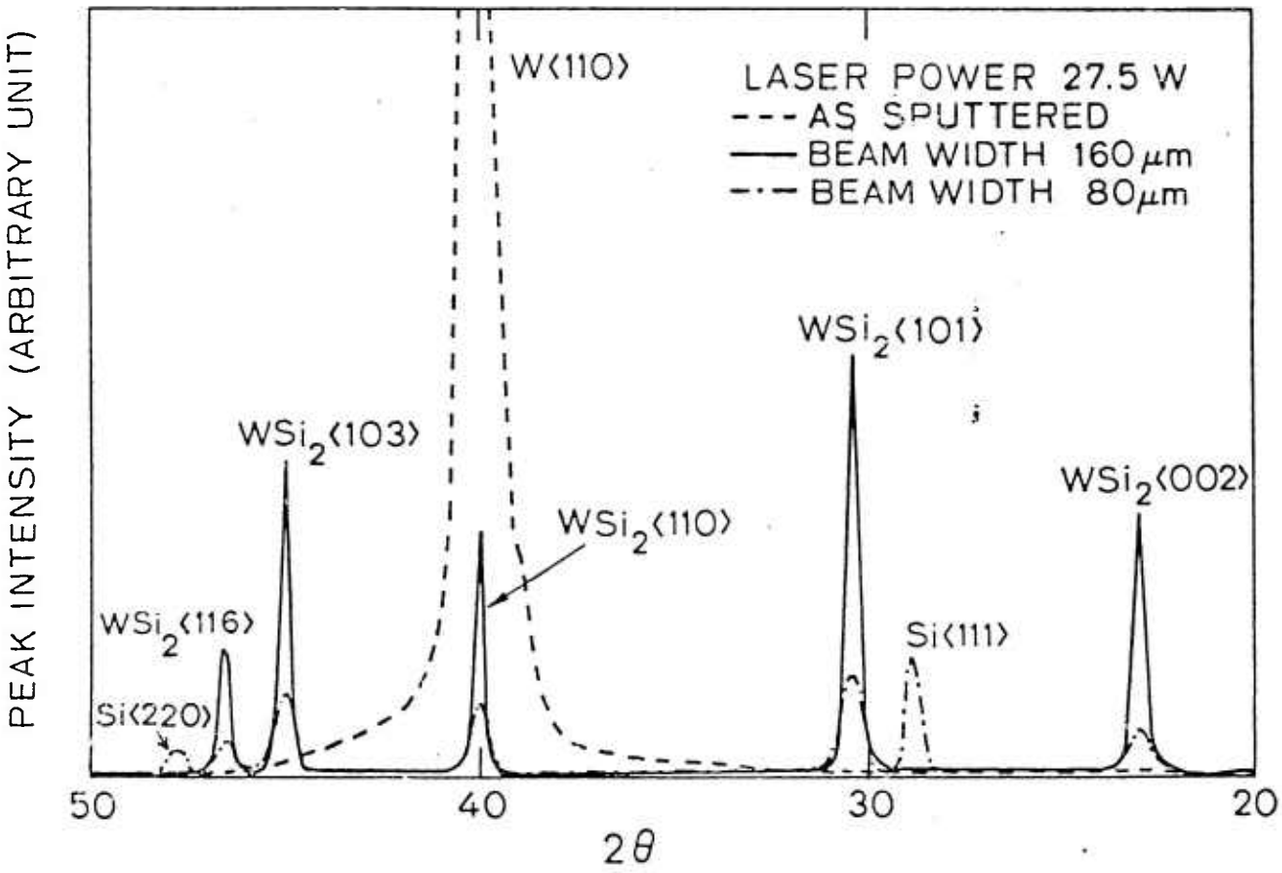


Fig. 4.2. X-ray diffraction spectra of as deposited and laser annealed films of 1000 Å thick tungsten

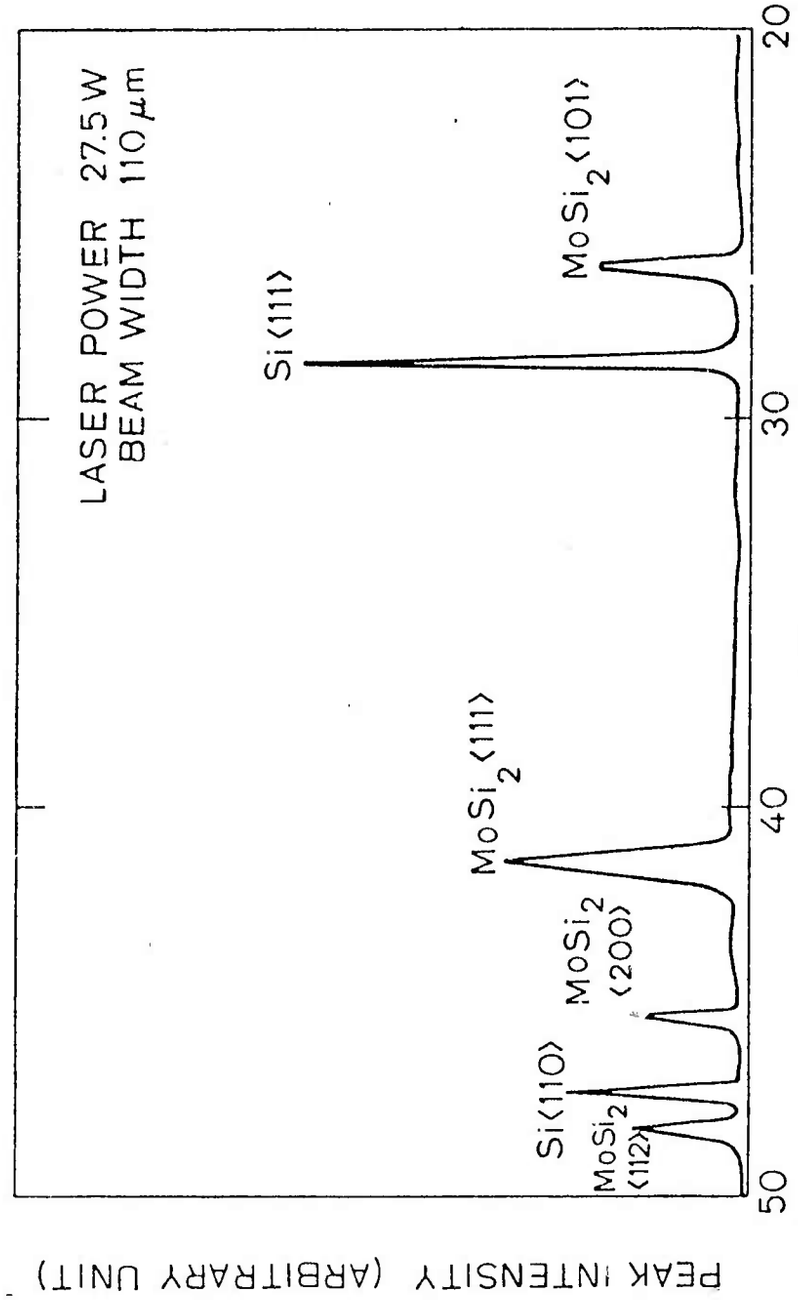


Fig. 4.3. X-ray diffraction spectra of laser annealed molybdenum film of 1000 Å thickness.

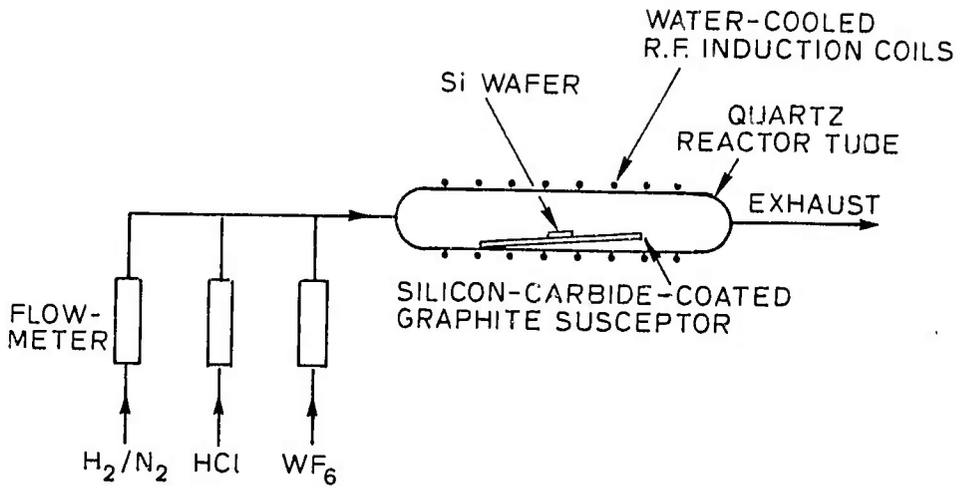


Fig. 4.4. Schematic diagram of the reactor used for the formation of WSi_2.

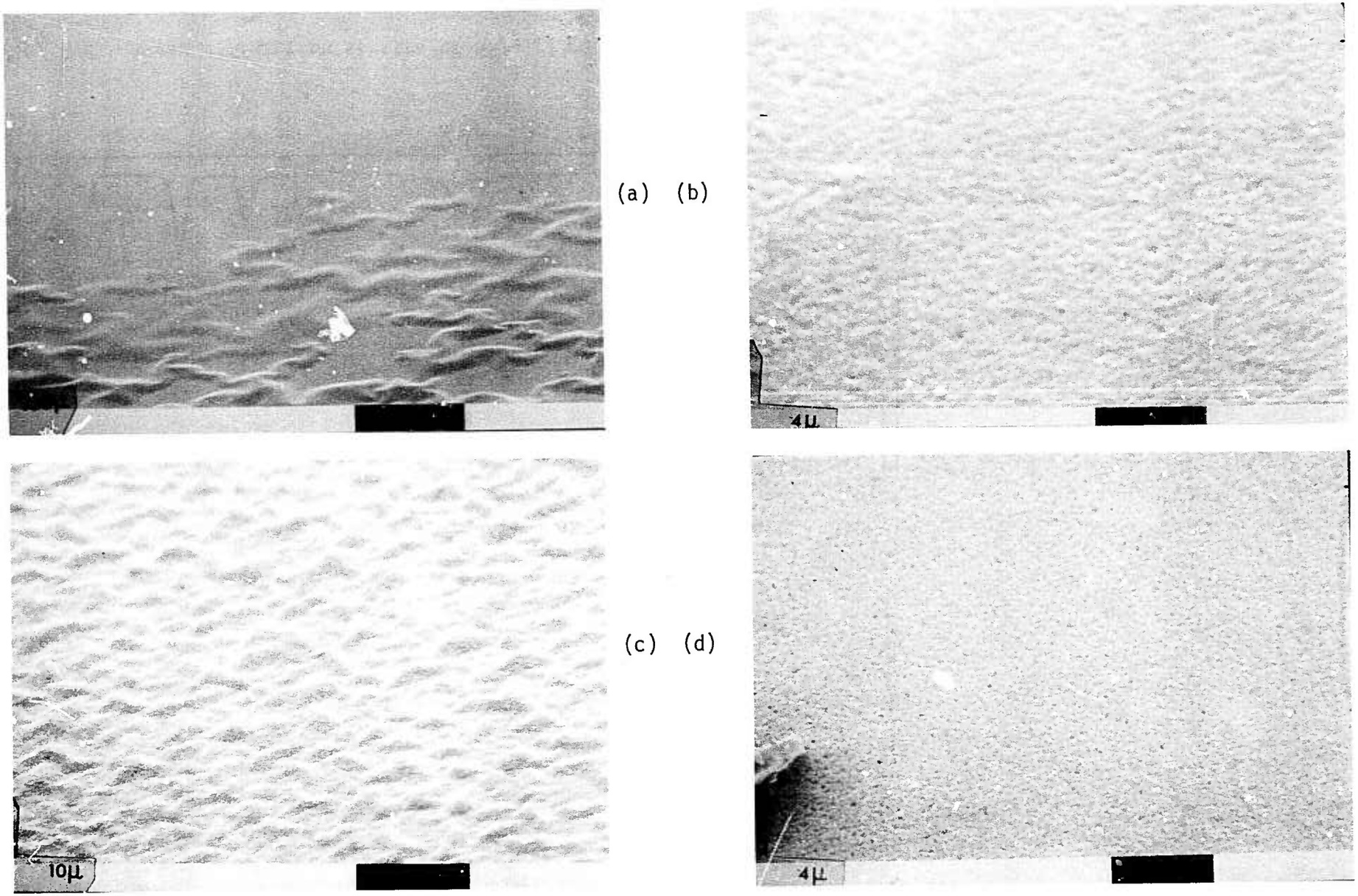


Fig. 4.5. WSi_2 films formed by CVD (a) 4000 Å thick film formed by W deposition in N_2 at 600°C (b) 1000 Å thick film formed by W deposition in N_2 at 600°C (c) 6000 Å thick film formed by W deposition in N_2 at 700°C (d) 2000 Å thick film formed by first HCl etching Si and then W deposition in H_2 at 600°C.

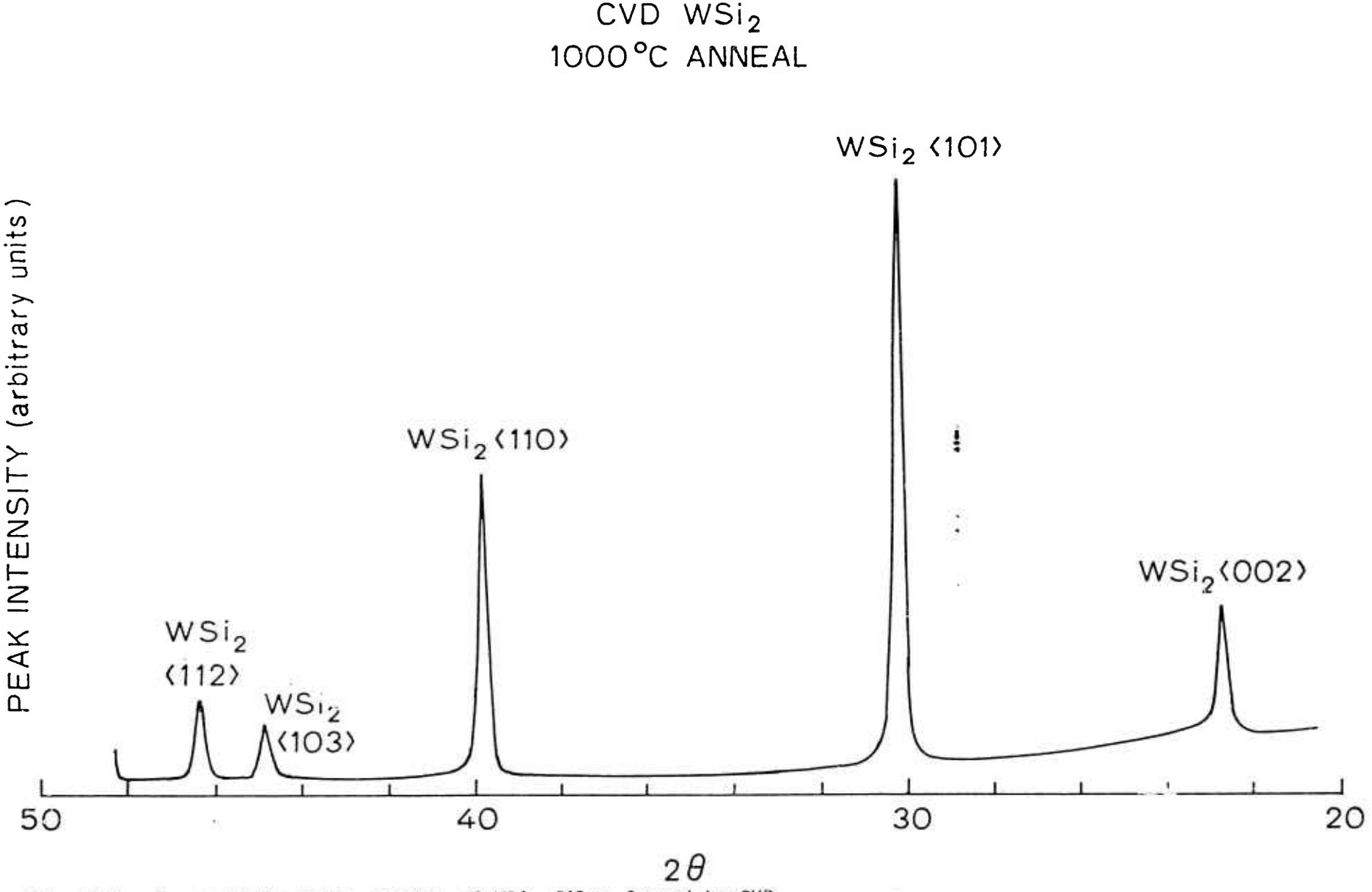


Fig. 4.6. X-ray diffraction spectra of WSi_2 films formed by CVD.

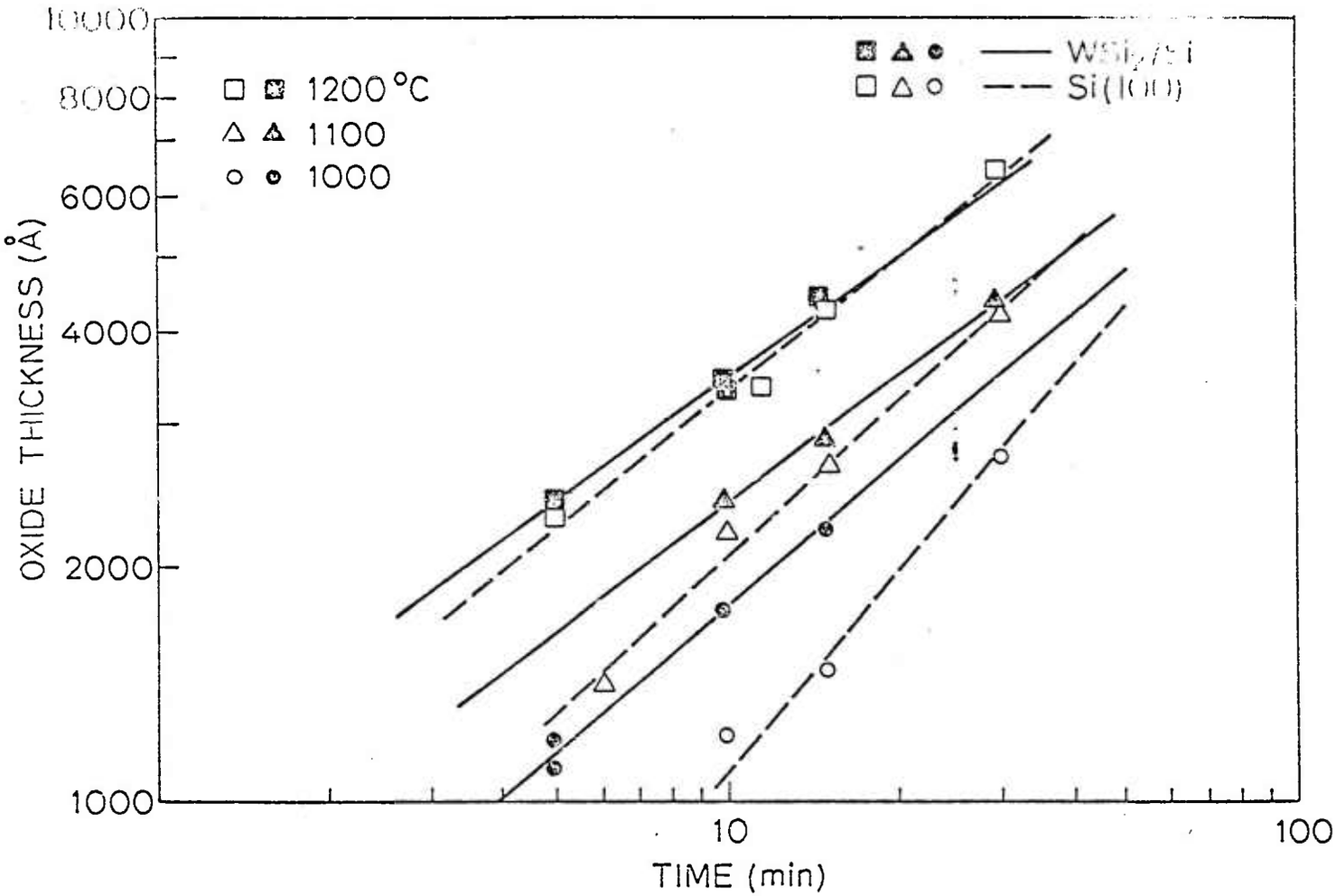


Fig. 4.7 Plot of oxide thickness vs oxidation time for WSi_2/Si films and Si (100) test samples.

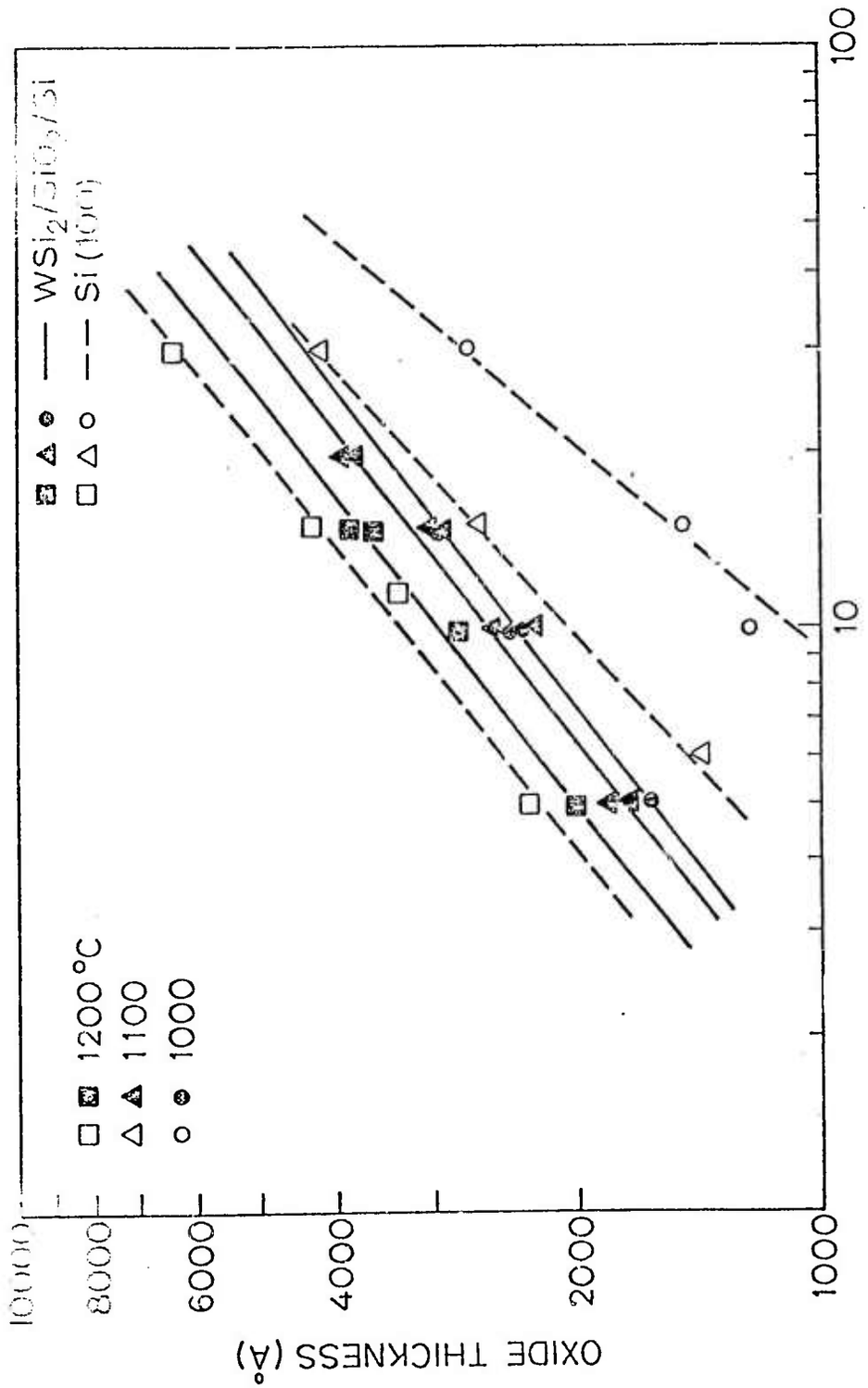


Fig. 4.8. Plot of oxide thickness vs oxidation time for $WSi_2/SiO_2/Si$ films and Si (100) test samples.

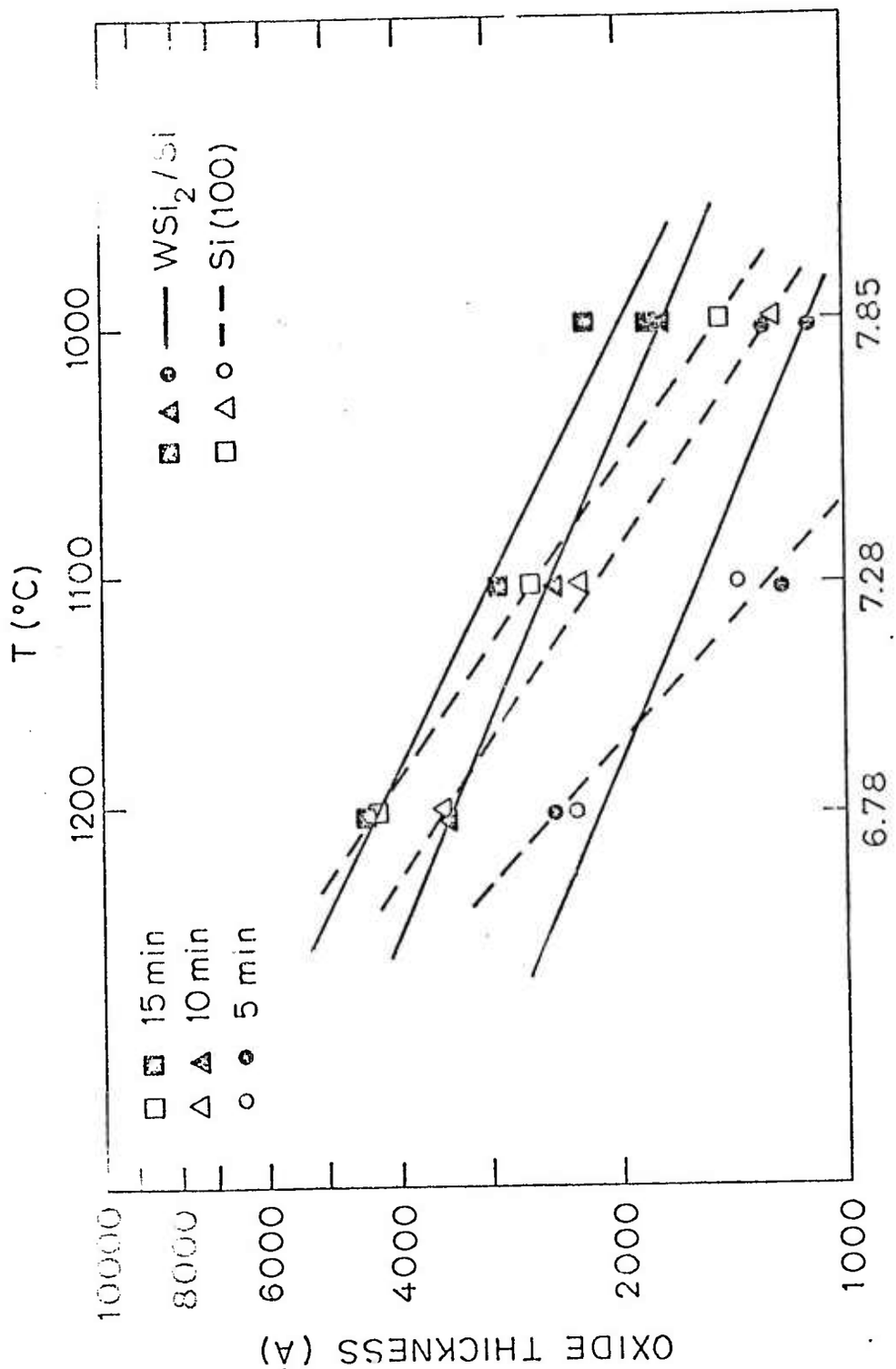


Fig. 4.9. Plot of oxide thickness vs 1/T for WSi_2/Si films and Si (100) test samples.

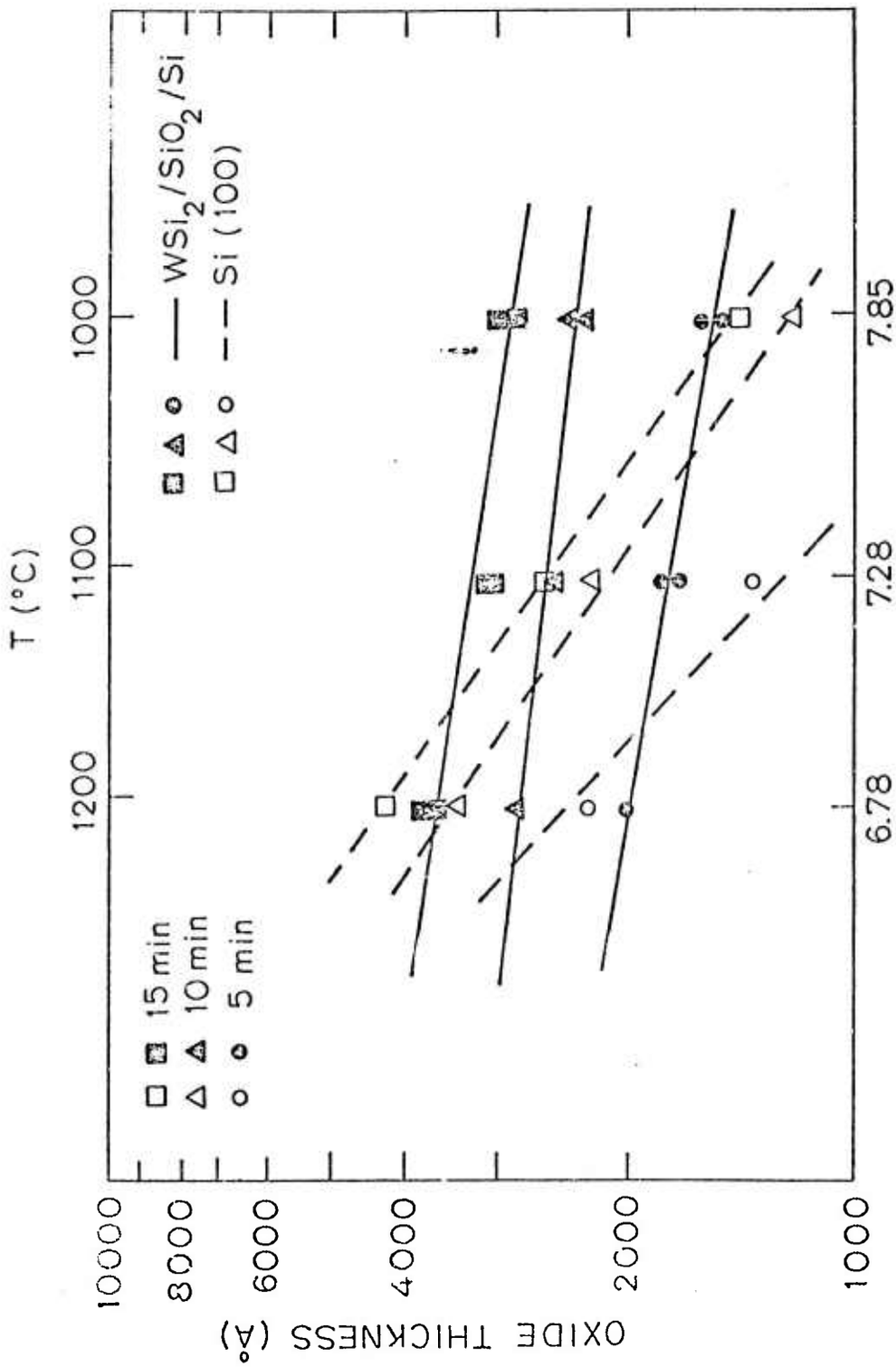
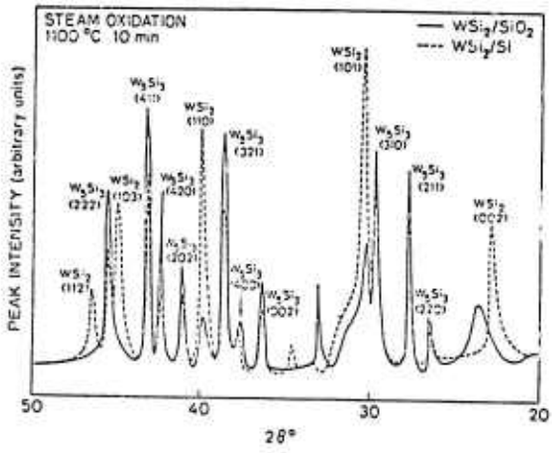
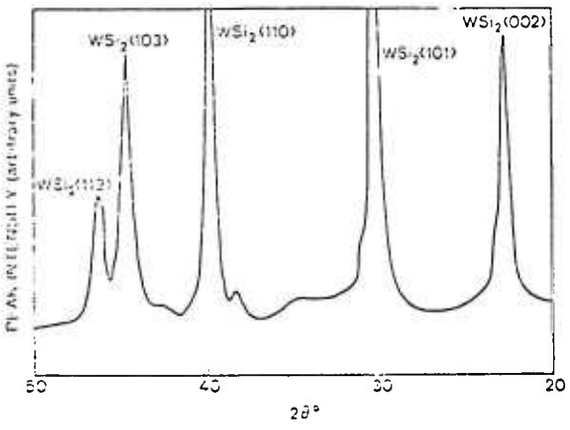


Fig. 4.10. Plot of oxide thickness vs 1/T for $WSi_2/SiO_2/Si$ films and Si (100) test samples.



(a)



(b)

Fig. 4.11. X-ray diffraction patterns of the (a) WSi_2/Si and $WSi_2/SiO_2/Si$ samples, oxidized at 1100°C for 10 min. (b) WSi_2 film annealed at 1000°C for 120 min. Cu-$K\alpha_1$ diffraction intensity vs 2θ.

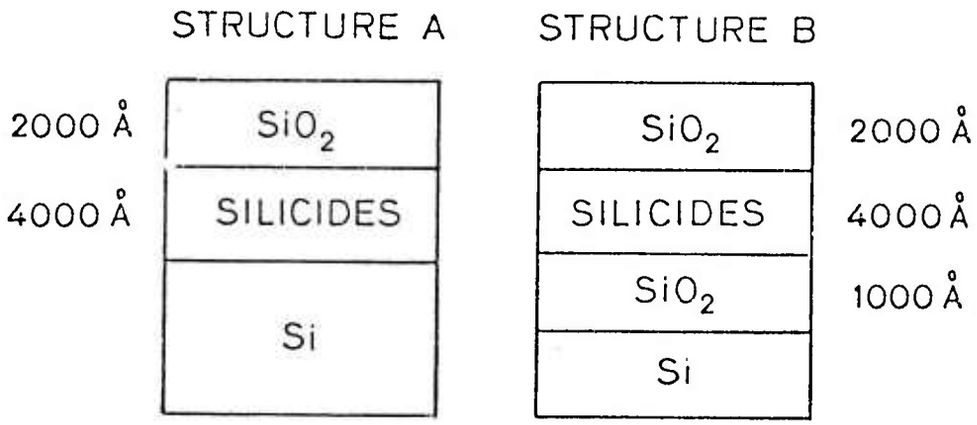


Fig. 4.12. (a) Structure A; WSi_2 deposited on Si and steam oxidized. (b) Structure B; WSi_2 deposited on SiO_2 and beam oxidized. Approximate layer thicknesses are shown.

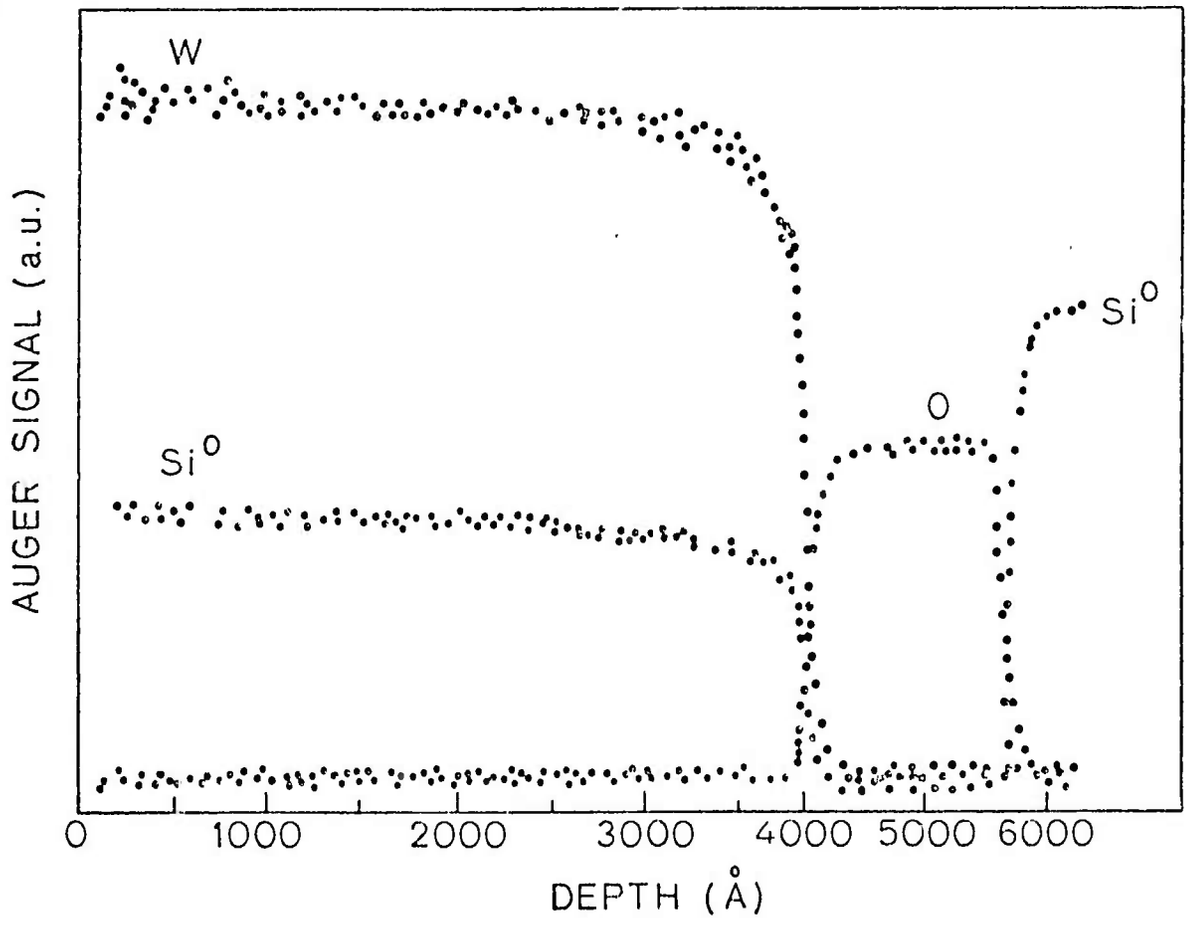


Fig. 4.13. Auger sputter profile of WSi_2 as deposited on SiO_2 substrate. Plotted is Auger signal height vs distance sputtered. This curve was used to calibrate Auger signal heights for W and Si.

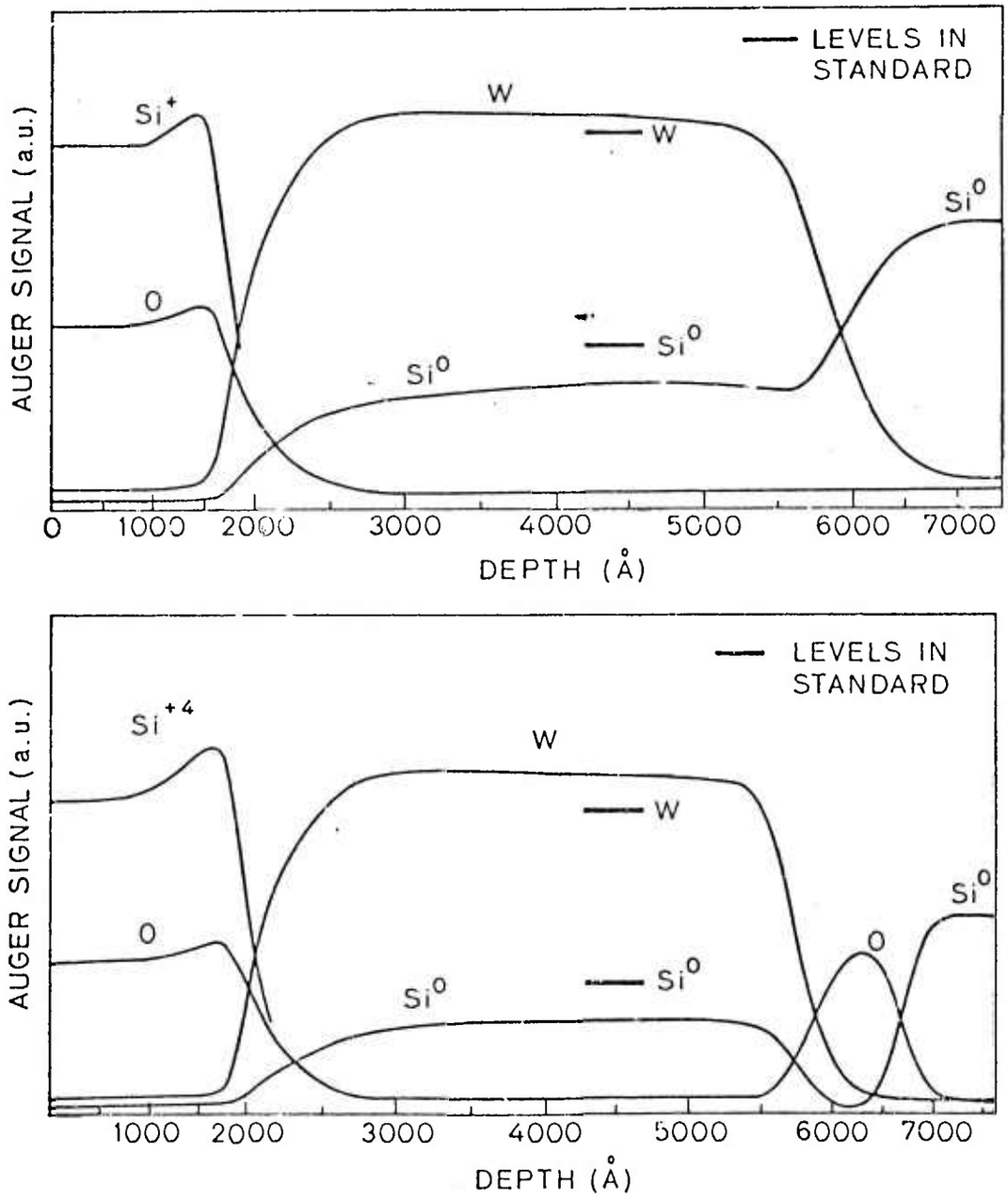


Fig. 4.14. (a) Auger sputter profile of structure A, with Auger signal height vs distance sputtered. Distance scale changes are due to differences in sputtering rate. Included for comparison are the W and Si levels from the WSi_2 standard (see Fig. 2). (b) Auger sputter profile of structure B. Again, W and Si levels from standard are included.

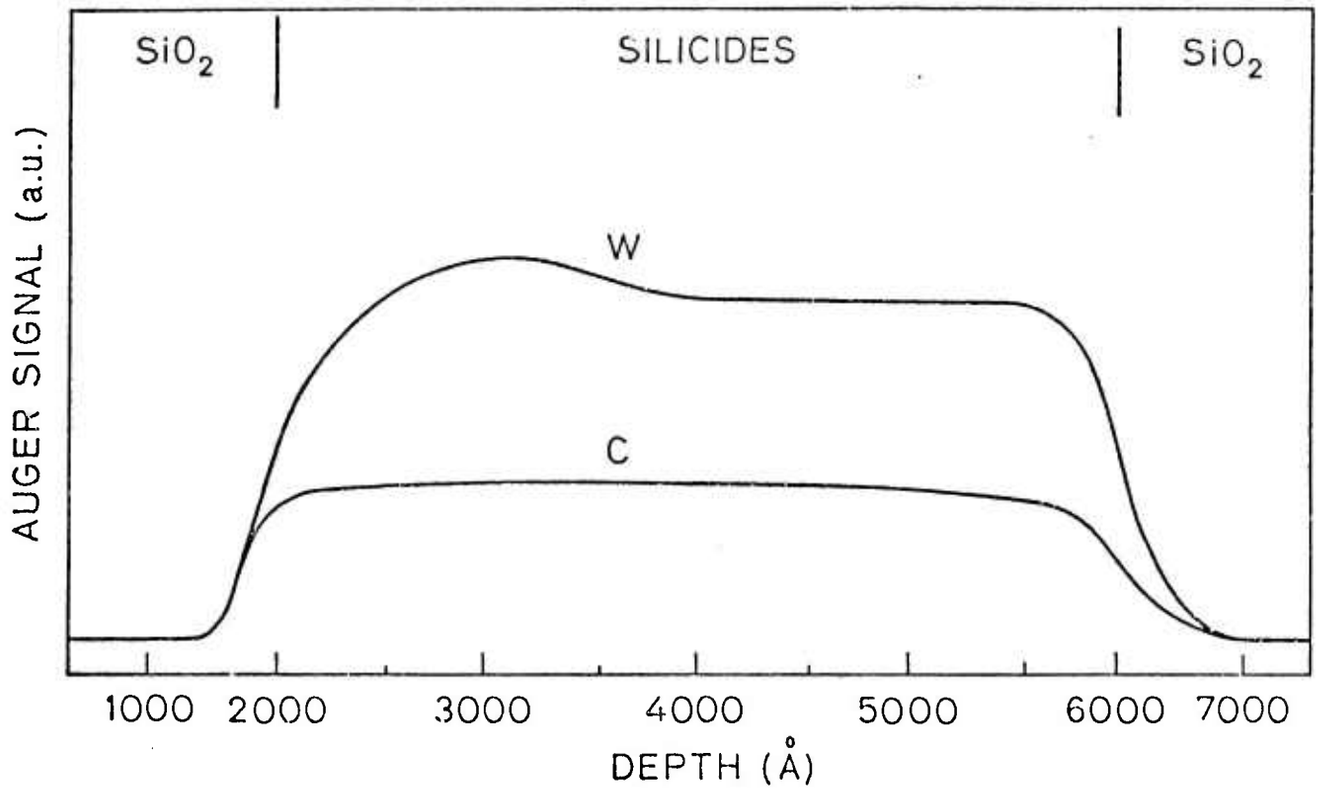


Fig. 4.15. Auger sputter profile of structure B, with Auger signal height vs distance sputtered. Shown are the W and C levels throughout the structure. Note that carbon appears only in the silicide region. The peak carbon concentration is approximately $2 \cdot 10^{21}$ cm^{-3}.

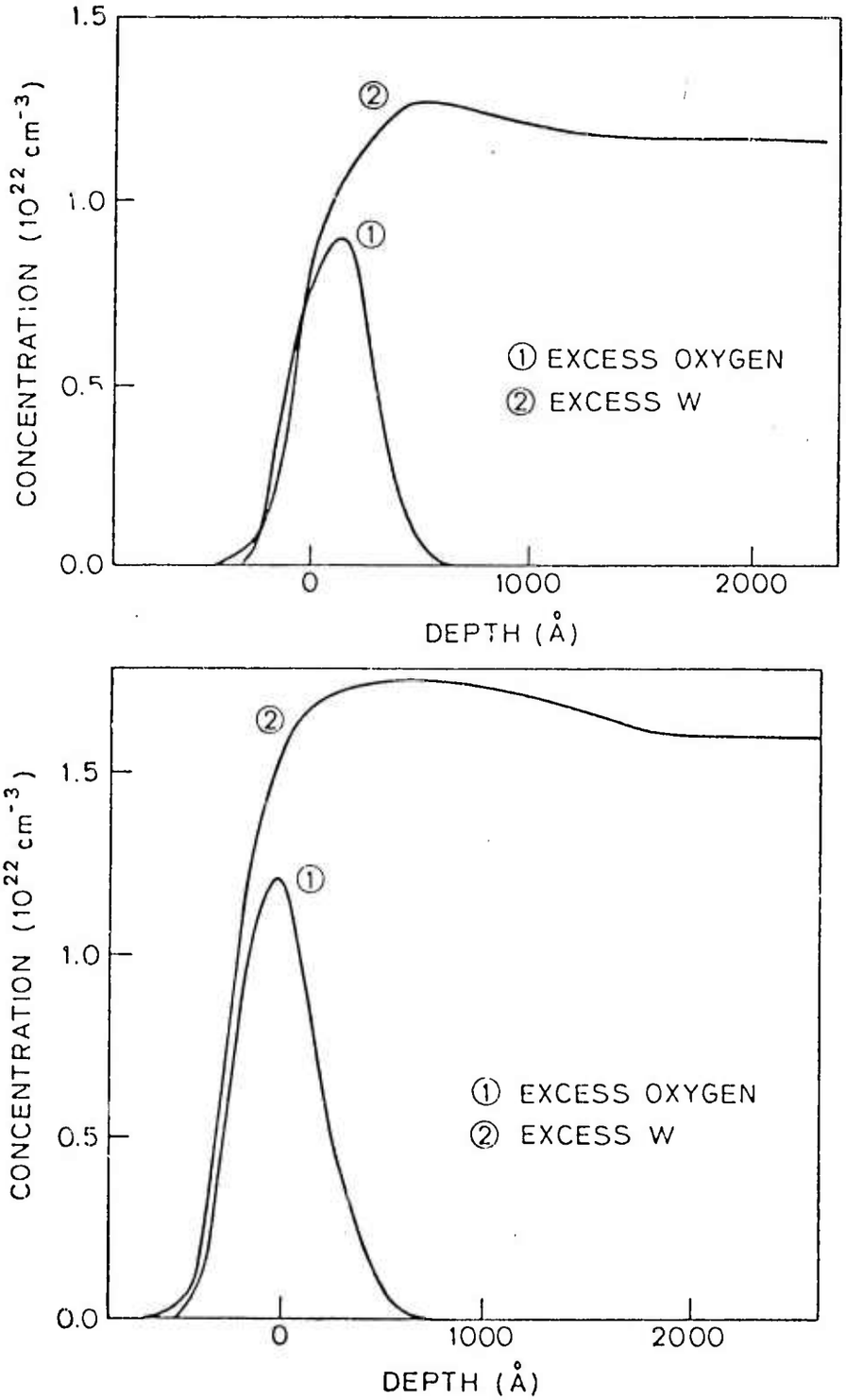


Fig. 4.16. (a) Calculated distribution of "excess" W and O in the SiO_2-WSi_2 interface region of structure A. The point x = 0 occurs where the SiO_2 signal is down 50%. (b) Calculated distribution of "excess" W and O in the SiO_2-WSi_2 interface region of structure B.

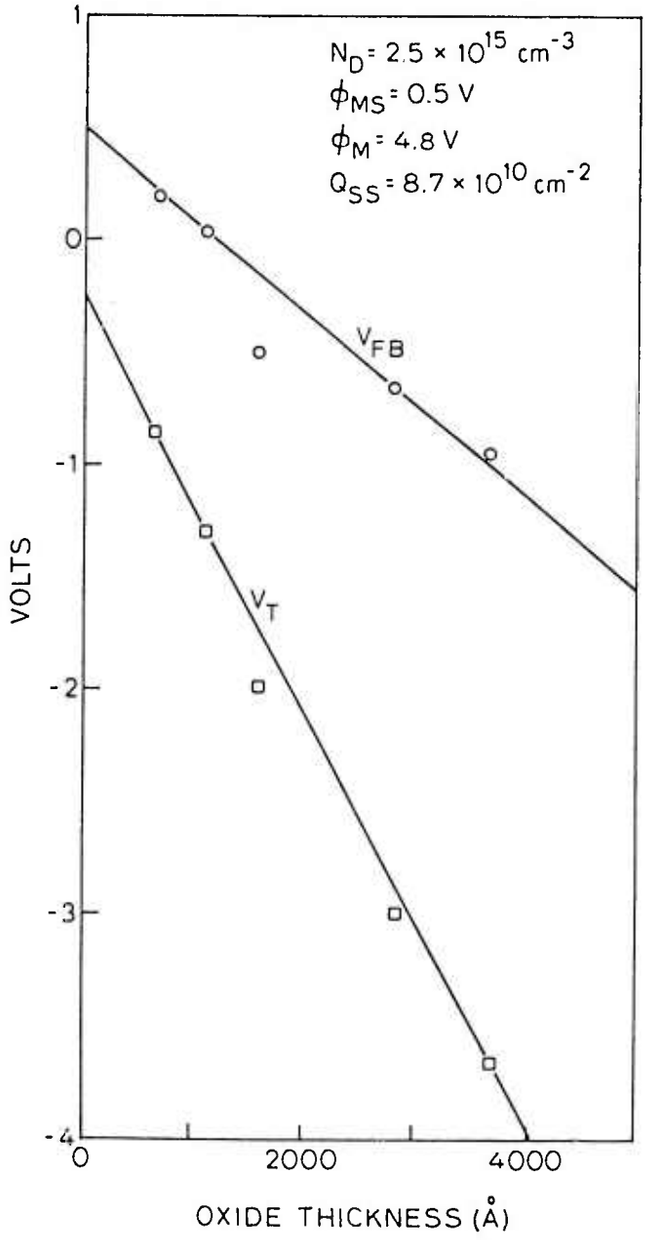


Fig. 4.17. V_{FB} and V_T as a function of gate oxide thickness.

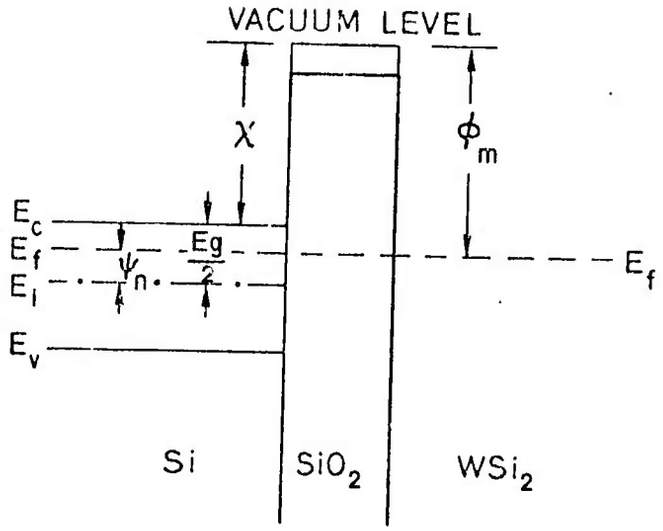


Fig. 4.18. Band diagram of WSi_2 gate MOS capacitor.

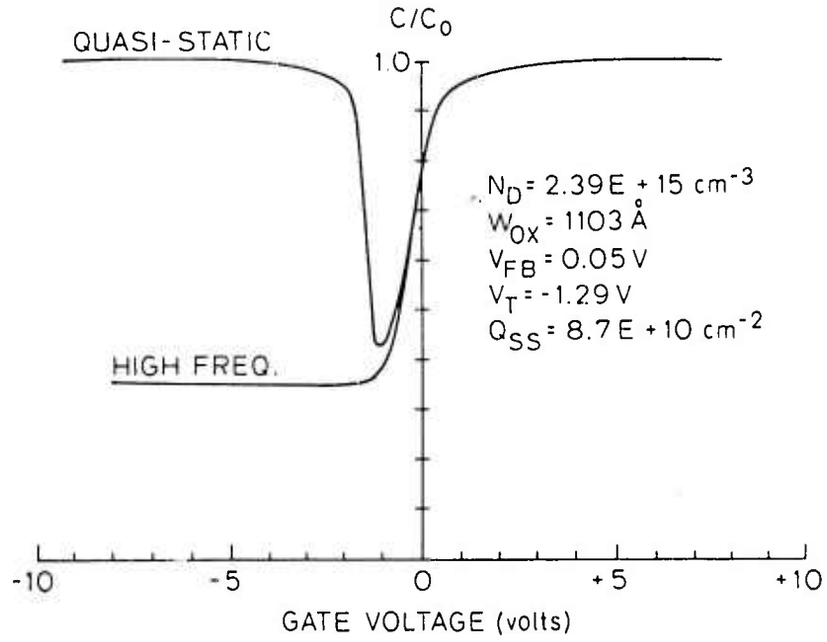


Fig. 4.19. Quasi-static and high frequency C-V measurements.

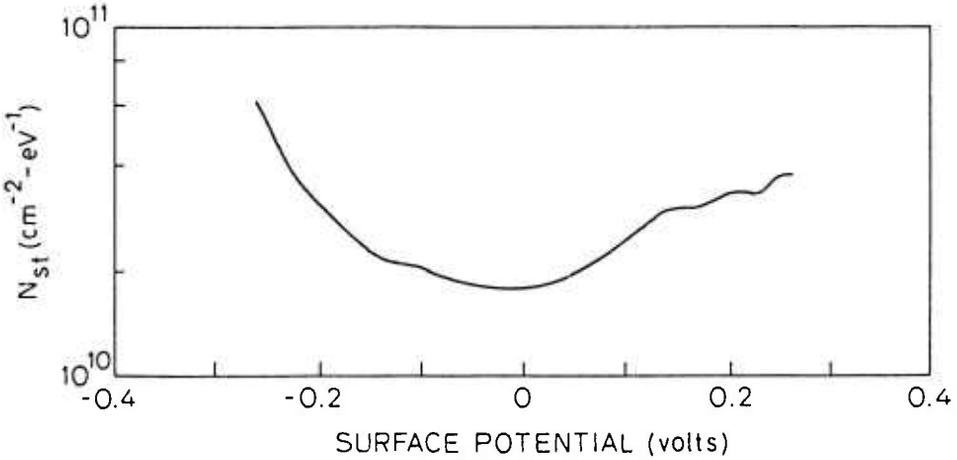


Fig. 4.20. Surface state density in the band gap.

5. MATERIALS ANALYSIS/INTERFACE PHYSICS

C. R. Helms, J. Rouse, R. W. Barton, S. A. Schwarz

5.1 Introduction

The past year has seen support for the Materials Analysis and Interface Physics Program described here transferred from the NBS/ARPA Semiconductor Technology Program to direct ARPA support through this contract. This transition has been smooth since we realized some time ago the value of close collaboration between this group and other groups in the program, especially thermal oxidation and chemical vapor deposition.

Within this framework, there are two major long-term objectives of this part of the program. First is the development and refinement of new measurement technology for the characterization of VLSI device structures. Second is the use of these measurement capabilities to obtain critical information, especially concerning interface structures on important materials problems in conjunction with other groups in this program.

The importance of the interfaces present in submicron device structures cannot be stressed too greatly. Even for larger device geometries, the performance of silicon integrated circuits is determined in large part by our ability to control and reliably reproduce various key interfaces present. Of prime importance are the interfaces between Si and SiO_2, both for the gate and field oxides and the interface between the electrodes and SiO_2 or Si.

These interfaces place even more severe constraints on our ability to build the small geometry devices contemplated for VLSI applications.

One major reason for this is the concept of scaling of larger geometry IC designs to smaller geometries. As an example, if we take an MOS structure, we desire to shrink a lateral dimension L_o to L_o/α where $\alpha > 1$. For the device performance to scale, channel doping levels would be increased by α and gate oxide thicknesses decreased by α. Neglecting edge effects, the capacitance of an MOS capacitor would be increased by α.

In scaling analysis, just described, the scaling of interface dimensions has not been taken into consideration. In fact, to first order, interface dimensions do not appear to scale at all; the widths of typical interfaces appear approximately constant. The effect of this on an MOS capacitor is shown in Fig. 5.1. The capacitance C of this structure is given by

$$C = \varepsilon_o A \left\{ \frac{\ell_1}{\varepsilon_1} + \frac{\ell_2}{\varepsilon_2} + \frac{\ell_3}{\varepsilon_3} \right\}^{-1} \tag{5.1}$$

or

$$\frac{C}{C_o} = \frac{1}{\alpha} \left\{ 1 + \frac{\varepsilon_2}{\ell_2} \left(\frac{\ell_1}{\varepsilon_1} + \frac{\ell_3}{\varepsilon_3} \right) \right\}^{-1} \tag{5.2}$$

where C_o is the capacitance for a large device so that the capacitance no longer scales as $1/\alpha$.

This brings up the question of tolerances on device processing to attain a given level of confidence in device electrical performance. For the above example for thin oxides, a small error in oxide growth time could lead to a much larger relative error in capacitance depending, of course, on the relative values of ε_1, ε_2, and ε_3 and ℓ_1, ℓ_2, and ℓ_3. A knowledge of the values of all these parameters is critical in device design and reliability.

Consistent with our long-term objectives, there are two major products of this program. First is advanced measurement technology which is directly disseminated to the industrial/government user. Second, the results of the interface studies performed lead to interface models for direct inclusion in SUPREM, which is normally accomplished in conjunction with other groups in the program. This synergestic approach is illustrated schematically in Fig. 5.2 and forms the framework for this part of the program.

5.2 Principal Accomplishments

5.2.1 Introduction

At the beginning of this program, a number of chemical microanalytic techniques existed for the determination of elemental depth profiles. These techniques, including Auger electron spectroscopy (AES) used with sputter profiling, photoelectron spectroscopy (PES), and secondary ion mass spectroscopy (SIMS), were, however, not sufficiently well understood to allow accurate depth profiling with subnanometer depth resolution.

Much of our activity has centered on determining what factors affect sensitivity, depth resolution, quantitative analysis, etc., so that the proper technique can be applied to the proper problem and each technique can be implemented to its full potential. Effects that have been investigated in detail include electron mean free path, electron stimulated desorption, ion knock-on mixing, sputtering-induced roughness, and problems associated with nonuniform ion beams. We have concentrated on AES used with sputter profiling. However, the results obtained concerning electron mean free path effects pertain equally well to PES and

the results obtained concerning ion beam effects pertain to SIMS or other techniques which employ ion beams for profiling.

In addition to studies of the capabilities and limitations of the above-mentioned microanalytical techniques, we have used these capabilities for critical measurements of interface properties in conjunction with other groups in the program.

A summary of our major accomplishments in this part of the program are listed in the next section. Following that, is a detailed description of these activities. Both of these sections will be divided into a materials analysis subsection and an interface physics subsection.

5.2.2 Summary

Materials Analysis

(1) Demonstrated that sensitivities down to 10^{18} cm^{-3} can be obtained routinely with commercially available AES equipment for elements such as B, C, N, O, F, Na, Al, Si, P, Cl, As, and Sb encountered in IC technology.

(2) Determined the limits on electron beam size and current for Auger analysis at which damage to Si or SiO_2 occurs; obtained maximum sensitivity versus beam size consistent with this.

(3) Developed models for sputtering and ion knock-on effects that provide the basis for deconvolving ion knock-on broadening from Auger sputter profiles and SIMS profiles.

Interface Physics (in collaboration with Oxidation group)

(1) Completed experimental study of the effect of oxidation time on the redistribution of phosphorus during the thermal oxidation of silicon. Showed that the chemical potential for phosphorus has a sharp minimum within 50 Å of the Si/SiO_2 interface.

(2) Completed simple theoretical model of dopant effect in (1), including interface field.

(3) Completed preliminary study of chlorine distribution in thermal SiO_2. Showed that Cl piles up within 50 Å of the Si/SiO_2 interface on the SiO_2 side. The amount of pileup monotonically increases with oxide thickness (or oxidation time); other results indicate that this increased Cl present at the interface is accompanied by a decrease in flat band voltage of so prepared MOS capacitors.

5.3 Detailed Discussion of Results

5.3.1 Sensitivity and Spatial Resolution

Questions of sensitivity and spatial resolution in AES must be dealt with together since, for a constant brightness source, an increase in spatial resolution can only be achieved through a reduction in beam current and therefore a loss in sensitivity for a given data collection time. By spatial resolution, we mean the dimensions parallel to the surface of the sample; the resolution perpendicular to the specimen's surface is determined by the escape depth which has been discussed in

previous reports. In order to ensure that the analysis of this section has the most general applicability, figures for source brightness and signal-to-noise ratio will be taken from currently available commercial equipment. In addition to the interdependence of these two factors, electron beam-induced damage will also affect sensitivity and spatial resolution limits. One such effect is electron stimulated desorption of oxygen from SiO_2.

In order to discuss sensitivity, we will consider a baseline case with the following parameters: Lock-in amplifier detection system providing an output of $d[E \cdot N(E)]/dE$, where E is energy and $N(E)$ is the distribution of electrons, 100 μA beam current, 2.4% analyzer resolution, 30 min data collection time, 10 V peak-to-peak modulation voltage, 50 V scan, and signal averaging using multichannel analyzer or minicomputer.

In general, the signal-to-noise ratio is given by

$$\frac{S}{N} = A \sqrt{It} \left(\frac{\Delta E}{E}\right)^{1/2} V_{ptp} \tag{5.3}$$

where A is a constant, I is the beam current, $(\Delta E/E)$ is the resolution of the analyzer, t is time, and V_{ptp} is the modulation voltage. This assumes that the value of $\Delta E/E$ chosen, as well as V_{ptp}, do not contribute any broadening to the line shape which limits both V_{ptp} and ΔE to 5 eV for most transitions. Also, the $\sqrt{t}$ dependence can only be realized for random noise and can be significantly affected by any coherent noise present. Within these assumptions, we will define the maximum AES sensitivity as the concentration of the element with the strongest Auger transition for which $S/N = 1$.

Remarkable as it may seem, no measurements of this maximum sensitivity are available. From extrapolating data taken at much less ideal conditions, a value for the maximum sensitivity of AES of ~10^{17} cm^{-3} can be estimated. This value of maximum sensitivity, however, does not correspond to all elements but to the maximum-strength AES transitions such as the Ag MNN transition or the Cl LMM transition. Maximum sensitivities for other elements of interest are listed in Table 5.1 within ± half order of magnetitude. These values are based on the assumption that there are no interference effects from neighboring transitions.

To obtain these sensitivities within a 30-min data collection time, a 100 μA beam current would be required. Presently, typical guns with LaB_6 emitters can provide this beam current with a minimum spot size about 10 μm in diameter. This corresponds to a current density of 100 A/cm^2, a power density of 5×10^5 W/cm^2, and a total power dissipation of ~1/2 W. These levels are sufficient to cause local melting of silicon under typical conditions. To bring beam effects to acceptable levels under these conditions, rastering the electron beam over a 100 × 100 to 200 × 200 μm area is required. Typical values for beam diameter versus beam current are shown in Fig. 5.3 as the solid curve; the dashed curve is the maximum current versus beam size to ensure that no significant damage is occurring in Si.

It should be stressed at this point that much instrumentation exists in the field today that does not possess the capabilities mentioned above. For example, instruments with beam currents of 10 μA, 0.3% analyzer resolutions, and without minicomputer facilities, are common. In this case, a factor of 10 degradation in sensitivity is probable. In Fig. 5.4, the variation of maximum sensitivity of AES as a function of

beam diameter is presented. The solid curve is for Si; the two dashed curves represent the limit in SiO_2 for beam energies of 4.5 keV with the indicated amounts of elemental silicon present in the spectrum due to electron stimulated desorption. Again, it should be stressed that these values are for maximum AES sensitivity for 30-min data collection times. For example, for As in Si and a 1 μm beam diameter, the minimum detectability limit (from Table 5.1) is approximately 5×10^{20} cm^{-3}.

5.3.2 An Ion Knock-On Mixing Model

Depth profiles obtained from ASP or SIMS experiments are broadened by ion beam-induced atomic mixing [5.1,5.2,5.3]. This effect is illustrated in Fig. 5.5. In this section, we develop a model for ion knock-on mixing [5.4] based on analogy-to-thermal diffusion theory. The model leads to a simple relationship which predicts the broadening observed in a sputter profiling experiment. We show that the broadening of the Si/SiO_2 interface for low energy Ne^+, Ar^+, and Xe^+ bombardment is consistent with the mixing model. We then extend the model to higher energies and compare it to Monte Carlo calculations [5.1] and experimental data.

The ion knock-on mixing model [5.4,5.5] developed here is based on an analogy to thermal diffusion theory. A similar approach has been taken by Haff and Switkowski [5.3]. In their paper, an effective diffusion coefficient D is derived in terms of various parameters including a cascade radius and the stopping power of the ion. A diffusion time t must be inferred in order to determine the broadening observed in a sputter profiling experiment. In our model, an effective broadening parameter $W^2 = Dt$ is derived and expressed in terms of the sputtering yield S, a measurable quantity. The stopping power effect in Haff's model is inherent in the sputtering yield.

Our derivation is based on the assumption that an atom in the bombarded solid undergoes a number of random collisions N and is detected immediately prior to or after its escape from the surface. In an ASP experiment, this assumption is valid if the escape depth L is small (~5 A) or is much less than the ion broadening parameter W. In a SIMS experiment, atoms are detected after their escape. We describe the path of the atom by a random walk as illustrated in Fig. 5.6. Monte Carlo calculations [5.1] confirm that the assumption of random collisions at low and moderate ion energies is reasonable. The net distance R travelled by the atom is given by

$$R^2 = N\overline{a^2} \tag{5.4}$$

where $\overline{a^2}$ is the mean square step length.

In order to determine N, we make use of the Kinchin-Pease formula which states that the number of atomic displacements occurring in the solid per incident ion is $E/2E_d$. The number of collisions which occur is twice the number of displacements since two atoms collide with each other in a displacement event. Suppose now that the specimen of interest has been sputtered for a time sufficient to reach equilibrium. N, the number of collisions suffered by an escaping atom, is simply the total number of collisions which have occurred, divided by the total number of escaping atoms. If we normalize to the number of incident ions, we obtain

$$N = \frac{\text{\# collisions/ion}}{\text{\# escaping atoms/ion}} = \frac{E/E_d}{S} \tag{5.5}$$

S, of course, is the sputtering yield.

In diffusion theory, a broadening parameter Dt is defined as $Dt = 1/6\ N\overline{a^2}$. We now define an analogous broadening parameter

$$W^2 = \frac{1}{6} N\overline{a^2} = \frac{\overline{Ea^2}}{6E_d S} \tag{5.6}$$

For low ion energies ($E_0 < 1$ keV), most collisions occur at energies of several E_d. a should then be roughly the nearest-neighbor distance d in the solid (0.3 nm). At higher energies, $\overline{a^2}$ is given by

$$\overline{a^2} = \frac{\int \ell^2(\varepsilon)\ N(\varepsilon)\ d}{N(\varepsilon)\ d} \tag{5.7}$$

where $N(\varepsilon)\ d\varepsilon$ is the number of collisions occurring in the interval $d\varepsilon$ about ε and ℓ is the mean free path. We will discuss higher energy considerations later in this section.

W can be determined experimentally by observing the broadening of a step-function profile, i.e., an abrupt change in concentration. The step function broadens into an erf profile with a 10% to 90% width of 3.6 W, as illustrated in Fig. 5.7.

From diffusion theory, we also know that, if the original profile is a Gaussian or an erf described by Dt_0, the diffused profile is also a Gaussian or erf described by $Dt_M = Dt_0 + Dt$. Similarly, in a sputter profiling experiment, if the original profile is described by the parameter W_0, the measured profile will be described by $W_M = \sqrt{W_0^2 + W^2}$. Thus, the profile broadens in rms fashion.

We have performed ASP experiments [5.4] to examine the broadening of the Si/SiO_2 interface as a function of ion energy for Ne^+ and Xe^+ bombardment. The experimental setup is discussed elsewhere [5.6]. The

measured sputtering yields of SiO_2 are shown in Fig. 5.8. The solid curves are obtained from our sputtering model with SiO_2 treated as a lattice of Ne atoms and with E_d = 25 eV. Neon, with an atomic mass of 20, has a suitable average mass for SiO_2. The sputtering curve for Ne^+ bombardment fits the data quite well. This may be due to the fact that the incorporation of Ne atoms into the SiO_2 lattice does not affect its average mass. The measured yields for Ar^+ and Xe^+ are larger than predicted by the model at low energies. This may result from a larger incorporation of noble gas atoms at low sputtering yields which makes the lattice effectively heavier. The range of the incoming ions is thereby reduced, more energy is released near the specimen surface, and the sputtering yield increases.

The measured interface widths are plotted in Fig. 5.9. The solid curves are calculated from the mixing model with an assumed original 10% to 90% interface width of $3.6\ W_0$ = 2.0 nm and with $(\overline{a^2})^{1/2} = a_{rms}$ = 0.33 nm. The divergence of the solid curves from the data at high energies is due to the increase of a_{rms} with energy. The dashed curves are a higher energy approximation which we discuss in the next section. If the value of $3.6\ W_0$ is increased beyond 2.0 nm, the curves in Fig. 5.9 will move closer together. Thus, at a particular ion energy, the different broadening parameters obtained for Ne^+, Ar^+, and Xe^+ provide good evidence for the actual interface width. Figure 5.9 illustrates that the mixing model is entirely consistent with the measured interface widths.

At energies greater than 1 keV, Eq. (5.7) must be used to evaluate $\overline{a^2}$. In that equation, $a(\varepsilon)$ is the mean free path between collisions, which is discussed in detail elsewhere [5.6]. We determined that, as a first approximation,

$$\ell(\varepsilon) \cong \gamma k \varepsilon = K\varepsilon \tag{5.8}$$

where γ and k are as defined elsewhere [5.6]. Note that γ and k refer to the atom of interest, a dopant atom, for example, striking a lattice atom and do not depend on the incident ion.

$N(\varepsilon)\, d\varepsilon$, the number of collisions occurring between ε and $\varepsilon + d\varepsilon$, may be roughly approximated as follows: Suppose that an ion of energy E_0 collides with a lattice atom producing two atoms of energy $E_0/2$ which then collide to produce four atoms of energy $E_0/4$ and so on. For such a process, $N(\varepsilon) \propto 1/\varepsilon^2$. Eventually, there will be $E_0/2E_d$ atoms with average energy $2E_d$ (recall the Kinchin-Pease formula), at which point no further displacements occur. In Eq. (5.7), we will use $2E_d$ as the lower limit of integration. We will also use $\gamma_i E_0$ as the upper limit since $\gamma_i E_0$ is the maximum energy that can be transferred to a lattice atom by the incident ion. We now choose

$$N(\varepsilon) = 2E_0/\varepsilon^2 \tag{5.9}$$

so that

$$N(\varepsilon)\, d\varepsilon \cong \int_{2E_d}^{\gamma_i E_0} \frac{2E_0}{\varepsilon^2}\, d\varepsilon \cong \frac{E_0}{E_d} \tag{5.10}$$

Recall that E_0/E_d is the number of collisions that occurred in the solid. Equation (5.7) may now be written as follows:

$$\overline{a^2} = \frac{E_d}{E_0} \int_{2E_d}^{\gamma_i E_0} (K\varepsilon)^2 \frac{2E_0}{\varepsilon^2}\, d\varepsilon \cong 2\gamma_i K^2 E_d E_0 \tag{5.11}$$

Substituting this value of $\overline{a^2}$ into Eq. (5.6), we obtain the following appealing result:

$$W^2 = \frac{\gamma_i K^2 E_0 E}{3S} \tag{5.12}$$

E_d has cancelled out of the equation. Ishitani and Shimizu [5.1] point out that, at these higher ion energies, the parameter E_d does not affect the broadening since low energy collisions (a few E_d) become unimportant.

At this point, we pause to consider the applicability of this equation. Clearly, the assumptions made above are of a very approximate nature. As the ion energy increases, the form of $N(\varepsilon) \propto 1/\varepsilon^2$ which we have used becomes inaccurate. The number of high energy collisions is exaggerated due, in part, to the electronic energy losses at higher energies. We see that Eq. (5.12) is only applicable when the electronic energy loss is small. We may, as a further simplification, let the energy of the ion E_0 equal the nuclear energy loss E in Eq. (5.12). Thus,

$$W^2 = \frac{\gamma_i (KE)^2}{3S} \tag{5.13}$$

Our assumptions also break down when the mass of the dopant atom of interest differs significantly from the mass of a lattice atom.

Equation (5.13) can only be expected to indicate trends with energy or sputtering yield and to roughly predict the broadening. We will demonstrate that Eq. (5.13) yields reasonable agreement with Monte Carlo calculations and experimental data. As a zeroth-order approximation, then, Eq. (5.13) is justified by its utility and its simplicity.

Ishitani and Shimizu [5.1] have performed Monte Carlo calculations for 5 and 10 keV Ar^+ bombardment of silicon at 0° and 60° incidence. In their paper, they tabulate the nuclear energy loss E, the sputtering

yield S, and the damage yield which conforms fairly well to the Kinchin-Pease formula. They also examine the broadening of a step-function profile of tagged Si atoms and tabulate the 20% to 80% width of the broadened profile. Their results are ideally suited for a comparison with Eq. (5.13).

For the interaction of lattice atoms, the expression for K from Eq. (5.8) reduces to [5.6]

$$K = \frac{0.051}{nZ^{5/3}} \qquad \text{Å eV}^{-1} \tag{5.14}$$

For Si, K = 0.012. γ = 0.97 for Ar^+ incident on Si. The 20% to 80% width of an erf profile is 2.4 W. Results are shown in Table 5.2a for the two potentials, Lenz-Jensen (LJ) and Thomas-Fermi, considered in the Monte Carlo calculations [5.1]. The magnitudes of broadening compare favorably. In particular, the trends with energy and angle predicted by Eq. (5.13) agree well with the Monte Carlo calculations, as shown in Table 5.2b.

Finally, we return again to our Si/SiO_2 data from Fig. 5.8. We apply Eq. (5.13) with the value of K = 0.012 previously calculated for Si. We also use the measured sputtering yields of Fig. 5.8 and again assume that the original interface width is 2.0 nm. The results are shown as dashed lines in Fig. 5.9. The energy and ion dependence is quite reasonable. The accurate prediction of the magnitude of the broadening is somewhat fortuitous.

We have developed a simple model of ion knock-on broadening based on an analogy to thermal diffusion theory. Broadening occurs in rms fashion, i.e.,

$$W_M = \sqrt{W_0^2 + W^2} \qquad (5.15).$$

where W_0 is an unperturbed width and W is the broadening parameter. At low energies, ($E \lesssim 1$ keV), we obtained

$$W^2 = \frac{Ea^2}{6E_d S} \qquad (5.16)$$

where a is the nearest-neighbor distance in the solid. The energy dependence of W is seen to be quite weak (as E increases, so does S). The sputtering yield dependence indicates that, for two different bombarding ions of the same energy,

$$\frac{W_1^2}{W_2^2} = \frac{S_2}{S_1} \qquad (5.17)$$

At higher energies (1 keV $\lesssim E \lesssim 10$ keV), provided that the electronic energy losses of the incident ion and of the lattice or dopant atoms are small, we used the following semi-empirical formula:

$$W^2 = \frac{\gamma_i (KE)^2}{3S} \qquad (5.18)$$

Comparisons with Monte Carlo calculations and with an ASP experiment performed on the Si/SiO_2 interface produced reasonable agreement.

5.3.3 Redistribution of Phosphorus During the Thermal Oxidation of Silicon

During the thermal oxidation of heavily phosphorus-doped silicon, the phosphorus, which is relatively insoluble in SiO_2, is rejected by the oxide and pushed into the silicon where it diffuses away from the

interface. This process was first described in detail by Grove et al [5.7]. If oxide growth and diffusive transport are the only operative mechanisms in the process, the phosphorus concentration near the Si/SiO_2 interface should be no more than twice the bulk concentration for typical conditions, as illustrated in Fig. 5.10a. However, a magnified view of the interfacial region obtained using Auger sputter profiling shows an anomalous buildup of phosphorus near the interface, as illustrated in the inset in Fig. 5.10. This phosphorus pileup has been observed by several groups [5.8-5.12]. An explanation of its mechanism is the goal of our current research.

The pileup phenomenon has implications for device performance. We have shown that the rapid uptake of phosphorus at the interface has a significant effect on the doping level near the interface, especially during the early stages of oxidation. This may be particularly important for VLSI devices in which thin oxides, high doping levels, and shallow junctions may be present. The phosphorus pileup is also one of a class of related interfacial phenomena which are of current theoretical interest. These phenomena include dopant-enhanced oxidation [5.13,5.14,5.15], oxidation-enhanced diffusion [5.16], and doped oxide-source diffusion [5.17].

In order to determine the mechanism for the phosphorus pileup, we have determined the dependence of the observed phosphorus distribution on several different processing conditions. We have measured the dependence of the phosphorus distribution on the time of oxidation. We have also determined the effect of post-oxidation anneals. And, finally, we have investigated the effect of an additional electrically compensating dopant: boron. We will discuss here, in order, the Auger sputter

profiling method used to measure the phosphorus distributions, the processing conditions employed in our matrix of samples, and the consequence of our observations for the several proposed models of the phosphorus pileup.

The Auger sputter profiling (ASP) technique has been discussed in detail elsewhere [5.18]. Here, we briefly describe the details of the ASP technique as it applies specifically to these studies.

Our experiments were performed in a Varian 2730 Spectrometer with a base pressure of ~10^{-9} torr. Electrons from the integral gun of the variable-aperture cylindrical mirror analyzer stretch the sample at an angle of 30° to the surface normal. The 4.5 keV electron beam was operated at currents less than 10 μA rastered over a 200 × 200 μm square to minimize effects of electron-stimulated desorption and electron beam-induced diffusion [5.8,5.10,5.19]. Sputtering was performed using a pseudorandom rastered ion beam at beam energies of 1 keV or less at an angle of incidence of 49°. Neon was used so that a liquid N_2 cryopanel could be employed to minimize the partial pressure of contaminant gases. We have previously shown that, with these experimental parameters, the measured interface width is not seriously affected by electron escape-depth broadening [5.11] or ion knock-on mixing [5.20].

A typical set of derivative spectra obtained during profiling of the phosphorus-rich Si/SiO_2 interfacial region are illustrated in Fig. 5.11. As we sputter through the oxide, two peaks appear in the featureless SiO_2 spectrum--a weak elemental Si peak at 108 eV and the P_{LMM} peak at ~120 eV. The phosphorus peak grows rapidly and then decreases to a steady value in the silicon bulk. The phosphorus peak position also shifts slightly in the interfacial region, which may be a chemical effect.

Although, generally, the Auger spectroscopy technique can be sensitive to elements in concentrations below 10^{19} cm^{-3}, the practical sensitivity in these studies is limited by the overlap of the Si 108-eV and P 120-eV spectra. Because the silicon derivative spectrum has a small positive slope in the vicinity of the P (LMM) spectrum, peak-to-peak measurements of this spectrum alone will not detect phosphorus below 4×10^{19} cm^{-3} (as calibrated by ion implanted standards). In fact, assuming that the silicon background remains the same, all peak-to-peak determinations of phosphorus concentration will be underestimated by the same amount; this is shown in Fig. 5.12. Most measurements presented here, however, are for concentrations in excess of 10^{20} cm^{-3}, for which the sloping background will add only a small correction to the peak-to-peak concentration.

When the peak-to-peak height of the Si_{LVV} (92 eV), P_{LMM} (120 eV), and O_{KLL} (502 eV) transitions are plotted as a function of depth, a profile [5.11] similar to that of the inset in Fig. 5.10 is obtained. The distance over which the Si_{LVV} signal rises from 10% to 90% of its final value is termed the interface width. In lightly doped samples, the interface width is ~25 Å and does not vary with the oxide thickness [5.21]. In heavily doped samples, we have shown that the interface width increases significantly [5.11].

It is interesting to compare the Auger sputter profile of phosphorus in Fig. 5.13 with depth profiles of the same sample obtained elsewhere using the SIMS (secondary ion mass spectroscopy) technique. In Fig. 5.13, we compare the data of Dobrott [5.22] (20 keV O_2^+ sputtered) and Evans [5.23] (8 keV Cs^+ sputtered) with our own (1 keV Ar^+ sputtered). It is observed that the SIMS technique, as applied in their studies, has

greater sensitivity to phosphorus but poorer depth resolution due to the high ion sputtering energies.

Heavily doped Si (111) and (100) samples were prepared by thermal deposition and drive-in from a $POCl_3$ source. The resultant phosphorus profiles were flat to within 10% to a distance of 1 μ into the silicon. Oxidations were performed in dry O_2 at 900° and 1150°C with 30 sec push-pulls in O_2.

Three sets of experiments were performed. First, the effect of oxidation time and therefore oxide thickness on the pileup was measured at an oxidation temperature of 900°C for (100) oriented samples. Second, the effect of subsequent post-oxidation N_2 anneals was determined. Finally, the effect of an additional electrically compensating dopant, boron, on the pileup effect was measured.

A series of experiments were performed to examine the evolution of the pileup during oxidation. Si (100) samples with bulk phosphorus doping of 2×10^{20} cm^{-3} (as measured by ASP calibration from ion-implanted standards) were oxidized in dry O_2 at 900°C for various lengths of time. The oxidation times used in these experiments are tabulated in Table 5.3. Oxide thicknesses were determined from previous calibrations [5.24] by the time required to sputter to the interface and are also listed in the table. The resultant interface widths, again the 10% to 90% width of the Si_{LVV} transition, appear in the third column of the table. As previously reported, the interface width appears to increase with oxide thickness. This effect is not observed in undoped samples [5.23].

The phosphorus profiles from the samples of Table 5.3 are collectively illustrated in Fig. 5.14. This figure shows the phosphorus profiles as a function of depth measured from the outside surface of each sample.

Each phosphorus peak is situated at the Si/SiO_2 interface, as shown in Fig. 5.12. The sample oxidized for 90 min, for example, has an oxide 550 Å thick and, so, shows the phosphorus peak at a depth of 550 Å. The peak heights shown were reproducible to within ~20%.

We see in Fig. 5.14 that a pileup is just beginning to form in the 1 min oxide. The next profile, for a 10 min oxidation, illustrates an important phenomenon. Much more phosphorus is present in the pileup region than can be accounted for by the displacement of phosphorus from 70 Å of oxide. The concentration of phosphorus on the silicon side of the pileup is observed to dip below the bulk doping level. Thus, we find that the interface is acting as a sink for phosphorus, with phosphorus diffusing from the bulk of the Si to the interface where it is immobilized. Succeeding profiles in Fig. 5.14 show that, for oxide thicknesses greater than 200 Å, the peak phosphorus concentration is no longer increasing, i.e., the pileup has saturated. As the oxide thickness increases, the phosphorus level on the silicon side of the pileup rises above the bulk doping level. Thus, any further phosphorus which is rejected by the growing oxide redistributes in the silicon bulk in a manner consistent with previous models [5.7] of phosphorus accumulation.

As a further check of the concentration dip effect, we examined a less heavily doped sample ($\sim 5 \times 10^{-19}$ cm^{-3}) which was oxidized concurrently with the 10-min heavily doped (2×10^{20} cm^{-3}) sample of Fig. 5.14. A comparison of these two profiles is shown in Fig. 5.15. The oxide thickness of the heavily doped sample is slightly larger, consistent with the oxidation enhancement model of Ho et al [5.14]. In the less heavily doped sample, we again observe a prominent concentration dip.

In order to determine whether or not the observed phosphorus distribution is related to the motion of the Si/SiO_2 interface, or is instead an inherent property of the interface, we measured the effect of N_2 annealing on phosphorus pileup. One sample with a bulk doping level of 2×10^{20} cm^{-3} was oxidized for 10 min at 1150°C (thickness of 450 Å) and later annealed for 14 min at the same temperature. Assuming conventional values for the phosphorus diffusion constant [5.24], the phosphorus that is observed at the interface after oxidation should be able to diffuse back into the bulk of the silicon after only 0.1 sec at 1150°C. As shown in Fig. 5.16, however, no significant change in the phosphorus distribution was detected after the post-oxidation anneal. The diffusion length of phosphorus under the conditions of the anneal was 3300 Å.

In a further test, a sample grown at 900°C for 10 min (thickness of 48 Å) was annealed at 900°C for 8.75 hr. The sample prior to the anneal showed a depletion of phosphorus on the silicon side of the phosphorus pileup (see Figs. 5.14 and 5.15). After the anneal, the phosphorus pileup had grown to the height of that in the sample oxidized for 20 min, and the depleted area had redistributed itself into the bulk so as to be undetectable. The diffusion length for this anneal was ~1500 Å.

One of the heavily phosphorus doped ($\sim 6 \times 10^{20}$ cm^{-3}) samples was further subjected to a 60-min boron deposition at 1100°C from a diborane source. The residual oxide was etched, and the sample was oxidized in dry O_2 for 4 hr at 900°C. An Auger sputter profile indicated a bulk boron doping level of $\sim 2 \times 10^{21}$ cm^{-3} (5 at.%), well above the solid solubility limit. The sputter profile also indicated that the phosphorus pileup had almost completely disappeared. The peak phosphorus concentration was roughly 10% larger than the bulk concentration. The interface width,

as measured by the 10% to 90% width of the Si_{LVV} profile, was ~80 A, much broader than the typical width of a similar sample with no boron present. Ho [5.15] has reported that the oxidation rate of this sample is much reduced by boron compensation.

As has been pointed out previously, the measured phosphorus profiles are not consistent with a simple diffusion-redistribution model, as these models would predict a much smaller peak concentration with the pileup region extending much further into the silicon.

A model put forth in a previous paper from this laboratory suggested that the pileup may be caused by the formation of a $P^+V^=P^+$ complex that would diffuse slowly and therefore cause a high concentration of phosphorus to build up near the interface. Our present data is inconsistent with this model in two ways. First, the depletion of phosphorus in the silicon in the near interface region for short oxidation times (see Fig. 5.15) clearly cannot be explained by this mechanism. Second, these complexes would tend to diffuse back into the bulk of the silicon during the N_2 anneal; the pileup, however, does not decrease during the anneal and, in fact, increases for some initial conditions. This mechanism therefore cannot account for the pileup observed.

Our data can only be explained by considering the interface as a sink for phosphorus either during oxidation or during anneals. This behavior is therefore a property of the interface and not a direct consequence of the oxidation process itself. The behavior of the Si/SiO_2 interface as a sink for phosphorus implies that the chemical potential for phosphorus has a sharp minimum at the interface.

There are several mechanisms which can account for this chemical potential sink at the interface. Since the phosphorus is positively

charged in the silicon, it might be attracted to a negative electrical potential at the interface. Since the phosphorus is a different size than silicon and will have, in general, a different bonding geometry, it might reduce the interfacial strain energy. The phosphorus might also form a chemical species at the interface (an SiP or an SiP_xO_y) which is not observable in the bulk. And, finally, the phosphorus might accumulate where vacancies, with which it is normally paired, have been depleted.

5.3.4. Auger Sputter Profiling Studies of SiO_2 Grown in O_2/HCl Mixtures

In recent years there have been many studies of the effects of introducing small amounts of chlorine into silicon dioxide during the growth process [5.26-5.33]. Chlorine allows the fabrication of oxides with reproducible breakdown voltages, immobilizes positive alkali ions in the oxide, and directly affects the oxidation kinetics. Some people believe that both oxygen and chlorine enter the SiO_2 lattice as negative ions. If this is true, chlorine may additionally provide insight into the mechanisms of oxide growth.

We have performed Auger Sputter Profiling (ASP) experiments [5.34, 5.35] on silicon oxides grown in HCl/O_2 ambients. Chlorine is observed to segregate to the SiO_2/Si interface. This chlorine pileup directly effects interface states and the morphology of the interface region. Chlorine may also behave in a fashion similar to the oxidizing species during the oxidation process, so an understanding of the evolution of the chlorine distribution is clearly desirable.

In this report we will introduce a simple model to predict the amount of interface chlorine for a given oxidation process and then present Auger data exhibiting the chlorine interface pileup as a function of time and temperature.

We have studied oxides of silicon grown in O_2/5% HCl at 1000°C and 1100°C using Auger sputter profiling. A typical oxide chlorine distribution profile from our studies appears as in Fig. 5.17. This sample was grown in O_2 5% HCl at 1100°C. In each sample of this type which we have studied, a large chlorine peak is observed in the oxide, adjacent to the SiO_2/Si interface. Other groups using ASP, SIMS, and other techniques have found similar results [5.36, 5.37]. Fig. 5.18 is a summary of the total interface chlorine as a function of the oxide thickness at both 1000°C and 1100°C. Note that the chlorine pileup increases monotonically with oxide thickness for a fixed temperature.

In their classic paper on the general relationship for silicon oxidation [5.38], Deal and Grove used a steady state approximation to the diffusion equation for the oxidizing species. We will take a similar approach here. In Fig. 5.19 we show a growing silicon dioxide layer containing a linear distribution of chlorine. This linear approximation arises essentially from the assertion that the oxide layer is sufficiently transparent to the chlorine so that the chlorine distribution can quickly adjust itself as the oxide thickness increases. This assumption of a steady state distribution is valid if [5.38]

$$t_{tr} = \frac{L^2}{2D_{Cl}} \ll t \qquad (5.19)$$

where t_{tr} is a transient time, L is the oxide thickness at time t, and D_{Cl} is the chlorine diffusion coefficient. L is given by the Deal-Grove expression [5.38]

$$L^2 + AL = B(t + t) \tag{5.20}$$

and approximating L^2 by Bt yields

$$\frac{t_{tr}}{t} \simeq \frac{B}{2D_{cl}} << 1 \tag{5.21}$$

as the criterion for validity of the steady state assumption.

Assuming that the steady state solution is valid (this will be examined later), then Fick's second law becomes

$$\frac{\partial C}{\partial t} = \frac{-\partial J}{\partial x} \simeq 0 \tag{5.22}$$

where C and J denote chlorine concentration and flux respectively. This implies that flux is a constant independent of position, and Fick's first law gives

$$J = J_o = - D \frac{\partial C}{\partial x} \tag{5.23}$$

then

$$C_{(x,t)} = \frac{(Cs-Ci)}{L(t)} x, \tag{5.24}$$

where Cs and Ci are the surface and interface chlorine concentrations respectively.

It has been observed that sufficient doses of chlorine will form blisters at the SiO_2/Si interface, and our data show no indication of a saturation behavior of the chlorine pileup. This indicates that a solid solution of chlorine at the interface will not be an equilibrium situation. This makes it reasonable to treat the SiO_2/Si interface as a very deep potential well for chlorine, with the consequence that Ci, the excess chlorine near the interface is zero for all times. Then our equations become

$$C(x,t) = \frac{Cs}{L(t)}\, x, \text{ and}$$

$$J(x,t) = \frac{-DCs}{L(t)} \tag{5.25}$$

We integrate J over time to determine the total interface chlorine Q at time t:

$$Q(t) = Q(o) + \int_{t=o}^{t2} (-J)\, dt \tag{5.26}$$

To express this in terms of L, note that

$$\frac{dL}{dt} = \frac{B}{2L+A}, \quad \text{so} \tag{5.27}$$

$$Q(L) = Q_o + \int_{L_1}^{L_2} \left(\frac{2L + A}{B}\right) \frac{DCs}{L}\, dL \tag{5.28}$$

$$Q(L) = Q_o + \frac{DCs}{B} \left[2(L_2-L_1) + A \ln {}^{L_2}/_{L_1}\right] \tag{5.29}$$

Where Q_o is the interface chlorine for an initial oxide thickness L_1. We will compare this relation with the results of our Auger data.

In Fig. 5.20 we show a comparison of the simple expression for Q, where DCs/B was evaluated at the 2000Å and 1600Å data points for 1000°C and 1100°C respectively. We set Q_o = 0 for the solid curves, and it is worth noting that a nonzero Q_o would slightly increase the values for Q for small values of L. The agreement of our Auger data with the model is reasonably good. It must be noted also that there is roughly a 10% standard deviation in the Auger data.

It is now necessary to check our assumption of steady state conditions during oxidation. That is

$$\frac{B}{2D_{Cl}} \ll 1 \tag{5.31}$$

This condition reduces to

$$Cs << 3x10^{18}\ cm^{-3} \text{ for } T = 1000^{o}C,$$
$$Cs << 8x10^{19}\ cm^{-3} \text{ for } T = 1100^{o}C \qquad (5.32)$$

This condition appears to be satisfied from Auger results at $1100^{o}C$, but at $1000^{o}C$ this condition may not be satisfied. It is important to note however, that the steady state criterion as stated is overly restrictive, and the assumption of a linear distribution of Cl in the oxide may be valid for all cases we have considered.

It remains to consider the effect of electron beam irradiation on the amount of chlorine pileup at the interface. Our studies have allowed us to quantitatively predict the amount of chlorine swept out of a given bulk distribution of chlorine. For the samples in this study, the contribution of electron beam irradiation to the total amount of interface chlorine is less than 10%. It is clear then, that our Auger results give an accurate determination of the chlorine distribution in the oxide after oxidation is complete.

REFERENCES

[5.1] T. Ishitani and R. Shimizu, Appl. Phys., 6, 241 (1975).

[5.2] J.W. Coburn, J. Vac. Sci. Tech., 13, 1037 (1976).

[5.3] P.K. Haff and Z.E. Switkowski, J. Appl. Phys., 48, 3383 (1977).

[5.4] S.A. Schwarz and C.R. Helms, J. Vac. Sci. Tech., in press.

[5.5] H.H. Andersen, Appl. Phys., 18, 131 (1979).

[5.6] S.A. Schwarz and C.R. Helms, to be published.

[5.7] A.S. Grove, O. Leisteko, Jr., and C.T. Sah, J. Appl. Phys., 35, 2695 - 2701 (1964).

[5.8] C.C. Chang, A.C. Adams, G. Quintana, and T.T. Sheng, J. Appl. Phys. 45, 252-256 (1974).

[5.9] N.J. Chou, R. Hammer, Y.J. Van der Meulen, and J. Cahill, Appl. Phys. Lett., 24, 200-202 (1974).

[5.10] J.S. Johannessen, W.E. Spicer, J.F. Gibbons, J.D. Plummer, and N.J. Taylor, J. Appl. Phys., 49, 4453-4458 (1978).

[5.11] S.A. Schwarz, C.R. Helms, W.E. Spicer, and N.J. Taylor, J. Vac. Sci. Tech., 15, 227-230 (1978).

[5.12] T.H. DiStefano, in ARPA/NBS Workshop: Surface Analysis for Silicon Devices (A.G. Lieberman, editor), NBS Special Publication 400-23, 197-210 (1976).

[5.13] C.P. Ho, J.D. Plummer, J.D. Meindl, and B.E. Deal, J. Electrochem. Soc. 125, 665-671 (1978).

[5.14] C.P. Ho and J.D. Plummer, J. Electrochem. Soc., 126, 1516-1530 (1979).

[5.15] C.P. Ho, PhD. Thesis, Stanford Univ., Stanford, Ca. (1978).

[5.16] D.A. Antoniadas, A.M. Lin, and R.W. Dutton, Appl. Phys. Lett., 33, 1030-1033 (1978).

[5.17] R.N- Ghoshtagore, Sol. St. Elec., 17, 1065-1073 (1974).

[5.18] C.C. Chang, in Characterization of Solid Surfaces (P.F. Kane and G.B. Larrabee, editors), Plenum Press, New York, 509-576 (1979).

[5.19] J.S. Johannessen, W.E. Spicer, and Y.E. Strausser, Appl. Phys. Lett. 27, 452-454 (1975).

[5.20] S.A. Schwarz and C.R. Helms, J. Vac. Sci. Tech., 16, 781-783 (1979).

[5.21] C.R. Helms, N.M. Johnson, S.A. Schwarz, and W.E. Spicer, in The Physics of SiO_2 and its Interfaces (S.T. Pantalides, editor), Pergamon Press, New York, 366-372 (1978).

[5.22] R.D. Dobrott, Materials Characterization Report, Texas Instruments, Inc., Dallas, Tx. (1976) unpublished.

[5.23] C.A. Evans, Jr., Technical Report, Univ. of Ill., Urbana, Ill. (1977) unpublished.

[5.24] C.R. Helms, N.M. Johnson, S.A. Schwarz, and W.E. Spicer, J. Appl. Phys., to be published.

[5.25] R.B. Fair and J.C.C. Tsai, J. Electrochem. Soc., 124, 1107-1118 (1977).

[5.26] Y.S. Van der Meulen, C.M. Osburn, and J.F. Ziegler, Journ. Electrochem. Soc., 122, 284.

[5.27] D.W. Hess and B.E. Deal, Journ. Electrochem. Soc., 124, 735 (1977).

[5.28] B.E. Deal, Journ. Electrochem. Soc., 125, 576 (1978)

[5.29] B.E. Deal, D.W. Hess, J.D. Plummer, and C.P. Ho, Journ. Electrochem. Soc., 125, 339 (1978).

[5.30] R.J. Kriegler, Semiconductor Silicon, 363 (1973).

[5.31] Y.J. Van der Meulen, J.G. Cahill, Journ. Electronic Materials, 3, 371 (1974).

[5.32] B.R. Singh, and P. Balk, Journ. Electrochem. Soc., 125, 453 (1978).

[5.33] A. Rohatgi, S.R. Butler, and F.J. Feigl, Appl. Phys. Lett., 30, 104 (1977).

[5.34] J.S. Johannessen, C.R. Helms, W.E. Spicer, and Y.E. Strausser, IEEE Trans. Electron Dev., ED-24, No. 5, 547 (1977).

[5.35] J.S. Johannessen, W.E. Spicer, and Y.E. Strausser, Journ. Appl. Phys., 47, 3028 (1976).

[5.36] N.J. Chou, C.M. Osburn, Y.J. Van der Meulen, and R. Hammer, Appl. Phys. Lett., 22, 380.

[5.37] B.E. Deal, A. Hurrle, and M.J. Schulz, Journ. Electrochem. Soc., 125, 2024 (1978).

[5.38] B.E. Deal and A.S. Grove, Journ. Appl. Phys., 36, #12, 3770 (1965).

Table 5.1

MAXIMUM ESTIMATED AES SENSITIVITIES

Element	Sensitivity	Comments
B	10^{18} cm^{-3}	Ar interference
C	3×10^{17} cm^{-3}	background contamination
N	3×10^{17} cm^{-3}	---
O	3×10^{17} cm^{-3}	background contamination
F	3×10^{17} cm^{-3}	---
Na	10^{18} cm^{-3}	beam induced drift
Al	3×10^{17} cm^{-3}	Si interference
Si	3×10^{17} cm^{-3}	---
P	3×10^{17} cm^{-3}	Si interference
Cl	10^{17} cm^{-3}	Ar interference
As	3×10^{18} cm^{-3}	---
Sb	10^{17} cm^{-3}	---

Table 5.2(a)

$Ar^+ \rightarrow Si$		Broadening Parameters (nm)			
		5 keV		10 keV	
Angle θ	Potential	Monte Carlo	$2.4\sqrt{W}$	Monte Carlo	$2.4\sqrt{W}$
0°	LJ	14.0	9.1	26.0	16.4
0°	TF	8.0	6.9	14.0	12.3
60°	LJ	9.0	5.5	15.0	8.9
60°	TF	6.0	4.0	10.0	7.6

Table 5.2(b)

Ratios		Monte Carlo	2.4 W
$\frac{10\text{ keV}}{5\text{ keV}}$	0° LJ	1.9	1.8
	60°	1.7	1.6
	0° TF	1.8	1.8
	60°	1.7	1.9
$\frac{60°}{60°}$	5 keV LJ	1.6	1.7
	10 keV	1.7	1.8
	5 keV TF	1.3	1.7
	10 keV	1.4	1.6

Monte Carlo calculations[5.1] of the broadening induced by Ar^+ bombardment of Si are compared to the predictions of the present model.

Table 5.3

DATA FOR 900°C DRY OXIDATION OF Si (100) SAMPLES WITH INITIAL P DOPING OF $\sim 2 \times 10^{20}$ cm^{-3}

Oxidation Time (min)	Oxide Thickness (Å)	Interface Width* (Å)
1	22	21
10	67	23
20	121	34
90	548	38
200	750	40
750	1760	51

*10% to 90% width of the Si_{LVV} (92 eV) profile.

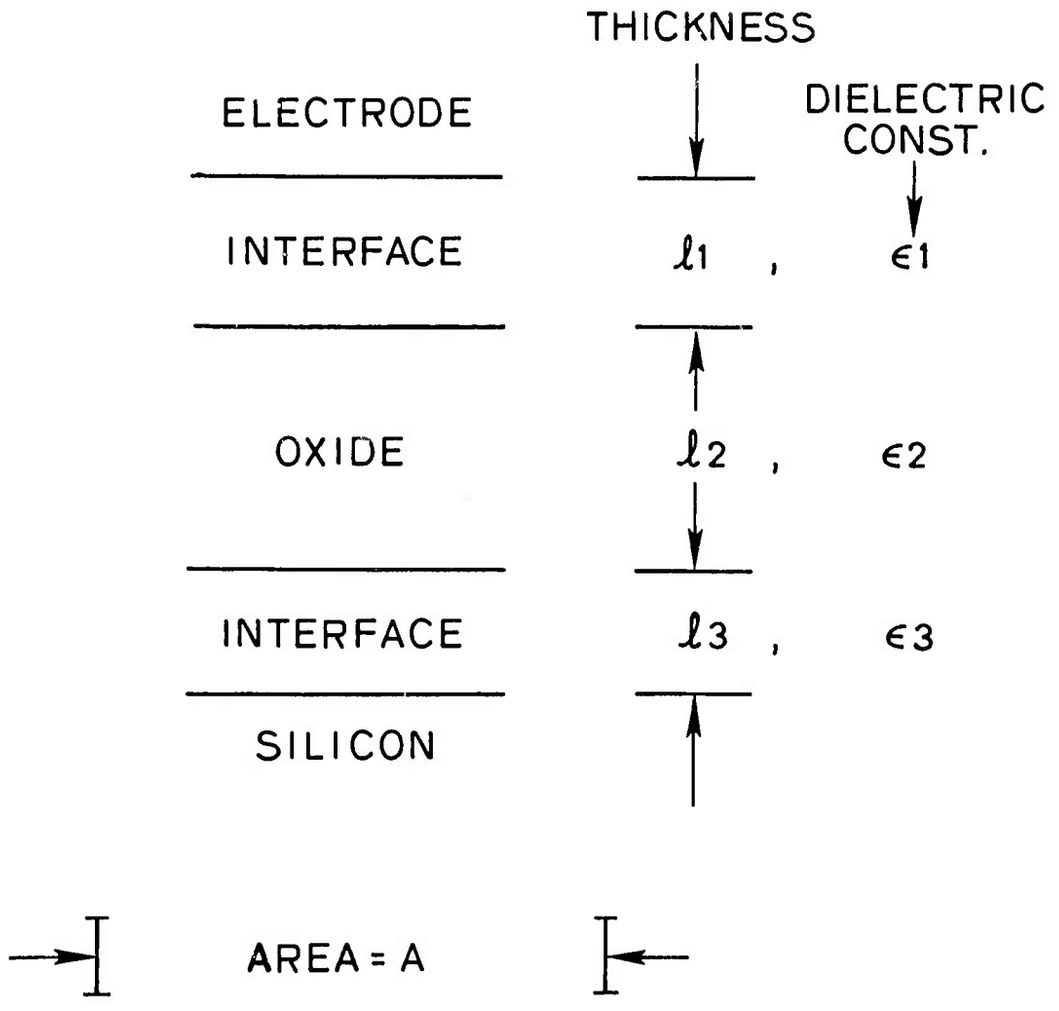


Figure 5.1 Schematic of an MOS structure indicating the importance of the interfaces present if the oxide thickness ℓ_2 is small.

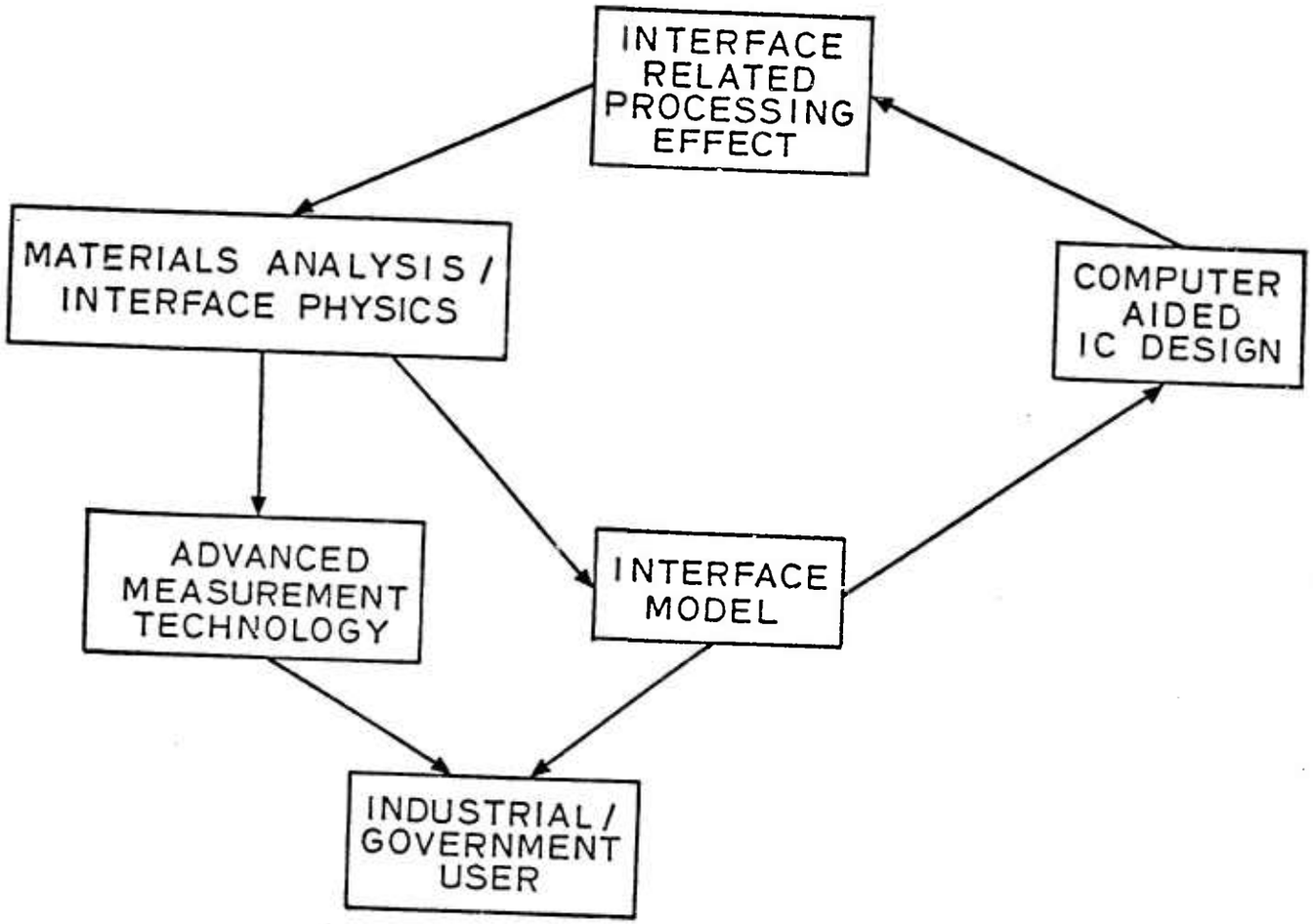


Figure 5.2 Block diagram of the relationship between materials analysis and interface physics section with other parts of this program.

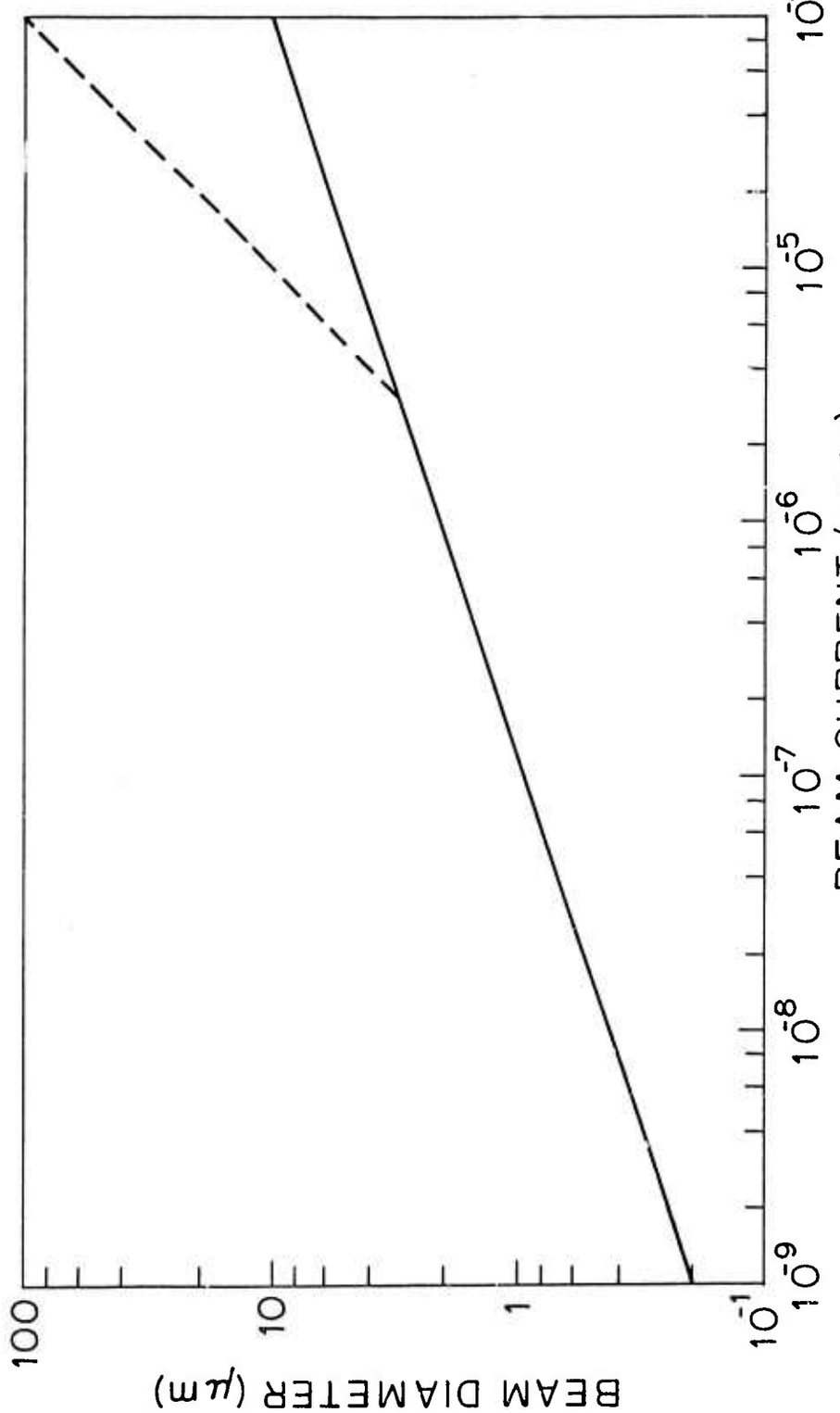


Figure 5.3 Typical beam diameter versus beam current with LaB_6 emitter--solid curve. Dashed segment corresponds to the approximate increase in beam size necessary to keep the local sample temperature down to reasonable levels in silicon

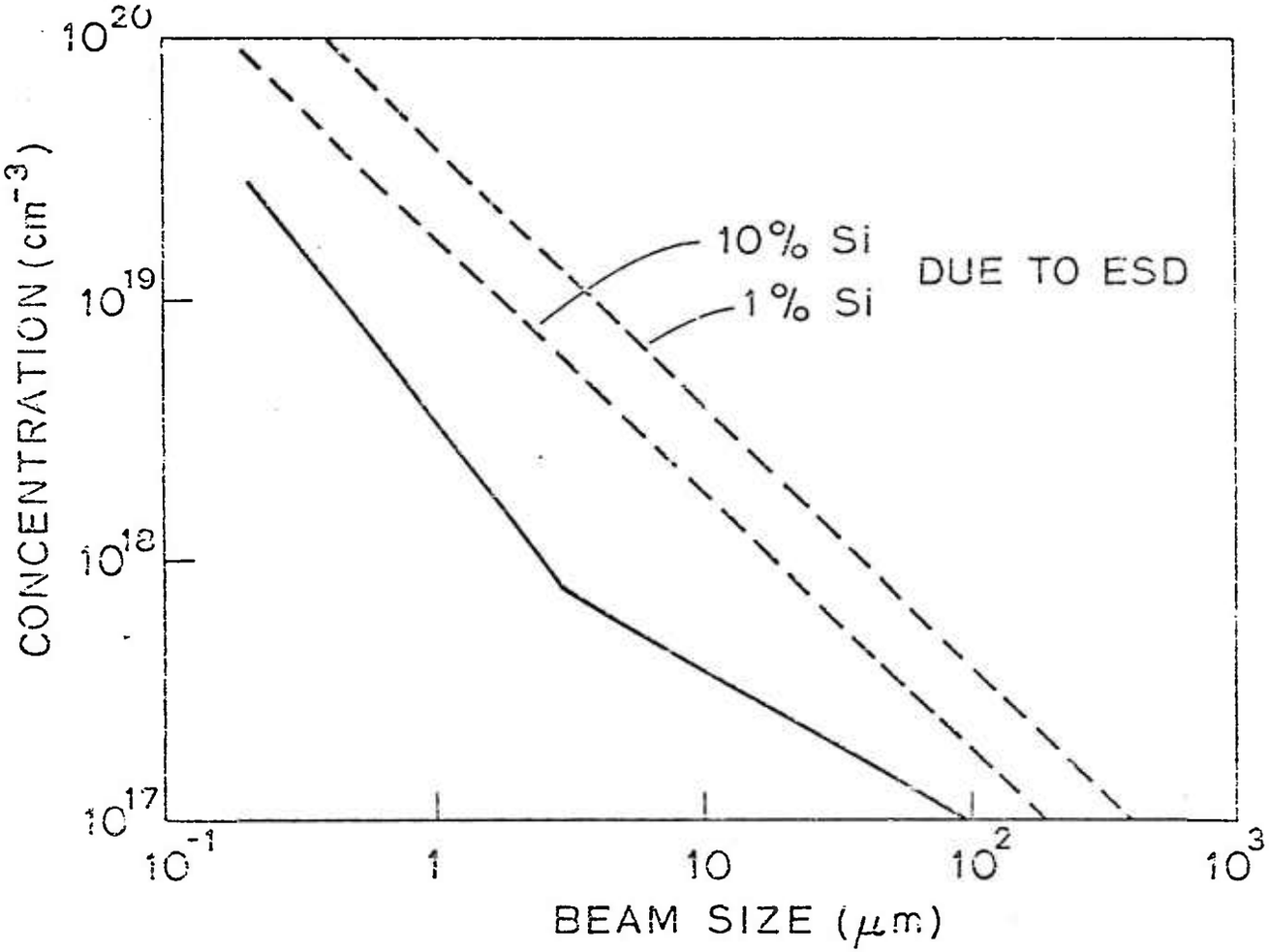


Figure 5.4 Maximum sensitivity of AES (to chlorine or silver transitions) as a function of electron beam diameter. Solid curve derived from Figure 5.3 Dashed curves correspond to the degradation in sensitivity in SiO_2 due to electron stimulated desorption esd. The 10% Si and 1% curves correspond to beam size and sensitivity such that the spectra will contain 10% or 1% contribution from elemental silicon due to esd.

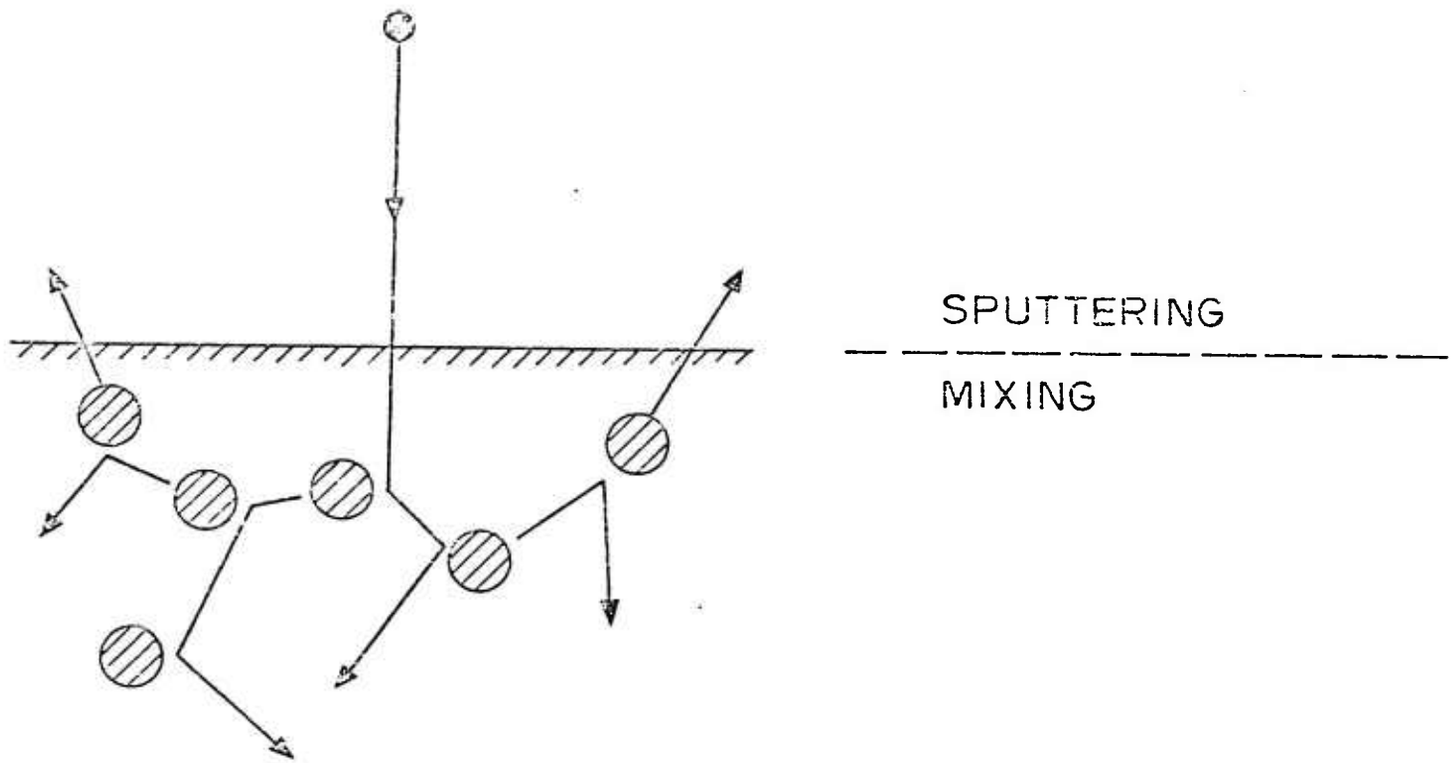


Figure 5.5 Schematic illustration of the sputtering and mixing processes resulting from ion bombardment.

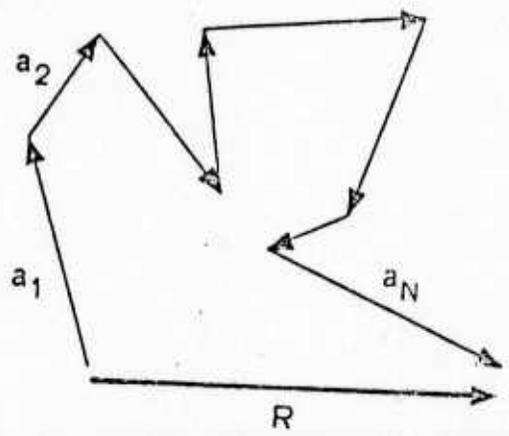


Figure 5.6 Schematic illustration of a random walk with N steps of lengths a_i and a net distance $\overline{\quad}$ travelled R. $R^2 = Na^2$

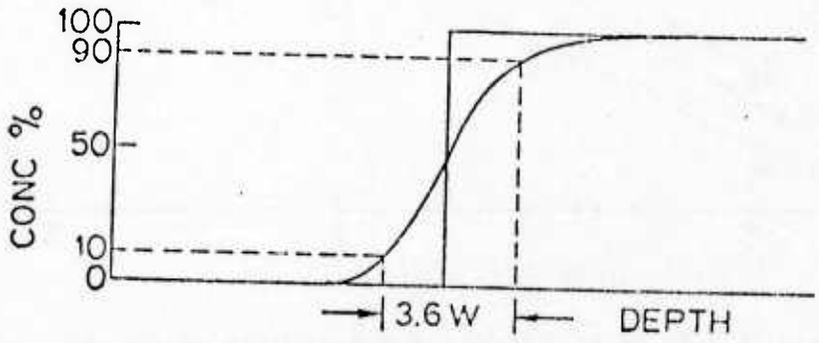


Figure 5.7 A step function concentration profile is broadened by ion knock-on mixing into an erf profile with a 10-90% width of 3.6 W.

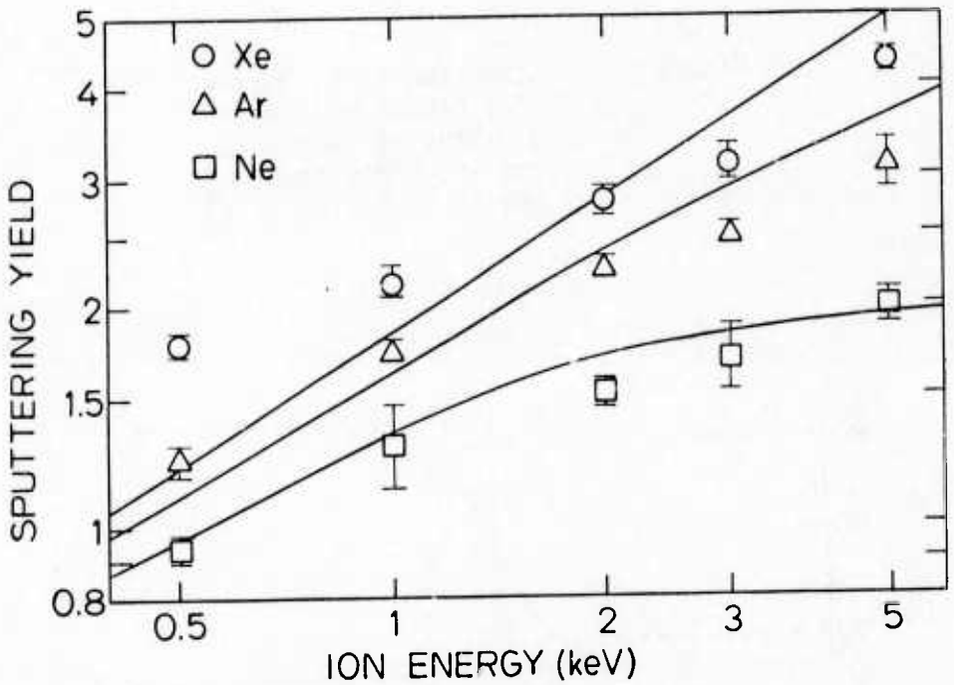


Figure 5.8 Average and standard deviation of sputtering yields (atoms/ion) in thermal SiO_2 as a function of ion energy for Ne^+, Ar^+, and Xe^+ bombardment. Solid curves are the predictions of our sputtering model for a lattice of Ne atoms (similar mass).

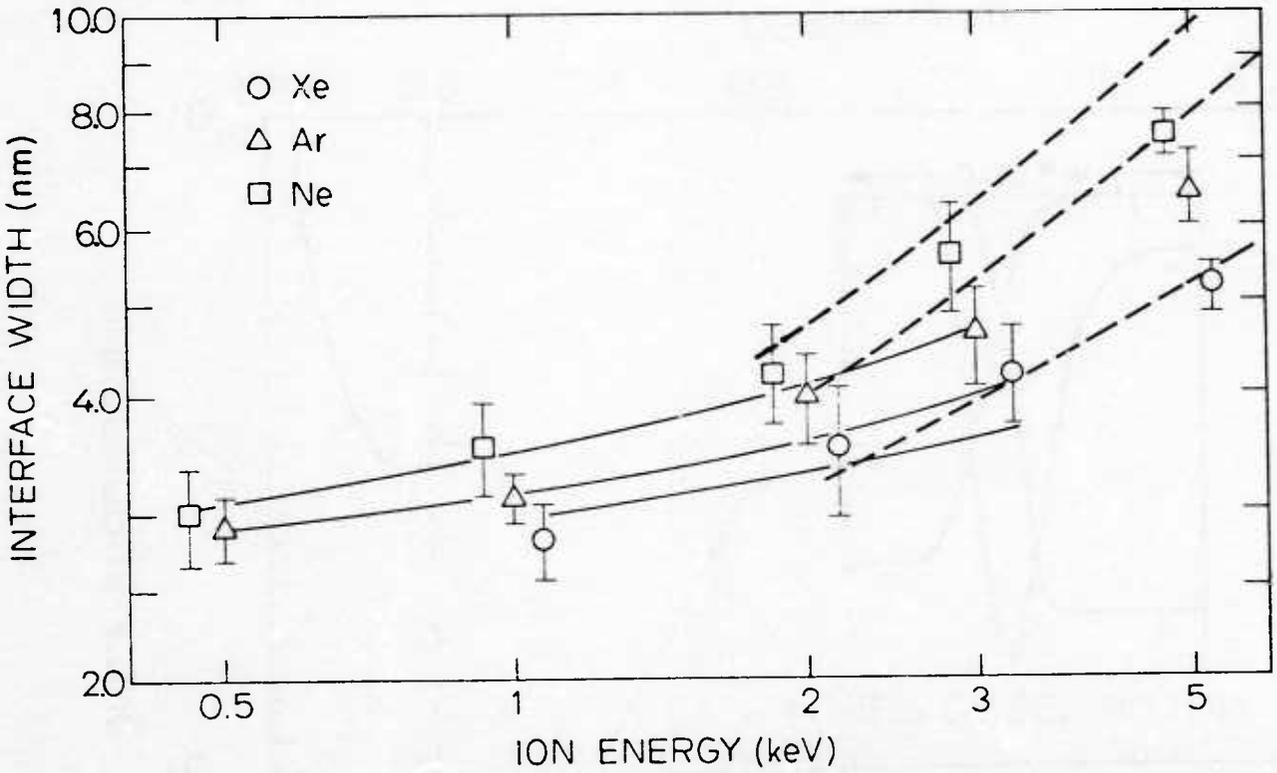


Figure 5.9 Average and standard deviation of the measured 10-90% Si-SiO_2 interface widths as a function of ion energy for Ne^+, Ar^+ and Xe^+ bombardment. Solid curves are the low energy predictions of our mixing model for an original interface width of 20 Å. Dashed curves are predicted by the high energy formula for W.

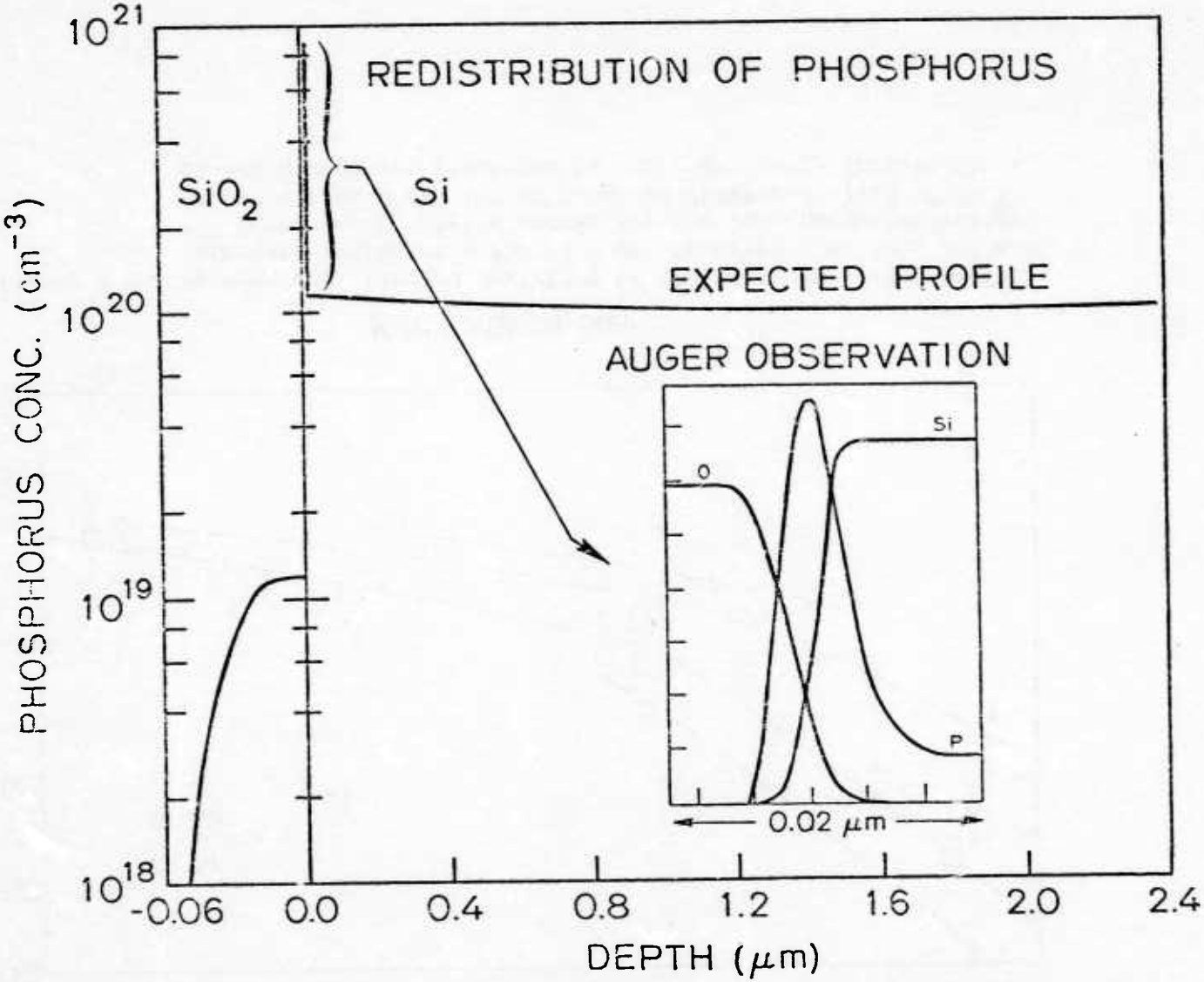


Figure 5.10 Illustration of the phosphorus pileup effect at the Si-SiO_2 interface for a bulk concentration of 10^{20} cm^2 and a sample oxidized at 1150°C for 10 min. Except for the delta-function-like contribution near the interface, the profile shown has been calculated using traditional diffusion redistribution theory. Note, the falloff in phosphorus concentration at the surface of the oxide is due to evaporation of phosphorus. The insert shows an actual measurement from this work of the near interface region, indicating the anomalous pileup observed.

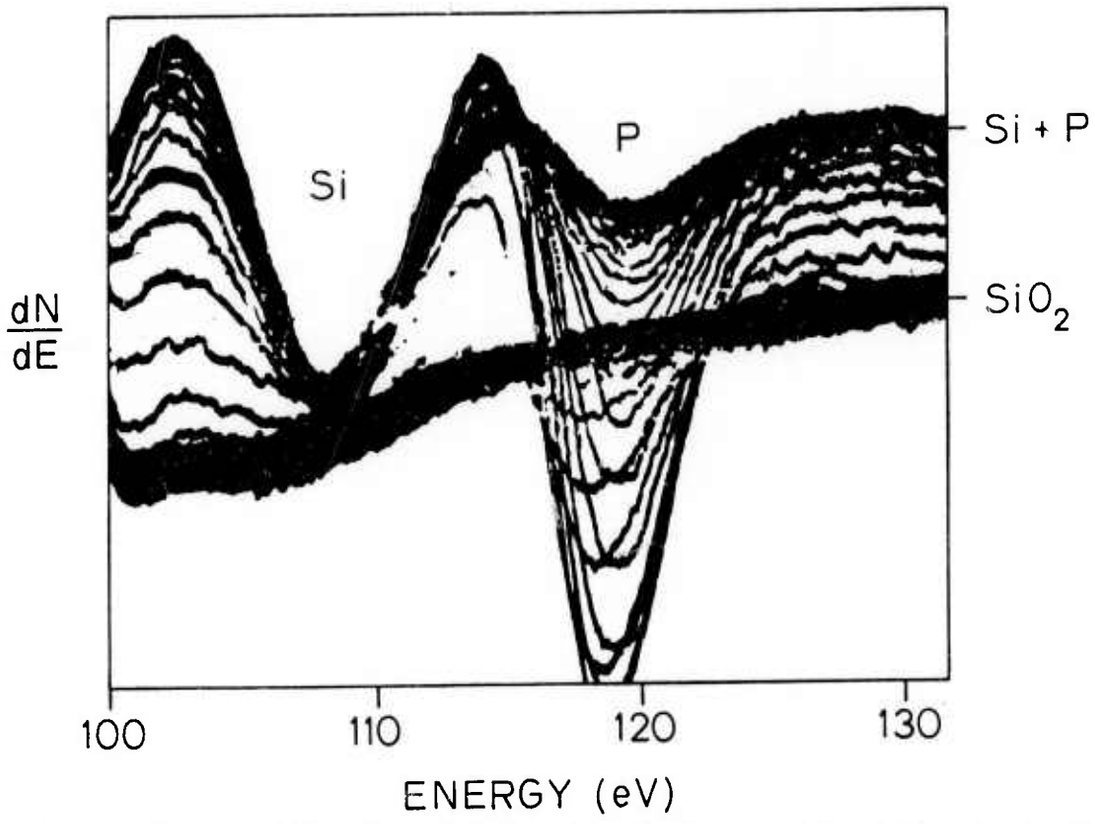


Figure 5.11 Oscil oscope photograph of Auger spectra being recorded during profile from an electron energy of 104 to 134 eV. For elemental silicon, there is a weak LVV peak that is present at ~108 eV, and the phosphorus LMM transition appears at 120 eV. The sloping background in the vicinity of the phosphorus peak cause errors in quantitative analysis that are illustrated in Figure 5.12

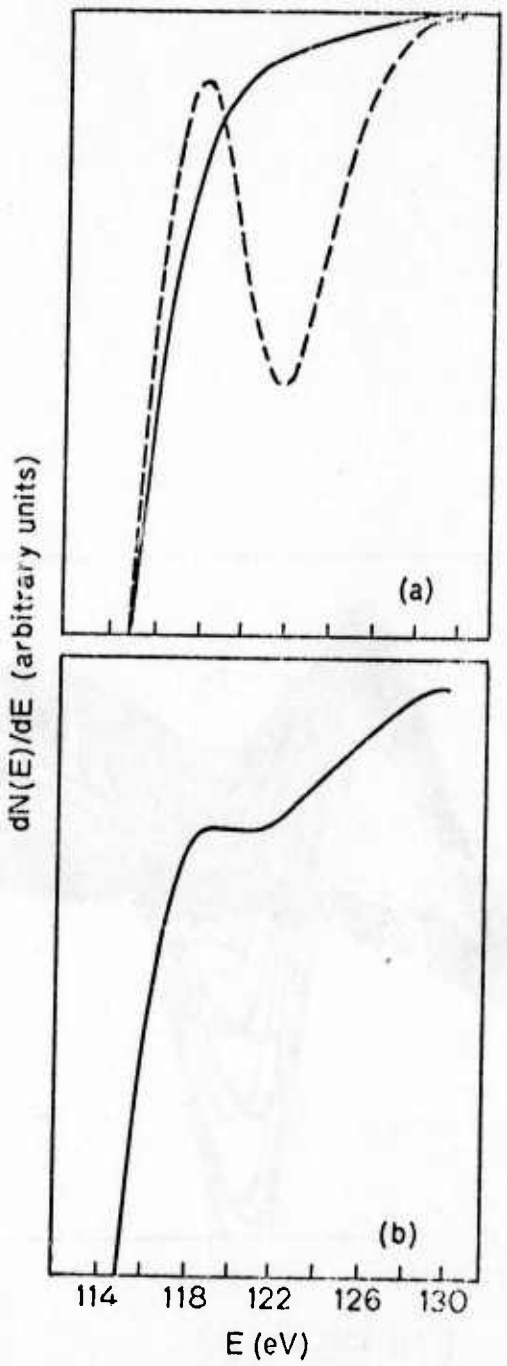


Figure 5.12 Examples of phosphorus Auger spectra taken from an implanted standard in silicon.

(a) The solid line is the silicon spectrum for a sample without phosphorus, the dotted line is a spectrum for a sample with 1×10^{20} cm^{-3} phosphorus.

(b) Auger spectrum for 4.0×10^{19} cm^{-3} phosphorus. Due to the sloping background, a simple peak-to-peak measurement would erroneously indicate zero concentration.

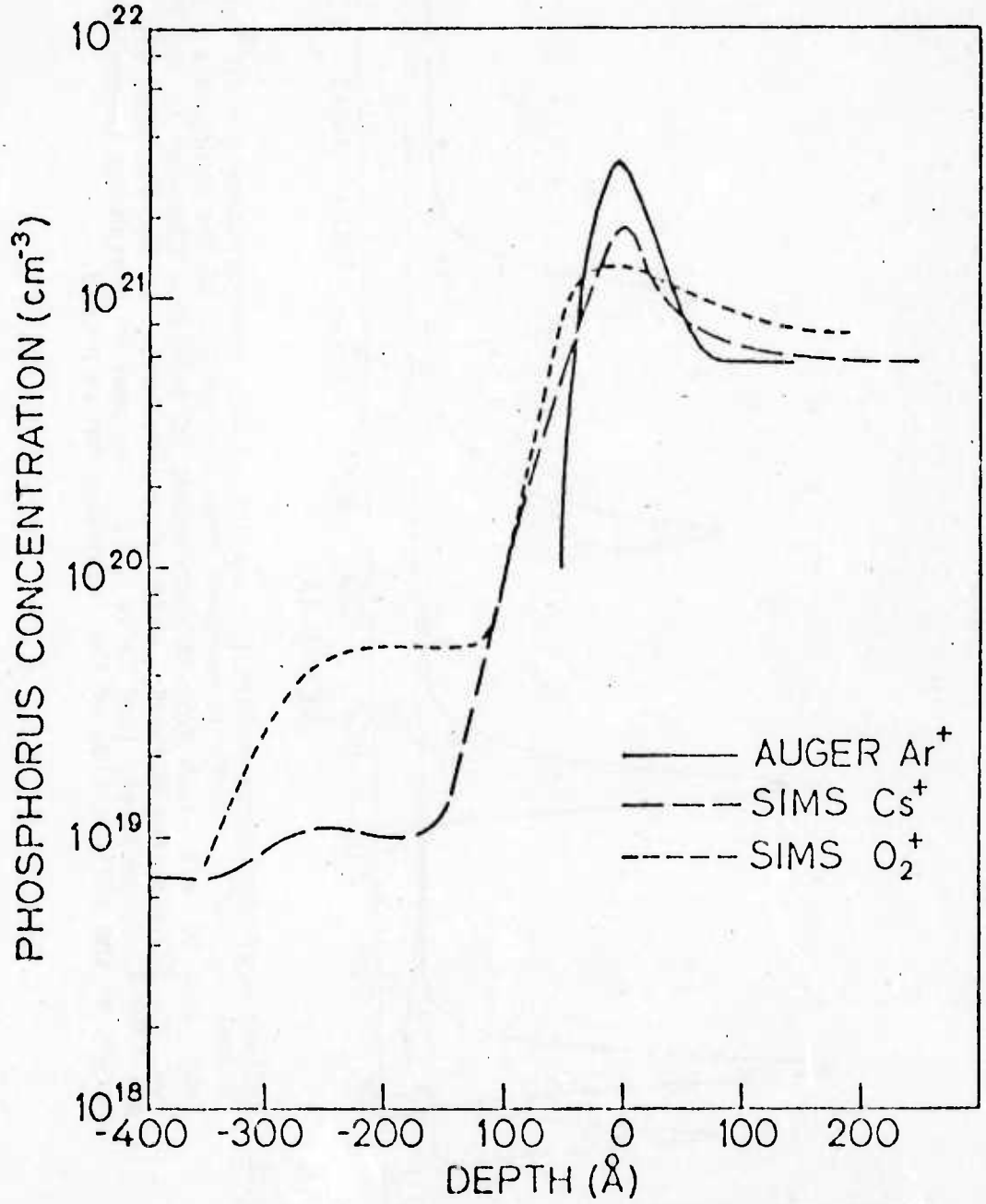


Figure 5.13 Comparison of SIMS and Auger profiles of phosphorus pileup. Auger data is shown as the solid line; SIMS data from Dobrott 5.22 (20 keV O_2^+ sputtering) is shown as the short dashed line; SIMS data from Evans 5.23 (8 keV Cs^+ sputtering) is shown as the long dashed line. The broadening observed in the SIMS experiments is due to the higher ion energies used. The SIMS technique does have higher sensitivity to phosphorus; the deviation in the phosphorus levels from the two SIMS profiles in the SiO_2 may be due to the presence of SiH which has the same mass as phosphorus.

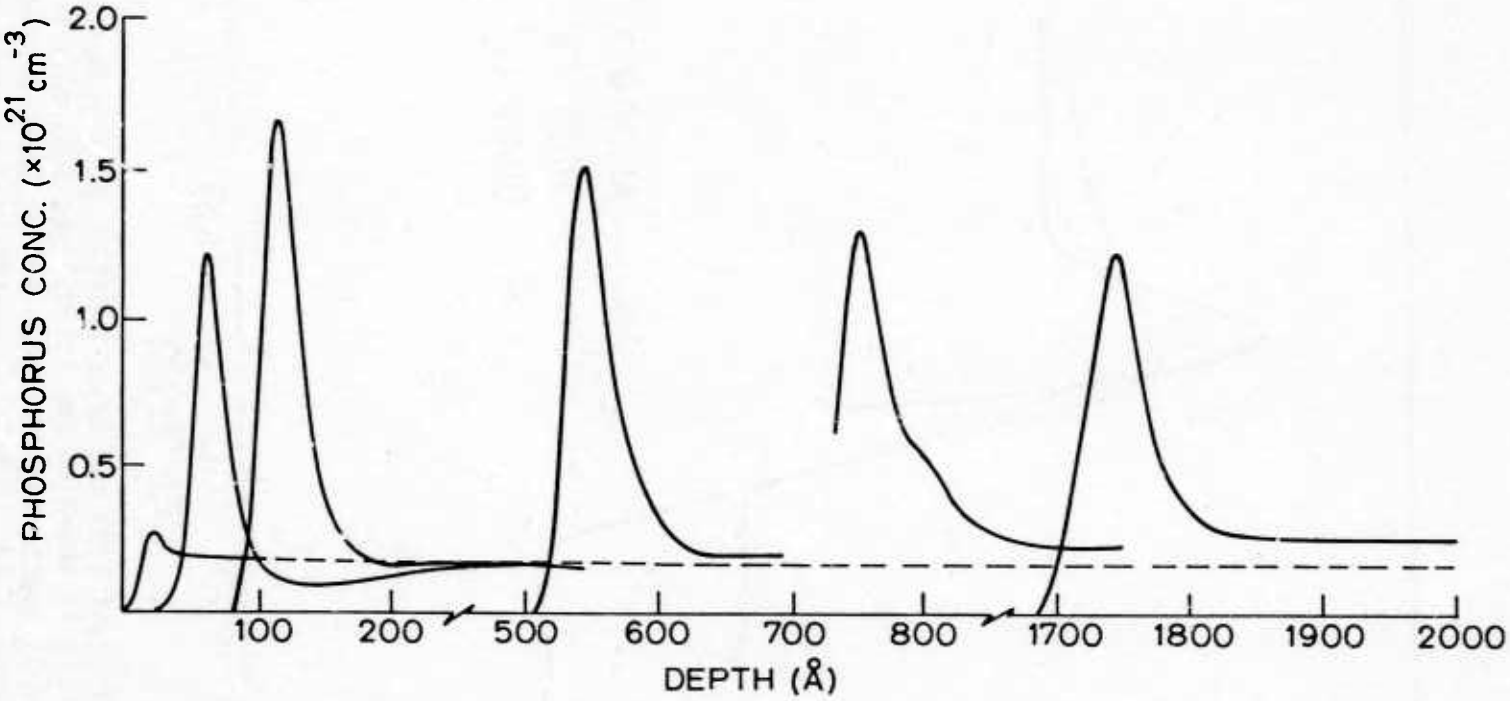


Figure 5.14 Evolution of phosphorus pileup for samples of bulk concentration 2×10^{20} cm^{-3} oxidized at 900°C. Depth is measured from the surface of the oxide in all cases. From right to left, each profile corresponds to oxidation times of 1, 10, 20, 90, 200, and 750 min, respectively, with oxide thicknesses of 20, 70, 120, 550, 750 and 1760 Å, respectively (see Table 5.3). The peak in phosphorus concentration occurs on the silicon side of the interface for each case.

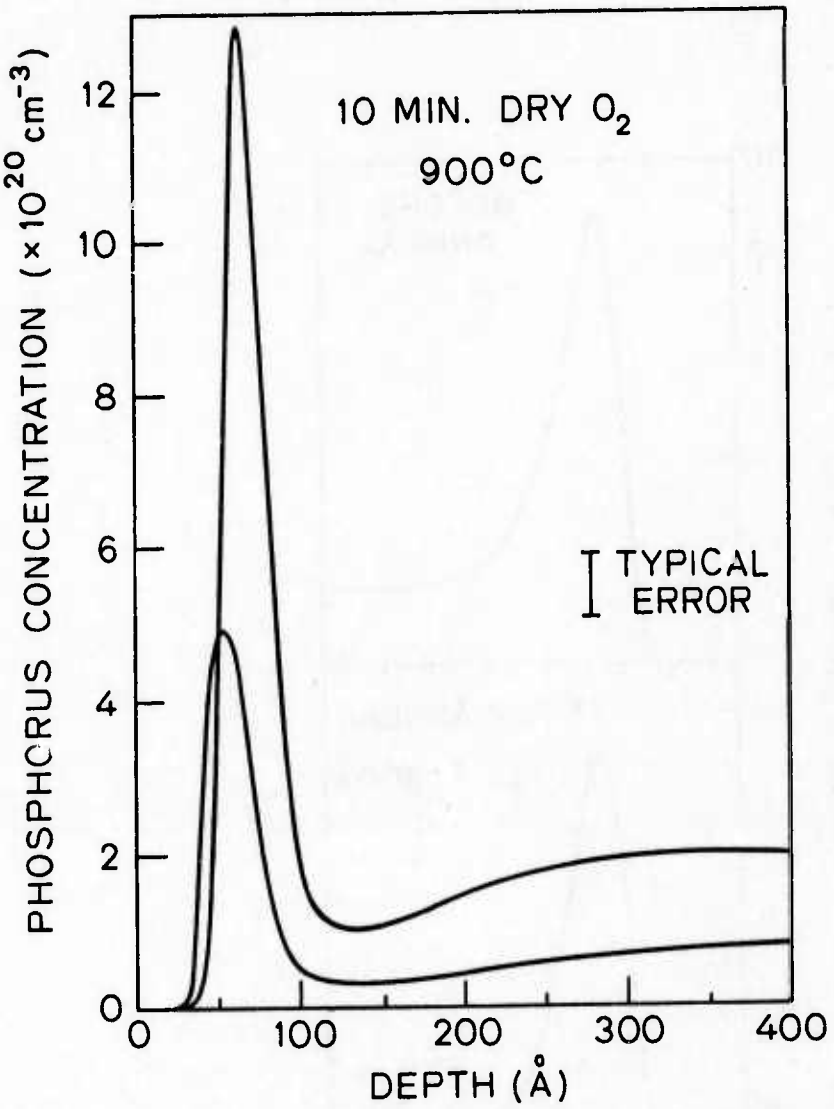


Figure 5.15 Illustration of the depletion of phosphorus that occurs in the near interface region for short oxidation times. Two curves are shown for samples of phosphorus concentration 2×10^{20} cm^{-3} and 8×10^{19} cm^{-3}, with oxides grown at 900°C for 10 min. (The 2×10^{20} curve is identical to the curve with peak second from the left in Fig. 5.14). An obvious depletion (dip) in the phosphorus is observed.

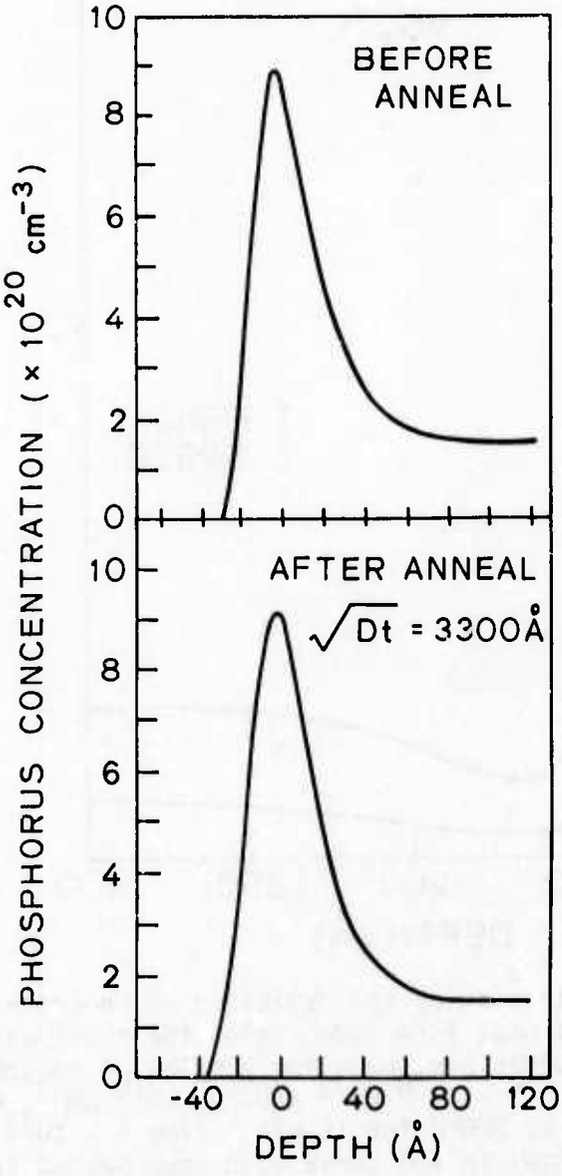


Figure 5.16 Illustration of the effect on annealing on the pileup effect for a sample of concentration 2×10^{20} cm^{-3} oxidized for 10 min at 1150°C in dry O_2 and annealed at the same temperature for 14 min. The profile before annealing is shown in (a); after the anneal is shown in (b). The annealing has had very little effect on the pile-up.

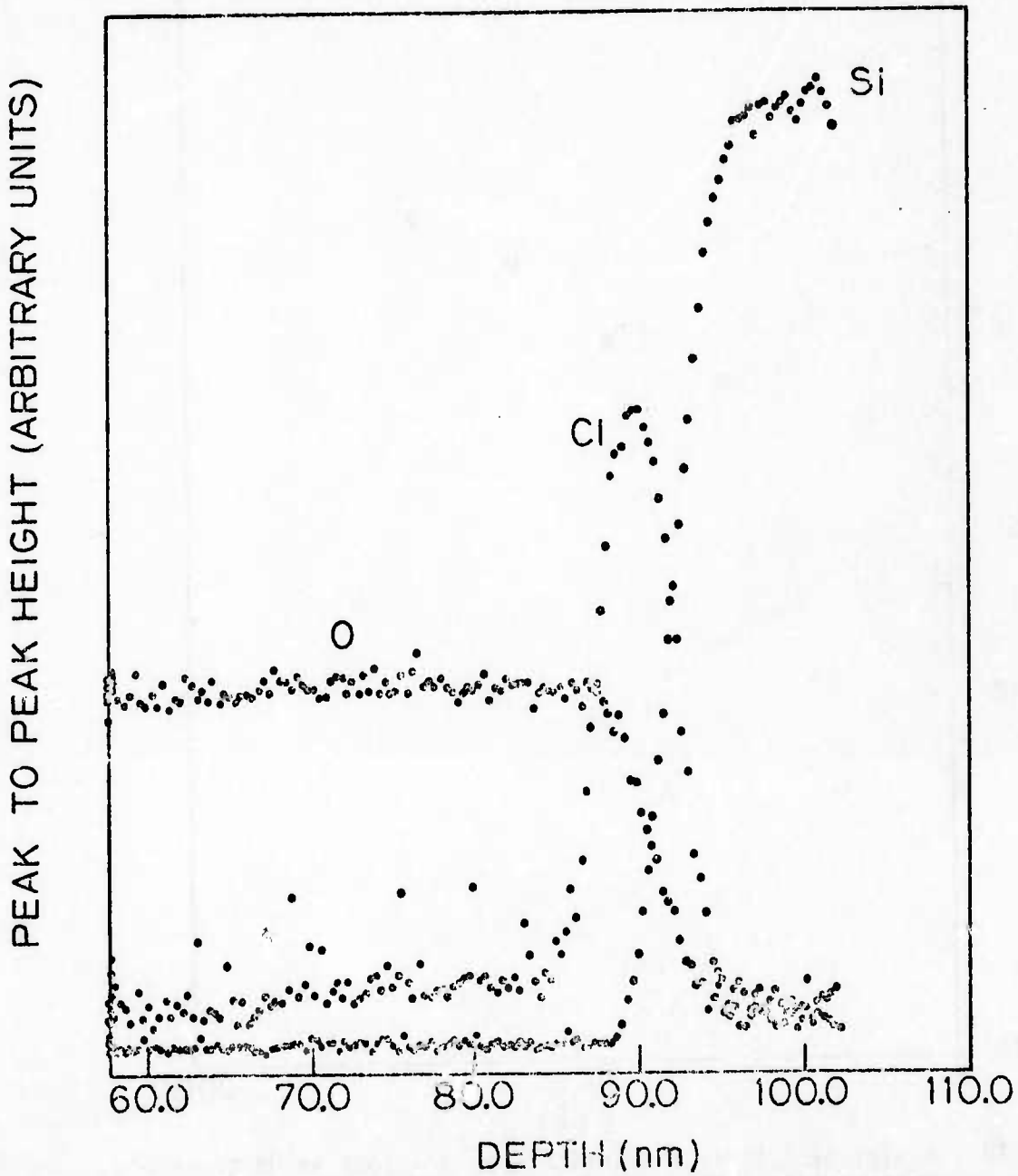


Figure 5.17 Auger sputter profile of SiO_2/Si. Peak height of the Auger signal is plotted as a function of depth into the sample. It is clear from this that there is a large amount of interface chlorine residing in the oxide.

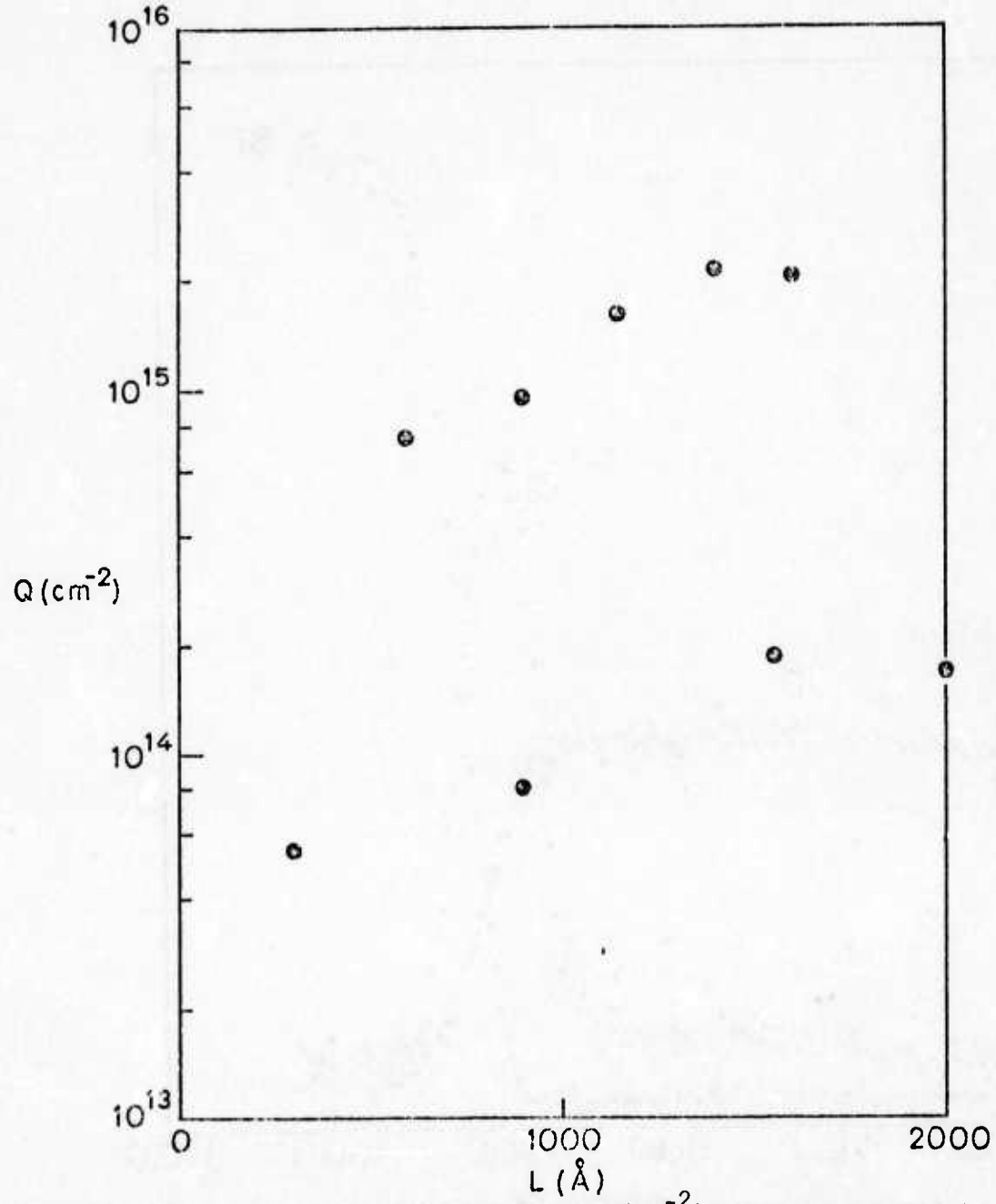


Figure 5.18 A plot of interface chlorine (cm^{-2}) versus oxide thickness. The upper set of points are for samples oxidized at 1100°C, the lower set for samples oxidized at 1000°C. Note the monotonic increase of interface chlorine with oxide thickness.

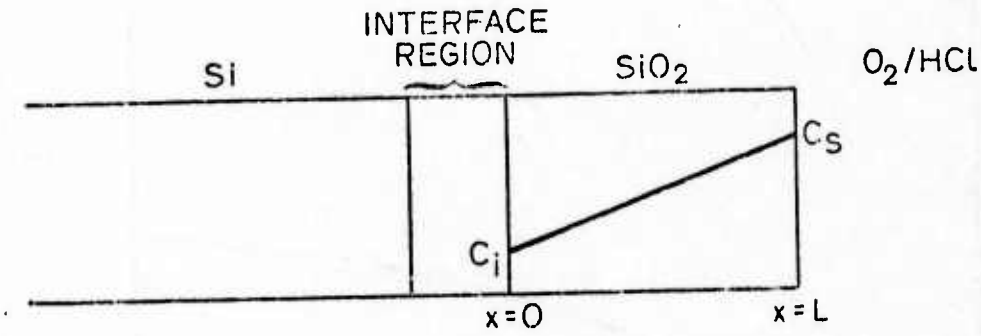


Figure 5.19 A plot of interface chlorine data of Figure 5.18 superimposed on solid curves predicted by the steady state model. One data point was used for each temperature to evaluate the DC_s product. The upper and lower curves are for 1100°C, 1000°C respectively.

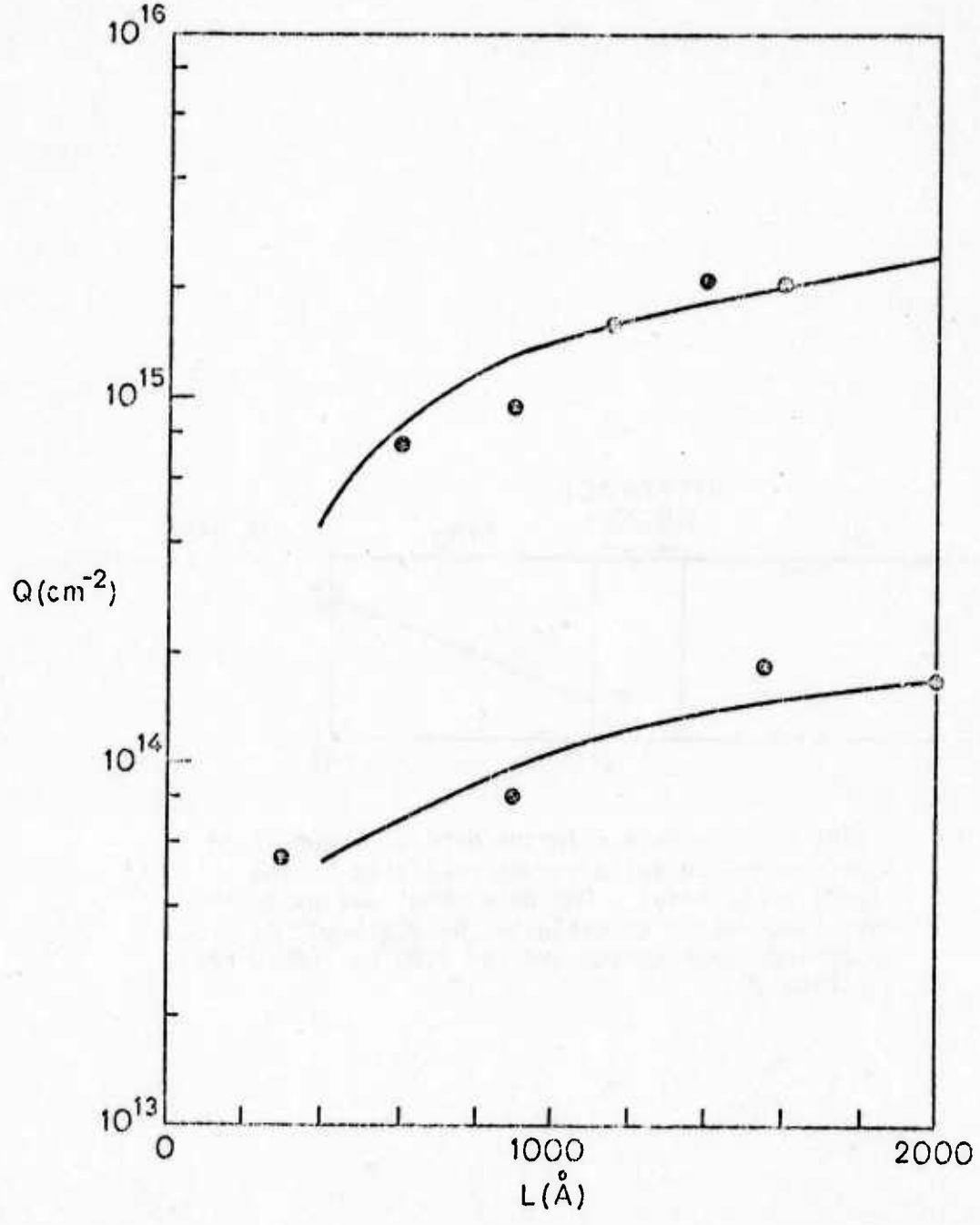


Figure 5.20 Illustration of the growing oxide layer and the interface region. The sample is broken into three phases, where the interface phase acts as a sink for chlorine. The steady state chlorine distribution in the oxide is also shown.

Part 6

COMPLETE PROCESS AND DEVICE SIMULATION

R. W. Dutton, J. A. Greenfield, H. G. Lee, S. Y. Oh,
C. Price, D. E. Ward

6.1 INTRODUCTION

Computer-aided engineering of semiconductor IC's has become a cornerstone of the VLSI era. At the outset of this research project the minimum device dimensions in production were greater than 3 μm whereas now submicron features have been demonstrated in many research laboratories. As minimum device dimensions shrink the need to understand the fabrication process in a quantitative way becomes critical. Thin oxide layers, polycrystalline silicon interconnections, shallow junctions and threshold implants each become more sensitive to process variations and statistical fluctuations. The development of the SUPREM program [6.1]–[6.2] and continued upgrading of the process models through the research discussed in preceding sections provides critical leverage in moving forward with technology scaling.

The overall goals of the research effort in computer-aided engineering to date have involved both the upgrading of process models in SUPREM and the coupling of the process simulation results to device and even circuit analysis capabilities. The major objectives of this research include:

(1) Development of Process Engineering Models which predict key

technological design parameters

(2) Development of methods to utilize geometrical and process information to specify simple device models including two-dimensional effects

(3) Development of circuit analysis capabilities which reflect statistical model effects including parameter correlations

(4) Dissemination of the results to industry.

The specific milestone accomplishments involving (1) have been discussed in the previous sections. In this section the results pertaining to the last three items are discussed. During the early phases of this research the technology spectrum being considered was broad and included bipolar, DMOS and CMOS as well as conventional NMOS technologies. In the present report the emphasis will focus primarily on the CMOS and NMOS technologies. However, highlights from previous efforts in other technologies will be summarized.

Major accomplishments can be identified in three main catagories—device modeling, analytical tools, and test structures. Highlight accomplishments from each of these areas are summarized. After enumerating the milestones on each topic a brief summary will be used to punctuate both the approach and the results which will be discussed subsequently.

A Device Modeling

Highlight accomplishments in this subsection include:

(1) A statistical bipolar model was developed including a methodology to determine correlation of all parameters [6.3]

(2) A two-lump bipolar model for high frequency circuit analysis was

developed based on numerical analysis of the emitter sidewall under high current conditions [6.4]

(3) A DMOS model was developed including velocity saturation and two-dimensional bulk impurity effects [6.5]. Two-dimensional device analysis using the CADDET program [6.6] was used to justify approximations and corroborate the results

(4) A CMOS latchup model has been developed based on a novel field-aided lateral pnp model as well as inclusion of all geometry dependent resistance, capacitance and transient radiation effects [6.7]

(5) A theory and method has been developed to extract subthreshold, punchthrough and linear region current-voltage characteristics for MOS devices using only two-dimensional Poisson solutions [6.10]

(6) A new MOS capacitance model was developed which includes the non-reciprocal time dependent effects of channel charge on gate capacitance [6.8]–[6.9].

Without exception, all model results presented have a clear and explicit dependence on technology variables and especially impurity profiles as determined from SUPREM. It should be emphasized that the model development activities were pursued to compliment the development of device simulators. In every instance it is found that at the performance limits—both for MOS and bipolar—two-dimensional device analysis is required in model formulation. Moreover, while the model equations can most often take a simple and physical form, the determination of coefficients frequently requires the use of device analysis.

The statistical and high frequency bipolar models have been formulated in terms of the familiar Gummel-Poon [6.11] integral charge relationship where the so-called Gummel number is given by:

$$GU_B = \int_{\text{Active Base}} \frac{N_A(x)}{D_n}\,dx \tag{6.1}$$

where N_A is the base acceptor concentration and D_n is the electron diffusivity—both are position dependent. The well-known dependence of bipolar parameters on GU_B leads naturally to the desired coupling of process specifications and device parameters. Moreover, through this research the basis has been established for modeling statistical [6.3] as well as high frequency effects [6.4] while maintaining a tight coupling to technology specifications and parameters such as GU_B.

For the DMOS work it was found that the two-dimensional nature of the bulk charge required that a space-averaged bulk charge must be computed in determining threshold as given by the following expression:

$$Q_{BO} = \frac{q}{L_E}\int_0^{L_E} N_e(x)W(x)\,dx \tag{6.2}$$

where L_E is the electrical channel length of the diffused region, N_e is an exponential approximation to the doping profile along the channel and W is the depletion width perpendicular to the channel direction. This equation form contains both the subsurface and lateral dependences of channel doping for computation of bulk charge. The DMOS model illustrates a typical condition in scaled devices where strongly two-dimensional effects make it increasingly difficult to formulate tractable analytical forms for models. The

numerical results shown in subsequent sections illustrate the capabilities of parameter extraction directly from computer simulations.

In the most recent results with dynamic charge effects in MOS devices [6.9] it has been determined that substrate doping effects are substantial in determining the minimum switching transit time delays. Previous workers have predicted a form for the delay given by [6.12]–[6.13]:

$$\tau_D = \sqrt{t_r \tau_o} \tag{6.3}$$

where t_r is the rise time of the gate signal and τ_o is the classical transit time expression given by $\tau_o = L^2/\mu_n V_{GSmax}$. As will be discussed shortly, this expression while qualitatively correct, substantially underestimates the delay since the bulk charge effect is neglected. In all of the modeling efforts discussed above there has been a concerted thrust to use SUPREM and enhance its further development. Moreover, the device simulation capabilities have been essential in understanding the two-dimensional manifestations of process modeling in the context of actual device structures.

B Analysis Tools

Highlight accomplishments in this subsection include:

(1) A statistical circuit analysis capability was developed including the means to handle model parameter correlations [6.3]

(2) A one-dimensional SEmiconductor DEvice ANalysis program (SEDAN) was developed [6.14] to couple directly with process modeling inputs from SUPREM

(3) The SEDAN program was used in conjunction with other device simulators to investigate two-dimensional MOS [6.5] and bipolar

[6.4] device effects

(4) A Two-dimensional ANalysis program for DEvice Modeling (TANDEM) has been developed to solve Poisson's equation and to calculate limited information concerning capacitance and low-level current flow [6.10]

(5) An experimental quasi two-dimensional MOS device analysis program has been developed using a boundary value (BV) mapping technique for the Poisson equation, coupled with a one-dimensional current transport solution [6.15]–[6.16]

(6) Demonstration of the applicability of the quasi two-dimensional analysis to determine capacitance and transient charge effects in MOS devices [6.9]

A dominant objective in developing device simulators such as SEDAN and TANDEM has been to use computed technology parameters in the most efficient way to predict device performance. For example, the SEDAN program accepts SUPREM files as input and generates output parameters such as GU_B and capacitance versus voltage as well as current versus voltage. As suggested earlier, two-dimensional effects tend to dominate high performance devices so that SEDAN is at present suited only for large geometry test structure work and as a vehicle to study process sensitivities of device parameters in a parametric way—that is, trends are shown but absolute values cannot be extracted because of small geometry (two-dimensional) effects.

In the two-dimensional device analysis area, the TANDEM program now provides a solid coupling of technology and surface topology information with solution of Poisson's equation in two dimensions. TANDEM uses SUPREM files directly as a means to input technology information. However,

it is necessary to convolve the one-dimensional SUPREM files with simple physical approximations for two-dimensional diffusion and local oxidation [6.18]–[6.19]. While these approximations limit the physical accuracy of two-dimensional technology modeling at present, the understanding of the process kinetics is still under investigation so that more complex models do not seem justifiable at this early stage. Results shown in the next sections indicate clearly the trends in process sensitivity of electrical device parameters based on measured and inferred physical constants. It is expected that as two-dimensional process modeling advances that a more direct profile and surface topology input for TANDEM will evolve. Beyond the efforts with TANDEM there have been two experimental programs developed each with a unique capability relative to the overall thrust in coupling process, device, and circuit modeling. The boundary value (BV) method for simplifying the two-dimensional analysis and substantially reducing computation time has lead to demonstration of quick curve-tracer type data generation as well as time-domain use of the method to investigate dynamic charge effects. The time domain applications are discussed in the next section. The focus of the BV method is directed towards the circuit applications end of the device modeling spectrum. The other experimental program development reported here is the status of TANDEM II, a one-carrier transport analysis capability. The emphasis of TANDEM II (called TANDII for convenience) is to provide a means to carefully investigate the effects of grid and non-planar surfaces on both carrier flow and impurity distributions in MOS devices. To date we have used the CADDET program extensively in device modeling work, however the benchmarks with TANDII show that grid plays a major role in solution accuracy and an in-house experimental program was essential for further investigations.

C Test structures

Highlights accomplishments in this subsection include:

(1) A CMOS test structure has been developed to characterize the lateral surface impurity diffused profile of p^+ channel stop and p-well regions [6.14]

(2) Further extensions of the structure for lateral surface impurity profile measurements have resulted in structures suitable for scaled NMOS [6.15] and DMOS [6.16]

(3) A set of CMOS latchup test structures have been developed including means to characterize all parameters contained in the latchup model [6.17].

It was realized from the outset of this project that many two-dimensional device effects depend critically on layout geometries—both macroscopically, as for latchup, and microscopically, as for specific technology test structures such as the lateral surface profile devices reported here [6.17]–[6.19]. The test structure work reported previously has emphasized the local layout effects associated with latchup. On the impurity profile level, the edge profile measurement test structures have been carefully confined to the specific MOS technologies under investigation. However, in the present report a broader set of conclusions are drawn concerning the utility of the test structure approach for impurity profile measurement.

The CMOS latchup test structure work as well as layout checking of general logic functions is a topic of increasing importance as greater performance is required at decreasing levels of power. The CMOS test structure work [6.17] and pnp model formulation [6.7] have contributed to the art of understanding and predicting the role of the base-emitter shunting

resistors on the susceptibility and holding conditions for latchup. In addition the calculation of component values as well as technological means to alter the values—including device scaling—have been considered. It is found that as surface dimensions are scaled-down the effect is to reduce the value of shunting resistance—and hence the latchup susceptibility—faster than any simultaneous increase in either npn or pnp current gain owing to changes in vertical dimensions of doping profiles [6.7]. The overall complexity of the analysis problem involving CMOS latchup has to date only been amenable to hand calculations of resistors based on layout geometries and first-order computations of bipolar device parameters using the STRAP program [6.7]. It is clear that further work is needed to carry the basic analysis methodology and computation techniques to a useable general purpose engineering tool.

The test structure area is a fertile one in need of substantial research efforts. On the one hand the packing density requirements of VLSI necessitate a better understanding of nearest-neighbor interactions including latchup. Moreover, even the seemingly straightforward tasks of parameter extraction on scaled-down devices now entail essentially test structure approaches to isolate the fundamental technology variables.

6.2 DEVICE MODELING

The emergence of very large scale integration has introduced small-geometry devices which violate many common assumptions used in device analysis. For example, the one-dimensional calculation of threshold voltages and the simple calculation of parasitic capacitances from mask layout information are often not sufficiently accurate. The analysis and understanding of two- and three-dimensional effects are essential to the design of modern integrated circuits. In the discussion which follows both numerical and analytical approaches to device characterization are presented. The first subsection describes two-dimensional analysis results obtained using solutions of Poisson's equation. In addition, simple model forms are presented which give an accurate representation of low level current in MOSFET's based only on parameters obtained from the Poisson solutions. The basic features of the analysis program TANDEM [6.25] are summarized—further details are contained in a separate document [6.25]—as well as further applications of the analysis in extracting parameters for oxide capacitance. The second subsection presents a detailed summary of the charge-oriented capacitance model. One particular new aspect of the capacitance model results in this report is the facility to use numerical results to obtain the needed channel charge parameters for the model. At present the capacitance model has limitations which limit it to cases where the quasi-static approximation applies.

6.2.1 Device Modeling using the Two-Dimensional Solution of Poisson's Equation (James A. Greenfield)

The work described in this section has been aimed at developing a complementary set of computer tools and theoretical techniques for analyzing small-geometry devices. The approach taken is to solve the electrostatic Poisson equation and infer many electrical device characteristics from the resulting potential distributions. Elimination of the simultaneous solution of the current-continuity equations significantly reduces the computational expense and complexity of the analysis. Emphasis is placed on maintaining close coupling with fabrication technology by accounting for the characteristics of realistic device structures.

The electrostatic solution of Poisson's equation alone provides the capability for analyzing the following important device characteristics:

linear region current

low level surface and bulk current

threshold voltage

potential barrier lowering effects

electrode capacitance

charge storage effects

These applications are particularly useful for the designer attempting to optimize a device to meet specific system requirements. The analysis also provides a means for evaluating the sensitivity of electrical characteristics to fabrication process variations.

The calculation of linear region and low level currents from the electrostatic solution of Poisson's equation relies on features of the device operation which allow the use of certain approximations. In the linear region of a MOSFET the drain-source bias V_{DS} is small (i.e. $V_{DS} < 0.1$ V) and the drain current I_{DS} is proportional to V_{DS} and the constant drain-source

conductance G_{DS}

$$I_{DS} = G_{DS} V_{DS} \tag{6.4}$$

G_{DS} is independent of V_{DS} because the current results from minority-carrier driftnn in the presence of the constant longitudinal electric field induced by V_{DS}. This allows G_{DS} to be calculated from the minority-carrier concentration provided by a solution of Poisson's equation with $V_{DS} = 0$. For an n-channel device the expression for the drain current is [6.10]

$$I_{DS} = qW\left[\int_0^L \left(\int_{y_1}^{y_2} \mu_n(x,y) n(x,y)\, dy\right)^{-1} dx\right]^{-1} V_{DS} \tag{6.5}$$

where q is the electronic charge, W is the effective device width, L is the electrical device length, and $\mu_n(x,y)$ is the electron mobility. $n(x,y)$ is the electron concentration obtained by solving Poisson's equation.

The calculation of low-level currents is based on the analysis of carrier transport over a potential barrier. In a MOSFET the surface or bulk current path between the source and drain contains a point P where the longitudinal component of electric field vanishes. The current is extracted from the electrostatic potential distribution near P, obtained by solving Poisson's equation and neglecting the minority carrier charge density responsible for the current. This approximation remains reasonably valid even when the neglected minority-carrier concentration near P exceeds the ionized impurity concentration because the minority carriers are localized near the current path.

Point P can be defined as the origin of a coordinate system (r, θ) with the r axis oriented along the current path and the θ axis oriented in the

orthogonal direction at P. This is illustrated in Fig. 6.1 for a bulk current path. For an n-channel device the drain current is given by [6.10]

$$I_{DS} = -qD_n W \frac{Z^*}{L^*} \frac{n_i^2}{N_B} \exp\left(\frac{q(\psi^* - V_S)}{kT}\right)\left[1 - \exp\left(-\frac{qV_{DS}}{kT}\right)\right] \qquad (6.6)$$

where D_n is the electron diffusion coefficient at P, W is the device width, n_i is the intrinsic carrier density, N_B is the bulk impurity concentration, ψ^* is the potential at P, and V_S is the source bias. The zero reference for ψ^* is defined by setting ψ equal to the substrate bias in the bulk region. Z^* is the effective width and L^* is the effective length of the region of the potential near P controlling the current.

Although Z^* and L^* can be defined rigorously [6.10], their approximate determination is illustrated in Fig. 6.1 for a bulk current path. L^* is the distance along the r axis between points A and B where $\psi = \psi^* + \pi kT/4q$ and Z^* is the distance along the θ axis between points C and D where $\psi = \psi^* - \pi kT/4q$. For a surface current path the definition of L^* remains the same and $Z^* = kT/q|E_s|$, where E_s is the vertical component of the electric field at P.

The technique described by (6.6) for calculating low-level current is approximate because the minority carriers are neglected in the channel region during the solution of Poisson's equation. The accuracy of this technique has been evaluated by performing comparisons with full two-dimensional simulations [6.10]. The comparisons were obtained by using the CADDET program [6.6], which simultaneously solves the Poisson and minority-carrier current-continuity equations.

The accuracy and applicability of the calculations described by (6.5) and (6.6) have been demonstrated by comparing simulated results with

various experimental measurements for short-channel MOSFETs [6.18]. Some important parameters for one of these n-channel devices are listed in Table 6.1. The effective channel length, oxide thickness, and surface mobility were determined experimentally from electrical measurements. The implanted channel region impurity profile was calculated by the process simulation program SUPREM [6.19].

Fig. 6.2 compares the simulated and measured subthreshold I_{DS} vs. V_{GS} for $V_{DS} = 0.1$ and $V_{BS} = 0, -2$. Good agreement is obtained over six current decades. The slopes of the simulated and measured data in the figure agree within several percent. The current was calculated using (6.5), although the low current portion could also be calculated with (6.6).

Punchthrough is induced in this device by increasing the reverse bias on the drain-substrate junction. The impurity concentration in the channel is sufficiently low that the current path occurs at the silicon surface. Fig. 6.3 compares the simulated and measured I_{DS} vs. V_{DS} for $V_{GS} = 0$ and $V_{BS} = 0, -2$. The current was calculated using (6.6) for a surface current path.

The utility of the described analysis techniques in the context of fabrication process design is illustrated by an example of the calculation of bulk punchthrough current. Two technological solutions to the punchthrough current problem for 1μm devices have been suggested. The use of a double implantation with a deep implant for threshold control was proposed by several workers [6.20]–[6.22]. This deep implantation should extend the full junction depth into the substrate, while not extending below the n^+ junctions such as to raise the n^+p junction capacitance. In addition, the bulk charge term added by this deep implant should not change with increasing V_{BS} more rapidly than the bulk charge due to the high resistivity substrate.

The use of one implantation in which the peak doping level is not at the surface and the tail is used to adjust the bulk punchthrough has also been proposed [6.23],[6.24]. Again, the proper placement of the implant is needed to minimize the n^+p junction capacitance and substrate sensitivity.

The expression for low-level current (6.6) was used to study the punchthrough properties of devices fabricated with the following enhancement implantation doses and energies

a) 5×10^{11} cm^{-2} @ 60 KeV

b) 5×10^{11} cm^{-2} @ 100 KeV

c) 3×10^{11} cm^{-2} @ 25 KeV
   2×10^{11} cm^{-2} @ 170 KeV

d) 5×10^{11} cm^{-2} @ 25 KeV

The process sequence was simulated for a typical n-channel Si-gate single-level polysilicon process, with the resulting impurity profiles shown in Fig. 6.4. Every case except the shallowest implant raises the impurity concentration above the bulk value (10^{15} cm^{-3}) for at least 0.9 μm into the substrate.

The punchthrough characteristics have been computed for the four profiles using junction depths of 0.3 and 0.5 μm with an effective channel length of 1.0 μm. Although the differences in the channel impurity profiles of Fig. 6.4 may seem slight, they have a significant effect on the punchthrough characteristics. For a junction depth of 0.5 μm, Fig. 6.5 demonstrates that the increase in implantation energy increases the punchthrough voltage and that the double implantation gives results close to the 100 KeV case. Fig. 6.6 shows that decreasing the junction depth to 0.3 μm gives greatly reduced current. These devices have curves shifted to higher V_{DS} by up to 7.5 V for the 60 KeV energy implant. The double implantation produces results

between the 60 and 100 KeV single implantations, as was the case for a junction depth of 0.5 μm.

The charge storage characteristics of a device are calculated by solving Poisson's equation in the presence of electron and hole charge. The charge concentrations vary nonlinearly with the potential and must be determined self-consistently with the potential distribution during the solution. The determination of the charge distribution within a device is of major use in the design of small geometry memory cells. A highly two-dimensional device structure can cause significant reduction in the charge storage capacity and consequent lowering of signal noise margins.

Electrode capacitance can be calculated from the variation of induced electrode charge caused by a change in electrode bias. Two solutions of Poisson's equation are performed for a small difference in the bias of one electrode. The change in the total charge induced on a second electrode is then calculated by integrating the component of electric field normal to the electrode.

$$C = \frac{\Delta\left(\int \epsilon \vec{E} \cdot \hat{n}\, dl\right)}{\Delta V} \tag{6.7}$$

The capacitance of the second electrode is the ratio of the change in induced charge over the potential change. This type of calculation can be of particular importance for devices where a highly two-dimensional structure makes estimation of the capacitance through simple approximations impossible. Junction capacitance could be determined by extending the analysis techniques to include the calculation of total depletion region space charge concentrations.

The TANDEM (Two-dimensional ANalysis for DEvice Modeling) computer program [6.25] was developed to provide a practical means for analyzing two-dimensional device structures. It is written in FORTRAN and implemented on a Hewlett-Packard 1000 series minicomputer running an RTE-IV operating system. The program handles standard enhancement and depletion mode field-effect transistors (FET's), DMOS FET's, and devices fabricated on insulating substrates (e.g. silicon-on-sapphire (SOS)).

TANDEM accurately models nonplanar surfaces and insulator-semiconductor interfaces. This allows the treatment of modern small-geometry devices fabricated with local oxidation and layers having thicknesses comparable in size to their lateral dimensions. The analysis of nonplanar structures is important for the accurate calculation of electrode capacitance, charge storage, and other electrical parameters.

The accuracy of the representation of the substrate impurity distribution in a device can have drastic effects when the electrical performance of a small-geometry device is simulated. TANDEM addresses this problem by allowing the specification of vertical impurity profiles through analytic functions or output data generated by the SUPREM computer program [6.19]. The impurity profiles can be extended into their two-dimensional form by the use of suitable functional products based on theoretical results [6.16]. The use of numerically calculated substrate impurity distributions provides a necessary coupling between the fabrication process and the final electrical device performance.

The input structure of TANDEM is keyword-driven in a manner similar to the input of SUPREM. However, the input interpretation has undergone significant improvement by providing for more flexibility in naming keywords and better reporting of syntactic errors. The user has the

ability to easily provide aliases for keywords and change parameter default values. The aliasing feature can be convenient when a user has specialized nonstandard terminology for some physical characteristics of a device.

The numerical techniques used by TANDEM to solve Poisson's equation have been described elsewhere [6.10],[6.26]. A one-step Newton-SLOR (successive line overrelaxation) iteration is used to solve a finite difference approximation to Poisson's equation. The discretization of the continuous equation is performed on a rectangular grid which is nonuniform in both spatial directions.

TANDEM provides extensive graphical output and parameter extraction capabilities. Contour plots can be generated for potential, electric field magnitude, impurity concentration, electron concentration, and hole concentration within the two-dimensional device cross section. The plots can also include the locations of junctions and depletion edges. One-dimensional plots can be displayed for the above quantities versus distance along arbitrary straight paths through the device. These plots can also be generated along the current path. The calculations described by (6.5) and (6.6) are implemented to couple the program with the theoretical analysis for extracting device currents.

6.2.2 A Charge-Oriented Capacitance Model for MOS Transistors (D. E. Ward)

Transient effects in MOS transistors can be simulated quite accurately using one-dimensional [6.12], [6.27], quasi-two-dimensional [6.28], or fully two-dimensional [6.29] numerical analysis. Circuit simulation, however, requires a model which is computationally fast—usually at the expense of model accuracy. Small-signal models [6.30]–[6.31], although valid to very high frequencies, are not applicable to large-signal analysis. Thus circuit simulators to date have used the gate-capacitance approach of Meyer and Klassen [6.32]–[6.33].

Previous reports have shown that the shortcomings of the conventional gate-capacitance models for MOS devices and circuits can be overcome by an improved numerical technique [6.8], [6.34]. This report will discuss the determination of the charges used in the numerical method and the characteristics and limitations of the resulting model.

A. The Method

Both the charge-oriented and gate-capacitance models are derived from the basic equation relating gate current i_G to gate charge Q_G:

$$i_G = \frac{dQ_G}{dt} \tag{6.8}$$

In the Meyer approach [6.32], [6.33] the right-hand side of (6.8) is expanded as

$$\frac{dQ_G}{dt} = C_{GD}\frac{d(V_G - V_D)}{dt} + C_{GS}\frac{d(V_G - V_S)}{dt} + C_{GB}\frac{d(V_G - V_B)}{dt} \tag{6.9}$$

where

$$C_{GD} \equiv -\frac{\partial Q_G}{\partial V_D} \qquad C_{GS} \equiv -\frac{\partial Q_G}{\partial V_S} \qquad C_{GB} \equiv -\frac{\partial Q_G}{\partial V_B}$$

This suggests the model of Fig. 6.7. The gate current is the sum of the currents through three (non-linear) capacitors, each current being given by

$$i = C\frac{dV}{dt}\,. \tag{6.10}$$

The circuit simulator program integrates this equation over one timestep to give

$$\int_{t_0}^{t_1} i\,dt = \int_{V_0}^{V_1} C\,dV = \overline{C}(V_1 - V_0) \tag{6.11}$$

where $\overline{C}$ is the capacitance averaged between voltages V_0 and V_1 at times t_0 and t_1 respectively. $\overline{C}$ is unknown and must be approximated. This approximation can lead to numerical errors (non-conservation of charge), slow convergence to a solution, and numerical instability. All capacitive currents are obtained from the three capacitors derived from the gate charge, thus implying a reciprocity which does not exist, e.g. that $\frac{i_G}{dV_D/dt} = \frac{i_D}{dV_G/dt}$, and omitting capacitive effects between bulk and drain, bulk and source, and drain and source.

The charge-oriented model works directly with (6.8). This is integrated by the circuit simulator to give

$$\int_{t_0}^{t_1} i_G\,dt = Q_{G1} - Q_{G0} \tag{6.12}$$

The right-hand side is evaluated by linearizing Q_{G1} about the operating point obtained at the most recent iteration. Thus at convergence the right-hand side of (6.12) is given exactly while that of (6.11) is only approximate.

Equations analogous to (6.8) and (6.12) 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 (6.12) provides for numerical speed and stability.

The gate and bulk charges used in the charge-oriented numerical method are physically distinct and may be obtained from the steady-state device equations. The source and drain charges, however, merge to form the channel charge and are not readily separated. It is possible, in fact, that unique source and drain charges do not exist which satisfy the conditions

$$i_S = \frac{dQ_S}{dt}\,, \quad i_D = \frac{dQ_D}{dt}\,, \quad Q_S + Q_D = Q_C = -(Q_G + Q_B)\,. \tag{6.13}$$

(Here i_S and i_D are the transient components of source and drain current and Q_C is the total channel charge). In order to investigate this possibility and to evaluate the charges required by the model, an analytical approach to the transient response of MOS transistors has been developed.

B. An Analytical Approach

In order to simplify the analysis of MOS transient behavior, it is assumed that channel current is one-dimensional (parallel to the surface) and is due to a single carrier type (electrons in an n-channel device). Thus the current transport equation may be written

$$I(y,t) = qW\mu(y,t)n(y,t)\frac{\partial\phi(y,t)}{\partial y} \tag{6.14}$$

where I is the channel current flowing in the minus y direction (from drain to source), n is the mobile carrier density in the channel, ϕ is the electron

quasi-Fermi potential at the silicon surface, and μ is the effective carrier mobility. W is the width of the device and q is the electron charge. The current continuity equation (neglecting recombination) is

$$\frac{\partial n(y,t)}{\partial t} = -\frac{1}{qW}\frac{\partial I(y,t)}{\partial y} \,. \tag{6.15}$$

Integrating (6.15) from source ($y = 0$) to position y along the channel and substituting $I(y,t)$ from (6.14) gives

$$I(0,t) - qW\mu(y,t)n(y,t)\frac{\partial \phi(y,t)}{\partial y} = qW\int_0^y \frac{\partial n(y',t)}{\partial t}\,dy' \,. \tag{6.16}$$

Integrating again from source to drain ($y = L$) and solving for $I(0)$ gives

$$I(0,t) = q\frac{W}{L}\int_0^L \mu(y,t)n(y,t)\frac{\partial \phi(y,t)}{\partial y}\,dy + q\frac{W}{L}\int_0^L\int_0^y \frac{\partial n(y',t)}{\partial t}\,dy'\,dy \,. \tag{6.17}$$

The second term on the right may be reduced by an integration by parts and by taking the derivative outside of the integral. The source current (positive flowing into the device) is then

$$I_S(t) \equiv -I(0,t) = -q\frac{W}{L}\int_0^L \mu(y,t)n(y,t)\frac{\partial \phi(y,t)}{\partial y}\,dy + \frac{d}{dt}\left\{-qW\int_0^L \left(1-\frac{y}{L}\right)n(y,t)\,dy\right\} \,. \tag{6.18}$$

The drain current is found by integrating (6.15) from source to drain and substituting $I(0,t)$ from (6.18):

$$I_D(t) \equiv I(L,t) = q\frac{W}{L}\int_0^L \mu(y,t)n(y,t)\frac{\partial \phi(y,t)}{\partial y}\,dy + \frac{d}{dt}\left\{-qW\int_0^L \frac{y}{L}n(y,t)\,dy\right\} \,. \tag{6.19}$$

For long-channel devices Poisson's equation may be written as a one-dimensional equation perpendicular to the surface. It is assumed that the electric field parallel to the surface is negligible (the gradual channel approximation) and that the device is uniform in cross-section between the source and the drain. This implies that the mobile carrier density may be written

$$n(y,t) = n\big(\phi(y,t), V_G(t)\big) = n(\phi, V_G)\,. \tag{6.20}$$

The time and space dependences of n have been replaced by dependences on the quasi-Fermi potential $\phi(y,t)$ and the gate voltage $V_G(t)$. The effective mobility is similarly assumed to be determined by ϕ and V_G. Thus for long-channel devices the first term of (6.18) and (6.19) becomes

$$q\frac{W}{L}\int_0^L \mu(\phi,V_G)n(\phi,V_G)\frac{\partial\phi(y,t)}{\partial y}\,dy = q\frac{W}{L}\int_{V_S}^{V_D}\mu(\phi,V_G)n(\phi,V_G)\,d\phi \equiv I_0\,. \tag{6.21}$$

The value of this term depends only on the instantaneous values of the terminal voltages. The source and drain currents are then

$$I_S = -I_0(V_G,V_D,V_S,V_B) + \frac{dQ_S}{dt} \tag{6.22a}$$

$$I_D = +I_0(V_G,V_D,V_S,V_B) + \frac{dQ_D}{dt} \tag{6.22b}$$

where

$$Q_S = -qW\int_0^L \Big(1-\frac{y}{L}\Big)n(y,t)\,dy \tag{6.23a}$$

$$Q_D = -qW\int_0^L \frac{y}{L}n(y,t)\,dy \tag{6.23b}$$

The first term on the right of (6.22) is clearly the DC transport current while the second term is the transient charging component. Thus the charges given

by (6.23) satisfy (6.13) and may be used as the source and drain charges in the charge-oriented model. Note that Q_S and Q_D are found by applying a linear partition function ($1 - \frac{y}{L}$ or $\frac{y}{L}$) to the total channel charge.

C. The Closed-Form Simulation Model

A closed-form model suitable for use in a circuit simulator may be obtained by finding $n(y,t)$ and performing the integrals of (6.23). In strong inversion the mobile carrier density may be approximated as

$$n(\phi, V_G) = \frac{C_{ox}}{q}\left[V_G(t) - V_B(t) - V_{FB} - \phi(y,t) - 2\phi_F - \gamma\sqrt{\phi(y,t) + 2\phi_F}\right] \tag{6.24}$$

where $\phi(y,t)$ is the quasi-Fermi potential along the channel, C_{ox} is the gate oxide capacitance, V_{FB} is the flatband voltage, and $\gamma = \sqrt{2q\epsilon_{Si}N_A}/C_{ox}$ is the body-effect coefficient. N_A is the effective substrate doping and ϕ_F is the bulk Fermi level. This equation may be substituted into (6.16) giving a partial differential equation for $\phi(y,t)$. The time derivative is assumed to be small, implying that the carrier density is given by the steady-state distribution determined by the terminal voltages. This is known as the quasi-static approximation. A closed-form relationship between y and $\phi(y,t)$ is thus obtained which, along with (6.24), may be used in (6.23) to obtain the source and drain charges.

The solutions for source and drain charge are quite complicated; simplified forms are desirable for computer implementation. One approximation (used for example in SPICE [6.35], [6.36]) is to partition the channel charge into equal source and drain components. This gives correct capacitances in the linear region of operation but not in saturation. Better approximations have recently been developed.

D. Accuracy and Limitations of the Charge-Oriented Model

Effective use of any simulation tool requires an understanding of its accuracy and limitations. In the case of the charge-oriented capacitance model these arise from two sources, the numerical method and the modeling of charge. It will be noted that the limitations of the charge-oriented model are generally parallel to those of the Meyer model, but are less restrictive in several significant aspects.

The numerical accuracy of the charge-oriented model has been mentioned previously. The main remaining source of error (present even in simulation of linear circuits) is in the integration of the current, and may be minimized by control of the simulation timestep. The error associated with integration of capacitance has been eliminated.

The capacitance characteristics predicted by the model have been verified by direct measurements on long-channel devices [6.34]. Typical results are shown in Fig. 6.8. (The rounding of the characteristics at low gate voltages is due to the use of a high measurement frequency (1 MHz) on a long-channel (100 μm) device). Note that the two capacitances $C_{GD} \equiv \frac{\partial Q_G}{\partial V_D}$ and $C_{DG} \equiv \frac{\partial Q_D}{\partial V_G}$, assumed identical in Meyer's approach, are not the same. The charge-oriented model is more complete in that it includes source-bulk, drain-bulk, and source-drain capacitances, non-reciprocal effects, and more detailed modeling of bulk doping effects.

The principal limitations of the model are associated with the combination of circuit speed and channel length constraints. The speed limitation stems from the use of the quasi-static approximation and is strongly dependent on circuit features, especially the shape of the input waveforms. The amount of error which can be tolerated is likewise dependent on the ap-

plication. In general, however, in digital circuits the quasi-static approximation may be considered to be valid if the input rise and fall times are large compared to the channel transit time, given by $\tau_0 = L^2/\mu V_{GSmax}$. If "large" is taken to mean a factor of 20, then the speed limitations shown in Table 6.2 result. (For the 1 μm device the transit time has been defined in terms of the saturation velocity of electrons). Note that for a given fabrication technology, the channel transit time and the input rise and fall times tend to scale in the same manner. Thus simple scaling of devices does not invalidate the quasi-static approximation. Reduction of parasitic capacitances will, however, tend to speed up circuit operation without improving model performance. State-of-the-art digital circuits operate well within the speed limitations of the model [6.39], [6.40]. It should be noted that the Meyer capacitance model also uses the quasi-static approximation and thus the speed limitations are identical.

The derivation of the charge expressions uses the gradual channel approximation and thus is strictly applicable only to long-channel devices. In devices shorter than 5–10 microns the accuracy of the charge-oriented model is reduced due to small-geometry effects. The significance of the resulting errors depends on the particular circuit application. In circuits dominated by parasitic capacitances the errors may be negligible, while for dielectrically isolated structures the requirements are much more stringent. The Meyer model uses the same gradual channel approximation and suffers the same channel length limitations. The conventional models, however, are not as readily modified to correct for short-channel effects.

The principal source of error in short-channel devices is the effect of the source and drain junctions on the gate, bulk, and channel charges. These effects appear as a reduction in gate-bulk, drain-bulk, and source-bulk

capacitances and an increase in gate-drain and gate-source capacitances. The capacitance characteristics also shift with changing threshold voltages in very short or narrow devices.

The simplest method of including these effects in the model is to modify the body-effect coefficient γ in the manner proposed by Poon and Yau [6.37], [6.38]. This accounts for the changes in threshold voltage and approximates the other changes in capacitances. The charges in short-channel devices can also be modeled directly. The use of two-dimensional numerical device modeling to otain corrections to the charge distributions has been previously reported [6.34]. Extremely accurate results may indeed be obtained by using a two-dimensional simulation to calculate the charges for the model ((6.23) would be used to obtain Q_S and Q_D). The difficulty lies in representing the charges in compact tabular or closed analytical form.

There is another more subtle error caused by velocity saturation and failure of the gradual channel approximation. For short-channel devices the first term of (6.18) and (6.19) can no longer be written as in (6.21). The result is that an extra term appears in (6.22):

$$I_S = -I_0(V_G, V_D, V_S, V_B) - I_{TT}(t) + \frac{dQ_S}{dt} \tag{6.25a}$$

$$I_D = +I_0(V_G, V_D, V_S, V_B) + I_{TT}(t) + \frac{dQ_D}{dt} \tag{6.25b}$$

I_{TT} is a transient transport current and is non-zero only for small devices under transient operating conditions. The properties of this current component have been studied using the two-dimensional transient simulation of Oh [6.28]. It has been shown that for slower inputs, I_{TT} is capacitive in nature (i.e. it scales with circuit speed). For a 3μm device ($t_{ox} = 1000$Å, $x_j = .5\mu$m, $N_A = 8 \times 10^{15}\text{cm}^{-3}$) I_{TT} typically comprises less than 10% of

the transient current. In most cases the transient transport current may be ignored entirely.

E. Summary and Conclusions

A fast, closed-form analytical model for MOS transitor capacitances has been developed which combines improved numerical simulation techniques with a complete description of the charges in a device. The model is well-behaved numerically and guarantees conservation of charge. The charge expressions used by the model are obtained analytically assuming one-dimensional current flow and the gradual channel approximation. The accuracy of the resulting capacitance characteristics has been verified by direct measurements on long-channel devices. The model is complete in that it includes non-reciprocal effects and capacitive couplings between source and bulk, drain and bulk, and source and drain.

The charge-oriented model, like the conventional Meyer model, is limited by speed and channel length constraints. Use of the quasi-static approximation introduces errors at high speeds. For digital circuits this occurs when the input rise and fall times are less than 10–20 times the channel transit time $\tau_0 = L^2/\mu V_{GSmax}$. State-of-the-art digital circuits generally operate within the speed limitations of the model.

For short-channel devices the charges used by the model must be corrected to account for two-dimensional and velocity saturation effects. These corrections may be obtained from analytical approximations or by two-dimensional numerical device analysis. In small devices a transient transport current appears in addition to the DC transport and transient charging currents. Although interesting, this component is of little consequence in practical circuit simulation.

6.3 ANALYSIS TOOLS

As a consequence of reduced device dimensions, highly two- and even three-dimensional device structures have resulted. New problems occur which are not significant in long-channel devices — for example short-channel [6.41] and narrow width effects [6.42]. Although these effects may not be a problem if the ideal scaling theory [6.43] is followed, new physical scaling limits are being encountered due to the practical limitations [6.39] — for example, no scaling of bias due to the standard 5-V level and the need for thicker oxides due to the breakdown-associated limits. Fig. 6.9 illustrates the field patterns of the long- and short-channel devices. In the long-channel device, there is clearly a one-dimensional region where the gradual channel approximation [6.44] holds. However, in the short-channel device, the edge region is significant where two-dimensional field coupling occurs. As a consequence, the device operation is predominantly controlled by the two-dimensional field. Hence, the device characteristics are highly dependent on the device structure. In this case, classical analysis based on the one-dimensional model is not valid and two-dimensional analysis is essential. However, the inherent generality of numerical device solutions in two dimensions has a cost in terms of computer time required to obtain a solution. This is especially true when both the potentials and current in a device are computed. In this section two aspects of the one-carrier transport solution in two dimensions will be considered. First, the evolution and performance evaluation of TANDEM II is discussed. Second, a novel boundary value solution method is discussed with special emphasis on its application to time-domain characterization. Both topics address the efficiency and

accuracy issues involved in computing current in MOS transistors.

6.3.1 Two-Dimensional Simulation of One-Carrier Steady-State Current Transport in MOS Devices (Craig Price)

Initial efforts in the development of a two-dimensional simulation program under this contract were aimed at finite element analysis [6.45]. The primary advantage of this method is the flexibility of the triangular finite element mesh versus the rectangular finite difference mesh. The finite element method was used in the program SUDAN written by S. Y. Oh, and R. J. Lomax. This program was a transient simulator and was written in two versions closely resembling the programs of Buturla and Cottrell [6.46], and Barnes and Lomax [6.47]. Version one solved Poisson's equation and the two-carrier continuity equation in an alternating manner. This program suffered from a severe time step limitation as a result of the decoupled equations. Version two solved both equations simultaneously, but it did not converge at nominal drain voltages.

Experience with these programs led directly to the development of TANDEM-II (called TANDII for convenience), a steady-state simulator which alternately solves Poisson's equation and the one-carrier continuity equation on a triangular mesh using a finite difference formulation of the equations. The sparse matrix equations are solved using the Vectorized General Sparsity algorithms package (VEGES) [6.48]. These routines utilize a direct matrix LU decomposition solution technique and, due to their vectorized structure, have the advantage that the inner loops of the forward and back substitutions may be efficiently executed using vectorized or assembly language coded instructions.

TANDII has been implemented on an HP-1000F minicomputer to take advantage of the ease of access and the shorter total time required for a solution as described in [6.26]. The parsed input structure and use of graphical output enhance the usability of this program. The input syntax is similar to that of TANDEM [6.25].

A. Results and Discussion

Although TANDII has been implemented on a minicomputer, a majority of the results have been obtained on the DEC-20 and all timing comparisons use this data. In order to obtain a benchmark on the status of this program, it has been compared against the Hitachi two-dimensional device simulation program CADDET [6.49][6.6] since it is the fastest two-dimensional device analysis program available.

Like TANDII, CADDET is a single carrier simulator and is capable of handling majority carrier devices. CADDET input allows definition of various types of MOSFET and JFET structures. The semiconductor surface (semiconductor-oxide interface for MOS) must be planar for these devices and the oxide surface must be planar or have at most one step. Doping profiles may be uniform or be two-dimensional Gaussian. Although many input structures can be accomodated by CADDET, these limitations generally make it difficult to accurately match the device structures found in practice. Such features as tapered oxides, non-Gaussian source and drain profiles and straggle in channel implants can have significant effects on device characteristics. Some of the limitations listed above may be overcome through expansion of the input definitions; however, other of the limitations are inherent in the methods used by CADDET. The limitation to planar

interfaces and oxide surfaces, for example, are imposed by the rectangular grid and the assumption of a one-dimensional field in the oxide.

The use of the triangular finite element type of grid in TANDII allows for a much more general device structure including nonplanar surfaces and tapered oxides. TANDII has the facilities to handle general impurity profiles, general specification of electrode locations, contact work function differences, degenerate statistics, and field and impurity concentration dependent mobilities.

CADDET solves the current continuity equations using stream functions. These have the advantage that the variable values range from zero to one, resulting in a very well behaved function. Stream functions have the disadvantage, however, that they represent conservative flow so that nonconservative effects such as recombination and generation cannot be modeled directly. Fortunately, this is not a great problem in majority carrier devices since the effects of recombination and generation are often small and can be modeled as perturbations to the final solution.

Some of the limitations of CADDET appear as inaccuracies in the solution. The accuracies of the numerical algorithms for both programs are equivalent with the exception that the convergence criteria affect the accuracy of solutions. The convergence criteria for TANDII can be selected by the user depending on the circumstances of the simulation being performed. The convergence criteria of CADDET appears to be insufficient for accurate simulations as illustrated by the discrepancies in the CADDET solutions for different voltage steps as shown in Figure 6.10.

The use of the regular rectangular mesh in CADDET offers a distinct advantage in speed since it simplifies the formulation of the matrix equations. Also, the matrix equation solution techniques allow the use of Stone's method

in solving the nonlinear Poisson's equation [6.50]. Further, CADDET's use of single precision arithmetic gives an increase in solution speed (depending on the machine) with a comensurate penalty in accuracy. TANDII's use of an irregular triangular grid requires larger storage and the equation formulation is more complex. In addition, the use of direct LU decomposition solution techniques further increases the storage and necessitates the use of double precision arithmetic. The major advantage is the elimination of the need for iterative solution of the matrix equations.

In order to compare the speeds of TANDII and CADDET, a timing comparison was made on the DEC-20 using equivalent convergence criteria for the two programs. The device simulated was a planar MOS transistor with an implanted channel and Gaussian source and drain profiles. The TANDII simulation was performed on a 323 node (17×19) graded rectangular mesh with diagonals added to yield a triangular mesh. CADDET was executed using a 1776 node (37×48) graded rectangular mesh. These two meshes are shown in Figure 6.11. The results of the comparison are shown in Table 6.5 and reveal that, with this grid ratio, TANDII is approximately half as fast as CADDET in the high gate and drain bias case and slightly faster in the low drain bias case.

Since a general triangular mesh was not available at the time the above simulations were performed, the TANDII mesh shown in Figure 6.11 does not take advantage of the triangular mesh. The node placement efficiency of a general triangular mesh is greater than that of a graded rectangular mesh so that fewer nodes are required for an accurate simulation [6.51]. Thus, using a general triangular mesh, TANDII would normally require only about one-third the number of mesh points of a graded rectangular mesh.

As described in [6.52] the alternating method of solving the Poisson's

and continuity equations suffers a decrease in the convergence rate at high gate and drain bias levels. Figure 6.12 shows a typical sample of drain current versus iteration number for an MOS device in the high gate and drain bias region (using the projection and single-Poisson methods described below). Note that in 25 iterations, the drain current has converged to within one percent of its fully converged value but that 30 more iterations are needed for full convergence. In these bias regions, acceleration of the convergence is important to minimize the solution time. Also, The least stringent convergence criteria necessary should be used.

The first type of acceleration found to be beneficial was to limit the number of iterations in the Newton loop of the Poisson's equation solution. Normally the nonlinear Poisson's equation is solved completely using several Newton iterations between continuity equation solutions. It was found that fewer total iterations were needed if the Poisson's equation iterations were only partially converged, and, in fact, excellent results were obtained when only 1 iteration on Poisson's equation was performed between continuity equation solutions.

The accuracy of the initial guess at each bias level also has an influence on the number of iterations required. The use of projection (extrapolation of potentials and quasi-Fermi levels from two previous solutions) was successful in reducing the total time for a solution.

The nearly linear convergence of potential seen in the first ten to 20 iterations was accelerated by overcorrecting the potential solved for by Poisson's equation. Instead of updating potentials by the ΔV solved for in Poisson's equation, $\alpha\Delta V$ is used where α is a number between 1 and 2 determined by the convergence rate. This may be interpreted as an over-relaxation method. The convergence acceleration resulting from this method

is illustrated in Figure 6.13.

Another method found to accelerate the initial convergence rate involved modification of the linearization term in Poisson's equation. Normally (using Boltzmann statistics) the linearization term in Poisson's equation is $(n+p)kT/q$ which arises from the fact that if the quasi-Fermi levels remain constant, the n and p rise and fall by $\exp(\Delta Vq/kT)$. It was found, however, that when using a single Poisson iteration per continuity solution, the quasi-Fermi levels tended to track the node potentials (Figure 6.14), thus the ΔV's solved for in Poisson's equation were consistently underestimated. Acceleration was achieved through the use of a linearization factor of $(\alpha n+p)kT/q$ where α ranged from .5 to 1. This method has essentially the same effect as the over-relaxation method except that it's influence is proportional to the free electron concentration of the node.

Timing comparisons of the above methods are given in Table 6.4. A more stringent convergence criteria was used for these simulations than was used in the simulations of Table 6.3.

As mentioned earlier, the use of Stone's method in CADDET for solving Poisson's equation has a speed advantage but restricts the discretization to a rectangular grid. More specifically, the method actually only restricts the grid to be rectangularly connected, i.e. the rectangular grid may be distorted to fit the problem as long as each node maintains its same general position with respect to it's neighbors. This modification would eliminate one of the greatest limitations of CADDET since it would permit nonplanar structures. The problem remains, however, of how to discretize the equations on this grid.

B. Conclusions

A two-dimensional device simulation program has been written which solves Poisson's equation and the one-carrier continuity equation. Although the program is somewhat slower than CADDET, it has the facilities to handle nonplanar structures, general impurity profiles, contact work function differences, general specification of electrode locations, degenerate statistics, and concentration and field dependent mobilities. These flexibilities make the program an excellent tool for the investigation of solution sensitivities to mesh, structure, and impurity profiles as well as for the investigation of mobility models. The program is also useful as an aid in the development of device models.

6.3.2 A Simplified Two-Dimensional Numerical Analysis of MOS Devices Including Transient Phenomena (Soo-Young Oh)

Two-dimensional MOS device problems have been analyzed using the finite difference (FD) [6.53]–[6.54] or finite element (FE) [6.55]–[6.56] methods. Both methods discretize the hole and electron concentrations and potential at each node point and require node allocations over the entire region; therefore, they require many nodes to solve two-dimensional problems. For example, 1500–2000 node points are necessary to solve a typical two-dimensional MOS device problem. Due to this large analysis grid, excessive CPU time and memory allocations result. Typically, 300 K–400 K bytes of memory are used and the CPU time for one bias point is on the order of 10 seconds † using an IBM 370/168. Due to these execution times and memory requirements, FD or FE methods are marginally practical for small numbers of DC simulations—specifically to enhance understanding of device operation mechanisms or to predict operational limits. However, they are totally impractical for large numbers of simulations (I-V characteristic generations, device parameter extractions, statistical simulations and transient simulations).

The purpose of this research is to develop a two-dimensional device analysis method which is sufficiently accurate and also faster than FD or FE by orders of magnitude so that it can be used as a practical tool for all aspects of the device simulation including two-dimensional transient simula-

---

†This figure is based on CADDET, a finite difference program developed by Hitachi[6.49] [6.58].

tions. Toward this goal, two milestones have been reached. In the first stage, a new two-dimensional boundary value method (BV method) has been developed. This BV method uses several simplifications which are appropriate to the MOS device analysis. The current transport is only considered one-dimensionally along the surface channel using the majority carrier of the source and drain. Poisson's equation in two dimensions is reduced to an initial solution followed by a series of the boundary value problem solutions for the incremental potentials with the relevant incremental boundary conditions. In this method, the number of nodes required is at least an order of magnitude smaller than FD or FE methods since nodes are allocated only along the channel and boundary. It is appropriate for most device analysis applications and simple circuit simulations, since it is orders of magnitude faster than the results of FD quoted above. The derivation of the new method was explained and its accuracy was discussed previously [6.34].

In the second stage, two-dimensional transient analysis has been performed to understand and characterize the MOS transient phenomena of long- and short-channel devices in a wide range of operating speeds. To date, no practical simulation tools are available that can include the distributed-channel and two-dimensional effects and perform time-domain analysis at the same time. In section A, MOS transients are measured and compared with simulations using the BV method to verify the applicability of the new method. In section B, MOS transient phenomena, transit-time delay and charge pumping, are analyzed using measurements and the BV method on long-channel devices. The short-channel behavior of devices is predicted using simulation. Section C summarizes the results of the research.

A. Measurements and Simulations of MOS Transients

In the analysis of MOS transients, the quasi-static model [6.32] [6.33] is usually used in the stray-capacitance limited mode and the non-linear transmission line model [6.12] [6.13] is used in the channel transit-time limited mode. However, neither model includes the two-dimensional effects in small geometry devices because the gradual channel approximation [6.44] is used in their derivation. In order to include the two-dimensional effects, the BV method is used in the analysis of MOS transients. In this section, MOS transient simulations using the BV method are compared with measurements to prove the validity and accuracy of the method in analyzing MOS transients.

Metal gate n-channel MOS test devices were fabricated on a p-type <100> substrate with a resistivity of 1 to 4 Ω cm using a conventional metal gate nMOS process. Fig. 6.15 shows a microscope picture of the test devices. The test devices consist of 3 MOS capacitors and 5 nMOS transistors. The size of the MOS capacitors is 500 X 500 microns. They are used in C-V measurements for the device parameter extractions. The 5 transistors are used in the measurements of MOS static and dynamic characteristics. The mask channel length of the transistors are 5, 10, 20, 40 and 100 microns while the widths are fixed to 100 microns.

A.1 Measurement System for MOS Transients

Measurements of MOS transients were made to verify the applicability of the BV method to MOS transients. The measurement set-up is shown in Fig. 6.16. The gate of the test device was driven by a ramp generator having adjustable rise time. The source and substrate are connected to ground. The drain is connected to 5 V supply voltage through a 1 K Ω resistor. The drain

current is measured using the voltage developed across the 1 K Ω resistor. The voltage developed across the resistor is few tenths of a volt and its effect is negligible in the saturation operation. An FET probe was used to minimized the probe capacitance. The input gate signal and output drain current (output of the FET probe) were displayed on an oscilloscope as shown in Fig. 6.17. A stripline probe card was used to minimized the stray capacitance in probing the transistors directly on the wafer. The measurements were made only on 100 × 100 micron n-channel transistors. For smaller devices, a faster input signal and measurement system are necessary to study the transients.

A.2 Comparison of Simulations with Measurements

Consider first the slow input case. Fig.6.18 shows the results obtained for an input ramp having a rise time of 500 nsec. Part (b) of the figure shows the measured and simulated drain currents and the simulated source current. The simulated drain current is in close agreement with the measurements—the slight discrepancy is attributed to the imperfections in the input ramp. The analytical quasi-static simulation[6.9] is also plotted for comparison.

Consider next the very fast input case. Figure 6.19 shows the results obtained for an input ramp having a 60 nsec rise time. Again the numerical simulation agrees well with the measurement except for a constant offset corresponding to a stray gate-to-drain capacitance of 0.25 pf. The quasi-static waveforms are qualitatively different from the measured and simulated results. The quasi-static charging current is capacitive and thus the total transient current is the sum of the steady-state current I_0 and a charging current which is proportional to dV_g / dt. The step changes in the source

and drain currents imply instantaneous changes in the charge and potential configuration in the channel. In the actual device, these changes are limited by the availability of channel charge and occur gradually, as evidenced by the linear increase in source current with time and exponential decay at the end of the input ramp.

These two comparisons show good agreement of the simulations and measurements and validate the use of BV method in MOS transient characteristics. A common feature of these two examples is the delay between the start of the input ramp and the appearance of the drain current, which is due to the non-zero transit-time of the channel charge. In the next section, this delay and other transient phenomena will be analyzed in detail.

B. MOS Transient Analysis

In this section, the transient phenomena of MOS devices are simulated and analyzed using the BV method. The delay time due to the non-zero channel transit-time is considered first. As shown in the previous section, there is a delay between the input gate signal and the output drain current. This delay becomes important in the turn-on transients of an inverter, since the delay time should be added to the output discharging time. The capacitance and transit-time contributions can be comparable when the stray capacitances are negligible as in the case of dielectric-isolated devices.

Charge pumping can occur in the channel of MOS devices during turn-off transients. The charge pumping [6.57] occurs when the excessive channel charge due to the decrease of the gate signal is larger than the charge removal capability to the source and drain. Then the excessive charge goes to the substrate. The charge pumping current can affect the actual

substrate bias in the scale-downed ion-implanted devices where a lightly doped substrate is usually used. In this case, the resistivity is high and significant negative bias can be developed between the real substrate and the back contact of the substrate. For a floating substrate as in the case of dielectric-isolated devices, the substrate potential is greatly affected by the charge pumping. Hence, it is necessary to characterize these effects more rigorously.

In the following analysis, the transient phenomena of the long-channel device are simulated and compared with the measurements. The behavior in short-channel devices is simulated and predicted using only device analysis, since the measurements of short-channel devices are very difficult and prone to inaccuracies.

B.1 Transit-time Delay

As illustrated in Fig. 6.18 and Fig. 6.19, there is a finite time delay, τ_D, between the input signal reaching threshold voltage and the output drain current starting to flow—a result of the finite channel transit-time. Fig. 6.20 shows the movement of the channel charge during this transient. The electron distributions are plotted along the channel with 0.2 rise time(t_r) interval. The plot shows that electrons take almost one t_r to cross the channel from the source to drain. During this time no drain current flows, which clearly indicates why there is a delay between input signal and output drain current.

Burns [6.12] and Swanson [6.13] analyzed these experimental conditions using the non-linear transmission-line model. Swanson derived the analytical expression for the delay time when t_r is longer than τ_D and the

device is in saturation.

$$\tau_D = \sqrt{t_r \tau_0} \qquad \tau_D < t_r \tag{6.26}$$

where

t_r = rise time of the gate signal

τ_0 = $L_{eff}^2 / \mu_n V_{gsmax}$

= characteristic channel-transit time

V_{GS} = 0 V to V_{gsmax}

As a limiting case of the ramp input, Burns obtained the result of τ_D for the step input

$$\tau_D = 0.38\,\tau_0 \tag{6.27}$$

Both analyses show the delay is a function of the rise time, effective channel length and gate bias. In order to find the relation between the delay and rise time in a long-channel device, the delay times from the measurements, simulations and Equation (6.26) are plotted against the rise time of the gate signal in log-log scale in Fig. 6.21. The measurements and numerical simulations agree whereas Equation (6.26) gives slightly faster delay times, since the substrate doping effects are neglected. τ_D is proportional to the square-root of t_r. However, the delay time deviates from this dependence and saturates at the value of 0.5 τ_0. The limiting value of 0.5 τ_0 is close to Burn's predictions for a step input. The deviation is also due to the neglect of the substrate doping effects.

The channel transit-time scales as the square of the channel length. This scaling is valid down to lengths of about 10 microns, at which point channel length modulation effects become important. In smaller devices, velocity saturation and two-dimensional effects must be considered. In Fig. 6.22, the delay times of a short-channel devices ($L_{eff} = 3$ microns) are simulated and plotted against t_r to determine quantatively the effect of these phenomena on the transient of the short-channel MOS transistors. Equation (6.26) and (6.27) are also plotted for comparison. Here, τ_D and t_r are normalized by τ_0 to compare the simulations of the short-channel case with those of the long-channel case. For a relatively slow input signal ($t_r \simeq 10\ \tau_0$), the delay times of the short-channel device are faster than Equation (6.26)—this is opposite to the long-channel case, which is due to the two-dimensional effects. Electron distributions are plotted along the channel for a long-channel case in Fig. 6.23(a) and for a short-channel case in Fig. 6.23(b). When time is zero, no electrons are in the channel of the long-channel device. However, electrons have already moved across approximately 0.2 L_{eff} of the channel in the short-channel device, due to the two-dimensional effects. The normalized delay time is faster in the short-channel device, since electrons cross only 80 % of the effective channel length. When the rise time becomes faster, the delay time in the short-channel device saturates much earlier than in the long channel case with a value of L_{eff} / v_{sat}. Fig. 6.23 (a) and (b) illustrate the movement of electrons for the long and short-channel devices when the rise time is 2 τ_0. It is clear that the electrons move more slowly in the short channel device than in the corresponding long-channel case owing to velocity saturation when the channel and time are normalized by the corresponding L_{eff} and τ_0. Due to these effects, the slope has decreased in the short-channel case. The results shows that the velocity saturation and

two-dimensional effects have a significant effect on the transit-time delay.

B.2 Charge Pumping

When the channel collapses during the turn-off transient, some charge in the channel goes to the bulk to recombine with holes. There are two components [6.57] in this charge pumping current. One component is due to the surface states. Some electrons are trapped in the surface states when the channel is inverted and then recombine with holes when it is accumulated. In n-channel devices, the channel is never accumulated, since the flat band voltage is usually negative and the gate voltage is varied between 0 V to positive bias. Hence, this component is negligible under the conditions described here.

The second component is due to the electrons which actually cross the depletion region between the channel and substrate and subsequently recombine with holes in the substrate. Fig. 6.24(b) shows the electron distributions in the channel during the turn-off transient. The source, drain and substrate are grounded and the gate voltage is changed from 5 V to 0 V. When the gate signal decreases rapidly, the excess charge due to the decrease of gate signal should be removed to the source and drain. If the removal capability is not sufficient, some electrons are trapped in the middle of the channel as illustrated in Fig. 6.24(b). These trapped electrons make the channel more negative with respect to the substrate. Some electrons cross the depletion region and recombine with holes in the substrate.

The average charge pumping current is measured in the circuit shown in Fig. 6.24(a). The source, drain and substrate are grounded and the pulse input is applied to the gate and the current is measured at the substrate

terminal. Devices with L_{eff} ranging from 98 microns to 3 microns are measured and simulated for different fall times to find the dependence of the charge pumping current on the fall time. The current is converted to the fraction of the pumped charge per cycle to compare the results for the different channel lengths and the results are plotted in Fig. 6.25. The graph shows the charge pumping increases when the fall time decreases—for a longer channel device a larger charge pumping is obtained for the same fall time. It is interesting to notice that the dependence of the charge pumping on the fall time is a power relation in the low current level (below 5 %). The slopes range from a value of -2.25 when L_{eff} is 98 microns to a value of -3.25 when L_{eff} is 3 microns. The critical fall time (t_{fc}) is defined as the fall time of the gate signal when 1 % of the channel charge is pumped to the substrate. Using t_{fc} the relationship between the charge pumping and L_{eff} is extracted from Fig. 6.25 and plotted in Fig. 6.26. Its dependence on L_{eff} is larger than the expected quadratic relation—based on the fact that the channel transit-time is a quadratic function of L_{eff}. The slope becomes steeper when L_{eff} becomes shorter due to the two-dimensional effects. This is consistent with the observed decrease of the delay time in the short-channel device.

C. Summary

A simplified two-dimensional device analysis method has been developed using the boundary value method. Hence, it reduces the computation time by orders of magnitude from those of FD or FE methods so that it can be used as a practical tool for all aspects of device simulation including two-dimensional transient simulations.

Simulations using the BV method agree with the measured static

and dynamic characteristics of MOS transistors over a wide range of device parameters and operating speeds. MOS transient phenomena including transit-time delay and charge pumping have been measured and simulated. The transit-time delay has the same functional dependence as Equation (6.26), however it predicts significantly slower switching than Equation (6.26) due to the substrate doping effects. Short-channel behavior of the transit-time delay has been simulated and analyzed. Results show that velocity saturation and two-dimensional effects play a significant role in the short-channel MOS transients. The charge pumping current has a slightly larger dependence on L_{eff} than the expected quadratic dependence. It is interesting to note the power law dependence of charge pumping on the fall time of the gate signal.

6.4 TEST STRUCTURES

Device structures have become highly two-dimensional (2-D) with the trend toward scaled-down geometries, and test structures for the measurement of 2-D profile and device effects are of increasing interest and importance for submicron technologies. Two-dimensional profile measurement techniques reported earlier [6.14] have proved to be accurate and useful for profiles such as CMOS p-well and lateral channel-stop diffusions. In this section, we will describe further developments in 2-D profile measurements. The necessity for a new approach applicable to VLSI devices is discussed. Next a generalized device structure which includes the generic features of many technologies is presented. This structure features a two-dimensional low-concentration boron profile near locally oxidized regions. Two groups of test structures for this general class have been built for the 2-D profile measurements and are described in the final subsections.

6.4.1 Test Structures for Extracting Two-Dimensional Impurity Profiles (H. G. Lee)

Test structures are designed to control and characterize IC manufacturing processes, to develop new process technologies, and to determine device electrical parameters. To meet these requirements, they must be upgraded and refined just as the device design rules themselves are changed.

Generally, test structures have been designed with large dimensions to minimize the peripheral effects. For example, large area capacitors are used to measure the oxide thickness or substrate impurity profiles. Long and wide channel MOS transistors are employed to measure carrier mobility

as a function of gate or substrate potential. The uncertainty over length or area can be reduced by these devices. Most frequently, the parameters measured by such structures are used in modeling small-geometry devices.

The large-geometry approach is not sufficient for the development of VLSI technologies. Small-geometry devices exhibit size-dependent characteristics, and test structures must also be scaled-down in order to measure the size-dependent parameters. In addition, computer simulation capabilities are becoming important for test structure evaluations. The following example demonstrates how the analysis of measured data can be enhanced by using a 2-D Poisson simulation.

In the 2-D profile measurement method [6.14], the drain current of the test transistor is measured against the gate potential from which the threshold voltage variation in the width direction is calculated. Based on a simple relationship between the threshold voltage and the channel doping, the surface profile is obtained. As the width of the lateral profile decreases, however, the validity of this method becomes uncertain due to increased 2-D electrical effects. A two-dimensional Poisson analysis [6.25] has been used, therefore, to study the accuracy of the method. For an MOS transistor with a changing surface concentration in the width direction, the surface inversion charge (and hence the channel current at a small drain voltage) can be calculated using the Poisson solution. The lateral profile is then calculated, which is compared to the original specified 2-D profile. Using this method, factors which affect the accuracy of measured lateral profiles are identified, and application range of the method can be determined.

A. Representative Cross Sections

A few key processing techniques and resulting device structures are commonly employed in many technologies; local oxidation [6.59], for example, is applied in both bipolar and MOS technologies for isolation. A special structure that contains the generic features of several technologies is defined below.

Figure 6.27(a) is the cross section of an n-channel MOS (nMOS) transistor; the direction of its channel length is also indicated. The n^+ layer is the source or drain, and the p^+ regions are the boron-doped channel stops. A bipolar transistor (Fig. 6.27(b)) has a similar cross section [6.60]. The shallow n^+ layer serves as an emitter below which is a thin p-type base region with a p^+ base contact on both sides; an n-type collector is at the bottom. The static-induction transistor [6.61] in Fig. 6.27(c) and the taper-isolated dynamic-gain cell [6.62] in Fig. 6.27(d) also have similar cross sections. The cross section in Fig. 6.28 shows a generic combination of profiles and oxide regions important to small-geometry devices. In this cross section, the one-dimensional characteristics are denoted by (1) through (3), where

(1) is the shallow n^+ layer implanted or diffused from the polysilicon layer on top

(2) is a low-concentration boron-diffused layer whose diffusion is affected by the n^+layer as in emitter-push with phosphorus and in emitter-pull with arsenic

(3) is boron diffusion which is affected by oxidation, where the parameters are segregation and oxidation-enhanced diffusion

The two-dimensional characteristics are indicated by (a) through (e), where

(a) is the *bird's beak* caused by lateral oxidation under the nitride layer in a LOCOS process [6.59]

(b) is the shallow implanted profile self-aligned to the bird's beak [6.62]; even after implantation, the resulting profile is highly two-dimensional

(c) is the lateral diffusion of channel-stop regions, which move in the silicon while simultaneously being removed from the oxide by preferential segregation [6.63]

(d) is the interaction of local oxidation with diffusion at a lateral distance where a certain amount of diffusion enhancement is expected to be distance-dependent [6.64], [6.15]

(e) is the self-aligned implantation of the n^+ layer with the edge of the polysilicon gate in the third dimension as in the edge of n^+ source and drain diffusions at the channel side

As can be seen, this generalized cross section contains many process and device effects that are important to small-geometry devices. Each of the device effects has to be measured by a test structure optimized for that purpose. In the following sections, we will focus our attention to test structures developed for low concentration two-dimensional distribution of boron.

B. Two-Dimensional Profile Measurement

Design of a test structure for two-dimensional profile measurements was started during this project. Results obtained and reported up to last year's report are briefly summarized.

(1) A method of calculating the surface impurity profiles of laterally diffused regions from measured transistor I_{DS} vs. V_{GS} characteristics has been developed. PROFILE-1 test chip was fabricated

using CMOS technology and surface impurity profiles of p-well and lateral channel-stop diffusion were successfully measured [6.65].

(2) Accuracy of the lateral profile measurement method has been examined using numerical device analysis. The result shows that lateral profiles with over a 2 μm transition length can be accurately measured using this method. For a rapidly changing profiles with half a micron transition length, two-dimensional potential distribution invalidates assumptions used in the method, and the method becomes inaccurate [6.66].

Previously reported measurement results were for laterally diffused profiles under inert ambient or under a uniform oxidation at surface. As shown in Fig. 6.28, however, two-dimensional diffusion associated with local oxidation has a significant impact for scaled-down devices. Two experiments have been initiated to study the effect of local oxidation on the diffusion of boron, as indicated in Fig. 6.29. It is believed that oxidation interacts with diffusion by generating point defects at the silicon-SiO_2 interface. These point defects enhance the diffusion of common dopant impurities like boron or phosphorus. Therefore, the amount of interaction will depend on the flux of point defects at the location of impurity diffusion. Oxidation-enhanced diffusion has been observed as deep as 20 μm below the surface (region A in Fig. 6.29). On the other hand, it is not well understood how the oxidation enhances diffusion of impurities in a 2-D structure (region B or C in Fig. 6.29). In the following sections, two experiments to study the oxidation effects on diffusion in regions E and C by means of electrical test structures are discussed.

C. Channel Stop Profiles near Local Oxidation Edges

Lateral channel-stop (LCS) diffusion in NMOS or CMOS technology creates substantial 2-D effects in scaled-down devices. The lateral encroachment of LCS diffusions together with the recessed oxide structure decreases the active channel area, and affects the threshold behavior of narrow width transistors. Proper control of the lateral diffusion will improve the device performance without a penalty in processing complexity. It is known that the lateral diffusion of LCS can be minimized by growing the field SiO_2 at lower temperature in wet ambient, since the ratio of the oxidation rate over the impurity diffusion constant is increased at lower temperature. More quantitative understanding of this process is essential for decreasing device dimensions.

Test structures were fabricated to measure the lateral penetration of LCS diffusions at various oxidation temperatures. P-type <100> silicon wafers with resistivity of 1–4Ω-cm were oxidized in dry O_2, growing about 450Å of buffer oxide. 1000Å of Si_3N_4 layer was deposited at 900°C, followed by another 1000Å of SiO_2 deposition which is used to pattern the Si_3N_4 layer as in Fig. 6.30. Boron atoms were implanted and were subjected to wet oxidation at temperatures between 900°C and 1050°C. Channel area of transistor A experienced uniform surface oxidation and that of transistor B included the locally oxidized area near the Si_3N_4 edge. After field oxidation, n^+ source and drain regions were diffused, the Si_3N_4 layer was stripped and p^+ channel stop implantation added around the devices. The gate oxide was grown at 850°C in wet ambient. Devices which used the Si_3N_4 as the gate dielectric were also fabricated in which case the channel profile would be less affected by gate oxidation. In the actual test devices, many identical

devices were fabricated side-by-side and connected in parallel to minimize edge effects [6.14]. In this experiment, the field oxide thickness for all cases was designed to be 6000Å after the oxidation step. Measured as well as calculated results are shown in Fig. 6.31. The theoretical ΔX's are generated by fitting the substrate and peak concentration values at the surface to the measured data, and by using the 2-D analytic diffusion models presented elsewhere [6.66]. Diffusion coefficients used in the theoretical calculations take the oxidation rate dependent enhanced diffusivity into account [6.67]. Increased lateral diffusion at higher oxidation temperature is apparent from this result. Differences of ΔX for cases A and B correspond to the length of the "bird's beak" and are in good agreements with the measured data [6.68].

D. Lateral Extent of Oxidation-Enhanced Diffusion

Oxidation of silicon not only enhances the diffusion of dopant atoms in the regions under oxidized surface but also in the laterally adjacent unoxidized regions (region C of Fig. 6.29). Lin et. al. [6.64] have shown that diffusivity of boron under the Si_3N_4 layer is enhanced by oxidation in neighboring regions, but the enhancement decreasing rapidly with increasing distance from the oxidized regions. Using angle lapping and staining methods, the lateral diffusion length of the oxidation-induced extrinsic point defects that cause the lateral enhancement was estimated to be 2 μm for temperatures from 900°C to 1100°C in dry O_2.

In this section, a new experiment is described using electrical test structures designed to measure the lateral effect of OED with the lateral profile measurement method. Cross section of the test structure is shown

after two processing steps in Fig. 6.32. ΔX_m defines the distance from Si_3N_4 mask edge to the edge of the implanted layer in the mask level. Si_3N_4 thicknesses used included both 500Å and 1000Å layers. Boron was implanted through the SiO_2 and Si_3N_4 layers into the silicon at doses of 2 and 3×10^{12} atoms/cm^2 and an energe of 180 keV. Additional implantation of 2×10^{13} atoms/cm^2 at 30 keV is added subsequently to prevent field inversion in the finished transistors. The wafers were then processed in a dry oxygen ambient at 1000°C for 8 to 10 hours. A similar process as described in the previous experiment was used to finish the fabrication sequence.

The measured surface profile is shown in Fig. 6.33. Enhanced diffusion of boron at a distance of 2.5 μm (mask level) from the Si_3N_4 layer edge is represented by a wider transition region, while substantially less enhancement is observed at distances over 5 μm (mask level). Due to the linewidth variations during processing and during formation of the Bird's Beak, the actual distance from the oxidized area to the edge of p-diffusion is decreased by approximately 1 μm. Enhancement of diffusivity for the three cases of ΔX_m, as defined by the ratio of boron diffusivity during oxidation to the diffusivity in inert ambient, D_{ox}/D_{inert}, is listed in Table 6.5. No apparent dependence on the Si_3N_4 layer thickness was observed for Si_3N_4 thicknesses in the range of 500Å to 1000Å.

E. Summary

An approach to VLSI test structures has been described. Unlike conventional test structures with large geometry, VLSI test structures must be scaled down together with the actual devices. The need for computer simulation tools has been discussed in order to deconvolve the profile-

dependent and size-dependent effects from the measured parameters of small-geometry devices. A generalized cross section is then introduced which includes generic features of many technologies.

Test structures have been designed to measure low concentration two-dimensional boron profiles, and two experiments are described. Lateral penetration of channel-stop diffusions in nMOS technology has been measured for a field oxide thickness of 6000Å, and the penetration increased at higher oxidation temperature as expected. The lateral extent of oxidation-enhanced diffusion has also been measured and the lateral diffusion length of the oxidation-induced extrinsic point defects was estimated to be 2 μm for temperatures in the range of 900°C to 1100°C in dry O_2.

REFERENCES

[6.1] D. A. Antoniadis, S. E. Hansen, R. W. Dutton, and A. G. Gonzales, "SUPREM I—A Program for IC Process Modeling and Simulation," *Stanford Electronics Laboratory Technical Report No. SEL 77-006,* May 1977.

[6.2] 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 No. SEL 78-020,* June 1978.

[6.3] D. A. Divekar, "Device Modeling for Statistical Circuit Simulation," *Stanford Electronics Laboratory Technical Report No. SEL 78-025,* July 1978.

[6.4] N. N. Chan, "Lump Partitioning of Integrated-Circuit Bipolar Transistor Models for High-Frequency Applications," *Stanford Electronics Laboratory Technical Report No. SEL 79-029,* Aug. 1979.

[6.5] D. C. D'Avanzo, "Modeling and Characterization of Short-Channel Double Diffused MOS Transistors," *Stanford Electronics Laboratory Technical Report No. G-201-6,* March 1980.

[6.6] T. Toyabe and S. Asai, "Analytical Models of Threshold Voltage and Breakdown Voltage of Short-Channel MOSFET's Derived from Two-Dimensional Analysis," *IEEE Trans. Electron Devices,* Vol. ED-26, pp. 453–461, Apr. 1979.

[6.7] D. B. Estreich, "Physics, Modeling, and Prevention of Latch-Up in CMOS Integrated Circuits," Ph. D. dissertation, to be published.

[6.8] D. E. Ward and R. W. Dutton, "A Charge-Oriented Model for MOS Transistor Capacitances," *IEEE J. Solid-State Circuits,* Vol. SC-13, no. 5, pp. 703–707, Oct. 1978.

[6.9] S. Y. Oh, D. E. Ward, and R. W. Dutton, "Transient Analysis of MOS Transistors," *IEEE Trans. Electron Devices,* Vol. ED-27, Aug. 1980.

[6.10] J. A. Greenfield and R. W. Dutton, "Nonplanar VLSI Device Analysis using the Solution of Poisson's Equation," *IEEE Trans. Electron Devices,* Vol. ED-27, Aug. 1980.

[6.11] H. K. Gummel and H. C. Poon, *Bell Sys. Tech. J.,* pp. 827-852, May–June 1970.

[6.12] J. R. Burns, "Large-Signal Transit-Time Effects in the MOS Transistor", *RCA Review,* Vol. 30, 1969.

[6.13] R. M. Swanson, "Complementary MOS Transistors in Micropower Circuits," *Stanford Electronics Laboratory Technical Report No. 4963-1,* Chapter 4, 1974.

[6.14] 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 J. Solid-State Circuits,* Vol. SC-13, p. 455, 1978.

[6.15] H.-G. Lee and R. W. Dutton, "Measurement of Two-Dimensional Profiles near Locally Oxidized Regions," *IEEE Int. Electron Devices Meet., Dig. Tech. Papers,* 1979.

[6.16] H. G. Lee, "Two-Dimensional Modeling of Fabrication Processes," Ph. D. Dissertation, to be published.

[6.17] D. B. Estreich, "Latch-Up and Radiation Integrated Circuit—LURIC: A Test Chip for CMOS Latch-Up Investigation," *Sandia Laboratories Report SAND 78-1540,* Albuquerque, NM, November 1978.

[6.18] E. Demoulin, J. A. Greenfield, R. W. Dutton, P. K. Chatterjee, and A. F. Tasch, Jr., "Process Statistics of Submicron MOSFET's," *IEEE Int. Electron Devices Meet., Dig. Tech. Papers,* pp. 34–37, 1979.

[6.19] D. A. Antoniadis and R. W. Dutton, "Models for Computer Simulation of Complete IC Fabrication Process," *IEEE Trans. Electron Devices,* Vol. ED-26, pp. 490–500, Apr. 1979.

[6.20] W. W. Johnson and P. P. Peressini, "Design of Short Channel Ion-Implanted MOSFET's with Relatively Deep Junctions," *IEEE Int. Electron Devices Meet., Dig. Tech. Papers,* pp. 550–553, Dec. 1974.

[6.21] P. P. Wang and O. S. Spencer, "The Threshold Voltage Characteristics of Double-Boron Implanted Enhancement-Mode MOSFET's," *IBM J. Res. Develop.*, Vol. 19, pp. 530–538, Nov. 1975.

[6.22] K. W. Yeh, J. Reuter, D. Heald, V. Dhaka, and A. Rangappan, "Optimum Short-Channel MOS Device Design for High Performance VLSI," *IEEE Int. Electron Devices Meet., Dig. Tech. Papers,* pp. 468–471, Dec. 1978.

[6.23] H. Nihira, M. Konoka, H. Iwai, and Y. Nishi, "Anomalous Drain Current in N-MOSFET's and Its Suppression by Deep Ion Implantation," *IEEE Int. Electron Devices Meet., Dig. Tech. Papers,* pp. 487–491, Dec. 1978.

[6.24] J. J. Barnes, K. Shimohigashi, and R. W. Dutton, "Short-Channel MOSFET's in the Punchthrough Current Mode," *IEEE Trans. Electron Devices,* Vol. ED-26, pp. 446–452, Apr. 1979.

[6.25] J. A. Greenfield, S. E. Hansen, and R. W. Dutton, "Two-Dimensional Analysis for Device Modeling (TANDEM)," *Stanford Electronics Laboratories Technical Report,* July 1980.

[6.26] J. A. Greenfield, "Two-Dimensional Modeling of Nonplanar Devices on Minicomputers," *Computer-Aided Engineering of Semiconductor Integrated Circuits,* pp. 270–300, July 1979.

[6.27] M. E. Zahn, "Calculation of the Turn-on Behavior of MOST," *Solid-State Electronics,* Vol. 17, p. 843, 1974.

[6.28] S. Y. Oh, Section 6.3.2 of this report.

[6.29] P. E. Cottrell and E. Buturla, "Two-Dimensional Static and Transient Simulation of Mobile Carrier Transport in a Semiconductor," *Proceedings of NASECODE I,* Dublin, June 1979.

[6.30] A. Van der Ziel, "Small-Signal High Frequency Theory of Field-Effect Transistors," *IEEE Trans. Electron Devices,* Vol. 11, pp. 128–135, Apr. 1964.

[6.31] I. Richer, "The Equivalent Circuit of an Arbitrarily Doped Field-Effect Transistor," *Solid-State Electronics,* Vol. 8, pp. 381–393, 1965.

[6.32] J. E. Meyer, "MOS Models and Circuit Simulation," *RCA Review,* Vol. 32, pp. 42–63, Mar. 1971.

[6.33] F. M. Klassen, "A MOS Model for Computer-Aided Design," *Phillips Res. Reports,* Vol. 3, pp. 71–80, 1976.

[6.34] J. D. Meindl et. al. , *Second Interim Report on Computer-Aided Engineering of Semiconductor Integrated Circuits, Stanford Electronics Laboratory Technical Report No. ICL 7-77,* pp. 387–399, July 1979.

[6.35] A. Vladimirescu, A. R. Newton, and D. O. Pederson, *SPICE Version 2F.1 User's Guide,* Department of Electrical Engineering and Computer Sciences, University of California, Berkeley.

[6.36] A. Vladimirescu, *The Simulation of MOS Integrated Circuits Using SPICE2,* Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, June 1979.

[6.37] H. C. Poon et. al., "DC Model for Short-Channel IGFET's," *IEEE Int. Electron Devices Meet., Dig. Tech. Papers,* pp. 156–158, 1973.

[6.38] L. D. Yau, "A Simple Theory to Predict the Threshold Voltage of Short-Channel IGFET's," *Solid-State Electronics,* Vol. 17, pp. 1059–1063, 1974.

[6.39] R. H. Dennard et. al. , "1 μm MOSFET VLSI Technology: Part II—Device Designs and Characteristics for High-Performance Logic Applications," *IEEE J. Solid-State Circuits,* Vol. SC-14, no. 2, pp. 247–255, Apr. 1979.

[6.40] L. M. Dang, "A Simple Current Model for Short-Channel IGFET and Its Application to Circuit Simulation," *IEEE J. Solid-State Circuits,* Vol. SC-14, no. 2, pp. 358–367, Apr. 1979.

[6.41] H. S. Lee, "An Analysis of the Threshold Voltage for Short Channel IGFET's," *Solid-State Electronics,* Vol. 16, p. 1407, 1973.

[6.42] W. P. Noble and P. E. Cottrell, "Narrow Channel Effects in IGFET's," *IEEE Int. Electron Devices Meet., Dig. Tech. Papers,* p. 582, 1976.

[6.43] R. H. Dennard et al, "Design of Ion-Implanted MOSFET's with very Small Physical Dimensions," *IEEE J. Solid-State Circuits*, Vol. SC-9, pp. 256–268, Oct. 1974.

[6.44] W. Shockley, "A Unipolar Field-Effect Transistor," *Proc. IRE*, Vol. 40, pp. 1365–1376, Nov. 1952.

[6.45] S. Y. Oh and R. J. Lomax, "Finite Elements for Semiconductor Device Simulation," *Stanford Electronics Laboratory Technical Report No. SEL 78-011*, Feb. 1978.

[6.46] E. M. Buturla and P. E. Cottrell, "Two-Dimensional Finite Element Analysis of Semiconductor Transport Phenomena," *Proc. Int. Conf. on Numerical Methods in Electric and Magnetic Field Problems, Italy*, Jun. 1976.

[6.47] J. J. Barnes and R. J. Lomax, "Finite Element Methods in Semiconductor Device Simulation," *IEEE Trans. Electron Devices*, Vol. ED-24, pp. 1082–1089, Aug. 1977.

[6.48] D. A. Calahan, et al., "Vectorized General Sparsity Algorithms with Backing Store," *Univ. of Michigan SEL Report 96*, January 15, 1977.

[6.49] M. S. Mock, "A Two-Dimensional Mathematical Model of the Insulated-Gate Field-Effect Transistor," *Solid-State Electronics*, Vol. 16, pp. 601–609, 1973.

[6.50] H. L. Stone, "Iterative Solution of Implicit Approximations of Multidimensional Partial Differential Equations," *SIAM J. Numerical Analysis*, Vol. 5, no. 3, Sep. 1968.

[6.51] C. H. Price, "Two-Dimensional Grid Generation for Semiconductor Device Simulation," *Stanford Electronics Laboratory Technical Report No. ICL 7-77*, July 1979.

[6.52] E. M. Buturla and P. E. Cottrell, "Simulation of Semiconductor Transport Using Coupled and Decoupled Solution Techniques," *Solid-State Electronics*, Vol. 23, pp. 331–334, 1980.

[6.53] M. B. Barron, "Computer Aided Analysis of IGFET Transistors," *Stanford Electronics Laboratory Technical Report No. 5501-1,* Nov. 1969.

[6.54] D. Vandorpe and J. Borel, "An Accurate Two-Dimensional Numerical Analysis of the MOS Transistor," *Solid-State Electronics,* Vol. 15, pp. 547–557, 1972.

[6.55] P. E. Cottrell and E. M. Buturla, "Steady-State Analysis of Field Effects Transistors via the Finite Element Method", *IEEE Int. Electron Devices Meet., Dig. Tech. Papers,* pp. 51–54, Dec. 1975.

[6.56] J. J. Barnes and R. J. Lomax, "Two-Dimensional Finite Element Simulation of Semiconductor Devices," *Electron. Lett.,* Vol. 10, pp. 341–343, Aug. 1974.

[6.57] J. S. Brugler and P. G. A. Jespers, "Charge Pumping in MOS Devices," *IEEE Trans. Electron Devices,* Vol. ED-16, pp. 297–302, Mar. 1969.

[6.58] T. Toyabe, K. Yamaguchi, S. Asai,and M. S. Mock, "A Numerical Model of Avalanche Breakdown in MOSFET," *IEEE Trans. Electron Devices,* Vol. ED-25, pp. 825–831, Jul. 1978.

[6.59] J. A. Appels and M. M. Paffen, "Local Oxidation of Silicon: New Technological Aspects," *Phillips Res. Reports,* Vol. 26, p. 157, 1971.

[6.60] G. Das and A. Phillips, "Emitter Compensation Effect," *IEEE Int. Electron Devices Meet., Dig. Tech. Papers,* p. 514, 1979.

[6.61] J.-I. Nishizawa, T. Takeshi, and J Shibata, "Field-Effect Transistor *versus* Analog Transistor (Static Induction Transistor)," *IEEE Trans. Electron Devices,* Vol. ED-22, p. 185, 1975.

[6.62] P. K Chatterjee, G. W. Taylor, and M. Malwah, "Taper Isolated Dynamic Gain RAM Cell," *IEEE Int. Electron Devices Meet., Dig. Tech. Papers,* paper 14.8 (Late news), 1978.

[6.63] R. B. Fair, and J. C. C. Tsai, "Theory and Direct Measurement of Boron Segregation in SiO_2 during Dry, Near Dry, and Wet O_2 Oxidation," *J. Electrochem. Soc.,* Vol. 125, p. 2050, 1978.

[6.64] A. M. Lin, R. W. Dutton, and D. A. Antoniadis, "The Lateral Effect of Oxidation on Boron Diffusion in <100> Silicon," *Applied Phys. Lett.*, 1979.

[6.65] J. D. Meindl et al., "Computer aided engineering of semiconductor integrated circuits," Interim report to ARPA, *TR 4969-3* Feb. 1978.

[6.66] H.-G. Lee and R. W. Dutton, "Minimum Geometry Design Considerations using a 2-D Implantation/Diffusion Model," Abstract no. 552, *Electrochemical Society Fall Meeting, Los Angeles, Calif.*, 1979.

[6.67] A. M. Lin, R. W. Dutton, and D. A. Antoniadis, "Oxidation Rate Dependence of B and P Oxidation-Enhanced-Diffusions in <100> Silicon," Abstract no. 133, *Electrochemical Society Spring Meeting, Boston, Mass.*, 1979.

[6.68] E. Bassous, H. N. Yu, and V. Maniscalco, "Topology of silicon structures with recessed SiO_2 " *J. Electrochem. Soc.*, Vol. 123, p. 1729, 1976.

Table 6.1

Device Parameters Used in the Calculation of Figs. 6.2 and 6.3

Oxide Thickness	500	Å
Channel Width (W)	10	μm
Effective Channel Length	0.89	μm
Substrate Impurity Concentration (N_B)	2.8×10^{15}	cm^{-3}
Junction Depth (X_j)	0.45	μm
Surface Electron Mobility (μ_n)	717	cm^2/V-sec
Threshold Voltage	0.380	Volt

Table 6.2

Minimum Rise and Fall Times for Quasi-Static Models

L_{eff}	minimum τ_R, τ_S
10 μm	0.8 ns
5 μm	0.2 ns
1 μm	0.3 ns*

$$\tau = 20\frac{L_{eff}^2}{\mu V_{GSmax}}, \quad \mu = 600 \text{ cm}^2/\text{v sec}, \quad V_{GSmax} = 4 \text{ V}$$

$$^*\tau = 20\frac{L_{eff}}{v_{sat}}, \quad v_{sat} = 6.5 \times 10^6 \text{ cm/sec} \quad \text{(see text)}$$

Table 6.3

Timing Comparisons of TANDII and CADDET Using Equivalent Convergence Criteria

V_G = 3.0 volts (V_T = 1.4 volts)

V_D	CADDET (1776 pts)	TANDII (323 pts)
1.5 volts	75 sec	115 sec
2.0	57	137
2.5	52	85
3.0	33	106
3.5	55	109
Total times	272 sec	552 sec

V_D = .1 volts (V_T = 1.4 volts)

V_G	CADDET (1776 pts)	TANDII (323 pts)
1.0 volts	25 sec	20 sec
1.5	45	23
2.0	83	59
2.5	27	35
3.0	24	19
Total times	204 sec	156 sec

Table 6.4

Timing Comparisons of Various Convergence Acceleration Methods

V_G = 3.0 volts (V_T = 1.4 volts)

V_D	None	1-Poisson	Projection 1-Poisson	Over-relax. Projection 1-Poisson
1.5 v	398 sec	274 sec	169 sec	139 sec
2.0	411	203	199	146
2.5	400	202	204	107
3.0	407	210	171	140
3.5	417	220	125	119
Totals	2033 sec	1109 sec	868 sec	651 sec

Table 6.5

ΔX_m (μm)	$\Delta X'_m$ (μm)	D_{ox}/D_{int}
2.5	1.5	2.50
5.0	4.0	1.36
10.0	9.0	1.14

$\Delta X'_m$ is the estimated actual distance from oxidized area to the edge of diffusion.

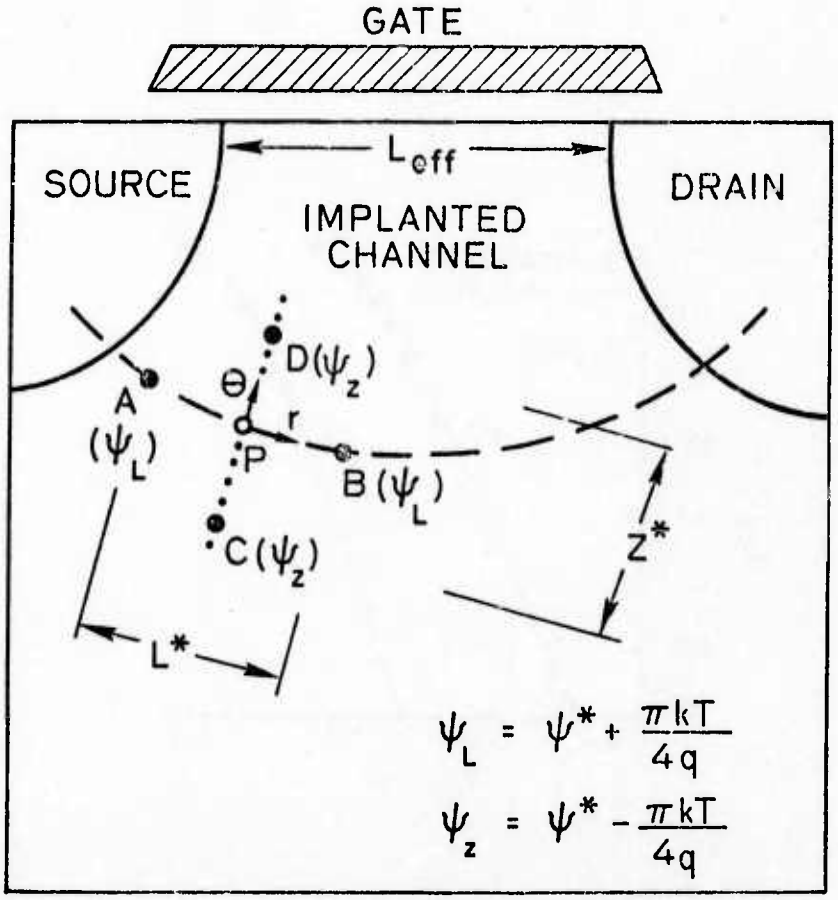


Fig. 6.1 Cross section of a MOSFET with a bulk current path showing the r and θ axes used to calculate the current. The points defining the approximate effective length L* and effective width Z* are shown.

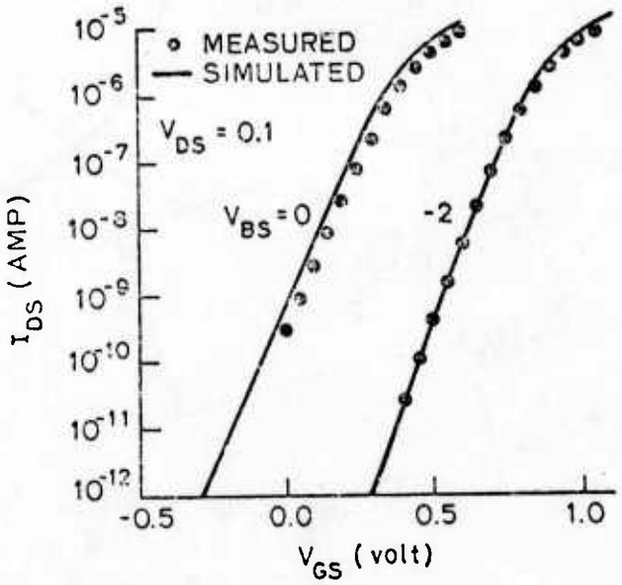


Fig. 6.2 Measured and simulated subthreshold I_{DS} vs. V_{GS} for a short channel MOSFET [6.18]. Device parameters are listed in Table 6.1. I_{DS} was calculated using (6.5).

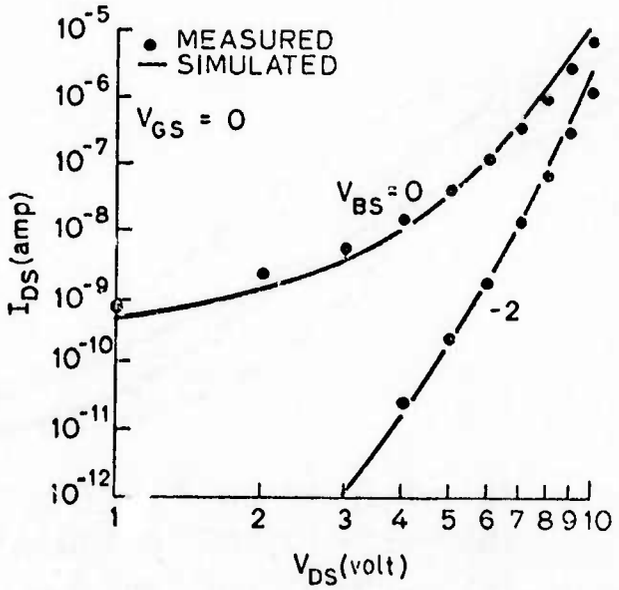


Fig. 6.3 Measured and simulated punchthrough I_{DS} vs V_{DS} for a short channel MOSFET [6.18]. Device parameters are listed in Table 6.1 I_{DS} was calculated using (6.6).

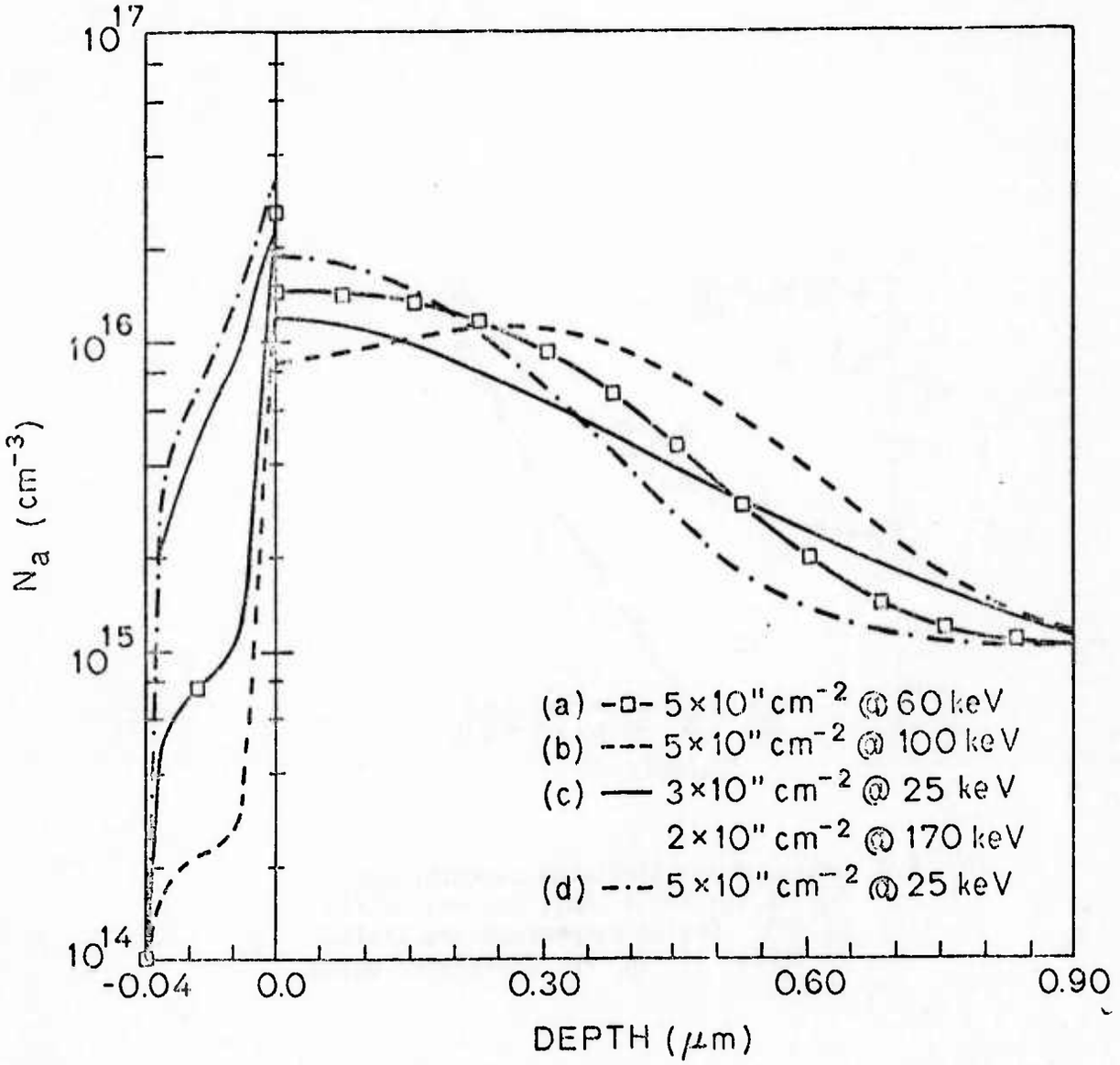


Fig. 6.4 Predicted impurity channel profile after the finished process for four channel implantation conditions of dose and energy. The profiles in the gate oxide appear on the left.

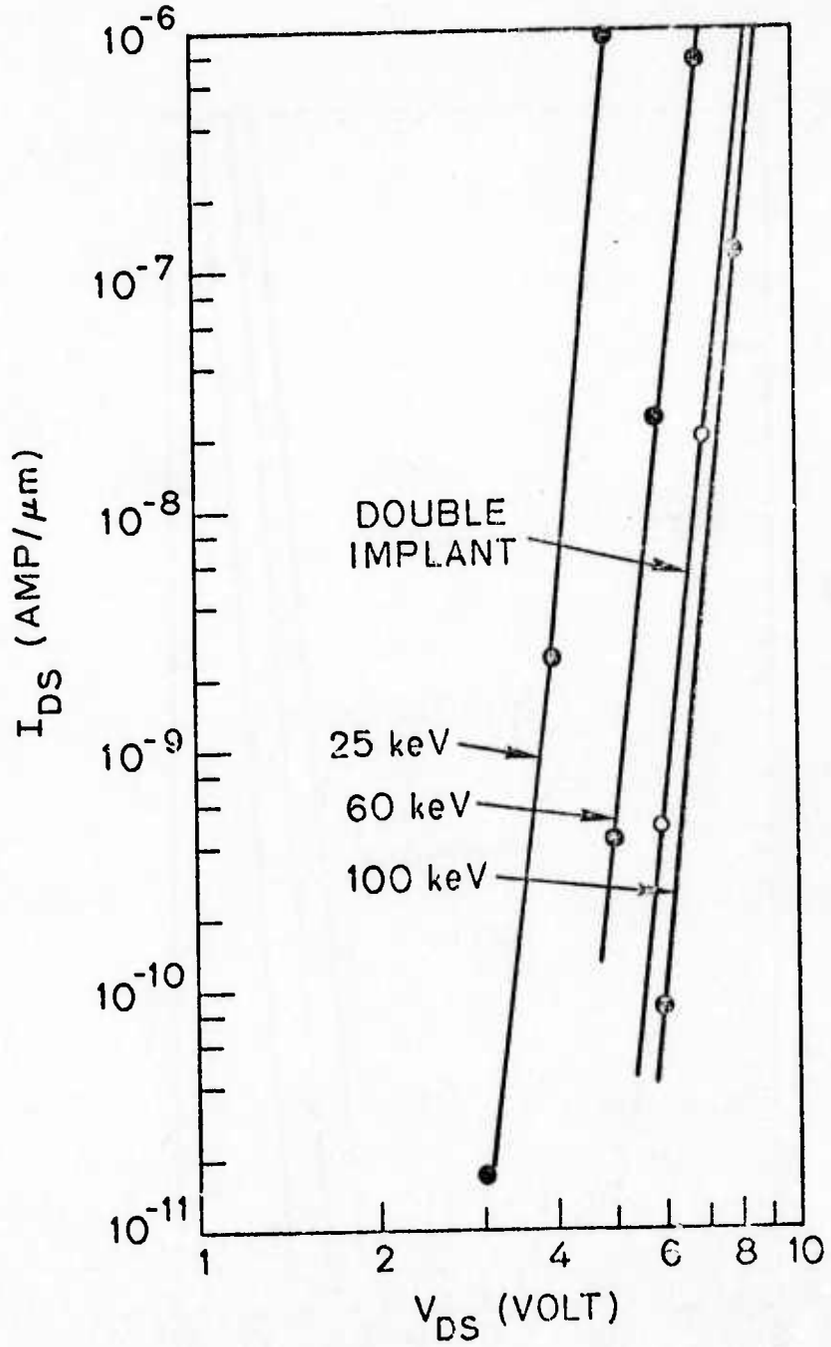


Fig. 6.5 Simulated current density for varying V_{DS}, V_{BS} = 0 V, and $V_{GS} = V_T - 1$ V. Four channel doping conditions were used as shown. T_{ox} = 0.04 μm, L_{eff} = 1.0 μm, W = 47 μm, and X_j = 0.5 μm.

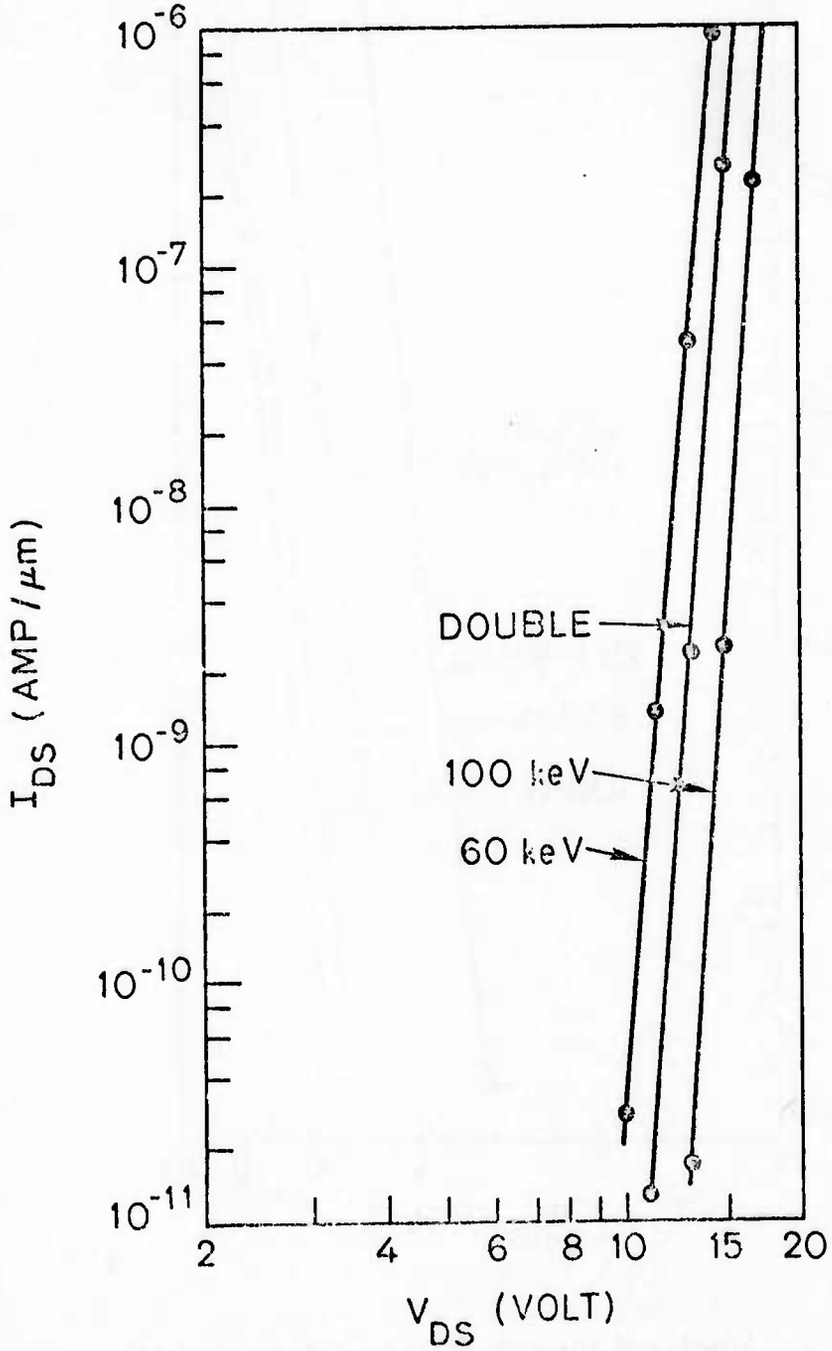


Fig. 6.6 Simulated current density for varying V_{DS}, V_{BS} = 0 V, and V_{GS} = V_T - 1 V. Three channel doping conditions were used as shown. T_{ox} = 0.04 μm, Leff = 1.0 μm, W = 47 μm, and X_j = 0.3 μm.

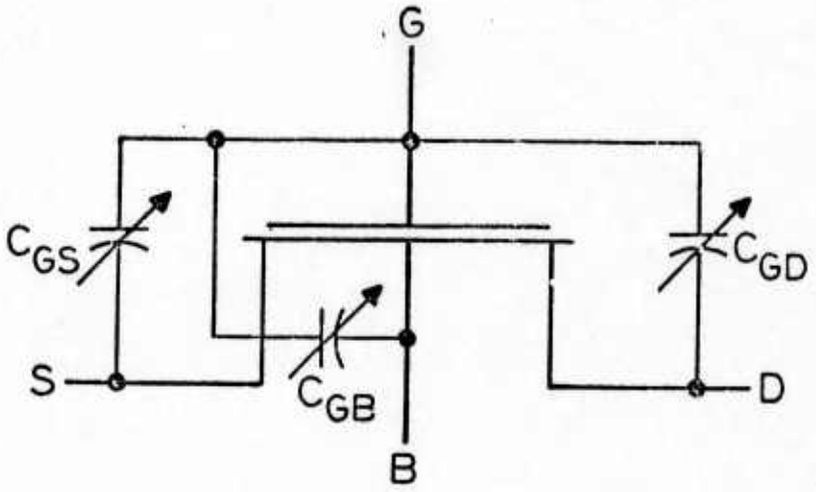


Fig. 6.7 Meyer gate-capacitance model of MOS transistor.

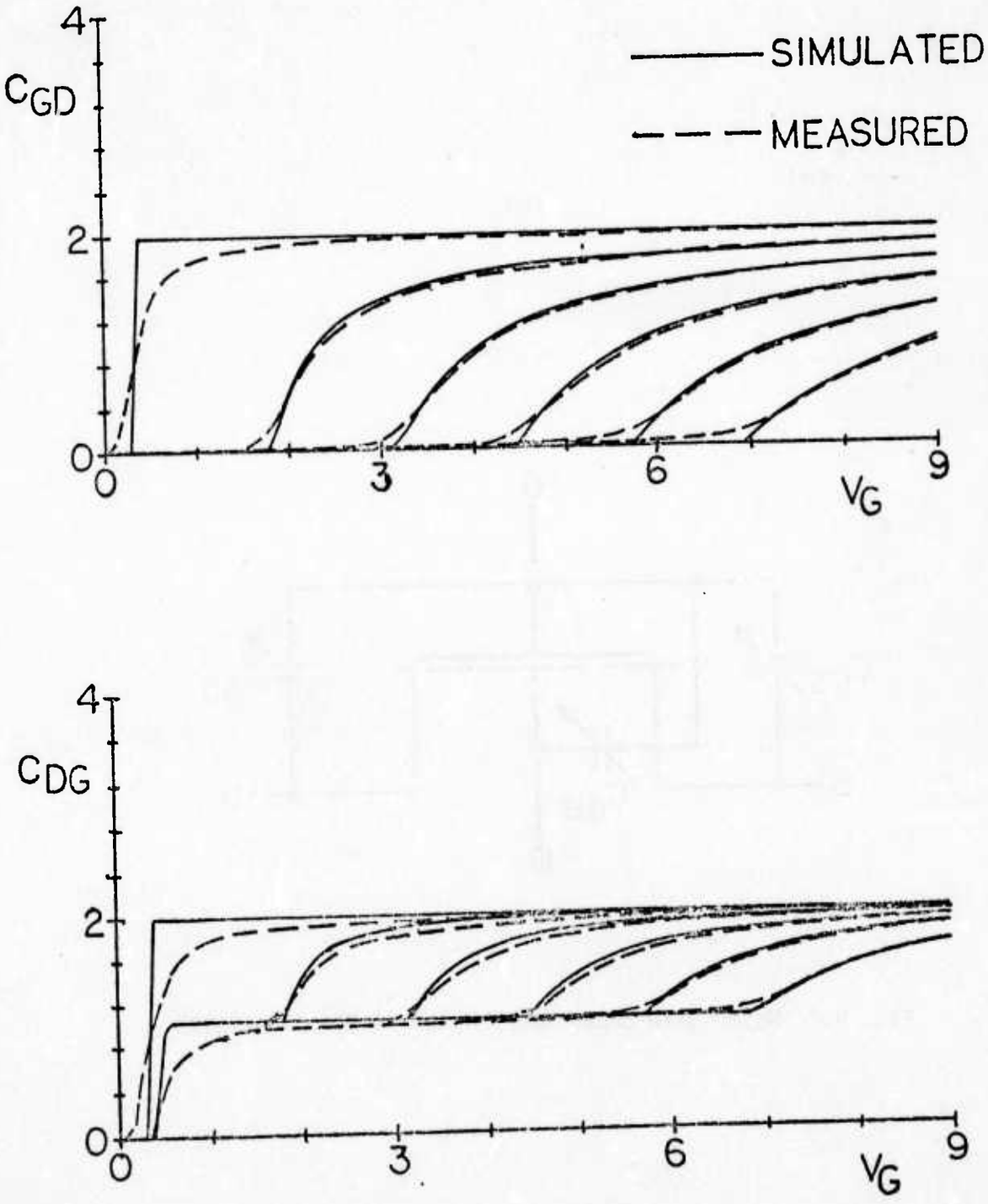
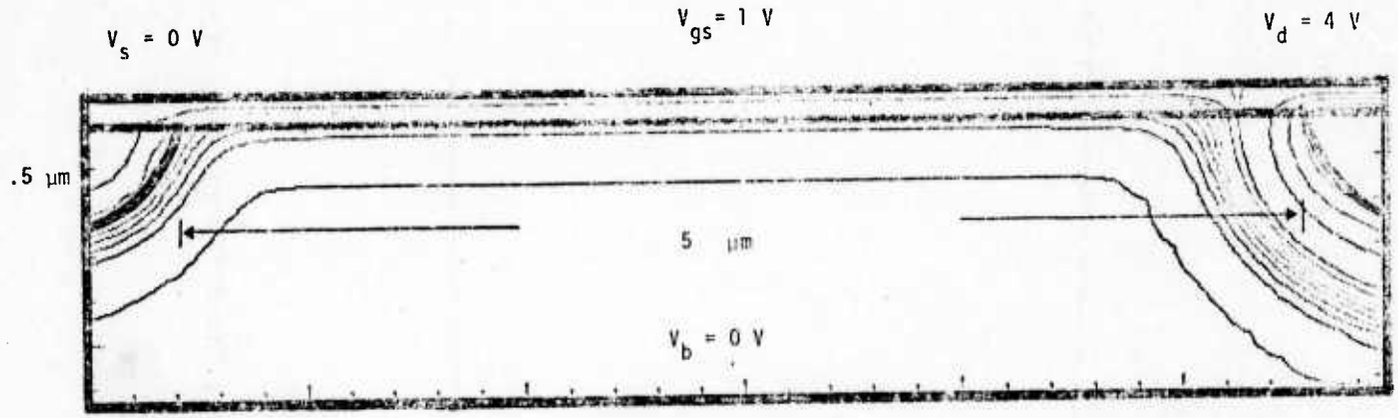


Fig. 6.8 Simulated and measured gate-to-drain capacitances.



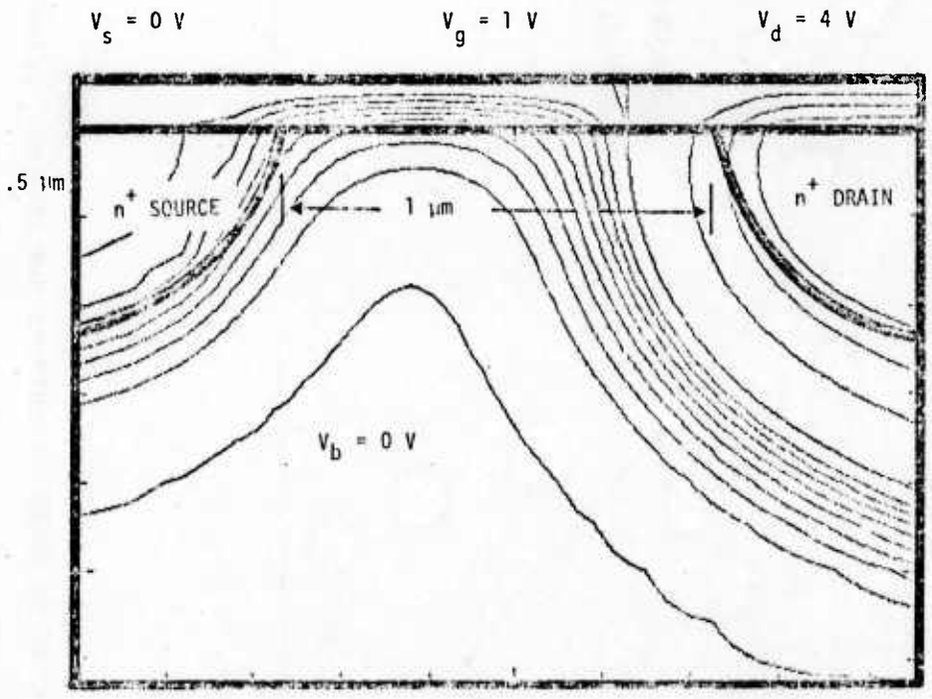


Fig. 6.9 The contour plots of the potential distributions of the long-channel (L_{eff} = 5 microns) and short-channel (L_{eff} = 1 micron) devices.

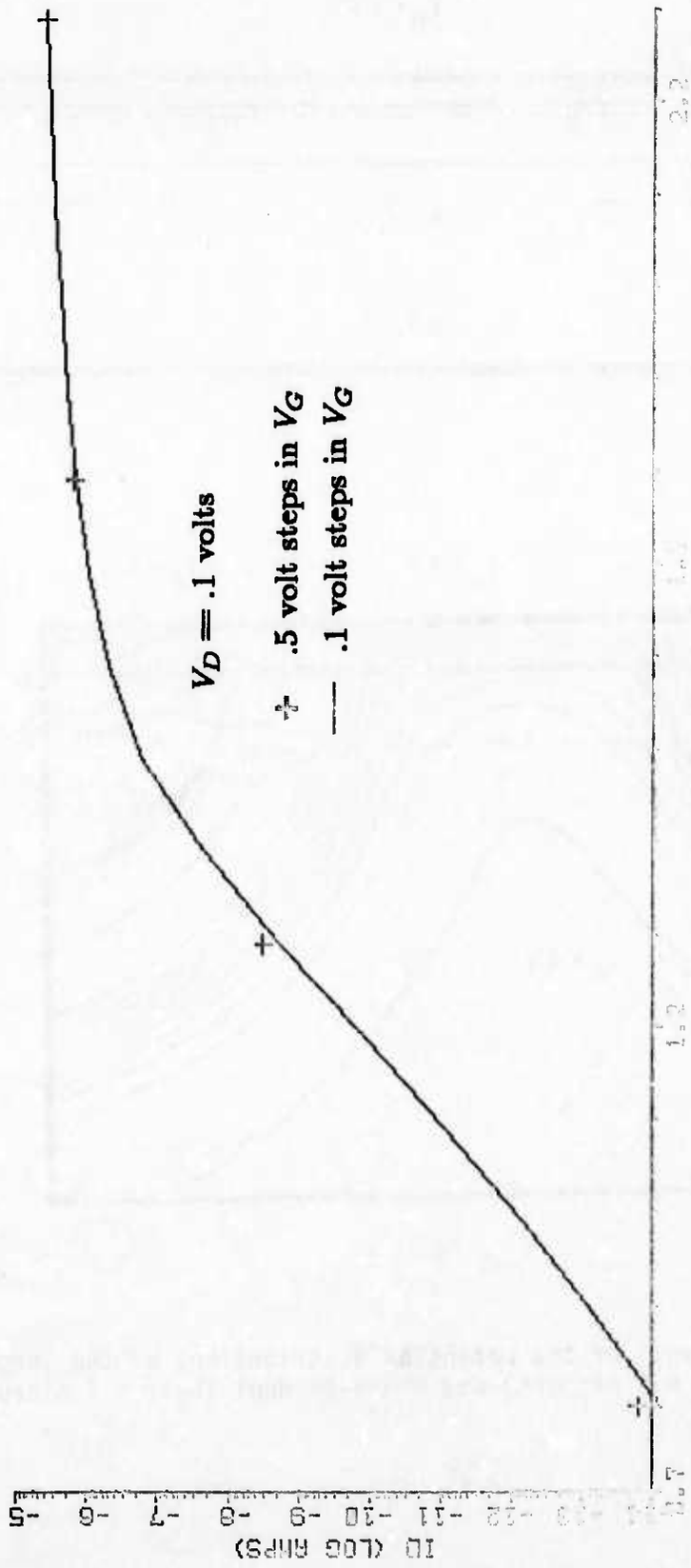
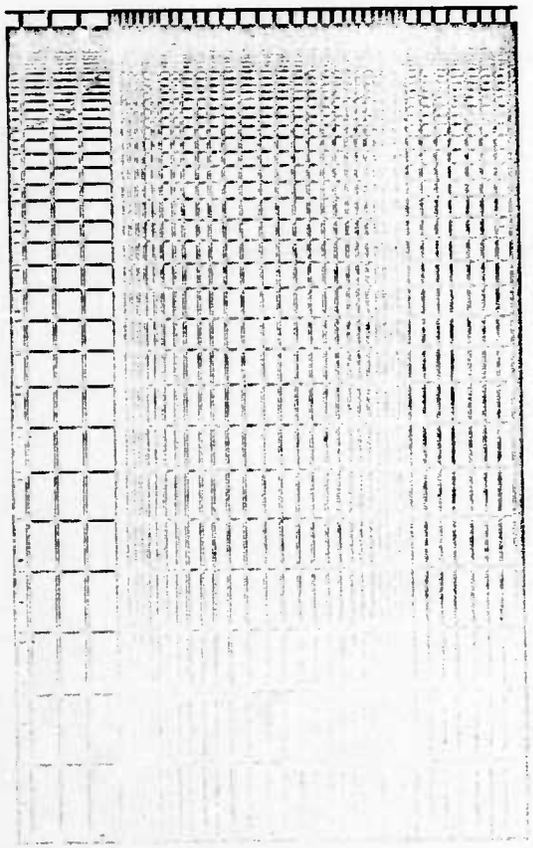


Fig. 6.10 Discrepancies in CADDET solutions due to insufficient convergence for different step sizes.



CADDET mesh

$t_{ox} = .04\mu$

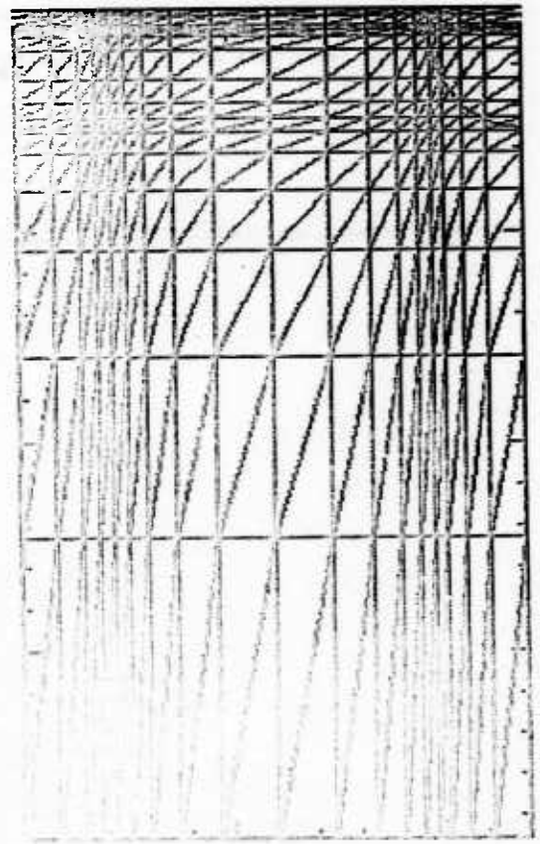
$h = 2.0\mu$

$L = 1.2\mu$

$N_{S,D} = 10^{19} cm^{-3}$

$\sigma_{S,D} = .09\mu$

$N_B = 10^{15} cm^{-3}$



TANDII mesh

Fig. 6.11 Comparison of CADDET and TANDII meshes for sample MOS transistor.

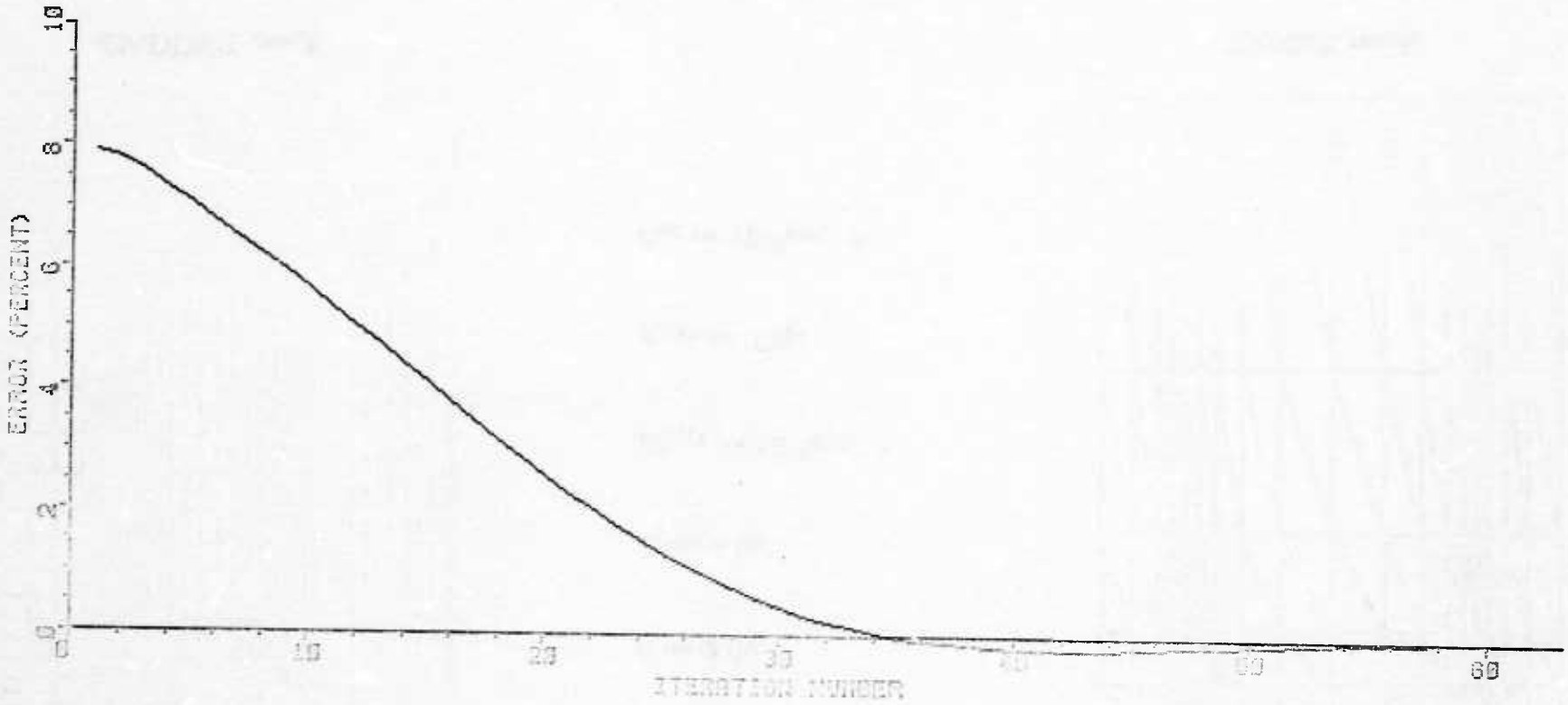


Fig. 6.12 Drain current convergence path (using projection and single-Poisson acceleration methods).

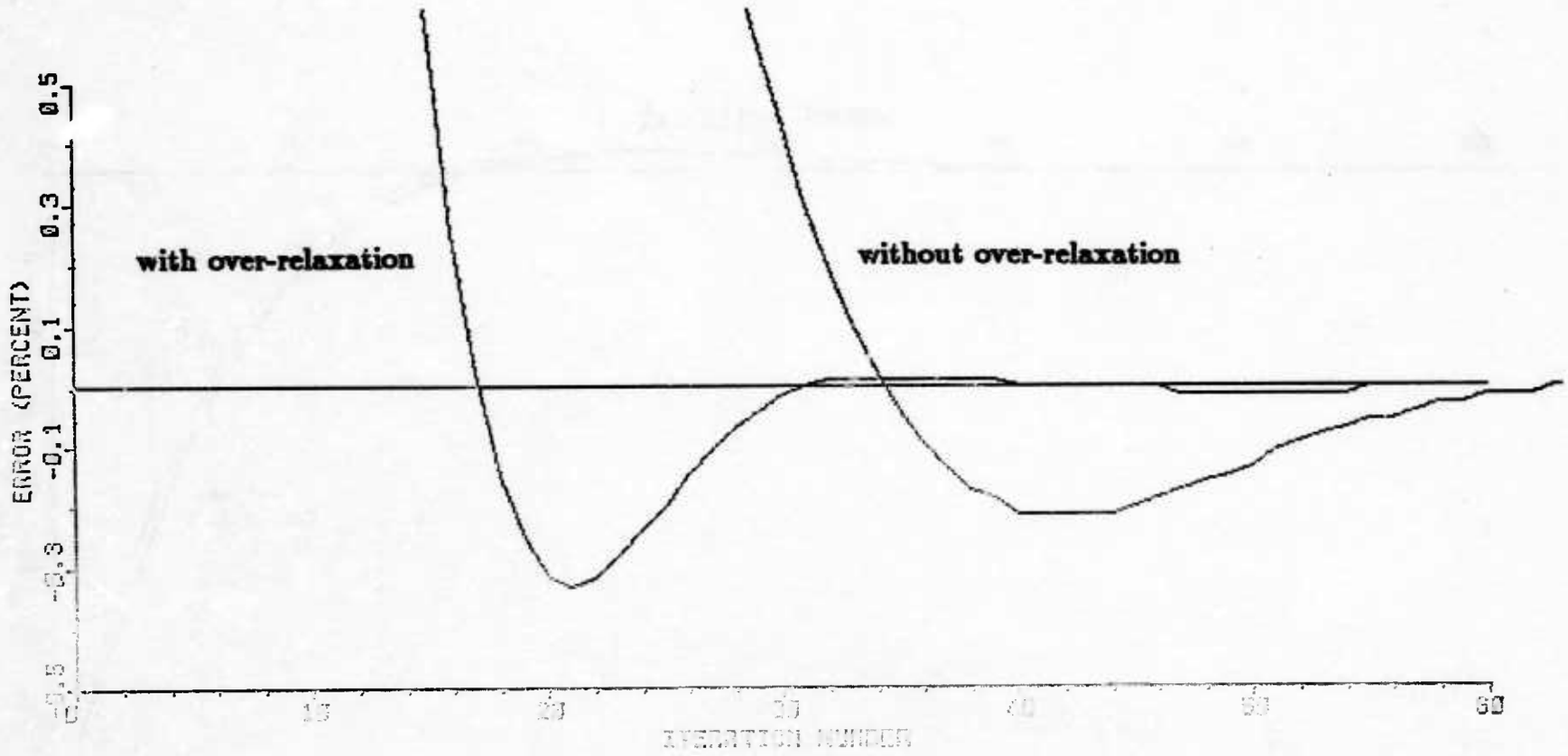


Fig. 6.13 Acceleration of the drain current convergence rate using over-relaxation.

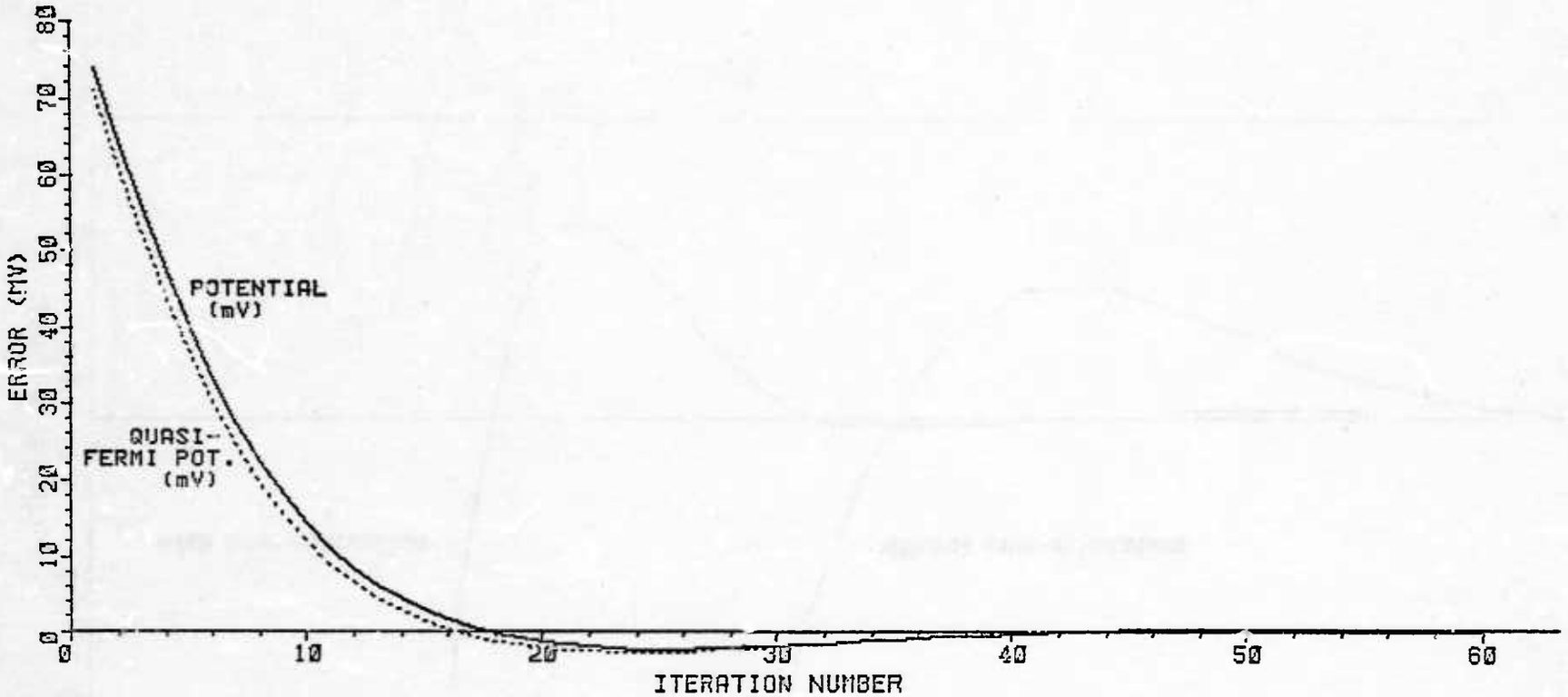


Fig. 6.14 Tracking of the potential and quasi-Fermi potential convergence at a node in the inverted channel.

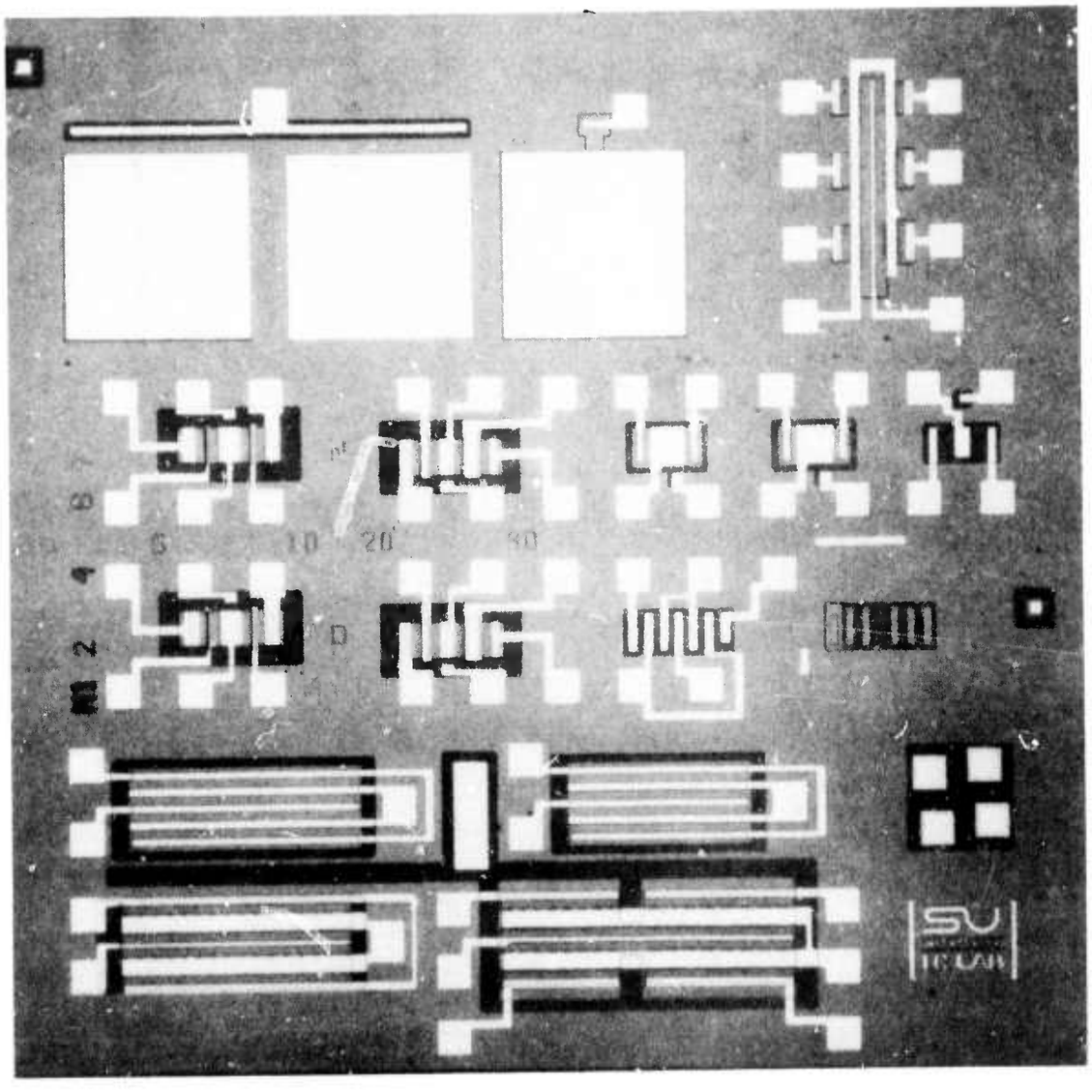


Fig. 6.15 Microscope picture of the test devices.

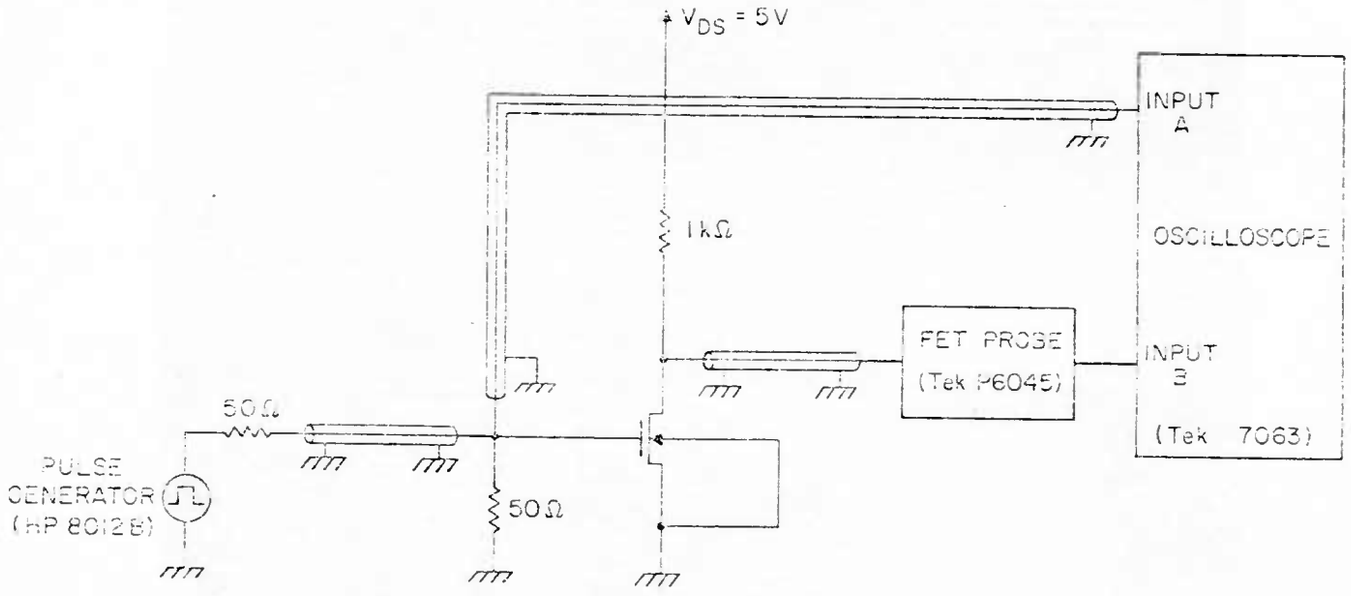


Fig. 6.16 The schematic diagram of the measurement system for MOS drain currents and delay times.

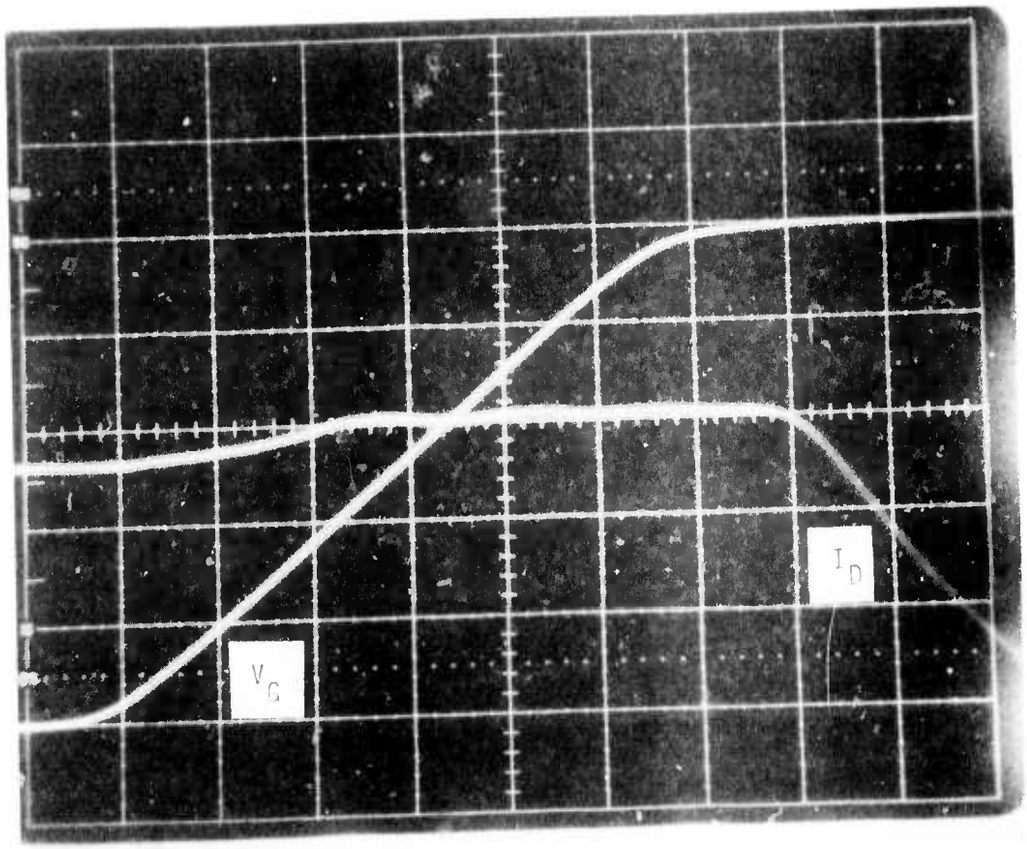
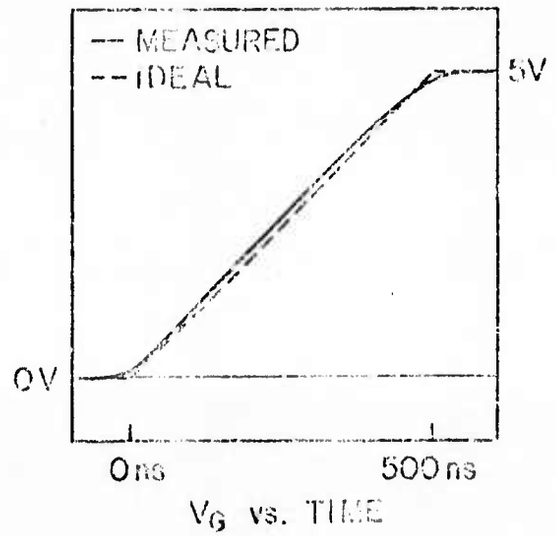
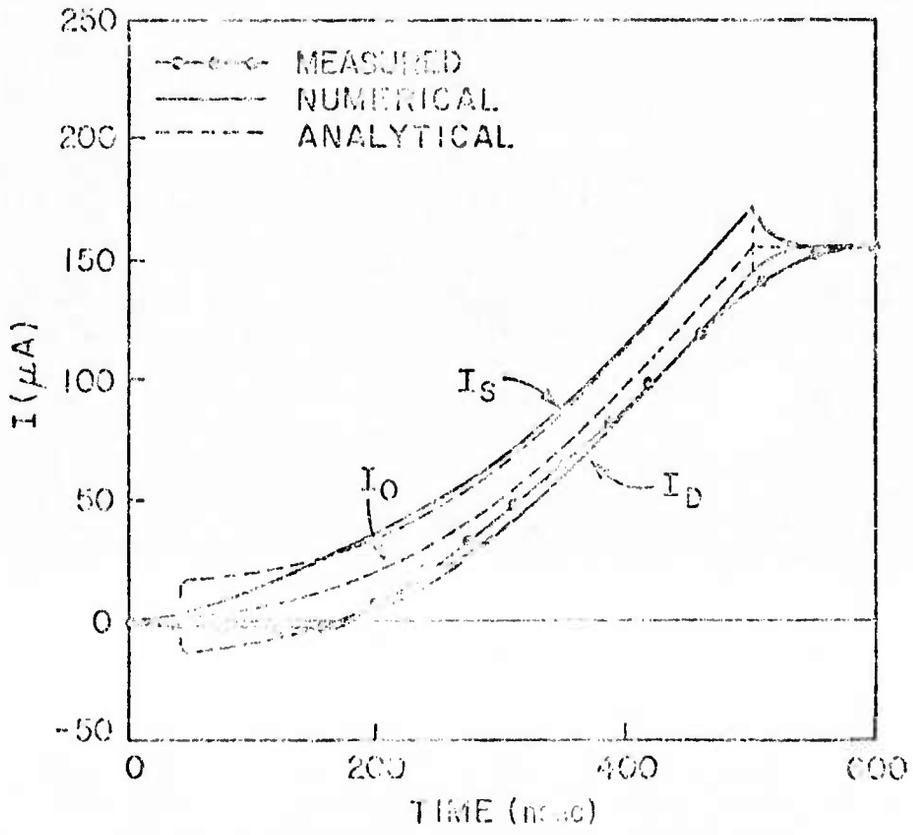


Fig. 6.17 Oscilloscope picture of the input gate signal and output drain currents. One division in the horizontal axis is 10 nsec. One division in the vertical axis is 1 V for the gate signal and 50 microamps for the drain currents.



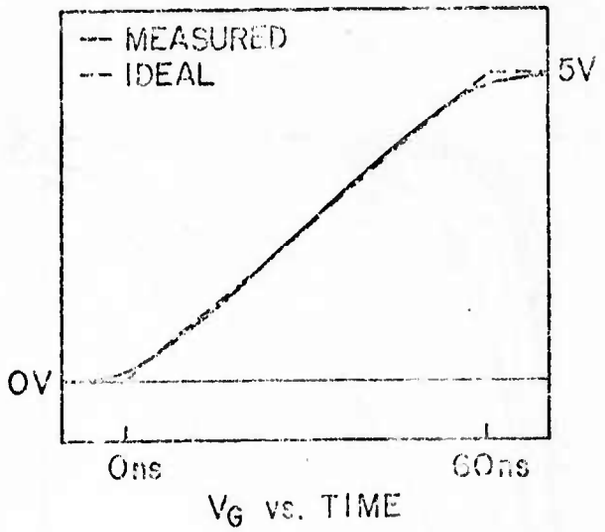
(a)



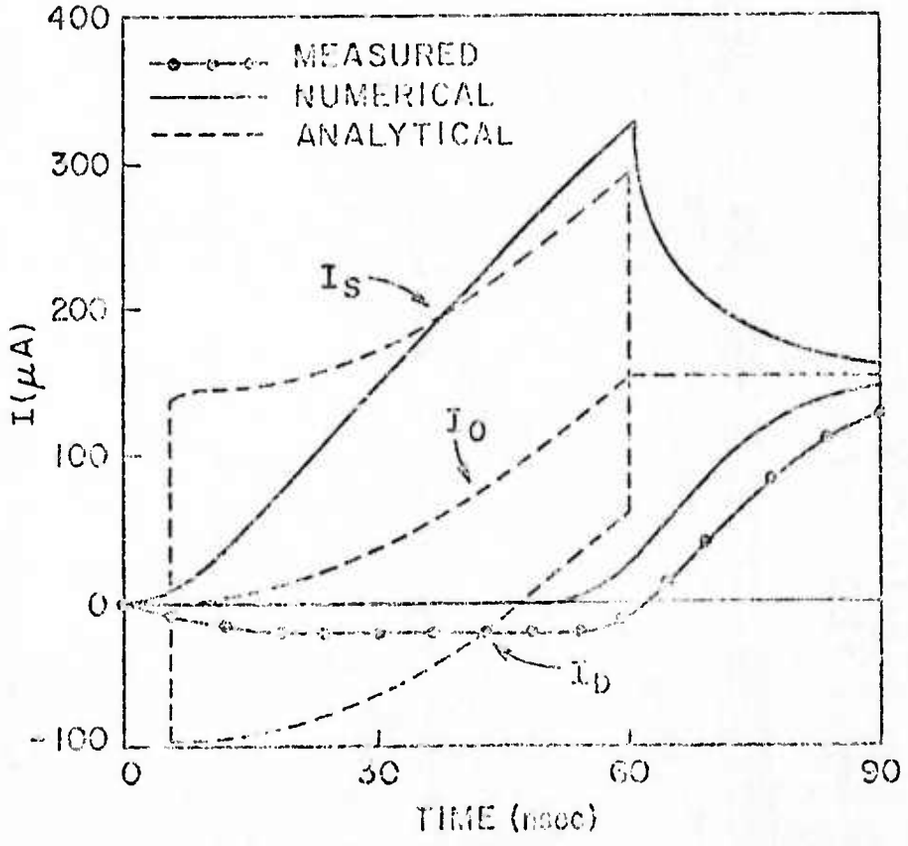
(b)

Fig. 6.18(a) Waveforms of the measured gate voltage and ideal ramp used for simulations.

(b) Comparisons of the measurements, simulations and analytical calculations of source and drain current waveforms for the 500 nsec rise-time input gate signal. I_O is the DC transport current.



(a)



(b)

Fig. 6.19(a) Waveforms of the measured gate voltage and ideal ramp used for simulations.

(b) Comparisons of the measurements, simulations and analytical calculations of source and drain current waveforms for the 60 nsec rise-time input gate signals. I_O is the DC transport current.

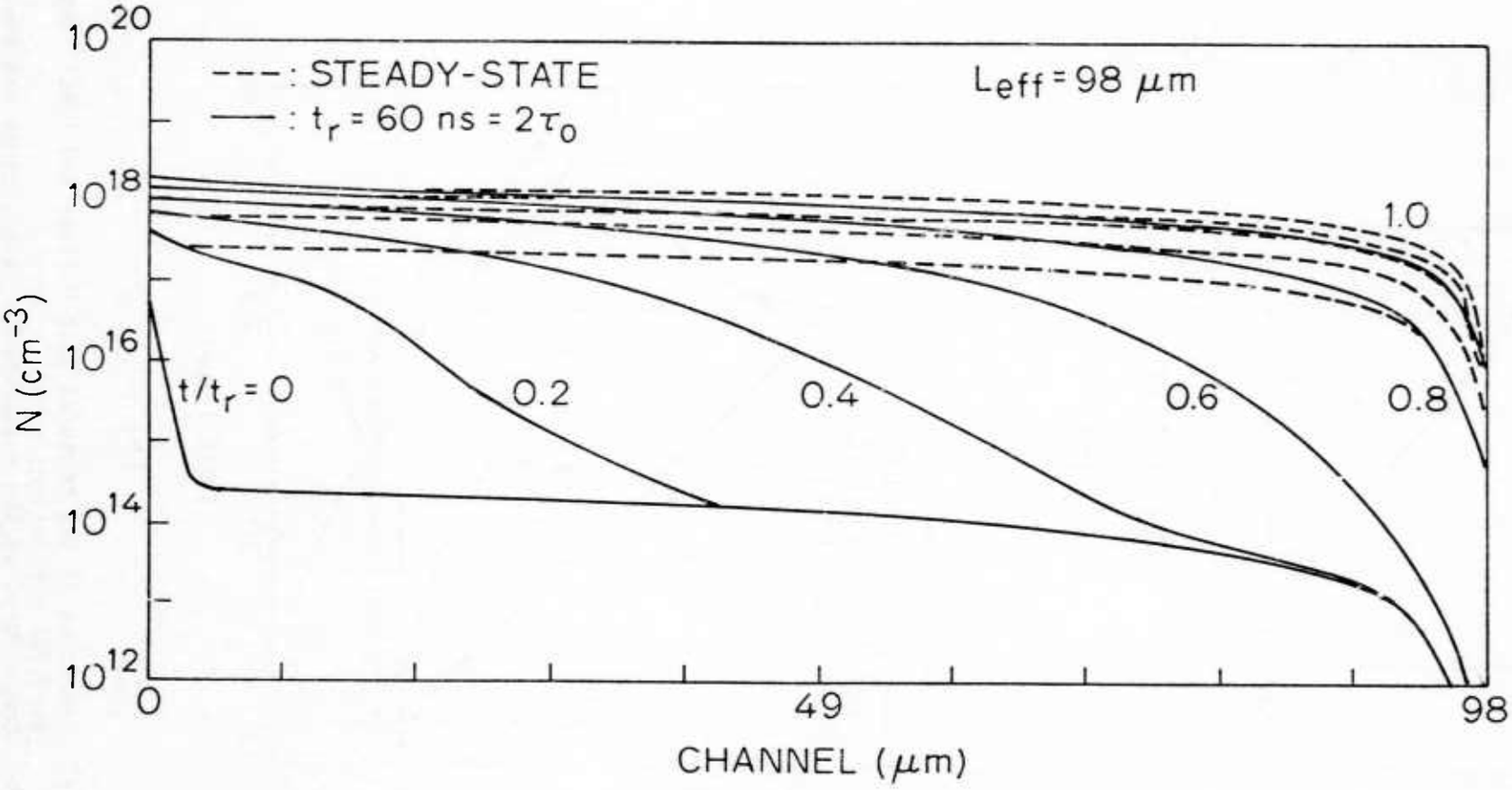


Fig. 6.20 Electron distributions of the static and dynamic cases are plotted along the channel in the log scale.

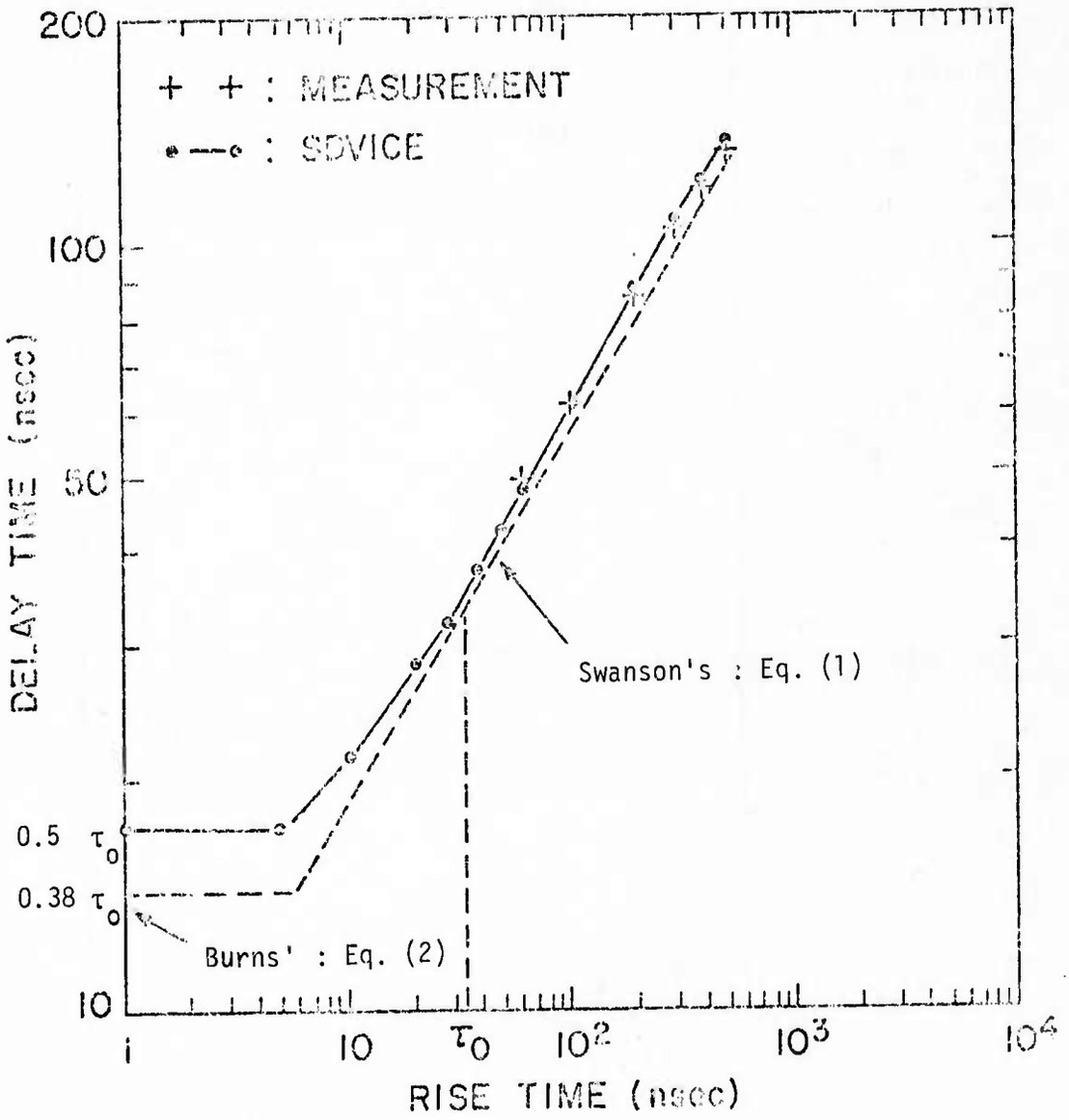


Fig. 6.21 The measurements and simulations of delay times are plotted against the rise time. The effective channel length of the device is 98 microns. Substrate doping is $8 \times 10^{15} cm^{-3}$. V_{DS} is 5 V. Predictions of Eq. (6.26) and (6.27) are also plotted for comparison.

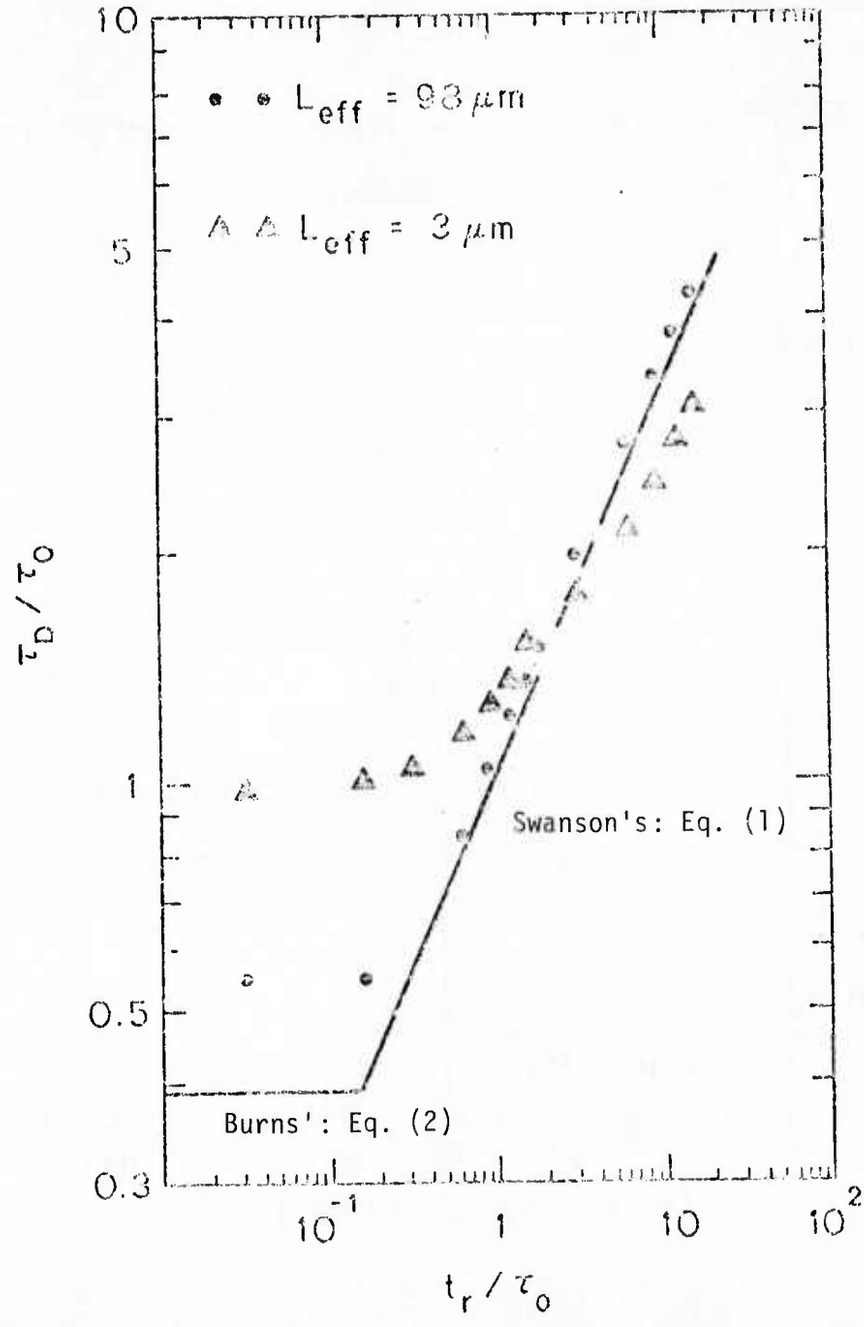
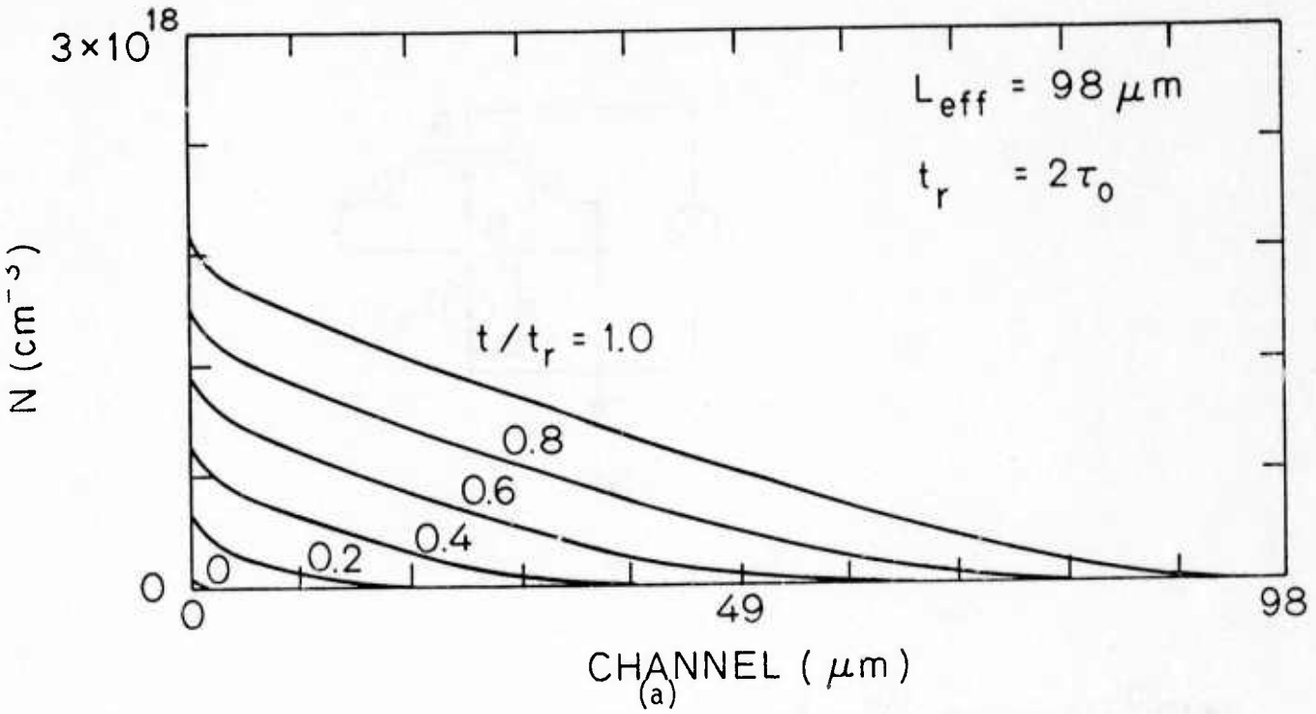


Fig. 6.22 Delay time is plotted against rise time. Both quantities are normalized by τ_o to compare the long and short-channel cases.



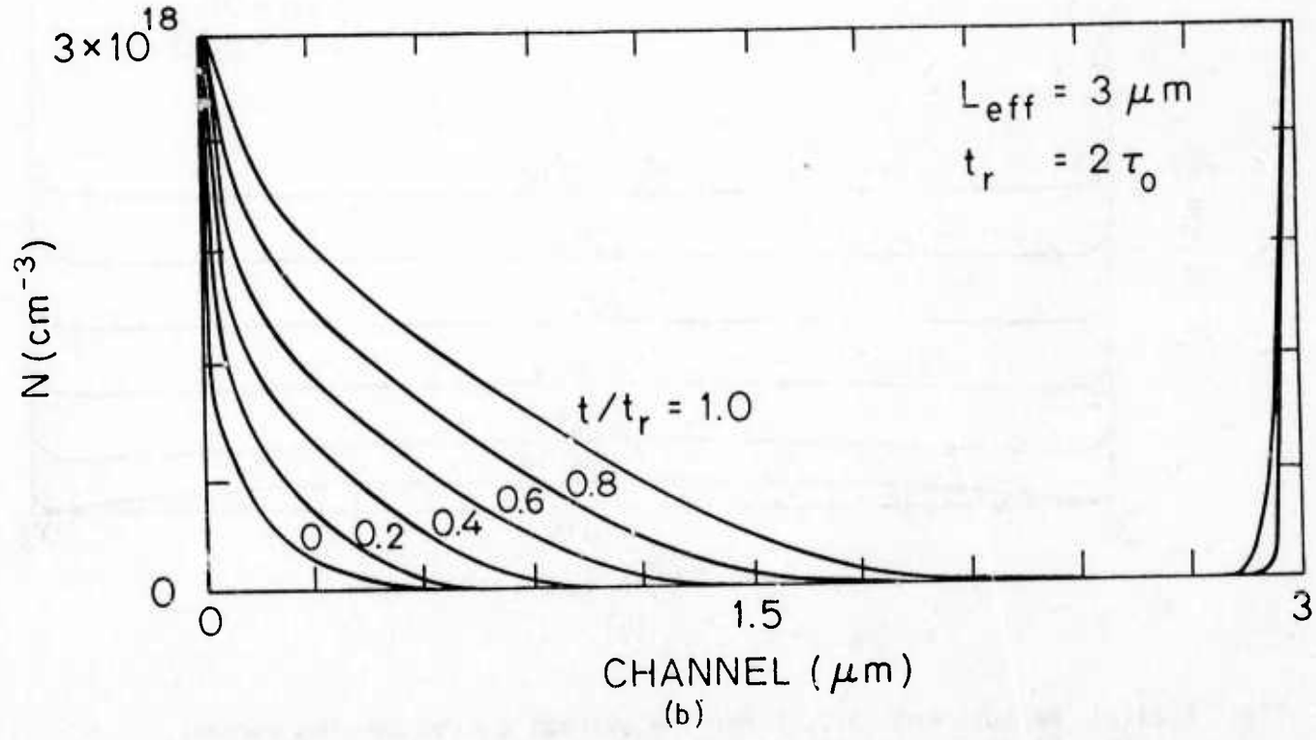
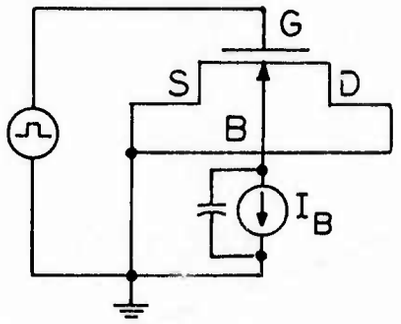
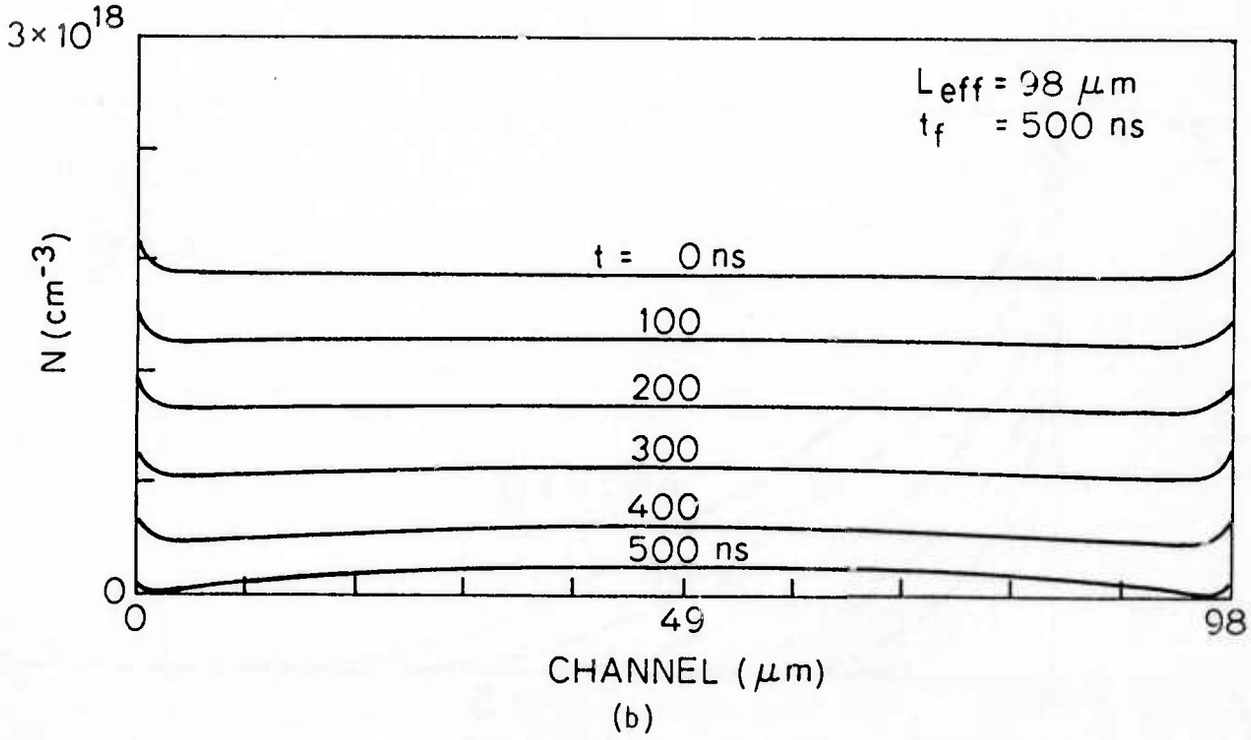


Fig. 6.23 Electron distributions in the turn-on transient are plotted along the channel in the linear scale for (a) the long-channel (L_{eff} = 98 microns) and (b) short-channel case (L_{eff} = 3 microns). It shows the electron movements during the channel transit.



(a)



(b)

Fig. 6.24(a) Measurement circuit for the average charge pumping current.
(b) Electron distributions in the turn-off transient are plotted along the channel in the linear scale. The bias conditions are same as (a).

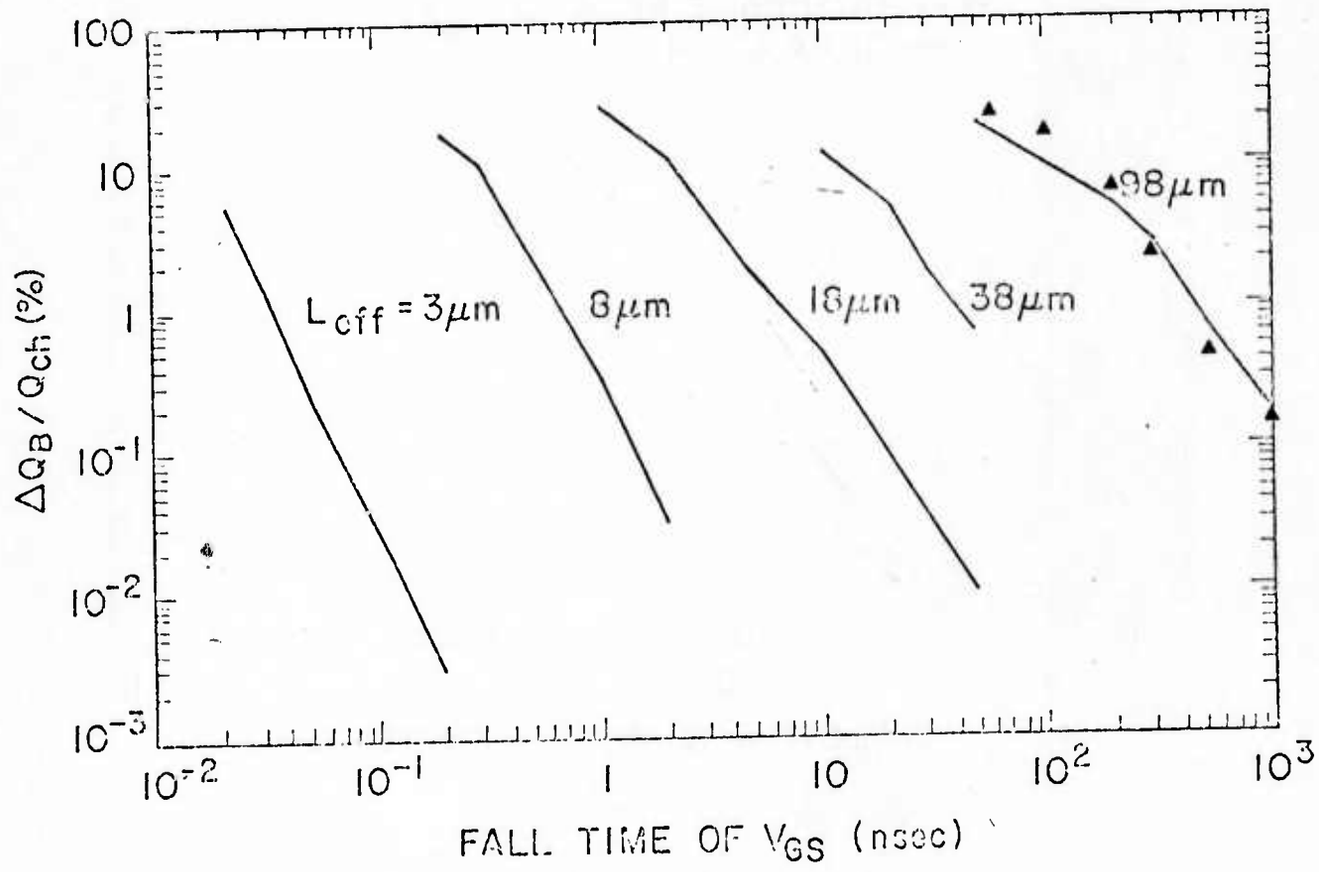


Fig. 6.25 The percentage of the pumped charge to the substrate per cycle is simulated and plotted against the fall time of the gate signal in log-log scale. The effective channel length are ranging from 3 microns to 98 microns. V_{GS} is from 0 to 5 V. The pulse frequency is 250 KHz.

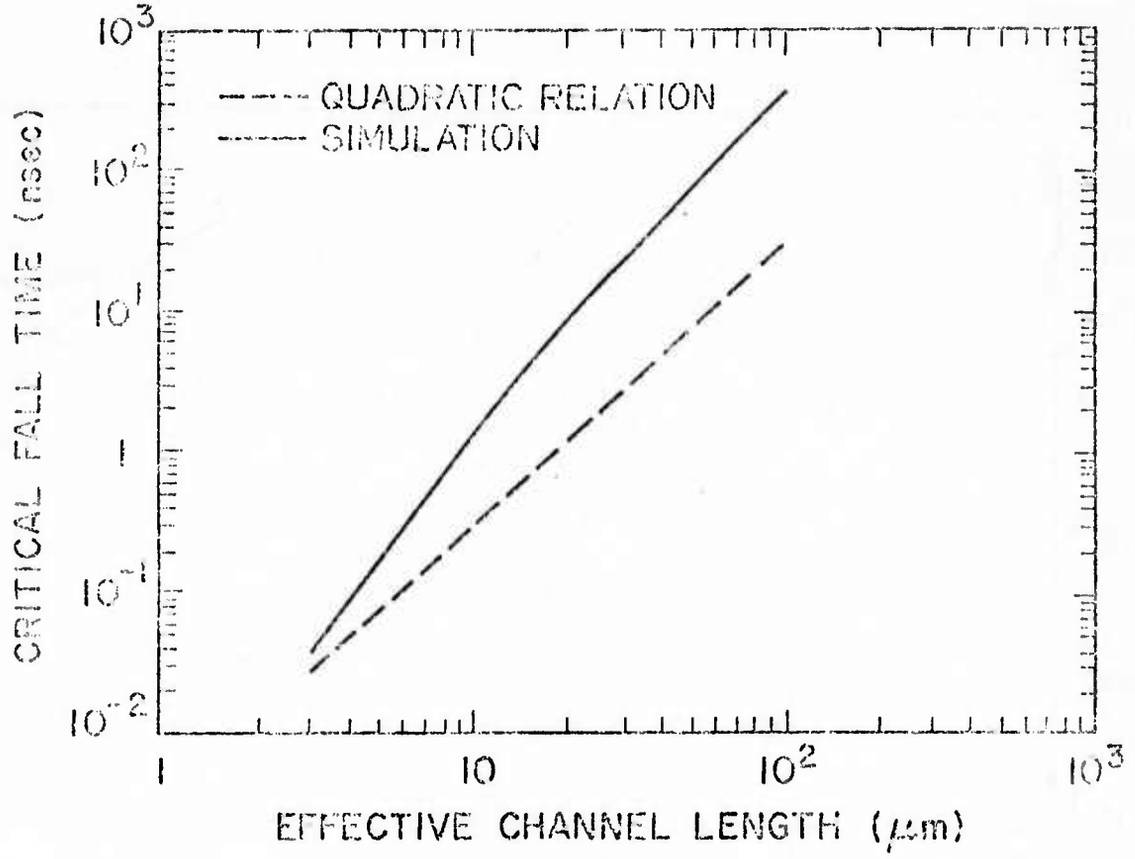
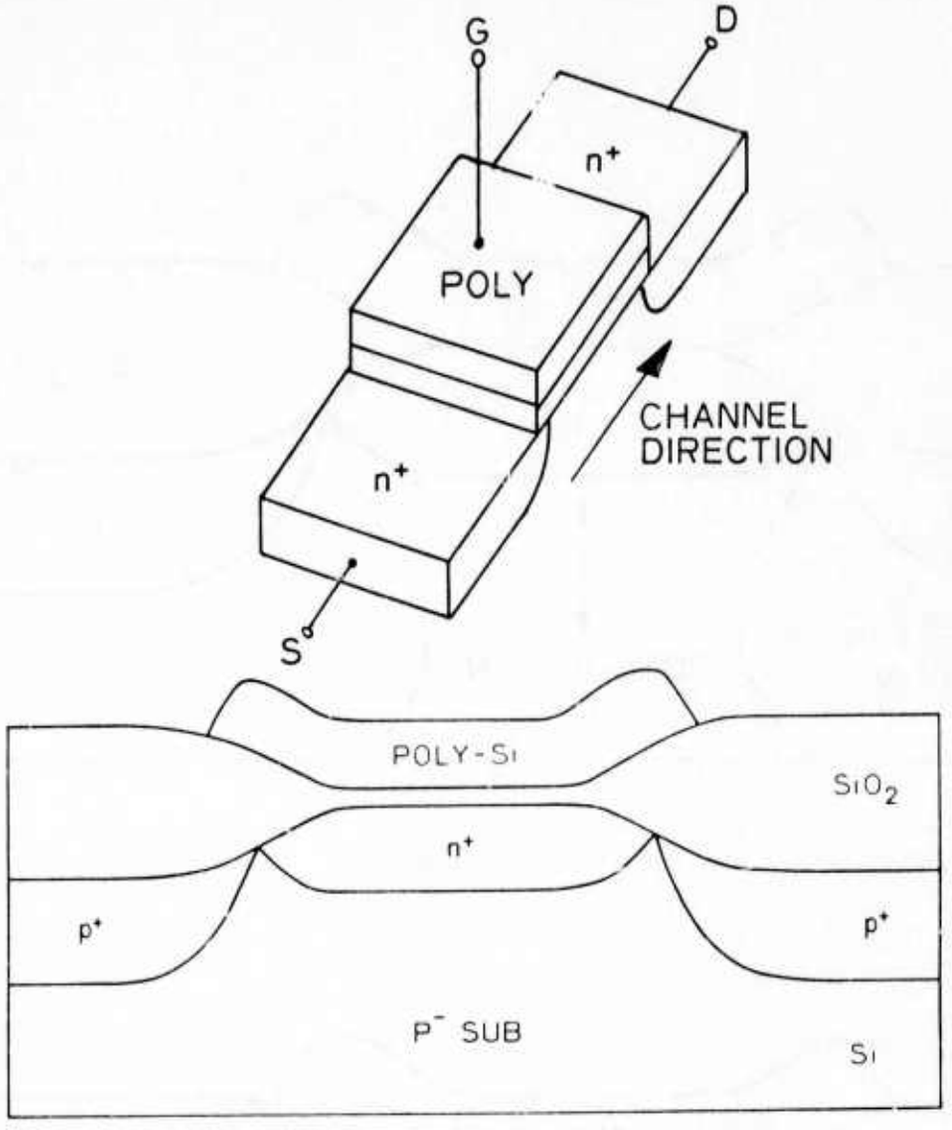
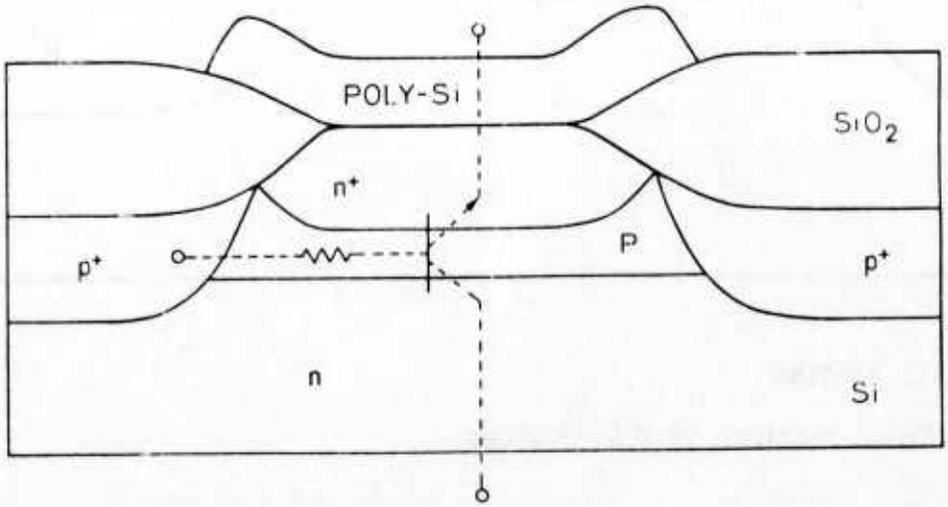


Fig. 6.26 Critical fall times are extracted from Fig. 12 and plotted against the effective channel length in the log-log scale. Quadratic dependence is plotted in a dotted line for comparison.

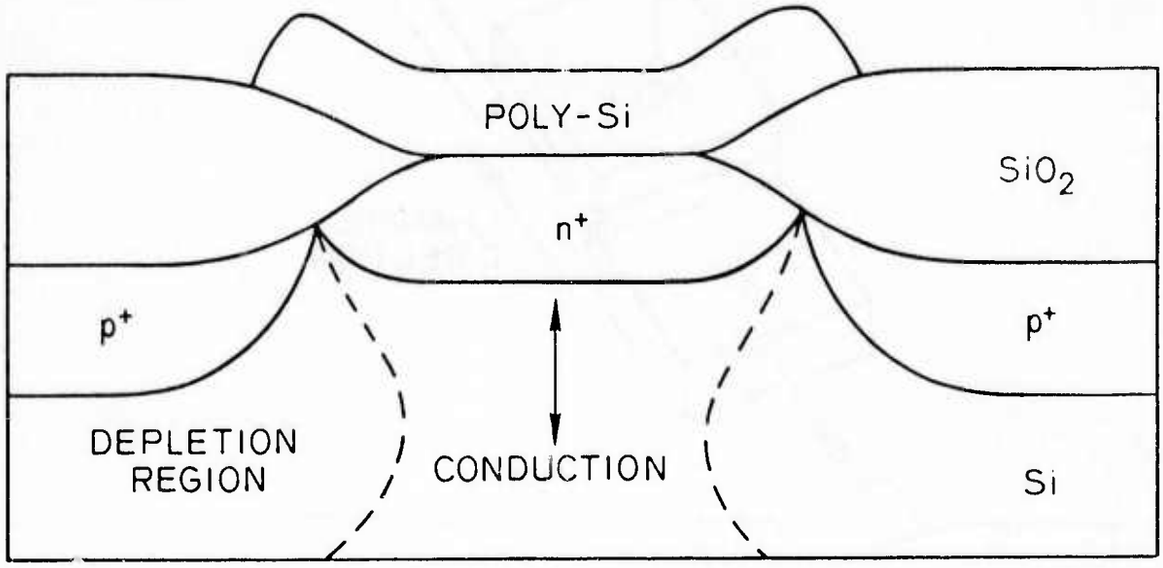


(a) NMOSFET Cross Section in the Width Direction

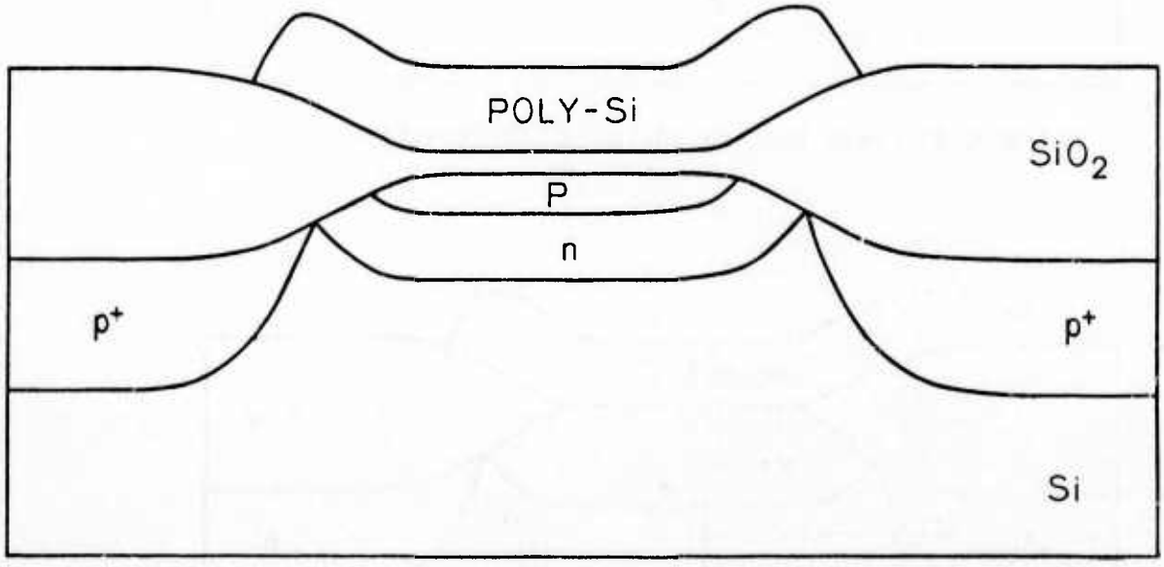


(b) Bipolar Transistor Cross Section

Fig. 6.27 Cross sections of VLSI devices.



(c) Static-Induction Transistor Cross Section



(d) TIDG Cross Section

Fig. 6.27 Cross sections of VLSI devices.

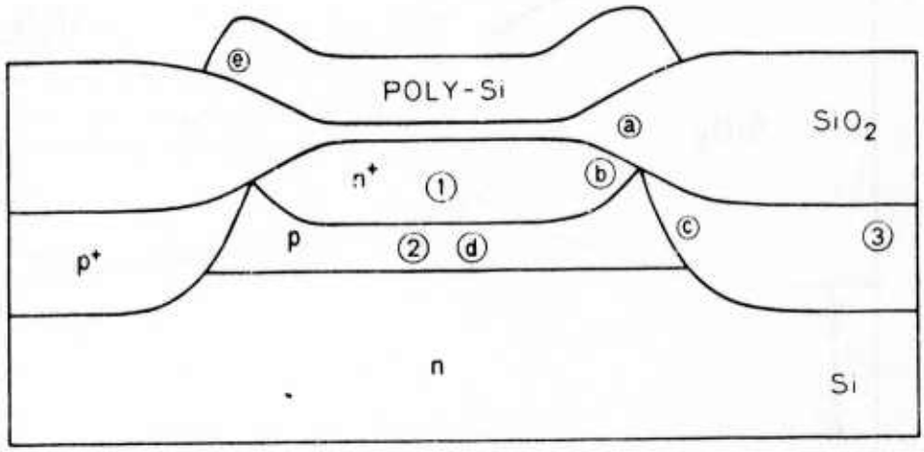


Fig . 6.28 Generalized Cross Section.

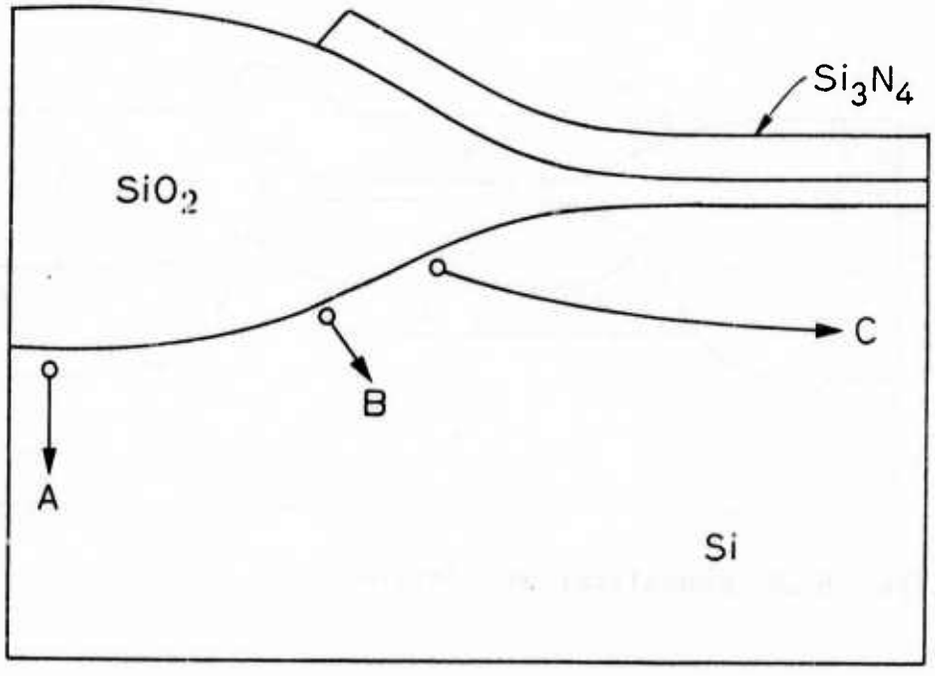


Fig. 6.29 Two-Dimensional Flow of Point Defects during Local Oxidation.

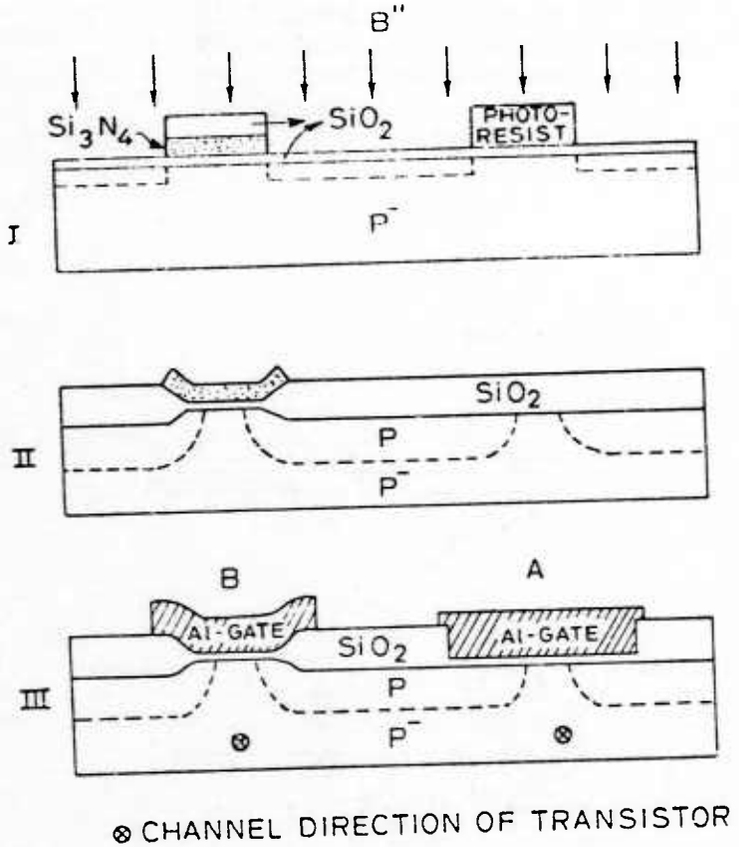


Fig.6.30 Fabrication sequence and final device cross section (I) Si_3N_4 island definition, implant mask definition (PR) and boron implant; (II) Field oxidation in wet O_2; after II, n^+ source, drain are diffused (not shown); (III) Gate oxidation and metal gate definition.

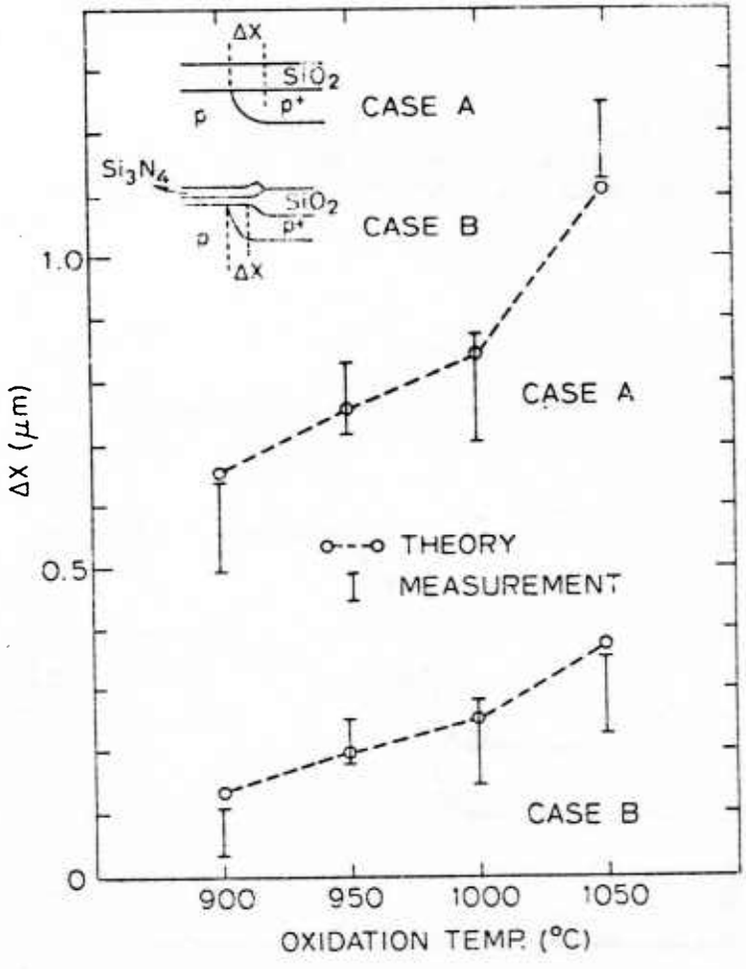


Fig. 6.31 Lateral profile transition length Δx vs. oxidation temperature. (Si_3N_4 = 1000 Å, SiO_2 = 6000 Å).

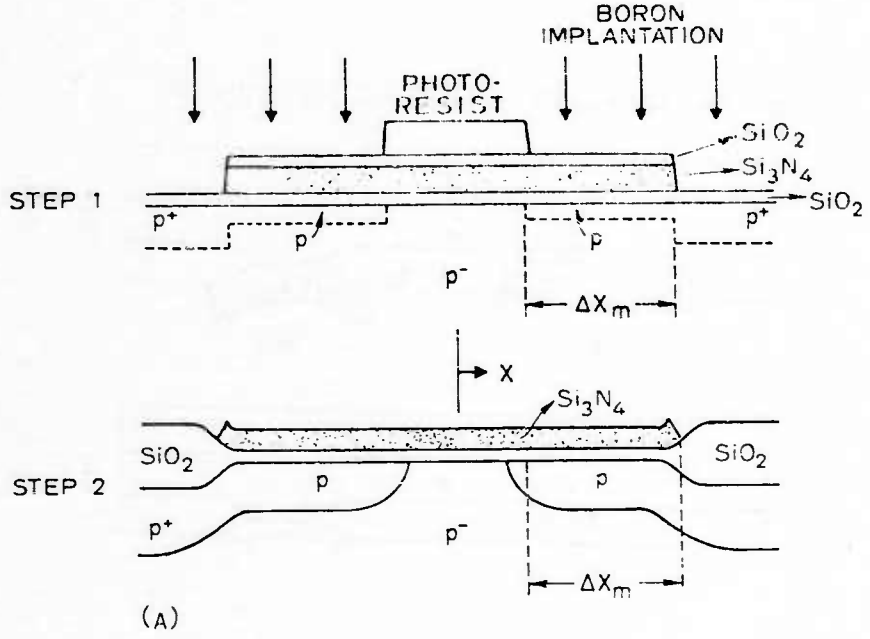


Fig. 6.32 Fabrication Sequence of Test Transistors; (I) Si_3N_4 islands definition, implant mask definition (PR) and boron implant; (II) Field Oxidation in dry O_2; after (II), a similar sequence as in Fig. 4 is followed.

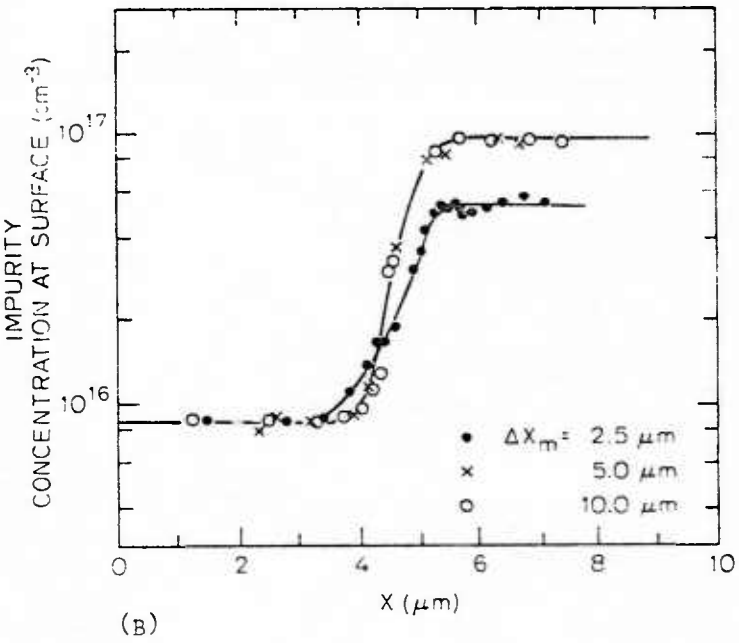


Fig. 6.33 Surface Impurity Profiles under Si_3N_4 after local oxidation at a distance X_m.