A UNITED STATES DEPARTMENT OF COMMERCE PUBLICATION

NBS SPECIAL PUBLICATION 337

Silicon Device Processing

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards

UNITED STATES DEPARTMENT OF COMMERCE • Maurice H. Stans, Secretary
NATIONAL BUREAU OF STANDARDS • Lewis M. Branscomb, Director

Silicon Device Processing

Proceedings of a Symposium
Held at Gaithersburg, Maryland
June 2-3, 1970

Charles P. Marsden, Editor

Institute for Applied Technology
National Bureau of Standards
Washington, D.C. 20234

Under the Sponsorship of Committee F-1
of the
American Society for Testing and Materials
and
The National Bureau of Standards

U.S. National Bureau of Standards Special Publication 337

Nat. Bur. Stand. (U.S.), Spec. Publ. 337, 467 pages (Nov. 1970)
CODEN: XNBSA

Issued November 1970

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402
(Order by SD Catalog No. C 13.10:337), Price $5.50

Library of Congress Catalog Card Number: 74-608543

FOREWORD

The 20th anniversary of the development of the alloy junction, and the 15th anniversary of both commercial silicon devices and the founding of ASTM Committee F-1, was an appropriate time for this Symposium. NBS welcomed the opportunity to cosponsor, with Committee F-1, a meeting that should produce better understanding of both currently used and newer methods of silicon material characterization.

The rate of development of silicon devices has often exceeded the ability to measure their properties. Committee F-1 has attempted to provide the leadership necessary for standardization of measurement methods, and, through symposia, vehicles for promoting discussion and understanding of such measurement methods.

Standard methods for material characterization are a major factor in the producer-consumer interface, and provide a common language in procurement. Standard methods also should provide the producer with greater product uniformity, and the consumer with increased reliability. The Federal Government, as a major consumer of silicon devices, has a profound interest in the advancement of measurement methodology, an area to which NBS contributes strongly through the efforts of its technical staff.

Lewis M. Branscomb
Director

PREFACE

This Symposium on Silicon Device Processing was held on June 2 - 3, 1970 at the National Bureau of Standards under the cosponsorship of this Bureau and Committee F-1 of the American Society for Testing and Materials. It consisted of seven sessions as detailed in the Index on page vi to viii.

The objective of the Symposium was to provide an opportunity for engineers and applied scientists actively engaged in the silicon device technology field to discuss the most advanced measurement methods for process control and materials characterization. The basic theme of the meeting was to stress the interdependence of measurements techniques, facilities, and materials as they relate to the overall problems of improving and advancing silicon device sciences and technologies.

Speakers were generally invited to talk on specific phases of the subject matter of the session, which usually concerned present work for which they had shown unusual competence and could be considered in the forefront of their field. The General Session consisted entirely of invited tutorial papers.

It is hoped that this proceedings will be an effective reference in its field. The information given in the evaluation of current measurement techniques as well as the exhaustive efforts in assembling evidence to prove their validity, should prove of lasting value.

Charles P. Marsden,
Editor

SYMPOSIUM COMMITTEE

D. E. KOONTZ, Symposium Chairman
Bell Telephone Laboratories
Murray Hill, N. J.

W. M. BULLIS, Member Organizing Committee
National Bureau of Standards
Washington, D. C.

W. B. BURFORD III, Committee F-1 Vice Chairman
Westinghouse Electric Corporation
Baltimore, Md.

J. W. CARLSON, Chairman Session II-A
Monsanto Co.
St. Peters, Mo.

J. A. DWYER, Society Representative
American Society for Testing and Materials
Philadelphia, Pa.

E. E. GARDNER, Chairman Session II-B
IBM
Essex Junction, Vt.

F. GITTLER, Chairman Session III-A
Bell Telephone Laboratories
Allentown, Pa.

D. C. GUPTA, Chairman Session III-B
General Telephone and Electronics Corp.
Woburn, Mass.

C. P. MARSDEN, Chairman Arrangements and Publications
National Bureau of Standards
Washington, D. C.

E. MENDEL, Chairman Session IV-B
IBM
Hopewell Junction, N. Y.

J. OROSHNIK, Symposium Secretary and Chairman
Session IV-A
National Bureau of Standards
Washington, D. C.

R. I. SCACE, Committee F-1 Chairman
General Electric Co.
Auburn, N. Y.

P. A. Schumann, Member Organizing Committee
IBM
Hopewell Junction, N. Y.

CONTENTS

Paper No.		Page No.
1.	Welcoming Remarks Dr. Lewis M. Branscomb, Director, National Bureau of Standards	1
2.	Some Thoughts on How We Might Improve Our Materials and Process Work Dr. D. G. Thomas, Executive Director, Bell Telephone Laboratories	3
	SESSION I - GENERAL, CHAIRMAN - D. E. KOONTZ	
3.	Crystallographic Imperfections as Related to Silicon Crystal Growth J. A. Lenard, IBM, Components Division	11
4.	A Review of Silicon Substrates Surface Preparation and Evaluation K. E. Lemons, Signetics Corporation Paper Withdrawn from Publication	-
5.	Epitaxial Growth of Silicon B. A. Joyce, Mullard Research Laboratories	19
6.	Diffusion in Silicon: Properties and Techniques C. F. Gibbon, Bell Telephone Laboratories, Inc.	21
7.	Measurement and Control of Dielectric Film Properties During Semiconductor Device Processing B. E. Deal, Fairchild Camera and Instrument	36
	SESSION II-A - EPITAXY-TECHNIQUES AND FACILITIES, CHAIRMAN - J. W. CARLSON	
8.	Equipment Considerations for Silicon Epitaxy Reactors M. L. Hammond and W. P. Cox, Hugle Industires, Inc.	51
9.	A Comparison of a Resistance Heated Reactor for Silicon Epitaxial Growth With Other Epitaxial Systems W. A. Kohler, Fairchild Camera and Instrument Corp.	60
10.	Techniques for Depositing Highly Uniform and Defect-Free Epitaxial Silicon D. C. Gupta, The Waltham Research Center of the General Telephone and Electronics Laboratories, Inc. and J. L. Porter, Sylvania Electric Products, Inc.	66
11.	Control of Thin Silicon Films Grown From Silane D. J. Dumin, RCA Laboratories	79
12.	The Growth of Submicron Single and Multilayer Silicon Epitaxy J. Simpson, A. C. Adams and M. H. Hanes, Bell Telephone Laboratories, Inc.	87
	SESSION II-B - DIFFUSION-PROPERTIES CHARACTERISTICS, CHAIRMAN - E. E. GARDNER	
13.	Techniques for Determining Surface Concentration of Diffusants J. C. Irvin, Bell Telephone Laboratories	99
14.	Current Status of the Spreading Resistance Probe and Its Application T. H. Yeh, IBM Components Division	111
15.	Incremental Sheet Resistivity Technique for Determining Diffusion Profiles R. P. Donovan and R. A. Evans, Research Triangle Institute	123
16.	Nuclear Methods for the Determination of Diffusion Profiles B. J. Masters, IBM Components Division	132
17.	Use of High-Energy Ion Beams for Analysis of Doped Surface Layers S. L. Chou, L. A. Davidson and J. F. Gibbons, Stanford Electronics Laboratories	141

Paper No.		Page No.
18.	Determination of Diffusion Coefficients in Silicon and Accepted Values M. F. Millea, Aerospace Corporation	156
	SESSION III-A - DIFFUSION-TECHNIQUES AND FACILITIES, CHAIRMAN - F. L. GITTLER	
19.	Diffusion Technology For Advanced Microelectronic Processing W. Greig, K. Cunniff, H. Hyman and S. Muller, RCA Solid State Division	168
20.	Diffusion From Doped-Oxide Sources M. L. Barry, Fairchild Camera and Instrument Corporation	175
21.	Capacitance-A Device Parameter and Tool for Measuring Doping Profiles B. R. Chawla, Bell Telephone Laboratories, Inc.	182
22.	Concentration Dependent Diffusion Phenomena P. E. Bakeman, Jr., Rensselaer Research Corporation and J. M. Borrego, Rensselaer Polytechnic Institute	184
23.	Orientation Dependent Diffusion Phenomena L. E. Katz, Bell Telephone Laboratories, Inc.	192
24.	Diffusion Inducted Defects and Diffusion Kinetics in Silicon M. L. Joshi, IBM, Components Division and S. Dash, Fairchild Semiconductor	202
	SESSION III-B - EPITAXY-PROPERTIES AND CHARACTERISTICS, CHAIRMAN - D. C. GUPTA	
25.	Limitations of Current Epitaxial Evaluations (Abstract Only) R. N. Tucker, Fairchild Semiconductor. Paper Withdrawn from Publication	223
26.	On the Interpretation of Some Measurement Methods for Epitaxially Grown Layers P. J. Severin, Philips Research Laboratories	224
27.	Thickness Measurement of Very Thin Epitaxial Layers by Infrared Reflectance P. A. Schumann, Jr., IBM Components Division	234
28.	Spreading Resistance Measurements on Buried Layers in Silicon Structures R. G. Mazur, Westinghouse Research Laboratories	244
29.	Variations of a Basic Capacitance-Voltage Technique for Determination of Impurity Profiles in Semiconductors W. C. Niehaus, W. vanGelder, T. D. Jones and P. Langer, Bell Telephone Laboratories	256
31.	Structural Faults in Epitaxial and Buried Layers in Silicon in Device Fabrication P. Wang, F. X. Pink and D. C. Gupta, General Telephone and Electronics Laboratories	285
32.	An Instrument for Automatic Measurement of Epitaxial Layer Thickness A. C. Roddan, Beckman Instruments, Inc. and V. Vizir, Fairchild Semiconductor	302
30.	A New Impurity Profile Plotter for Epitaxy and Devices B. J. Gordon and H. L. Stover, University of Southern California and R. S. Harp, California Institute of Technology	273
	SESSION IV-A - INTERDEPENDENCE OF UNIT PROCESSING OPERATIONS, CHAIRMAN - J. OROSHNIK	
33.	A Statistical Approach to the Design and Fabrication of Diffused Junction Transistors D. P. Kennedy, IBM, Components Division Previously published in IBM J. of R and D 8, 482 (1964)	-
34.	Defects Induced in Silicon Through Device-Processing M. L. Joshi and J. K. Howard, IBM, Components Division	313
35.	A Study Relating MOS Processes to a Model of the Al-SiO_2-Si System M. H. White, F. C. Blaha and D. S. Herman, Westinghouse Corporation	365

Paper No. | Page No.

36. Activation Analysis in Silicon Device Processing 375
G. B. Larrabee and H. G. Carlson, Texas Instruments, Inc.

37. The Use of the Scanning Electron Microscope as a Semiconductor Device Production Line Quality Control Tool 384
J. W. Adolphsen and R. J. Anstead, NASA Goddard Space Flight Center.

38. Metallization Deposition Parameters and Their Effect on Device Performance 398
J. R. Black, Motorola, Inc.

39. Methods for Determination of the Characteristics of Hyper-pure Semiconductor Silicon and Their Information Content for Device Production 409
Fritz G. Vieweg-Gutberlet, Wacker Chemitronic GMBH

SESSION IV-B - SURFACE PREPARATION, CHAIRMAN - E. MENDEL

40. Mechanical Damage-Its Role in Silicon Surface Preparation 412
R. B. Soper, Semiconductor Processing Co., Inc.

41. Crystallographic Damage to Silicon By Typical Slicing, Lapping, and Polishing Operations 419
T. M. Buck and R. L. Meek, Bell Telephone Laboratories

42. The Preparation of Practical, Stabilized Surfaces for Silicon Device Fabrication 431
A. Mayer and D. A. Puotinen, RCA Corporation

43. Surface Contamination 436
J. W. Faust, Jr., University of South Carolina

44. The Precipitation of Oxygen in Silicon and Its Effect on Surface Perfection 442
W. J. Patrick, IBM, Components Division

45. Auger Spectroscopy and Silicon Surfaces 450
J. H. Affleck, General Electric Company

46. Characterization of Semiconductor Surfaces and Interfaces by Ellipsometry 457
N. M. Bashara, University of Nebraska

AUTHOR INDEX

Author	Page
Adams, A. C.	87
Adolphsen, John W.	384
Affleck, J. H.	450
Anstead, Robert J.	384
Bakeman, P. E., Jr.	184
Barry, M. L.	175
Bashara, N. M.	457
Black, J. R.	398
Blaha, Franklyn C.	365
Borrego, J. M.	184
Branscomb, L. M.	1
Buck, T. M.	419
Carlson, H. Gary	375
Chawla, B. R.	182
Chou, S. L.	141
Cox, W. P.	51
Cunniff, K.	168
Dash, S.	202
Davidson, L. A.	141
Deal, Bruce E.	36
Donovan, R. P.	123
Dumin, D. J.	79
Evans, R. A.	123
Faust, J. W., Jr.	436
Gibbon, C. F.	21
Gibbons, J. F.	141
Gordon, B. J.	273
Greig, W.	168
Gupta, D. C.	66, 285
Hammond, M. L.	51
Hanes, M. H.	87
Harp, R. S.	273
Herman, David S.	365
Howard, J. K.	313
Hyman, H.	168
Irvin, J. C.	99
Jones, T. D.	256
Joshi, M. L.	202, 313
Joyce, B. A.	19
Katz, L. E.	192
Kohler, William A.	60
Langer, P.	256
Larrabee, Graydon B.	375
Lenard, J. A.	11
Masters, B. J.	132
Mayer, A.	431
Mazur, R. G.	244
Meek, R. L.	419
Millea, M. F.	156
Muller, S.	168
Niehaus, W. C.	256
Patrick, William J.	442
Pink, F. X.	285
Porter, John L.	66
Puotinen, D. A.	431
Roddan, Allison	302
Schumann, P. A., Jr.	234
Severin, P. J.	224
Simpson, J.	87
Soper, R. B.	412
Stover, H. L.	273
Thomas, D. G.	3
Tucker, R. N.	223
Van Gelder, W.	256
Vieweg-Gutberlet, F. G.	409
Vizir, Vitali	302
Wang, P.	285
White, Marvin H.	365
Yeh, T. H.	111

Silicon Device Processing

Abstract

This Proceedings contains the information presented at the Symposium on Silicon Device Processing held at the National Bureau of Standards on June 2 - 3, 1970.

This Symposium covered the measurement field and was purposefully restricted to the generic topics of diffusion, epitaxy, surface preparation and interdependence of unit processing operations. This emphasis on measurement during the processing operation or on the characteristics of the processed material, showed the necessity and more important, the ambiguities of current methods of measurement. The application of some new techniques to measurement were also discussed.

Key words: Analysis; device processing; diffusion; epitaxy; junctions; resistivity; surface preparation.

Introduction

The Symposium objective was to provide a critical review of present characterization methods and disclosure of newer techniques with ample opportunity for discussion among the participants. The material was directed to the needs of the applied scientists and the engineers concerned with the development and production of devices. The registration (446) for this Symposium confirmed the need for this information. A secondary objective was to increase the awareness of the program of the sponsoring Committee F-1 both for the recruitment of new members and the introduction of new ideas.

Most of the authors of the 46 papers delivered at the Symposium, were selected to speak on a subject in which they had acquired unusual competence. This selection process was accomplished by the Chairman of each of the seven sessions. Furthermore to assure a comprehensive program, all authors were acquainted with the subject matter of other authors in the session and several were requested to emphasize certain aspects in their paper.

This Proceedings provides a current review of measurement technology in these limited fields of silicon device processing.

WELCOMING REMARKS BY DR. LEWIS M. BRANSCOMB, DIRECTOR
NATIONAL BUREAU OF STANDARDS
U. S. DEPARTMENT OF COMMERCE
DELIVERED BEFORE THE SYMPOSIUM ON SILICON DEVICE PROCESSING
AT THE NATIONAL BUREAU OF STANDARDS
GAITHERSBURG, MD.
JUNE 2, 1970

We are especially happy to welcome you to this Symposium on Silicon Device Processing at the National Bureau of Standards because of our continuing interest in cooperating with the American Society for Testing and Materials and particularly because of the common objectives of Committee F-1 and the Electronic Technology Division in this field. These two groups have cooperated over the years in their technological support of electronics in reviewing and disseminating information of value to science and technology. Support of the electronics industry is an important NBS activity. We are particularly proud of the accomplishments of the Electronic Technology Division of the Institute of Applied Technology.

As you are well aware, there is still much to be done in processing technology of these devices. There is a misunderstanding in the country of the state of processing. Generally, this area is taken for granted both by the consuming industry and in government procurement. The Bureau is proud to play a part in strengthening process technology. Even though it is a small part, it is the major organized effort in the Government in your field of interest. Many comments have been made elsewhere which justify need for such work.

Indeed, it has been estimated by an industry representative that there is a loss of $100,000,000 annually in this field because of the inaccuracy of measurements. One must always be careful in evaluating statements like this, but the intent is to emphasize that much needs to be done to improve appropriate measurement abilities. However, a large part of the required work is within the scope of interest of both ASTM and NBS and the challenges are being met, at least in part.

Many of you are not active in this Committee and may not be familiar with ASTM, whose goal is the promotion of knowledge of the materials of engineering and the standardization of specifications and the methods of testing. As a matter of fact, this phrase also describes much of the activity of NBS.

Committee F-1 is concerned with material for electron devices and microelectronics and covers a very broad range of topics in this field. It has issued over a hundred voluntary industrial standards and has a roster of over 600 members and guests who represent 200 different company organizations. Participation in this Committee provides a unique opportunity for both users and suppliers in this field to meet to discuss measurement problems and develop carefully documented test methods with demonstrated relevance, precision, and accuracy. The ASTM approach is especially important because the user viewpoint is so carefully considered.

The National Bureau of Standards has an active program in the field of interest to Committee F-1. This program was established to focus National Bureau of Standards' efforts to enhance the performance, interchangeability, and reliability of semiconductor devices through improvements in methods of measurement for use in specifying materials and devices and in control of device fabrication processes. These improvements are intended to lead to a set of measurement methods which have been carefully evaluated for technical adequacy which are acceptable to both users and suppliers, and which can provide a common basis for the purchase specifications of government agencies. In addition, such methods will provide a basis for controlled improvements in essential device characteristics. Cooperation with industrial users and suppliers of semiconductor devices is a vital factor in the success of the program. The program is jointly supported by the National Bureau of Standards and several other government organizations, including Defense Atomic Support Agency, U.S. Navy Strategic Systems and the National Aeronautics and Space Administration.

A word of warning is often raised about the depressing effect on the industry that can result from standardization activities that drag on over long periods of time. The opposite effect on the industry can result when Government, as a consumer, does its standardization job well. The development of performance standards can provide an incentive for industrial innovation, as has been done in some areas of MILSPEC development.

Electronics is a $25 billion industry at the factory sales level and about half of this is purchased by the Federal Government. Semiconductor sales are over $1 billion, with about 25 percent of this output going to the Department of Defense alone.

Committee F-1 provides valuable guidance for assignment of priority to new work to be undertaken

by NBS in meeting national needs in electronic technology. It acts as a forum for prompt dissemination of the output of our work and immediate feedback for guidance during its conduct to maintain the relevance of the work. Especially important is the participation in the industry round-robins where interlaboratory precision of the measurements under development in the industry can be established.

NBS is consequently an active participant in Committee F-1. Twenty-nine NBS staff members are members or guests of F-1. Eight hold positions as elected or appointed officers. These numbers may seem large, but they still represent less than five percent of the members and guests of this large Committee. Furthermore, the Committee Chairman is a member of the Advisory Committee from the National Academy of Engineering for the NBS Electronic Technology Division.

The best known example of the synergism resulting from this cooperation is the work done by NBS on the four-probe measurement of resistivity of silicon. This was an undertaking proposed to us by this Committee, carried on throughout in cooperation with the Committee members, and evaluated in round-robins conducted by the Committee. Even the success of the cost benefit analysis which disclosed the widespread use and the very significant value of the resultant measurement method was dependent to a considerable extent on contributions made by Committee members and others in their companies.

The Committee is a healthy, growing organization. This is the 15th anniversary of its founding as a separate ASTM Committee. It has recently added new activities in ceramics for electronics and has under consideration new activities in measurements for process control and for semiconductor devices.

I am sure this fine work will not only continue but will grow both in scope and importance.

Some Thoughts on How We Might Improve Our Materials and Process Work

D. G. Thomas

Bell Telephone Laboratories
Murray Hill, New Jersey
07974

Why Should We Improve Quality?

Before trying to grapple with the problem of how we might improve the quality of our silicon processing work, it might be worthwhile spending a few moments on precisely why we want to do this.

One obvious advantage of better processing is to allow us to increase the size of an individual integrated circuit chip which can be made with a tolerable yield. Only if the average defect density/sq. cm is small enough can we successfully manufacture a chip of a certain size. Now to have chips larger than we can make at present might certainly be a good thing - circuits could be made faster, there would be economies that would flow from the reduced need of interconnections between several chips, and indeed the vision of being able to have a small computer on a single chip has an almost mesmeric attraction.

There are a few notes of caution to sound however. One is the problem of heat dissipation. Particularly for fast, large circuits it will become increasingly difficult to remove the heat without very special bonding techniques. Furthermore the question of economics becomes involved with such factors as how much does it cost to bond a chip, what is the cost of an interconnection pattern, what are the economic penalties to be paid if a chip is broken so that there has to be some sort of repair procedure, how can the system be repaired if it fails in use, what are chip test costs and how do they vary with chip size, what is the unused chip border width and how much does it cost? The answers to these questions depend on the particular strategy used for fabricating and mounting the chips.

As I daresay many of you know we in the Bell System are using a hybrid technology in which we have beam lead-sealed junction silicon devices bonded to a ceramic substrate. Some essentials of this technology are:

(a) Carefully sealed junction, so that no hermetically sealed can is necessary

(b) Batch bonding of the SIC chip

(c) Ta circuit on Al_2O_3 substrate for precision resistors and capacitors

(d) A faulty chip can be replaced, so that a whole ceramic board is not wasted.

In contemplating this arrangement our people believe that there will be little or no advantage in going to chips larger than about 100 mils on an edge. These chips would have an area 1.5 - 4 times what we are using today, and this is decided by our present defect rate, so we certainly need to improve our process work. But from this viewpoint alone you might conclude that the problem was not one of the highest urgency requiring priority attention.

There are however other advantages both economic and technical which can flow from improved processing, without increasing the size of the chip.

First, there is the obvious one that if yields can be improved for a given size chip, the chips will become cheaper - not only will there be less handling required on the production line, but we will need less crystal growing apparatus, fewer epi machines, less polishing of slices and so forth. At present in development facilities overall probe yields are in the range 15-20% including large as well as conventional-sized chips. In production facilities yields may be somewhat higher, but not for chips 100 mils on a side, and as I said there is economic incentive to achieve such sizes.

Furthermore, if quality can be improved then design tolerances can be lowered. For instance many devices are now made typically with 0.3 mil tolerances. If this could be reduced to 0.1 mil then 9 times as many devices could be made per slice as are made at present, again with great saving of cost.

In addition to this, lower tolerances would mean finer lines, and finer lines mean lower capacities. Thus with finer lines we can achieve the same speed with less power, and a reliable source of power is not cheap, to say nothing of heat dissipation and other problems associated with high power levels.

So I have come to the right conclusion, we do have great incentive for increasing the quality of our work, but it is not just for the sake of making the chips larger.

Organization at BTL

Now the need for this improvement was recognized some time ago at Bell Laboratories and it may be of interest to you to know what was done about it from an organizational point of view.

It was decided over two years ago to create a new Division within the Device Development Vice Presidential Area, devoted to Materials and Process work. (Figure 1) Previous to this change there had been only two Divisions. You can imagine that this was a major disruption; previously everyone had been connected directly to a device organization. Now, some 180 professionals were picked up and transferred to a new organization, the success of which would ultimately be judged by how well they served the other two Divisions, each of which has about 200 professionals in it.

Stated simply, the purpose of this move was to bring together the people concerned with Materials and Processes into a strong group which could specialize and develop the knowledge and understanding required for this field of activity. One had to have professional people who regarded this end of the business as their function in life, and to give this group sufficient management power that success in this field could be properly recognized and not eclipsed by the impact which the development of a working device might have. They would also have the obligation to study materials work in depth, to some degree free of pressures normally attendant upon getting a particular device to work. Only in this way, it was felt, could we be sure that adequate attention would be paid to this increasingly important and complex and technically deep part of our business.

Now the danger of this sort of thing is that this organization might take on a life of its own, and publish papers and do all sorts of gratifying things only to find one day that no one cares, and in these rather austere times this will mean, quite properly, its extinction. Being responsible for this organization, this is not one of my goals.

So a great deal of attention has to be paid to coupling together the parts of the organization. How do we do this?

First of all, our people are mixed together geographically. Figure 2 illustrates how at one of our main locations, Murray Hill, we have the Research Department (Area 10), quite separate organizationally from the Fundamental Development (part of Area 20). However they are linked geographically by being in the same building. Our final development work is done at two locations in Pennsylvania, some 64 and 100 miles away from Murray Hill. The reason is that it is at Allentown and Reading that the Western Electric Company has its manufacturing locations, and we believe strongly that if technology is to be smoothly transferred to the Western Electric Company, then the Bell Laboratories must have strong efforts in the same building as the Western Electric Company. In this way a genuine working relation is established with those who have the problem of manufacture.

Do not think however that at each of the three locations Area 20 has one of its three Divisions. Quite the contrary. Generally at each location there will be strong representation from each Division. Thus the Materials and Processes Division has roughly half its people at Murray Hill, and the other half distributed between Allentown and Reading in Pennsylvania. Thus we have established organizational bonds where geographic barriers exist, to quote J. A. Morton, one of the chief architects of this scheme.

Now I may say that in some respects this is not the tidiest organization to manage. For any sort of meeting almost always involves travel. How much simpler to have the Materials and Process people at Murray Hill with good contact with Research, and an

ability to publish papers which could be handed on the Device people. But how obviously deadly. So, as I say, the Materials and Processes people are scattered throughout the operation thereby making close contact easy where there is no geographic barrier.

Second, in our effort to couple our Materials and Processes people and their device colleagues, we usually structure our work so that more than one of the Divisions is vitally involved in any one project, and a mutual interdependence is built up. For instance if we want GaP light-emitting diodes, the Materials and Processes people may grow crystals and do some liquid phase epitaxy, but the devices are fabricated and designed by the Device organizations. Another way of saying this is that generally we do not arrange things on a project basis but that, in principle at least, we have groups of experts who can be invited to occupy themselves with various specific projects as these come along. In this way we expect to maintain a body of people who can adapt to the rapid changes which take place in our technology. We will not build up groups who suddenly find themselves out of a job as they are working on a technologically obsolete project.

Finally, I do think that the management of such a Materials and Processes organization has a role to play which must be very sensitive to the professional standards of the people involved, but which can be important in realizing that certain things should be done and then persuading people of the wisdom of doing them. Let me mention two related areas that exercise us at present.

1. There often is a need to make a serious attempt to understand and control certain generic areas of our technology. For instance we do a great deal of metallization by a variety of techniques both on silicon and on ceramic substrates. We are concerned that these processes should work as well as possible, and be compatible with one another and also compatible with economic manufacture. It turns out that metallization is a remarkably complex materials field and there is a great lack of understanding of such things as corrosion resistance, adhesion to substrates, intermetallic compound formation and so on. This lack of understanding makes rational choices and decisions difficult. So here is an area of work which, from the point of view of any one individual, may not seem to be particularly vital, but which management can recognize as being in need of overall attention. Other examples are not hard to find. Our way of approaching such problems is first to attempt to define them, and then to present them to a group of individuals who should make recommendations and take appropriate action. We are currently active in this area on several fronts. Studies are underway not only of metallization but also in the fields of gold plating, photoresists, silicon oxidation and cleaning, and water purification. I think these subjects are all worthy of deep study but I am sure the list is not exhaustive. Nevertheless I am not anxious to increase it further without completing some of the topics, for these studies do really require the full-time attention of several people, and management urges that they be considered as very solemn undertakings. Obviously therefore management can be irresponsible by asking for too many such studies, with the result that some, at least, will not be done properly.

2. On a rather less grand scale there is the necessity of curbing the very natural tendency that seems to exist for any group of people who are doing one thing, to do it differently from another closely-related group of people doing the same thing elsewhere. In other words we should try to persuade people to adopt a measure of standardization and commonality in their process work, and not solve the same problem in many different ways. Examples of what might appear to be unnecessary duplication are not hard to find. There is a multiplicity of methods used for diffusion of impurities into silicon, probably more photoresists are in use than are necessary. I doubt if cleaning procedures are understood as well as they should be so that again there is agreat diversity of practices. So we are trying to pick, on a rational basis, a preferred process, and then to trumpet it forth with appropriate fanfare, and trust that by offering help and pointing out the advantages, it will be widely adopted.

One must recognize of course that this sort of policy should not act to curb useful innovation, and should not prevent improvements from being adopted as they are discovered. The answer, I think, to this danger is to make sure that the people involved in attempting to select preferred processes, do not just go away after they have made their initial recommendations. Rather they must maintain a watching brief and when something new comes along they should do their best to publicize it and point out its virtues. In this way, one can argue, the adoption of innovations will be helped by being made more rapid, not hindered as one had feared.

One further point worth emphasizing is that while it may not be easy to get a development line to change its practices, it will probably be even harder to get a high level production line to change. The reasons are obvious. If something is going

well there will be great reluctance to interfere in any way, and there is much reason for this. After all a preferred process may have hidden difficulties which may become apparent in high level production with disastrous consequences. I believe therefore that it is particularly important to work hard to make sure that the development and pilot lines are using the best possible techniques, so that when something passes on to the next level of production, it does so with the best possible processes in use.

Some Examples

Well, these are fine words. To make them seem a little more credible I would like to describe to you briefly a few specific examples where I believe progress is being made.

The first example concerns gold plating. At all three of the locations I described to you we have groups of professional electrochemists, and these people work well together. Now over the years gold electroplating for electronic device development and manufacture at Bell Laboratories and the Western Electric Company, has come to be carried out in three different bath types. These are alkaline cyanide, citrate and phosphate. It was not clear that such duplication was necessary and so our electrochemists were invited to sit down and reason together. They concluded that where pure soft gold is required, the phosphate bath, which had been under development at our Reading laboratory is the preferred bath. The alkaline cyanide bath is too unstable, and produces porous deposits. Furthermore the high pH attacks most photoresists. The citrate bath, which was also under development at Bell Labs, tends to degrade more rapidly than the phosphate bath and metallic impurities such as iron, nickel or cobalt will plate out and harden deposits from the citrate solution. So we have a drive on to standardize as much as we possibly can on the phosphate bath. Such a thing can only be done responsibly if we have available the advice of experts. For only then can we be sure that the choice is sensible, and furthermore if, when other people do make the conversion, difficulties are encountered the experts are available for help, and it is likely that such help will be forthcoming for the experts, presumably, have some interest in maintaining their reputations. Finally it may need a reasonably strong organizational hierarchy to see that the standardization really has a good chance of being adopted.

As it becomes known that there are experts around willing to help, questions may come in. One such question concerned the nuisance that arose from small gold nodules formed during the gold plating of integrated circuits. It was pointed out that these nodules may arise when foreign particles are present in the plating solution. These can be eliminated by continuous filtration, and again we have a drive on to institute this quite simple procedure wherever critical gold plating is carried on. When this step was described it became apparent that continuous filtration might be generally useful when integrated circuits are processed in a solution. We are, after all, very careful to filter dust particles out of the air that surrounds our integrated circuits during manufacture, should we not be similarly careful with our solutions? So again the Materials and Processes people are investigating this possibility for improving the quality of our work.

While these sort of activities are in train our professional electrochemists are innovating and inventing in relevant areas. For instance to be able to do metal plating outside a vacuum station offers economic advantages, and this of course is what electroplating does. However on occasion we may wish to plate a surface to which there is no ready electrical contact. At this point an electroless process is useful in which metal deposition from a labile solution is catalyzed by the surface and by the deposited metal itself. The trouble is that there are very few metals which can be deposited this way. Recently Okinaka, one of our electrochemists, has devised a truly autocatalytic, electroless gold bath in which borohydride is used as the reducing agent. and cyanide as the complexing and stabilizing agent. The deposit obtained is pure, soft gold and does not contain boron. The conductivity and density are good.

We expect this bath to be useful, but perhaps of even greater use would be an electroless platinum bath which was catalyzed by a silicon surface. Then we would be able to lay down Pt selectively on areas at which we wanted to make PtSi contacts to the silicon. We are not there yet but there are hopes for developing this process too.

So I think there are resources offered by a group of people such as these electrochemists which can have benign effects upon the work of all of us. And further, impact of these effects will be greater if there is an organization devoted to this class of thing which has some measure of authority.

Let me turn to another example of our Materials and Processes work.

From time to time completely new processing possibilities may present themselves. One of these which seems to us to have great importance is ion implantation.

As you know this technique allows one to shoot a very pure beam of ions at a silicon surface. If the energy of the beam is high there can be appreciable penetration into the crystal, with the simultaneous creation of lattice damage. Why is one enthusiastic about this? There are perhaps two main classes of silicon processing which can be improved.

The first of these is a replacement for the predeposition that is often done before a diffusion drive in. If you ask people who are nominally studying diffusion in silicon what they are really doing, as often as not you will find they are actually studying predeposition. The reason is that for many devices one needs to control the diffusion profile rather precisely, and this will depend on the quantity of impurity that is present at the surface before diffusion begins. It is not easy to chemically or mechanically deposit such small quantities as are usually required, but to put them there by ion implantation is very simple. One merely has to measure the beam current and the time. Furthermore it is often necessary to remove a residual glass after diffusion if further processing is needed, and this step is simplified with ion implantation since glass formation is held to a minimum.

MacRae and others at Bell Laboratories are for instance making a hyperabrupt diode using this predeposition technique. This is a backward-biased Varactor diode in which one wants the capacity to vary as the inverse square root of the applied reverse bias. There is a Schottky barrier at the surface formed on an epi layer into which impurities have been diffused. Clearly the C-V properties depend critically on the precise impurity distribution. Suffice it to say that whereas with conventional diffusion techniques this device was made with a yield of <1%, with ion implantation a yield of 80% has been achieved with high quality. This device is being introduced into Western Electric based on this technology. I am sure it will not be long before the advantages of this method of predeposition become more and more apparent and its use, involving as it does quite modest apparatus, will grow explosively.

Ion implantation is more usually thought of as a method for shooting ions into a crystal, and controlling their depth of penetration by means of their energy. In principle it is possible to make all sorts of cunning structures, with impurity profiles tailored to a fare-thee-well.

In practice there are problems with lattice damage and with "tails" on the impurity distributions. These tails may arise from channeling of the ions down hollow crystallographic channels in the solid, or perhaps from diffusion enhanced by the presence of lattice damage. These problems can however be understood and as the depth of understanding increases it is likely that they will become unimportant.

Already it is possible to devise annealing schedules which can result in almost 100% of the implanted ions finding their way to the desired crystal sites. At the same time the mobility can be restored to its normal value. The maximum annealing temperature varies with the doping level; it can be as low as 600°C but high concentrations may require 1000°C.

It is remarkable too that even the bulk lifetime of silicon may be restored to reasonable values. This is illustrated in Figure 3. MacRae and his associates have here bombarded Si with Si ions and then annealed. In this way they avoided complications due to electrically active impurities affecting the lifetime. You can see that except for the very highest doping levels, equivalent to about 10^{19} atoms/cc, the lifetime can be restored to the respectably long value of about 50 microseconds.

An example of a device made using deep ion implantation is shown in Figure 4. This depicts a double avalanche diode, or Impatt structure made by Seidel and others at Bell Telephone Laboratories, in which there is a high field drift space for both holes and electrons. In most Impatt structures only one type of particle is accelerated. The structure is made by implantation into an epi layer at 200 and 100 KeV energies to get a smooth distribution up to the junction at a concentration of 6×10^{16}/cc. The depth of the junction is 0.45μ. It is important that $N_A \simeq N_D$ and it is just this that can be achieved with ion implantation. The heavily-doped contact is diffused in. I believe this structure currently holds a record for microwave generation; at a frequency of 50 Gc/s it generates 650 milliwatts with an efficiency of 13.4%.

So I think that ion implantation is going to be very important to us. But we are going to have to persuade people to use it, and to develop simple inexpensive machinery which can be used on the factory floor. Our strategy is to recognize this as a generic materials system with special problems which must be understood and solved before it becomes a useful process. So it was put as a whole with the Materials and Processes people, and we rely on our couplings and motivations to see that it does become useful. So far I think the arrangement is working well, but we have ahead of us large scale factory use.

To change the subject again I will say something concerning the recently developed silicon target for use in the Bell System's PICTUREPHONE(R) camera tube. You may perhaps know that this year in the city of Pittsburgh, PICTUREPHONE(R) will be available on a commercial basis. Clearly every set not only has to have the display but it must also have an image pick-up device. In contemplating this pick-up element it was concluded that the commercially available units fell short of the needs for this particular application. Perhaps foremost among the troubles was the fact that the commercial tubes could be damaged by strong lights; that is they suffered from optical, and also electron beam, "burn in." These troubles are not very serious when the tubes are used in a television studio under the direction of experts. Such conditions do not exist however on the average telephone users premises, and the Telephone Company was not looking forward to having to replace the camera tube every time sunlight accidentally fell onto a customer's PICTUREPHONE(R) set. Because of the imperfect way in which we understood these pick-up devices we did not really see how improvements could be made.

Consequently it was decided to fabricate a pick-up tube so far as possible using the silicon technology which we understood well, and which we expected would not be subject to the difficulties just described.

A target was made which contained about 500,000 diodes roughly 10μ in diameter, on a single slice of silicon 18μ thick. There had to be no serious defects. One or two bright or dark spots on such a target can make it unacceptable to the human eye. In addition the reverse leakage current of the diodes has to be extremely low.

This device is of such elegant simplicity that it undoubtedly gave the device designer much pleasure. It did however present the device fabricator with some rather severe challenges. As you can imagine Materials and Processes people were involved at a very early stage and are still keeping an eye on things now that the targets are being made by the Western Electric Company.

Having a strong Materials and Processes group did, I think, help in the successful fabrication of these targets.

One area that caused a lot of trouble was white defects which appeared on the video display. Oversimplifying somewhat these WVD's appeared to be of two types, high level and low level. At one stage of our work we were anxious to see if this was really so or if we were really dealing with a distribution of one type of defect. One approach to this problem was to use a mask with numbers down the edge. Using this mask to fabricate the diode array, and using reduced area electron scanning one could sometimes make a positive geographical identification of a particular defect. Now under a scanning electron microscope the same spot could be examined. For high level defects a chip had often broken out from the surface. This allows direct contact between the scanning electron beams and the n type substrate. There is in effect a dead short and a very bright white defect appears. Similar fractures in the oxide have been seen and they lead to similar results. In general we have found that careful processing will reduce this type of defect to manageable proportions.

Now if the same procedure was applied to the low level white defects, no visible trouble was apparent in the scanning electron microscope. Clearly therefore something different is involved. A good deal of work has been devoted to understanding these defects. At present our best guess is that it is a region inside the crystal at which there is a high electron hole generation rate. It appears to be associated with the precipitation of an impurity or some crystalline defect for its presence is affected by the rate at which samples are quenched.

These are a couple of examples of investigations that went on for the PICTUREPHONE(R) target. There was a great deal besides. In-process controls were developed, making use especially of MOS measurements on test slices that ran through with the regular targets. In this way a close watch was kept on the fixed charge in the oxide and the surface state densities, both of which are important parameters for

the working of the device. Such tests were made at 4 points in the processing sequence and so trouble could be identified early, and the piece of apparatus or the material supply quickly corrected.

In addition some of our people from Murray Hill went on temporary assignment to Reading, where the devices were being produced. They did this both to bring skills to the job at Reading and subsequently to bring back a knowledge of real problems to Murray Hill.

All together it was an exciting time and I am glad that production of these devices is now beginning.

Well, I have described why and how we have set up a group of people concerned with Materials and Processes, and something of our structuring to try to make this organization useful, and finally I have given a few examples, selected you may be sure, to demonstrate that things are going well.

I might conclude by offering a few thoughts concerning how we are to keep the operation on the rails, to make sure that we discharge our responsibilities, and to keep morale and standards high. Like most other groups we are a service group and our ultimate success is measured by the contribution we are able to make to our colleagues who are directly concerned with devices. The words "service group" may however in the minds of some have pejorative overtones, and indeed we do not think of ourselves as "hardware merchants" who hand out stock items or materials on request. Rather, to quote again Jack Morton, we regard ourselves as "professional servants," people who not only can provide others with what they need today, but can anticipate future needs and indeed create new opportunities by improving and inventing ourselves. This of course we will only be able to do by having a broad knowledge of the field, so that future needs may be appreciated, and at the same time having a deep knowledge of our own subject. Once one is able to create the situation in which people are doing work which is challenging their intelligence and using their professional training, and leading to a result which can be clearly seen to be an improvement, ideas about whether or not they are working in a "service group" tend to become unimportant.

Recently some of our very best people set out to determine why some furnaces produced high quality silicon oxide films on silicon and why others did not. Something a Materials and Process group could take the time to do. First of all thoughtful experiments had to be designed to measure the critical properties in a statistically meaningful way, and then a series of carefully controlled experiments had to be devised to track down the critical steps. At present it seems as though the way the slices are cleaned before oxidation is one of the most important steps. If now we have a way of really measuring the effectiveness of cleaning we will be able to define the most efficient and foolproof cleaning technique. The economic consequences of this could be enormous, since there are in use a huge variety of cleaning methods, many of them requiring several hours to complete. I have high hopes that we will be able to adopt a simple, quick method that is highly effective. The study will not end there however for we are determined to find out the chemical nature of the impurities which can lead to trouble, for only then will we feel comfortable about having solved the problem. Unless a solution is based on understanding then always one will be at the mercy of something unexpected turning up which will appear to be a great mystery. This sort of thing is often referred to as "Murphy's Law" -- if something can go wrong, it will, and unless one understands, it will be hard to diagnose the trouble. At any rate we have a self-starting, highly professional program conducted by talented exploratory development people investigating, of all things, how to clean a silicon surface. This seems to me to be a good thing.

So I think that once one is reasonably sure that the problem being attacked is a relevant problem, the best strategy which management can adopt is to simply ask the question, "Do you understand what you are doing?" If the answer is "yes" you may be reasonably sure that real progress is being made towards the solution of the problem, and at the same time that the people will be satisfied because they will be able to describe their results to others, to write papers, to have an impact on technology, and so ultimately to derive satisfaction from their work. And as most of us know, while salary and such like matters have their place in the scheme of things, what we really want to feel as we go home at night is that we have done a day's work in which we can take pride, and which commands the respect of our colleagues.

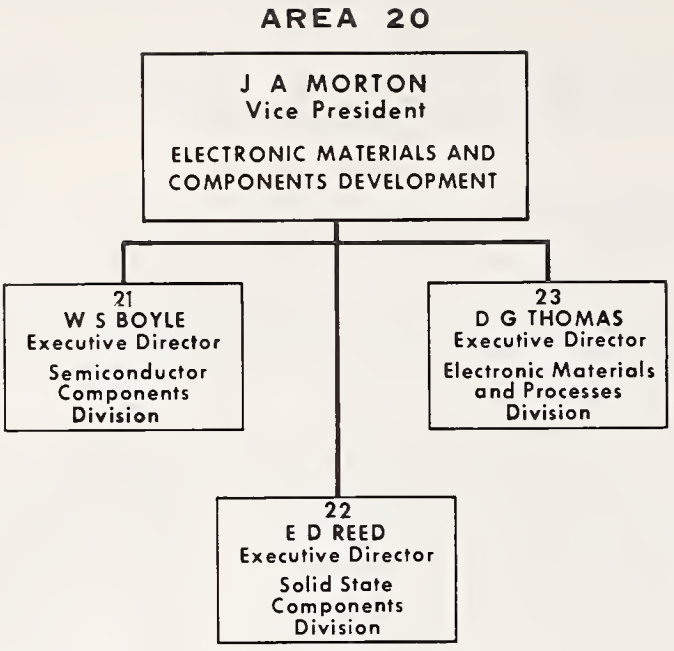


FIGURE 1

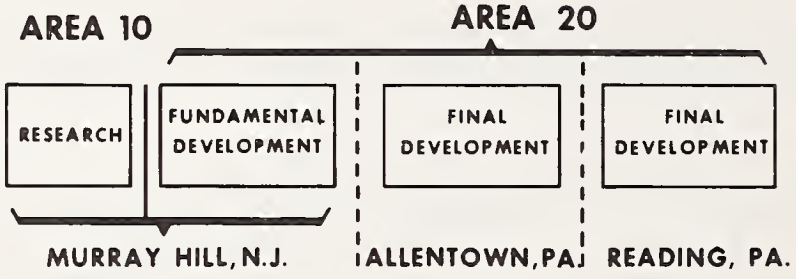


FIGURE 2

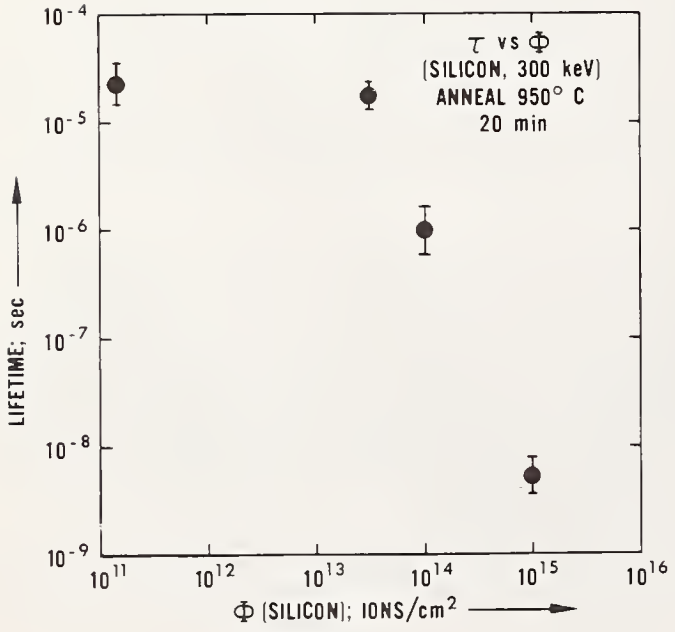


FIGURE 3

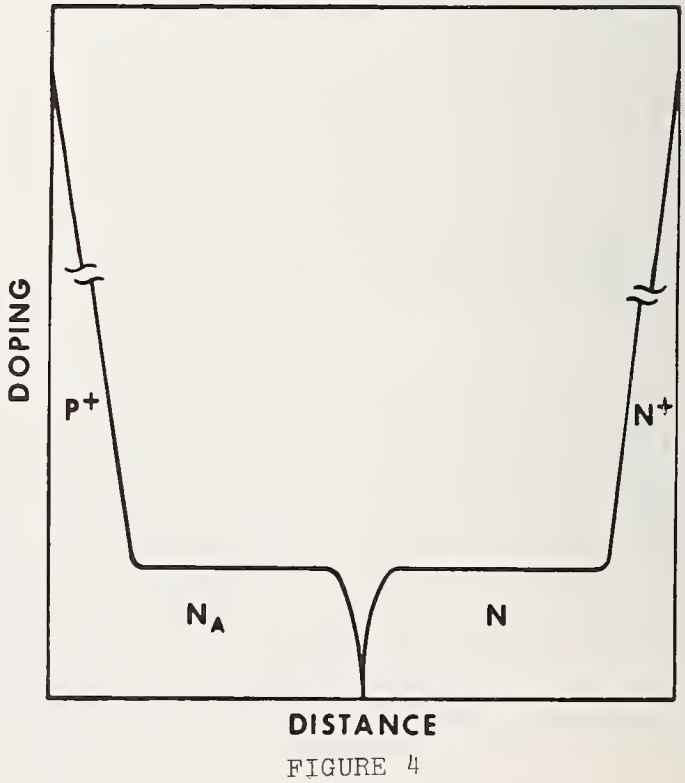


FIGURE 4

Crystallographic Imperfections as Related to Silicon Crystal Growth

J. A. Lenard

IBM Components Division
East Fishkill Laboratory
Hopewell Junction, N.Y. 12533

Problems of growing silicon crystals while keeping crystallographic imperfections to a minimum are described, with emphasis on Czochralski-grown crystals. The increase in silicon crystal diameter is attributed to improved analytical techniques for assessing the structure of crystals produced in a given system. Among these techniques are x-ray diffraction microscopy and the scanning oscillator technique. Since growth-rate fluctuations are the major cause of nonhomogeneous impurity distributions, thermal gradient variations, and crystal diameter changes, it is necessary to obtain tighter control of the growth conditions. Use of the IBM 1800 process control computer to grow silicon crystals of uniform diameter and superior structure is briefly described.

Key Words: Crystal growth, crystallographic defects, control systems, Czochralski crystals, grown-in defects, process-induced defects, silicon crystals.

In 1948 (1)[1] the semiconductor age was launched at the Bell Telephone Laboratories with the invention of the transistor. This invention led to the need for single-crystal semiconductor material, which has continued to expand to meet market requirements. The early work in the growth of single-crystal germanium by Teal and Little (2) provided a sound basis for a large portion of the later material requirements for discrete transistors and diodes, more advanced hybrid and monolithic integrated circuits, and other devices, such as light-emitting diodes, tunnel diodes, and solid state lasers, among others.

Although device performance depends on characteristics primarily associated with materials in Group IV of the periodic table, III-V and II-VI compounds were found to have desirable energy gap, carrier mobility, and thermal stability, even though the crystal perfection could not, in most cases, match that of germanium and silicon. Gallium arsenide, for example, attracted considerable interest in the early 1960's because of its inherently favorable energy gap and carrier mobility. However, it never became a competitor to elemental semiconductor materials since other factors, such as economics and availability, prevented its widespread implementation.

Theory predicts faster device performance for germanium and gallium arsenide, as well as other compounds, as compared with silicon, but ingenious developments in device design and geometry (such as planar technology) have led to the domination by silicon throughout the 1960's. It appears that silicon's domination will continue for the electronics industry throughout the 1970's, barring some unforeseen technological breakthrough.

Most of the silicon crystals produced in the world today are produced by the method reported by Czochralski in 1916. He did not envision, I am convinced, the fantastic growth and almost universal application of his technique to the growth of silicon crystals, since his interest was in measuring the solidification rates of lead, tin, and zinc. Many features of his original experiment are being rediscovered daily in crystal-pulling systems. Teal and Little first applied the method to germanium, as mentioned earlier, and Dash (3) improved the method further to enable one to grow dislocation-free crystals rather routinely in crystal diameters common at the time. Since most crystals grown today are prepared by the Czochralski method, the remainder of this discussion will concentrate on crystal characteristics of Czochralski-grown crystals.

[1] Figures in parentheses indicate the literature references at the end of this paper.

With the advent of monolithic integrated circuit technology, use of the Czochralski method has become more popular, mainly because:

- Large-diameter crystals are needed to process more circuits with a minimum of handling per circuit.
- Economic factors affect cost of crystal growth.
- Resistivity tends to be uniform in a crystal doped with boron.
- High-quality material is needed with a minimum of defects during crystal growth.

The crystal quality requirement being imposed on material suppliers becomes necessary since device dimensions and densities cause microscopic crystal defects to become critical from a yield, performance and reliability standpoint.

The Czochralski method is not without problems, however. The need to contain the molten silicon in a quartz crucible because of the lack of a better crucible material leads to the incorporation of oxygen into the silicon in levels as high as 10^{18} atoms/cc. Oxygen causes problems in further characterization of crystals because of crystal resistivity changes, sometimes to the extent of altering the conductivity type. Oxygen precipitates, under prolonged heat treatment, can also cause dislocations and inhomogeneous distribution of doping impurities, as seen in Fig. 1. Crystal rotation is known to have a marked influence on oxygen concentrations in Czochralski-grown crystals, as reported by Taylor (4). The concentration of incorporated oxygen can be reduced by a factor of two by slowing the crystal rotation rates in Czochralski crystals.

The floating zone method is popular for producing silicon crystals with low oxygen concentrations when zero dislocations and exceptionally large diameters are not required. Another method of producing silicon without crucible contamination has been successfully employed by Sterling and Warren (5) who used water-cooled crucibles to melt high-temperature materials. The use of electron beam heating is also becoming increasingly popular to eliminate environmental conditions that contribute to crystal contamination.

The expansion of crystal diameters, which makes more area available per slice for device processing, has been dramatic in the the last decade. Technology limitations constrained crystal diameters to less than 1-in. in the late 1950's. Today, 3-in. diameters are becoming common in device processing, as evidenced by the availability of 3-in. device processing equipment in the trade journals. Figure 2 shows a crystal of 4-1/2-in. diameter being grown. Much larger silicon single crystals have been grown in the industry for several years. Severe limitations, however, inhibit the use of this type of crystal for device processing.

Along with the increase in silicon crystal size, developments in recent years have permitted the growth of silicon crystals with a much higher perfection level than was available in the past. These improvements, such as adjusted thermal gradients, improved thermal stability, improved mechanical and electrical stability, etc., have been made possible by advanced analytical techniques that enable one to accurately assess the structure of crystals produced in a given system. X-ray diffraction microscopy has played a vital role in aiding crystal growth development. A nondestructive analytical technique particularly useful in studying the defect structure of crystal, known as the scanning oscillator technique (SOT), was developed by G. H. Schwuttke (6). This method has been used successfully for analyzing large silicon wafers up to 6-in. diameter with relative ease. The experimental arrangement is shown in Fig. 3. Some of the early work at our laboratory on crystal diameters 2-in. and larger yielded results shown in Fig. 4. High dislocation densities, thermal slip, and lineage characterized some of this early work and was of concern with respect to device performance, yields, and ultimate circuit reliability. Although the role of grown-in dislocations in device performance had not been established, their occurrence was generally considered undesirable. Adjusted thermal gradients and improved control techniques, discussed below, improved crystalline perfection so that defect densities are minimized, as evidenced in Fig. 5.

Grown-in dislocations are generally near zero, approximately four or five orders of magnitude less than process-induced dislocations introduced during high-temperature solid state diffusion and oxidation operations necessary in fabricating planar transistor structures (7). Thermal slip introduced as a result of differential heating and cooling before and after heat treatment is shown in the topograph in Fig. 6. Figure 7 is a yield map showing device failures in regions of process-induced defects. With larger crystal diameters demanded for device requirements, it becomes more important to understand the origin of crystallographic defects, either process-induced or grown-in.

The growth of silicon crystals with varying impurity distributions and subsequent resistivity nonuniformity is a problem that cannot be overcome until crystal growth methods are improved. The increased complexity of integrated circuits in conjunction with larger crystal diameters complicates the problem by requiring tighter resistivity gradient specifications. Localized resistivity variations, which result from impurity concentration changes, give rise to localized stresses. These stresses can be photographed by the SOT. Figure 8 shows the striations that result from instabilities in growth and rotation rates. A. F. Witt (8) has shown by anomalous x-ray transmission that extensive lattice strain is associated with impurity striations and core boundaries. He also showed a relationship between growth rate fluctuations and core boundaries. Goetzberger et al.(9) have found that low breakdown voltage is correlated with glowing diodes, which correspond to striations that result from impurity concentration variations.

Grown-in dislocations in silicon crystals can be minimized by proper control of growth variables. When various <111> grown crystals are cut with slices parallel to the growth axis in the <112> plane through the center of the crystal "neck" region of the initially grown crystal and the crystal seed, topographs of the slices show dislocations in the seed and their propagation into the growing crystal, as well as the initial growth interface between seed and crystal. Figure 9 illustrates a heavily dislocated seed from which a crystal was grown with a dislocation density several orders of magnitude lower. It is clear that the propagation of dislocations across the seed-crystal interface is inhibited. Figure 10 shows a high density of dislocations in the neck region becoming considerably reduced as the body of the crystal continues to grow. After only a few inches of growth the density of dislocations is relatively low. Figure 11 illustrates further evidence of the reduction in grown-in defects as the crystal grows. By carefully controlling the growth parameters during the initial phases of flaring out to the desired crystal diameter, one can completely eliminate the grown-in dislocations in silicon crystals, as shown in Fig. 12. Although the initial crystal was heavily dislocated, a dislocation-free crystal emerged after only a few millimeters were grown. The mechanism for the reduction in dislocations during silicon crystal growth has been proposed by Schwuttke (10). Low dislocation mobility in silicon is one of the unique features that allow one to suppress dislocation generation entirely during crystal growth. If proper controls and a sufficiently fast growth rate are imposed, dislocation-free crystals 75mm and more in diameter can be grown.

It seems then that the primary approach for achieving quality silicon crystals is adequate control of the critical growth parameters: crystal and crucible rotation rates, crystal growth rate, temperature and temperature gradients, and environmental conditions within the furnace. Many schemes to automate these critical parameters have been used with varying degrees of success. A computerized scheme for controlling the process to produce synthetic quartz was described by D. W. Rudd et al. (11) Computerized methods have been used successfully on silicon crystals also.

An IBM 1800 process control computer has been used successfully by D. Jen, R. Slocum, and C. Valentino to grow silicon crystals of uniform diameter and superior structure. A schematic of the crystal growth system is shown in Fig. 13, and the 1800 DAC control system configuration in Fig. 14.

Use of the 1800 system begins with developing a mathematical model and programming the system for various crystal types and sizes. The digital signal is converted to analog and fed to the controls for heater power, crystal lift, crucible lift, crystal rotation, and crucible rotation. The crystal puller now functions according to the predetermined program. The process operator console allows the operator to override the program if necessary. Signals are then picked up by various sensors and fed into the multiplex terminal unit, amplified, converted back to a digital signal, and returned to the central processing unit. The prime advantage of this type of control system is that infinitesimal adjustments can be made every few seconds. Measurements of existing growth conditions can be compared with data stored in memory. If deviations from desired conditions are anticipated from the data being accumulated, necessary corrections are made on a real-time basis. Since growth-rate fluctuations, as noted previously, are the major cause of nonhomogeneous impurity distributions, thermal gradient variations, and crystal diameter changes, the way to improved crystal quality is by tighter control of the growth conditions. A photograph of crystals grown by computer control, Fig. 15, shows that diameter control is extremely good for Czochralski crystals.

The future of silicon crystal growth will continue to require effort to further improve control and to understand more thoroughly the basic mechanisms involved. Requirements for multifunction large chips with decreasing device size dictate better crystallographic perfection, high purity, and improved resistivity uniformity. This means larger wafer sizes are required to achieve desirable processing economies. High-quality, large-diameter crystals will require scaling up and considerably adjusting the thermal design of most crystal-growth equipment. Methods will have to be devised to reduce the contamination of Czochralski-grown crystals by oxygen, oxygen complexes, carbon, and other metallic impurities. No doubt other than conventional induction and resistance heating systems will have to be developed, along with improved controls to eliminate impurities and defects in Czochralski-grown crystals.

I would like to thank Dr. G. H. Schwuttke for supplying all topographs.

REFERENCES

(1) J. Bardeen and W. H. Brattain, Phys. Rev. 74, 230 (1948).

(2) K. Teal and J. B. Little, Phys. Rev. 78, 647 (1950).

(3) S. C. Dash, "Method of Growing Dislocation-Free Semiconductor Crystals," Canadian Patent No. 713, 524, issued July 13, 1965.

(4) H. L. Taylor, "Imperfections in Silicon," Fall Meeting, Electrochemical Society, Columbus, Ohio, September 1959.

(5) H. F. Sterling and R. W. Warren, "Cold Crucibles for the Processing of Silicon and Applied Reactive Materials," Semiconductor Products and Solid State Technology, pp. 33-36, August 1965.

(6) G. H. Schwuttke, J. Appl. Phys. 36, 2712 (1965).

(7) G. H. Schwuttke, "Semiconductor Junction Properties as Influenced by Crystallographic Imperfections," Air Force Cambridge Research Laboratory Report 66-192. March 1966.

(8) A. F. Witt, "Microdistribution of Impurities in Semiconductor Single Crystals," ACCG Conference on Crystal Growth, Gaithersburg, Md., August 11-13, 1969.

(9) A. Goetzberger, B. McDonald, R. Hast, and R. M. Scarlett, J. Appl. Phys., 34, 1591 (1963).

(10) G. H. Schwuttke, "Crystal Properties as Influenced by Crystallographic Imperfections," AFCRL-70-0110, March 2, 1970.

(11) D. W. Rudd, A. R. Fiore, and N. C. Lias, "Computerized Process Control for Synthetic Quartz Growth," ACCG Conference on Crystal Growth, August 11-13, 1969.

Fig. 1. Oxygen precipitates after prolonged heat treatment as shown by SOT.

Fig. 2. Large diameter (4-1/2") silicon crystal being grown.

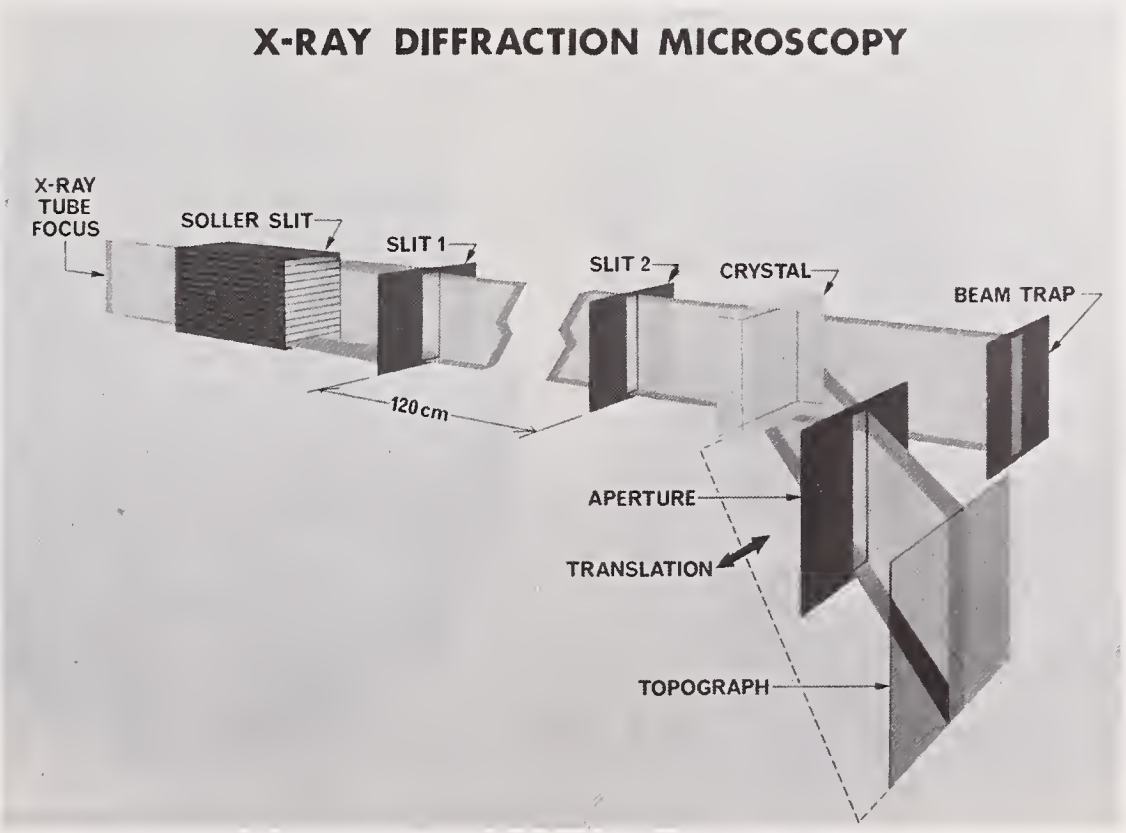


Fig. 3. SOT experimental arrangement.

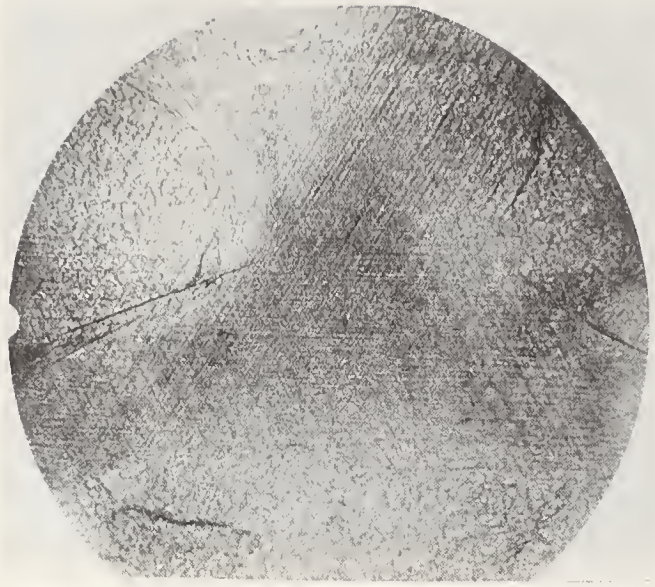


Fig. 4. Topograph of large-diameter <111> crystal exhibiting high dislocation densities, thermal slip, and lineage.

Fig. 5. Topograph of perfect silicon crystal.

Fig. 6. Topograph of silicon wafer showing process-induced defects.

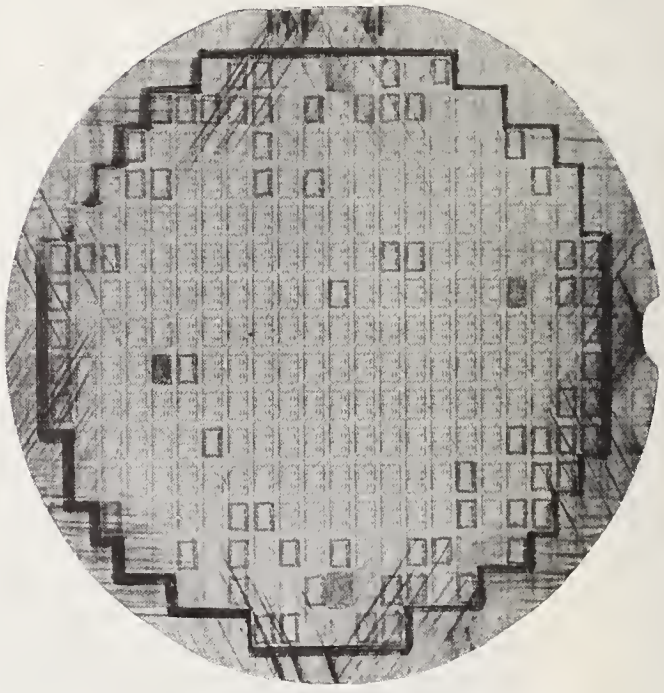


Fig. 7. Yield map showing device failures in regions of process-induced defects.

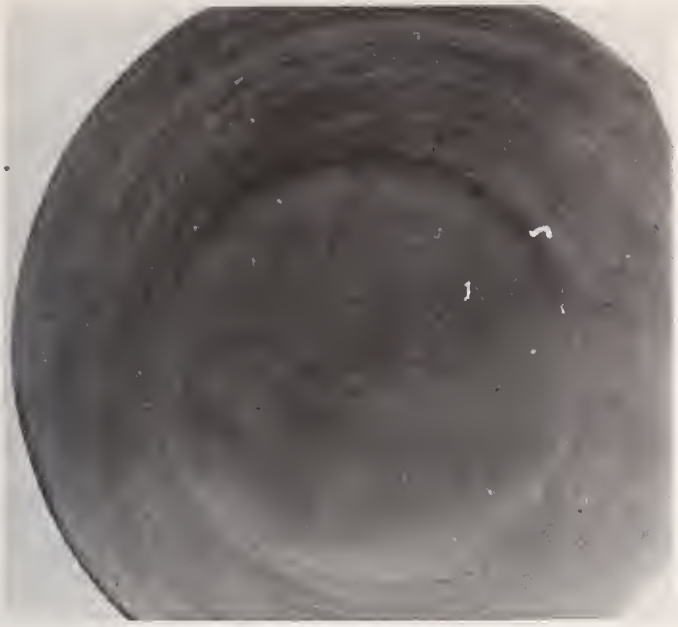


Fig. 8. Topograph showing localized stresses due to impurity concentration changes.

Fig. 9. Topograph of crystal section showing dislocations propagating from seed into body of crystal.

Fig. 10. Topograph of crystal section showing dislocations propagating from seed into body of crystal.

Fig. 11. Topograph of crystal section showing dislocations propagating from seed into body of crystal.

Fig. 12. Topograph of crystal section showing dislocations propagating from seed into body of crystal.

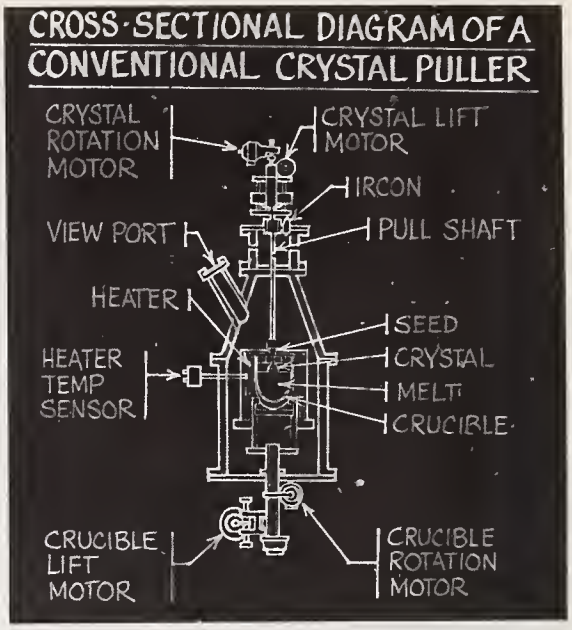


Fig. 13. Schematic of Czochralski crystal growth system.

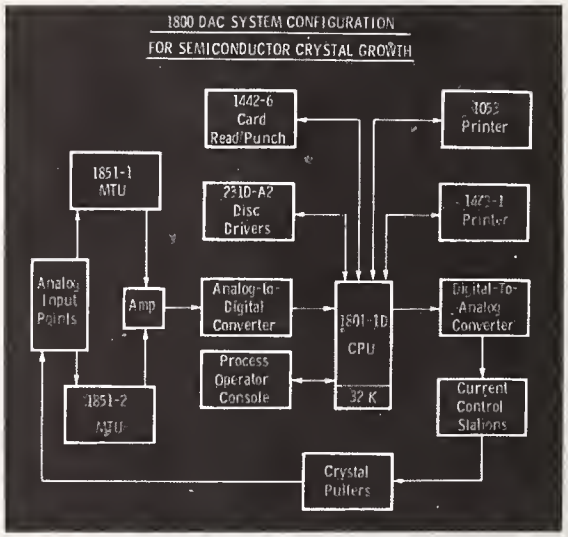


Fig. 14. 1800 DAC system configuration for control of the silicon crystal growth process.

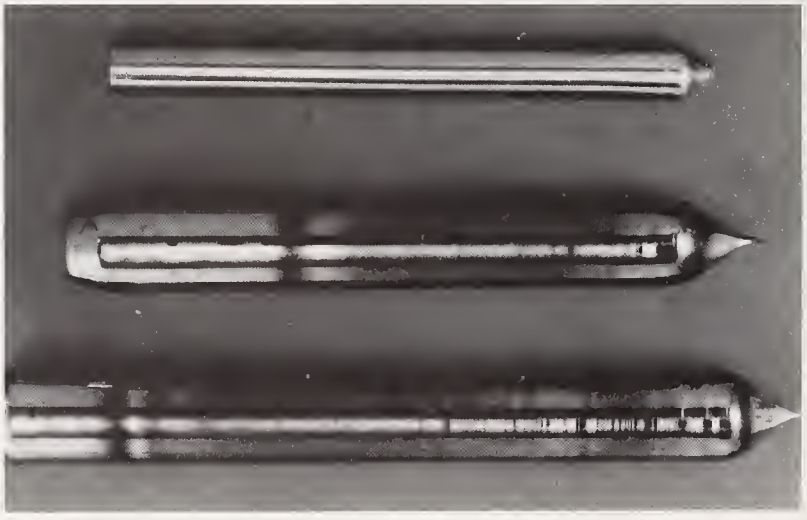


Fig. 15. Silicon crystals grown by use of computer control.

Epitaxial Growth of Silicon

B.A. Joyce

Mullard Research Laboratories
Redhill, Surrey, England

The aim of this review will be to discuss work that has been reported on the fundamental aspects of the growth of epitaxial silicon films and to relate this to practical device processing technology. An attempt will also be made to predict future developments in processing, and to determine the extent to which these may be limited by lack of basic information.

It was clearly illustrated in a recent review by Runyan* that the development of epitaxial silicon deposition technology has reached a stagnant phase, probably because the available technology is adequate to meet the specifications required for the present generation of devices. It is, therefore, perhaps appropriate at this stage to consider what is known of the basic mechanisms involved in growth, and to what extent this information is utilised in process technology. In order to simplify presentation of this material, it will be discussed under five fairly arbitrary topic headings: (i) Growth systems; (ii) Nucleation and growth mechanisms, including substrate surface effects; (iii) Crystallographic defects; (iv) Dopant incorporation and redistributions; and (v) Deposition on insulating substrates.

Growth Systems

Conventionally, for process technology, atmospheric pressure flow systems are used with low partial pressures of the active gas ($SiHCl_3$, $SiCl_4$ or SiH_4) in hydrogen. The limitations on ultimate gas purity and system dynamics imposed by this situation will be discussed in relation to growth temperature restrictions and boundary layer models. A comparison will then be made with the possible alternative methods of deposition, i.e. low absolute pressure chemical systems and vacuum sublimation techniques.

Nucleation and Growth Mechanisms

The nucleation behaviour observed when a molecular beam of silane is made to impinge on a heated substrate will be described, paying particular attention to the influence of surface impurities on both nucleation, and post-nucleation early growth stages. The use of Auger electron spectroscopy in the determination of surface impurities will be outlined with reference to the effectiveness of various substrate surface preparation techniques, including vapour phase etching and thermal treatment under U.H.V. conditions. The relationship between fundamental growth behaviour and layer morphology will be considered, and from this the development of the 'ideal' chemical deposition system, based on the understanding of the growth mechanism, will be described.

Crystallographic Defect Studies

It has been quite clearly established that, with a few possible exceptions, crystallographic defects in epitaxial silicon films occur as a result of substrate surface contamination, and from a technological viewpoint they no longer represent a serious problem for growth at comparatively high temperatures ($> 1000^{o}C$). However, for growth at low temperatures, which is becoming increasingly important, crystallographic defects do form, so a knowledge of the mechanism of their formation is important and the evidence obtained by transmission electron microscopic studies of the early growth stages of layers containing defects will be briefly reviewed.

The problems involved in obtaining defect free growth at low temperatures will be considered in relation to the type of growth system which is used and the critical impurity problems (both in the gas phase and on the substrate surface) of the three basic deposition methods will be compared. An attempt will be made to show that the basic limitation on obtaining defect free growth, at least down to room temperature, is simply one of adequate gas phase and substrate surface purity, most particularly involving the elimination of water vapour, oxygen and carbon-containing materials.

---

* W.R. Runyan, Semiconductor Silicon, Edited by R.R. Haberecht and E.L. Kern p.169, 1969.

The very sparse information relating defects to device failure will be briefly reviewed and the apparent influence of precipitation at defects will be stressed.

Dopant Incorporation and Redistribution

The deliberate introduction of dopants into growing layers is almost universally carried out at present by addition of the hydrides of Groups III and V (B_2H_6, PH_3, AsH_3) to the gas phase reactants. This basic system is probably adequate for all foreseeable requirements, but possible refinements in absolute control will be discussed. A more important problem is the redistribution of impurities originally present in the substrate, particularly since present device specifications require thinner layers than previously. Some discussion of transfer mechanisms will be given, together with the possible influence of process parameters on these.

Silicon Films on Insulating Substrates

No attempt will be made to review the large volume of literature on this topic, but the potential advantage of such a technology, especially for the fabrication of fast access memory arrays, will be considered in relation to the basic limitations imposed by material quality and processing difficulties.

Finally, some consideration will be given to possible future developments in this branch of silicon technology. These are seen as occurring in two directions. One towards a greater degree of automation and control of thickness, resistivity and profiles, particularly for thin (< 3μm) layers by 'on stream' monitoring and computer-controlled feed-back loops, but with no change in the atmospheric pressure type of system (except for the replacement of silicon chlorides by silane). The other towards the development of low pressure systems to facilitate low temperature growth and which, in the less immediate future, may be required for compatibility with ion implantation so that ultimately all processing from polishing to metallisation could be carried out in a single chamber.

Key words: Autodoping, autoepitaxy, crystal defects, deposition systems, epitaxy, growth mechanism, insulating substrates, nucleation, silicon.

Diffusion in Silicon: Properties and Techniques

C. F. Gibbon

Bell Telephone Laboratories, Incorporated
Murray Hill, New Jersey 07974

This paper contains a review of a number of areas of interest in the physics and technology of impurity diffusion into silicon. Some fundamental concepts of the kinetics of diffusion in solids are presented and the mathematics of diffusion for the generally utilized approximations is outlined. The physical interpretation of the diffusion coefficient and its potential dependence on concentration are discussed. Features of the chemistry of a number of diffusion sources such as high temperature trihalide depositions, predeposited doped SiO_2 films and doped silicon in vacuo are described. A brief review of diffusion coefficient data for the common dopants in silicon and SiO_2 is presented. Consideration is then given to the various mechanisms which have been proposed for the diffusion of dopants in silicon. The paper is concluded with a description and discussion of several special or anomalous effects associated with diffusion in silicon.

Key words: Diffusion, silicon

1. Introduction and General Concepts

The study of atom movements in silicon has been and continues to be one of the most important fields associated with silicon device processing. In this paper a brief review of diffusion in silicon will be presented with most emphasis given to the in-diffusion of dopants from the surface such as occurs in the formation of junctions and channels in integrated circuits [1].[1] The first section introduces the subject of diffusion kinetics and notes the various diffusion systems commonly used, the second section concerns the chemistry of diffusion sources, a brief survey of available diffusion data for selected dopants constitutes the third section, a discussion of diffusion mechanisms follows this survey and some anomalous diffusion effects are considered in the fifth and final section.

1.1 Diffusion Kinetics

Quantitative discussions of diffusion kinetics in silicon, or any other material, are based on the assumption that the flux of atoms, J, at a point is proportional to the gradient of concentration, $\vec{\nabla}C$, at that point, this is commonly known as Fick's law and in an isotropic material for a gradient in one direction, x, only is given by

$$J = -D \frac{\partial C}{\partial x} \tag{1}$$

When this flux equation is inserted into the continuity equation, we obtain the differential equation which must be solved in most diffusion problems; in one dimension again

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(D(C)\frac{\partial C}{\partial x}\right), \tag{2}$$

[1] Figures in brackets indicate literature references at the end of this paper.

where t is time. Note that the diffusion coefficient, D, is not assumed to be independent of concentration. In many cases it does appear to be independent of C but in some conditions of importance in silicon processing D is a strong function of C. Several solutions to this equation with D constant are of particular importance. If the indiffusion of an impurity into silicon occurs from a surface source phase of infinite extent and initially uniform composition then one obtains a solution in terms of the error function and the surface concentration, C_s. The solution is:

$$C = C_s\left(1 - \text{erf}\left(\frac{x}{2\sqrt{Dt}}\right)\right) = C_s\text{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) \tag{3}$$

If on the other hand an amount of the dopant $Q(cm^{-2})$ is applied to the surface in a very thin film such that all dopant enters the silicon at t = 0, one obtains

$$C = \frac{Q}{\sqrt{\pi Dt}}\exp\left(-\frac{x^2}{4Dt}\right). \tag{4}$$

This yields the Gaussian or instantaneous source dopant distribution. Figures (1) and (2) compare the profiles, and variations of diffused layer properties with time, for the two source conditions discussed above.

A somewhat better solution for conditions frequently encountered in diffusion technology has been given by Churchill [2] and Owen and Schmidt [3]. In this case a finite source film of thickness ℓ is considered having an initially uniform doping, Cox, and concentration independent diffusion coefficient, Dox, such as would be the case for a doped predeposited SiO_2 source film on silicon. The solution is a series of error functions which reduces to an expression similar to equations (3) for short times, $\sqrt{D_{ox}t} \ll \ell$, and equation (4) for long times $\sqrt{D_{ox}t} \gg \ell$.

1.2 The Nature of the Diffusion Coefficient

Although strictly speaking the diffusion coefficient is simply the constant of proportionality between the flux and the concentration gradient, some further physical significance can be usefully given to it [4]. Comparison of equation (1) with the predictions for the movements of atoms on a lattice given by random walk theory shows that we can set

$$D = ka^2\Gamma \tag{5}$$

where k is a geometrical factor ~ 0.1, a is the distance moved by an atom in a single jump on the lattice and Γ is the jump frequency. In the case of diffusion of a substitutional impurity, where a vacancy or other defect is required on an adjacent site before the impurity can jump, Γ is the product of a second "jump attempt frequency", Γ', and the mole fraction of the required defects N_d. The parameter Γ' in turn can be thought of as the product of the mole fraction of the diffusing atoms in "activated complexes" N_m, i.e., regions having the proper energy and configuration to allow an atom to reach the mid point between two sites, and an atomic vibration frequency ν. Thus

$$D = ka^2\nu N_d N_m, \tag{6}$$

N_d and N_m can be approximated with Boltzman equations containing the entropy and enthalpy of formation of the necessary point defects, ΔS_d and ΔH_d, and the entropy and enthalpy of the activated complexes ΔS_m and ΔH_m, thus

$$D = ka^2\nu\exp\left(\frac{\Delta S_d + \Delta S_m}{R}\right)\exp -\left(\frac{\Delta H_d + \Delta H_m}{RT}\right) \tag{7}$$

The experimentally determined values of D generally obey an Arrhenius equation of the form

$$D = D_o e^{-H/RT} \tag{8}$$

Comparing (7) and (8) we can see that the D_o and H values can be associated with thermodynamic quantities as follows

$$D_o = ka^2\nu \exp\left(\frac{\Delta S_d + \Delta S_m}{R}\right) \tag{9}$$

$$H = \Delta H_d + \Delta H_m. \tag{10}$$

1.3 Diffusion Practice in Silicon

The solutions to the diffusion equation (2) described in sec. 1.1 relate to situations which seldom if ever strictly hold during silicon processing, although they are useful approximations in many cases. The real dopant source phase conditions used in current technology vary widely but in most if not all a film of oxide, generally identified as a glassy mixture of SiO_2 and dopant oxides, is formed on the surface of the silicon; the dopant diffuses from this film into the silicon. Table I presents a tabulation of the essential features of a number of currently used diffusion systems including those systems in which the source is pre-deposited and those in which the source film grows at the diffusion temperature.

TABLE I. Tabulation of Diffusion Source Systems.

General Type	Specific Source	Comments
High Temperature Sources	Dopant Trihalides	Source film forms in oxidizing ambient at diff. temp.
	Dopant Hydrides	Source film forms in oxidizing ambient at diff. temp.
	Boron Nitride	Source film deposited from oxidized BN disks
Pre-deposited Oxides	Anodic Oxide	Source grown by anodic oxidation of silicon
	Doped Ethyl Silicate	Source pyrolytically deposited at 500-700°C
	Silane and Dopants	Source deposited by oxidation of silane at 300-400°C
	Doped Silicon	Source applied by reactive sputtering
Other	Spun-On Liquid	Proprietary liquid applied with photoresist spinner
	Doped Silicon	Source either powder or wafer; diffusion in vacuo
	Ion Implantation	After implantation, dopant activated and driven-in

2. Diffusion Source Systems

2.1 High Temperature Deposition Systems

In a wide range of diffusion practices the silicon slices are placed directly into a furnace, usually above 800°C, through which a chemical dopant is being flushed in a carrier gas. The dopant is generally in the form of an oxide, although frequently the presence of the dopant oxide in the vicinity of the slice is due to the reaction of some other compound of the dopant with oxygen which is included in the carrier stream. At least three important kinetic processes then occur at the slice surface during such an operation: (1) dopant oxide deposits on the surface (2) the silicon surface is oxidized (perhaps by reduction of the dopant oxide) and dilutes the source and (3) dopant diffuses into the silicon. Generally the process parameters are adjusted so that the concentration of dopant in the source layer formed exceeds that needed to insure that the surface of the slice is saturated with the dopant, once this occurs the surface concentration is insensitive to moderate variations in some processing parameters. However, if this is done, and a lower surface concentration is desired, a second "redistribution" diffusion must be carried out.

Introducing the dopant into the furnace as a trihalide is a frequent practice in these systems, however, a surprisingly limited published literature exists on these sources. Nakamura [8] has studied the boron trichloride system, particularly with regard to the uniformity over a slice. His work indicated that the oxygen flow rate is the most important parameter. He found that at each temperature there exists a critical oxygen flow rate beyond which the resistivity variation across a slice decreases sharply (see Figure 3). He interprets this result in terms of the equilibrium SiO_2-B_2O_3 phase diagram, concluding that at this oxygen flow rate the rate of B_2O_3 deposition is just sufficient to completely liquify the source film and allow complete and continuous mixing.

In two related systems, using PBr_3 and BBr_3, Schmidt [9] has noted that the sheet resistance of the diffused layer drops to a constant value at low (~ 5 cm^3/min. for PBr_3) carrier gas flow rates through the trihalide liquid. Above this flow rate the source glass thickness increases without altering the sheet resistance and raising the flow rate simply pointlessly increases the amount of glass which must be subsequently removed.

Another class of dopant compounds which have received attention recently for high temperature depositions are the gaseous hydrides. Heynes [10] and Duffy, Fay and Armstrong [11] have published work on the use of diborane as a diffusion source. Heynes argues that the H_2O produced by the reaction of oxygen with the diborane is important in providing good uniformity for boron diffusions done this way. The B_2O_3 formed has a very low vapor pressure but the boric acids formed by reaction with H_2O are volatile and ensure uniform dopant distribution. In other papers Heynes and van Loon [12] and Hsueh [13] have described the use of phosphine and arsine, respectively, as high temperature sources of n-type dopants. In general the use of these gaseous compounds seems to allow convenient and precise dopant flow metering, however the need to handle high pressure gas bottles of toxic and potentially explosive gases has perhaps limited their acceptance.

2.2 Low Temperature Deposited Oxide Systems

In stead of introducing the dopant and forming the source film at the diffusion temperature, the doped source layer can be applied onto the silicon surface at low temperature prior to putting it into the drive-in furnace. In this way a uniformily doped film of easily controlled composition can be applied to the slice surfaces without competition from other complicating reactions and the drive-in can be performed in a neutral ambient. Two of the first ways suggested of doing this are by formation of an anodic oxide on the silicon in a solution such that this oxide is appropriately doped as suggested by Schmidt and Owen [14] and by a low temperature pyrolysis of appropriate metal-organic compounds as described by Scott and Ohlmstead [15]. Since these techniques were suggested several others have been utilized with some success. Table I includes a compilation of the more common methods for low temperature formation of doped oxide diffusion sources. Barry and Olofsen [5] have discussed the relative merits of these systems and conclude the low temperature (300 - 400°C) oxidation of appropriate hydrides (silane, diborane, etc.) appears to be the most potentially advantageous technique. This process occurs at lower temperatures than the pyrolysis scheme and less organic contamination is likely, while there is not the limitation on source thickness inherent in the anodization technique.

The deposition of doped oxides by oxidation of silane has received considerable attention in the recent literature [16]. In particular Strater [17] has investigated the controlled oxidation of silane to deposit SiO_2. He finds that the stoichiometry of the reaction is $SiH_4 + O_2 \rightarrow SiO_2 + 2H_2$, another interesting feature of his results is that increasing oxygen flow rate inhibits the reaction. Finally he makes the surprising observation that there is no surface effect on the reaction. Barry and Olofsen [18] have extensively investigated diffusion from such oxidized silane sources (containing boron from diborane). They have shown [5] that boron diffused resistors can be made utilizing predeposited oxides in which the standard deviation of resistor values is only 2% of the median value over a 1-5/8" wafer (the median resistance was not stated). They attribute this 2% principally to mask width variations.

Another interesting technique for applying doped oxides is by reactive sputtering. Nagano and coworkers [19] have published results which indicate that this technique may allow moderate control (±50%) over surface concentrations in a range difficult for chemical source techniques to produce, namely 10^{15}-$10^{17}/cm^3$.

2-3. Diffusions in Evacuated Systems

Another diffusion system which has received some attention recently is diffusion in vacuo, generally in a sealed quartz tube. As pointed out by Armstrong and Duffy [20] this technique is advantageous since it allows a large number of slices to be diffused simultaneously under in a static quasi-equilibrium system in which high uniformity might be expected. However the technique does have the added complication of requiring the evacuation, sealing and opening of quartz tubes with every diffusion.

Armstrong and Duffy [20] utilized powdered silicon appropriately doped as a source in the evacuated tube. They find good correlation between the surface concentration they obtain and source powder doping. Figure 4 shows the effect of the tube seal off pressure on the sheet resistance and its uniformity in this case. A notable variation in these parameters occurs above 10 torr. Workers at Toshiba [21] have used this sort of process to diffuse silicon rectifiers; they report that as many as 300 slices can be diffused at one time with variations of ±1% in surface concentration and junction depth; bulk life-times of 100-300μsec are maintained in this process. Dahlberg et al. [22] have discussed the use of a dynamic vacuum in place of sealed tubes for vacuum diffusions; in addition they have shown that doped silicon wafers can be used as sources in place of powders.

2-4. Other Diffusion Sources

Very briefly two other potentially important source, application techniques should be mentioned. One is the so-called spin-on source; in its most commonly used form it is a proprietary liquid which is spun on to the surface of a silicon slice and hardens into a film. Although it is apparently being evaluated in a number of laboratories, the author knows of no published literature on this technique aside from commercial circulars.

The use of ion implantation as a diffusion source as is the case where subsequent high temperature treatment is used to redistribute ions after an implantation, is an increasingly important subject of investigation. An interesting observation in this regard is that of Pavlov and his coworkers [23] who have noted an anomalously rapid p-n junction movement in diffusions of boron at low temperatures ($\sim 900°C$) after implantation. In the area of application of ion implantation as a diffusion source in device processing Foxhall and Moline [24] have demonstrated the high degree of uniformity and control which can be obtained with this technique.

3. Diffusion Data in Silicon and SiO_2

3.1 Diffusion of P, As, B and Si in Silicon

In making calculations of the rates of diffusion processes in silicon accurate values of the diffusion coefficients of the various species are required. Compilations of such values have been made by several authors [1] in the past and most recently by Kendall and DeVries [25]. It is of interest, however, to consider the diffusion coefficient data available for several common dopants to observe its consistancy and give the reader an insight into the state of knowledge of these parameters.

Figure 5 presents measurements by three workers of the diffusion coefficient of phosphorus in silicon. The diffusion coefficient of phosphorus is known to be a strong function of concentration [26] as are the coefficients of several other common dopants. The values presented in Fig. 5 are for surface concentrations below the limit of about 3×10^{19} cm^{-3} for which this concentration dependence apparently disappears [26]. Thus these values should be reproduced from system-to-system. The values given by Maekawa [27] were found by studying the changes sheet resistance and junction depth with diffusion time and assuming that the profile of equation (3) holds; the data by Mackintosh [28], taken in the same way, agree well with Maekawa's measurements. However, Tannenbaum's [26] data, obtained by applying Boltzman-Matano analysis [4] to conductivity profiles in the diffused layers at concentrations below the level of doping where D depends on C, are at considerable variance with the previous data. The origin of this difference is not clear although it indicates that our knowledge of the diffusion coefficient of a dopant as common as phosphorus at low concentrations is still not complete.

Figure 6 contains data from several workers on diffusion of arsenic at relatively low concentrations. Included are measurements by the observation of junction depth and sheet resistance variation with time [13, 29]; also included is a recent tracer measurement by Masters and Fairfield [30]. Agreement here is relatively good but a variation of $\sim$ 50% can still be seen around 1150°C. The data for boron diffusion (Fig. 7) at surface concentrations of 10^{18} - 10^{19} cm^{-3}, measured indirectly by the technique of sheet resistance and p-n junction observation [18, 31] are consistent with the three D values determined by proton activation of boron by Maekawa and Oshida [32]. However, data taken at much lower surface concentrations (10^{15}-$10^{17}/cm^3$) by Nagano and coworkers [19] gives much lower D values suggesting that even in the low concentration range boron diffusion is still a strong function of concentration.

Finally recent tracer measurements [33, 34, 35] of silicon diffusion in intrinsic silicon are assembled in Fig. 8.

3.2 Diffusion of Light Elements into Silicon

Although diffusion of light elements such as carbon, nitrogen and oxygen into silicon is seldom done intentionally, the diffusion behavior of these species should be of considerable interest since they are common processing contaminants. Very little has been done however on the direct measurement of the diffusion of these elements. One set of measurements of the diffusivity of carbon has been made by Newman and Wakefield [36]. They give $D = 1.9 \exp(-3.1 \pm 0.2 eV/KT)$ cm^2/sec for carbon diffusing into silicon. The author knows of no measurements of nitrogen in silicon, although Kendall and DeVries [25] cite a Russian paper from which they obtain an activation energy of 4.55 eV, however, no actual D values are tabulated. Somewhat more data exists for oxygen diffusion; Logan and Peters [37] found $D = 135 \exp(-3.5 \text{ eV/KT})$ cm^2/sec for oxygen by observing the formation of p-n junctions by oxygen related donors. Corbett and his coworkers [38] have also measured the diffusion coefficients of O in Si by studying the relaxation of stress induced dichroism in the infrared spectrum of the oxygen doped silicon; they find $D = 0.23 \exp(-2.561 \pm 0.005 \text{ eV/KT})(cm^2/sec)$.

3.3 Diffusion in Thin Films of SiO_2

The characteristics of diffusion of dopants in thin films of SiO_2 are of interest in diffusion masking and, of more recent importance, in diffusion from predeposited oxides. However, very limited data exists in this area. The diffusion of phosphorus in SiO_2 films has been studied by observing masking behavior on silicon by Sah, Sello and Trēmere [39]. They find that the phosphorus reacts with the SiO_2 to form a sharply defined phosphosilicate glass phase whose boundary migrates into the SiO_2; masking is lost when this boundary reaches the silicon surface. They found the diffusion coefficient of phosphorus in the phosphosilicate glass to be about one-tenth of its value in the silicon. Thurston et al [40] studied P diffusion in SiO_2 with radiotracers and noted that it strongly depended on the phosphorus concentration.

Horiuchi and Yamaguchi, [41] using mask failure studies of the type first suggested by Sah and his associates [39], find the diffusivity of boron in SiO_2 to be about 10^4 smaller than in Si. They note the formation of a distinct borosilicate layer in the SiO_2 film at high B_2O_3 activities.

4. Mechanism of Diffusion of Atoms in Silicon

The mechanism of diffusion of both silicon and impurity ions in silicon has been the subject of a reasonable amount of recent discussion, for instance by Kendall and DeVries [25] and by Seeger and Chik [42]. Three mechanisms of diffusion are generally considered for the common group III and V dopants although others have been invoked for certain impurities. The first and simplest of these is the vacancy exchange mechanism in which the diffusing ion advances by exchanging with a single lattice vacancy. The second mechanism is a similar one except that the exchange is with one vacancy in a divacancy or vacancy pair. The last mechanism relies on an extended defect of a type which has received little attention in the study of defects in crystals. This is the extended or strongly relaxed interstitial which can be portrayed as a locally melted area involving perhaps 10 atoms in which one more atom is present than would be in a similar undisturbed volume of the crystal. Passage of such a defect through the crystal would transport atoms in the same sense an ordinary interstitial would.

Fairfield and Masters interpret their silicon self-diffusion results [33] and their arsenic diffusion data [30] in terms of a simple vacancy exchange mechanism. However, they note [33] that they cannot exclude the possibility that vacancy complexes (such as divacancies) play an important role. The dependence of the arsenic diffusivity on background arsenic doping which they find is interpreted as indicating that the migration of the arsenic (donor) ion is enhanced by the presence of negatively charged (acceptor) vacancies.

Seeger and Chik [41] note that the principle difficulty in accounting for self-diffusion data in silicon is the need for the establishment of a mechanism which yields the very high value of the entropy of self-diffusion, ($\Delta Sd+\Delta Sm$) in equation (9), implied by the high D_o values found. (D_o is $\sim 10^4$ cm^2/sec for Si [33] while for face centered cubic metals it is $\sim 10^{-1}$ cm^2/sec [42].) They propose the motion of the extended or relaxed interstitial described above as explaining these results and give estimates of the entropies involved which indicate that such a mechanism could provide an explanation for the high D_o values. They also suggest that the migration of group III and V impurities could occur by this mechanism.

Finally, Kendall and DeVries [25] have attempted to interpret a considerable mass of diffusion data in silicon. In particular they argue that the activation energy for self-diffusion of silicon ($\sim$ 5 eV [33]) is far too large to be reconciled with theoretical estimates of the enthalpies of single vacancy formation and motion. The offer the divacancy as an alternative candidate for the defect active in migration and support their contention with an interpretation of the activation energy on the basis of this model which yields the theoretically estimated enthalpy of formation of single vacancies. They show that the apparent dependence of the boron diffusion coefficient on boron doping is consistent with a divacancy mechanism for the migration of that impurity.

5. Diffusion at High Concentrations and Other Effects

5.1 Dependence of Diffusion Coefficient on Concentration

At several points above the fact that the diffusivity, D, might be a function of concentration was briefly alluded to. It was noted that constant D solutions of Eq. (2) might not be applicable in some cases; the diffusion data given in section 3 was for relatively low surface concentration (and substrate dopings) in an effort to minimize concentration as a source of inconsistancy; and account was taken of certain concentration dependent effects in discussing diffusion mechanisms in the last section. One might expect the diffusion coefficient to be truly a constant at concentrations low enough that the Fermi level, and thus the carrier densities, in the silicon are controlled by thermally generated holes and electrons at the diffusion temperature and not by the doping density, and low enough that the diffusing atoms are sufficiently separated that they would be unlikely to interact directly with each other through elastic or electric disturbances. Since the intrinsic carrier density is $\sim 10^{19}$/cm^3 at 1000°C this might be a reasonable, but hardly absolute, upper limit of concentration independent diffusion to expect in normal diffusion processing.

Above this it is to be expected that diffusion behavior may vary significantly with concentration. Perhaps the most classic demonstration that this is true is found in the work of Tannenbaum [26] previously discussed. Figure 9 gives her data on the dependence of the phosphorus diffusion coefficient on phosphorus concentration at 1050°C. It can be seen that the diffusion coefficient rises sharply above $\sim 5\times10^{19}$/cm^3. There are roughly three categories of effects which might account for this type of behavior:

(1) Mechanical Effects. Queisser [43] and Prussin [44] discovered nearly simultaneously that heavy diffusions of phosphorus and boron into silicon generated dense arrays of dislocations. These authors relate this generation to the relief of elastic strain introduced by the impurity ions which in this case are smaller than the lattice ions. Such a network of dislocations could effect diffusion in two ways as discussed by Lawrence [45]. First, dislocation cores may act as high diffusivity paths for dopant ions, i.e., diffusion "short circuits"; and second, the nonconservative movement of dislocations can locally enhance the concentrations of point defects such as vacancies which may influence the diffusion coefficient. Lawrence then goes on to show experimentally that only dislocation networks freshly generated during diffusion lead to diffusion enhancement and concludes that the second mechanism above is the principal one active.

(2) Electrical Effects. Several authors have pointed out that, if no space charge is to form during in-diffusion of electrically active ions, then a field must be present which tends to accelerate the ions and retard the associated carriers [46]. Such a field would give rise to an anomalously high diffusivity at high concentration gradients. Kennedy [47], however, has recently criticized this model pointing out that at normal diffusion temperatures the separation between ions would be the same as or greater than the Debye length (or electrostatic shielding distance) even at high concentrations. He concludes that it is doubtful that this built-in field has any significant effect on diffusion of impurities into silicon.

(3) Electrochemical Effects. Millea and others [46,48] have discussed the potential influence of the Fermi level on the concentration of certain defects important in diffusion. They point out that, if one assumes that vacancies have acceptor levels in the band gap of silicon, increasing the n-doping (raising the Fermi level) will tend to increase the concentration of negative charged (acceptor) vacancies. Masters and Fairfield [30] invoke this mechanism to explain the enhancement of the diffusion coefficient of arsenic with increasing background arsenic doping.

5-2. Emitter-Dip Effect

One particularly troublesome manifestation of the enhanced diffusion effects just discussed is the so-called emitter dip effect. It has been found by a number of workers [46] that diffusion of the emitter of a transistor into the previously diffused base layer causes an anomalous outward movement of the base-collector junction immediately under the emitter. An excellent example of the effect is shown in Fig. 10 (supplied through the courtesy of R. J. Jaccodine). The presence of this effect has limited the production narrow base widths in microwave transistors. All three of the mechanisms described in the preceding section have been invoked to explain this effect.

One of the most recent and satisfying of these explanations (although perhaps not the only reasonable one) is that of Hu and Yeh [49]. They base their mechanism on the production of vacancies by the network of dislocations generated by the in-diffusing emitter impurities. The vacancies produced enhance the effective base diffusivity under the emitter diffusion causing the anomalous base-collector junction movement.

6. Summary

In summary perhaps it would be useful to point out several areas where more extensive work on diffusion might be profitable.

(1) In order to improve the control, range, uniformity and capacity of chemical diffusion source systems much more systematic work is required to understand the fundamentals of deposition kinetics in flow systems and apply this knowledge to the engineering of diffusion systems than, judging from the published literature, has been done.

(2) A good deal better characterization of dopant diffusion in SiO_2 and typical silicate thin film diffusion sources would be valuable in learning how to obtain better control of, for instance, surface concentration, in chemical source diffusions.

(3) Isoconcentration diffusions of tracers into equilibrated silicon slices such as those carried out by Masters and Fairfield [30] could certainly be profitably done for a variety of dopants to untangle effects due to uniform doping from those due to concentration gradients. Such systematic studies might clarify the precise nature of the diffusion coefficient dependence on concentration. Extension of these measurements to lower temperatures (600-900°C), though difficult, would be valuable as more processing is carried out at low temperatures, e.g., in conjunction with ion implantation.

7. References

[1] Both Runyan, W. T., Silicon Semiconductor Technology, McGraw-Hill, New York, 1965, and Grove, A.S., Physics and Technology of Semiconductor Devices, J. Wiley and Sons, New York, 1967, have useful and somewhat more extensive treatments of diffusion in silicon.

[2] Churchill, R. V., Operationial Mathematics, McGraw-Hill, New York, 1948, p. 141.

[3] Owen, A. E., and P. F., Schmidt, Diffusion from a Plane, Finite Source into a Second Phase with Special Reference to Oxide-Film Diffusion Sources on Silicon, J. Electrochem. Soc., 115, 548 (1968).

[4] Further discussion and references can be found in the excellent textbook: Shewmon, P. G., Diffusion in Solids, McGraw-Hill, New York, 1963, p. 40.

[5] Barry, M. L., and P. Olofsen, Advances in Doped Oxides as Diffusion Sources, Sold State Tech. , 39 (Oct., 1968).

[6] Grove, A. S., Physics and Technology of Semiconductor Devices, J. Wiley and Sons, New York, 1967, p. 38

[7] Irvin, J. C., Resistivity of Bulk Silicon and of Diffused Layers in Silicon, Bell Syst. Tech. J. 41, 387 (1962).

[8] Nakamura, T., Diffusion of Boron in Silicon, N.E.C. Research and Development No. 11, 93 (April, 1968).

[9] Schmidt, R., IEEE Intern. Elect. Device Conf., Washington, D. C., 1966, also unpublished data by R. A. Maher using Schmidt's ideas.

[10] Heynes, M. S. R., Boron Duffusion into Silicon Using Diborane, Electrochem. Tech., 5, 25 (1967).

[11] Duffy, M. C., D. W. Foy and W. J. Armstrong, Diborane for Boron Diffusion into Silicon, Electrochem. Tech., 5, 29 (1967).

[12] Heynes, M. S. R., and P. G. G. van Loon, Phosphorus Diffusion into Silicon Using Phosphine, J. Electrochem. Soc., 116, 890 (1969).

[13] Hsueh, Y. W., Arsenic Diffusion in Silicon Using Arsine, Electrochem. Tech., 6, 361 (1968).

[14] Schmidt, P. F., and A. E. Owen, Anodic Oxide Films for Device Fabrication in Silicon. 1. The Controlled Incorporation of Phosphorus into Anodic Oxide Films on Silicon. J. Electrochem. Soc., 111, 682 (1964).

[15] Scott, J., and J. Ohlmstead, A Solid-to-Solid Diffusion Technique, R.C.A. Review 26, 357 (1965).

[16] See for example, Fisher, A. W., J. A. Amick, H. Hyman and J. H. Scott, Jr., Diffusion Characteristics and Application of Doped Silicon Dioxide Layers Deposited from Silane, (SiH_4), R.C.A. Review 29, 533 (1968); and Kern, W., and R. C. Heim, Chemical Vapor Deposition of Silicate Glasses for Use with Silicon Devices. 1. Deposition Techniques, J. Electrochem. Soc. 117, 562 (1970).

[17] Strater, K., Controlled Oxidation of Silane, R.C.A. Review 29, 618 (1969).

[18] Barry, M. L., and Olofsen, P., Doped Oxides as Diffusion Sources. 1. Boron into Silicon, J. Electrochem. Soc., 116, 854 (1969).

[19] Nagano, K., S. Iwachi and T. Tanaka, Solid-Solid Diffusion of Boron in Silicon Using Reactive Sputtering, Japan J. Appl. Phys. 7, 1361 (1968).

[20] Armstrong, W. J., and M. C. Duffy, A Closed Tube Technique for Diffusing Impurities into Silicon, Electrochem. Tech. 4, 475 (1966).

[21] Muraoka, H., T. Kato and M. Nakamura, Recent Diffusion Technique for High Power Silicon Rectifier and Silicon Controlled Rectifier, Toshiba Review, No. 37, 1968.

[22] Dahlberg, R., R. Gereth, A. Kostka, and K. Kreuser, Technique and Kinetics of Solid-Solid Vacuum Diffusion Processes in Silicon, Semiconductor Silicon (R. R. Haberecht and E. L. Kern, Eds.), The Electrochemical Society, New York, 1969.

[23] Ravlov, P. V., V. A. Uskov, E. I. Zorin, D. I. Tetelbaum and A. S. Baranova, Investigation of the Diffusion of Boron in Silicon from a Layer Doped by Ion Bombardment. Soviet Physics-Solid State 8, 2221 (1967).

[24] Foxhall, G. F., and R. A. Moline, Advantages of Ion Implantation in Fabricating Hyperabrupt Diodes, IEEE Intern. Elect. Dev. Conf., Washington, D. C., 1969.

[25] Kendall, D. L., and D. B. DeVries, Diffusion in Silicon, Semiconductor Silicon (R. R. Haberecht and E. L. Kern, Eds.) The Electrochemical Society, New York, 1969.

[26] Tannenbaum, E., Detailed Analysis of Thin Phosphorus Diffused Layers in P-type Silicon, Solid-State Elect. 2, 123 (1961).

[27] Maekawa, S., Diffusion of Phosphorus into Silicon, J. Phys. Soc. Japan 17, 1592 (1962).

[28] Mackintosh, T. M., The Diffusion of Phosphorus in Silicon, J. Electrochem. Soc. 109, 392 (1962).

[29 Armstrong, W. J., The Diffusivity of Arsenic in Silicon, J. Electrochem. Soc. 109, 1065 (1962).

[30] Masters, B. J. and J. M. Fairfield, Arsenic Isoconcentration Diffusion Studies, J. Appl. Phys. 40, 2390 (1969).

[31] Kurtz, A. D., and Yee, R., Diffusion of Boron into Silicon, J. Appl. Phys. 31, 303 (1960).

[32] Maekawa, S., and T. Oshida, Diffusion of Boron into Silicon, J. Phys. Soc. Japan 19, 253 (1964).

[33] Fairfield, J. M., and B. J. Masters, Self-Diffusion in Intrinsic and Extrinsic Silicon, J. Appl. Phys. 35, 3148 (1967).

[34] Peart, R. F., Self-Diffusion in Intrinsic Silicon, Phys. Stat. Sol. 15, K119, (1966).

[35] Ghostagore, R. N., Method for Determining Silicon Diffusion Coefficients in Silicon and in Some Silicon Compounds, Phys. Rev. Lett. 16, 890 (1966).

[36] Newman, R. C. and J. Wakefield, Diffusion and Precipitation of Carbon in Silicon, Metallurgical Society Conf., Vol. 15, Metallurgy of Semiconductor Materials J. Wiley, New York, 1962, p. 201.

[37] Logan, R. A., and Peters, A, J , Diffusion of Oxygen in Silicon, J. Appl. Phys. 30, 1627 (1959).

[38] Corbett, J. S., R. S. McDonald and G. D. Watkins, The Configuration and Diffusion of Isolated Oxygen in Silicon and Germanium, J. Phys. Chem. Solids 25, 873 (1964).

[39] Sah, C. T., Sello, H., and Tremere, D. A., Diffusion of Phosphorus in Silicon Oxide Film, J. Phys. Chem. Solids 11, 288 (1959).

[40] Thurston, M. D., F. E. Battocletti and G. M. Rodgers, Impurity Profile in the Oxide Region, Final Report by Ohio State University Research Foundation on Contract DA-36-039-SC-87426, 1964.

[41] Horuichi, S., and Yamaguchi, J., Diffusion of Boron in Silicon Through Oxide Layer, Japan J. Appl. Phys. 1, 314 (1962).

[42] Seeger, A., and K. P. Chik, Diffusion Mechanisms and Point Defects in Silicon and Germanium, Phys. Stat. Sol. 29, 455 (1968).

[43] Queisser, H. J., Slip Patterns on Boron-Doped Silicon Surfaces, J. Appl. Phys. 32, 1776 (1961).

[44] Prussin, S., Generation and Distribution of Dislocations by Solute Diffusion, J. Appl. Phys. 32, 1876 (1961).

[45] Lawrence, J. E., Solute Diffusion in Plastically Deformed Silicon Crystals, Brit. J. Appl. Phys. 18, 405 (1967).

[46] See the recent review by Willoughby, A. F. W., Anomalous Diffusion Effects in Silicon (A Review), J. Mats. Sci. 3, (1968).

[47 Kennedy, D. P., On the Ambipolar Diffusion of Impurities into Silicon, Proc. IEEE 57, 1202 (1969).

[48] Millea, M. F., The Effect of Heavy Doping on the Diffusion of Impurities in Silicon, J. Phys. Chem. Solids 27, 315 (1966).

[49] Hu, S. M., and T. H. Yeh, Approximate Theory of Emitter-Push Effect, J. Appl. Phys. 40, 4615 (1969).

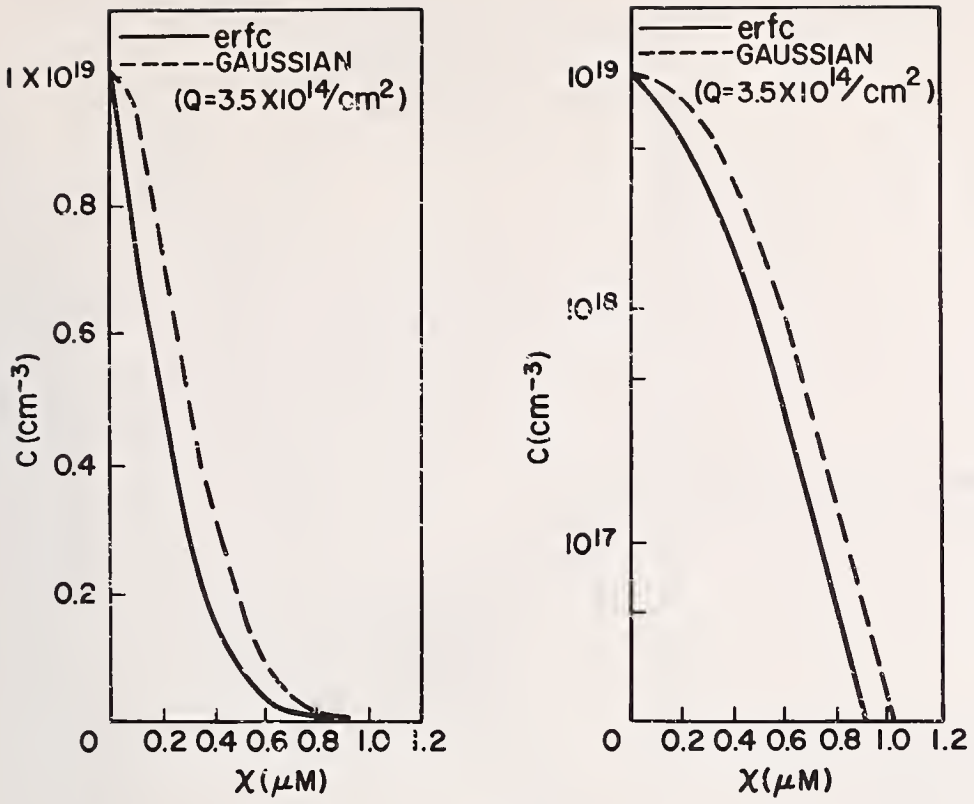


Fig. 1. Calculated diffusion profiles for boron diffused into silicon at 1100°C for 1 hour, D taken from reference [6], Q is the initial source concentration in the Gaussian case.

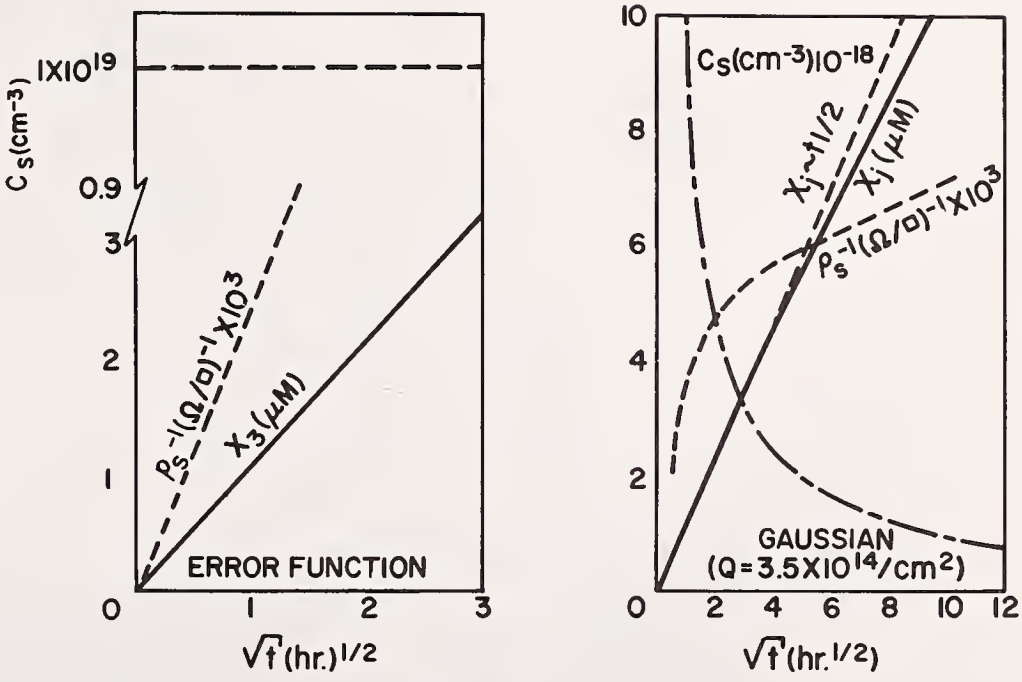


Fig. 2. Calculated diffused layer properties for boron diffused into silicon at 1100°C, sheet resistance calculated using data from reference [7]. "ρ_s" is the sheet resistance, x_j the junction depth, and C_s the surface concentration.

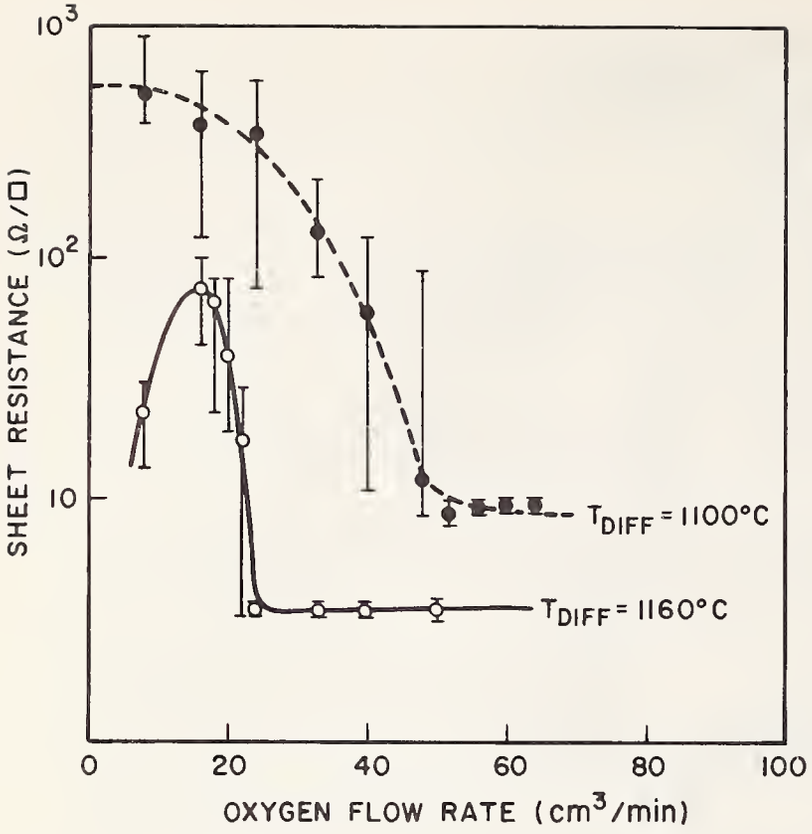


Fig. 3. Sheet resistance versus oxygen flow rate (in a 60 mm I.D. tube) in a diffusion using a BCl_3 source, from Nakamura [8].

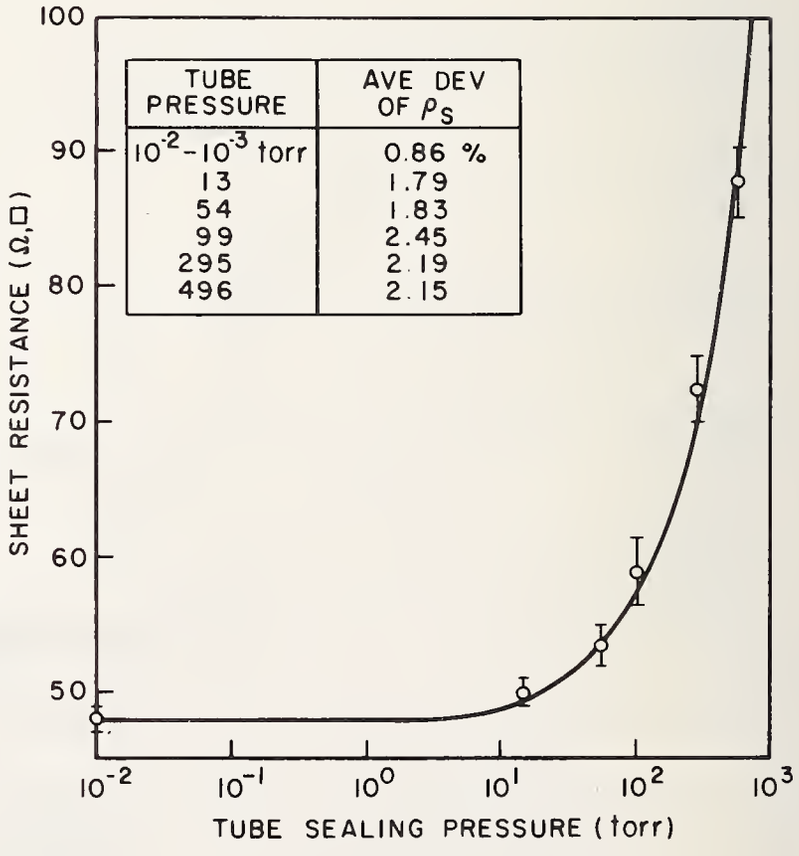


Fig. 4. Effect of background pressure of dry nitrogen on the sheet resistance value and uniformity in a sealed tube diffusion, from Armstrong and Duffy [20].

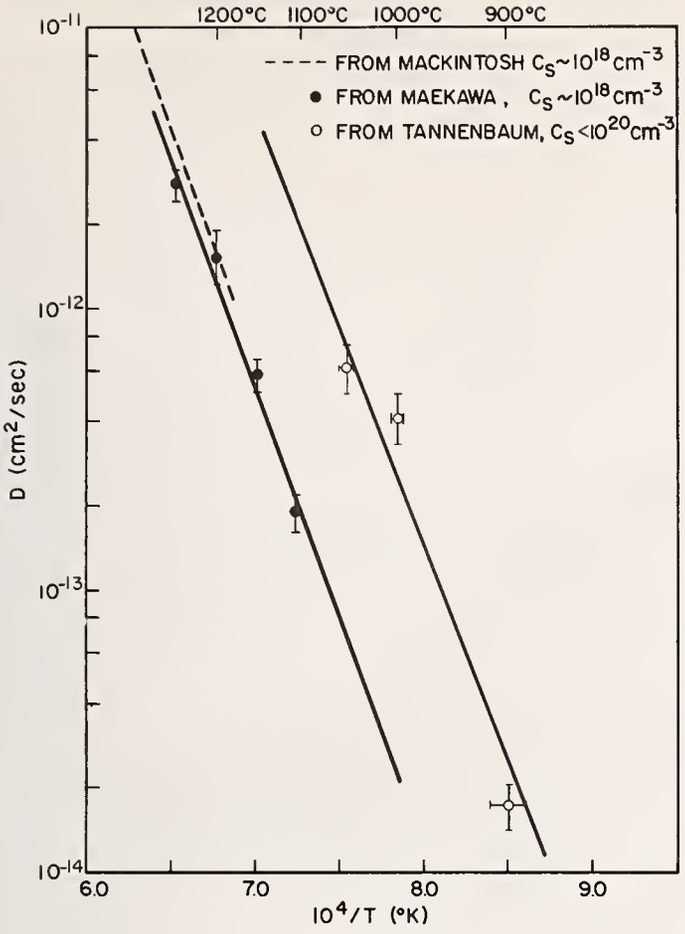


Fig. 5. Diffusion coefficients of phosphorus at low concentrations; measurements by Mackintosh [28], Maekawa [27] and Tannenbaum [26].

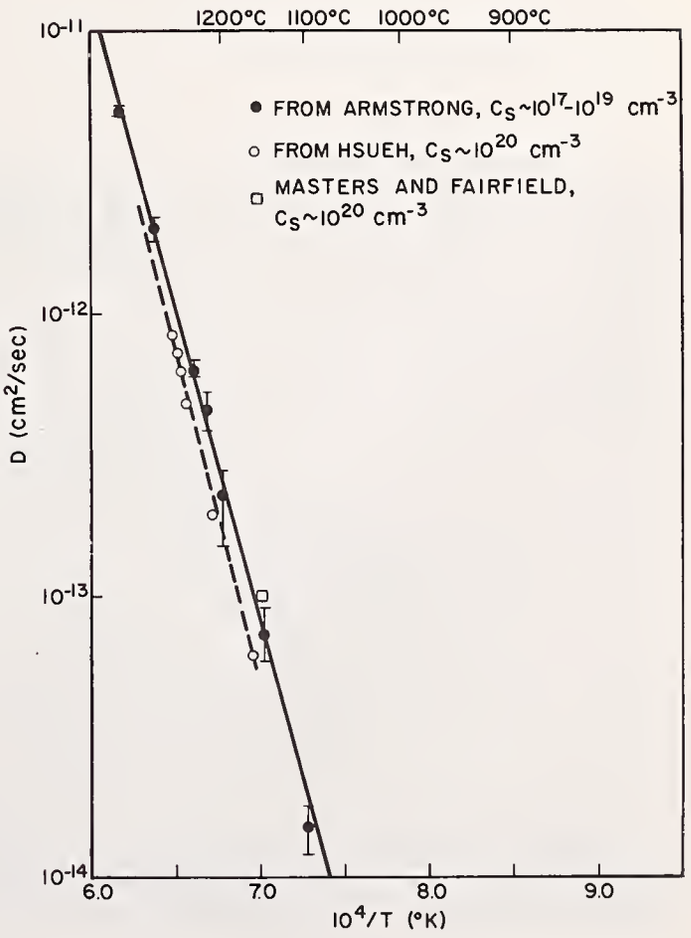


Fig. 6. Diffusion coefficients of arsenic at low concentrations, measurements by Armstrong [29], Hsueh [13] and Masters and Fairfield [30].

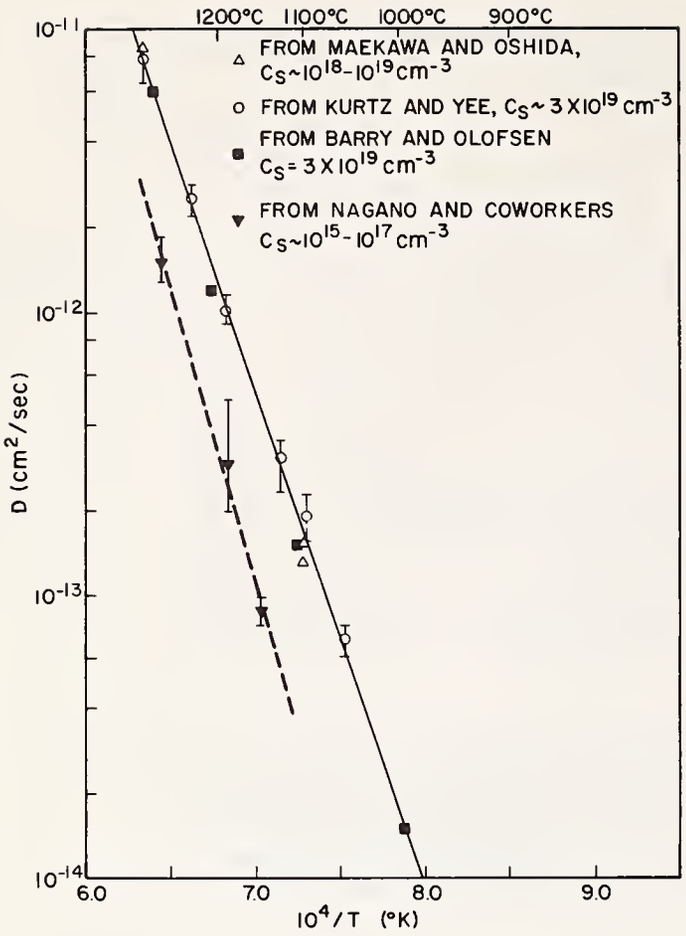


Fig. 7. Diffusion coefficients of boron at low surface concentrations; measurements by Maekawa and Oshida [32], Kurtz and Yee [31], Barry and Olofsen [18], and Nagano and coworkers [19].

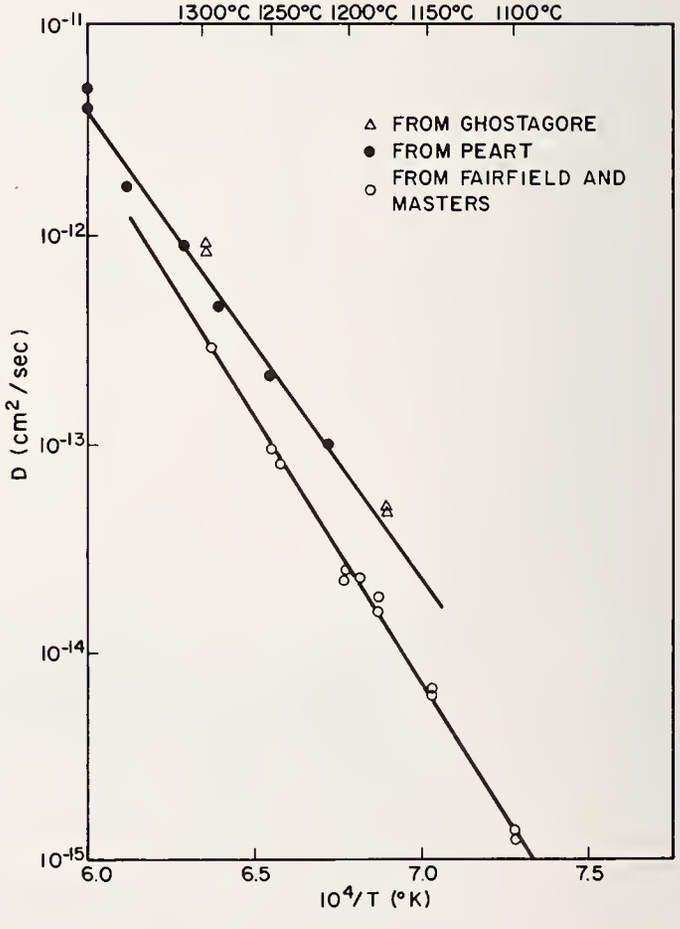


Fig. 8. Diffusion coefficient of silicon in high resistivity silicon; measurements by Ghostagore [35], Peart [34] and Fairfield and Masters [33].

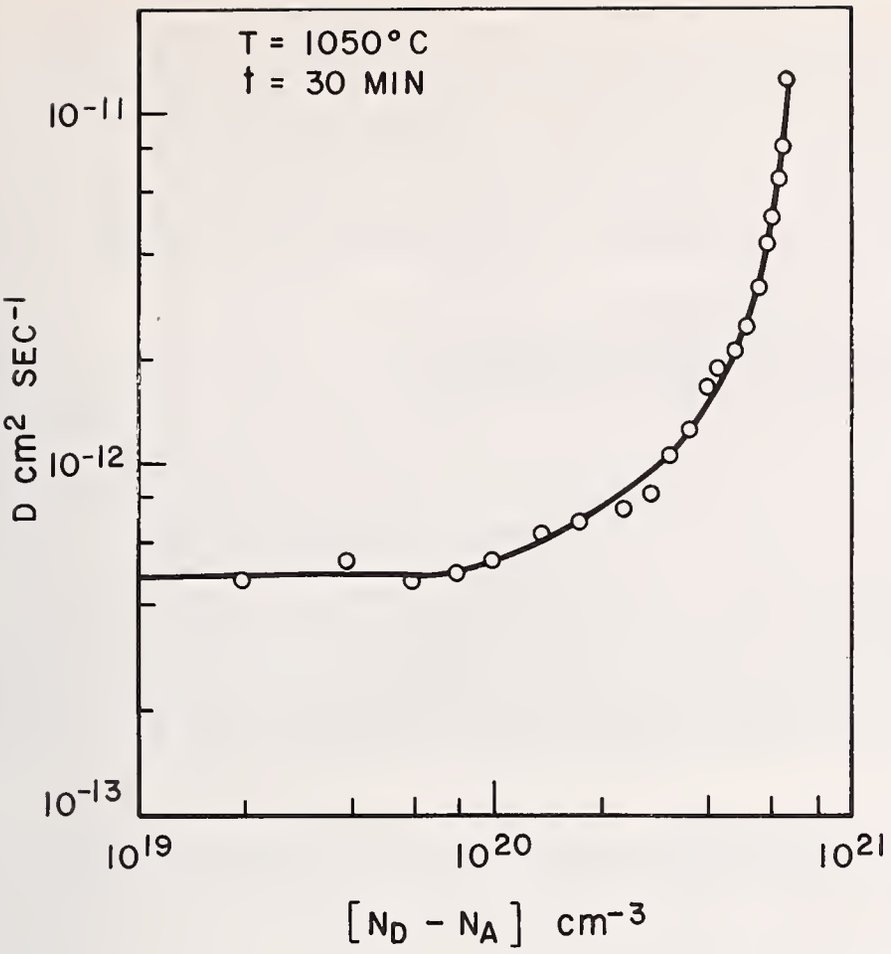


Fig. 9. Diffusion coefficient of phosphorus in silicon as a function of concentration at 1050°C, from Tannenbaum [26].

Fig. 10. Emitter-dip effect in n-p-n structure, photograph courtesy of R. J. Jaccodine.

Measurement and Control of Dielectric Film Properties During Semiconductor Device Processing

Bruce E. Deal
Research and Development Laboratory
Fairchild Camera and Instrument Corporation
Palo Alto, California 94304

The methods for measuring and controlling chemical, physical, and electrical properties of the various types of dielectric films during the fabrication of semiconductor devices are reviewed. Measurement and control procedures for both thermal oxide and combination dielectric films, as well as control procedures for processing materials, are summarized. Electrical properties of dielectric films have been found to be the most significant in regard to resulting device characteristics, and the MIS capacitance-voltage technique is shown to be very effective for monitoring and controlling these properties. Finally, observations are made of the current status of dielectric films in semiconductor technology and of requirements they must satisfy in the future.

Key words: Dielectric films, measurements, passivation, semiconductors, silicon dioxide.

1. Introduction

Dielectric films of various types are used extensively in semiconductor technology. The use of these films — going back more than a decade — includes: masking against doping impurities, passivating junctions, actual components in device structures, dopant sources, and protection of the entire chip against mechanical damage and corrosion. With the development of more complex device and circuit structures which often include several dielectric layers of various types, the measurement and control of film properties — especially during device processing — becomes of primary importance. The film properties which must be reproducibly controlled are chemical, physical, and electrical in nature. Any of these properties can adversely affect device yield and/or reliability. In many cases improper control of dielectric film properties can lead to device failure, either during fabrication or in later operation.

This paper reviews methods of measuring and controlling dielectric film properties — especially during device fabrication. Special emphasis will be placed on the interdependence of the dielectric film properties and other materials and processes used during the fabrication procedure. The types of dielectric films currently being used in silicon semiconductor technology will first be summarized. Next the important properties of these films — as related to device characteristics — will be discussed along with the more important methods for measuring these properties. The significance of the analysis and control of dielectric film properties during device processing will be discussed, with special emphasis on the MIS capacitance-voltage analysis method for thermal oxides as well as combination dielectric films. Also, procedures used to evaluate processing materials will be reviewed.

The present and future status of dielectric films as used in semiconductor technology will be assessed. For instance, what are the trends for their use, and what additional problems will future device requirements cause in terms of formulation, analysis, and control? It is hoped that this review will provide a background for possible solutions to some of these problems.

2. Types of Dielectric Films

Thermal silicon dioxide, formed by reacting silicon with oxygen or water vapor in the temperature range 900°—1250°C, has been used for over ten years to passivate silicon devices [1].* This oxide has several basic characteristics: First, it is almost always stoichiometric SiO_2 and therefore its dielectric properties are very good, e.g., low conductivity, no structural polarization, and little trapping.

*Figures in brackets indicate the literature references at the end of this paper.

Second, since the new oxide is continuously being formed at the Si-SiO_2 interface, the resulting interface properties are reproducible and are not generally dependent on preoxidation surface conditions. Because thermal oxidation depends on solid-state diffusion of the oxidant, the various characteristics of this type of film can be reproducibly controlled, including the physical properties such as thickness. Several reviews of thermal oxide kinetics and properties are available [2-4]. The main disadvantage of thermal oxides is that the open structure permits migration of mobile alkali ions, such as sodium, at fairly low temperatures (200°C). Consequently, surface-sensitive device characteristics have been found to be unstable unless ultra-clean fabrication procedures were followed. Thus it has been found advisable to employ a second dielectric over the thermal oxide to block the alkali ion migration.

A number of types of deposited dielectrics have been evaluated for their ability to prevent ionic migration — both directly on the silicon and over thermal oxide layers. In the majority of cases, the directly deposited dielectric-silicon interface properties were found unsatisfactory for stable device fabrication. Thus most uses of these deposited dielectrics involve combination films, with thermal oxide directly over the silicon device. A number of papers have been written about various deposited dielectrics used in semiconductor technology — many of these are referenced in three reviews [2,5,6], as well as being included in a special Symposium on Dielectric Films sponsored and published by The Electrochemical Society [7]. Following is a brief discussion of the more important types of dielectric films including the advantages and disadvantages of each.

Deposited Silicon Oxides. Silicon oxide films have been deposited using sputtering, vacuum evaporation, and vapor deposition techniques [5]. Although each may have particular advantages, the main problem appears to be stoichiometry control. Deficiencies in either the silicon or the oxygen result in poor electrical properties (see later section). These films have been successfully used, however, to increase the total dielectric thickness or to isolate low-temperature metal (aluminum) interconnections, since all three types of deposition can be carried out below 500°C.

Phosphosilicate Glass. The P_2O_5-SiO_2 glass obtained over thermal oxide regions during predeposition and diffusion processes has been shown to provide two advantages. First, the glass can "getter" impurities from the underlying thermal oxide [8-10]. Second, it can prevent medium-level concentrations of ionic impurities from migrating into the thermal oxide [11,12]. Phosphosilicate glass also has some distinct disadvantages. For one, it exhibits structural polarization which has the same effect in an MIS structure as ionic migration. This polarization is proportional to the amount of phosphorus in the film [13]. Also, phosphosilicate glass is relatively soluble in water and etching solutions. Thus, difficulties arise during photoresist operations and during device operation due to corrosion effects.

Silicon Nitride. The use of silicon nitride films for masking against ionic contamination is now widely known [14]. It is also well known that for most applications the silicon nitride-silicon interface properties do not permit the film to be used directly on silicon devices. Rather a double layer (Si_3N_4 over thermal SiO_2) is the preferred structure, leading to MNOS devices. While the main advantage of a silicon nitride film is its dense structure, disadvantages include the difficulties arising from the control of the stoichiometry. The exact ratio of Si to N in the film is very sensitive to process conditions and thus control of these conditions is extremely important. As electrical characteristics change with composition, so do such properties as etch rate — also a variable that must be controlled. The sensitivity of the nitride layer to trace amounts of oxygen is also an important consideration in its preparation. Another important factor of MNOS structures is the effect of the nitride layer on fast surface state density in the underlying SiO_2-Si interface region.

Other Dielectric Films. In addition to the above-mentioned films, other dielectrics have been used for passivating device structures. Usually deposited over thermal oxides, these films include aluminum oxide, tantalum oxide, boron nitride, organic dielectrics, and sedimented glasses. Anodic silicon oxide has also been used in specific applications. Each of these films has its own peculiar properties, but the control procedures are in general very similar to the ones used for the more common dielectrics described above.

3. Film Properties and Measurement Techniques

As mentioned in the Introduction, the properties associated with characterization and control of dielectric films may be classified as chemical, physical, or electrical. Following is a brief discussion of some of the more important properties and methods of measuring them. Many of these were discussed in a special Electrochemical Society Symposium [15].

Chemical Properties. The etch rate was first used to characterize the dielectric film's structure as well as the degree of defective nature. The measurement is generally made as a function of thickness [5] or by general appearance in the case of defect analysis [16]. Closely related to etch rate is the corrosion resistance of dielectric films. This property is important from the standpoint of protecting

the device structure against various packaging materials and/or ambients. The primary composition of a film can significantly affect its ability to mask against contamination, as well as electrical characteristics of the entire device structure. Methods have been developed for determining the composition of dielectric films. These include infrared analysis [17], X-ray fluorescence [18], ion microprobe analysis [19], and ion backscattering techniques [20]. The latter method, recently applied, appears to be quite promising for determining small deviations from stoichiometry and as a function of depth into the film. In addition, trace impurities can also result in adverse device characteristics [21,22]. For the determination of trace impurities, such tools as spark source mass spectrometry [23] and radiochemical analysis [24] have proved especially useful. One other chemical property of some importance is the nature of the dielectric film surface [25], which can affect photoresist masking operations, the effectiveness of cleaning treatments, and properties of metal and secondary dielectric films deposited over the primary dielectric.

Physical Properties. The most important physical properties of dielectric films in relationship to semiconductor technology are thickness and density. These can directly affect many device properties. Improved control of thickness has been a continuing goal and many methods — destructive and nondestructive — have been developed. Twenty-three of these methods are discussed in Ref. 26. Density is interrelated with thickness and can provide information about other properties such as composition. Mechanical stress of dielectric films may affect device properties [27,28]. Optical properties are also related to thickness determinations, and their general significance is discussed in Ref. 29. An especially useful tool for evaluating optical properties of dielectric films is ellipsometry [30]. Defect structure of dielectric films may also be a physical property as well as chemical. Certainly, irregularities such as cracks and pinholes are important factors in the control of dielectric films [16,31]. The cleanliness of dielectric films may also be considered physical in nature — whatever the classification, it certainly is among the most important properties when considering semiconductor fabrication and passivation. The measurement of surface cleanliness is still, however, one of the most difficult techniques [16,32,33].

Electrical Properties. Of the three classes of film properties described here, the ones that have the most impact and effect on device characteristics are electrical properties. They can result in degraded junction performance in bipolar devices through the formation and variation of charges in the dielectric bulk or at the semiconductor-dielectric interface. In addition, charges in the dielectrics can determine basic characteristics in devices which use the dielectric as an active (MOS transistor) or passive (capacitor) component.

Charges, and instabilities thereof in passivated semiconductor devices, have been the primary subject of hundreds of publications in the past few years. These have been concerned mainly with MIS structures — especially for evaluation purposes [34]. It has been adequately demonstrated that the MOS (or MIS) capacitance-voltage method of analysis is one of the most powerful yet simplest tools for evaluating charge formation and other electrical phenomena in passivated semiconductor structures. In addition to the charges (which will be tabulated below) other electrical characteristics of importance are conductivity or resistivity, dielectric strength or voltage breakdown, and dielectric constant. The measurement of all of these can be made using the basic MIS (metal-insulator-semiconductor) structure.

Four basic types of charges have been shown to be associated with thermally oxidized silicon. These are:

Q_{ss} — Fixed Surface Charge. This positive charge is located in the oxide near the SiO_2-Si interface and is believed to be due to excess silicon ions. Its density can be varied by high-temperature processing treatments and silicon orientation.

Q_o — Mobile Ionic Charge. Alkali ions — Na^+, K^+, and Li^+ — are the main source of this charge. They can migrate in thermal oxide at low temperatures — 100°C or less — and can originate from any processing step.

N_{st} — Fast Surface States. The charges in these states may be positive or negative and will be a function of surface potential. The states are due to the structure of the SiO_2-Si interface, and their density depends on processing conditions and silicon orientation.

N_{ot} — Radiation-Induced Charge. Generation of electron-hole pairs by ionizing radiation results in a positive space charge due to trapped holes.

In addition to these, other types of charges have been found to occur in deposited dielectric films (and thermal oxides under severe conditions). These charges can be electronic, such as trapped electrons or holes, or ionic, such as metallic ions, and result usually because of film nonstoichiometry. They may be fixed or stable charges, or they may charge under normal device operating conditions. As mentioned above, all of these charges can be monitored using the MIS C-V method of analysis. Examples of the measurement of these various types of charges and how they and other electrical properties can be controlled during device fabrication are presented in the next section.

4. Control of Dielectric Film Properties During Device Processing

Some of the important properties — chemical, physical, and electrical — of dielectric films in relationship to device performance were listed in the preceding section. General methods of measuring these properties were also mentioned. In some cases the methods are destructive, difficult and time-consuming; others may be carried out as a part of the device processing. The importance of the various properties varies widely, depending on the particular device and its requirements.

In the author's experience, two areas of control of dielectric films are of prime importance: (1) control of electrical properties of the films, and (2) control of cleanliness during processing of passivated device structures. As it also turns out, both of these can be monitored during processing using MIS structures — mainly involving C-V analyses. Such control procedures have been discussed in part in Refs. 35-38.

In the following section, the measurement of the four general types of charges found in thermally oxidized silicon using C-V plots will be reviewed. Typical procedures and effects arising during the evaluation of thermal oxides during device processing will be given. Finally, the use of C-V analysis for multilayer dielectrics will be discussed.

Measurement of Charges in Thermal Oxides. The theory of MOS physics has been adequately described in the literature [39-41]. For this discussion, it will be assumed that a standard structure, shown in Fig. 1, involves a metal (usually aluminum) field plate, a thermal oxide layer in the thickness range 0.1–0.2 μ, and a chemically-polished silicon substrate. The DC voltage is plotted as a function of capacitance at about 0.1–1.0 MHz. The resulting theoretical characteristics for p- and n-type silicon are also shown in Fig. 1, with the curve shape and exact values of C_{min}/C_o depending on the silicon surface impurity concentration and the oxide thickness [42].

In the absence of any other charge, the value of Q_{ss}, the fixed charge, can be determined from the relationship:

$$\frac{Q_{ss}}{q} = (-V_{FB} + \phi_{MS})\frac{C_o}{q} = (-V_{FB} + \phi_{MS})\frac{k_o\varepsilon}{qx_o} \qquad (1)$$

where V_{FB} = flatband voltage, ϕ_{MS} = metal-semiconductor barrier energy difference, C_o = oxide capacitance, q = electronic charge, k_o = oxide dielectric constant, ε = permittivity of free space, and x_o = oxide thickness.

The lower portion of Fig. 1 shows a typical C-V plot for an MOS structure exhibiting a Q_{ss}/q value of 3×10^{11} cm^{-3} and a ϕ_{MS} value of -1 V. Note that the curve is displaced horizontally in the negative direction, indicating a positive Q_{ss} value. Also note that the experimental curve is exactly the same shape as the theoretical one — indicating no fast states nor nonuniform contamination.

If impurity ions get into the oxide, the additional charges may or may not be observed through the C-V characteristics. The case where impurity ions, such as sodium, are initially at the outer oxide surface is shown in Fig. 2a. Here the C-V plot reflects only the fixed charge Q_{ss}. If, however, the MOS structure is biased positively at 300°C for a minute or two, the sodium ions are driven to the SiO_2-Si interface. The difference between the two C-V plots, ΔV, can be used to obtain a density of the sodium ions Q_o by the relationship

$$\frac{Q_o}{q}\ (\text{charges/cm}^2) = (\Delta V)\frac{C_o}{q} = \frac{\Delta V}{q}\left(\frac{k_o\varepsilon}{x_o}\right) \qquad (2)$$

Figure 2b shows the appearance of the C-V curve after drifting when a nonuniform amount of contamination was initially present. Under these conditions, it is difficult to obtain a quantitative measure of Q_o.

For cases where mobile ions are initially located in the oxide away from the field plate, the amount of Q_o can be determined by measuring the voltage difference ΔV between the C-V plots after both positive and negative drift at 300°C, as shown in Fig. 2c. The value of Q_{ss} can also be obtained from the plot after the negative drift.

Fast surface states, which result from the charging of traps or defects at the SiO_2-Si interface, can also be observed using C-V plots. It has been shown that fast states at particular levels in the silicon band gap will result in distortions in the plot, as indicated in Fig. 3. These states can be annihilated by treatment in H_2 at 300°–500°C [43] or by thermal annealing in an inert ambient with

aluminum or other active metal over the oxide in the same temperature range [44]. Some hydrogen species is believed to be produced by the reaction of the aluminum with residual water on the oxide surface. (This process has been designated "Alneal," the combination of aluminum + anneal.) A more quantitative indication of fast state density is obtained by a C-V plot of an MOS structure at low temperature, e.g., 77°K. Here, due to the change in the Fermi level with temperature, the resulting (A)-state appears at a different position on the C-V curve, as indicated in Fig. 3b. Other more sophisticated methods are available for measuring fast state densities, but the two above are better suited for rapid comparisons. Continuous distributions of fast states will be discussed below.

The fourth charge, N_{ot}, due to ionizing radiation can also be measured using C-V plots. After radiation the positive space charge will cause a negative shift in the C-V curve, similar to ionic contamination, Q_o. This charge will anneal out at 350°C or higher, and unlike the fast states above, neither hydrogen nor an aluminum field plate is required. An example of the C-V shift due to the radiation-induced charge N_{ot} is shown in Fig. 4. It has also been demonstrated that additional fast surface states are produced by ionizing radiation [45]. These appear to differ from those described above in that a more continuous distribution is observed, as indicated by the C-V curve in Fig. 4. In addition, these states can be annealed without the presence of hydrogen at temperatures of 350°C or higher.

Extensive studies of the above charges associated with thermally oxidized silicon have been carried out in the past few years. It has been demonstrated in the author's laboratory that each of these charges can be varied by different processing conditions — leading to device instabilities or changes in device characteristics. These instabilities are locally referred to as the "XVI Drifts of SOT" (Silicon-Oxide-Technology). Thus measurement and control of these charges is very important if stable devices are to be obtained. Some of these effects are discussed in the next section.

Monitoring Charge Instabilities in Thermal Oxides During Processing. The MOS C-V method of analysis has been demonstrated to be unique in its ability to monitor many oxide properties during device processing. In addition to measuring charge densities, it can be used to determine film thickness, surface impurity concentration and redistribution characteristics, film leakage or conductance, and dielectric constant, and it even provides a measure of defect structure or pinhole density. It will be impossible in this review, because of space limitations, to describe each and every effect due to processing steps. However, certain examples considered especially important or representative of the capability of this method will be given.

The basic description of an MOS test structure was described earlier and shown in Fig. 1. An even simpler, although less quantitative, method involves the use of a gold (or mercury) probe. Here, an oxidized silicon slice can be probed directly and reasonable C-V plots obtained. Semiquantitative values of Q_{ss} can be determined as shown in Fig. 5. In addition, larger quantities of ionic contamination in the oxide bulk can be observed; generally this will be nonreproducible, as also shown in Fig. 5. The advantage of the gold probe is that measurements can be made immediately after oxidation or annealing and without additional effects due to metallization and/or photoresist processing.

Measurement of fixed charge Q_{ss} was discussed above. The origin of this fixed charge is believed to be due to excess silicon ions remaining from the thermal oxidation process. Its density can be reproducibly controlled by the final high-temperature thermal treatment in O_2, H_2O, argon, etc. The C-V measurement thus will permit this control, and the relation of Q_{ss} and temperature is shown in Fig. 6 for several ambients and (111) silicon. Other orientations provide correspondingly lower values, with Q_{ss} for (100) silicon being one-third that of (111).

Every device processing step is a potential source of Q_o (Na^+, Li^+, K^+), the mobile positive ions. Each of these steps, i.e., surface preparation, oxidation, metallization, photomasking, assembly, etc., may be monitored using standard MOS structures. An example of this type of evaluation is given in Refs. 46 and 47, where sodium in oxidation tubes and furnaces was monitored for various conditions. In addition, excellent correlation between MOS C-V analysis and actual sodium content in processing materials has been obtained [35]. The sodium analysis of processing materials will be discussed later. It has also been found that negative ions, which are not mobile at 300°C, can contaminate oxidized devices at high temperatures (>900°C) and result in p-type silicon surfaces. The fact that the negative ions do not migrate at 300°C is demonstrated by C-V plots in Fig. 7. After stress, the positive mobile ions are pulled to the Al field plate, leaving the negative ion in the oxide.

The annihilation of fast states, N_{st}, by hydrogen or the "Alneal" process was mentioned earlier. These states, indicated in Fig. 3a as being at somewhat discrete levels, are believed to be due to "nonsatisfied" silicon bonds. These defects can be charged as a function of surface potential and will be eliminated by reacting with hydrogen. The radiation-induced states, on the other hand (Fig. 4b), have a continuous distribution and are believed to be due to the breaking of a silicon-oxygen bond during radiation. A low-temperature anneal is all that is needed to repair the broken bond. The difference between these two annealing mechanisms is readily ascertained by the use of C-V analysis.

Monitoring Instabilities in Combination Dielectric Films. Combination dielectric films are being used more and more to take advantage of the excellent SiO_2-Si interface properties of thermally oxidized

silicon and the added protection of overlying deposited dielectrics such as silicon nitride. All of the possible charge instabilities associated with thermally oxidized silicon still have to be controlled, but additional effects due to the deposited dielectric itself or its interaction with the thermal oxide have to be considered. These additional possible charges and instabilities are indicated in Fig. 8. These charges may result from fields across the structures as in Fig. 8a–c, or they may be present after the deposition of the second dielectric as in Fig. 8d. When these additional effects are added to those possible in the thermal oxide by itself, it is evident that the control procedure may be very complex and difficult to accomplish. Certainly the interpretation of the various effects becomes quite important.

Some examples of measuring and controlling properties of combination dielectric films are presented below. They show the interdependence of film properties and also point out some of the problems of interpretation.

The first example involves the C-V drift testing of a phosphosilicate glass-thermal oxide combination structure. The total polarization ΔV as a function of applied field (at 300°C for 2 minutes) is plotted in Fig. 9 (open circles) for (111) silicon. The plot shows the expected polarization instability under positive bias, but very little under negative bias (normally the polarization should be symmetrical). The picture is clarified, however, if (100) silicon is used. Under the same drift conditions, the expected symmetrical relationship is obtained (closed circles in Fig. 9). The trapping effect of the high negative field (also called Drift VI) on the (111) substrate cancels out the negative-bias polarization. (If higher values of Q_{ss} are associated with the thermal oxide, a greater negative instability is obtained with negative applied fields.) Thus two instabilities operating in opposite directions can provide misleading results, and may or may not result in reliable device characteristics.

A second example demonstrates the effect of dielectric (Si_3N_4) composition on conductivity and thus on the charge buildup instability due to conductivity differences. It has been shown that variations in reactant composition will affect conductivity through the film [48], an example being shown in Fig. 10 [49]. The two extremes of conductivity will provide wide variations in device stability as evidenced by C-V stress tests shown in Fig. 11 [49]. Note that the stress for this case is at 25°C. In addition, the conductivity variation can be correlated with etch rate and stoichiometry of the film as shown in Fig. 12. These composition measurements were made using the backscattering and channeling technique [50]. It is readily apparent that as the film becomes richer in silicon, the conductivity increases and the etch rate decreases.

Still another effect of a deposited silicon nitride film on charge properties of an underlying thermal oxide is indicated in Fig. 13. In this case a thin layer of silicon nitride is deposited over 0.2 μ thermal oxide, an aluminum field plate is deposited using filament heating, and a C-V plot made (Curve 1). Next, the sample is subjected to ionizing radiation and a second C-V plot is obtained (Curve 2) which shows the voltage displacement due to the space charge N_{ot} and the distortion in the plot due to the continuous distributed fast states (similar to Curve 2 in Fig. 4). Finally the structure is annealed at 500°C. Now the C-V plot (Curve 3) shows that most of the radiation-induced charges and fast states are annealed out, but not the original discrete-level states. The dense nitride did not permit a hydrogen species to migrate through to these fast states and annihilate them. These examples are only a few out of many demonstrating the effects and interactions between dielectric films in combination structures.

5. Analysis and Control of Impurities in Processing Materials

In addition to monitoring properties of dielectric films themselves, materials used to process the films can be analyzed for impurity content. This has been done especially effectively for sodium [35, 46,47], which is a chief cause of electrical instabilities and film defects. In these studies, the sodium content was measured independently and the results compared with MOS instability measurements. Very good correlations have been obtained by such comparisons.

The types of processing materials may be divided into three groups — gases, liquids, and solids. Gases include the common oxidizing and annealing ambients such as O_2, N_2, He, argon, and H_2 (also H_2O, but for purposes of this paper, it is treated as a liquid). It has been found that the primary consideration is to use gases from liquid sources if the ultimate electrical stability in dielectric films is to be obtained. MOS C-V tests have actually been used to check the purity of inert gases (argon) by annealing oxidized silicon at 600°C for several hours. In these tests impurities in compressed gaseous argon shift the C-V plot a number of volts. Recently, experiments in the same temperature range have shown a similar effect due to hydrogen [51], which is undoubtedly one of the impurities present (by itself or in hydrocarbons) in compressed gas cylinders.

Liquids include the usual organic solvents such as trichloroethylene, acetone, alcohols, inorganic acids (HF, HNO_3, H_2SO_4), and water. Analyses have shown that good deionized water is the best solvent from the standpoint of not contaminating semiconductor devices, while acetone is one of the worst [35]. C-V monitoring of solvents has been an excellent way of minimizing problems in dielectric films and device structures.

There are many different types of solids used to fabricate semiconductor devices. All can indirectly affect dielectric film properties, since they can be a source of ionic contamination. These solids

include quartz and ceramics used for oxidation tubes and boats, metals such as aluminum used for metallized contacts and interconnections, photoresist materials, construction materials used for furnaces, evaporators, reactors, etc., packaging materials, and dopant sources, to name a few. The effect of impurities on film properties was discussed in an earlier section, and the use of MOS C-V analysis was recommended. However, chemical or radiochemical analysis of these materials is also recommended as a cross-check. All of the possible techniques and results of such analyses are too long to be included in this paper. However, the maintenance of pure processing materials is an ever-present problem.

6. Discussion

Over the past several years, it has been shown that the measurement and control of thermally oxidized silicon dioxide films is very important if the underlying device properties are to be stable and reproducible. A number of film properties — chemical, physical, and electrical — are important in this respect. It has also been demonstrated that the most versatile single evaluation for many of these properties is the MIS C-V method of analysis.

Until a couple of years ago, the thermal oxide was used mainly to mask against dopants during diffusion and then to passivate the junction. Because of its excellent and reproducible interface properties, the thermal oxide was satisfactory for most applications and control of its properties was possible. However, as more complex semiconductor devices and circuits were developed, additional requirements were placed on the dielectric film system. In certain cases the thermal oxide by itself was not sufficient. Combination dielectric film structures resulted.

The use of combination dielectric films in the past two years has made necessary many more analysis and control procedures for the properties of these systems. In addition to the properties of the underlying thermal oxides, which still must be controlled more closely than ever, the properties of the secondary deposited films must be controlled. Especially critical in these deposited films is stoichiometry, which was never a problem with thermal oxides but which affects the electrical, chemical, and physical properties of the passivating system. One especially good measurement that reflects film composition is conductance. This measurement may become as standard as the common C-V "drift" test.

As if the additional effects due to less desirable properties of the deposited dielectric films were not enough, a third general effect to be considered is that due to interaction of the two films. For instance, the deposition treatment of the secondary film may alter the electrical charge properties of the thermal oxide. The nature of the thermal oxide surface may also lead to defects or pinholes in the deposited film. It is also known that certain secondary dielectrics may "getter" or mask against alkali contamination — thus decreasing the control problems.

Another factor that increases the requirements for controlling dielectric film properties is the additional use of the films. The nonhermetic packaging concepts now being developed for economical reasons are such that the entire metallized device must be sealed or passivated against corrosion. This corrosion may be caused by the ambient or the packaging material. Thus the moisture resistance of the dielectric structure is or will be an important factor, whereas previously it was not. Another type of evaluation must therefore be carried out.

These additional requirements placed on dielectric films may lead to departures from conventional inorganic oxide or nitride structures. Already organic polymer films are being investigated. This does not mean that the present films will be eliminated — they will merely be added to. And this just adds more variables to the never-ending control requirements for dielectric film systems.

7. Acknowledgments

The author wishes to thank all those individuals who have contributed to the overall understanding of dielectric film analysis and control through individual or cooperative research programs at Fairchild R&D. In particular, I wish to thank Pat Castro, Phil Fleming, Pat Kodama, Ed MacKenna, Val Rodriguez, Maija Sklar, Ed Snow, and Bernie Yurash.

8. References

[1] Frosch, C. J. and Derick, L., *J. Electrochem. Soc.* 104, 547 (1957).

[2] Donovan, R. P., "Integrated silicon device technology. Vol. VII. Oxidation," Technical Report No. ASD-TDR-63-316, Research Triangle Institute (June 1965).

[3] Revesz, A. G. and Zaininger, K. H., *RCA Review* 29, 22 (1968).

[4] Gray, P. V., *Proc. IEEE* 57, 1543 (1969).

[5] Pliskin, W. A., Kerr, D. R., and Perri, J. A., "Thin glass films," in *Physics of Thin Films*, Vol. 4, p. 257 (Academic Press, 1967).

[6] Chu, T. L., *J. Vac. Science Tech.* 6, 25 (1969).

[7] Vratny, F., *Thin Film Dielectrics* (The Electrochemical Society, 1969).

[8] Yon, E., Ko, W. H., and Kuper, A. B., *IEEE Trans. Electron Devices* ED-13, 276 (1966).

[9] Buck, T. M., Allen, F. G., Dalton, J. V., and Struthers, J. D., *J. Electrochem. Soc.* 114, 862 (1967).

[10] Carlson, H. G., Brown, G. A., Fuller, C. R., and Osborne, J., *Phys. Failure Electron.* 4, 390 (1966).

[11] Kerr, D. R., Logan, J. S., Burkhardt, P. J., and Pliskin, W. A., *IBM J. Res. Develop.* 8, 376 (1964).

[12] Snow, E. H., Grove, A. S., Deal, B. E., and Sah, C. T., *J. Appl. Phys.* 36, 1664 (1965).

[13] Eldridge, J. M., Laibowitz, R. B., and Balk, P., *J. Appl. Phys.* 40, 1922 (1969).

[14] The Electrochemical Society Symposium on "Silicon Nitride," Oct. 12-13, 1966, Philadelphia, Pa.

[15] Schwartz, B. and Schwartz, N., *Measurement Techniques for Thin Films* (The Electrochemical Society, 1967).

[16] Stickler, R. and Faust, J. W., *Electrochem. Tech.* 2, 298 (1964).

[17] Pliskin, W. A., *Thin Solid Films* 2, 1 (1968).

[18] Pink, F. X. and Lyn, V., *Electrochem. Tech.* 6, 258 (1968).

[19] Colby, J. W., in *Thin Film Dielectrics*, p. 491 (The Electrochemical Society, 1969).

[20] Gyulai, J., Meyer, O., Mayer, J. W., and Rodriguez, V., *Appl. Phys. Letters* 16, 232 (1970).

[21] Barson, F., Armstrong, W. J., and Mutter, W. E., *J. Electrochem. Soc.* 111, 1263 (1964).

[22] Mathews, J. R., Griffin, W. A., and Olson, K. H., *J. Electrochem. Soc.* 112, 899 (1965).

[23] Willardson, R. K., in *Measurement Techniques for Thin Films*, p. 58 (The Electrochemical Society, 1967).

[24] Dalton, J. V. and Drobek, J., *J. Electrochem. Soc.* 115, 865 (1968).

[25] Lussow, R. O., *J. Electrochem. Soc.* 115, 660 (1968).

[26] Gillespie, D. J., in *Measurement Techniques for Thin Films*, p. 102 (The Electrochemical Society, 1967).

[27] Fairfield, J. M., Howard, J. K., and Schwuttke, G. H., *Electrochem. Tech.* 6, 111 (1968).

[28] Jaccodine, R. J. and Schlegel, W. A., *J. Appl. Phys.* 37, 2429 (1966).

[29] Axelrod, N. N., *Optical Properties of Dielectric Films* (The Electrochemical Society, 1968).

[30] Zaininger, K. H. and Revesz, A. G., *RCA Review* 25, 85 (1964).

[31] Besser, P. J. and Meinhard, J. E., *Proc. Symposium on Manufacturing In-Process Control and Measuring Techniques for Semiconductors*, vol. 2, p. 16-1 (1966).

[32] Ing, S. W., Morrison, R. E., and Sandor, J. E., *J. Electrochem. Soc.* 109, 221 (1962).

[33] Fisher, A. W. and Amick, J. A. *J. Electrochem. Soc.* 113, 1054 (1966).

[34] Schlegel, E. S., *IEEE Trans. Electron Devices* ED-14, 728 (1967); ED-15, 951 (1968).

[35] Yurash, B. and Deal, B. E., *J. Electrochem. Soc.* 115, 1191 (1968).

[36] Szedon, J. R. and Handy, R. M., *J. Vac. Science Tech.* 6, 1 (1969).

[37] Hofstein, S. R., *Solid-State Electron.* 10, 657 (1967).

[38] White, M. H. and Cricchi, J. R., *Solid-State Electron.* 9, 991 (1966).

[39] Grove, A. S., *Physics and Technology of Semiconductor Devices* (Wiley, 1967).

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

[41] Goetzberger, A. and Sze, S. M., in *Applied Solid State Science*, Vol. 1 (Academic Press, 1969).

[42] Goetzberger, A., *Bell Sys. Tech. J.* 45, 1097 (1966).

[43] Balk, P., Paper presented at the San Francisco Meeting of The Electrochemical Society, May 9-13, 1965, Abstract No. 109.

[44] Balk, P., Paper presented at the Buffalo Meeting of The Electrochemical Society, Oct. 10-14, 1965, Abstract No. 111.

[45] Snow, E. H., Grove, A. S., and Fitzgerald, D. J., *Proc. IEEE* 55, 1168 (1967).

[46] Fewer, D. R. and Carlson, H. G., "Report on Surface Studies," RADC Final Report TR-66-776, March 1967.

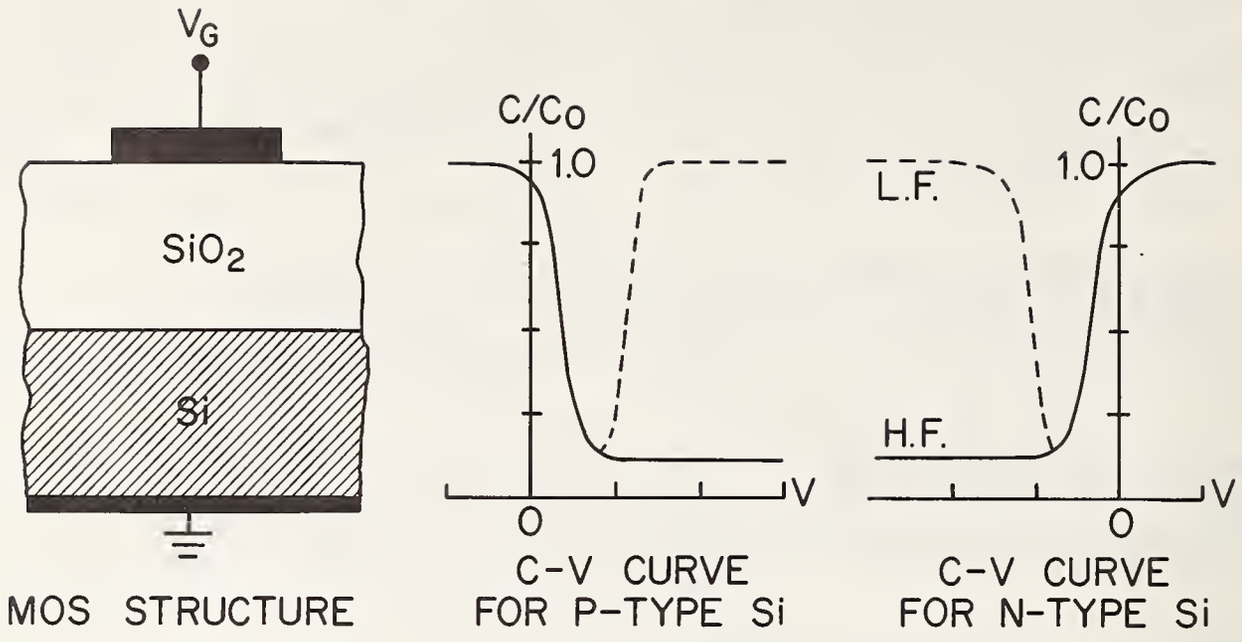
[47] Burgess, T. E. and Donega, H. M., *J. Electrochem. Soc.* 116, 1313 (1969).

[48] Doo, V. Y., Kerr, D. R., and Nichols, D. R., *J. Electrochem. Soc.* 115, 61 (1968).

[49] Rodriguez, V., MacKenna, E. L., and Fleming, P. J., To be published.

[50] Gyulai, J., Meyer, O., Mayer, J. W., and Rodriguez, V., To be published.

[51] Castro, P. L. and Deal, B. E., To be published.



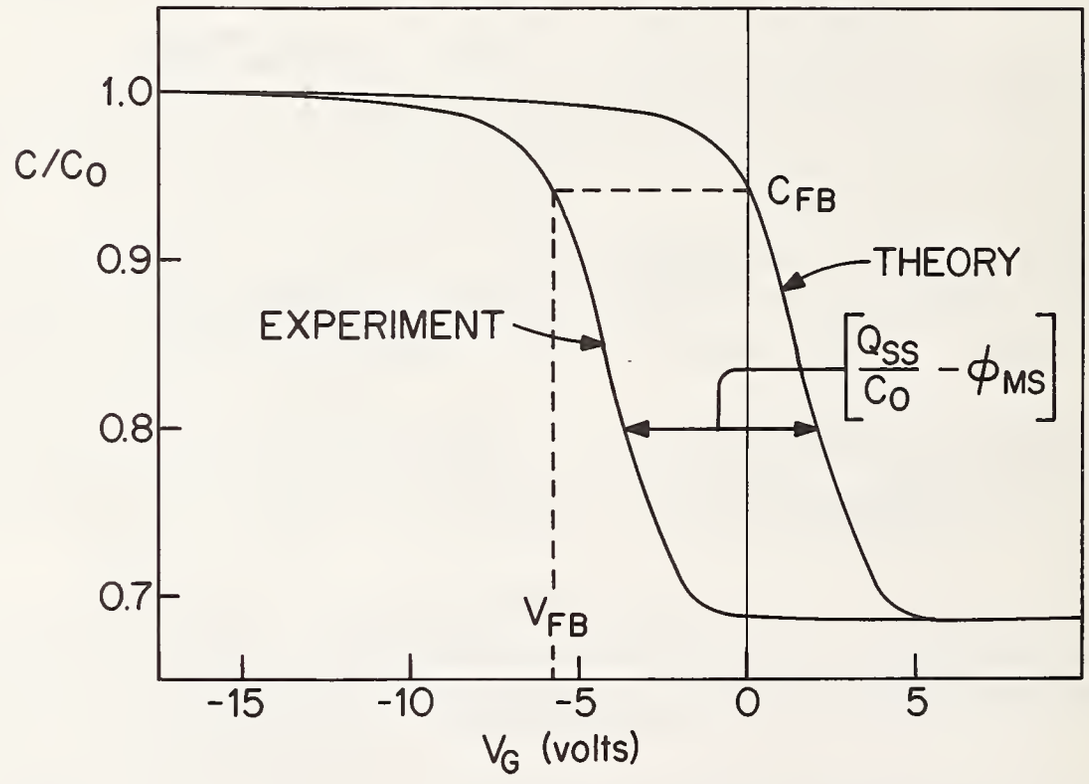


Fig. 1. Typical examples of MOS structures (upper left), C-V curves for p- and n-type silicon (upper right), and measurement of Q_{ss} using MOS C-V plot (bottom).

Note that for this and subsequent figures, $C_B \sim 1 \times 10^{16}$ cm^{-3}, silicon orientation is (111) and $X_0 = 0.2$ μ, unless otherwise specified.

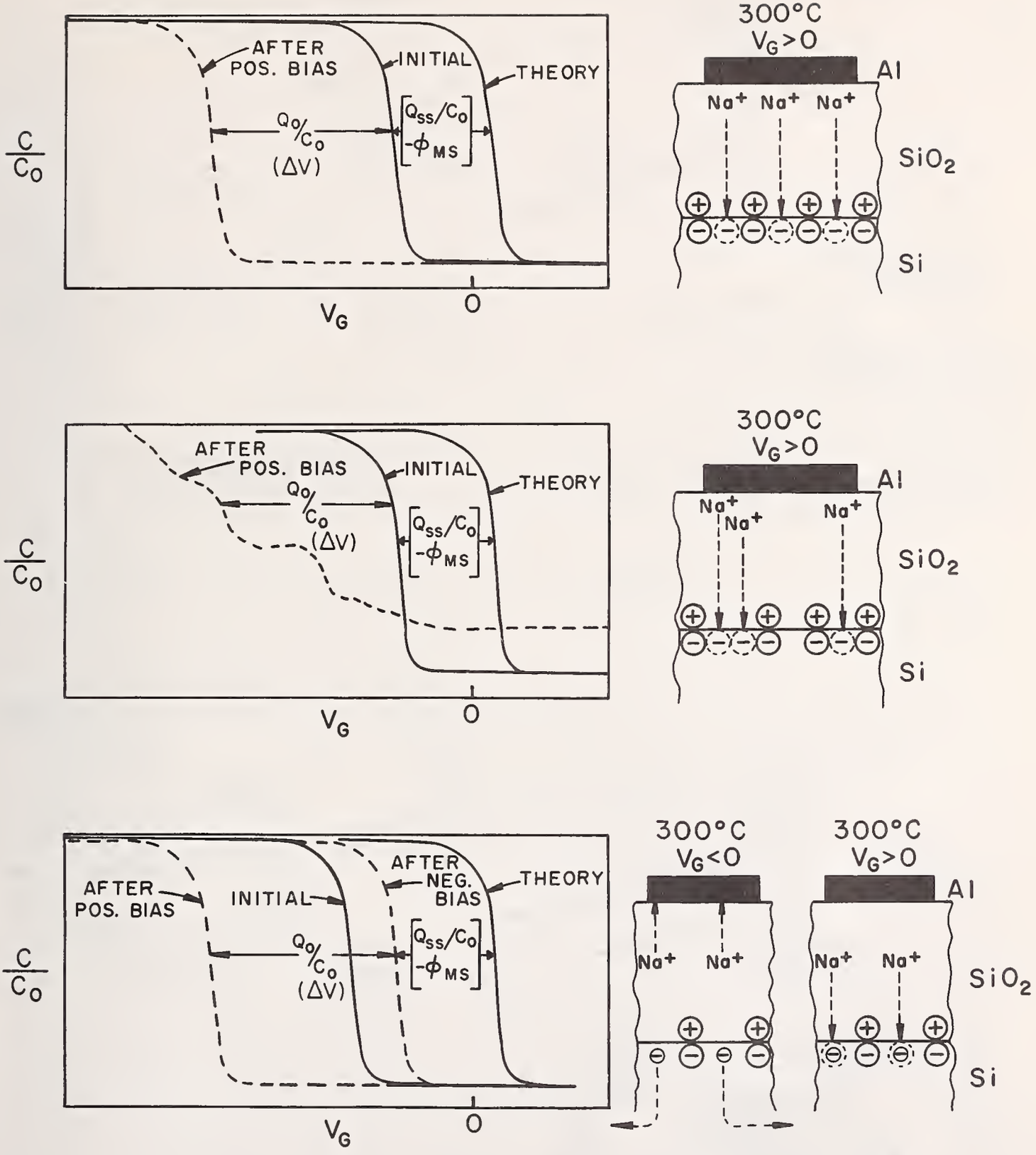
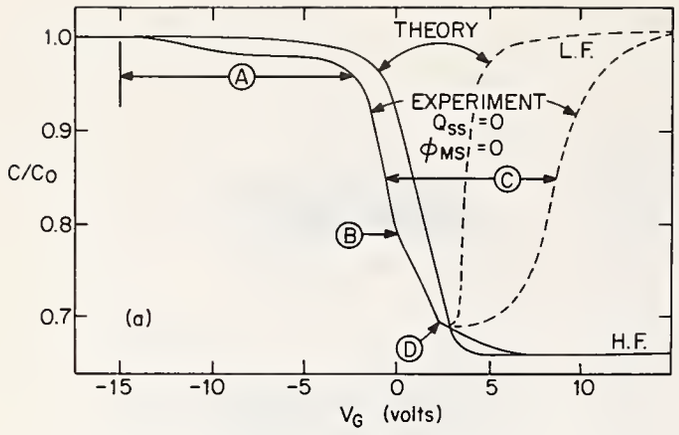


Fig. 2. Examples of three types of ion (Na^+) drift as indicated by MOS C-V plots.



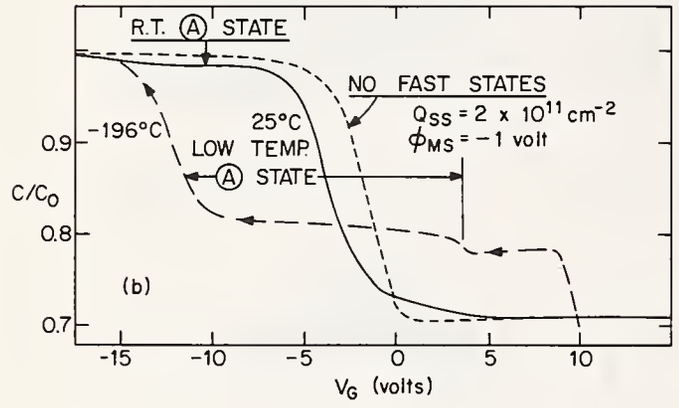


Fig. 3. Examples of fast surface states as indicated in MOS C-V plots (a) 25°C, (b)77°K.

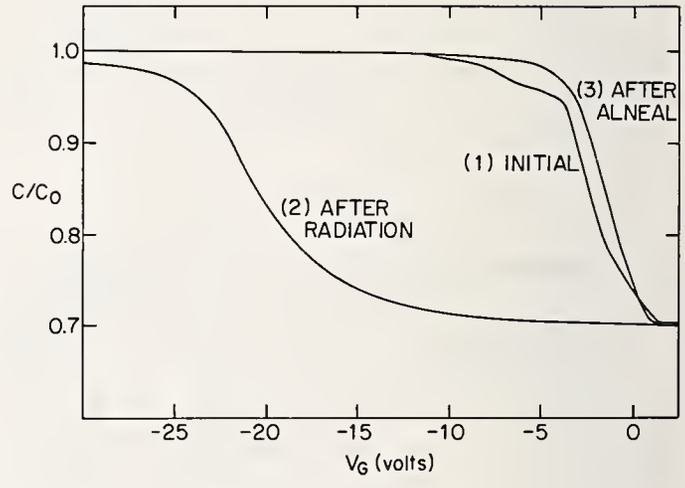


Fig. 4. Effect of radiation on charges in MOS structure. Curve (1) shows initial condition, Curve (2) shows formation of space charge N_{ot} as well as additional fast states N_{st}, Curve (3) shows that both N_{ot} and N_{st} are eliminated by "Alneal" process.

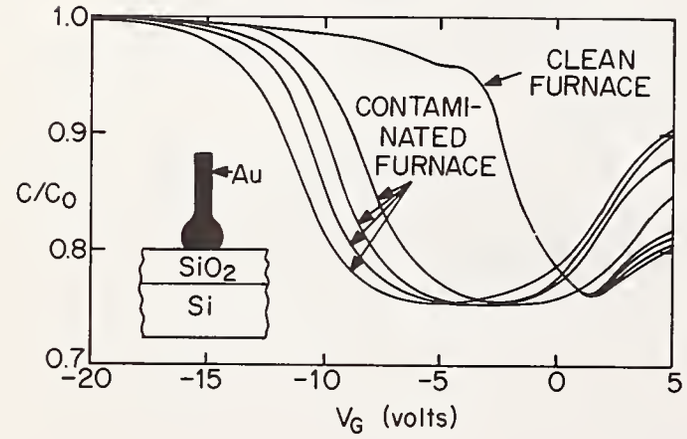


Fig. 5. Example of MOS C-V plots using the gold probe for silicon samples oxidized in clean and contaminated furnaces.

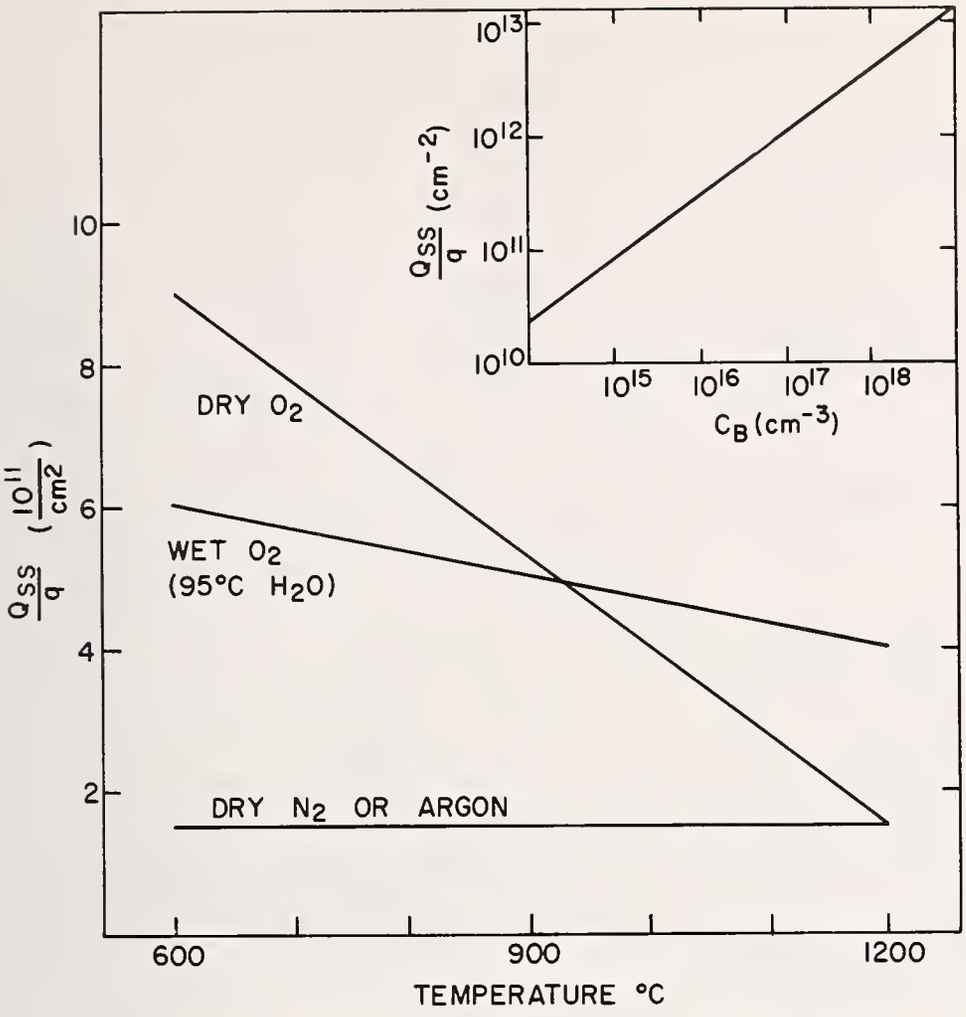


Fig. 6. Dependence of Q_{ss} for (111) silicon on final oxidation and/or annealing temperature. Insert indicates charge required to invert p-type silicon surface for varying doping concentrations.

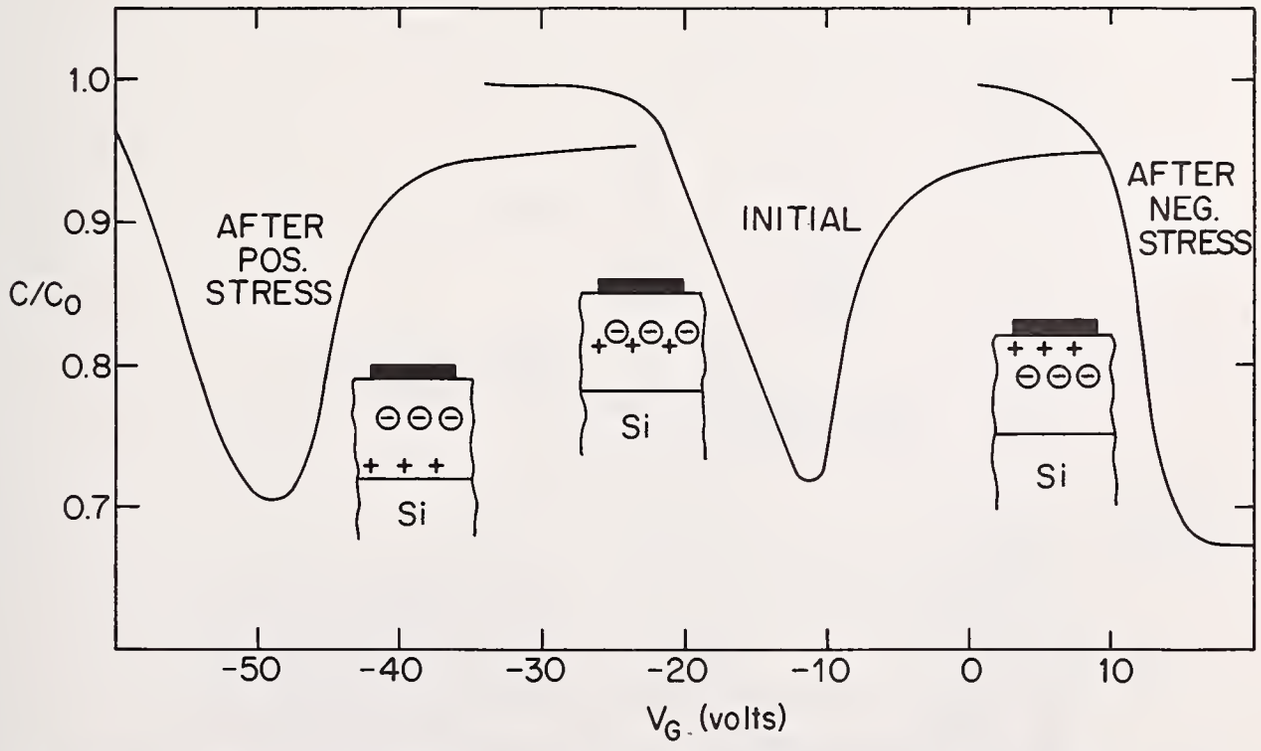


Fig. 7. Example of negative charges remaining in thermal oxide after high temperature contamination. C-V plots show that negative ions do not drift at 300°C resulting in a positive "turn-on" voltage after negative stress at 300°C.

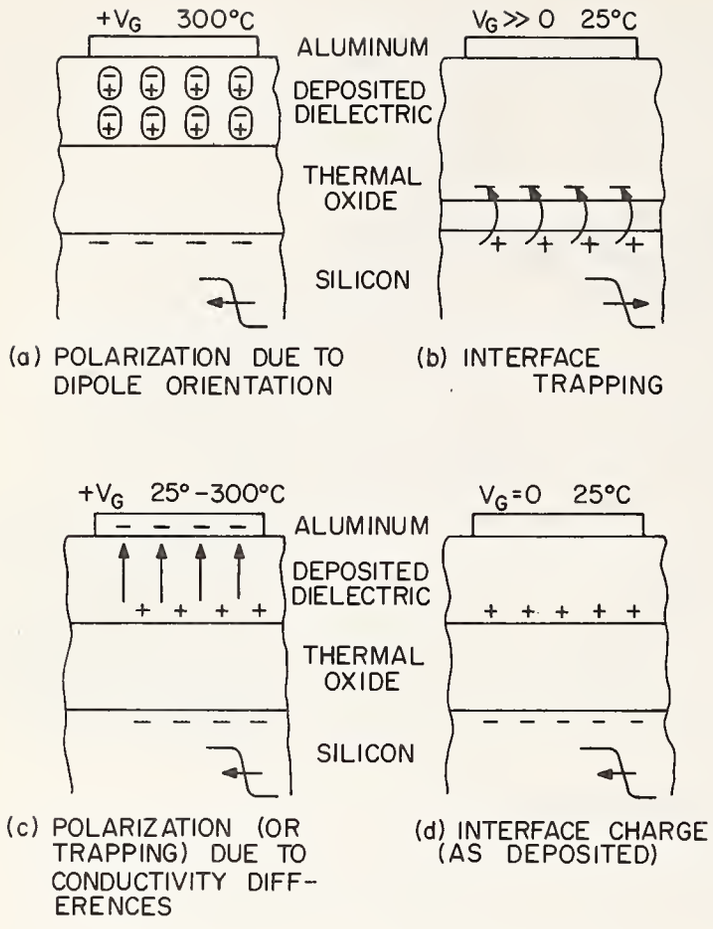


Fig. 8. Four different types of instabilities or charges associated with MIOS structures.

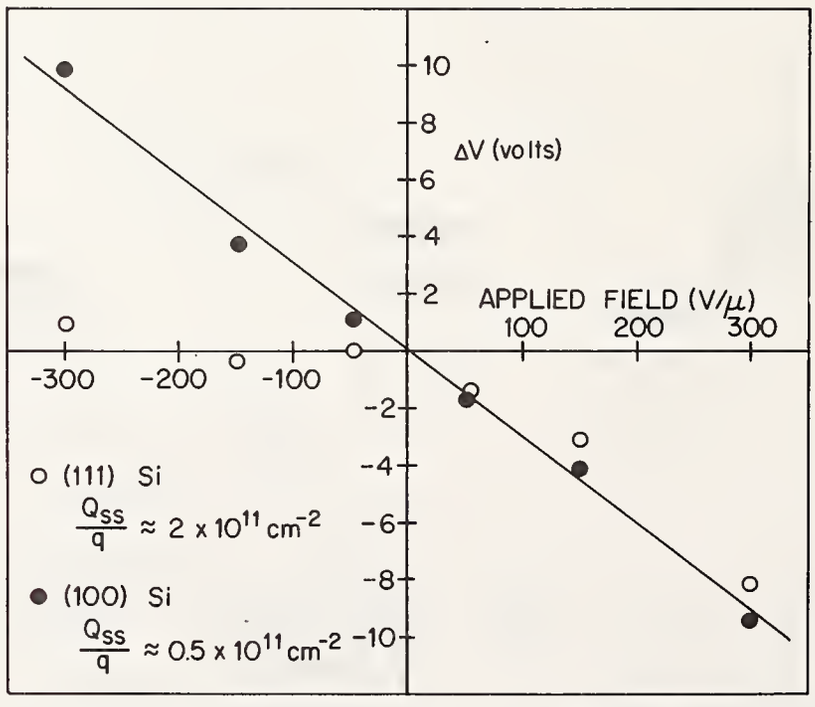


Fig. 9. Effects of two different (and opposite) instability mechanisms on C-V plots resulting from orientation (and Q_{ss}) differences in Phosphosilicate/Thermal Oxide MIOS structure. (X_G = 0.04 μ, x_0 = 0.18 μ.)

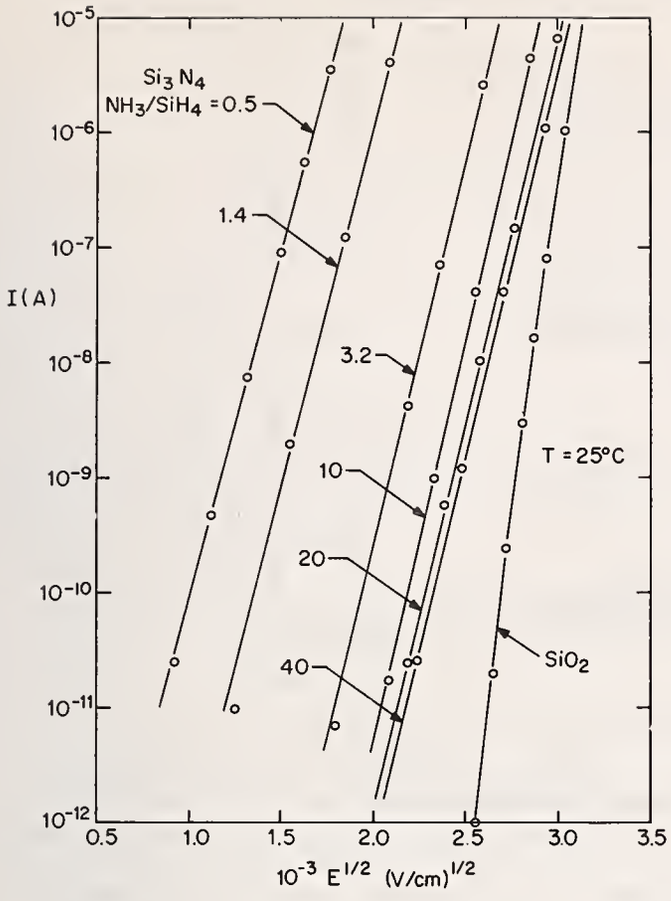


Fig. 10. Differences in I-E characteristics for MNS structures as affected by NH_3-SiH_4 reactant composition during silicon nitride deposition [49]. MOS data also included for comparison.

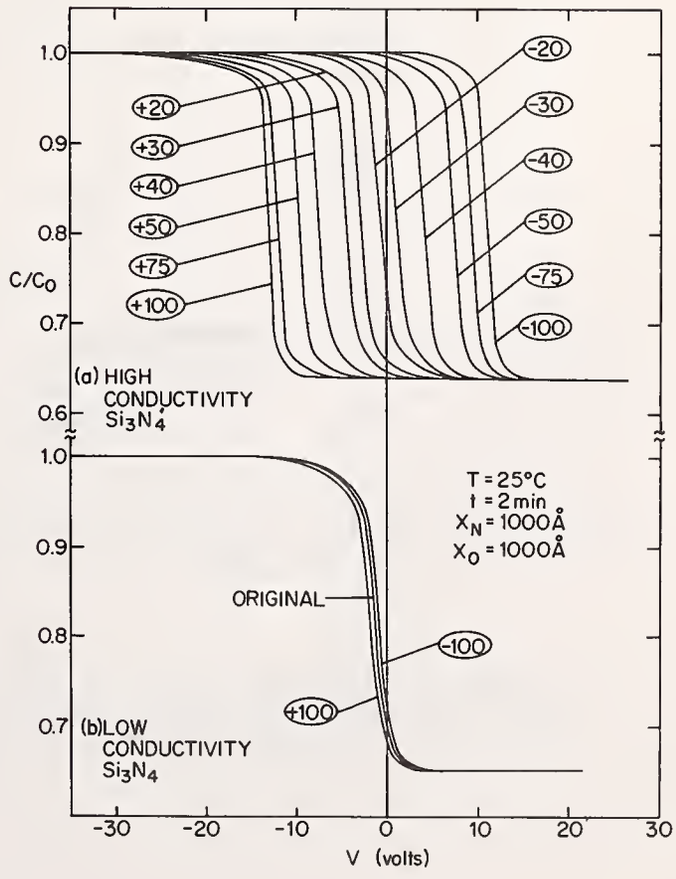


Fig. 11. Effect of silicon nitride conductance on MNOS electrical stability after stressing at 25°C with various positive and negative voltages [49].

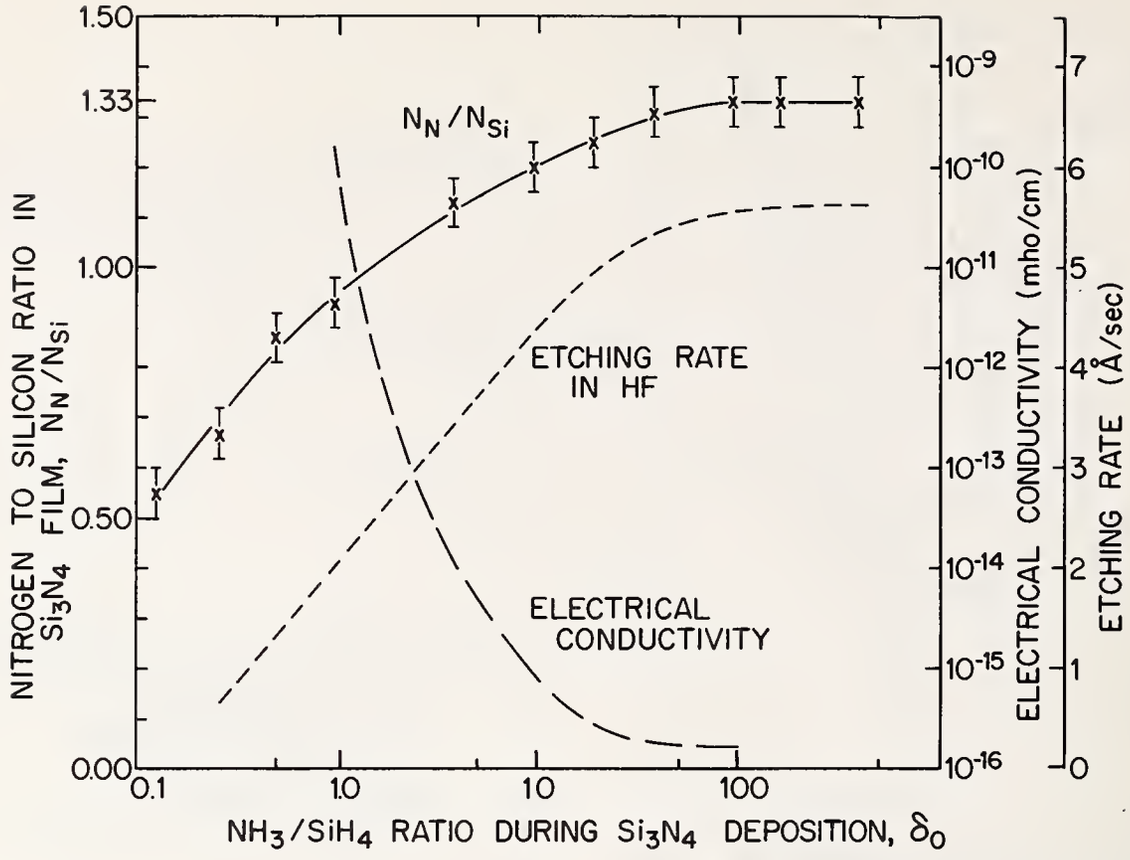


Fig. 12. Relationship among silicon nitride film composition, etch rate, and conductivity as influenced by NH_3/SiH_4 ratio during nitride deposition [50].

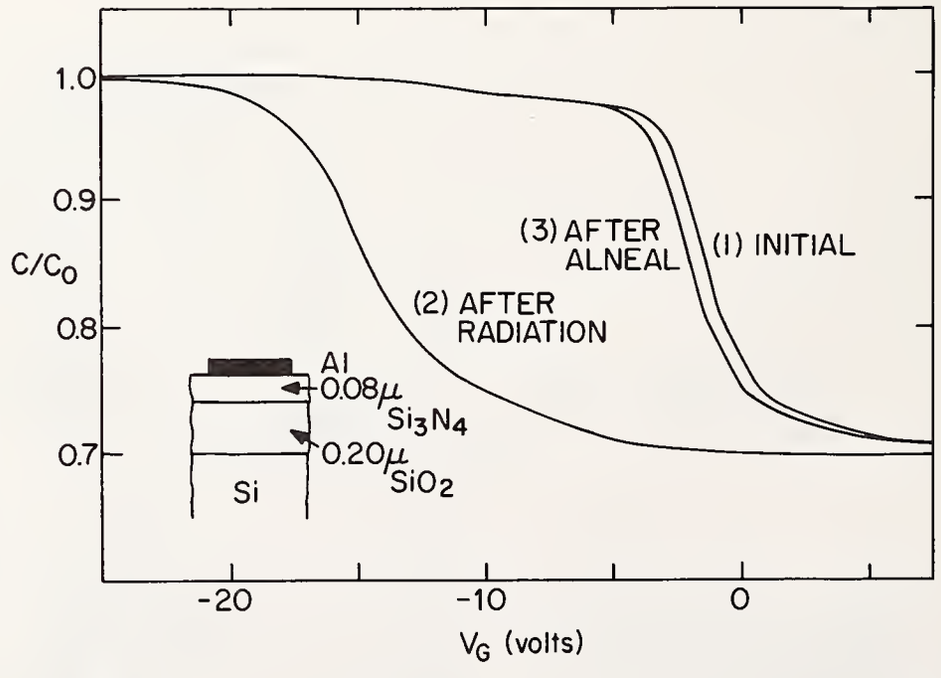


Fig. 13. Example of silicon nitride layer in MNOS structure preventing the reduction of initially present fast surface states at the Si-SiO_2 interface during the "Alneal" process but permitting the reduction of radiation-induced states and charges.

Equipment Considerations for Silicon Epitaxial Reactors

M.L. Hammond and W.P. Cox[1]

Hugle Industries, Inc.
625 N. Pastoria Avenue
Sunnyvale, Ca. 94086

The subsystems of a batch silicon epitaxy reactor--power supply, cabinetry, reactor geometry, gas flow control, and automatic control-are discussed. The effect of reaction chamber geometry on reactor performance is reviewed and fully automated epitaxial reactors using closed loop automatic gas flow control are described.

Key Words: Automatic processing, barrel reactor, epitaxy, gas flow, horizontal reactor, mass flow control, resistivity range, silicon, vertical reactor.

1. Introduction

Silicon epitaxy developed from a laboratory project to full scale manufacturing technique between 1955 and 1964. Since that time, the number of published papers on silicon epitaxy has declined (1)[2] each year while the manufacturing volume has increased enormously. The techniques of silicon eiptaxy are reasonably well established now and the major emphasis is on utilizing the techniques for new devices, increasing control and production rates, and increasing versatility.

As will be discussed, batch processing is presently favored over continuous operations. A batch silicon epitaxy reactor system can be divided into five subsystems:

1. power supply
2. cabinetry
3. reaction chamber
4. gas flow control
5. automatic control

The present and near-future status of each of these subsystems will be reviewed.

2. Power supplies

The silicon epitaxy process dictates a cold wall furnace with 1000-1300°C operating temperatures (2). The principal heat loss mechanism in a cold wall furnace is radiation and, according to the Stefan-Boltzman law, the power required is proportional to the fourth power of the absolute temperature. At 1200°C, 150-180 watts/in^2 are required to maintain temperature.

Table 1 compares the various power supplies which have been used for silicon epitaxial furnaces. The high power density and rapid response time of induction heated systems more than offsets their initial high cost. As a consequence, induction heated systems are the most widely used power supplies.

---

[1] Hammond, Division Manager, Fabrication Equipment
Cox, Manager, Applications Laboratory

[2] Figures in parentheses indicate the literature references at the end of this paper.

Power supplies are usually time-shared with two reaction chambers to provide maximum efficiency. By independently controlling the pre- and post-purges of H_2 and N_2, a single power supply can be used virtually 100% of the time.

3. Cabinetry

The reactor cabinet must meet the following conditions:

1. house the subsystems
2. provide easy access for operation and maintenance
3. provide for personnel safety
4. provide thermal dissipation
5. provide dust-free loading area
6. provide corrosion protection for critial subsystems such as the electronic packages

4. Reaction Chamber

The reaction chamber in an epitaxial reactor provides atmosphere control, defines the gas flow characteristics, and provides a means of observing the samples at temperature.

All the commonly used reactors are "open" systems in that the gases continuously enter and leave the reaction chamber during the deposition cycle. The geometry of the reaction chamber strongly influences the flow characteristics which, in turn, affect the uniformity of the epitaxial layer.

The substrates ordinarily rest on or near the hottest surface of the reactor chamber; deposition temperatures are in the 1000-1300°C range. The chamber walls are kept cool to minimize deposition on the walls and to prevent contamination from the walls. As noted in figure 1, the deposition rate on the tube walls is a minimum for 200-300°C wall temperatures. At low temperatures, polymers of various silicon compounds form and polycrystalline silicon forms at higher temperatures. Quartz and water-cooled stainless steel are commonly accepted materials for silicon epitaxial furnace walls.

Table 1. Power supplies for silicon epitaxy.

	Induction		Other	
	450KHz	10KHz	I^2R	Radiation
Transmission lines	20'	200'	-	-
General efficiency	50%	90%	90%	-
Induction penetration (graphite)	0.1 in.	0.3 in.	-	Surface
Solid state available	No	Yes	Yes	Yes
Maximum temperature	No limit	No limit	No limit	1100°C
Resonse time	Short	Short	Medium	Medium
Cost	High	Medium	Low	Med.-Low
Maintenance	Medium	Low	Low	Medium
Cooling water purity	High	Low	Low	Low

Many different reaction chamber configurations have been used but only a few basic designs are in widespread use for volume production of silicon epitaxial wafers. The three most common production reactor designs use induction heating and are shown schematically in figures 2a, b, and c. The designations-vertical pancake, horizontal slab, and barrel-are the names conventionally attached to the respective systems. However, it is more useful to categorize the various configurations according to their gas flow characteristics. Two types of gas flow chemical reactor vessels are recognized:

1. displacement or slug flow reactors in which the individual gas molecules move in an approximately streamlined fashion through the reaction chamber

2. mixed flow reactors in which the incoming gas molecules mix with the partially reacted gas before going out the exhaust.

The vertical pancake reactor in figure 2a is considered a mixed flow reactor and the horizontal reactor in figure 2b is a displacement flow reactor. The barrel reactor could be either depending upon the method of gas flow inside the reaction chamber.

Mixed flow vertical pancake reactors have historically had lower wafer capacities than displacement flow horizontal reactors and, when displacement flow horizontal reactors are operated at a reduced capacity equivalent to that of vertical pancake systems, horizontal reactors generally have better wafer-to-wafer uniformity.

Vertical pancake reactors are generally limited by radial temperature gradients and radial gas depletion effects.

Variations from wafer-to-wafer are more predictable in a displacement flow reactor because the gas flow characteristics at each point in the reactor are more reproducible. This observation can be used to advantage when working with very close tolerances.

Autodoping effects or the transfer of dopant from the substrate to the epitaxial layer in mixed flow reactors are different from those in displacement flow systems. Heavy autodoping occurs when the substrate is doped with high vapor pressure impurities such as P, As, and B. In displacement flow reactors, the effect of autodoping can be measured by noting how the resistivity changes along the gas stream. The wafer capacity of the system is then determined by the tolerance permitted for a given product. In a vertical pancake reactor the effects of autodoping are not easily noticed because of the mixed gas flow. The amount of dopant contributed by the substrate decreases with time as the epitaxial layer grows. All of the wafers in the mixed gas flow are uniformily exposed to a gas stream whose dopant level is changing with time. As a result, all the wafers have the same apparent resistivity which is only the average value of a resistivity gradient. Fortunately, autodoping effects can be minimized by using low vapor pressure Sb for low resistivity substrates.

Barrel reactors, both displacement and mixed flow, offer the largest capacity and are used by several manufacturers for high production rates. With proper design, the barrel reactors can provide high production rates and excellent deposition uniformity.

5. Gas Flow Control

Gas compositions in an epitaxial reactor are usually controlled by gas flow rates. The usual epitaxial process consists of the following general steps: purge, heat, etch, grow, cool, purge.

Silicon etchants include water, hydrogen halides (3), and sulfur hexafluoride (4). Anhydrous HCl diluted to 1-3% in H_2 is the most commonly used etchant, and the necessary degree of control is easily obtained with a ball-in-tube flowmeter.

5.1 Deposit and Dopant Gas Control

Controlling resistivity and deposition rate requires a much greater degree of gas flow control than etching. The common deposition gases are $SiCl_4$ and $SiHCl_3$ (obtained from bubbler systems) and SiH_4 (gas at room temperature). The concentration of Si-bearing gas in the reactor is controlled by varying the H_2 flow rate through a constant temperature bubbler or by controlling the flow rate of SiH_4 directly.

The resistivity depends upon the ratio of dopant to Si-bearing gas. The mole fraction of impurity at the point of injection into the main stream is in the 10^{-5}-10^{-9} range. Control at these low concentration ranges is achieved by multiple dilution schemes in which a dynamic equilibrium is established through a vent and then the gas is allowed to go to the reactor.

Figure 3 shows schematically the kind of empirical correlation observed between the impurity/silicon tetrachloride ratio and the resistivity of the resulting epitaxial layer. Data for this figure was obtained with $SiCl_4$; however, similar curves would be obtained with SiH_4 or $SiHCl_3$.

Plots of resistivity vs. impurity/silicon ratio in the gas stream follow the curves such as shown in figure 3, but the absolute values depend upon the reactor geometry, the total flow rate, and the deposition temperature. This schematic plot can illustrate the range of resistivities easily accessible with the various gas dilution schemes indicated in figure 4. If we suppose that the main H_2 flow rate is set to obtain optimum uniformity over all the wafers, and the mole fraction of $SiCl_4$ is set to give the desired deposition rate, then one can calculate the range of resistivities that can be obtained

without changing the concentration of the dopant gas or the floats or flowmeters.

The data in figure 3 were obtained on a horizontal displacement flow reactor with 50 ppm concentrations of both PH_3 and B_2H_6 diluted in H_2. A double dilution system such as shown in fugure 4b was used. Experimentally, the system proved capable of reproducibly depositing both N and P-type layers with resistivities in the 0.05 to 30 ohm-cm range.

The ratio $PH_3/SiCl_4$ in figure 3 is the ratio of the moles per minute of PH_3 to the moles per minute of $SiCl_4$ being delivered to the reaction chamber. For a particular fixed rate of flow or $SiCl_4$, this ratio can be shown to be (5)

$$\frac{PH_3}{SiCl_4} = 2.5 \times 10^{-8} \frac{V_2 V_3 V_5}{(V_1 + V_2)(V_3 + V_4)}$$

where V_1 = first dilution H_2 flow rate

V_2 = doping gas flow rate

V_3 = flow rate of diluted mixture of V_1 and V_2

V_4 = second dilution H_2 flow rate

V_5 = flow rate of diluted mixture injected into the main stream

With the particular set of flowmeters used in this experiment[1] the ratio $PH_3/SiCl_4$ could easily be varied from 3×10^{-5} to 8×10^{-10} while keeping the floats on all the flowmeters in the upper 2/3 of the flow range to assure reasonable accuracy. This doping range clearly covers the desired resistivity range in figure 3 with ample resolution.

If a single dilution doping system such as that in figure 4a is used with R215AAA Dopant and Inject flowmeters and an R215D Dilution flowmeter (glass floats), the gas ratio could be controlled over a 100:1 ratio which would correspond approximately 0.1 to 5 ohm-cm using PH_3. The resolution over this range would not be as fine as using a double dilution system which also offers a wider resistivity range.

The ball-in-tube flowmeter is being replaced in many critical areas with high response mass flowmeters (6). In a mass flowmeter, the incoming gas temperature is sensed, the gas is heated by a small fixed thermal input, and the temperature rise is related to the thermal properties of the gas and its mass flow rate. If the gas composition is fixed, then the temperature rise is a direct function of mass flow.

Mass flowmeters are 2-10 times more sensitive than ball-in-tube flowmeters and they offer an electrical signal which can be used to control a servo-valve. Automatic Mass Flow Control systems are available (6) in a single unit which incorporates the sensor, the control circuitry, and a unique flat-plate orifice valve. Complete recovery times range from 2 - 20 seconds for large deviations, and the repeatability of the set point is 0.25%.

By virtue of their increased precision, mass flow control systems offer a considerable improvement in doping gas control. For example, using the mass flowmeters in a single dilution doping system (figure 4a) with 5-1000 cc/min Dilution flow rate and 2-100 cc/min Dopant and Inject, the gas ratio can be controlled over a 3500:1 ratio with good resolution. By similarly controlling the deposition gas, improvements can be made in the control of deposition rate and, consequently, the resistivity.

Mass flowmeters can also be used to sense the incremental gas composition change through a $SiCl_4$ bubbler and, with proper calibration, can eliminate the need for maintaining $SiCl_4$ bubblers at constant temperature. Such a system is called a Differential Automatic Flow Controller (6). These closed-loop automatic mass flow control systems are being supplied on commercial epitaxial reactors (7). The obvious improvement lies in more reproducible gas flow flow rates; however, these systems will also respond to discrete or continuously variable programmed set points. Such systems offer improved control for multilayer and graded layer epitaxy.

---

[1] Brooks Rotameters, Glass Floats, Emerson Electric Co., Brooks Instr. Div., Hatfield, Pa. 19440
first dilute H_2 (R215D), Dopant (R215AAA), Mix (R215AA),
second dilute H_2 (R215D), Inject (R215AAA).

5.2 Main Gas Flow

Because silicon epitaxy requires precise control of the main gas flow with dopant concentrations in the 10^{-9} mole fraction range, the main gas is usually also vented prior to deposition so that a dynamic equilibrium can be established. Two schemes for accomplishing the vent/deposit selection on a dual chamber reactor system are shown in figure 5. The system with multiple vent lines, has the obvious advantage of providing independent vent selection for each flow line at the cost of more valves. For complex processes such as multilayer epitaxy, the independent vent capability is very important; fortunately, highly reliable solenoid valves are available.

6. Automation

Many parts of the batch silicon epitaxial reactor are amenable to automation and complete program control. The parameters of the silicon epitaxial process which require control are:

1. wafer temperature
2. gas composition and flow rate
3. process times

Figure 6 illustrates how these parameters can be fully program controlled.

Automatic temperature control using remote sensors are commonly used. Commercial epitaxial reactors are available with closed loop temperature control, digital temperature readout calibrated in °C, and multiple program-selectable set points (7).

Automatic gas flow control systems are available as discussed above, and both fixed and program-variable set points have been demonstrated (7). Process time control is the straight-forward manipulation of valves.

There is sufficient process knowledge to write the necessary computer program and rapid evaluation techniques have already been developed using infrared thickness and four-point probe electrical measurements.

Systems as in figure 6 are possible today and are under evaluation in several semiconductor facilities. When combined with automatic wafer handling, these automatic batch process systems offer reduced chances for human error and considerable reduction in labor cost.

The next levels of sophistication in epitaxial process are:

1. actual gas composition control
2. in-situ evaluation
3. continuous processing

The technology for gas analysis is quite advanced and, with sufficient time and money, gas control could be based on actual composition measurements and closed loop flow controllers. At the present, the gain in performance may not be worth the additional cost in instrumention.

In-situ thickness measurement using infrared detectors has already been demonstrated in heteroepitaxy (8) and a sapphire or spinel test wafer might be used to monitor growth rates in a homoepitaxial reactor. Unfortunately, the characterization of doping levels at deposition temperatures presents some apparently insurmountable problems, due to the thermal excitation of carriers.

Continuous processing is a straight-forward extension of the automated systems shown in figure 6. Mechanical gating or gas curtains would be required to maintain the gas composition control in the reactor and methods of moving the wafers through the deposition zone would have to be devised. At this point, it is not obvious that a continuous system would have more productivity than a comparably priced system of batch reactors.

In summary, the performance of the popular silicon epitaxial reactors can be characterized by the nature of the main gas flow, with displacement flow system showing certain process advantages over mixed flow systems. Dilution and vent/deposit schemes are required for gas composition control at low concentrations and recent innovations in closed loop gas flow control have greatly improved the precision and versatility of gas control systems. The techniques necessary to provide fully automatic batch processing of silicon epitaxial wafers are presently available and complete computer controlled systems are within the state-of-the-art.

7. References

(1) Runyan, W., Review, The Status of Silicon Epitaxy, p. 169, Semiconductor Silicon, (The Electrochemical Society, New York, 1969).

(2) Shephard, W.H., Vapor Phase Deposition and Etching of Silicon, J. Electrochemical Society., 112, 988 (1965).

(3)a Integrated Silicon Device Technology, Vol. IV-Diffusion, Research Triangle Institute, (Clearing House for Federal Scientific and Technical Information, Feb. 1964, Report No. AD 603176).

b Ibid, Vol. IX, Epitaxy, (Report No. AD 624520).

(4) Matheson Gas Products, P.O. Box 85, East Rutherford, N.J. 07073.

(5) Cox, W.P., Resistivity Calculations for Hugle Industries' Horizontal Silicon Epitaxial Reactors, Hugle Industries Technical Note 7001, Hugle Industries, Inc., 625 North Pastoria Avenue, Sunnyvale, Ca. 94086.

(6) Tylan Corporation, 4203 Spencer Street, Torrance, Ca.

(7) Hugle Industries, Fabrication Equipment Division, 625 North Pastoria Avenue, Sunnyvale, Ca. 94086.

(8) Cullen, G.W., et al, Epitaxial Growth and Electrical Properties of Silicon Films on Spirel, p. 291, Semiconductor Silicon (The Electrochemical Society, New York, 1969).

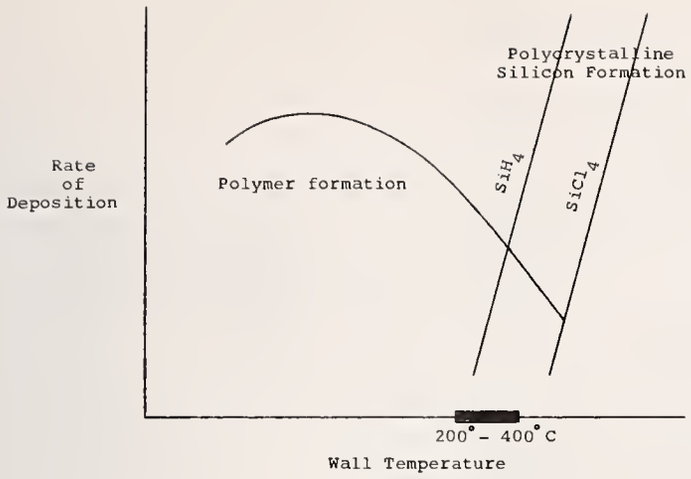
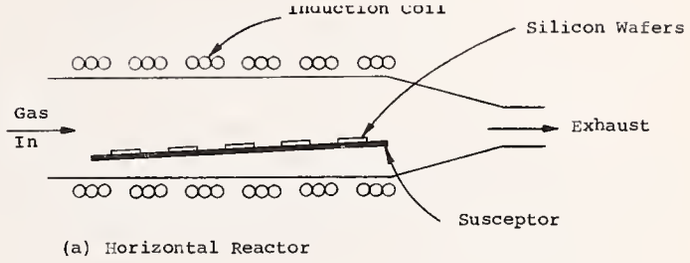
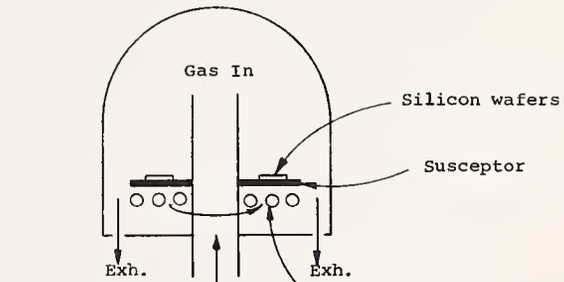


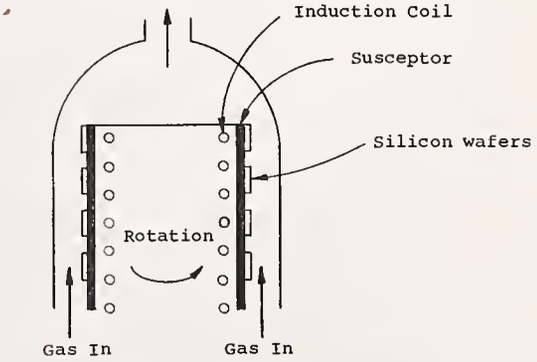
Fig. 1. Effect of wall temperature on the rate of deposition on reaction chamber walls.



(a) Horizontal Reactor



(b) Vertical (Pancake) Reactor



(c) Cylinder Displacement Flow Reactor

Fig. 2. Common Reactor Geometries:
- a) Horizontal Displacement Flow Reactor
- b) Vertical Pancake Mixed Flow Reactor
- c) Barrel Reactor with Displacement Gas Flow Pattern

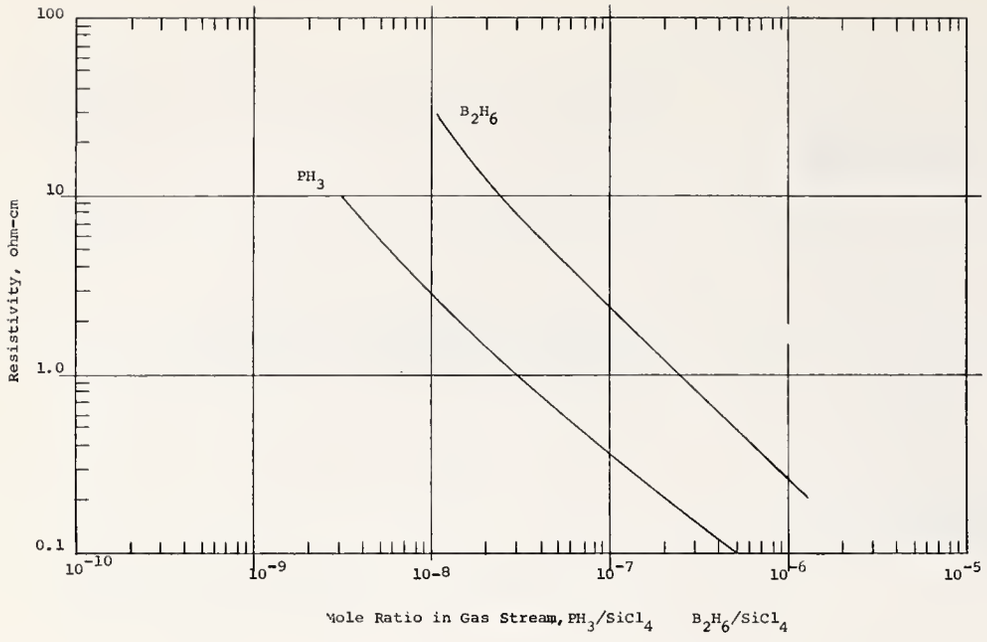
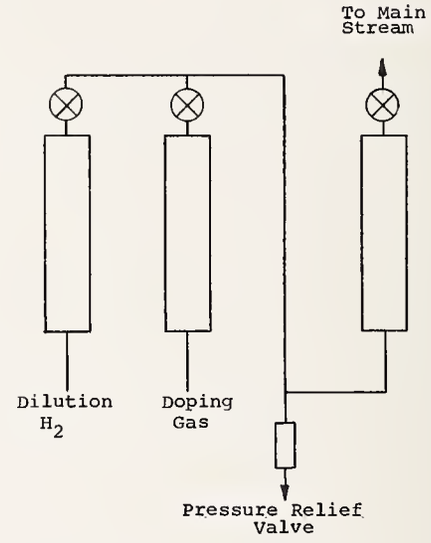
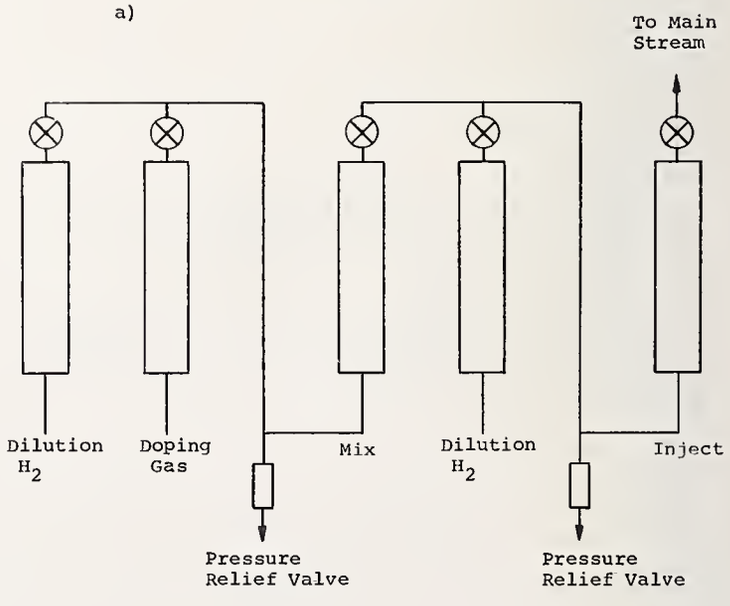


Fig. 3. Film Resistivity vs Dopant gas to $SiCl_4$ ratio for Hugle Industries Horizontal Displacement Flow Reactor



a)



b

Fig. 4. Dopant Gas Dilution System:

a) Single Dilution
b) Double Dilution

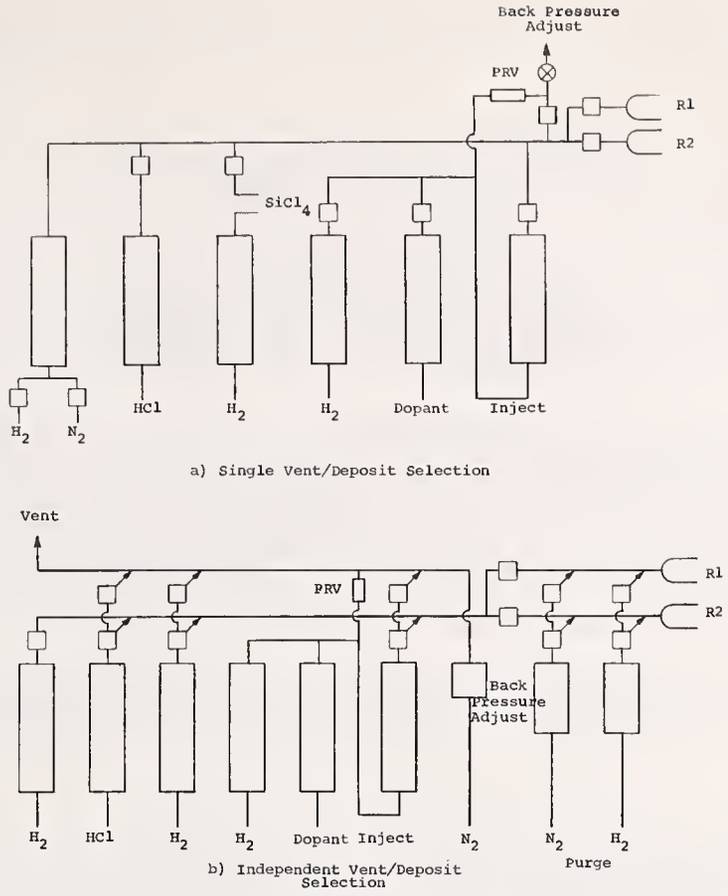


Fig. 5. Techniques for Main Gas Vent/Deposit Selection

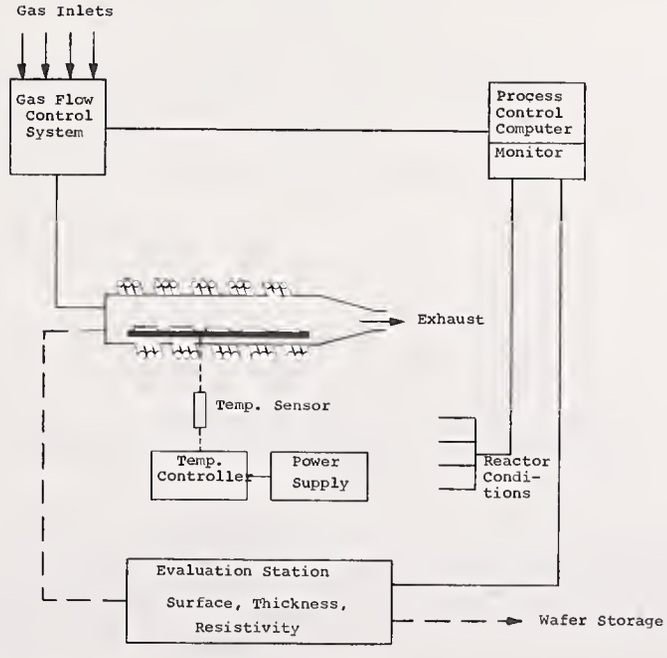


Fig. 6. Automatic Program Controlled Batch Epitaxial Reactor System.

A Comparison of a Resistance-Heated Reactor for Silicon Epitaxial Growth with Other Epitaxial Systems

Willem A. Kohler
Research and Development Laboratory
Fairchild Camera and Instrument Corporation
Palo Alto, California 94304

A special reactor has been developed which makes possible the growth of epitaxial silicon films in a resistance-heated system. The reactor is specially designed to minimize the deposition of silicon on the tube walls, thus eliminating the major disadvantage of resistance-heated systems. The construction of this simple, inexpensive reactor is described, and its advantages and disadvantages compared to other approaches.

Key words: Resistance-heated epitaxial reactor, silicon epitaxial reactors.

1. Silicon Epitaxial Technology

For the deposition of epitaxial silicon on single crystalline silicon substrates, several approaches can be used, depending on the requirements. In this paper, these methods will be compared with special emphasis on the properties necessary for a good production process, that is, good reproducibility, process control, and possibilities for automation.

Silicon may be deposited by vacuum deposition or by chemical vapor deposition. The most elementary method is vacuum evaporation or sublimation from a heated silicon source onto a single crystalline substrate at a vacuum of better than 10^{-5} Torr. For high quality films, special substrate cleaning by ion bombardment or high-temperature heat treatment is necessary before growth. A great advantage of vacuum deposition is the flexibility of the process, e.g., single crystalline silicon growth on fused quartz was first obtained by high vacuum evaporation [1],* and recently epitaxial films on silicon were obtained by high vacuum sublimation at 650°C substrate temperature [2]. This method is unsuitable for production at present, however, because of the difficulty in controlling doping.

While vacuum deposition has some special advantages, chemical vapor deposition processes are generally used for reproducible production processes. Such processes have several advantages:

(1) Deposition rate and doping levels can be made reproducible and uniform by controlling the gas flows and temperature setting.

(2) If desired, a chemical etch of the substrate in the reactor can be performed prior to silicon deposition.

For heating, an RF generator is generally used since it heats the susceptor without directly heating the walls of the quartz reactor tube. With additional (e.g., forced air) cooling of the tube, the deposit on the walls can be negligible. This heating system is commonly used in production processes because it simplifies reactor design. The most widely-used approach is a horizontal system consisting of a rectangular flat reactor tube with matching flat susceptor inside and an RF coil outside the system for heating.

A variety of vertical systems are also available commercially. Some vertical systems have a rotating disc as the susceptor with a pancake RF coil inside the system, others have a cylindrical rotating susceptor. Rotation of the susceptor gives better uniformity of the films but introduces, in the case of an inside RF coil, a possible impurity source. In addition, vertical systems are generally more complicated in construction and for this reason can give a less reliable system and more maintenance problems.

---

*Figures in brackets indicate the literature references at the end of this paper.

A description of epitaxial processes and equipment is given by Doo et al. [3] and Gupta et al. [4]. In these articles, as in most survey publications on epitaxial growth, the emphasis is on RF-heated chemical vapor deposition systems.

RF-heated reactors in general, however, have several shortcomings:

(1) The whole system is expensive in installation cost and power consumption.

(2) Temperature control is generally done by optical pyrometry. This method, however, is irreproducible, mainly because of transmission losses during growth due to unavoidable thin deposits on the quartz tube wall and structure changes in the quartz.

(3) Undesired temperature gradients can occur if the position of the susceptor with respect to the coils is not completely reproducible or if the wafer position on the susceptor is changed.

In contrast, resistance-heated systems are much less expensive and offer a more controllable temperature profile since the heating elements are kept under constant operation, and temperature control by thermocouple is very reproducible. In some instances [5], hot wall resistance furnaces have been adapted for epitaxial growth, but two problems have been encountered:

(1) Silicon deposition on the reactor tube.

(2) Impurity outgassing from the heating elements and from the quartz wall through the hot wall.

These disadvantages can be overcome in a specially designed resistance-heated system that is still a magnitude lower in cost than an RF-heated system of the same capacity.

2. Resistance-Heated Epitaxial Reactor

Contrary to the hot wall tube as generally used in diffusion processes, the resistance-heated system as described in this paper* has heating elements only in the lower half of the clamshell-type furnace, Fig. 1. Five semicircular heating elements consisting of Kanthal heating wire imbedded in alumina cement and high-temperature ceramic are used. Element size must be sufficient to provide the temperature and cycle time requirements. All five elements can be separately controlled to obtain the desired temperature profile. The absence of heating elements in the top half of the furnace results in the top and side of the reactor tube remaining relatively cold. This, together with some additional gas cooling on top of the tube, as shown in Fig. 2, makes the top and side walls of the reactor tube cold enough to minimize the deposit on it. For this reason, many runs can be done before there is enough deposit to cause particles to fall on the wafers and consequently create defects in the epitaxial films. Deposits can be completely eliminated when using a reactor etch either to etch the wafer in the reactor prior to deposition or after a number of runs to avoid a heavy buildup of silicon deposit. Extremely suitable for this purpose is SF_6 gas which etches very fast at low temperatures.†

Figure 4 is a diagram of the gas flow system. 100% silane gas is mixed with H_2 diluted dopant gas, either AsH_3 or B_2H_6 and H_2. These gases are introduced into a separate inlet on top of the reactor tube (Figs. 2 and 3). The main H_2 carrier gas is introduced along the horizontal axis of the reactor, is preheated in the cylindrical part of the reactor tube, and only mixes with the dopant and silane gases near the point in the tube where the temperature profile has reached the required deposition temperature. To avoid decomposition of the silane in the inlet, this tube is cooled by a mantle consisting of a concentric tube around the inlet tube through which cooling gas (N_2) is blown. The reactor tube has a rectangular cross section and only the flat bottom is directly heated (Fig. 3). To avoid deposition on the bottom and to minimize outgassing and diffusion of impurities through the hot wall, the wafer boat covers the bottom completely. The boat consists of graphite coated with silicon or silicon carbide. Pyrolytic graphite also can be used. The graphite boat, on which the wafers are placed, is placed on a quartz boat to obtain a slight angle to improve uniformity along the tube.

High-quality epitaxial films have been reproducibly obtained over a 20-inch long zone at 1030°C growth temperature using silane with H_2 as the carrier gas. Feasibility has been shown for the deposition of good-quality epitaxial films using the hydrogen reduction of $SiCl_4$. However, with the reactor as described in its present dimensions, the heat-up time for a 1150°C temperature range as required for $SiCl_4$ reduction increases considerably, resulting in more out-diffusion and a longer cycle time as compared to an RF-heated system. The out-diffusion and mass transport of dopants (autodoping) from the substrate cause the silicon deposition generally to shift toward the lower temperature deposition ranges [4,6]. In this respect it makes the SiH_4-H_2 system as presently used a more desirable system. An additional benefit of SiH_4 is the absence of a hydrogen chloride reaction which reduces mass transport effects.

---

*A patent is pending on a part of this work.

†To etch off any silicon deposits after a number of runs, the cooling on top of the tube is switched off so that etching also occurs at normally cooler parts of the reactor tube.

A heating time of 7 minutes causes out-diffusion and autodoping of a negligible amount compared to the impurity redistribution determined by growth and purging time at actual growth temperature.

Thickness and resistivity uniformity of films grown with silane have been controlled and reproduced within 7% precision over eight 2-inch wafers along the boat. This was obtained after optimizing the flow ratio, tilting the boat, and using a baffle in front of the reactor tube to guarantee good mixing inside the rectangular portion of the reactor tube. Both n-doped (arsine) and p-doped (diborane) films have been grown up to 20 Ω-cm resistivity. Film quality has been excellent.

For high growth temperatures such as in the epitaxial growth by hydrogen reduction of $SiCl_4$ and for high growth rates such as for thick polycrystalline growth in dielectric isolation processes, RF generators provide the best heating systems. However, for the silane decomposition process a resistance-heated reactor has proved to work satisfactorily. The silane process is preferred in many cases, not only because of lower growth temperatures but also because of the absence of pattern shift using <111> material [7]. Chemical reactor etching can be done at low temperatures using SF_6 gas or high purity HCl. For some special devices deposition temperatures lower than 1030°C are required. Epitaxial silicon deposition in the 900°C range can be obtained using silane with helium as a carrier gas, as reported by Richman and Arlett [8]. A heat-up time of less than 5 minutes is sufficient to obtain this temperature with the reactor described.

Other processes such as low-temperature polycrystalline silicon growth and silicon nitride growth [9] can be done in an inexpensive and very reproducible way in this kind of reactor. Especially for processes in these low temperature ranges, thermocouples as used in this reactor are the most accurate temperature monitors.

The future trend for semiconductor processes is directed toward automation. Epitaxial growth is a complex process to automate because of its many variables. An automated system using a resistance-heated furnace and continuously moving susceptors offers many possibilities: The temperature profile is easily kept constant, eliminating one important variable. With the help of effective gas curtains, the reaction zone can be separated from the inlet zone and the cooled outlet zone so that a continuous gas flow of the silicon compound and doping gases can be maintained in the reaction zone. A separate zone for vapor etching could be introduced preceding the reaction zone. The thickness could be controlled by the rate at which the susceptor is moved through the reactor. Such a furnace could produce extremely uniform and reproducible films as each wafer would experience the same environment. Such an approach appears very desirable for very high volume production of silicon epitaxial wafers.

3. Summary

Although RF-heated reactors are presently the most common type of epitaxial reactors, resistance-heated reactors as described offer an inexpensive and simple alternative for epitaxial silicon growth. Their special advantages lie in the temperature range up to 1100°C. Here, the low cost and excellent temperature control of the resistance-heated reactor make this system most attractive. At higher temperatures, the fast heat-up time of the RF reactor makes it preferable. The resistance-heated reactor also shows promise for use in a fully automated, continuous epitaxial growth process.

4. Conclusion

It has been found that high-quality epitaxial silicon films can be grown in a specially designed resistance-heated reactor at temperatures up to 1100°C. The installation and maintenance costs of such a reactor are an order of magnitude lower than an RF-heated reactor of the same wafer capacity while the epitaxial films are of comparable quality.

5. Acknowledgment

The author wishes to acknowledge the helpful discussions with C. Plough, the very valuable contribution in the development and construction of the reactor by J. Flood, and the assistance of C. Fullmer in the deposition and evaluation of numerous epitaxial films.

6. References

[1] Haidinger, W., Courvoisier, J. C., and Kohler, W. A., *Le Vide, Special Semiconductor Issue*, pp. 59-65 (Oct. 1966).

[2] Thomas, R. N. and Francombe, M. H., *Solid-State Electron.* 12, 799 (1969).

[3] Doo, V. Y. and Ernst, E. O., *SCP and Solid-State Technol.* 10, 31 (1967).

[4] Gupta, D. C. and Wang, P., *Solid State Technol.* 11, 48 (1968).

[5] Deal, B. E., *J. Electrochem. Soc.* 109, 514 (1962).

[6] Frieser, R. G., *J. Electrochem. Soc.* 115, 401 (1968).

[7] Enomoto, T., Yukawa, T., Iwata, Y., and Ohkubo, T., Paper presented at the Boston Meeting of The Electrochemical Society, May 5-9, 1968, Abstract No. 78.

[8] Richman, R. and Arlett, R. M., in *Semiconductor Silicon*, R. R. Haberecht and E. L. Kern, Editors (The Electrochemical Society, 1969).

[9] Kohler, W. A., *Metal. Trans.* 1, 735 (1970).

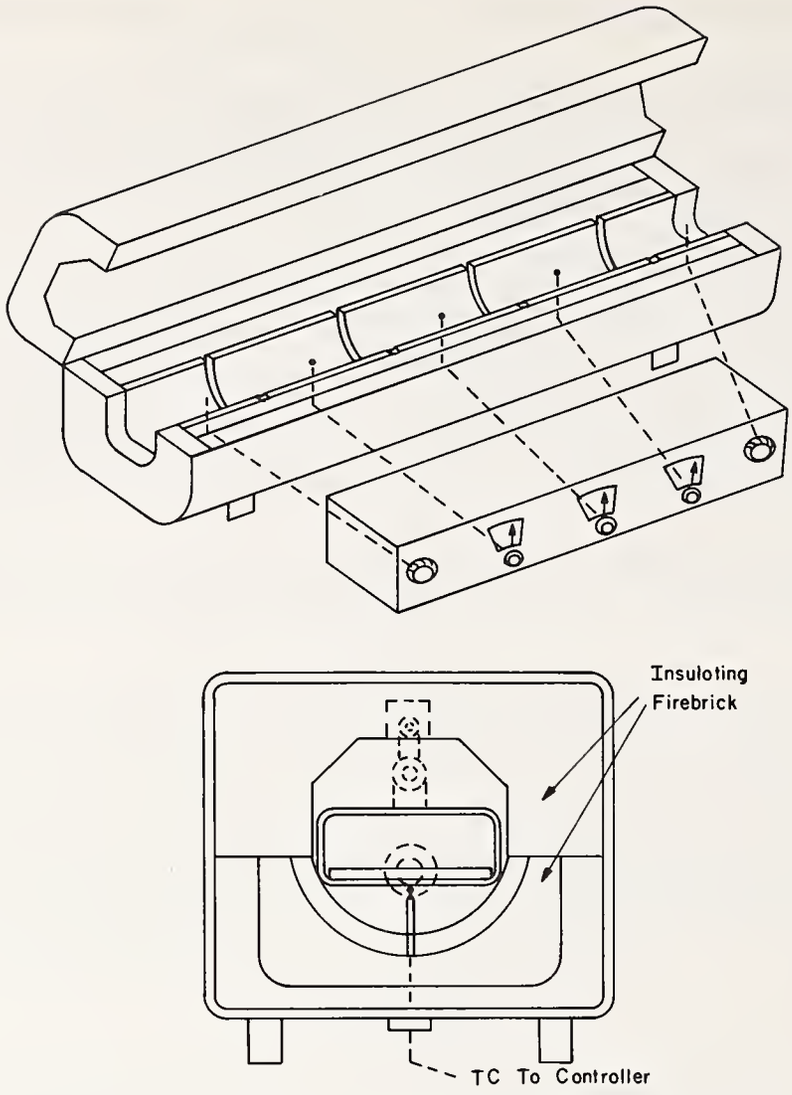


Fig. 1. Reactor furnace and controller.

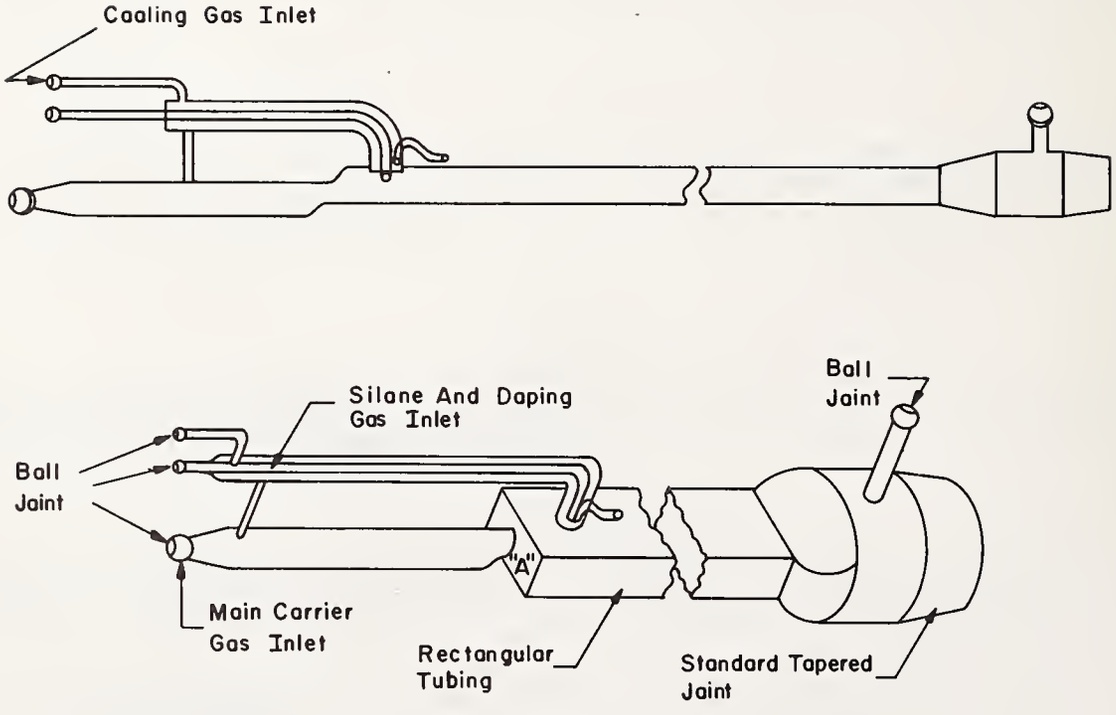


Fig. 2. Reactor tube.

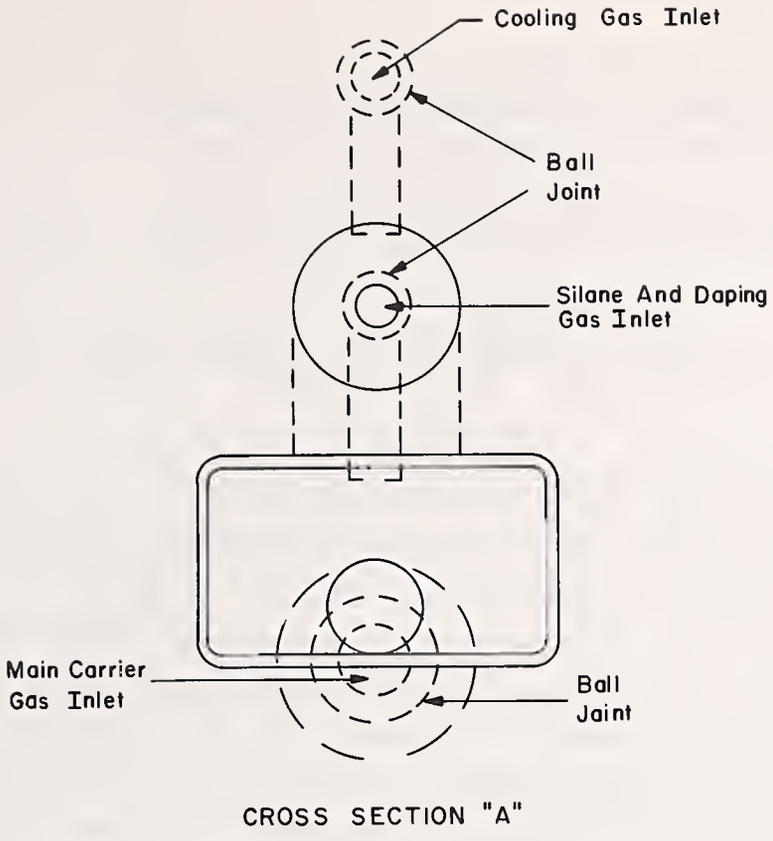


Fig. 3. Cross section "A" of the reactor tube in Fig. 2.

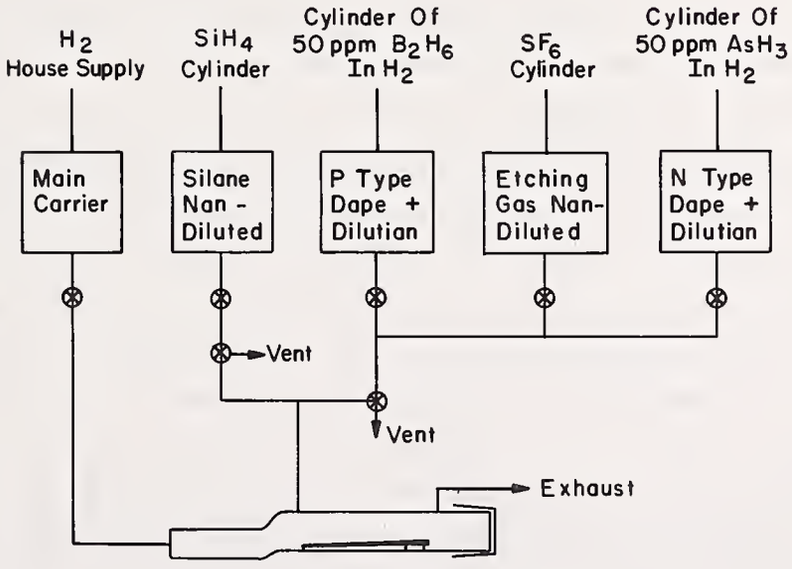


Fig. 4. Schematic of the gas flow system.

TECHNIQUES FOR DEPOSITING HIGHLY UNIFORM AND DEFECT-FREE EPITAXIAL SILICON

Dinesh C. Gupta
General Telephone and Electronics Laboratories, Inc., Woburn, Mass.

and

John L. Porter
Sylvania Electric Products, Inc., Woburn, Mass.

Introduction

In the fast developing field of semiconductor devices and integrated circuits using silicon epitaxial layers, it is important to produce layers with uniform thickness and resistivity control and high surface quality. A number of improvements in the deposition techniques have been advanced to achieve this uniformity and surface quality. Most of these involve reactor tube configuration and optimizing the mechanical design of the reactor. Basically, three reactor designs are used for silicon vapor depositions, namely, vertical, barrel and horizontal types. The vertical reactor (1) uses a pedestal-type susceptor, rotating or non-rotating, which is heated from outside the chamber. Reactant gases may be introduced in the chamber from top or bottom of the chamber. In either case, the gases undergo a turbulent motion while approaching the susceptor and strike the substrate in different vertical planes. The barrel reactor (2) allows a large number of wafers deposited in the same run. The wafers are placed on a rotating cylindrical susceptor heated from outside the chamber. In horizontal configuration, the gas is directed laterally over the wafers which are placed on the susceptor in the plane parallel to the gas flow. The susceptor is heated again from outside the chamber.

The horizontal system has been most widely used for silicon epitaxial deposition. The reasons for this may be because it is relatively simple and inexpensive to put together and the kinetics of the system is rather easy to understand. However, there are many obvious disadvantages in this configuration such as the one posed from placing the wafers in the plane of the gas flow. This contributes to the depletion of the gas and to the dopant escaping from the substrate into the flow and redepositing on the down-stream wafers.

A great many improvements have been made in recent years on the geometry of the horizontal reactors. These have included optimizing the tilt in the susceptor (3) and constructing a planar boundary layer controlled system (4). These help in achieving uniform depositions but have been of limited success. A few authors (5,6) have suggested achieving better uniformity using high temperatures and low deposition rates. The interface impurity gradient is degraded using these techniques however (7), and therefore, not preferred. This paper will discuss alternate techniques to achieve uniform layers with better surface quality.

Low Temperature Deposition

Considerable effort has been made to achieve depositions at lower temperatures and by a process where mass transport is reduced or may be completely eliminated. One of these processes is the pyrolysis of silane (8-10). In the last two years, the pyrolysis of silane has been improved to the extent that its use in production is now possible (7,11). Using a 5 percent SiH_4 mixture in hydrogen, low deposition rates are obtained (Fig. 1) and deposition times become unsatisfactory for a thick layer. By increasing the concentration of silane however, the deposition rate may be obtained comparable to the high temperature processes (Fig. 2). Depositing a good quality layer at 2 to 4 μm/minute using silane at 950 or 1000°C is not unusual. Also the dependence of the deposition rate on temperature is found to be well behaved from 850°C to 1100°C (Fig. 3 and 4) and layers of good quality may be obtained as low as 900°C by ordinary techniques and about 800°C using the special techniques such as a high-low temperature cycle (7) or SiH_4-He mixtures (12).

Effects of Gas Velocity on Deposition

Shepherd (13) and Bradshaw (14,15) developed a theory for depositions in a horizontal system at low deposition efficiencies. Both theories show that for silicon tetrachloride or trichlorosilane deposition rates are determined by the transport properties of the gas stream. After the reactor geometries are optimized, use may be made of increasing the gas velocities to achieve uniform layers. Fig. 5 shows the change in deposition rate along the length of the susceptor at different gas velocities for the decomposition of $SiCl_4$ in a round tube configuration. As the gas velocity increases, the change in deposition rate, in general, decreases. Also at high velocities, a constant deposition rate is obtained in a portion along the center of the susceptor. For silane depositions, the curves look almost similar except that they are shifted downward by about half-an-order magnitude (Fig. 6) and there is no region of constant deposition rate even at high velocities. Figure 7 shows the longitudinal taper along the length of the susceptor for both the $SiCl_4$ and SiH_4 depositions at different gas velocities. At gas velocity of 10 cm/sec , a taper of $\pm$ 7 percent for $SiCl_4$ and $\pm$ 5 percent for SiH_4 is obtained in the round tube configuration over a length of about 20 cm. These taper figures reduce to about $\pm$ 2.5 percent and $<$1 percent respectively at a gas velocity of 45 cm/sec . Also shown in this figure is the effect of gas velocity on depositions on substrates of different orientations. Both in $SiCl_4$ and SiH_4 depositions, the taper is slightly higher for (111)- and (110)-oriented substrates at low gas velocities. As the gas velocity increases, the taper becomes the same for all orientations.

The use of a planar boundary layer configuration is definitely an advantage, e.g., using a rectangular tube instead of the round tube, the longitudinal taper is reduced to about 1 percent for depositions using $SiCl_4$ and about $<$0.5 percent for depositions

using SiH_4 (Fig. 8). Again the same differences are found for the differently oriented substrates at low gas velocities. Fig. 9 shows the change in deposition rate along the length of a susceptor for depositions using $SiCl_4$ in rectangular tube configuration. The change in the rate is reduced by about half-an-order of magnitude for similar depositions in round tube configuration. If the silane depositions are made in the rectangular tube configuration (Fig. 10), the change in rate is only .002 μm/minute over a length of 20 cm at the gas velocity of 43 cm/sec .

Lateral taper is also improved as the gas velocity is increased, both for $SiCl_4$ and SiH_4 depositions. At a gas velocity of 43 cm/sec , a lateral taper of only about $\pm$ 0.5 percent is found over 6 cm across the susceptor for SiH_4 depositions (Fig. 11). Lateral variations in resistivity and thickness are also due to the temperature variations on the susceptor. The variations in temperature may be as much as 15 to 20°C across the susceptor. The lateral taper may further be improved by decreasing these temperature variations. Sutton (16) suggested the use of baffles to adjust the lateral temperature gradients.

Effect of Gas Velocity on the Impurity Distribution in the Layer

High gas flows tend to move the impurities originating from autodoping, reaction chamber or the hot susceptor faster through the tube which helps in obtaining sharper gradients in the layer. Various techniques (7,17) to improve the epitaxial layer impurity profiles have been suggested such as reducing the temperature of deposition, carrying out high-low temperature cycle deposition, and sealing the susceptor surface and the backside of the substrates. Further improvement in the profile may be obtained by increasing the gas velocity. Shown in Fig. 12 are two profiles, one for the layer deposited with a gas velocity of 11.2 cm/sec and the other with a gas velocity of 43 cm/sec . These depositions were made at 1000°C using silane as the source. The surface of the susceptor was sealed with a layer of high purity silicon but the substrate back-surface was not sealed. Also interesting to note are the measurements of thickness using infrared reflectance (IR) and bevel-stain (B&S) techniques. One may detect very small differences in flatness of the profiles using these measurements (18). The differences between IR and B&S measurements was 0.25 μm for one layer and 0.05 μm for the other layer. The latter of these two layers is more abrupt.

A Novel Technique To Improve Layer Uniformity

Recently Ipri (19) and Bloem (20) suggested the deposition of the layers using a mixture of silane and silicon tetrachloride or SiH_4. They reported better uniformities in these layers. The introduction of an etchant with the source helps in obtaining better uniformity in layer thickness. The required concentration of HCl in H_2 is very small, e.g., of the order of 0.001 to 1 percent. Fig. 13 shows the percent longitudinal taper over a length of 20 cm as a function of HCl concentration in H_2 for silane depositions made in a rectangular tube configuration. It is possible to reduce the

taper by about ± 0.15 percent (i.e., from ± 0.72 percent to ± 0.57 percent at a gas velocity of 25 cm/sec) but only at some expense. Even the slight introduction of HCl in the tube degrades the quality of the surface of the layer (Fig. 14). Further degradation is noted with increase in concentration of HCl. Also the impurity profile in the layer degrades with an increase of HCl concentration during deposition (Fig. 15). The quality of the surface improves if the temperature is raised to 1200°C, however then, the impurity profile degrades further.

We suggest the introduction of sulphur hexafluoride which is also an etchant for silicon and works at low temperatures. Use of SF_6 gives good surfaces, if used in very low concentrations at 900-1000°C. The longitudinal taper is comparable to that obtained with the introduction of HCl and the impurity profile of the layers is not as much degraded.

Conclusions

After the reactor geometries are optimized, the only way to achieve more uniform layers is by changing gas flow conditions. Increasing the gas velocity reduces both the longitudinal and the lateral taper. Changing the round tube configuration to the rectangular tube further reduces the taper. The quality of the surface of the layers deposited with high gas velocity is found to be slightly better than the layers deposited with low gas velocity (Fig. 16). No degradation in the surface due to increases in gas velocity has been noted. With all other conditions the same, the abruptness of the impurity profile in the layer increases with an increase in gas velocity. An improvement in uniformity for silane depositions is possible if a low concentration of HCl is introduced with SiH_4 during deposition. However, this change is small and the introduction of HCl degrades both the quality of the surface and the interface impurity gradient. The results are better when SF_6 is used, however, more work needs to be done to improve these processes.

Acknowledgments

The authors wish to acknowledge the assistance of P. Keating, J. Weltchek, R. Lederman and Mrs. E. Kelley in experimental work and to J. Chan for measurements.

REFERENCES

1. H. C. Theurer, J. Electrochem. Soc., 108, 649 (1961)
2. E. Ernst, D. Hurd, G. Seely and P. Olshefski, Paper presented at the Electrochem. Soc. meeting, Buffalo, N. Y., Oct. 10-14, 1965.
3. F. C. Eversteijn, P. J. W. Severin, C. H. J. v.d. Brekel and H. L. Peck, Paper presented at the Electrochem. Soc. meeting, New York, N. Y., May 4-10, 1969.
4. J. G. Wilkes, Paper presented at the Electrochem. Soc. meeting, Oct. 1966.
5. V. Y. Doo and E. O. Ernst, SCP and Solid State Technology, P.31, Oct. 1967.
6. P. C. Rundle, Int. J. Electronics 24, 405 (1968).
7. Dinesh C. Gupta and Roy Yee, J. Electrochem. Soc., 116, 1561 (1969).
8. B. A. Joyce and R. R. Bradley, J. Electrochem. Soc. 110, 1235 (1963).
9. S. R. Bhola and A. Mayer, RCA Review, 24, 511 (1963).
10. S. E. Mayer and D. E. Shea, J. Electrochem. Soc., 111, 550 (1964).
11. Dinesh C. Gupta and Pei Wang, Solid State Technology, P.48, Oct. 1968.
12. D. Richman and R. H. Arlett, J. Electrochem. Soc. 116, 872 (1969).
13. W. H. Shepherd, ibid, 112, 1988 (1965).
14. S. E. Bradshaw, Int. J. Electronics, 21, 205 (1966).
15. S. E. Bradshaw, ibid, 23, 381 (1967).
16. R. P. Sutton, "Semiconductor Silicon", The Electrochemical Society, New York, N. Y., P.214, (1969).
17. S. Nakanuma, IEEE Trans. on Electron Devices, ED-13, 578 (1966).
18. P. Schumann, Jr., J. Electrochem. Soc., 116, 409 (1969).
19. A. C. Ipri, Solid State Technology, 13, 38 (April 1970).
20. J. Bloem, J. Electrochem. Soc., 117, 197 C (1970).

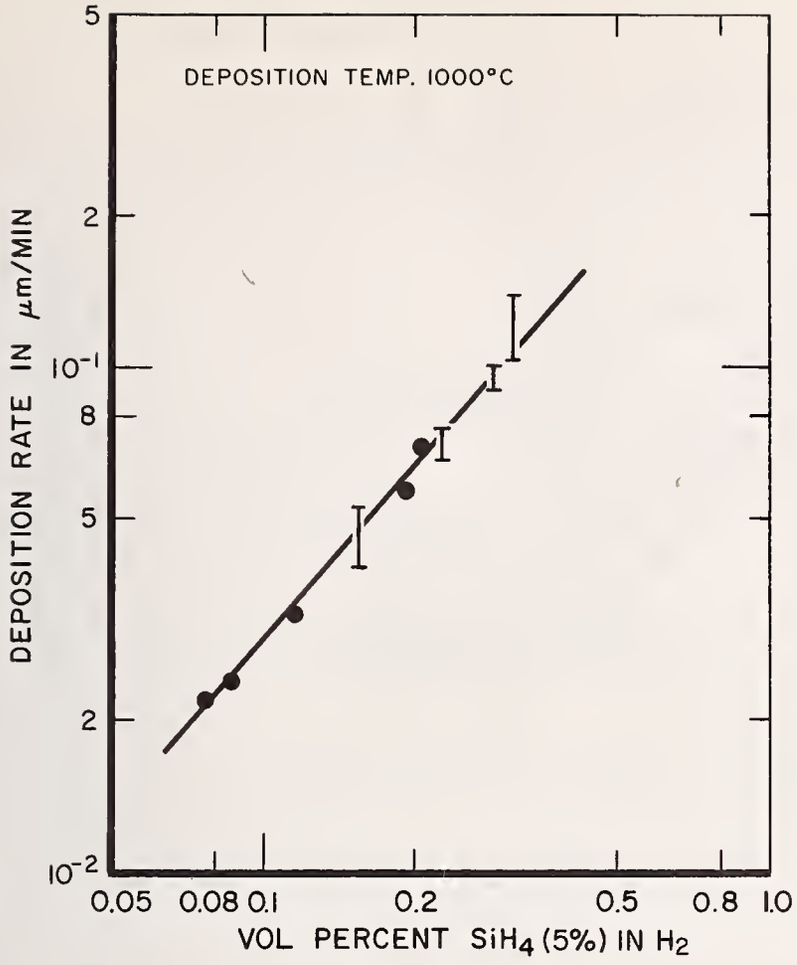


Fig. 1. Effect of SiH_4-H_2 ratio on the silicon deposition rate using 5 percent SiH_4 mixture in hydrogen (7).

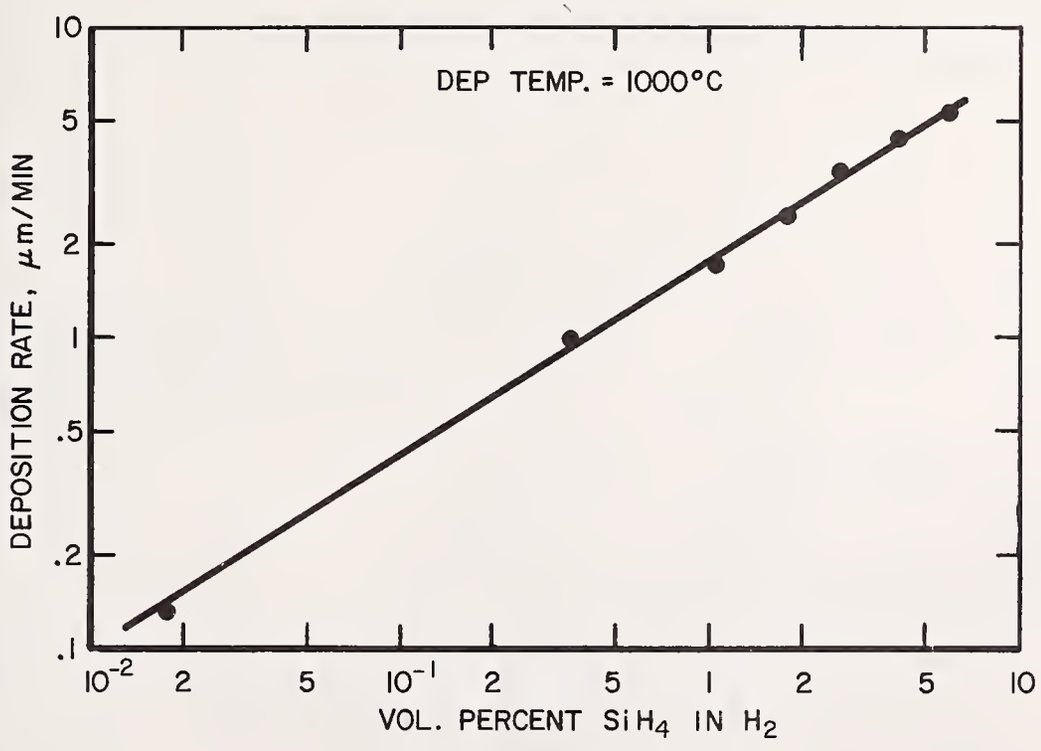


Fig. 2. Effect of SiH_4-H_2 ratio on the silicon deposition rate using 100 percent SiH_4.

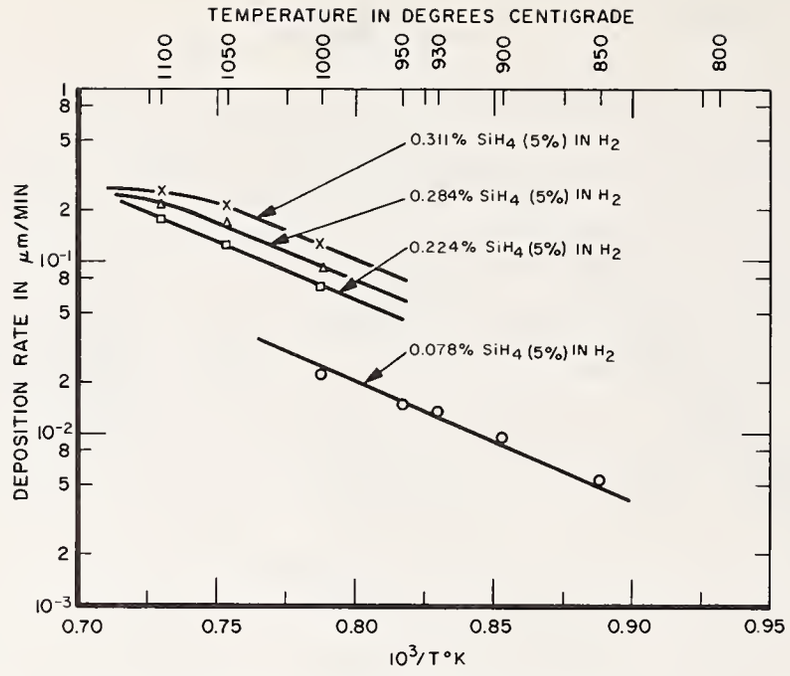


Fig. 3. Temperature dependence on silicon deposition rate using 5 percent SiH_4 mixture in hydrogen (7).

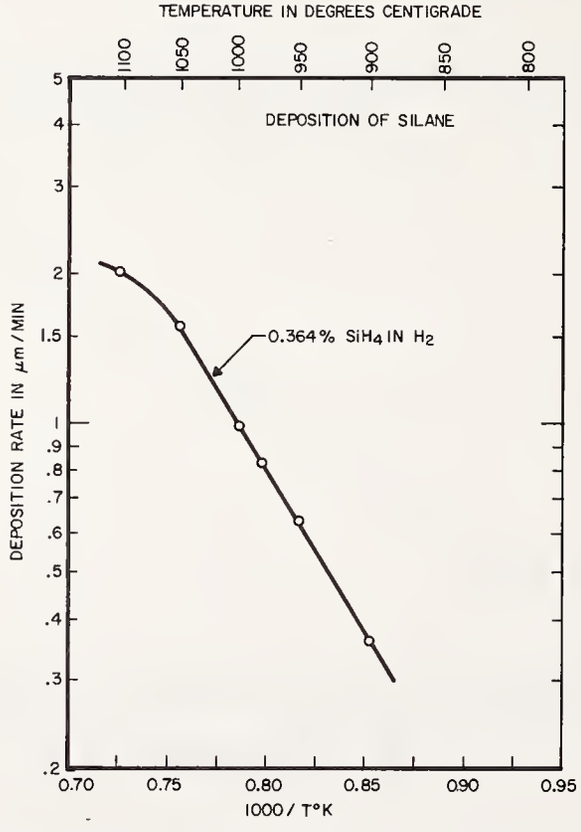


Fig. 4. Temperature dependence on silicon deposition rate using 100 percent SiH_4.

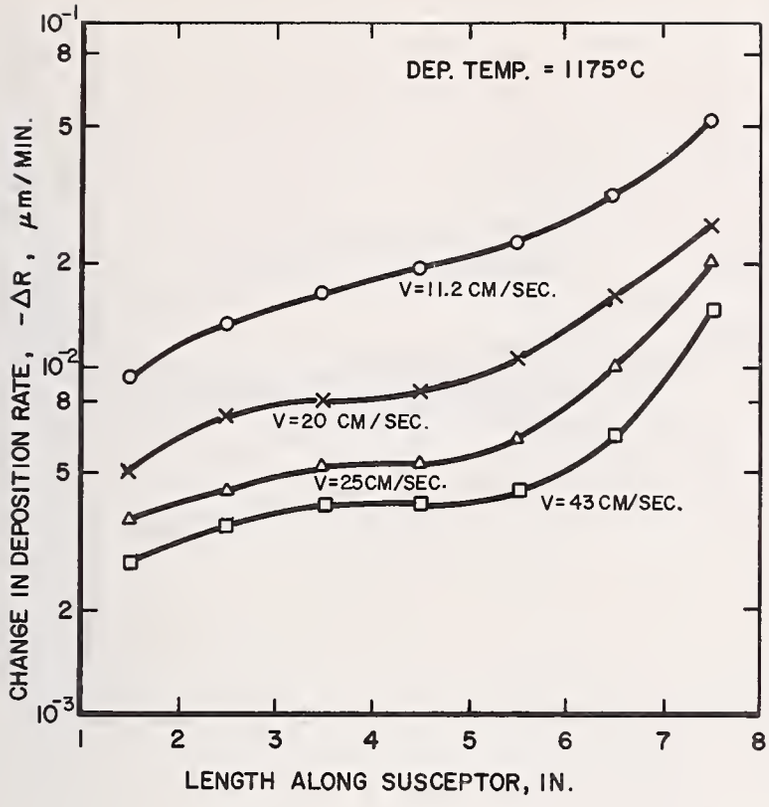
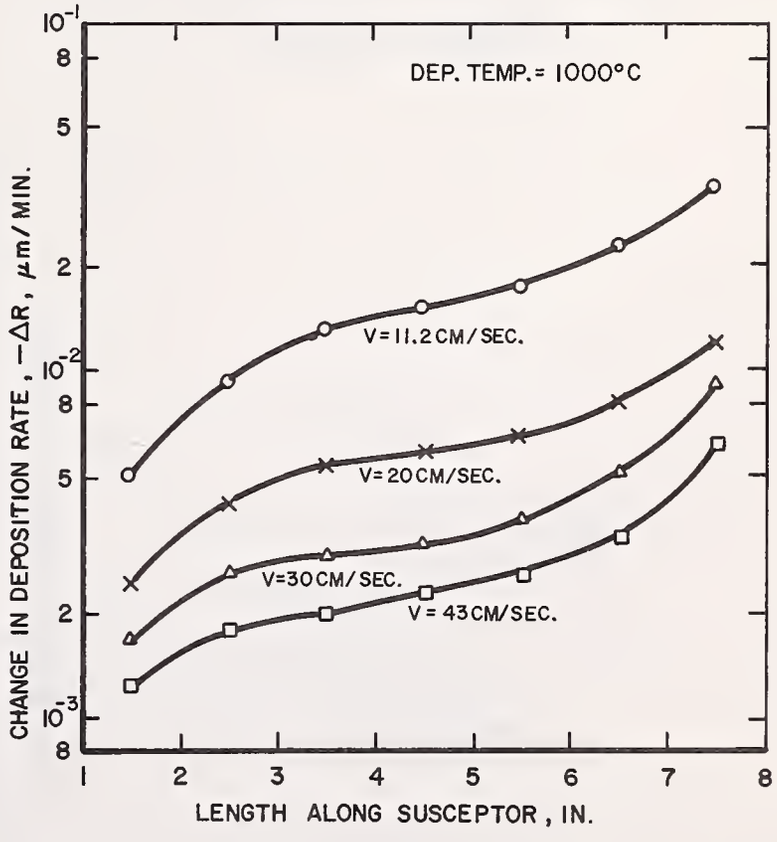


Fig. 5. Longitudinal uniformity as function of gas velocity for depositions using $SiCl_4$ in round tube configuration.

Fig. 6. Longitudinal uniformity as a function of gas velocity for depositions using SiH_4 in round tube configuration.



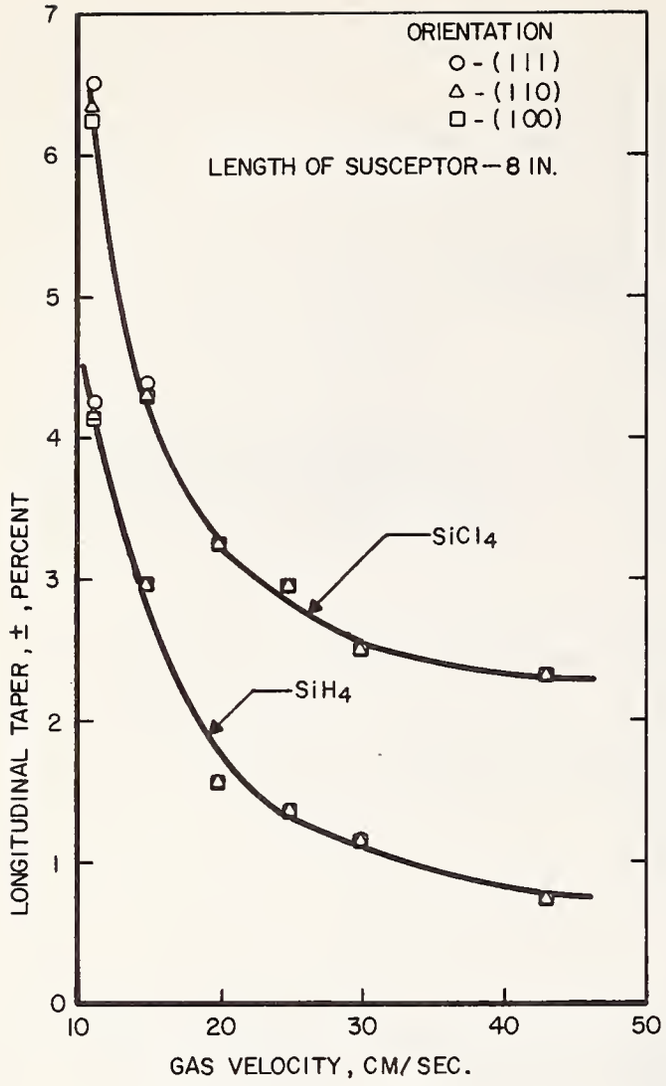


Fig. 7. Longitudinal taper vs. gas velocity in the round tube configuration.

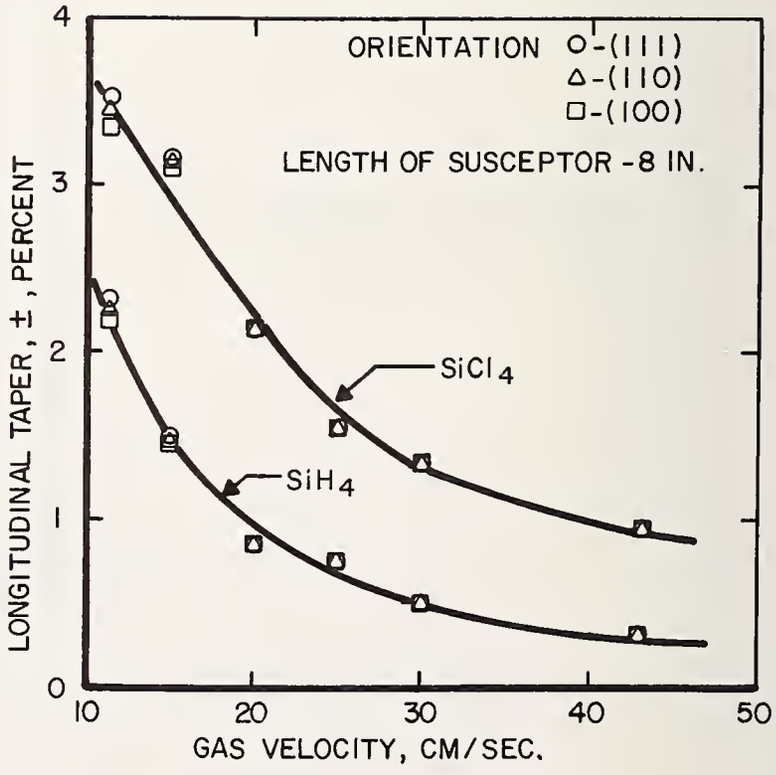


Fig. 8. Longitudinal taper vs. gas velocity in the rectangular tube configuration.

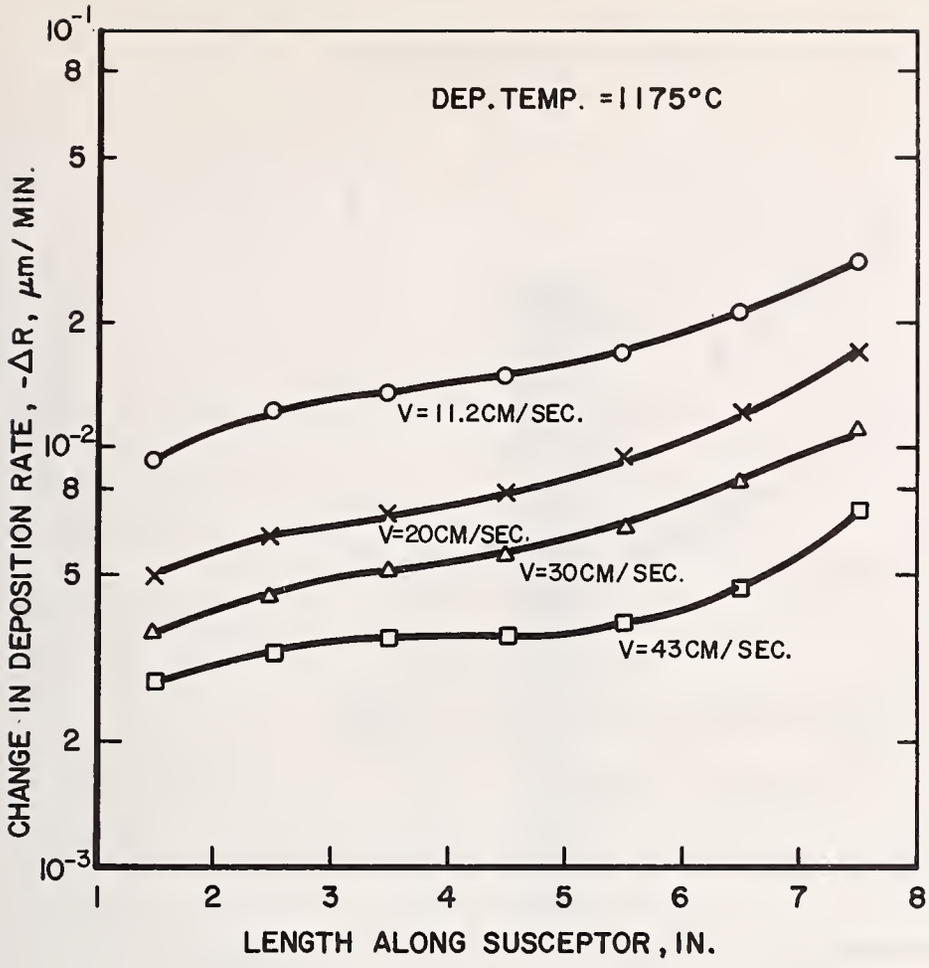
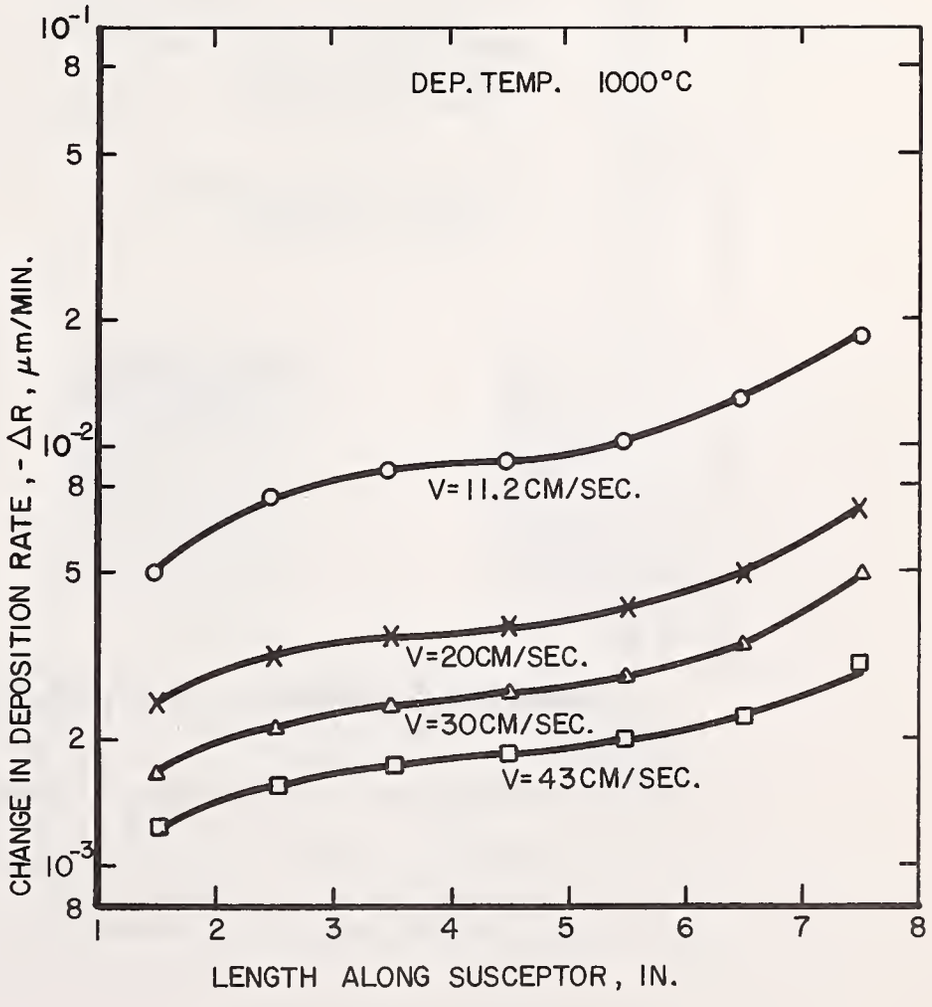


Fig. 9. Longitudinal uniformity as a function of gas velocity for depositions using $SiCl_4$ in rectangular tube configuration.

Fig. 10. Longitudinal uniformity as a function of gas velocity for depositions using SiH_4 in rectangular tube configuration.



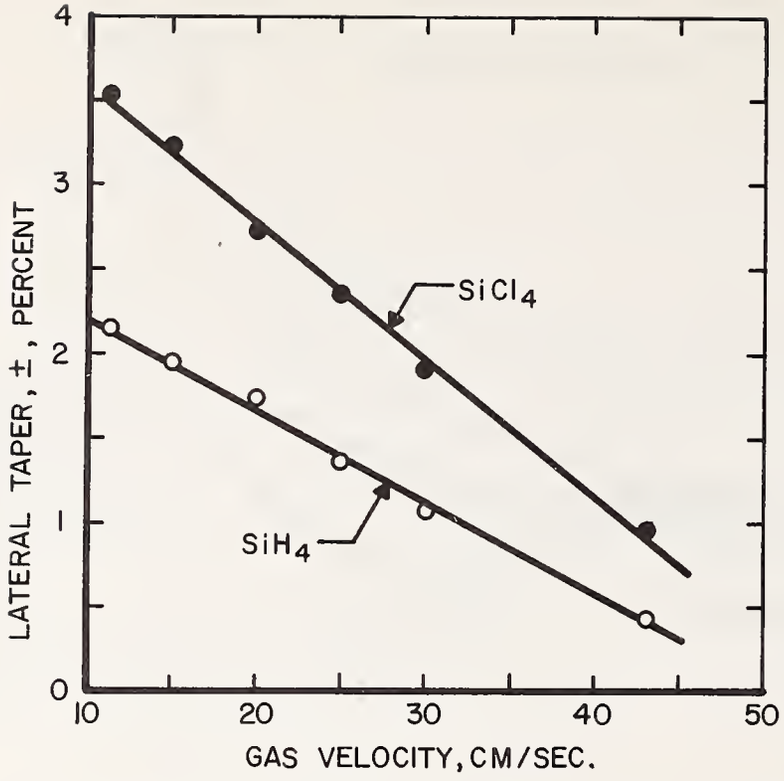


Fig. 11. Lateral taper vs. gas velocity (round tube configuration).

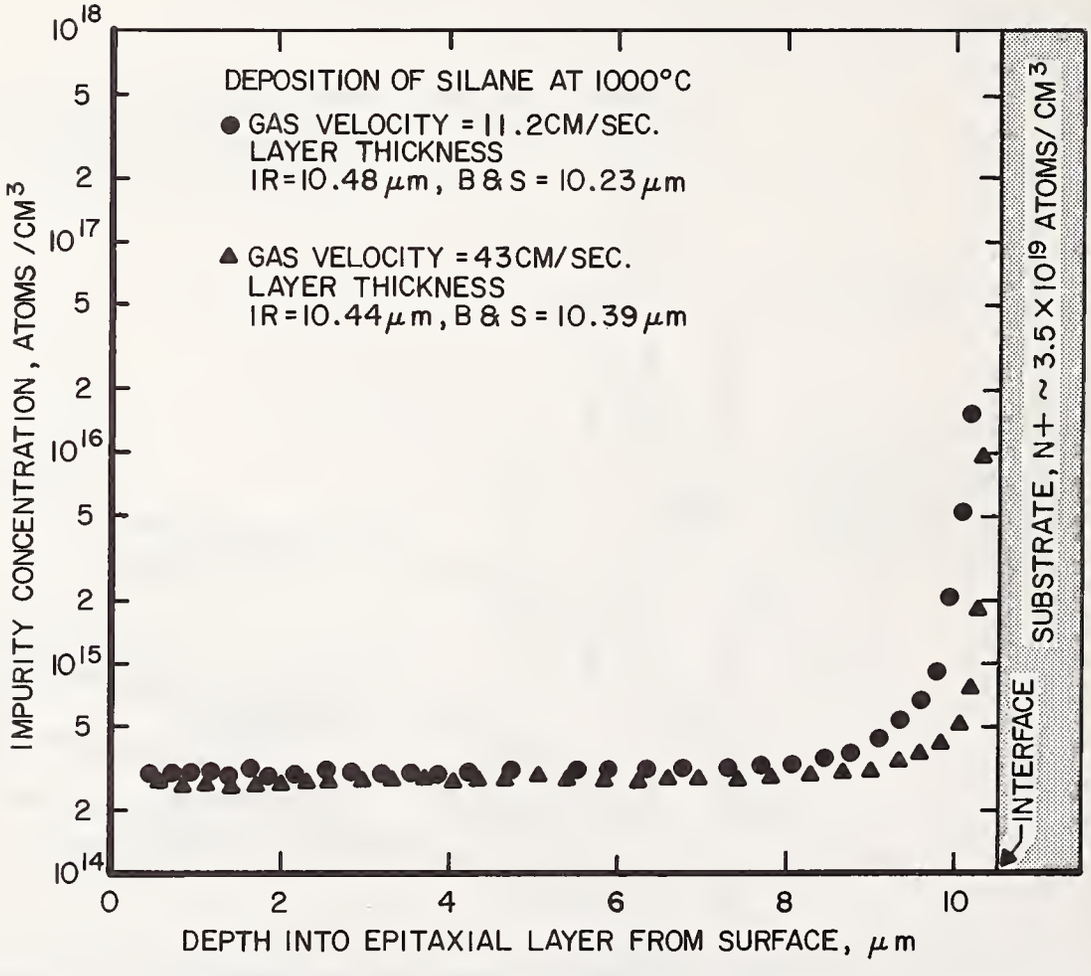


Fig. 12. Effect of gas velocity on the impurity distribution in the layer.

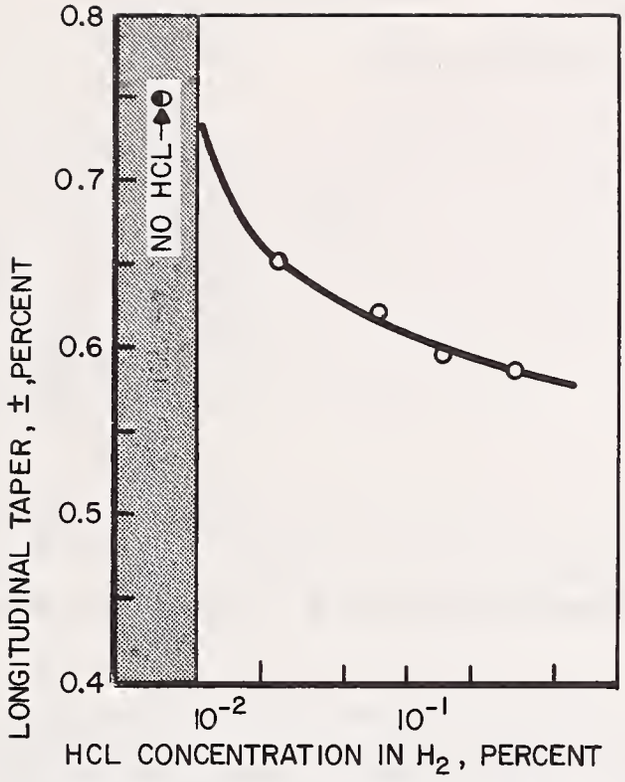


Fig. 13. Longitudinal taper vs. HCl concentration in H_2 (HCl introduced during deposition).

(a) 0.025% HCl (b) 0.16% HCl (c) 0.33% HCl in H_2

Fig. 14. Quality of the surface for various HCl concentrations introduced during deposition using SiH_4 (a) 0.025% HCl, (b) 0.16% HCl, (c) 0.33% HCl in H_2.

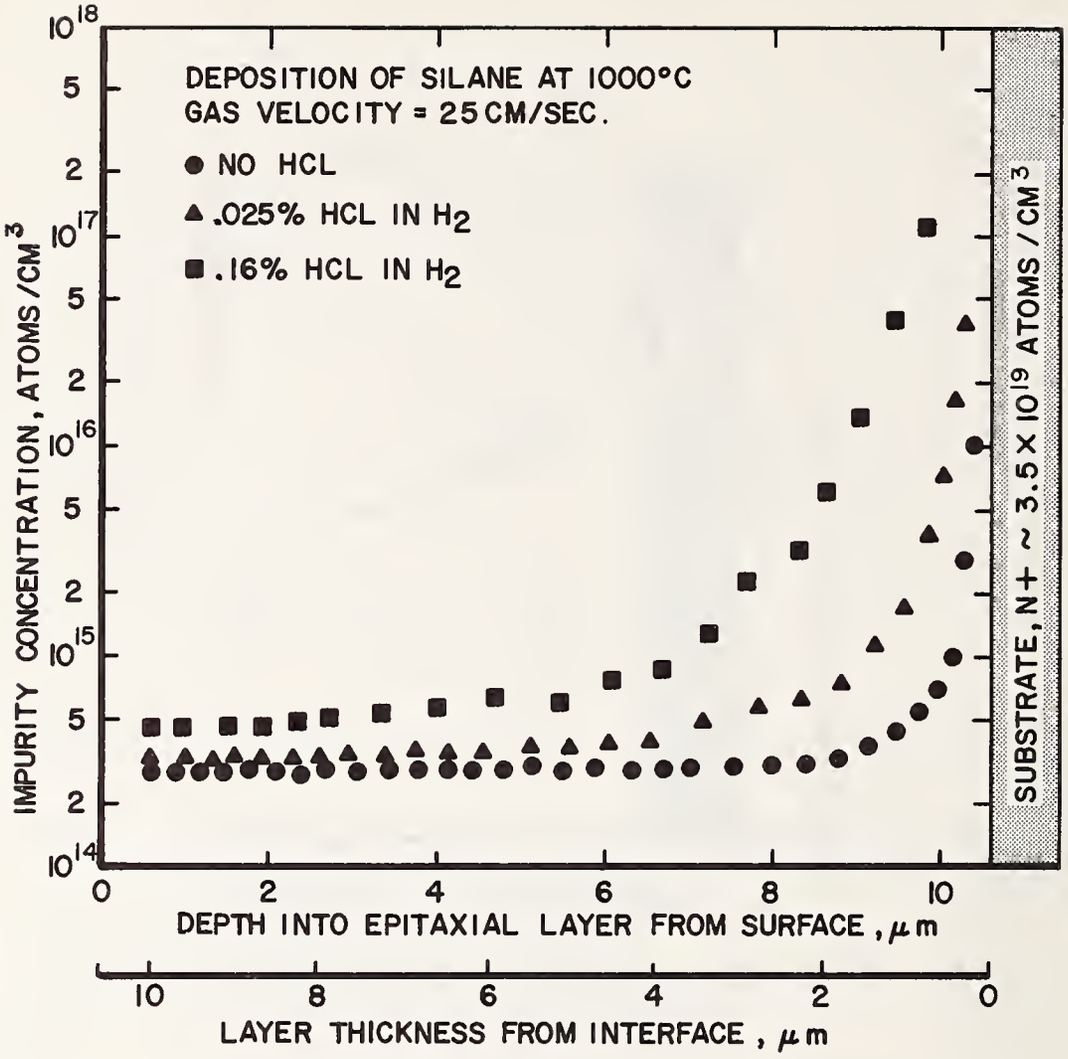


Fig. 15. Effect of introduction of HCl vapors during deposition on the impurity profile.

Fig. 16. Photomicrographs of the surface of the epitaxial layer deposited with (a) 11.2 cm/sec, (b) 25 cm/sec, (c) 43 cm/sec.

Control of Thin Silicon Films Grown from Silane

D. J. Dumin

RCA Laboratories
Princeton, N. J. 08540

The electrical properties of thin films of silicon grown epitaxially on sapphire or spinel can be controlled to a considerable extent by control of the growth parameters. Unlike the case of homoepitaxial silicon growth, however, the electrical properties of heteroepitaxial silicon can be considerably altered by relatively small variations in the growth conditions. The variations in growth rates and growth temperatures combined with variations in the carrier ambient and purity of the source gases, makes analysis of the growth-dependent, electrical properties somewhat complicated. Several aspects of the characterization of the electrical properties of the silicon films will be discussed. Particular attention will be paid to the effects of variations in growth temperature, growth rate, film thickness, and defect structure in the thin silicon films.

Key Words: Thin silicon films, silicon-on-sapphire, electrical properties, growth dependent properties, defect structure, silane, silicon-on-insulators, silicon-on-spinel.

1. Introduction

Recently thin films of silicon-on-sapphire [1][1] and silicon-on-spinel [2] have become available. These films have been utilized in the fabrication of diodes [3], MOS transistors [4], and bipolar transistors [5]. The MOS transistors have found applications in relatively complex CMOS integrated circuits [6].

In order to facilitate fabrication of the MOS transistors and integrated circuits, it has proven necessary to extensively study the electrical properties of the thin silicon films [7]. It was found that the electrical properties of the films were a strong function of the growth parameters. In this paper the influence of several of the growth parameters on the electrical properties of the thin silicon films will be discussed.

2. Experimental Conditions

The substrates were single crystal wafers, 0.010 to 0.020" thick of sapphire or spinel. Silicon with (100) orientation was grown on ($1\bar{1}02$) sapphire and (100) spinel, while (111) silicon was grown on (0001) sapphire and (111) spinel [8]. The substrates were mechanically polished and the work damaged layer left by the mechanical polishing was removed from the substrate surface prior to growth by a short high temperature firing in hydrogen [9].

The films were grown in both vertical and horizontal, water cooled, RF heated, reaction chambers. Silicon and silicon carbide coated graphite have been used as the susceptors. The silicon was grown via thermal decomposition of SiH_4-H_2 mixtures and p-type and n-type impurities have been grown into the films via the addition of B_2H_6-H_2, PH_3-H_2, or AsH_3-H_2 mixtures to the gas stream during growth. Palladium diffused H_2 was used as the carrier gas for growth of most of the films discussed in this paper. However, dry He was used as a carrier gas for growth of some of the films.

A typical growth cycle was:

a) Prefire in H_2 at 1300-1500°C for 15 min to 60 min.
b) lower the temperature to the growth temperature--usually between 1000°C and 1200°C.
c) grow the film at rates from 0.2 to 15 μm/min
d) post fire the film in H_2, H_e, or O_2. This step was <u>not</u> performed on all of the wafers, only on a selected portion of them. This step was used to simulate steps involved in device processing.

[1] Figures in brackets indicate the literature references at the end of this paper.

The electrical characteristics of the films have been determined by measuring the Hall mobility, μ, and the resistivity, ρ, on 6-terminal etched Hall bars. The carrier concentration, n, was determined from the measured properties using $n = 1/\rho e\mu$, where e is the electronic charge. On numerous samples the Hall data were taken before and after a post-growth heat treatment.

The film thickness was determined during growth by using IR interference techniques [10]. The usefulness of this method of measuring film thickness can not be over emphasized. The IR interference scheme allowed control of film thickness under variable growth conditions and was an immediate indicator of problems in the growth system.

A plot of the output from an IR pyrometer with peak sensitivity at 2.4 μm is shown in Fig. 1 as a function of time. When the film started to grow, an oscillating pattern appeared in the pyrometer output. The film thickness corresponding to constructive interference is given by

$$t = \frac{\lambda}{2\eta} \tag{1}$$

where t is the film thickness corresponding to one cycle in Fig. 1, λ is the wave length of the pyrometer and η is the index of refraction in the silicon. For silicon at 1000°C to 1200°C, one cycle corresponds to about 0.34 μm. The oscillations eventually become clamped due to absorption in the film and surface roughness or faceting in the surface of the growing film.

3. Growth-Temperature Dependent Properties

Rather intuitive, and possibly naive, arguments can be given for suspecting that there might be an optimum temperature or temperature range over which good quality films could be grown. Growth at excessively high temperatures would be expected to lead to high atomic surface mobilities and agglomeration of the films. Also, gas-phase decomposition and autodoping would be expected to be more severe at higher growth temperatures. At lower growth temperatures poor crystal structure due to low atomic surface mobilities could be expected and possible incomplete decomposition of the SiH_4 could occur.

In order to illustrate some of the types of effects that occur as the growth temperature was changed, data have been presented in Fig. 2 and Fig. 3 showing the variations in carrier concentration and Hall mobility as a function of growth temperature. Data taken on different film thicknesses and on films grown on (0001) and ($1\bar{1}02$) sapphire and (111) spinel have been presented. All of the data shown in Fig. 2 and Fig. 3 were taken on p-type films.

The carrier concentration-growth temperature data shown in Fig. 2 indicate that as the growth temperature was increased the hole concentration rose. Spectrographic analysis of these films showed that the films were doped with aluminum [11]. The increased aluminum concentration at higher growth temperatures was due to aluminum autodoping from the Al_2O_3 or $MgAl_2O_4$ substrates. The substrate was reduced by either the hydrogen or the silicon or a combination of the two. Evidence has pointed to hydrogen reduction of the Al_2O_3 as being the dominant mechanism at 1000°C and silicon reduction as being significant above 1250°C. The portion of the aluminum due to contamination from the reverse side of the wafer has not been determined but evidence for doping from the backs of the wafers has been presented [12].

It is significant that the films grown on (111) spinel have been doped less heavily than films grown on sapphire, and that the reduction in the amount of autodoping can be several orders of magnitude. It is also significant that films grown in He as an ambient were less heavily doped than the films grown in the H_2 ambient by over an order of magnitude.

Accompanying the variations in aluminum autodoping were changes in the Hall mobility as shown in Fig. 3. The simplest explanation of this data would be; a) there is an optimum temperature at which the highest mobility films are grown; b) at high growth temperature the mobilities dropped due to both increased impurity scattering and, possibly, poorer crystal structure; c) at lower growth temperatures the mobilities dropped due to poorer crystal structure; d) as the film thickness rose, the mobilities rose, due to partial self-annihilation of crystalline defects: e) The optimum growth temperature, like the autodoping, was different for each substrate material and orientation.

As far as these explanations go, they are all correct. However, further experiments and attempts to grow films at temperatures of less than 1000°C brought up two other effects acting to lower the mobilities at lower growth temperatures. As the growth temperature was lowered below 1100°C, the density of deep trapping levels increased [13]. These levels have been identified as a donor 0.3 ev above the valence band edge and an acceptor 0.25 ev below the conduction band edge. These two levels can act as both hole and electron traps and can thus lead to depletion of carriers. They also act as scattering centers and can lower the Hall mobility. These levels have been associated with a defect-impurity interaction, although the impurity species has not yet been identified.

Careful examination of the sources of SiH_4 used to grow the films described in Fig. 2 and Fig. 3 showed that all of these sources were contaminated with N-type impurities which doped the films in the range $10^{13}/cm^3$ to $10^{15}/cm^3$, depending on the tank of gas used. Thus a portion of the drop in Hall mobility observed in the films grown at lower growth temperatures was due to the presence of unwanted donors acting us compensating centers. These donors were most effective in lowering the mobility of films doped less than $10^{16}/cm^3$.

4. Growth-Rate Effects

The growth rate is linked to the growth temperature in determining the electrical properties of the films. At a given growth temperature, the films will be more heavily doped with aluminum as the growth rate is lowered. Thus, in general, faster growth rates lead to less autodoping and slower growth rates lead to more autodoping. However, the crystalline perfection of the films is also related to the growth rate, and, in general, the optimum growth rate appears to decrease as the temperature is lowered. More important than the growth rate, however, is the nucleation rate. It is important that when the insulator surface has been completely covered, this layer should have a high degree of crystalline perfection.

An example of the variations in crystalline perfection that can occur in films grown at different growth rates is illustrated in Table 1. The Hall mobilities of several 0.15 μm thick films grown on $(1\bar{1}02)$ sapphire at 1150°C have been measured. The doping density of all of the films was kept nearly constant at $2\text{-}5\cdot10^{17}/cm^3$. As the growth rate was increased from 0.7 μm/min to 10 μm/min, the Hall mobilities continually rose. Growth rates above 10 μm/min resulted in lowered mobilities. Thus at growth temperatures of about 1150°C, the optimum nucleation rate was about 10 μm/min. This optimum nucleation rate was a function of growth temperature and dropped as the growth temperature dropped. The nucleation rate and growth rate of these films were varied and the Hall mobility and carrier concentration were measured. The Hall data taken on these are described in Table 2. These films were grown with a source of SiH_4 containing N-type impurities and produced 10 Ωcm, N-type, silicon-on-silicon. As seen in Table 2, as the nucleation rate was increased, the amount of aluminum autodoping dropped significantly. In fact, the film grown with a nucleation rate of 4 μm/min and a growth rate of 1 μm/min was nearly perfectly compensated and thus the Hall mobility was very low, about 45 $cm^2/Vsec$. When the growth rate and nucleation rate were raised to 4.5 μm/min, the aluminum concentration in the films dropped and the films became reasonably high mobility N-type films. It should be noted that a significant portion of the autodoping was introduced in the nucleation step, which indicates the importance of covering the surface of the insulator both quickly and with a layer of high crystalline perfection. This data also emphasizes the fact that it is incomplete to measure only carrier concentration and Hall mobility on a sample and attempt to draw conclusions concerning the crystallinity of the film. The low mobility film described in Table 2 had a low mobility caused by unintentional compensation and not by poor crystalline perfection.

Table 1. Hall mobilities of 0.15 μm thick silicon films grown at 1150°C on $(1\bar{1}02)$ sapphire. $N_A \simeq 2\text{-}5\cdot10^{17}/cm^3$

Growth Rate (μm/min)	Hall Mobility (cm^2/V sec)
0.7	40
1.5	59
4.0	68
10	100
15	61

Table 2. Mobility and carrier concentration of (111) silicon films grown on (111) spinel at 1150°C

Nucleation Rate (μm/min)	Growth Rate (μm/min)	Hall Mobility (cm^2/v sec)	Carrier Concentration ($^1/cm^3$)
0.3	1.0	300	$1.1\cdot10^{16}$
1.0	1.0	322	$7.8\cdot10^{15}$
4.0	1.0	45	$4.5\cdot10^{15}$
*4.5	4.5	510	$1.4\cdot10^{14}$

* N-type film

5. Thickness Dependent Properties

One of the problems associated with the interpretation of the electrical properties of the thin film is the fact that the electrical characteristics of the film are a function of the film thickness. This occurs primarily because as the films grows thicker, the defects generated at the silicon substrate interface propagate through the film. These defects are partial self-anihilating. Thus, the density of defects intersecting the free surface decreases as the film thickness increases. A similar effect has been demonstrated in germanium thin films [14]. When determining the quality of films grown under different growth conditions, it is most important to assure that the films are all of comparable thicknesses.

It has been determined that the defects in silicon-on-sapphire films can act as deep donors and acceptors [13]. The donor is located about 0.30 ev above the valence band edge and the acceptor about 0.25 ev below the conduction band edge. The energy relationship of these levels is shown in Fig. 4. These levels appear to occur in nearly equal densities in the films and are thus referred to as a donor-acceptor pair. The density of the donor-acceptor pair is proportional to the defect density, but since an $HCl-O_2$ gettering [15] step can be used to reduce the pair density, the donor-acceptor pair apparently is related to a defect-impurity interaction. The position of these levels makes them ideal as recombination centers for minority carriers in either P-type or N-type films, and probably causes the low minority carrier lifetimes in these films [3].

The defect generated levels are also important as trapping and scattering centers. Several P and N-type films were grown in which the doping density was kept fixed throughout the film. The electrical properties of these films were measured as the films were sucessively reduced in thickness [16]. The measured values of carrier concentration and Hall mobility as a function of film thickness are shown in Fig. 5 and Fig. 6. The curves labeled "Measured Total Concentration (mobility)" were the measured values, while the curves labeled "Calculated Layer Concentration (mobility)" were the calculated values of the removed layer. It is seen that as the silicon-sapphire interface was approached the hole and electron concentrations dropped due to the trapping of carriers at the deep levels. The carrier scattering also rose as the film thickness dropped, resulting in the lower Hall mobilities near the silicon-sapphire interface. Similar results have been reported for silicon-on-spinel films [17].

6. Summary

Some of the growth-dependent, electrical properties of thin silicon-on-insulator films have been described. The importance of growth temperature, growth rate and film thickness have been discussed. While this paper has been a very incomplete discussion of the problem of the control of thin silicon films, it points out some problems that arise in the control of thin films that are not so significant in the control of bulk homoepitaxial layers.

7. Acknowledgements

The work described in this paper was largely the result of a joint effort involving many people at the RCA Laboratories. The author gratefully acknowledges the assistance and cooperation of C. W. Mueller, P. H. Robinson, G. W. Cullen, and G. E. Gottlieb.

8. References

[1] R. W. Bicknell, J. M. Charig, B. A. Joyce, and D. J. Stirland, Phil. Mag. 9, 965 (1964); H. M. Manasevit and W. I. Simpson, J. Appl. Phys. 35, 1349 (1964); C. W. Mueller and P. H. Robinson, Proc. IEEE 52, 1487 (1964).

[2] H. M. Manasevit and D. H. Forbes, J. Appl. Phys. 37, 734 (1966); P. H. Robinson and D. J. Dumin, J. Electrochem Soc 115, 75 (1968).

[3] D. J. Dumin and R. S. Silver, Solid-State Electronics 11, 353 (1968).

[4] C. W. Mueller and P. H. Robinson, Proc. IEEE 52, 1487 (1964).

[5] F. P. Heiman and P. H. Robinson, IEEE Trans. ED-11, 411 (1968).

[6] J. F. Allison, J. R. Burns, and F. P. Heiman, IEEE J. Solid-State Circuits SC-2, 208 (1967); J. F. Allison, D. J. Dumin, F. P. Heiman, C. W. Mueller, P. H. Robinson, Proc IEEE 57, 1490 (1969).

[7] D. J. Dumin, J. Appl. Phys 38, 1909 (1967); D. J. Dumin and P. H. Robinson, J. Appl Phys 39, 2759 (1968).

[8] H. M. Mansevit, et.al, Trans. Met. Soc. AIME 242, 465 (1968).

[9] D. J. Dumin, J. Appl. Phys. 38, 1909 (1967).

[10] D. J. Dumin, Rev. Sci. Instr. 38, 1107 (1967).

[11] D. J. Dumin and P. H. Robinson, J. Electrochem. Soc 113, 469, (1966).

[12] J. D. Filby and S. Nielson, Brit, J. Appl. Phys. 18, 1357 (1967). P. B. Hart, et. al, Brit.J. Appl. Phys. 18, 1389 (1967).

[13] D. J. Dumin, Bull. Am. Phys. Soc, 14, 394, (1969); D. J. Dumin, Solid-State Electronics, To be published.

[14] B. W. Sloope and C. O. Tiller, J. Appl. Phys 37, 887 (1966).

[15] S. W. Ing, R. E. Morrison, L. L. Alt, and R. W. Aldrich, J. Electrochem. Soc. 110, 533 (1963).

[16] D. J. Dumin and P. H. Robinson, J. Crystal Growth 3, 4, 213 (1968).

[17] G. W. Cullen, G. F. Gottlieb, C. C. Wang, and K. H. Zaininger, J. Electrochem. Soc. 116, 1444 (1969).

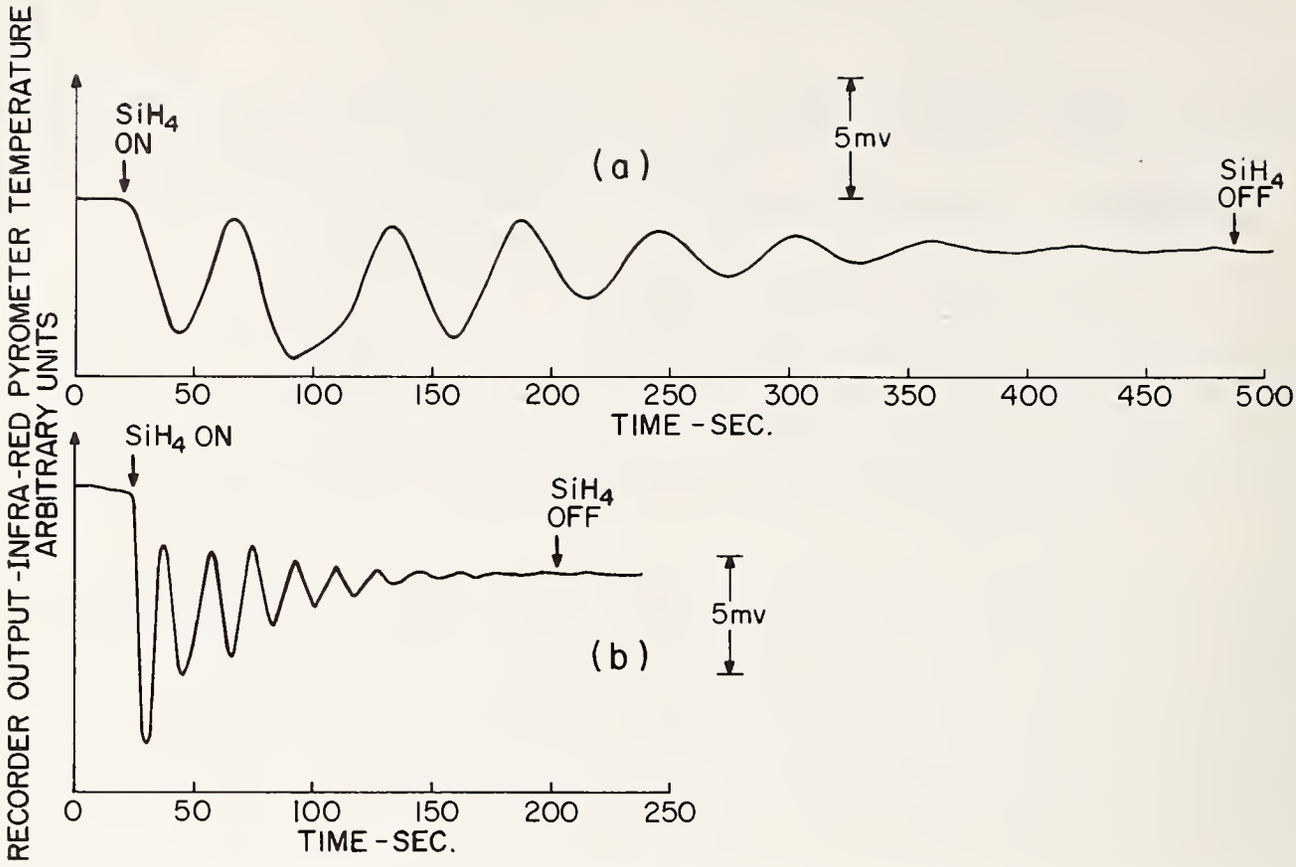


Fig. 1: Output voltage from the IR Pyrometer as a function of time for two growth rates of silicon-on-sapphire

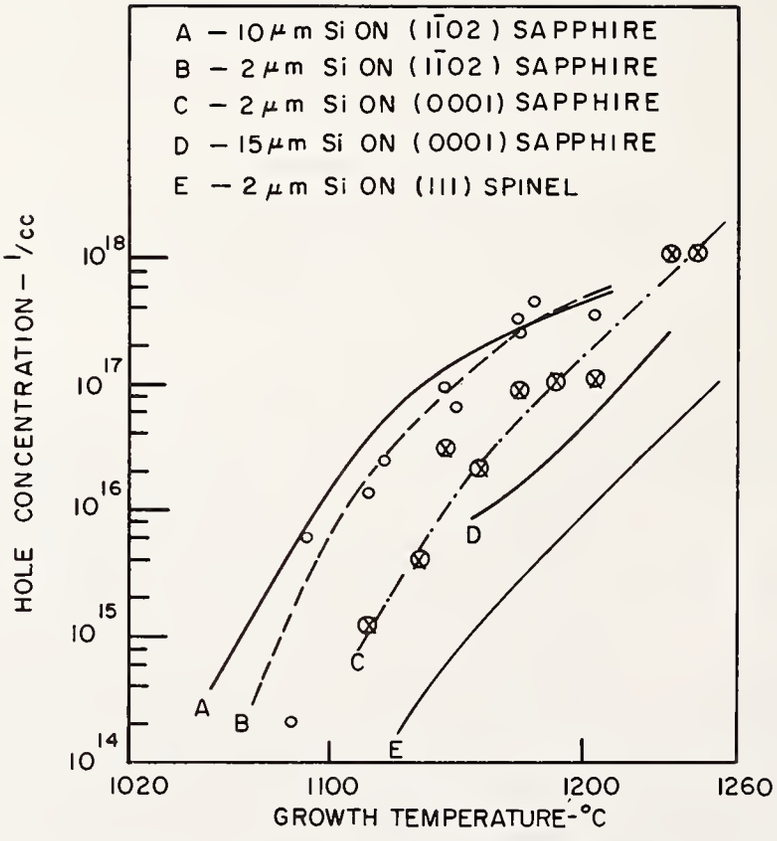


Fig. 2: Carrier concentration vs growth temperature for P-type films on silicon and spinel

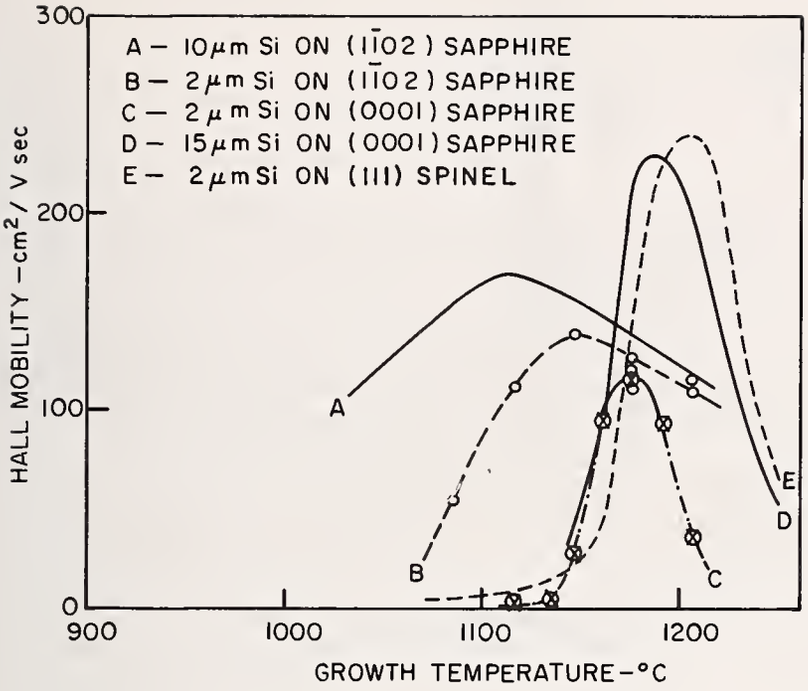


Fig. 3: Hall mobility vs growth temperature for P-type films on sapphire and spinel

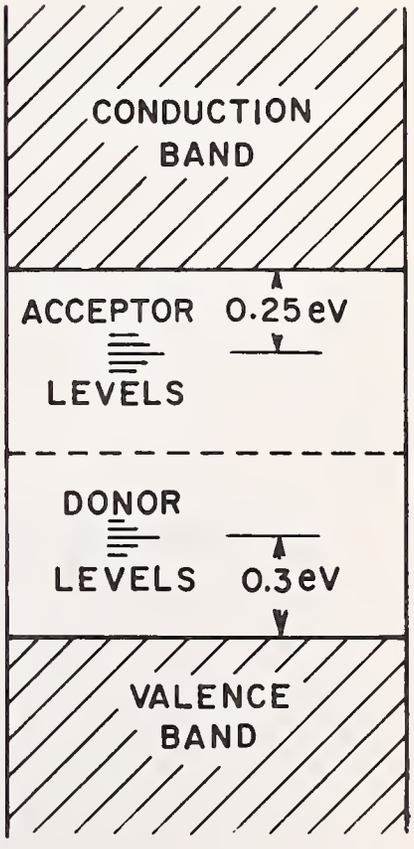


Fig. 4: The energy positions of the deep donor and acceptor pair in silicon-on-sapphire. This pair is due to an impurity-defect interaction.

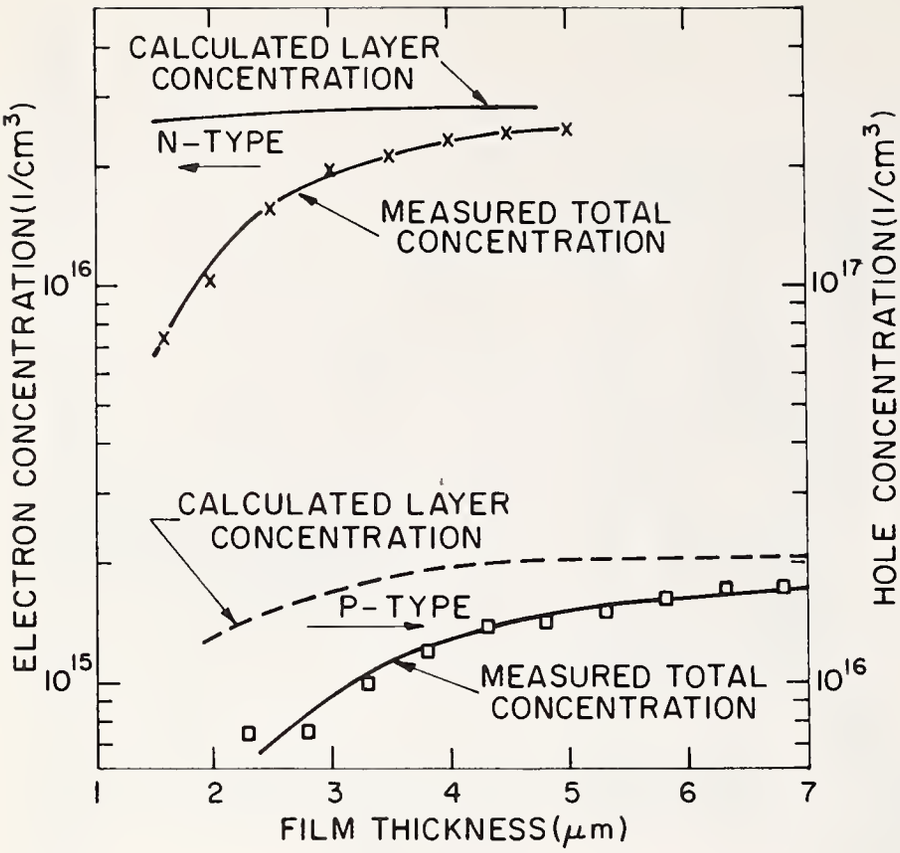


Fig. 5: Carrier concentration vs. thickness in P-type and N-type films

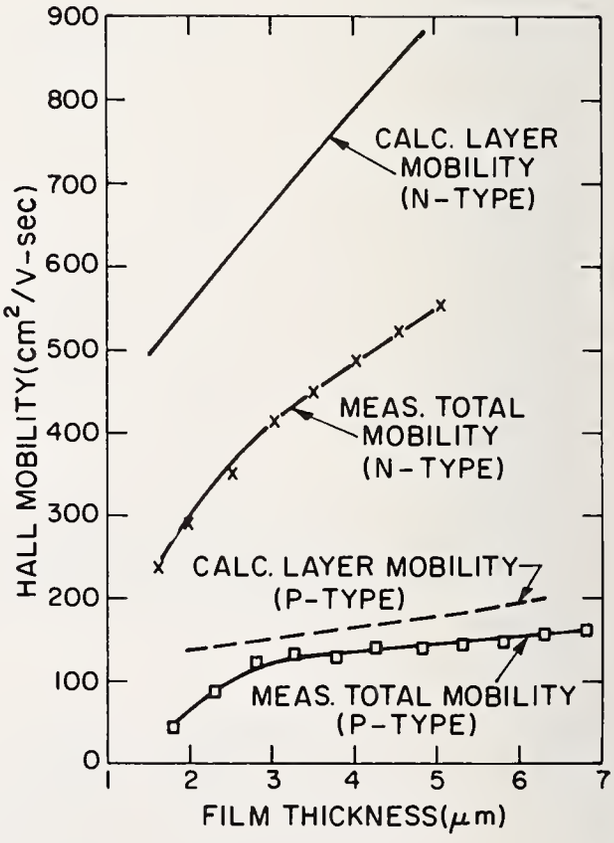


Fig. 6: Hall mobility vs. thickness in P-type and N-type films

The Growth of Submicron Single and Multilayer Silicon Epitaxy

J. Simpson, A. C. Adams and M. H. Hanes

Bell Telephone Laboratories, Incorporated
Murray Hill, New Jersey 07924

A study of the growth of submicron single and multiple epitaxial silicon layers deposited by the silane pyrolysis method has been made. The structural and electrical properties of such layers are discussed in relation to the silicon deposition parameters and to the substrate surfaces. The silane decomposition process has been examined over the temperature range 950°-1150°C and the ability to control reproducibly both thin film uniformity and thickness is shown to be strongly dependent on the linearity of the process parameters studied. The problems associated with the characterization of substrate surfaces and epitaxial layers are discussed in terms of the techniques used to measure surface cleanliness and the thickness and distribution of impurity atoms in epitaxial layers.

Key Words: Characterization, epitaxial thickness, epitaxy, flow systems, impurities, multilayers, profiling, silane, silicon, surfaces.

1. Introduction

The fabrication of high frequency electron devices and integrated circuits requires that submicron single and multilayer silicon epitaxial layers be grown with rigorous control over both uniformity and reproducibility of epitaxial thickness and resistivity. The need to improve the abruptness of the impurity atom distribution profile at electrical junctions has led to the examination of various methods of depositing epitaxial silicon at reduced temperatures such as evaporation, sputtering, liquid epitaxy, etc., methods which have been successful but which have not so far been developed sufficiently for device fabrication purposes. The present work on silane pyrolysis was undertaken on the basis that the development of silane technology at temperatures of approximately 1000°C would be compatible with existing silicon tetrachloride reduction processing and thus could be readily put into manufacturing capability. In particular, two problems associated with the 1200°C reduction of silicon tetrachloride should be considerably minimized, namely autodoping of the epitaxial layers by the chemically reactive hydrogen chloride reaction by-product and outdiffusion of impurity atoms from heavily doped substrates.

Chemically, the silane pyrolysis produces only silicon and hydrogen as end products and control of epitaxial layer doping, especially for lightly doped layers over heavily doped substrates, should be significantly improved. In the case of submicron epitaxial layers on heavily doped substrates, it is essential that outdiffusion be reduced to a minimum so that flat portions of the impurity profile may exist for device purposes. Figure 1 illustrates this point by showing plots of erfc calculations for the outdiffusion of arsenic from a substrate, doping concentration 10^{20} carriers/cm^3, into an epitaxial layer doped to 8×10^{16} carriers/cm^3, for two temperatures, namely, 1000°C and 1200°C. The important features to note are the time scales and the simulated junction profiles -- at 1000°C only 0.1 micron outdiffusion will occur over a five-minute interval, whereas at 1200°C approximately 0.3 micron outdiffusion will occur over a one-minute interval, a significant fraction of a one-micron epitaxial layer. In a technology which relies on controlling thickness by the growth time in a flow system, the factor of 5 in time should provide considerable control over thickness reproducibility in thin epitaxy. Previous work on silane

pyrolysis [1-6][1] has shown that reasonable epitaxial growth rates may be achieved in the temperature range of 1000°C - 1100°C. This paper presents an overview of those characteristic properties of silane decomposition which enable submicron layers to be readily grown in a controllable manner and also those problems encountered in measuring such layers.

2. Apparatus

The system consisted of a standard single slice vertical flow water cooled quartz chamber with the 1-1/4 inch silicon slice held on a high density graphite susceptor which was r.f. heated. The susceptor was coated with SiC and HCl vapor etched before each run. The susceptor face was angled such that the silicon surface lay at an angle of 30° to the vertical downward flow of reaction gases, this geometry having been found to yield the best thickness uniformity across the slice. The gas source was a 3-4% commercial mixture of silane in hydrogen, both this mixture and the dopant gases, arsine and diborane, having been mass spectrometrically analyzed for purity and concentration. The carrier gas, hydrogen, was purified by passing through a palladium diffuser and then passed through the several dilution stages and flow meters associated with the other gases. The system tubing and flow meter hardware were made of stainless steel, the system being continuously flushed with nitrogen when not in use and leak checked prior to each run. A three-way valve was built into the doping train so that by simple switching the doping gas could be rapidly changed to form vapor phase grown junctions for multilayer structures.

3. Substrate Characterization

Epitaxial structural perfection being so dependent on substrate surface characteristics, it was of concern to determine whether epitaxial layers several tenths of a micron thick would develop electrical and structural properties different from thicker layers. For this reason particular attention was paid to silicon polishing and cleaning procedures, and to the defects, such as precipitates, dislocations, etc., present in heavily doped substrates. Both $\langle 100 \rangle$ and $\langle 111 \rangle$ oriented substrates were used ranging in resistivity from 0.0006 Ω-cm to 100 Ω-cm. Techniques involving ellipsometry, electron diffraction, X-ray topography and scanning electron microscopy (SEM) were used to examine the substrates and to correlate substrate defects with subsequent epitaxial flaws.

3.1 Substrate Examination

a. Ellipsometry

This technique was found to be powerful in determining the cleanliness of silicon surfaces [7] and also indicative of residual damage at the surface. Typically, residual film thicknesses on silicon surfaces which were "Syton" polished ranged from 5 Å to 200 Å. These thicker oxide films, which could not be readily removed even by a hydrogen chloride vapor etch at 1200°C, were found to be characteristic of detergent contamination from commercial sources and were removed by adding a peroxide boil to the cleaning procedure. Surfaces examined immediately after the predeposition hydrogen chloride vapor etch, or as freshly deposited epitaxial surfaces, were found to have the thinnest oxide films, typically 4-6 Å due to oxide growth on the surface during the ellipsometer evaluation. Measurement of the refractive index of the substrate showed that this property was sensitive to polishing variables. Thus, ellipsometric determinations of film contaminant and substrate refractive index have been a useful method of quantitatively monitoring substrate surface quality and reproducibility.

b. X-ray Topography

The use of Berg-Barrett reflection topography with a resolution of 5 microns indicated that two types of defects could occur occasionally in the substrates; a) dislocation networks related to saw damage which had not been totally removed by polishing and b) discrete precipitates of doping impurities occurring in heavily doped

[1] Figures in brackets indicate the literature references at the end of this paper.

substrates where the solubility limit of the impurity in silicon was approached. Such defects propagated into the epitaxial layer as dislocations, stacking faults and pyramid structures. It was found that a combination of chemical etching and "Syton" polishing gave adequate surfaces to grow high quality thin epitaxy. Differential X-ray topography of the epitaxy and underlying substrate was used as a nondestructive technique, the Cr K^{α} reflection from the silicon (220) penetrating approximately 1.2 microns into $\langle 111 \rangle$ silicon and the Fe K^{α} reflection from the silicon (331) penetrating 7.0 microns such that epitaxial defects could be related to substrate effects.

c. Electron Diffraction and Microscopy

The possible existence of surface defects formed during processing, eg. SiO_2, SiC, etc., was examined by both these techniques [8]. Low incident angle high energy electron diffraction measurements on samples which had been hydrogen annealed and HCl vapor etched indicated that no SiC particles formed as contrasted to the appearance of β-SiC epitaxial particles on samples which had been chemically cleaned only. SEM surface examination showed the surface to be free of defects within the 200 A resolution of the instrument.

3.2 Surface Cleaning

To minimize the number of stacking faults in thin epitaxy considerable attention was paid to cleaning procedures as outlined above. The most reproducible surfaces were prepared by taking freshly "Syton" polished samples washed free of residual "Syton", giving these a trichloroethylene boil followed by a HNO_3-HF etch. Although such surfaces could be used to grow epitaxy the addition of a 10 minute HCl vapor etch at 1000°C removed any natural oxide formed on the surface after chemical cleaning ensuring defect free epitaxy. Originally a 1200°C anneal and etch were used to remove approximately 1 micron/minute of silicon but to make the temperature compatible with the growth conditions this was reduced to 1000°C. At 1000°C etch rates of 0.01-0.05 microns/minute for HCl concentrations of 0.25-1.0% can be used to obtain controlled etching of buried layers and oxide patterns.

4. Silicon Deposition

4.1 Temperature Dependence

To obtain good control over thickness reproducibility and yet to minimize outdiffusion to approximately 0.1 micron in a 1.0 micron layer growth rates of about 0.3-0.5 microns at 1000°C would be desirable. Using flow rates of 2.1 liters/minute and a silane concentration in hydrogen of 0.14%, the effect of varying the deposition temperature on the growth rate was determined and the results plotted in Fig. 2. These show that over the temperature range of 950-1100°C, under the flow conditions employed, the growth rate is virtually independent of temperature indicating that the reaction mechanism is mass transfer dominated. Above 1100°C the growth rate dependence becomes negative and at temperatures below 950°C the deposition rate also falls off markedly and mixtures of polycrystalline and single crystal silicon are obtained. The flat portion of the curve is of most interest to our work -- to grow the epitaxial layers in conditions where the thickness is relatively unaffected by minor fluctuations in temperature yet achieve good quality epitaxy with low outdiffusion 1000°C was chosen as a standard deposition temperature. Although in concept the choice of a mass transfer dominated reaction mechanism should make epitaxial thickness reproducibility dependent principally on system geometry it has been our experience with four epitaxial silane systems that this has not been a problem for thin layers.

The present results are compared to some silane kinetic data culled from the literature [1-6] and it is clear that the kinetic behavior is determined to an extent by the flow conditions employed by the previous workers, i.e., gas concentrations and flow rates in addition to temperature. The highest deposition rates were obtained by Joyce and Bradley [1] under reduced pressure conditions, the lowest temperature results by Richman and Arlett [6] using helium carrier gas. These latter results are of especial interest for thin layers but our present stations are not sufficiently oxygen leak tight to reproduce this work. Attempts to grow an initial epitaxial layer of 1000 Å of similar doping to the substrate at 1000°C and then reducing the deposition temperature to 800°C gave structurally good epitaxy but the deposition rate dropped by two orders of magnitude.

In Fig. 3 curve A illustrates the fact that a linear relationship exists between deposition time and epitaxial thickness using a growth rate of 0.24 micron/minute at 1000°C. This curve can be extrapolated through zero time and it is this apparent lack of induction period plus linearity with time that enables thin layers to be grown reproducibly. In fact layers as thin as 500 Å have been grown reproducibly and uniformly by using low growth rates, and electron diffraction measurements have shown these to be single crystal material.

4.2 Silane Flow

In Fig. 3 curve B is a plot of the effect of increasing the silane flow on the deposition rate at 1000°C. Again, over the range of flow conditions examined, namely 1.0-10.0 micromoles silane per minute, a linear relationship exists, the highest growth rate at which good epitaxy could be obtained being 1.5 microns/minute. The inherent advantage of this linear relationship when combined with the relative temperature independence is obvious -- the growth rate may be readily adjusted in a linear manner and controlled closely to give excellent thickness reproducibility and uniformity across a slice.

4.3 Impurity Doping

To achieve comparable reproducibility in doping uniformity gaseous sources of impurity dopant were selected for use in this work utilizing dilution stages to premix the gases before they entered the reaction chamber. Arsine and diborane were chosen, phosphine being eliminated because of the fact that the diffusion coefficient for phosphorus at 1000°C is two orders of magnitude higher than for arsenic, which would reduce the ability to grow sharp multilayer junctions. Figure 4 is a plot on a log-log scale of arsenic concentration in the gas phase against resistivity in the doped epitaxial layer. These data were taken by four-point probe measurements of one micron epitaxy grown on P-type substrates and have been confirmed by capacitance-voltage measurements. A linear relationship with a slope of approximately unity exists for both arsine and diborane (not shown). This again is an asset in controlling doping accurately but only for epitaxial doping concentrations above 10^{15} carriers/cm^3. For higher resistivity epitaxy two experimental complications have been encountered. Firstly, variations occur in the intrinsic doping of the commercial silane gas which has been found to range in resistivity from 0.5 Ω-cm to 250 Ω-cm N-type but occasionally P-type is noted. This problem has been improving recently as more attention has been paid to gas purification and packaging. Secondly, autodoping has been a problem in controlling the doping of epitaxial layers in the 10^{13}-10^{15} carriers/cm^3 range especially over heavily doped arsenic substrates although the mechanism is not yet understood. In summary, the deposition capability for the silane epitaxial process at 1000°C can be shown in Table 1.

Table 1. Summary of Epitaxial Silicon Growth on 1 1/4-Inch Substrates

Property	Range	Control
Thickness	0.1-2.0 microns	±5%
Doping	(carriers/cm^3)	
N-type (As)	5×10^{15} -5×10^{18}	±5%
P-type (B)	1×10^{16} -1×10^{19}	±5%

5. Epitaxial Characterization

5.1 Impurity Atom Distribution

There was shown in Fig. 1 the anticipated profile for a one micron layer grown at 1000°C. In Fig. 5, there are plotted three actual measurements of impurity atom concentration as a function of epitaxial depth, measured by the inverse capacitance profile method, superimposed on a calculated outdiffusion curve based on a one micron layer, doped to 4×10^{17} carriers/cm^3, over a substrate of doping concentration 7×10^{19} carriers/cm^3. Three diffusions were performed at 850°C to depths of 0.3, 0.6, and 0.9 microns for the respective capacitance measurements. As may be seen, the measured

impurity levels fall in good agreement on the calculated curve, no autodoping occurring at these levels. It has been our experience that over the thickness range examined, 0.2-2.0 microns, similar correlations exist between calculated and measured impurity profiles for moderate to heavily doped epitaxy.

5.2 Multilayer Structures

Multilayer structures have been used for both device fabrication purposes and for the capacitance-voltage characterization of epitaxy by replacing diffused junctions by vapor grown junctions 0.25 micron thick. Comparison of such epitaxial and diffused junctions showed that good junction properties exist, e.g., low junction leakage, and that the fast epitaxial growth times yield more realistic measurements of the as grown impurity distributions. Figure 6 is a plot of a simple N^+/N epitaxial double layer grown on a P-type substrate. The design required a 1.0 micron layer doped to 2×10^{17} carriers/cm^3 over a 2.0 micron layer doped to 10^{15} carriers/cm^3 over a 10^{19} carriers/cm^3 boron doped substrate and both the calculated and measured profiles are shown. Angle lap and stain measurements gave thicknesses of 1.1 and 2.1 micron respectively, whereas breakdown in the capacitance measurement occurred at 1.8 microns depletion depth. Various methods of fabricating multilayers have been explored by varying flush times on changing the dopant gas concentration or type from zero to several minutes. We have also attempted to sharpen the junction profile by dropping the pedestal temperature during the flush-out period. In practice, the most reproducible results were obtained at constant pedestal temperature and by flushing out the reaction system for a two-minute period between setting the dopant flow levels. Multilayer structures as shallow as 0.25 micron thick in each layer have been grown but their characterization requires the use of a spreading resistance probe whose accuracy does not approach that of the profiler.

5.3 Epitaxial Lifetime

Alternative methods of discussing epitaxial perfection and reproducibility were examined in terms of mobilities and lifetimes. Mobility proved to be rather insensitive to epitaxial variation but the measurement of minority carrier lifetime by a pulsed MOS technique [9] has proved to be a viable method. Figure 7 is a composite plot of lifetimes measured as a function of deposition temperature and epitaxial doping with the spread in the measurements shown as part of the plot. Curve A shows that the lifetimes vary only slightly from somewhat under a microsecond when grown at 950°C to several microseconds when grown at 1150°C. These measurements were made on one micron thick layers, 1 Ω-cm. N-type material, with the growth rate in all cases being approximately 0.25 micron/minute. Curve B indicates that the lifetime is inversely proportional to epitaxial doping, as would be expected from bulk silicon values, and begins to shorten appreciably for heavily doped layers. These results would indicate that the epitaxial material is of high structural perfection, even at low depostion temperatures, but results for higher resistivity epitaxy have proved to be more variable, perhaps due to autodoping effects.

6. Epitaxial Characterization Methods

It has been pointed out earlier that the measurement of submicron layers has required the extension of existing methods and the exploration of other approaches. In some cases where very thin layers could not be characterized for a particular property thicker layers had to be grown as substitute specimens. A short discussion of the problems encountered and some of the solutions is given next.

6.1 Epitaxial Perfection

A significant result of this work has been to show that even on heavily doped substrates and for layers as thin as 0.1 micron high quality defect free epitaxy can be obtained. Examination of one micron epitaxy by phase contrast microscopy and by etching procedures has shown that the cleaning and substrate examination outlined previously can ensure epitaxy free of stacking faults and pyramids. Such defects as do occur appear at the edge of the wafers and appear related to stresses developed there by ingot grinding and etching to meet diameter tolerances. When defects do occur near the center of the epitaxial layer these have been shown by differential X-ray topography to be related to substrate imperfection. SEM examination of epitaxial surfaces has shown very few defects existing below the resolution limit of the phase contrast microscope.

6.2 Thickness Measurements

Apart from data generated by capacitance-voltage measurements on test diodes, three physical methods were used in the 0.2-2.0 micron thickness range, namely infrared reflectance, angle lapping and stain and disc lapping and stain. For measurements of thickness less than 0.2 microns both the talysurf and SEM have been used.

a. Infrared Reflectance

Curves computed by P. A. Schumann [10] to cover the range 0.5-5.0 microns thickness for substrates doped more heavily than 10^{16} carriers/cm^3 were extended down to 0.2 microns for both N and P epitaxial layers. It has been found that with slow scanning through the requisite spectral reflectance maxima and minima a precision of ±0.02 microns may be obtained by this method using an I.R. spectrophotometer covering the wavelength range 2-50 microns and with an accuracy of ±0.5 cm^{-1}. It is probable that this method could be extended to even thinner layers were sufficient precision built into the computational data and very slow scans employed. This method is limited at present to single layer epitaxy and where multilayer structures were measured lap and stain methods have been employed.

b. Lap and Stain Methods

Standard angle lap and stain techniques were initially employed to examine multilayers and single layers where the epitaxial doping was greater than that of the substrate or the substrate doping was less than 10^{18} carriers/cm^3. For thin layers the need to lap with extremely shallow angles proved too irreproducible and the interference fringe method has been used. The best precision attainable was ±0.15 microns. For greater precision and improved reproducibility modification was made to the cylinder lap method [11,12] by replacing the cylinder with a disc of precisely known-radius. This disc lap is considerably faster and more reproducible than cylinder lapping and has been found to give a precision of ±0.06 micron with standard silicon stains. This precision could probably be improved as sharper staining techniques are devised, especially for use with multilayer structures. An upper limit of 5 micron thickness of epitaxy is put on this method because lapping time becomes inordinately long as thickness increases. A summary of the comparative merits of these thickness measuring techniques is given in Table 2.

Table 2. Comparison of Thickness Measurement Techniques

Method	Precision	Application
I. R. reflectance	±0.02 microns	single layer epitaxy substrate doping $> 10^{18}/cm^3$.
Angle lap	±0.15	single, multilayer
Disc lap	±0.06	single, multilayer
Talysurf	±0.005	single layer steps

6.3 Impurity Doping Measurements

The determination of impurity atom distributions in thin epitaxial layers has been the most difficult part of this program and no unique solution has been developed. The major problems encountered have been the possibility of punch through by mechanical probes in thin layers, the built-in depletion layer in lightly doped thin layers and the inability to profile multilayer epitaxy. Table 3 lists the various techniques which have been utilized illustrating the relative advantages of each measurement.

Table 3. Comparison of Epitaxial Resistivity Measurement Techniques

Measurement Type	Resistivity Range (carriers/cm^3)	Conductivity Type	Precision
Schottky barrier, C-V	10^{13}-10^{16}	N/N	10%
Diffused junction, C-V	10^{13}-10^{18}	N/N or P/P	10%
Epitaxial junction, C-V	10^{13}-10^{17}	N/N or P/P	10%
Four-point probe	10^{12}-10^{20}	N/P or P/N	2%
Spreading resistance probe	10^{14}-10^{20}	All	?
Capacitance profiler (CIP)	10^{13}-10^{18}	N/N or P/P	10%
MOS capacitance	10^{12}-10^{16}	All	50%

Several of these methods have been used to determine the impurity profile as grown without further changes induced by processing such as diffusion or oxide growth at high temperatures. Both the Schottky diode and the epitaxial formation of a 0.3 micron junction have been found particularly useful for this purpose, whereas attempts to use step etching combined with MOS capacitance methods proved abortive due to the poor resolution and precision of the method. 20 mil Schottky diodes, prepared by evaporation of gold through a mask onto epitaxial surfaces, were measured by a capacitance inverse profiler (CIP) to an upper epitaxial doping level of 10^{16} carriers/cm^3 before diode leakage effects became too great. In the case of junction diode measurements, whether by CIP or standard capacitance-voltage measurements, epitaxially formed junctions were necessary for evaluating 0.5 micron and less epitaxy where a diffused junction would destroy a large part of the layer of interest. It has also been found for submicron layers that mesa diodes gave consistently more reproducible results than planar diodes with the additional advantage again that the planar diodes would use up a significant fraction of the epitaxial layer.

The four point probe method of measuring resistivity on substrates of opposite type conductivity has been useful at resistivities less than 1 Ω-cm where autodoping effects are not significant. At higher resistivities the Schottky diode method is preferable. For multilayer evaluation the CIP measurement is useful for structures, such as Fig. 6, where the voltage can sweep through the N-N^+ region. In many instances this is not applicable and the use of the spreading resistance probe has been investigated. For probing across P-N junctions and for determining doping profiles of several submicron layers in a multilayer structure this appears to be the only feasible method. It has been our experience that the reproducibility is reasonably good and that even uncorrected profiles are of considerable use when combined with disc lap measurements of thickness, but considerable refinement of the technique is required.

7. Summary

It is safe to state that in the study of submicron single and multilayer epitaxial silicon layers the silane process has enabled structures to be designed and fabricated which could not previously be achieved by other methods on a reproducible basis. It is also true that the problems inherent in measuring such thin layers are only partially solved, particularly for multilayer epitaxy, and that investigation of optical and electrical methods should prove valuable in the near future.

8. References

[1] B. A. Joyce and R. R. Bradley, J. Electrochemical Soc., 110, 1235, (1963).

[2] S. B. Bhola and A. Mayer, R.C.A. Review, 24, 511, (1963).

[3] S. E. Mayer and D. E. Shea, J. Electrochemical Soc., 111, 550, (1964).

[4] Y. Nishi and M. Watanabe, Japanese J. Applied Phys., 6, 550, (1967).

[5] D. C. Gupta and R. Yee, J. Electrochemical Society, 116, 1561 (1969).

[6] R. Richman and R. H. Arlett, Electrochemical Society Meeting, NYC, May 4-9, 1969.

[7] A. C. Adams and R. H. Kaiser, Electrochemical Society Meeting, Los Angeles, May 10-15, 1970.

[8] R. C. Henderson, W. J. Polito and J. Simpson, Applied Physics Letters, 16, 15 (1970).

[9] F. P. Heiman, IEEE Trans. Electrical Devices, ED-14, 781 (1967).

[10] P. A. Schumann, J. Electrochemical Soc., 116, 409 (1969).

[11] W. W. Hopp and W. Shockley, Bull. Am. Phys. Soc., II 1, 382 (1956).

[12] A. Goetzberger and B. MacDonald, J. Electrochemical Soc., 109, 141, (1962).

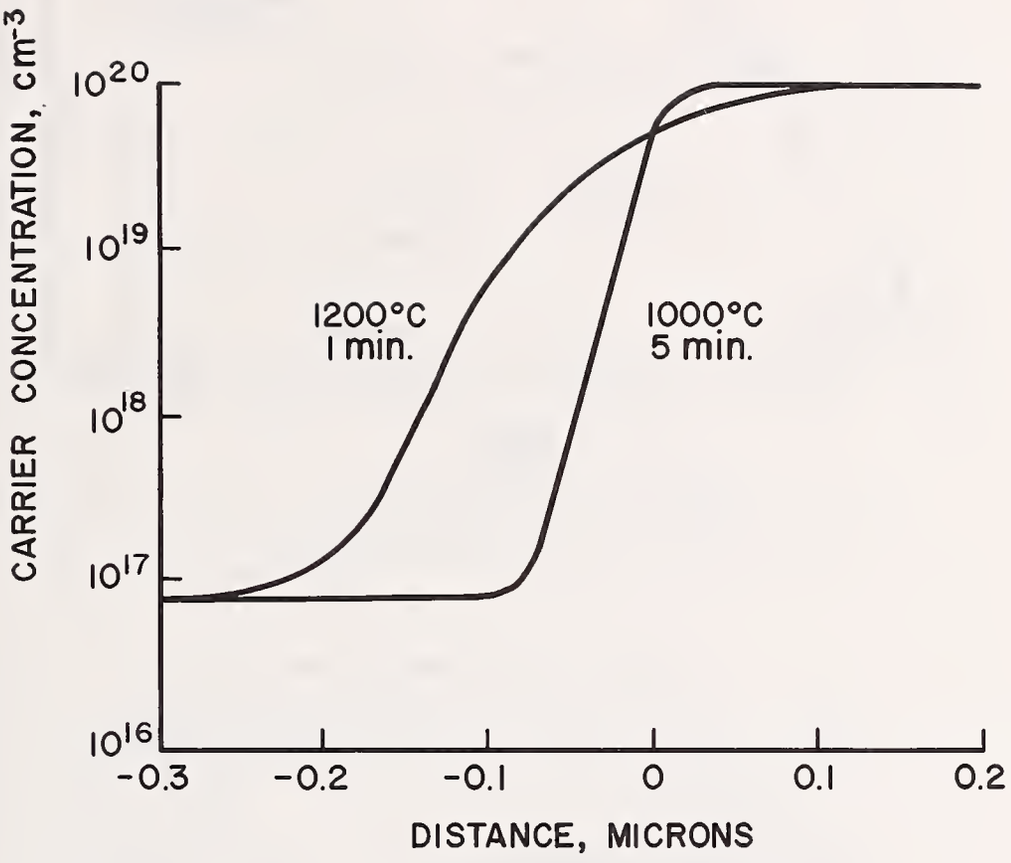


Fig. 1. Calculated Outdiffusion Profiles for One Micron Epitaxy

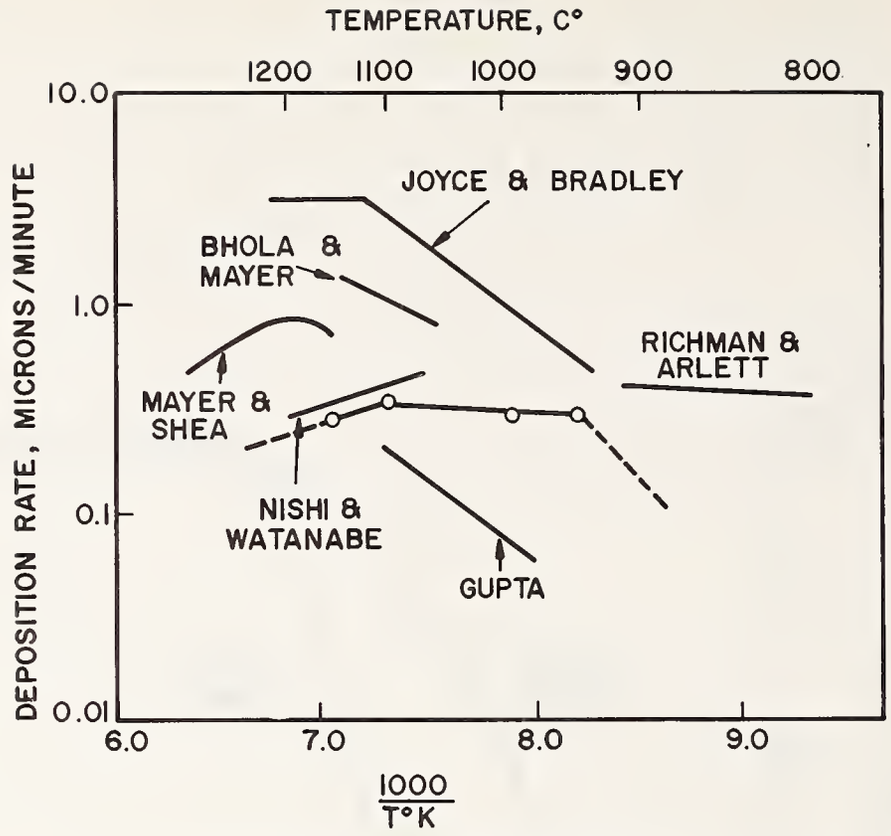


Fig. 2. Comparison of Temperature Dependence of Growth Rate of Epitaxial Silicon from Silane

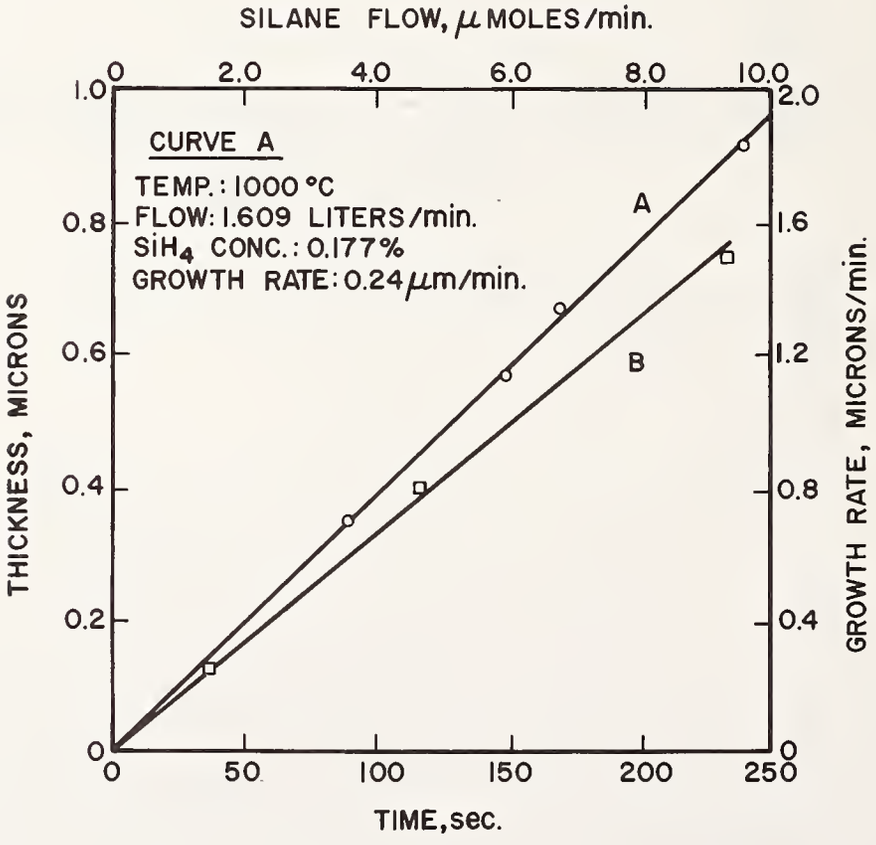


Fig. 3. Growth Rate Dependence of Silane Epitaxy
Curve A: Thickness vs. Time
Curve B: Rate vs. Silane Flow

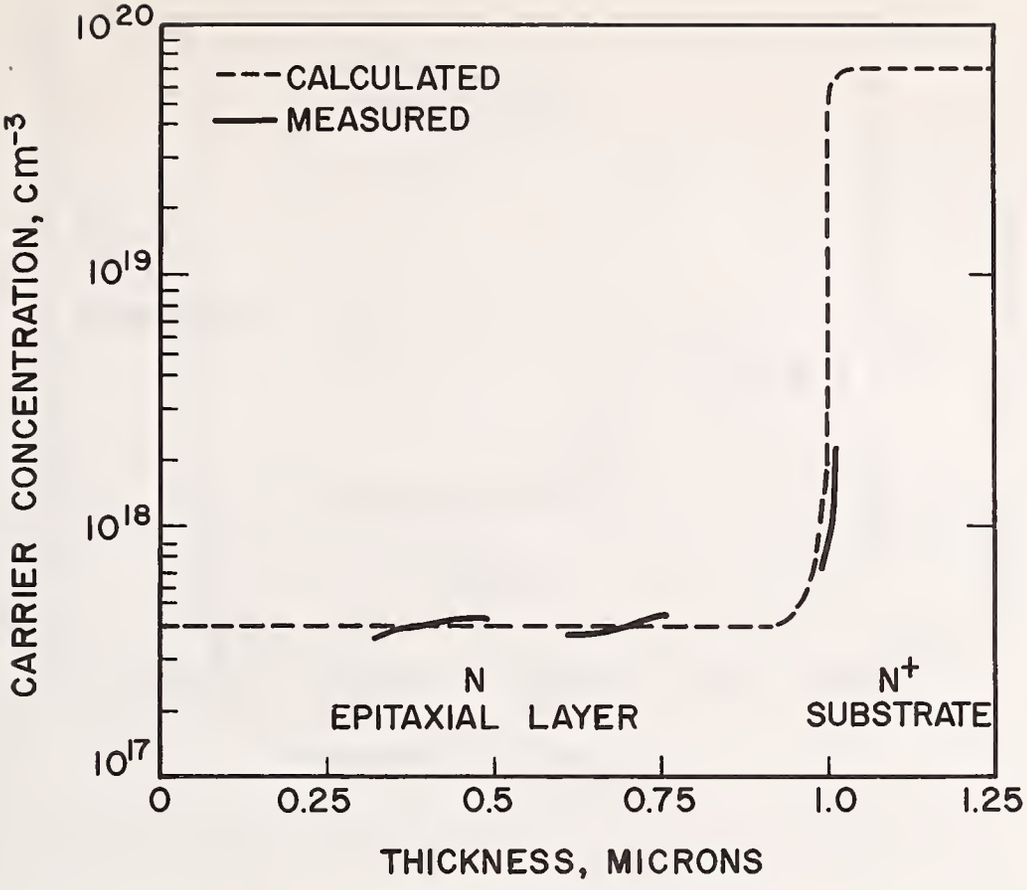


Fig. 4. Relationship of Epitaxial resistivity on Arsine Concentration

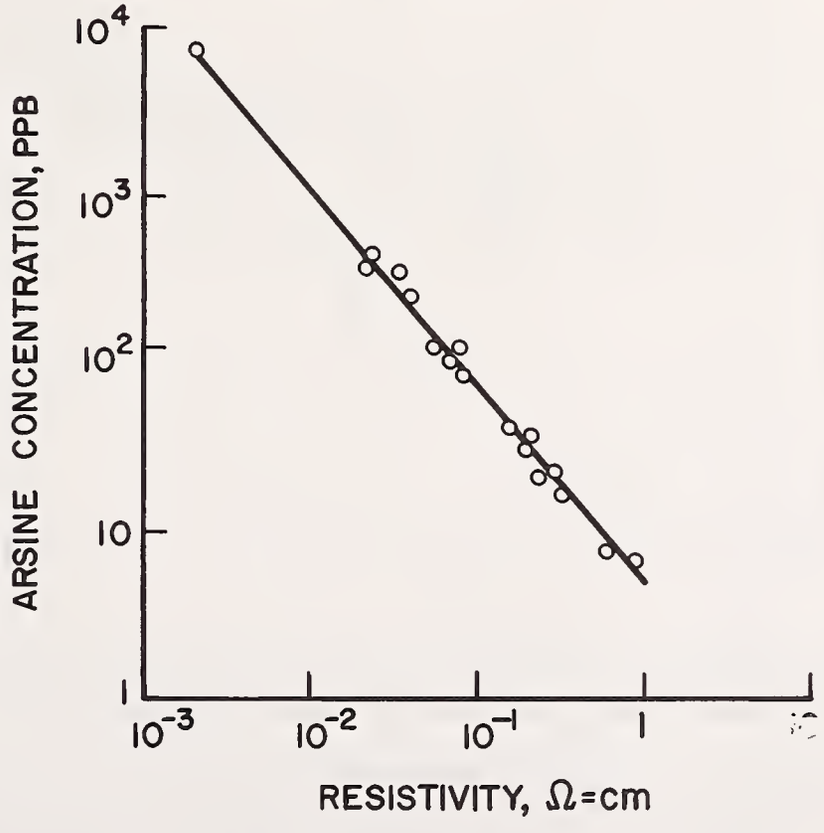


Fig. 5. Doping Profile for One Micron Epitaxial Layer

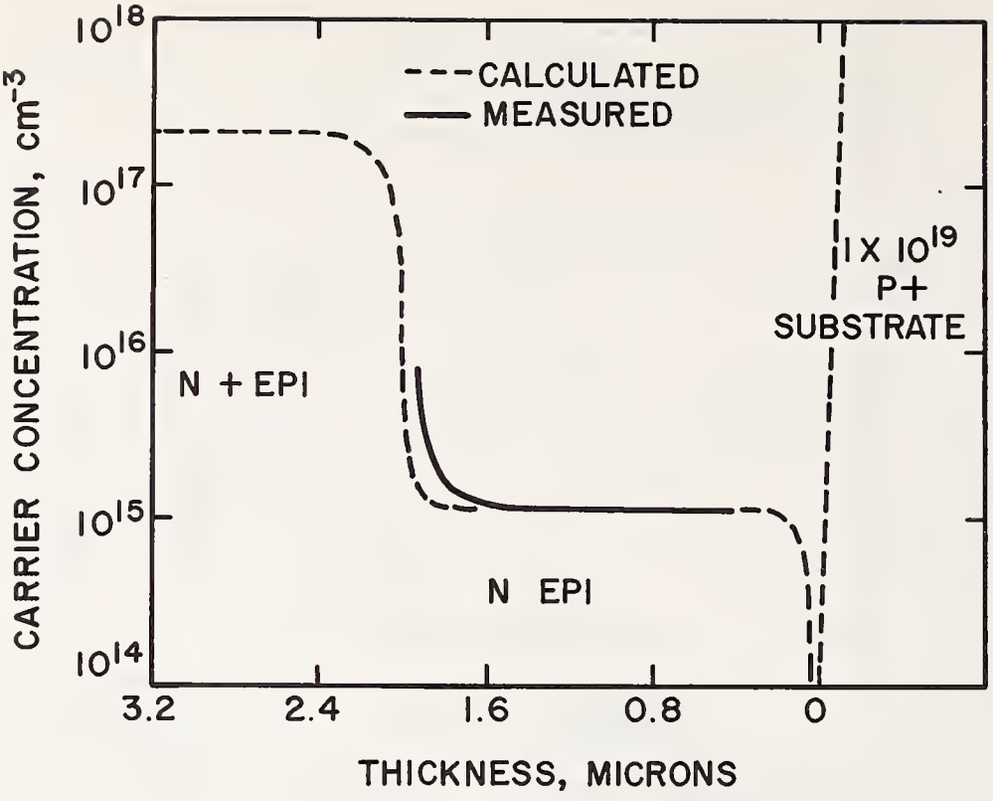


Fig. 6. Doping Profile for Double Layer Epitaxy.

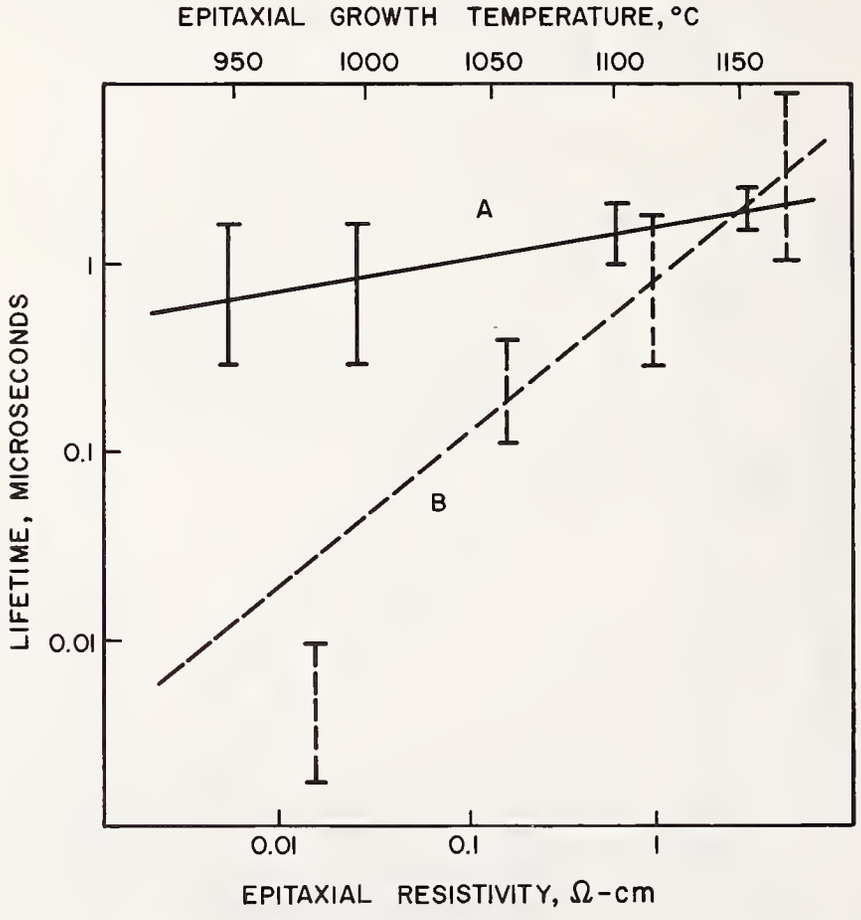


Fig. 7. Comparison of Minority Carrier Lifetime in Epitaxial Silicon
Line A - As a Function of Growth Temperature
Line B - As a Function of Epitaxial Resistivity

Techniques For Determining Surface Concentration of Diffusants

J. C. Irvin

Bell Telephone Laboratories
Murray Hill, N. J. 07974

Eight techniques for the determination of the surface concentration, C_s, of diffused layers in silicon are reviewed. The techniques are 1) use of published curves relating junction depth and sheet resistance to C_s; 2) use of published curves relating C_s to measured Hall effect; 3) differential sheet resistance due to removal of a surface lamina; 4) differential radioactive analysis; 5) differential Hall effect; 6) the spreading resistance probe; 7) plasma resonance; and 8) ratio of maximum and minimum capacitances of a MIS capacitor. The limitations and range of application of each method are discussed.

Key Words: Diffusion profile, Hall effect, Irvin's curves, MIS capacitance, plasma resonance, radioactive analysis, sheet resistance, silicon, spreading resistance, surface concentration.

1. Introduction

An accurate knowledge of the surface concentration of diffusants is required in the development of new diffusion methods and especially in investigations of the thermodynamics of the reactions that take place at the surface during diffusion. Numerous methods are available for the determination of the donor or acceptor concentration at the surface of a diffused layer. Most of these methods are destructive and the more reliable ones are not simply executed. Fortunately, for the monitoring of device processing it is not necessary to measure surface concentrations. The sheet resistivity and junction depth, both rapidly and easily measurable parameters, are sufficient to assure that routine diffusions are within specifications.

In this paper, eight of the more common methods of measuring surface concentration will be reviewed and the advantages, limitations, and range of application of each method will be discussed. These eight techniques may be divided roughly into three groups:

1. Methods valid only when certain distributions of the diffused impurities prevail; these are the sheet resistance-junction depth method using Irvin's curves [1] and the Hall effect method using Tufte's curves [2].[1]

2. Incremental techniques requiring removal of a lamina of material and measurement of the resultant change in sheet resistance, radioactivity, or Hall effect.

3. Techniques which assume that the penetration of the observed phenomenon is limited to very shallow depths; these include the spreading resistance, plasma resonance, and MIS methods.

In principle the members of the third group could be transferred to the first group; however, this would require recourse to certain corrections which generally have not been computed and published. The eight measurement methods will be discussed in the following sections in the above-cited order.

[1] Figures in brackets indicate the literature references at the end of this paper.

2. Description of Methods

2.1 The Sheet Resistance-Junction Depth Method

As shown by Backenstoss [3], Irvin [1], and others the product of the sheet resistance, ρ_S, (ohms/square) and junction depth, x_j, (cm) of a diffused layer is uniquely determined by the surface concentration, C_S, and background concentration, C_B, of the slice, provided the distribution, $C(x) = f(x)$, of the diffused impurities is known. (To measure ρ_S, the original slice and diffused layer must be of opposite conductivity types). These relationships have been computed and published in graphical form by Irvin [1] for a wide range of background and surface concentrations, for complementary error function (erfc) and for Gaussian profiles of donors and acceptors in silicon. An example of these curves for an erfc distribution is shown in Fig. 1. Thus a measurement of ρ_S and x_j and reference to the mentioned graphs enable simple determination of C_S, provided the diffused layer in question is known to be either erfc or Gaussian in profile.

This provision is a considerable limitation. A diffusion process obeying Fick's law with constant concentration at a stationary surface will yield an erfc distribution. A constant total number of impurities per unit area yields a Gaussian profile. In actuality, many diffusions are non-Fickian due to a concentration-dependent diffusivity and the built-in electric field of the impurity gradient. The constant concentration boundary condition may not prevail due to lack of equilibrium at the surface or due to movement of the surface from oxidation or evaporation. The calculated C_S-ρ_S-x_j relationships [1] are based on observed resistivity-impurity concentration data and thus approximately take into account the deionization and precipitation of impurities at high concentrations. However, these data are obtained primarily from bulk material. The lattice strain associated with the concentration gradient in a diffused layer can produce a high dislocation density near the surface [4-7] leading to greater precipitation than in the bulk and altering the diffusion mechanism and also the carrier mobility near the surface. (An electron mobility decrease from 40 to 2 cm^2/volt-sec in the top 0.1 micron of a diffused layer has been suggested by some measurements [8]). The misfit ratios in Si of the Pauli tetrahedral ionic radii are 0.746 for B, 0.932 for P, and 1.00 for As [6]. Thus fewer dislocations are expected and observed in an As-diffused layer [6]. However, the diffusivity of As is still found to be strongly concentration dependent due to electronic interactions between the diffusing impurity and the host lattice [9].

All the above causes for non-Fickian diffusion and hence departure from expected erfc or Gaussian profiles are aggravated by high surface concentration. As a rule of thumb, it would appear that donor or acceptor diffusions in Si with surface concentrations over 1×10^{20} cm^{-3} are not apt to conform very accurately to erfc or Gaussian distributions. Hence, Irvin's C_S-ρ_S-x_j curves [1] should not be employed for the measurement of C_S values over 10^{20} cm^{-3}. This can only be a broad generalization, since the particular diffusion process chosen determines the boundary conditions and hence the profile. Good erfc profiles have been reported with C_S as high as 8×10^{20} cm^{-3} (of phosphorus), [10] while significant departures from erfc have been reported with C_S as low as 1×10^{19} cm^{-3} [11].

2.2 The Hall Effect Employing Tufte's Curves

The Hall constant of a diffused layer is [2]

$$R_H = e x_j \int_0^{x_j} n(x)\mu^2 dx \Big/ \left(|e| \int_0^{x_j} n(x)\mu \, dx \right)^2 \qquad (1)$$

where e is electronic charge, n(x) is the carrier density at depth x and depends on the distribution, and μ is the mobility (a function of n and thus of x also). Due to form of this expression, R_H is relatively insensitive to the mobility and its concentration dependence. However R_H is sensitive to the form of the distribution, n(x), and in particular to n(0), i.e., the surface carrier concentration. If an

effective mobility is used, $\mu' \equiv 1/\rho e N_I$, where ρ is the resistivity of silicon with an impurity density N_I (as opposed to carrier density), then the concentration in the expression for R_H may be reinterpreted as impurity density rather than carrier density. R_H then becomes a function of the surface impurity concentration, C_S, and the background concentration, C_B, for a given distribution. Tufte[2] has calculated this relationship for p-type and n-type erfc diffusions in silicon. His n-type curve is shown in Fig. 2. Thus a measurement of the Hall constant of a diffused layer and the junction depth in the material of known background concentration (of opposite type) is sufficient to determine C_S through reference to Tufte's curves [2], provided the profile is an erfc distribution.

This Hall effect method suffers the same disadvantage that besets the ρ_S-x_j technique, namely the failure of nature to conform to the assumed erfc distribution under some circumstances. In addition there is the greater complexity of measuring a Hall constant, instead of sheet resistance. Also Tufte's curves are based on older effective mobility data than are Irvin's ρ_S-x_j-C_S curves. The more recent data show a significant error in the older values for p-type material. The result of these limitations is that this technique is best confined to n-type diffusions, performed under conditions expected to yield an erfc profile, and in particular, therefore, with surface concentrations not over 10^{20} cm^{-3}. The probable departures from erfc profile at higher surface concentration are such that the application of Tufte's Hall effect curves leads to estimates of C_S which are too high, as is also the case with Irvin's ρ_S-x_j-C_S curves.

2.3 Incremental Techniques: Sheet Resistance

The incremental sheet resistance method of profiling diffused layers is considered in detail in another paper in this issue [12]. It is one of several incremental techniques frequently used to measure impurity or carrier distributions resulting from diffusions or epitaxial deposition [13,14]. Other properties which may be used in a similar manner are the Hall constant and the radioactivity. In at least one case [10] colorimetric chemical analysis of successively dissolved laminas has also been used for profiling. In any of these incremental profiling methods, the results obtained from the first increment are a measure of the surface concentration. Thus in the case of the incremental sheet resistance, three measurements suffice to determine the surface concentration: the initial sheet resistance, ρ_{S1}, the sheet resistance after removal of a thin lamina, ρ_{S2}, and the thickness t of the lamina removed. Since $1/\rho_S = \int_0^{x_j} \mu(x,n)en(x)dx$, the carrier density is given by $n(x) = (1/\mu e)d(1/\rho_S)/dx$, or in terms of the parameters measured, $n(0) = (1/\mu et)\,(1/\rho_{S2} - 1/\rho_{S1})$. This expression gives the carrier density at the surface and requires a value of of carrier mobility μ, appropriate to the surface lamina. Due to the varying degree of precipitation and dislocation at the surface, mobilities there are very uncertain, as noted earlier [8]. This is the largest source of error in the technique. In order to measure ρ_S of the diffused layer it must be electrically isolated from the bulk, as in the previous two methods; i.e., layer and original slice must be of opposite conductivity types.

An alternate and sometimes more reliable calculation, which yields impurity rather than carrier concentration, is obtained from the relation $d(1/\rho_S)/dx = \mu en = 1/\rho$. In terms of the measured parameters, then $\rho(0) = t/(1/\rho_{S1} - 1/\rho_{S2})$ where $\rho(0)$ is the resistivity characteristic of the surface lamina. Using the published curve by Irvin [1] relating resistivity to impurity concentration in Si, a value for the impurity concentration at the surface is thus obtained. This ρ-N_I curve is shown in Fig. 3. The high concentration end of the n-type curve of Fig. 3 is based to a large degree on data from phosphorus diffused layers Hence the effects of precipitation and dislocations associated with diffused layers are already partially included in that curve. The p-type ρ-N curve, on the other hand, is derived entirely from bulk measurements. Therefore this manner of calculating C_S from incremental sheet resistance measurements, when applied to n-type diffusions, is relatively free of the errors arising from the uncertainty in mobilities. Calculations for p-type diffusions, while not free of this hazard, at least give direct measures of impurity concentration rather than carrier concentration. When impurity concentrations are desired, whether donor or acceptor, this calculation is preferable since it obviates conversions involving the degree of ionization.

Another source of error in the incremental-ρ_S method is one inherent in any incremental method: for small increments, the difference between initial and final values of the observable is small and the corresponding error consequently large

The relative change in ρ_s, due to removal of the lamina, may be shown to be $\Delta\rho_s/\rho_s = t\rho_s/\rho$. If t is a given fraction, γ, of the junction depth, $t = \gamma x_j$, then $\Delta\rho_s/\rho_s = \gamma x_j \rho_s/\rho$. Reference to Irvin's ρ_s-x_j-C_s curves [1] shows that the product $\rho_s x_j$ is approximately inversely proportional to surface concentration, C_s, for either an erfc or Gaussian profile. The resistivity, ρ, of the surface lamina removed is also approximately inversely proportional to C_s. Thus $\Delta\rho_s/\rho_s = \gamma K_1 C_s/K_2 C_s = \gamma K_3$, where K_1, K_2, and K_3 are proportionality constants, fixed for a given type of profile and background concentration. For n-type erfc diffusions into 1 Ω-cm p-type material, $K_3 \approx 3$. Thus for $\gamma = 0.1$, $\Delta\rho_s/\rho_s \approx 30\%$. Since the precision in measuring ρ_s is about 1%, this 30% change can be detected with fair accuracy. If, however, $\gamma = 0.01$, then $\Delta\rho_s/\rho_s \approx 3\%$, the measurement of which would be subject to considerable error. For p-type erfc diffusions in 1 Ω-cm n-type material, $K_3 \approx 6$, affording somewhat greater accuracy in the measurement of $\Delta\rho_s$ in p-type diffusions as compared to n-type layers.

These values of K_3 are based on an assumed erfc profile. For surface concentrations over 10^{20} cm^{-3} as pointed out before, the erfc profile is often not obtained. The region adjacent to the surface is apt to be characterized by a high dislocation density, a large degree of precipitation, unusually low carrier mobility, and a carrier density which is constant with depth over a large fraction of the total junction depth. Under these circumstances, K_3 will be smaller than for an erfc profile and the error in $\Delta\rho_s$, correspondingly larger. Analysis of published profiling data [13,14] suggests K_3 = 1 to 2.

The methods used for removing lamina from the surface of a silicon slice are either a slow etch (such as 50 HNO_3:1 HF) or oxidation and subsequent dissolution of the oxide. In the latter case the oxidation is usually anodic. Due to the nature of the surface (strain, dislocation and precipitation), etch rates are not reproducible. The lamina is typically 0.1 micron or less in thickness and hence not accurately measurable by most techniques. Therefore oxidation techniques are preferred since this provides some "leverage" (1000 Å SiO_2/360 Å Si) [16,17] and the oxide thicknesses are more accurately measured.

In view of the errors in both $\Delta\rho_s$ and t which increase as t becomes smaller, it is desirable to avoid exceptionally thin laminas. However, (for an erfc profile) the concentration falls 30% at $x = 0.1x_j$ for a typical diffusion ($C_s/C_B = 10^4$); hence a thick lamina gives a rather coarse average of the actual concentration at any point. The best solution to these many complications is to profile a considerable portion of the diffused layer, even if only C_s is desired. Smoothing of the resultant data will help to cover the aforementioned sins.

2.4 Incremental Techniques: Radioactive Analysis

Radioactivity techniques are frequently used for profiling diffused layers in silicon [13,17,18]. Either radioactive material may be employed for the diffusion source, or more commonly, the slice may be neutron activated after diffusion. Using beta-ray or gamma-ray spectrometry, a particular element can be monitored with great selectivity and sensitivity. The decrease in activity of the remaining slice and the activity of the solution are both measures of the desired impurity concentration in a dissolved lamina. Obviously, the surface concentration of the diffused layer is obtained from the top lamina or by extrapolation from several laminas. Radiochemical techniques yield a direct measure of the total concentration of the impurity in question. Consequently, the results are not encumbered by a complex and uncertain translation from resistivity to carrier density to impurity density. However, the total impurity concentration will of course include precipitated atoms as well as those in solution. It is usually only the latter number which is desired for the investigation of thermodynamic and diffusion processes or in predicting electrical characteristics.

The sensitivity of thermal neutron activation analysis [19,20] is typically 10^{-9} gram for most of the common donors or acceptors in Si, using reasonable neutron fluxes (10^{13} n/cm^2-sec for 1 hour). The accuracy of measuring an impurity concentration in a 1000Å thick lamina is determined primarily by the standard, which is irradiated simultaneously, and is 2-3%. Boron, however, is quite recalcitrant radiochemically. It is undetectable by slow neutron activation analysis and has an inconveniently short half life with fast neutron activation. It is better analysed by charged particle activation. The common donors and acceptors which are feasibly analyzed by thermal neutron activation include P, Ga, As, In, and Sb. Al, with a half life of 2.3 minutes, requires much celerity. Thermal neutron activation analysis is also applicable to a number of other impurities which have important effects in

Si but which are electrically inactive (except as traps) and thus otherwise difficult to measure quantitatively. These include Au, Co, Cu, Fe, Ni, and Sn, the last three less sensitively than the others. The remarks made in the last section regarding the accuracy of measuring the lamina thickness, t, and the averaging of the distribution over that depth apply equally here. Using a 0.1 micron thick lamina, surface concentrations as low as $10^{16}/cm^3$, typically, are measurable by activation analysis with 10% accuracy.

Except for the few elements to which it is not amenable (B, Li, and O among them) thermal neutron activation analysis is certainly one of the most reliable means of determining the surface concentration of impurities, even the electrically inactive ones. Where greater sensitivity is required, larger fluxes, larger samples, or analysis at the neutron generator are possible. In any case, however, radioactive determinations are an elaborate undertaking, and do not distinguish between ionized, unionized, and precipitated impurities.

2.5 Incremental Techniques: Hall Effect

The incremental Hall effect method of measuring surface concentration, or of profiling an entire diffused layer, offers definite advantages over the other two incremental methods discussed here. It measures directly the number of mobile carriers in the removed lamina, which is also the number of ionized donors or acceptors. It does not include those impurities which are not ionized, but still in solution in the lattice, nor those that have precipitated out of the lattice. No assumption regarding carrier mobility is required in applying the incremental Hall effect, nor any assumption regarding the shape of the distribution. This method is more easily performed than a radioactive analysis, though not as simply as the incremental sheet-resistance method. However, some precautions are in order.

The surface Hall coefficient, R_S, [21,22] is given by the expression $R_S = R_H/\bar{x}_j$, where R_H was given in eq 1. Now $R_S = 1/e\overline{N_S}$, where $\overline{N_S}$ is a weighted average of the area carrier density in cm^{-2}. $\overline{N_S}$ is not the actual total number of carriers per unit area, unless the mobility is constant throughout the layer. Hence, the difference in successive values of $\overline{N_S}$ is not a reliable measure of the carriers contained in a removed lamina, especially at the surface of diffused layers where mobility is effected by rapidly varying impurity density as well as other scattering processes. To allow for mobility variation, the sheet conductance $\sigma_S = 1/\rho_S$ must also be measured. Then, as shown by Baron, Shifrin, and Marsh [23],

$$C_s = \tfrac{1}{2}\left(\frac{\mu_H}{\mu_C}\right)\left(\frac{\Delta\sigma_s}{t}\right)^2 \Big/ \frac{\Delta(R_s\sigma_s^2)}{t} \qquad (2)$$

where $\Delta\sigma_S$ is the change in σ_S resulting from the removal of the lamina of thickness t, and similarly for $\Delta(R_S\sigma_S^2)$. (μ_H/μ_C) is the ratio of Hall and conductivity mobilities. Wolfstirn [24] has found this ratio to be 1.25 for electrons and 0.73 for holes in highly doped silicon. R_S is obtained from the measured Hall voltage V, sampling current I, and magnetic field B: $R_S = V/IB$.

The Hall constant, like the sheet resistance (or its inverse, sheet conductance) can be measured with an accuracy of 1-2%. The problems associated with very thin lamina, discussed earlier, both with regard to the effect on the accuracy of differences as well as the accuracy of measuring t, apply here also.

2.6 The Spreading Resistance Technique

The spreading resistance method [25, 26, 27,] of profiling diffused or epitaxial structures is discussed in detail in another paper of this volume [25]. Therefore its application to the measurement of surface concentration will be treated only very briefly in this review. The spreading resistance at the surface of Si wafer can be measured with either a two-point or three-point probe, the latter being preferred. The measurement is based on the principle that the resistance of a circular contact of radius r on a semi-infinite slab of uniform resistivity ρ is $R = (\rho/4r)C_1$. C_1 is a correction factor dependent on the nature of the contact, the slice surface preparation and its resistivity, as well as geometry (probe radius, probe spacing, and slice thickness). For a given set of probes, C_1 is determined empirically by means of a set of slices of known resistivity. Using high quality apparatus [28] for

controlling probe spacing and pressure and velocity of contact, the resistivity of bulk material may be measured with ±1% reproducibility by a skilled operator. (The ρ-N curve [1] which translates the measurements into concentration values, however, is only accurate to 10%).

A typical radius of contact is one micron. If the resistivity of the slice is not uniform, then the measured resistivity will be some average of the resistivity over a depth of several microns. Hence the spreading resistance probe does not readily yield good estimates of surface concentration unless the junction is very deep indeed ($x_j > 25\mu$ for an error < 40%). If the form of the impurity distribution is known, as well as junction depth and background concentration, then in principle a correction factor can be computed. However, such factors have been published for only a very limited number of cases [27]. If the slice is angle lapped and the entire diffused layer is profiled, then an analysis [27] may be employed which involves successive corrections to these data. The resulting profile will of course include the surface concentration, limited however in accuracy by the same uncertainties regarding carrier mobility when $C_s > 10^{20}$ cm^{-3} as have been mentioned in previous sections.

2.7 The Plasma Resonance Technique

An ideal technique for measuring surface concentration would 1) be nondestructive i.e., require neither sectioning nor angle lapping, 2) require no electrical contacts, since probes also produce damage and poor contacts adversely affect some measurements, 3) provide high accuracy, and 4) be useful on small areas. With some reservations about junction thickness, and within the concentration range fixed by available instrumentation and standards (5×10^{18} to 5×10^{20} cm^{-3}), the plasma resonance technique possesses all these features.[29-34]

At wavelengths above the absorption edge, the dielectric constant of a semiconductor is a function of free carrier density and at the plasma resonance frequency the infrared reflectivity (a function of dielectric constant) goes through a minimum. In the case of negligible index of absorption, which case prevails for high carrier mobility, the wavelength corresponding to minimum reflectivity λ_m is simply related to the inverse square root of the carrier density n. For low mobility carriers, such as in p-type Si or highly-doped n-type, the relation is more complicated and inverse fourth-root dependence appears. In applying the plasma resonance technique to the measurement of carrier concentration, it is most convenient to refer to published curves [30,32] which are statistical fits to observed λ_m - N_I values from bulk material. Such curves are shown in Fig. 4. Note that these curves have been calibrated against N_I, total impurity concentration, though the physical mechanism involves only carrier concentration, n.

In silicon, λ_m falls at about 2.5μ for $N_I = 5\times10^{20}$ cm^{-3}. At this and higher concentrations, theory and observation depart significantly (due to increasing effective mass or other factors), and bulk material is not available for empirical calibrations. Hence, this is the upper limit of usefulness for the technique at present. As the carrier concentration decreases, the reflectivity minimum becomes broader and broader. Resolving the position of this minimum poses a problem in instrumentation which presently places the lower practical bound at 5×10^{18} cm^{-3} for which $\lambda_m = 25\mu$.

In applying the plasma resonance technique to diffused structures, however, another problem becomes paramount as the surface concentration is lowered. That is the penetration depth of the infrared beam. In p-type Si, for example, the penetration depth (where the transmitted intensity has fallen by 1/e) at the plasma resonance frequency is 2500 Å for a concentration of 10^{21} cm^{-3}, 7000 Å for 10^{20} cm^{-3} and 20,000 Å for 10^{19} cm^{-3}. If the junction depth is large (300-400%) compared to this depth, then the bulk curves can be applied directly. Such uncorrected plasma resonance measurements on diffused layers are reported [30,33] with accuracies of ±10% claimed. The depths of these layers, however, were not reported.

For shallow diffusions corrections are possible, but these corrections require an assumption regarding the form of the distribution. Corrected λ_m - N_I curves, which allow for certain limited choices of junction depth, have been published for erfc and Gaussian n-type diffusions [31].

Thus for diffused junctions of suitable depth, and with surface concentrations between 5×10^{18} cm^{-3} and 5×10^{20} cm^{-3} of either type, the plasma resonance measurement

provides a quick, nondestructive, and reasonably accurate determination of surface carrier concentration. For shallower junctions, however, which means one micron or less for $C_s \approx 10^{20}$ cm^{-3}, or 8 microns with $C_s \approx 10^{19}$ cm^{-3}, corrections to bulk λ_{m_2}-N_I curves are required and these corrections in general have not been published.[2]

2.8 MIS C_{max}/C_{min} Technique

The final technique for measuring surface concentrations, to be described in this review, is based on the capacitance of a metal-insulator-semiconductor (MIS) structure. A typical realization of this structure consists of an evaporated gold dot on an SiO_2 layer on the surface of a Si slice. When an electrical bias is applied such as to cause accumulation of majority carriers at the semiconductor surface under the dot (positive bias for n-type material), the high-frequency ($\geq$1 MHz) capacitance of the structure acquires a value, C_{max}, given by $\varepsilon_I A/W$, where ε_I and W are the permittivity and thickness, respectively, of the insulating layer, and A is the area of the metal dot. When the bias is reversed, a depletion region is produced in the semiconductor region under the dot and the capacitance decreases until it saturates at some lower value, C_{min}. The latter situation is attained when the reverse bias is just large enough to start the onset of inversion, i.e., accumulation of minority carriers at the semiconductor-insulator interface. The measured capacitance then is $C_{min} = (1/C_{max}+1/C_s)^{-1}$ where $C_s = \varepsilon_{Si}\, A/d$ and ε_{Si} is the permittivity of Si and d is the width of the depletion region. Now d is a function of the ionized impurity density in the depletion region. Hence a measurement of C_{max} and C_{min} is sufficient to determine the average ionized donor or acceptor concentration in a depth d of a Si slice [35-37]. The relations between n,d, and C_{max}/C_{min} are shown in Fig. 5.

The depletion width at inversion bias decreases with increasing n, and thus C_s increases. When C_s is too large compared to C_{max}, it can no longer be measured with accuracy, i.e., C_{max}/C_{min} approaches unity. This is the case for $N \geq 10^{19}$ cm^{-3}, at which doping $d \approx 130$ Å. There is no lower limit on the doping for measurements in bulk material, but of course d increases as the doping is lowered and the technique becomes less suitable for measuring surface concentrations. At $N = 10^{18}$ cm^{-3}, d = 360 Å while at $N = 10^{17}$ cm^{-3} d = 1076 Å.

Since MIS structures can be built on and subsequently removed from a silicon surface, this technique is nondestructive. It therefore offers an appropriate complement to the plasma resonance technique, provided the junction depths are large compared to the above-mentioned values of d. In principle, a correction factor could be calculated for arbitrary junction depth if the distribution is known. None have been published, however.

The capacitance of a p-n junction in series with the MIS capacitance complicates measurement of the latter. The MIS technique is particularly suited to diffusions of the same conductivity type as the original slice. Opposite type surfaces may also be analysed if the p-n junction capacitance is very much larger than C_{max}, of if the junction can be circumvented by ohmic connection to the diffused surface.

3. Summary

The table below summarizes the features and limitations of each of the eight measurement techniques discussed. In this table n stands for carrier density, presumed to be the same as net ionized impurity density. N_I refers to total impurity density including unionized and precipitated portions. Other symbols have been previously identified in the text or are identified in the table footnotes. As a summary, this table of course does not show all the reservations and restrictions mentioned in the text.

---

[2] "Junction" is used figuratively here as the plane where diffused and background concentrations are equal. The diffused layer and the original slice need not be of opposite conductivity types for the plasma resonance technique to be applied.

Table 1. Features of various techniques for measuring C_s

Technique	Quantity Determined	Assumptions	Range and Limitations	Destructive
1. $\rho_s - x_j - C_s$	N_I	erfc or Gaus.; $\rho - N_I$	$C_s < 10^{20}$ .cm^{-3}	yes
2. $R_H - C_s$	N_I	erfc; $\rho - N_I$	$C_s < 10^{20}$ cm^{-3}	yes
3. $\Delta\rho_s$	n or N_I (a)	$\rho - N_I$		yes
4. Δ-radioactive	N_I			yes
5. ΔR_H, $\Delta\sigma_s$	n	μ_H/μ_c		yes
6. Spread. -Resis.	n or N_I (a)	$\rho - N_I$	$xj > 25\mu$	
7. λ_m	n (b)		$5\times10^{18} < C_s < 5\times10^{20}$ cm^{-3} $x_j > 4/\alpha$ (c)	
8. MIS	n		$C_s < 10^{19}$ cm^{-3} $x_j >> d$ (d)	

(a) The quantity determined is ρ which may be translated into either n or N_I.
(b) Physically, λ_m depends on n, but the published calibration curves employ a mixture of n and N_I.
(c) α is the absorption coefficient in Si at λ_m and C_s
(d) d is width of the depletion layer at C_{min}.

Measurements of solid-solubilities in Si are no better off than those of C_s, and generally for the same reasons. However, most reports [4,5,7,38,39] place the maximum solubility of either P or B in Si at 2 to 5×10^{20} cm^{-3}. Likewise, it has been observed [13,39,40] that the resistivity of n-type Si cannot be made less than 0.00035 ohm-cm (however much the donor density is increased) which corresponds to $N_I = 4\times10^{20}$ cm^{-3}. (Minimum resistivities for p-type Si have not been reported). It would appear, therefore, that the maximum value of C_s that should be encountered is not more than 5×10^{20} cm^{-3} (carrier or ionized impurity density). This means that the upper boundary on application of the plasma resonance technique is no practical restriction at all. Thus this technique gets three stars as the method of choice for high surface concentrations provided the junction is not too thin.

A common diffusion technique which produces lower surface concentrations is solid-solid diffusion from a doped oxide. The resulting surface concentrations are almost always less than 10^{20} cm^{-3}, and usually much less (10^{15} - 10^{19} cm^{-3}). Under common boundary conditions, the diffused impurity distribution follows an erfc profile [41,15]. For such layers, the ρ_s-x_j-C_s method is appropriate and is simpler to apply than other techniques. Gaussian diffusions produced by long drive-ins from shallow predeposits, also frequently possess $C_s < 10^{20}$ cm^{-3} and are therefore amenable to ρ_s-x_j-C_s analysis.

As noted in Table 1, four of the measurement techniques involve direct or indirect use of the Si ρ-N_I relationship, either the curves reproduced here as Fig. 3, or in the case of Tufte's R_H-C_s curves, older data. In addition, the calibration curves [30,32] of the plasma resonance (λ_m) technique lean to some degree on the Si ρ-N_I curves. Thus any errors in the ρ-N_I curve are propagated into these C_s measurements. The accuracy of the ρ-N_I curves [1] is estimated to be 10% up to $N_I = 10^{20}$ cm^{-3}, and less accurate in the last decade, 10^{20} cm^{-3} to 10^{21} cm^{-3}. Furthermore, except for the last decade of the n-type curve where some data from diffused layers were used, the ρ-N_I curves are based on and strictly apply only to uncompensated, bulk material. The additional scattering mechanisms and precipitation near the surface of diffused layers may produce a radically different ρ-N_I relationship in that region. All four of these methods for determining C_s are therefore quite circumspect in the region $C_s \geq 10^{20}$ cm^{-3}. The ρ_s-x_j-C_s and R_H-C_s methods, using Irvin's [1] and Tufte's [2] curves, are already disqualified for $C_s \geq 10^{20}$ cm^{-3} due to untenable assumptions regarding the mathematical forms of the profiles.

Fortunately, the ρ_s-x_j-C_s method always errs on the high side (by as much as a factor of 10) when the error is due to the non-Fickian diffusion associated with high surface concentrations. Thus if the value of C_s as obtained by this method is near

or greater than 10^{20} cm^{-3}, the result should be considered unreliable. Any values over 5×10^{20} cm^{-3} obtained by this method should be rejected outright. As in tasting wine, discerning between the good and the great takes some expertise, but the poor results are immediately obvious.

4. Acknowledgement

The author is grateful to L. A. Murray, P. A. Schumann, Jr, and O. N. Tufte for private contributions and to R. M. Ryder and S. M. Sze for critical comments.

5. References

[1] J. C. Irvin, Bell Sys. Tech. J., 41, 387 (1962).

[2] O. N. Tufte, J. Electrochem. Soc., 109, 235 (1962)

[3] G. Backenstoss, Bell Sys. Tech.,J. 37, 699 (1958).

[4] T. H. Yeh and M. L. Joshi, J. Electrochem. Soc., 116, 73 (1969).

[5] A. Rupprecht and C. H. Schwuttke, J. Appl. Phys., 37, 2862 (1966).

[6] M. L. Joshi and Wilhelm, J. Electrochem. Soc., 112, 185 (1965).

[7] M. L. Joshi and S. Dash, IBM J., 10, 446 (1966).

[8] P. A. Iles and B. Leibenhaut, Solid-State Electron., 5, 331 (1962).

[9] B. J. Masters and J. M. Fairfilld, J. Appl. Phys., 40, 2390 (1969).

[10] A. G. Petrova, Ind. Lab., 33, 226 (1967).

[11] S. M. Hu and S. Schmidt, J. Appl. Phys. 39, 4272 (1968).

[12] R. P. Donovan, this issue.

[13] E. Tannenbaum, Solid-State Electron., 2, 123 (1961).

[14] R. A. Evans and R. P. Donovan, Solid-State Electron., 10, 155 (1967).

[15] M. L. Barry and P. Olofsen, J. Electrochem. Soc., 116, 854 (1969).

[16] T. M. Golovner and V. V. Zadde, Ind. Lab., 34, 1323 (1968).

[17] M. C. Duffy, F. Barson, J. M. Fairfield, G. H. Schwuttke, J. Electrochem. Soc., 115, 84 (1968).

[18] T. H. Yeh, S. M. Hu, and R. H. Kastl, J. Appl. Phys, 39, 4266 (1968).

[19] A. A. Smales, V. P. Guinn, Trace Characterization, NBS Monograph 100 (1967).

[20] A. S. Gillespie and W. W. Hill, Nucleonics, 19, 170 (1961).

[21] R. L. Petritz, Phys. Rev., 110, 1254 (1958).

[22] V. K. Subashiev and S. A. Poltinnikov, Sov. Phys. Solid St., 2, 1059 (1961)

[23] R. Baron, G. A. Shifrin, O. J. Marsh, J. Appl. Phys., 40, 3702 (1969).

[24] K. B. Wolfstirn, J. Phys. Chem. Solids, 16, 279 (1960).

[25] T. H. Yeh, this volume.

[26] R. G. Mazur and D. H. Dickey, J. Electrochem. Soc., 113, 255 (1966).

[27] P. A. Schumann, Jr., and E. E. Gardner, Solid-State Electron., 12, 371 (1969).

[28] J. M. Adley, M. R. Poponiak, C. P. Schneider, P. A. Schumann, Jr., A. H. Tong, First International Symposium on Silicon Material Science and Technology (Semiconductor Silicon), Electrochem. Soc. 1969, p. 721.

[29] L. A. Murray, J. J. Rivera, P. A. Hoss, J. Appl. Phys., 37, 4743 (1966).

[30] E. E. Gardner, W. Kapallo, and C. R. Gordon Appl. Phys. Lett., 9. 432 (1966).

[31] T. Abe and Y. Nishi, Japan J. Appl. Phys., 7, 397 (1968).

[32] P. A. Schumann, Jr., Solid-State Tech., 13-1, 50 (1970).

[33] L. A. Murray, R. A. Duclos, M. Goldsmith, A. Bomberger, Electrochem. Soc. Meeting, Oct. 1966, Abs. 194.

[34] L. A. Murray, Private Communication.

[35] Deal, Grove, Snow, and Sah, J. Electrochem. Soc., 112, 308 (1965).

[36] C. Jund and R. Poirier, Solid-State Electron., 9, 315 (1966).

[37] Y. Nishi and M. Konaka, Japan J. Appl. Phys., 5, 1116 (1966).

[38] G. L. Vick and K. M. Whittle, J. Electrochem. Soc., 116, 1142 (1969).

[39] A. G. Nassibian and G. Whiting, Solid-State Electron., 8, 843 (1965).

[40] E. Kooi, J. Electrochem. Soc., 111, 1383 (1964).

[41] K. Nagano, S. Iwauchi, T. Tanaka, Japan J. Appl. Phys., 7, 1361 (1968).

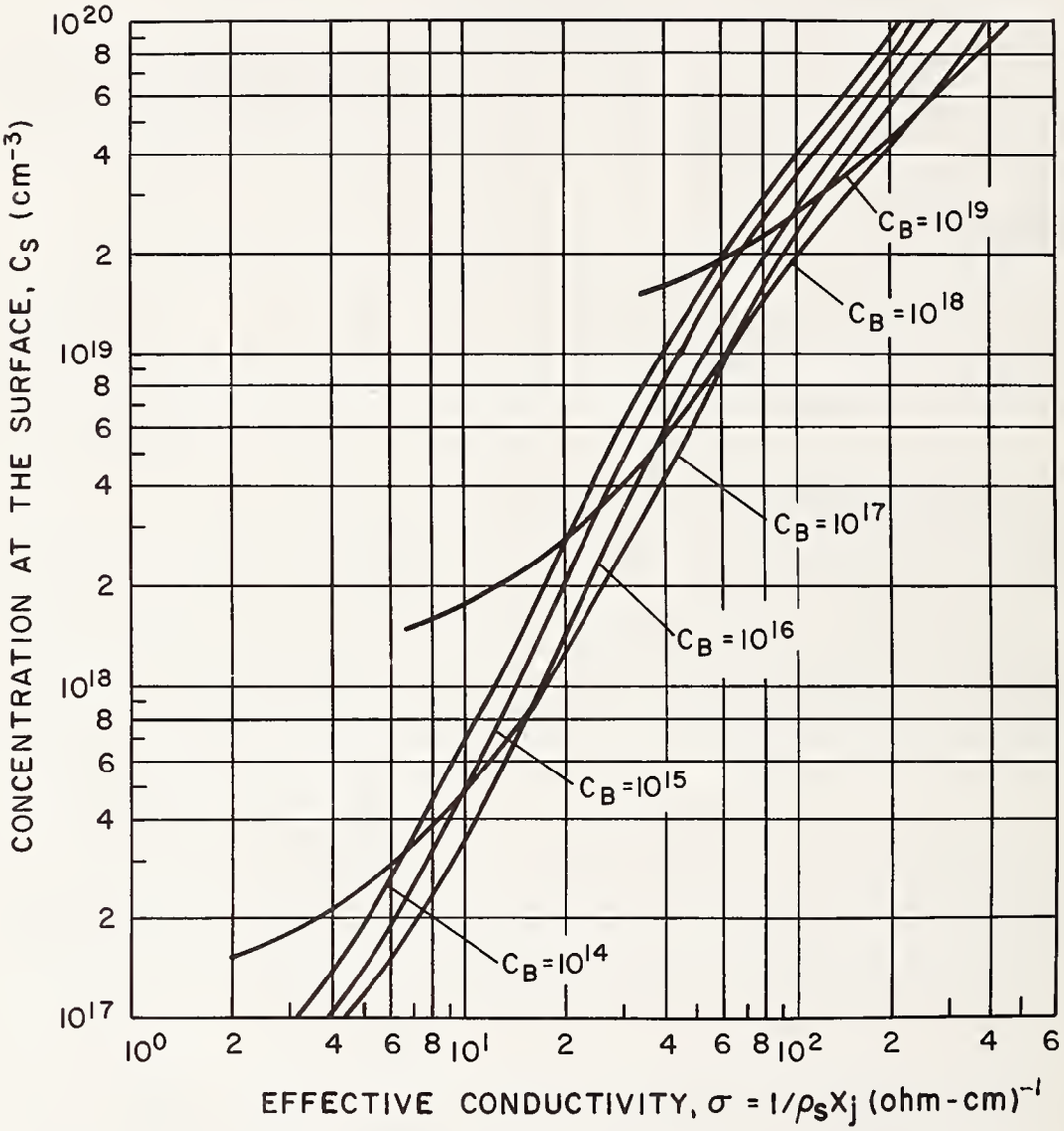


Fig. 1 Calculated relation between effective conductivity and surface concentration of diffused n-type, erfc layers in p-type Si. C_B is the background acceptor concentration.

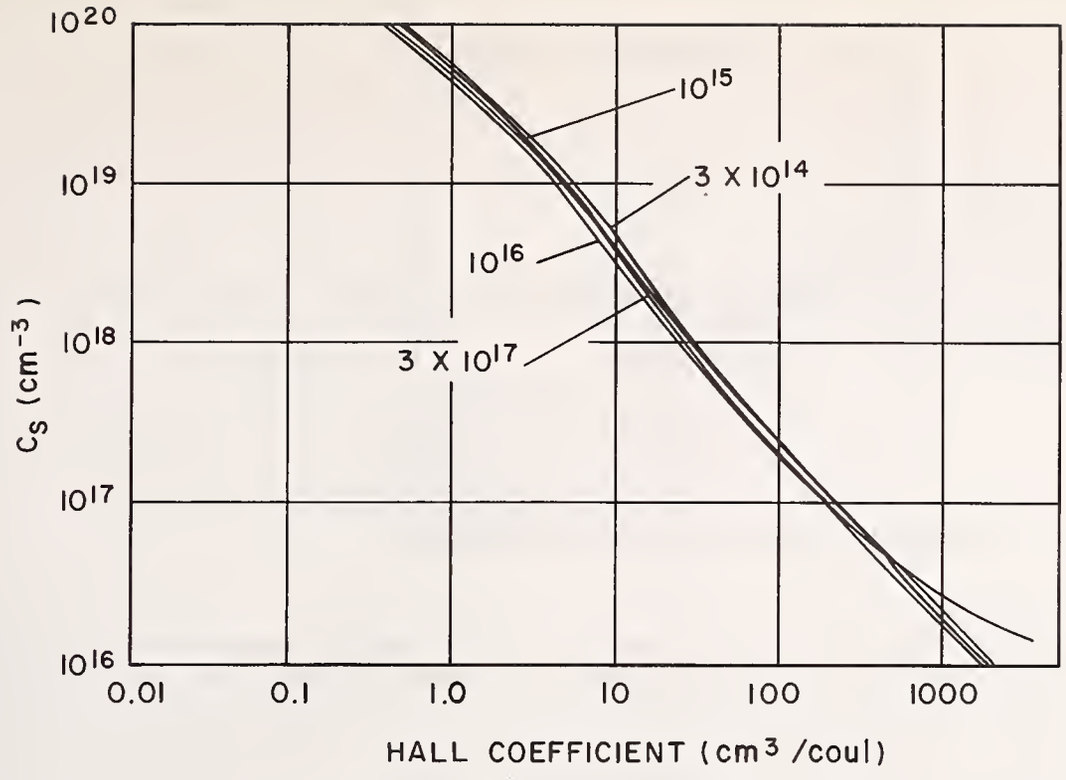


Fig. 2 Calculated relation between the Hall coefficient and surface concentration of diffused, n-type, erfc layers in p-type Si. The labels on each curve refer to background acceptor concentration. (O. N. Tufte, J. Electrochem. Soc. 109, 235 (1962).)

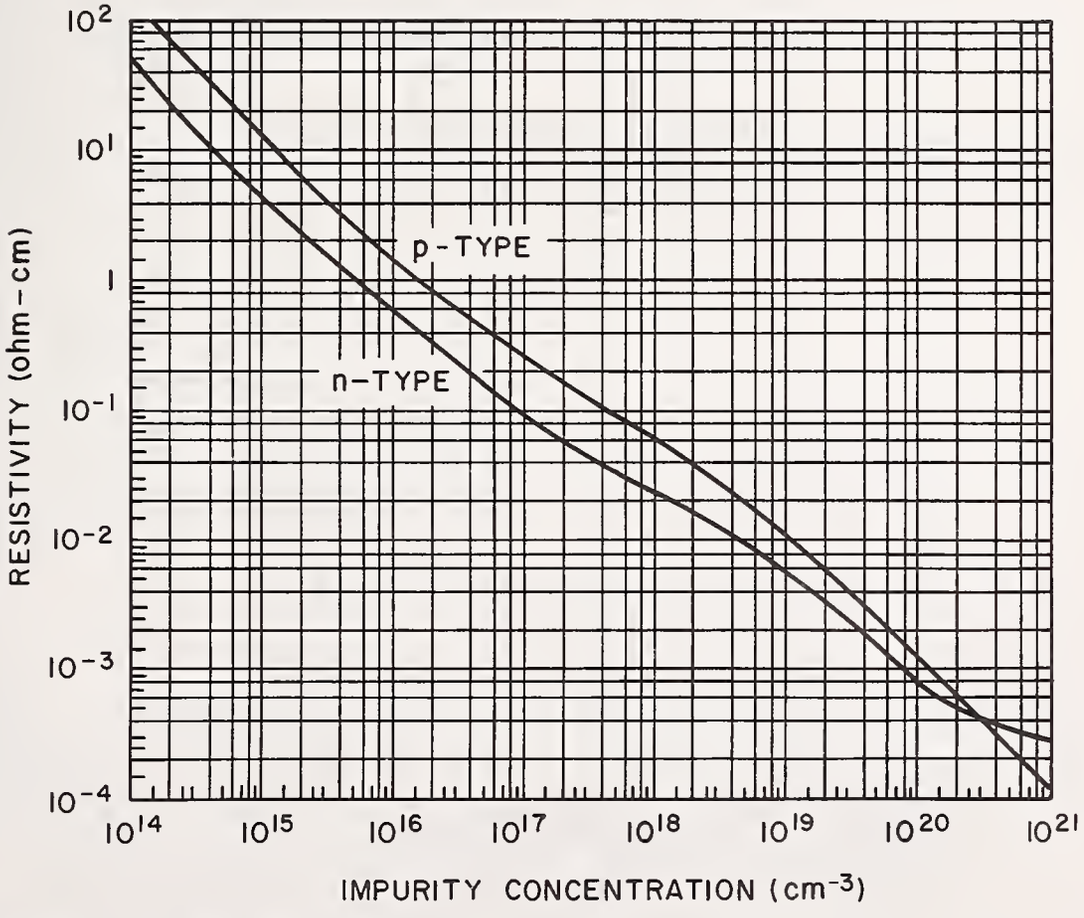


Fig. 3 Resistivity of Si at 300°K as a function of net acceptor or donor concentration. (J. C. Irvin, Bell Sys. Tech. J. 41, 387 (1962).)

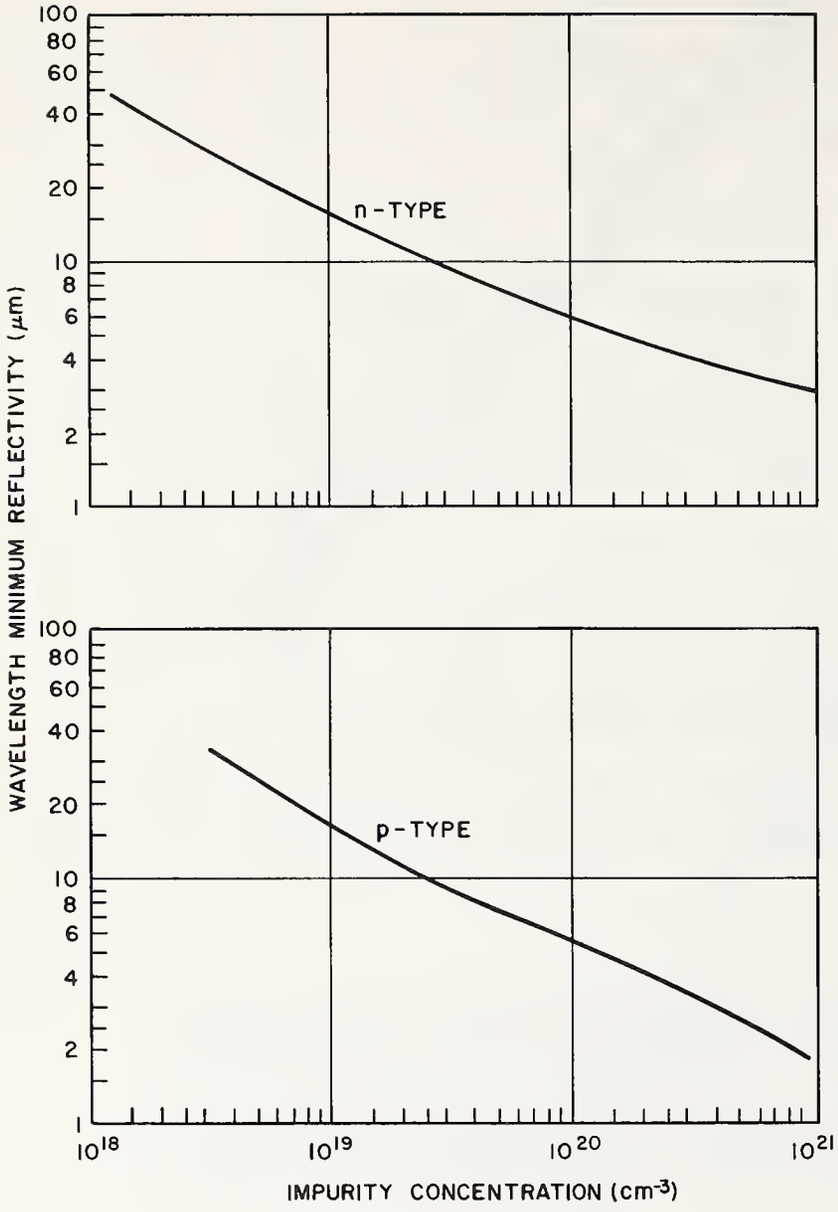


Fig. 4. Plasma resonance calibration curves for Si showing wavelength of reflectivity minimum as a function of impurity concentration. (P. A. Schumann, Jr., Solid-State Technol. 13, No. 1, 50 (Jan. 1970).)

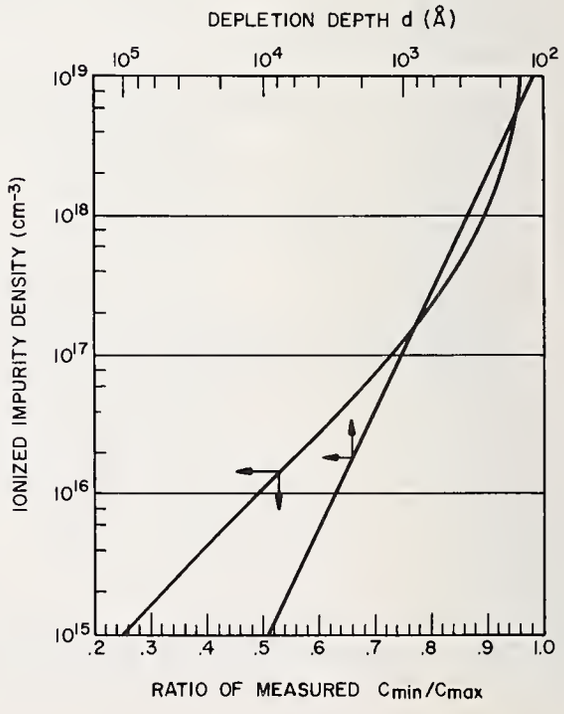


Fig. 5. Calculated ratio of measured C_{min} to C_{max} and also depletion depth of an MIS capacitor on Si with 1000 Å SiO_2 as a function of ionized impurity density.

Current Status of the Spreading Resistance Probe and its Application

T. H. Yeh

IBM Components Division
East Fishkill Facility
Hopewell Junction, N.Y. 12533

The spreading-resistance technique has shown the versatility of determining either the thickness of difffused or epitaxial layers, or establishing the impurity profiles for various multi-layered silicon structures. However, correction factors must be applied to the measured spreading-resistance values in order to get the correct resistivities. The calculation for the correction factors based on either the unilayer step-junction theory has been discussed by the author and the others. The utilization of a computer for such calculations to obtain the corrected resistivities, thus the impurity profiles, on a typical N/N+ structure has also been demonstrated.

In this paper, the calculations of the correction factors for spreading-resistance probe measurements made on a N/N+/P- structure, based on the multi-layer step-junction theory and the utilization of a computer for such calculations is presented. An iteration scheme is also adopted in the process of obtaining the correction factor, hence the corrected resistivity, for each measured value. In the end, a figure of corrected impurity profile vs depth is plotted-out by the computer automatically.

In addition, use of the spreading resistance probe to obtain the impurity profiles for the diffused structures of N/P or P/N, based on the multi-layer step-junction theory and the utilization of a computer for the calculation of the correction factors, will be discussed.

Key Words: Spreading resistance, diffused layers, epitaxial layers, multi-layer, silicon structures, correction factors, unilayer, step-junction, probe measurements, impurity profiles.

The spreading-resistance probe has been proposed some time ago by Mazur and Dickey[1] and Gardner, Schumann and Gorey[2] for determination of either the thickness of diffused or epitaxial layers of silicon or establishing the impurity profiles for various multi-layered structures. These determinations are still considered the probes' major applications. Other applications include the mobility of N/N+ epitaxial layers,[3] the effect of high temperature heat treatment on silicon resistivity,[4] resistivity striation in silicon caused by interstitial oxygen,[5] and the study of the electrical properties of imperfections in bulk and epitaxial silicon.[6] At present, ASTM is in the process to draft a procedure for this technique to measure the resistivities of epitaxial layers.

The development of the spreading resistance probe was started primarily to overcome some of the difficulties existing at the time for several well known techniques to measure the resistivity of certain diffused or epitaxial structures. The four-point probe can be used to establish the resistivity (or impurity profile) of a diffused structure[7] but can not be used to establish the resistivity (or impurity profile) of a lightly doped epitaxial layer grown on heavily doped substrate of the same conductivity type. This method requires an isolating junction between the layer being measured and the underlying substrate. The three point breakdown[8,9] or differential capacitance techniques[10,11] may be used to determine the layer resistivity on the substrate of the same conductivity type. But, neither of these techniques are useful when an isolating junction exists between them.

The spreading resistance probe can be used to establish resistivity (or impurity) profiles and thickness of multi-layered silicon structure because it measures the localized resistivity on the sample in an extremely small volume, in the order of $10^{-10}cm^{-3}$. If a flat circular voltage probe makes contact to a semi-infinite conducting material,[12] the potential distribution in the vicinity of the probe occurs as shown in Fig. 1. It is observed that practically all the potential drops occurs within a distance of a few probe radii of the probe center. For an ideal metal-semiconductor contact, having no barrier resistance, the potential of the contact is directly related to the semiconductor resistivity ρ by

$$V_0 = \frac{\rho I}{4r_0} \qquad (1)$$

Gardner, Schumann and Gorey[2] have found it convenient to use three probes -- two of which are connected to a current source and the potential difference is measured between one of these and the third probe as shown in Fig. 2. For this experimental arrangement, the probe connected to both the current source and the voltmeter

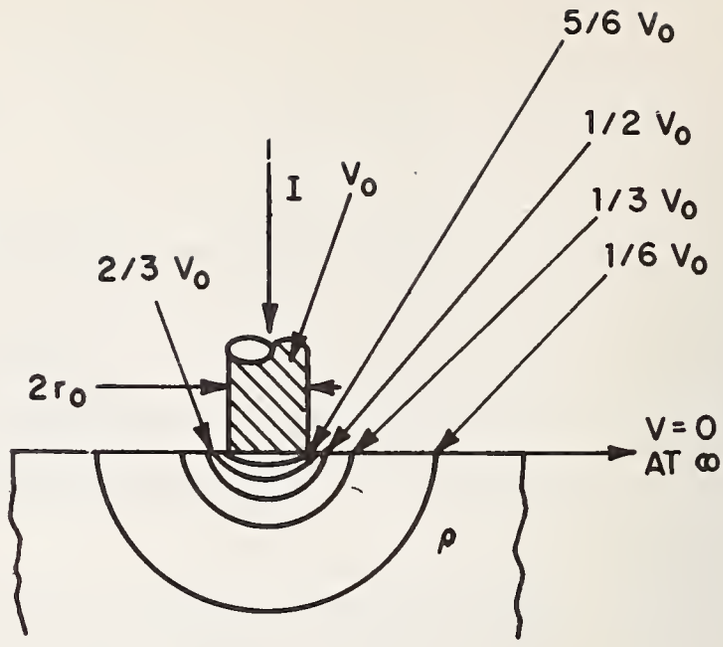


Fig. 1. Potential distribution of flat circular probe on semi-infinite medium.

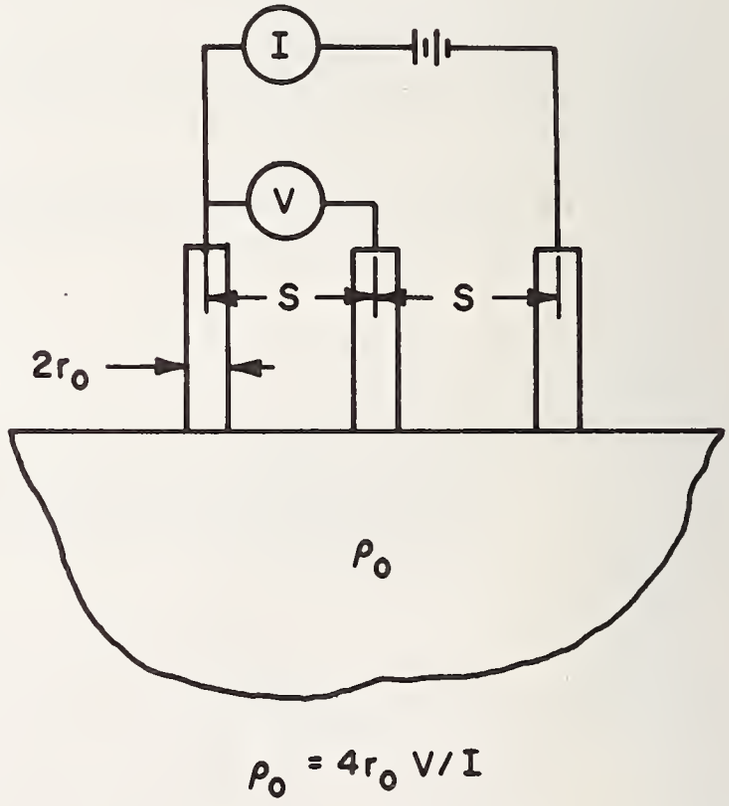


Fig. 2. Spreading resistance technique on semi-infinite medium (after Gardner et al[2]).

is the critical probe. A second configuration is to place the left hand current connection of Fig. 2 on the center probe -- in this case the center probe becomes the critical probe. The chief consideration in the design of the apparatus is the control of the common probe's contact parameters; contact area, depth of penetration, contact pattern and surface damage. To regulate these parameters, the probe tip loading, velocity of contact, and deformation must be controlled precisely. Schumann et al have discussed these points in detail.[5] Some of the important design considerations are summarized below. The probe arms are arranged as balanced beams, pivoted on precision bearing, as shown in Fig. 3. Each of the probe arms are individually weight loaded and adjustable from 0 to 100 grams.

The vertical movement of the probe arms is controlled by lifters riding on a cam plate, Fig. 4. When the arms are descending, the motion of the cam plate is controlled by an adjustable precision dash pot which varies the rate of descent of the probe arms from approximately four mils per second immediately after release to near zero at the time the common probe engages the sample. The actual velocity at which the common probe engages the sample is determined experimentally. The cam plate also serves to bring the three probes into engagement with the sample in proper order. The common probe must engage the sample last to avoid any movement or vibration caused by impact of the current and voltage probe.

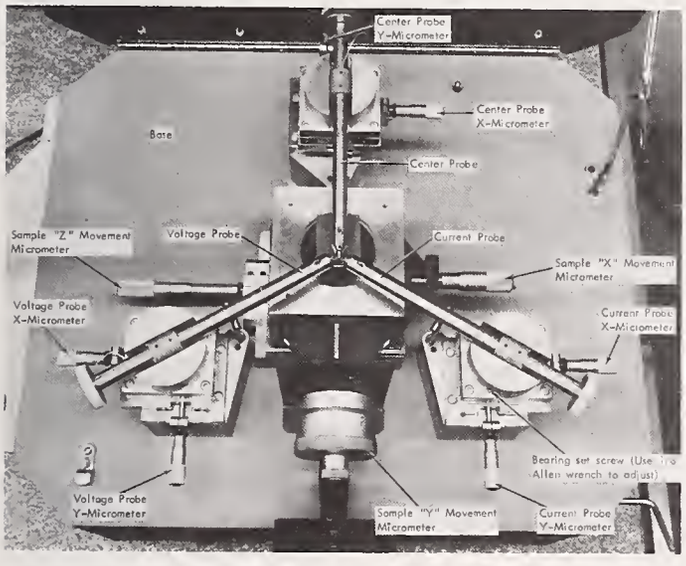


Fig. 3. Probe arms (after Schumann et al[5]).

The probe tips are a Ruthenium-Tungsten Alloy* with a tip radius of 0.0007 ± 0.00025 inch with an included tip angle of 40 to 50 degrees. The probe tip is bonded to an Inconel shaft which mounts in a chuck on the end of the probe arm.

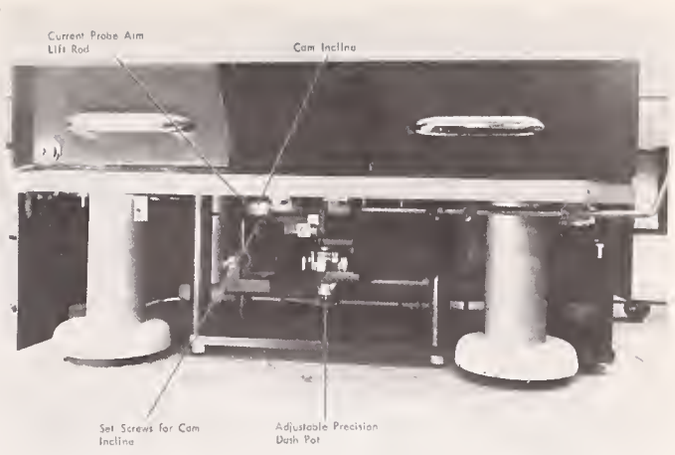


Fig. 4. Underside view of probe apparatus (after Schumann et al[5]).

The electronics for the spreading resistance measurements are shown in Fig. 5. The bipolar integrating digital voltmeter is used to measure the voltage drop between the common probe and voltage probes. This same voltmeter is also used to measure the current flow through the sample by switching the input to the voltmeter from the sample to a standard resistor connected in series with the sample.

A logic circuit composed of IBM SLT logic devices is used to control the switching circuits as follows. When the probe is raised the current supply is automatically shorted. When the probes are lowered and contact the sample, the logic circuit is activated. After a five second interval the current is measured, reversed, and measured again. The voltage is then measured in the same sequence. Twenty seconds are needed for the entire cycle. All readings may be recorded by the card punch. Forward and reverse readings of the current are identical, and of the voltage are within 5%. This voltage must not exceed thirty millivolts or the reading will be in error.

It has been emphasized by Schumann et al[5] that a spreading resistance probe will function properly if a particular type of probe point is used. A good probe point must have the following characteristics as discussed by Gorey et al;[13] the effective penetration of the probe into the silicon surface must be kept to a minimum, and the spreading resistance measured in the foreward and reverse direction of current must not differ by more than 5% of the average. The best probe point is prepared by an abrasive blasting technique[13] in the following manner. A series of probe points* to be prepared are positioned in a holder with their points exposed. An air abrasive unit is then used to abrade the point with 27.5μm aluminum oxide with a 50 psi pressure. The unit has a 1/6

*K-75 pivot, Fideltone, Inc.

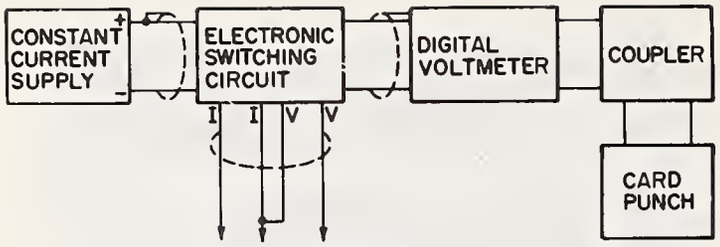


Fig. 5. Block diagram of spreading resistance probes' electronics (after Schumann et al[5]).

inch nozzle orifice. The nozzle of the unit is placed approximately one inch from the probe points. After abrading, each probe is immersed in trichlorethylene for five minutes and cleaned with acetone. This procedure gives the probe tips the proper finish to make good contact to the sample.

A convenient method to ensure a good probe point is to make measurements on three wafers; first on a thin (approx. 2.0μm thick) n-type epitaxial layer of certain resistivity on an N+ substrate exhibiting a certain spreading resistance value, secondly on an n-type bulk wafer with approximately the same spreading resistance value as the epitaxial layer, and thirdly on a p-type mechanical polished bulk wafer of approximately 20Ω-cm. The operating range for the probe point is decided when both the ratios of the spreading resistances of the N/N+ wafer to the n-type bulk wafer and the spreading resistances in the reverse to the forward direction of current of the p-type wafer have a slope of zero when they are plotted against loads as shown in Fig. 6. The exact loading of the probe point should be obviously chosen from the middle of this operating range, and experimental data showed that a minimum of 250 measurements are required to stabilize the probe point.

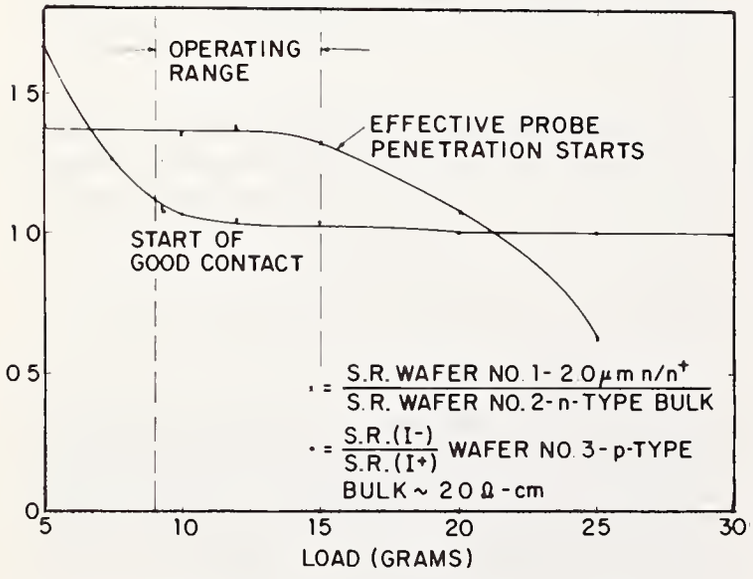


Fig. 6. Ratios of spreading resistances of N/N+ to N-type bulk wafers and spreading resistances in the reverse direction to the forward direction of current as a function of loading for a typical prepared K-75 probe point (after Gorey et al[13]).

The apparatus described in Mazur and Dickey's early work[1] consisted of one or two probes (Jensen A-70LP phonograph needles) supported in such a manner that they could be carefully placed with a known, fixed force on a sample without dragging. The total resistance between a pair of probes or between one probe and a low resistance base contact was measured with a Kiethley 610A electrometer-ohmmeter. The probes had an osmium tip of 2.5×10^{-3} cm radius of curvature and were mounted on separate boom, which were loaded with calibrated 50 gm weight. Recently, Mazur[14,15] has improved his apparatus into an automatic machine.

Most recently, Gupta and Chan[16] have described a semiautomatic spreading resistance probe which can make contact to the specimen with the same force during each measurements, operates in the constant output voltage mode, and the mechanical vibrations are minimized, such that the probe's effective contact area to the specimen is kept constant. Their probe's tips were made of tungsten carbide with a tip radius of 0.04 mm. The probes were spring loaded with a force of 15 grams and had no lateral movement. Their test set could be operated in two modes: constant input current (CC) and constant output voltage (CV). The CC mode, which a constant current of 1 μA to 100 MA is passed through the specimens, depending upon its resistivity, and the potential difference is measured on a sensitive digital millivoltmeter. The CV mode is operated at a constant voltage of 4mV for all ranges of resistivity. They claimed that the reproducibility of the measurements is better for the CV mode, especially for the higher resistivity specimens, as shown in Table I for the comparative data, probably because the damage on the specimen in the contact area is significantly less when the test set is operating in the CV mode. Furthermore, they pointed out that in the CV mode the reproducibility of the measurement is essentially independent of resistivity, while in the CC mode it varies with the resistivity and is poor for higher resistivity specimens as shown in Fig. 7.

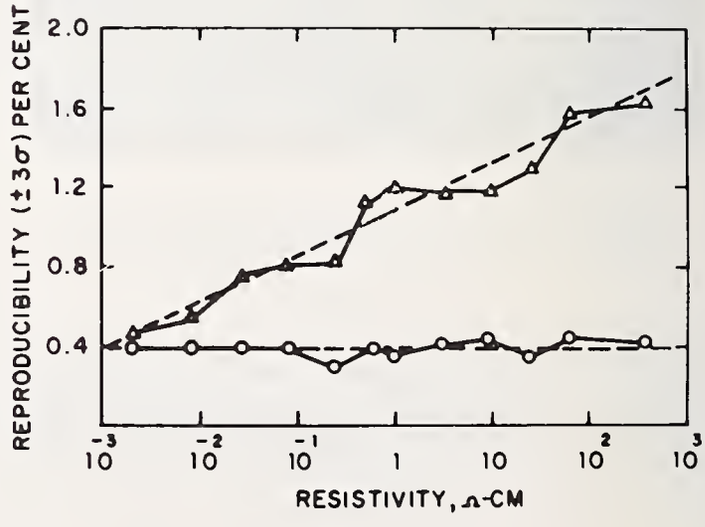


Fig. 7. Reproducibility of the measurements • CV mode Δ CC mode (after Gupta and Chan[16]).

TABLE I. Reproducibility of the S-R probe measurements in CC and CV modes.

Resistivity of N-type silicon (Ω-cm)	Reproducibility (± 3σ), percent (CV mode)	(CC mode)
0.002	0.39	0.43
0.2	0.31	0.80
2.0	0.39	1.21
20.	0.38	1.28
200.	0.41	1.60

If the semiconductor material is semi-infinite in extent and the probe spacing, S, is large compared to the contact radius, r_0, then the potential is the same as given by Eq. (1). Thus the measured spreading resistance, $R_{S.R.}$, is related to the resistivity of the material and the effective radius of contact of the probe as follows:

$$R_{S.R.} = \frac{V}{I} = \frac{\rho_0}{4r_0} \quad . \qquad (2)$$

But for a diffused structure or a lightly doped epitaxial layer grown on heavily doped substrate of the same conductivity type i.e., the semiconductor material is not semi-infinite in extent but a layered structure as shown in Fig. 8. The spreading resistance, $R_{S.R}$, is given by the expression[17]

$$R_{S.R} = \frac{\Delta V}{I} = \frac{\rho_1}{4r_0}\left\{\frac{4}{\pi}\int_0^\infty \left[\frac{1 + K_1 e^{-2H_1 X}}{1 - K_1 e^{-2H_1 X}}\right]\cdot \left[\frac{J_1(X)}{X^2} - \frac{J_0\left(\frac{S}{2}X\right)}{X} + \frac{J_0(SX)}{2X}\right] dX\right\} \qquad (3)$$

where

$R_{S.R.}$ = Spreading resistance

$K_1 = (\rho_2 - \rho_1)/(\rho_2 + \rho_1)$

$S = 2s/r_0$

$H_1 = h_1/r_0$

s = Probe spacing

r_0 = The effective radius of contact of the probe

X = Integration parameter

$J_1 J_0$ = Bessel function .

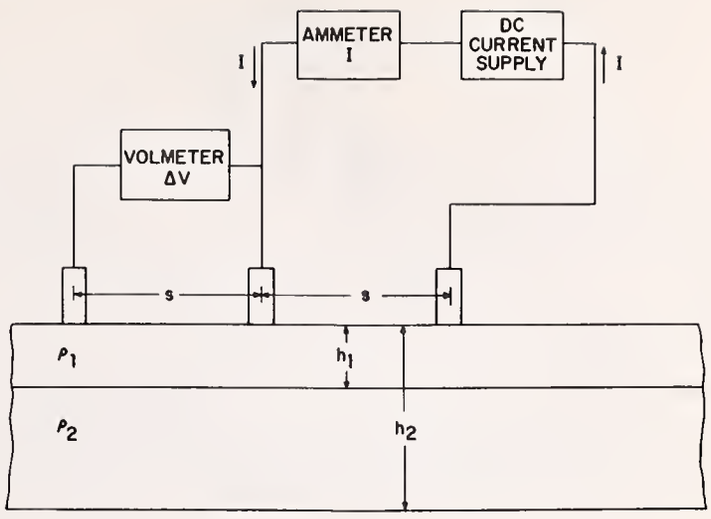


Fig. 8. Spreading resistance technique on a two-layered structure.

The terms within the curl bracket of Eq. (3) are generally labeled as a correction factor, C.F. The corrected resistivity value is then given by the following expression:

$$\rho_{1_{corrected}} = \frac{\rho_1}{C.F.} \qquad (4)$$

The calculation of the correction factors based on the assumption of the unilayer step-junction theory was discussed in the earlier publications.[2,18] The unilayer step-junction theory assumes that the resistivity changes at the interface are abrupt, and was used extensively to establish the impurity profile for the measurements made on N/N+ silicon structure.[2,18]

To determine an impurity profile, one can use a small test chip and bevel at a small angle as shown in Fig. 9. The probes are aligned as shown and are moved down the bevel in regular increments. The thickness, t, needed for the correction factor is shown in the same figure as the distance from the interface of the two layers.

For a N/N+/P⁻ silicon structure, the measured spreading resistance values are converted to resistivity values from the calibration curve[2] and plotted against distance as shown in Fig. 10. The calculation of correction factors is based on two distinct structures, N/N+ and N+/P⁻, according to Eq. 3. Each structure is considered to be a two layer problem as shown in Fig. 8.

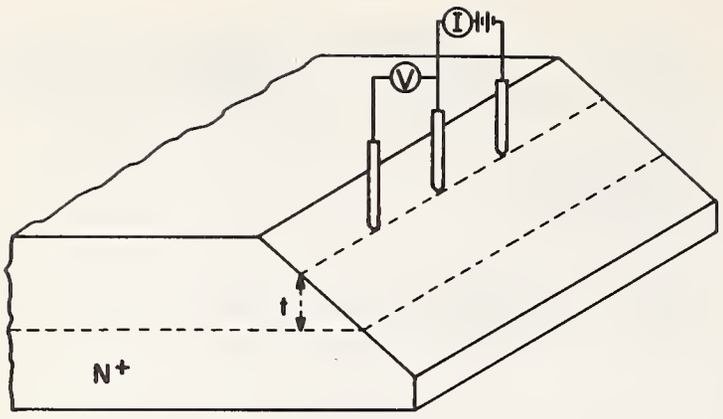


Fig. 9. Determination of resistivity profiles by moving probes down a beveled test chip.

The first measured resistivity to be corrected is labeled ρ_1, which actually is the last measurement in the N+/P⁻ structure, and this resistivity is assumed to be uniform throughout the layer thickness of $x_0 - x_1$ ($=h_1$). The resistivity change at the interface N+/P⁻ is abrupt, and an infinite resistivity can be assumed for ρ_0. Then the K_1 value in Eq. 3 for this problem is equal to $(\rho_0-\rho_1)/(\rho_0+\rho_1)$. Upon determining the correction factor, the resistivity is corrected using Eq. 4. This corrected resistivity value $\rho_{1\text{corrected}}$ may then be used to find a new effective radius of contact, r_0, from the calibration curve,[2] a new correction factor, and a new corrected resistivity. This procedure is repeated to get an exact value of resistivity, (i.e., until either the r_0 value or the correction factor does not change), because the r_0 value initially obtained from the calibration curve for the first correction factor calculation is not necessarily exact.

Because of the uni-layer step-junction assumption, neither this corrected resistivity, $\rho_{1\,\text{corrected}}$, nor its initial resistivity ρ_1 are utilized when the second measured resistivity ρ_2 is corrected. For correcting ρ_2, it is assumed that the resistivity, ρ_2, is homogeneous throughout its thickness $x_0 - x_2$ ($=h_1$), even though we know that within its thickness one of the layers has a different resistivity value, $\rho_{1\text{corrected}}$. The calculation of the correction factor for ρ_2 is again evaluated by Eq. 3 using the K_1 value of $(\rho_0-\rho_2)/(\rho_0+\rho_2)$.

Undoubtly, certain accuracy of the correction factor is sacrificed for each of the subsequent resistivity values, such as $\rho_3, \rho_4 \ldots \rho_{14}$ for the N+/P⁻ structure because of the uni-layer step-junction assumption. This is especially true for the corrected value of ρ_{14}, because in the calculation of its correction factor, this resistivity, ρ_{14}, is assumed to be uniform throughout the total thickness of $x_0 - x_{14}$, even though there are thirteen layers of corrected resistivity values, $\rho_{1\text{corrected}} \cdots \rho_{13\text{corrected}}$, below ρ_{14}.

Obviously, if one uses the corrected resistivity values to obtain a correction factor for any one of the measured values, e.g. utilizing ρ_1 corrected and ρ_2 corrected to correct ρ_3, then the value of ρ_3 corrected should be more accurate than the one without the consideration of those values of ρ_1 corrected and ρ_2 corrected. Likewise, this is true for the corrected values of $\rho_{15} \ldots \rho_{34}$ in the N/N+ structure, i.e., the value of ρ_{15} corrected $\ldots \rho_{34}$ corrected should be more accurate than the one without the consideration of those values of ρ_1 corrected $\ldots \rho_{14}$ corrected and ρ_{15} corrected $\cdots\cdots \rho_{33}$ corrected.

However, if one wants to use the preceeding corrected resistivities for the calculation of the correction factor for the subsequent initial resistivity, one must use the multi-layer step-junction theory to solve the problem. The geometry is shown in Fig. 11. The solutions to La Place's equation in cylindrical coordinates for this problem is:

$$R_n = \frac{V_n(R,Z)}{I} = \frac{\rho_n}{2r_0}\left\{ \int_0^\infty \frac{e^{-XZ} \sin X\, J_0(RX)}{X}\, dX + \int_0^\infty \frac{N^{\theta}n \left(\frac{X}{r_0}\right) e^{-XZ} \sin X\, J_0(RX)}{X}\, dX \right.$$

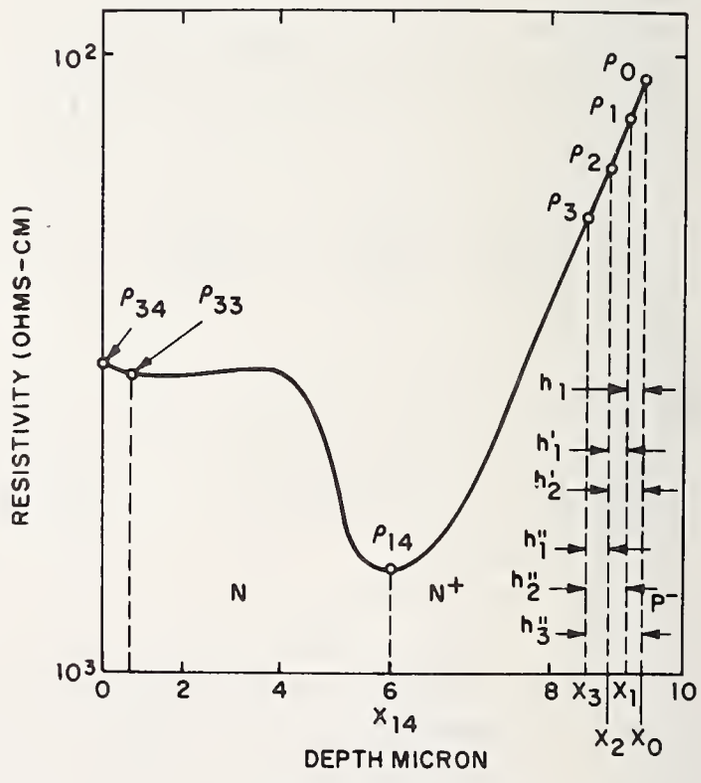


Fig. 10. Initial resistivity values vs. depths for N/N+/P⁻ structure.

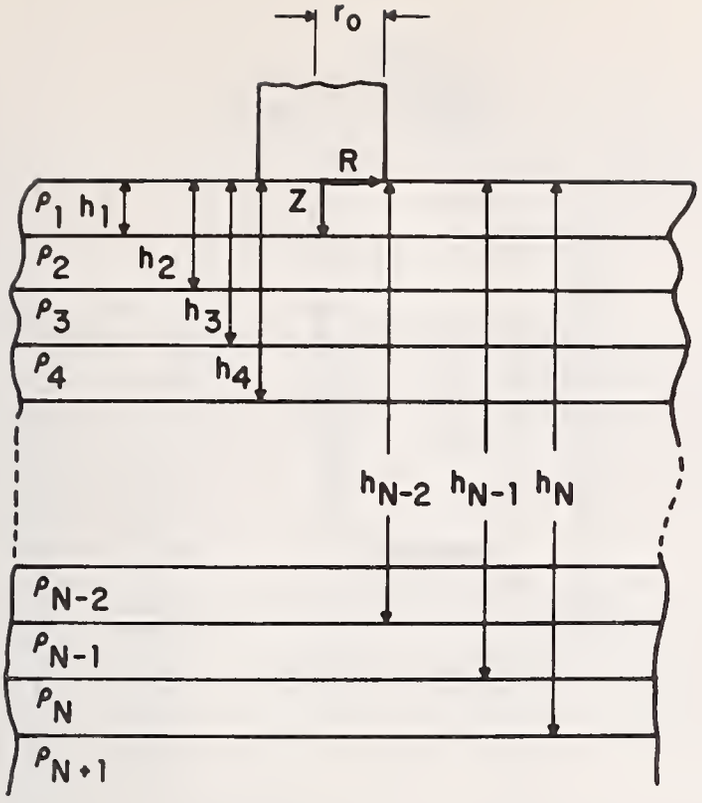


Fig. 11. Spreading resistance technique on a multi-layered structure.

$$+ \int_0^\infty \frac{{}^N\psi_n\left(\frac{X}{r_0}\right) e^{XZ} \sin X J_0(RX)}{X} \, dX \Bigg\}$$

$$n = 1,2,3 \ldots N \quad . \tag{5}$$

The boundary conditions for this type of problem are relatively standard with the exception of the current distribution under the contact.

The application of the boundary conditions[17,18,19] result in a series of linear equations:

$${}^N\theta_1\left|\frac{X}{r_0}\right| - {}^N\psi_1\left|\frac{X}{r_0}\right| = 0 \tag{6}$$

$$e^{-2H_nX} \cdot {}^N\theta_n\left|\frac{X}{r_0}\right| + {}^N\psi_n\left|\frac{X}{r_0}\right| - e^{-2H_nX} \cdot {}^N\theta_{n+1}\left|\frac{X}{r_0}\right| - {}^N\psi_{n+1}\left|\frac{X}{r_0}\right| = 0 \tag{7}$$

$$-e^{-2H_nX} \cdot {}^N\theta_n\left|\frac{X}{r_0}\right| + {}^N\psi_n\left|\frac{X}{r_0}\right| \cdot \beta_n e^{-2H_nX} \cdot {}^N\theta_{n+1}\left|\frac{X}{r_0}\right| - \beta_n {}^N\psi_{n+1}\left|\frac{X}{r_0}\right| = (1-\beta_n) e^{-2H_nX} \tag{8}$$

$${}^N\psi_{n+1}\left|\frac{X}{r_0}\right| = 0 \tag{9}$$

where N = Number of layers of different resistivity and finite thickness

n = Index, from 1 to N

$$\beta_n = \frac{\rho_n}{\rho_{n+1}}$$

$$H_n = \frac{h_n}{r_0}$$

and region N is infinitely thick.

Using the simplified notation

$${}^N\theta_n\left|\frac{X}{r_0}\right| = {}^N\theta_n \tag{10}$$

$$e^{-2H_nX} = E_n \tag{11}$$

The above equations become:

$${}^N\theta_1 - {}^N\psi_1 = 0 \tag{12}$$

$$E_n \cdot {}^N\theta_n + {}^N\psi_n - E_n \cdot {}^N\theta_{n+1} - {}^N\psi_{n+1} = 0 \tag{13}$$

$$-E_n \cdot {}^N\theta_n + {}^N\psi_n + \beta_n E_n \cdot {}^N\theta_{n+1} - \beta_n \cdot {}^N\psi_{n+1} = (1-\beta_n)E_n \tag{14}$$

$${}^N\psi_{n+1} = 0 \tag{15}$$

For example, if N = 2, the set of linear equations becomes:

$${}^2\theta_1 - {}^2\psi_1 = 0 \tag{16}$$

$$E_1 \cdot {}^2\theta_1 + {}^2\psi_1 - E_1 \cdot {}^2\theta_2 - {}^2\psi_2 = 0 \tag{17}$$

$$-E_1 \cdot {}^2\theta_1 + {}^2\psi_1 + \beta_1 E_1 \cdot {}^2\theta_2 - \beta_1 \cdot {}^2\psi_2 = (1-\beta_1)E_1 \tag{18}$$

$$E_2 \cdot {}^2\theta_2 + {}^2\psi_2 - E_2 \cdot {}^2\theta_3 - {}^2\psi_3 = 0 \tag{19}$$

$$-E_2 \cdot {}_2\theta_2 + {}_2\psi_2 + \beta_2 E_2 \cdot {}_2\theta_3 - \beta_2 \cdot {}_2\psi_3 = (1-\beta_2)\, E_2 \quad (20)$$

$${}_2\psi_3 = 0 \quad (21)$$

There are six equations and six unknowns. The matrix is, with the columns given by:

${}_2\theta_1$	${}_2\psi_1$	${}_2\theta_2$	${}_2\psi_2$	${}_2\theta_3$	${}_2\psi_3$	Constant
1	-1	0	0	0	0	0
E_1	1	$-E_1$	1	0	0	0
$-E_1$	1	$\beta_1 E_1$	$-\beta_1$	0	0	$(1-\beta_1)\, E_1$
0	0	E_2	1	$-E_2$	-1	0
0	0	$-E_2$	1	$\beta_2 E_2$	$-\beta_2$	$(1-\beta_2)\, E_2$
0	0	0	0	0	1	0

Determinants may then be formed to solve for any of the θ's and ψ's. In general, the matrix will be 2N + 3 by 2N + 2. The determinants will be 2N + 2 by 2N + 2.

Prior to this paper, the multi-layer theory has been tested for a N/N+ structure with the aid of a computer.[20] For this N/N+/P⁻ structure, a sample having thirty-four layers to be considered, one must solve an ever increasing number of determinants, starting with a 4 x 4 determinant then steadily expanding to a 70 by 70 determinant. With the aid of an IBM 360 Model 91 computer, it took about three minutes to obtain thirty-four corrected resistivities. These values were then converted to net impurity concentration according to Irvin's curve[21] and plotted against the depth as shown in Fig. 12. In the same figure, the un-corrected impurity profile and the corrected profile obtained by the unilayer step-junction approach are also included for comparison.

The output of the computer program includes one page of calculations based on simple multi-layer theory without iteration, as shown in Table II. For iterative approach, (i.e., upon determining the correction factor, the corrected resistivity is obtained by Eq. 4, and this corrected resistivity value is used to find a new effective radius of contact, r_0, from the calibration curve, a new correction factor, and a new corrected resistivity. This procedure is iterated until either the r_0 value or the correction factor does not change), the output for one data point at a time is printed on each page until the value for that data point has converged. The maximum number of iterations for each data point is limited to fifteen, so that iteration will be stopped if values are oscillating

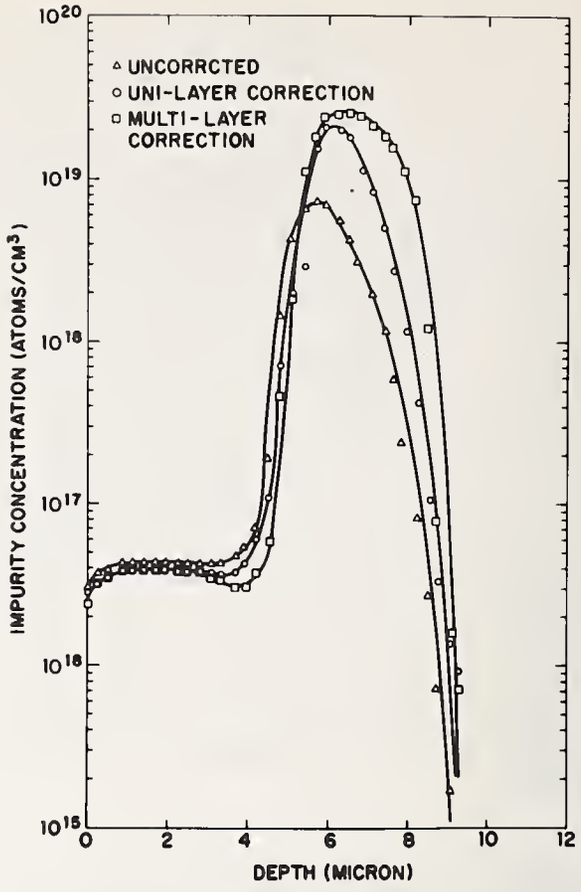


Fig. 12. Impurity concentration as a function of depth for N/N+/P⁻ structure.

within a narrow range. The final output includes the iterated and converged values of resistivity and their corresponding impurity concentrations, as shown in Table III.

The N/N+/P⁻ structure discussed so far really includes a diffused structure of N+/P⁻. And the calculation of the correction factors based on the multi-layer step-junction theory and the utilization of a computer for such a calculations are already done. The corrected impurity profile for this N+/P⁻ diffused structure is in part already represented in Fig. 12 for the N/N+/P⁻ structure, only if one has to realize that the surface of the diffused layer begins at 5.96 micron.

Figure 13 shows another corrected impurity profile for a N⁺/P⁻ diffused structure, and Fig. 14 shows the corrected impurity profile for a P-type diffused layer in a N/N+/P⁻ structure. All of these corrected resistivities, thus impurity concentrations, are based on the multi-layer step-junction theory calculations.

TABLE II. Evaluations based on multi-layer theory. Non-iterative approach.

SAMPLE NUMBER AR-2197 N(111) TYPE

PROBE SPACING = 0.6350E-01 CMS.
STEP SIZE = 0.2830E-04 CMS.
MATERIAL TYPE IS N111

NUMBER OF LAYER	DEPTH FROM SURFACE	SPREADING RESISTANCE	RESISTIVITY	CORRECTED RESISTIVITY	IMPURITY CONCENTRATION	EFFECTIVE RADIUS OF CONTACT	CORRECTION FACTOR
1	0.00000E-38	0.75000E 03	0.22684E 00	0.26256E 00	0.25872E 17	0.75612E-04	0.86393E 00
2	0.28300E-04	0.66000E 03	0.19792E 00	0.21849E 00	0.32222E 17	0.74970E-04	0.90584E 00
3	0.56600E-04	0.62000E 03	0.18588E 00	0.20238E 00	0.35311E 17	0.74952E-04	0.91846E 00
4	0.84900E-04	0.58000E 03	0.17388E 00	0.18426E 00	0.39499E 17	0.74948E-04	0.94368E 00
5	0.11320E-03	0.58000E 03	0.17388E 00	0.18614E 00	0.39021E 17	0.74948E-04	0.93412E 00
6	0.14150E-03	0.57000E 03	0.17089E 00	0.18282E 00	0.39869E 17	0.74950E-04	0.93470E 00
7	0.16980E-03	0.57000E 03	0.17089E 00	0.18470E 00	0.39385E 17	0.74950E-04	0.92519E 00
8	0.19810E-03	0.57000E 03	0.17089E 00	0.18709E 00	0.38786E 17	0.74950E-04	0.91340E 00
9	0.22640E-03	0.57000E 03	0.17089E 00	0.19021E 00	0.38028E 17	0.74950E-04	0.89843E 00
10	0.25470E-03	0.57000E 03	0.17089E 00	0.19444E 00	0.37039E 17	0.74950E-04	0.87884E 00
11	0.28300E-03	0.57000E 03	0.17089E 00	0.20053E 00	0.35700E 17	0.74950E-04	0.85216E 00
12	0.31130E-03	0.57000E 03	0.17089E 00	0.21002E 00	0.33782E 17	0.74950E-04	0.81368E 00
13	0.33960E-03	0.56000E 03	0.16789E 00	0.21915E 00	0.32107E 17	0.74953E-04	0.76612E 00
14	0.36790E-03	0.54000E 03	0.16192E 00	0.23151E 00	0.30071E 17	0.74963E-04	0.69942E 00
15	0.39620E-03	0.47000E 03	0.14088E 00	0.22072E 00	0.31835E 17	0.74936E-04	0.63829E 00
16	0.42450E-03	0.34000E 03	0.10251E 00	0.16962E 00	0.43605E 17	0.75377E-04	0.60437E 00
17	0.45280E-03	0.18500E 03	0.59567E-01	0.95781E-01	0.86312E 17	0.80495E-04	0.62191E 00
18	0.48110E-03	0.43000E 02	0.19714E-01	0.16325E-01	0.20774E 19	0.11462E-03	0.12076E 01
19	0.50940E-03	0.21000E 02	0.11344E-01	0.54342E-02	0.12896E 20	0.13505E-03	0.20875E 01
20	0.53770E-03	0.15500E 02	0.87524E-02	0.29969E-02	0.24289E 20	0.14117E-03	0.29204E 01
21	0.56600E-03	0.14500E 02	0.82497E-02	0.23354E-02	0.31670E 20	0.14224E-03	0.35325E 01
22	0.59430E-03	0.15000E 02	0.85000E-02	0.20832E-02	0.35764E 20	0.14167E-03	0.40802E 01
23	0.62260E-03	0.17000E 02	0.94919E-02	0.21045E-02	0.35380E 20	0.13959E-03	0.45104E 01
24	0.65090E-03	0.20000E 02	0.10896E-01	0.21788E-02	0.34097E 20	0.13621E-03	0.50011E 01
25	0.67920E-03	0.25500E 02	0.13248E-01	0.24534E-02	0.30052E 20	0.12988E-03	0.53999E 01
26	0.70750E-03	0.35000E 02	0.16898E-01	0.29462E-02	0.24735E 20	0.12070E-03	0.57354E 01
27	0.73580E-03	0.52000E 02	0.22629E-01	0.37534E-02	0.19118E 20	0.10879E-03	0.60289E 01
28	0.76410E-03	0.85000E 02	0.32500E-01	0.51630E-02	0.13618E 20	0.95588E-04	0.62947E 01
29	0.79240E-03	0.16000E 03	0.52733E-01	0.81585E-02	0.74525E 19	0.82396E-04	0.64636E 01
30	0.82070E-03	0.33000E 03	0.99617E-01	0.14230E-01	0.26752E 19	0.75467E-04	0.70003E 01
31	0.84900E-03	0.82000E 03	0.25125E 00	0.32171E-01	0.59561E 18	0.76600E-04	0.78098E 01
32	0.87730E-03	0.26000E 04	0.87041E 00	0.10359E 00	0.78594E 17	0.83694E-04	0.84023E 01
33	0.90560E-03	0.76000E 04	0.29050E 01	0.29666E 00	0.22360E 17	0.95559E-04	0.97923E 01
34	0.93390E-03	0.20000E 05	0.87617E 01	0.60519E 00	0.95399E 16	0.10952E-03	0.14478E 02
35	LAST LAYER AT JUNCTION, SPREADING RESISTANCE VALUE TAKEN AS INFINITY						

TABLE III. Evaluations based on multi-layer theory. Iterative approach.

SAMPLE NUMBER AR-2197 N(111) TYPE

PROBE SPACING = 0.6350E-01 CMS.
STEP SIZE = 0.2830E-04 CMS.
MATERIAL TYPE IS N111

NUMBER OF LAYER	DEPTH FROM SURFACE	SPREADING RESISTANCE	RESISTIVITY	CORRECTED RESISTIVITY	IMPURITY CONCENTRATION	EFFECTIVE RADIUS OF CONTACT	CORRECTION FACTOR
1	0.00000E-38	0.75000E 03	0.22684E 00	0.28038E 00	0.23920E 17	0.77675E-04	0.80903E 00
2	0.28300E-04	0.66000E 03	0.19792E 00	0.22582E 00	0.30977E 17	0.75484E-04	0.87645E 00
3	0.56600E-04	0.62000E 03	0.18588E 00	0.21000E 00	0.33785E 17	0.74997E-04	0.88514E 00
4	0.84900E-04	0.58000E 03	0.17388E 00	0.18521E 00	0.39257E 17	0.74949E-04	0.93884E 00
5	0.11320E-03	0.58000E 03	0.17388E 00	0.18815E 00	0.38524E 17	0.74952E-04	0.92415E 00
6	0.14150E-03	0.57000E 03	0.17089E 00	0.18319E 00	0.39775E 17	0.74948E-04	0.93285E 00
7	0.16980E-03	0.57000E 03	0.17089E 00	0.18458E 00	0.39418E 17	0.74949E-04	0.92584E 00
8	0.19810E-03	0.57000E 03	0.17089E 00	0.18630E 00	0.38982E 17	0.74951E-04	0.91727E 00
9	0.22640E-03	0.57000E 03	0.17089E 00	0.18851E 00	0.38437E 17	0.74954E-04	0.90652E 00
10	0.25470E-03	0.57000E 03	0.17089E 00	0.19149E 00	0.37724E 17	0.74959E-04	0.89242E 00
11	0.28300E-03	0.57000E 03	0.17089E 00	0.19602E 00	0.36685E 17	0.74968E-04	0.87180E 00
12	0.31130E-03	0.57000E 03	0.17089E 00	0.20336E 00	0.35107E 17	0.74984E-04	0.84031E 00
13	0.33960E-03	0.56000E 03	0.16789E 00	0.21011E 00	0.33765E 17	0.75038E-04	0.79909E 00
14	0.36790E-03	0.54000E 03	0.16192E 00	0.22841E 00	0.30558E 17	0.75656E-04	0.70889E 00
15	0.39620E-03	0.47000E 03	0.14088E 00	0.22712E 00	0.30765E 17	0.75500E-04	0.62028E 00
16	0.42450E-03	0.34000E 03	0.10251E 00	0.18902E 00	0.38313E 17	0.74953E-04	0.54235E 00
17	0.45280E-03	0.18500E 03	0.59567E-01	0.13027E 00	0.59771E 17	0.74868E-04	0.45725E 00
18	0.48110E-03	0.43000E 02	0.19714E-01	0.36694E-01	0.46746E 18	0.92222E-04	0.53726E 00
19	0.50940E-03	0.21000E 02	0.11344E-01	0.17663E-01	0.17969E 19	0.11914E-03	0.64227E 00
20	0.53770E-03	0.15500E 02	0.87524E-02	0.61949E-02	0.11218E 20	0.14503E-03	0.14128E 01
21	0.56600E-03	0.14500E 02	0.82497E-02	0.39936E-02	0.17897E 20	0.14274E-03	0.20657E 01
22	0.59430E-03	0.15000E 02	0.85000E-02	0.30780E-02	0.23609E 20	0.14074E-03	0.27615E 01
23	0.62260E-03	0.17000E 02	0.94919E-02	0.30443E-02	0.23888E 20	0.14066E-03	0.31179E 01
24	0.65090E-03	0.20000E 02	0.10896E-01	0.29071E-02	0.25089E 20	0.14026E-03	0.37483E 01
25	0.67920E-03	0.25500E 02	0.13248E-01	0.30554E-02	0.23795E 20	0.14069E-03	0.43359E 01
26	0.70750E-03	0.35000E 02	0.16898E-01	0.33973E-02	0.21256E 20	0.14158E-03	0.49739E 01
27	0.73580E-03	0.52000E 02	0.22629E-01	0.38271E-02	0.18726E 20	0.14250E-03	0.59127E 01
28	0.76410E-03	0.85000E 02	0.32500E-01	0.45396E-02	0.15616E 20	0.14374E-03	0.71592E 01
29	0.79240E-03	0.16000E 03	0.52733E-01	0.60982E-02	0.11408E 20	0.14499E-03	0.86474E 01
30	0.82070E-03	0.33000E 03	0.99617E-01	0.80901E-02	0.75688E 19	0.14253E-03	0.12313E 02
31	0.84900E-03	0.82000E 03	0.25125E 00	0.22132E-01	0.11861E 19	0.10909E-03	0.11352E 02
32	0.87730E-03	0.26000E 04	0.87041E 00	0.11736E 00	0.67707E 17	0.74932E-04	0.74165E 01
33	0.90560E-03	0.76000E 04	0.29050E 01	0.38738E 00	0.16257E 17	0.79189E-04	0.74992E 01
34	0.93390E-03	0.20000E 05	0.87617E 01	0.76219E 00	0.72420E 16	0.82634E-04	0.11495E 02
35	LAST LAYER AT JUNCTION, SPREADING RESISTANCE VALUE TAKEN AS INFINITY						

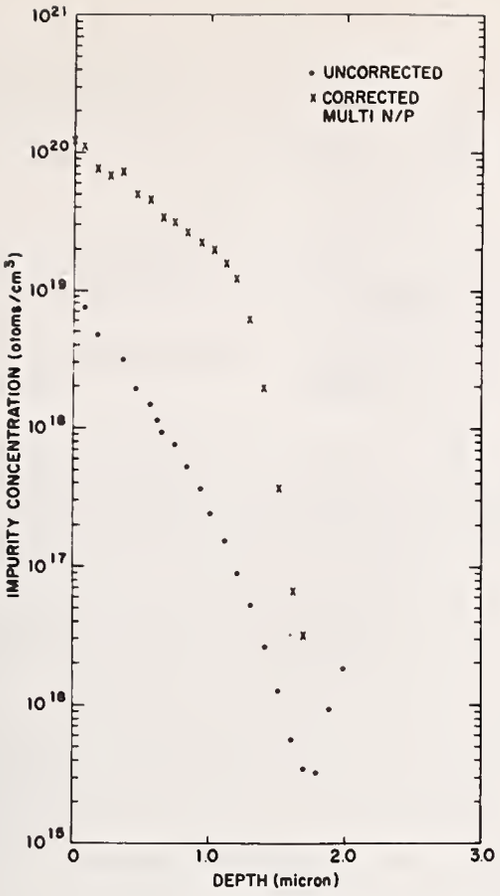


Fig. 13. Impurity concentration as a function of depth for N^+/P^- structure.

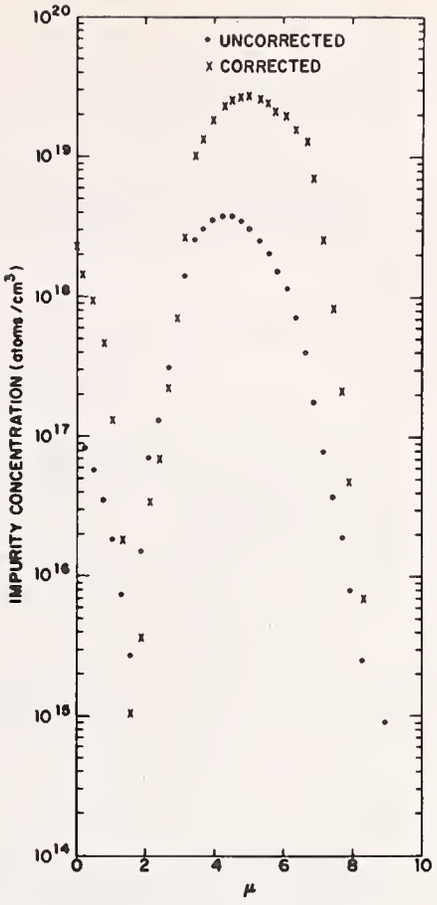


Fig. 14. Impurity concentration as a function of depth for a P-type diffused layer in a N/N+/P⁻ structure.

References

[1] R. G. Mazur and D. H. Dickey, J. Electrochem. Soc. Vol. 113, 255 (1966).

[2] E. E. Gardner, P. A. Schumann, Jr., and E. F. Gorey, Electrochemical Society Symposia Proceedings "Measurement Techniques for Thin Films April 1967.

[3] D. C. Gupta, ibid, Vol. 116, 670 (1969).

[4] C. K. Chu, ibid, Vol. 115, 192C (1968).

[5] P. A. Schumann, Jr., J. M. Adley, M. R. Poponiak, C. P. Schneider and A. H. Tong, ibid, Vol. 116, 150C (1969).

[6] D. C. Gupta, J. Y. Chan, and P. Wang, ibid, Vol. 116, 301C (1969).

[7] E. Tannenbaum, Solid State Electronics Vol. 2, 123 (1961).

[8] J. Brownson, J. Electrochem. Soc. Vol. 111, 919 (1964).

[9] E. E. Gardner and P. A. Schumann, Jr., Solid State Electronics Vol. 8, 165 (1965).

[10] J. Hilibrand and R. D. Gold, RCA Review Vol. 21, 245 (1960).

[11] C. O. Thomas, O. Kahng and R. C. Manz, J. Electrochem. Soc. Vol. 109, 1055 (1962).

[12] R. Holm, Electrical Contacts Handbook, Springer-Verlag, Berlin, 1958.

[13] E. E. Gorey, C. P. Schneider, M. R. Poponiak, J.Electrochem. Soc. Vol. 117, 721 (1970).

[14] R. G. Mazur, J. Electrochem. Soc. Vol. 116, 149C (1969).

[15] R. G. Mazur, ibid, Vol. 117, 95C (1970).

[16] D. C. Gupta and J. Y. Chan, The Review of Scientific Instruments Vol. 41, p. 176 (1970).

[17] P. A. Schumann, Jr. and E. E. Gardner, J. Electrochem. Soc. Vol. 116, p. 88, (1969).

[18] T. H. Yeh and K. H. Khokhani, to be published.

[19] P. A. Schumann, Jr. and E. E. Gardner, Solid State Electronics Vol. 12, p. 371 (1969).

[20] T. H. Yeh and K. H. Khokhani, J. Electrochem. Soc. Vol. 116, 1461 (1969).

[21] J. C. Irvin, Bell System Tech. Journal Vol. 41, 387 (1962).

Incremental Sheet Resistivity Technique for Determining Diffusion Profiles

R. P. Donovan and R. A. Evans

Research Triangle Institute
Post Office Box 12194
Research Triangle Park, N. C. 27709

Measurements of incremental sheet resistivity constitute a straightforward technique for determining impurity distribution following diffusion. The measurements require no special apparatus or equipment not commonly found in a semiconductor laboratory; they employ procedures that silicon technologists are familiar with. The method is destructive, however, and tedious. It is the type of measurement that can be employed on a sampling basis only. The method consists simply of measuring the sheet resistivity of the diffused layer, removing a thin uniform increment from the surface, remeasuring the sheet resistivity and repeating the procedure until the desired depth is reached. The success of the method depends upon the ability to: 1) accurately measure sheet resistivity of the wafer or sample, and 2) uniformly remove a measurable increment of silicon from the surface. Uncertainties in the method exist because of the assumption that carrier mobility in the thin increments removed is the same as the carrier mobility measured in bulk samples. For many diffused layers this assumption is valid. When the assumption is questionable, the sheet resistivity measurements can be easily supplemented by measurements of sheet Hall coefficient. Incremental measurements of sheet resistivity and Hall coefficient together yield profiles of both carrier concentration and mobility and have been valuable in evaluating the influence of annealing upon the electrical properties of ion implanted layers.

Key Words: Anodic oxidation, four point probe, impurity profile, incremental sectioning, mobility determination, resistivity measurements, sheet Hall coefficient, sheet resistivity.

1. Introduction

The incremental sheet resistivity method for determining impurity profiles in silicon was employed initially by Fuller and Ditzenberger [Ref. 1] and has been popularized by Tannenbaum [Ref. 2].

The method consists of measuring the sheet resistivity, removing a thin uniform increment, remeasuring the sheet resistivity and repeating the procedure until the profile is complete. Figure 1 illustrates the method.

1.1 Theory

The relationship between the bulk conductivity (≡ reciprocal bulk resistivity) at any depth and the sheet conductivity (≡ reciprocal sheet resistivity) is:

$$\sigma_s(x) \equiv \int_x^{x_j} \sigma(x)\,dx \qquad (1)$$

x is the depth measured perpendicularly to the surface;

$[\rho_s(x)]^{-1} \equiv \sigma_s(x)$ is the sheet conductivity at plane x;

$[\rho(x)]^{-1} \equiv \sigma(x)$ is the bulk conductivity at x;

x_j is the bottom boundary of the conducting layer.

Differentiating eq. 1 yields

$$\sigma(x) = -\frac{d\sigma_s(x)}{dx} \qquad (2)$$

The derivative $\frac{d\sigma_s(x)}{dx}$ can be approximated from measured data by

$$\frac{d\sigma_s}{dx} \simeq \frac{\Delta(\sigma_s)_i}{\Delta x_i} = \frac{(\sigma_s)_{i-1} - (\sigma_s)_i}{x_{i-1} - x_i} = -\bar{n}_i q \bar{\mu}_i \qquad (3)$$

where

Δx_i is the thickness of the i*th* increment removed (x_i is the distance from the original surface to the surface after removal of i*th* increment)

$\Delta(\sigma_s)_i$ is the change in sheet conductivity brought about by the removal of the i*th* increment $(\sigma_s)_i$ is the sheet conductivity measured after the removal of the i*th* increment.

$\bar{n}_i$ is the average carrier concentration in the i*th* increment

$\bar{\mu}_i$ is the average mobility of carriers in the i*th* increment

The numerator of eq. 3 is always positive, while the denominator is negative. Assuming that the carrier concentration in the increment removed is equal to the net impurity concentration ($n \simeq |N_D - N_A|$) eq. 3 becomes;

$$|N_D - N_A| = -\frac{\Delta(\sigma_s)_i}{q\bar{\mu}_i \Delta x_i} \qquad (4)$$

N_D = donor impurity concentration

N_A = acceptor impurity concentration

The only unknown on the right-hand side of eq. 4 is $\bar{\mu}_i$. This value can be estimated from an empirical curve of conductivity mobility versus impurity concentration such as shown in Fig. 2 [Ref.3]. Initially one guesses what the impurity concentration is and uses the corresponding mobility value to calculate a net impurity concentration from eq. 4. If the calculated value corresponds to the guessed value, the problem is solved. Most likely the values differ. In this case the mobility value corresponding to the value of impurity concentration just calculated is used in eq. 4 to calculate a new value of impurity concentration. By this iterative process, calculated values of $\bar{\mu}_i$ and $|N_D - N_A|$ giving a satisfactory fit to the measured values of $\Delta(\sigma_s)_i$ and Δx_i are obtained.

As evident in eq. 4, the calculated value of net impurity concentration depends solely on the difference between successive measurements of sheet resistivity and measurements of the incremental thicknesses removed. How one chooses to derive these values from the measured data can influence the result. Consider the data presented in Fig. 3. The raw data are those labeled sheet conductivity, σ_s. If one simply followed the procedure outlined previously, he would have a curve of high unrealistic scatter. When the sheet conductivity does not change between successive removals, the conclusion must be that the impurity concentration is zero. Rather than accept this kind of scatter in points through which the impurity profile is drawn, an averaging technique utilizing several nearest neighbor points to determine slope is employed or else a curve is drawn through the entire set of points, using some method of determining the best fit. This has been done in reducing the data of Fig. 3 as indicated by the dashed line drawn through the data points. The consequent impurity distribution curve is also shown. Note the discontinuity in the calculated impurity distribution that appears at 0.8μm. This discontinuity reflects the boundary between the region of constant slope and that of continuously changing slope. As evident in eqs. 2 and 4, the impurity concentration depends only on the slope $d\sigma_s(x)/dx$.

An alternative expression for determining impurity concentration from measurements of incremental sheet resistivity results from plotting the data on semilog paper [Ref. 5]. The method retains the simplicity and speed of that just described but involves the measured value of sheet resistivity directly. From eq. 2,

$$\rho(x)^{-1} = -\frac{d[\rho_s(x)]^{-1}}{dx} = \frac{d\ell n[\rho_s(x)]^{-1}}{\rho_s dx} = \frac{d\ell n\rho_s(x)}{\rho_s dx} = \frac{1}{0.4343\rho_s}\frac{d \log \rho_s(x)}{dx} \qquad (5)$$

The major variable on the right-hand side of eq. 5 is the measured sheet resistivity itself. In a typical profile it changes by nearly as many orders of magnitude as the impurity concentration itself, while the derivative changes less than 1 order, usually.

The procedure for converting incremental sheet resistivity measurements into profiles of impurity concentration is as follows:

1) Plot log ρ_s (s) vs. x; fair a curve through these points.

2) Calculate the slope of this curve, $d(\log \rho_s)/dx$ and plot it vs. x; fair a curve through the points.

3) Calculate $\rho(x)$ from the relationship $\rho = \dfrac{0.4343\ \rho_s}{d(\log \rho_s)/dx}$.

4) From a curve of resistivity versus impurity concentration (such as published by Irvin [Ref. 6])find the net impurity concentration, $|N_D - N_A|$, corresponding to each value of x. Plot log $|N_D - N_A|$ versus x and fair a curve through the points. This is the desired profile of impurity concentration versus depth.

Figure 4 shows a plot of both sheet resistivity and the corresponding impurity concentration as calculated by this procedure, using the same data as used in Fig. 3. Figure 5 shows the slope values (measured from Figure 4) that were used to calculate the impurity concentration curve. When plotted by this procedure, the impurity profile differs from that shown in Fig. 3. The data points used are the same, and the same analysis has been used. What has been changed are the data smoothing procedures. The iterative determination of mobility has also been avoided by using Irvin's curves but this is merely a labor-saving convenience rather than any improvement in accuracy. It does not change the results.

In both procedures the mobility values employed are determined by using the relationship between bulk impurity doping levels and mobility. The implicit assumption is that at high impurity concentrations mobility is determined primarily by the impurity concentration. Carrier scattering caused by structural damage, precipitates or the surface can cause variations in the relationship between mobility and impurity scattering. The significance of surface scattering is particularly important to consider in the analysis of thin increments such as described by Tannenbaum [Ref. 2]. Surface scattering is most noticeable when removing thin, lightly-doped sections of silicon. Figure 6 (reproduced from Tannenbaum) compares the impurity profile determined radioactively with that determined by the incremental resistivity method. Below about 1×10^{20} cm^{-3}, the curves essentially coincide. The increments used in this sectioning were about 40 nm (400 Å) thick. For this thickness and concentration level, (10^{18} - 10^{20} cm^{-3}), surface scattering is unimportant. As the incremental thickness is reduced and as the doping level decreases, more uncertainty arises. For MOS transistors the mobility of carriers in the surface channel is approximately half their value in the bulk [Ref. 7]. This effective surface channel mobility is relatively independent of perpendicular electric field up to 1.5×10^5 volts/cm. The doping level characteristic of these channels is 10^{15} - 10^{16} cm^{-3} and the maximum channel thickness is on the order of 50 to 100 nm (500 to 1000 Å) [Ref. 8]. Most carriers, however, move in a layer much narrower than this maximum width. At doping levels above 10^{18} cm^{-3}, the maximum channel width is less than 5 nm (50 Å).

For the commonly investigated doping levels and the incremental thicknesses used, the assumption of bulk mobility values does not introduce additional error. A far more significant problem is that the bulk mobility itself is not well known at high values of impurity concentration and many impurities enter the silicon crystal in electrically inactive sites. The consequence is a large difference between the impurity profiles determined by resistivity and those determined by radioactivity at high doping levels (Fig. 6). When the carrier concentration no longer equals the net impurity concentration, the method breaks down and one is no longer plotting impurity concentration but carrier concentration.

1.2 Summary of Method

1. The method yields carrier concentration vs. depth which under certain conditions is impurity concentration vs. depth.
2. It assumes that the dominant carrier scattering mechanism in each increment is the same as in the bulk silicon samples used to prepare empirical curves of mobility or resistivity vs. impurity concentration. This statement in turn implies negligible impurity compensation since the published curves are restricted to such samples.

2. Techniques

The success of the method depends upon the ability to: 1) accurately measure the sheet resistivity at each increment; 2) uniformly remove a measurable increment of silicon.

2.1 Sheet Resistivity

To determine sheet resistivity, the collinear four point probe method is most commonly employed [Ref. 9]. This technique is a standard laboratory procedure and with suitable correction factors for sample geometry can be used to derive sheet resistivity directly. These procedures and precautions are well documented [Ref. 9, 10].

A less standard but increasingly popular geometry for measuring sheet resistivity is that due to Van der Pauw [Refs. 11, 12]. Van der Pauw's theorem shows that the sheet resistivity of an arbitrarily shaped specimen can be determined by two measurements of resistivity using four, small peripheral contacts as shown in Fig. 7. The sample thickness is assumed to be uniform and the doping, homogeneous. When a current I, is passed through contacts a and b, a voltage V, appears across contacts d and c. The ratio of the voltage to the current is defined as a resistance R_1. A similar resistance R_2 is determined by passing a current I_2 through contacts b and c and measuring the voltage V_2 across contacts a and d. In Ref. 10 the following relationship is shown to exist:

$$\exp\left(-\frac{\pi R_1}{\rho_s}\right) + \exp\left(-\frac{\pi R_2}{\rho_s}\right) = 1 \qquad (6)$$

The only unknown in eq. 6 is ρ_s. Equation 6 can be expressed as:

$$\rho_s = \frac{\rho}{x_j - x_i} = \frac{\pi}{\ell n 2}\,\frac{R_1 + R_2}{2}\, f\left(\frac{R_1}{R_2}\right) \qquad (7)$$

where f (R_1/R_2) is graphed in Fig. 8. When there is a line of symmetry, two contacts can be placed on the line and the other two placed symmetrically with respect to it. In this case $R_1 = R_2$ and f = 1 so that eq. 7 reduces to the familiar four-point probe relation:

$$\rho_s = 4.5324\,\frac{V}{I}\,. \qquad (8)$$

Other non-peripheral probe geometries can be used along with appropriate correction factors [Ref. 13]. A square four point probe array is one such possibility. Such modified geometries permit resistivity measurements to be made on odd shaped chips or small geometries--samples that might be difficult to evaluate by the collinear four-point probe. Aside from this consideration, the major reason for seeking a resistivity measuring method other than the collinear four-point probe is that more data than sheet resistivity are desired. An argument will be made for including sheet Hall coefficient measurements in the analysis. For such measurements a geometry capable of measuring both sheet resistivity and Hall coefficient is desirable.

2.2 Incremental Sectioning

The second necessary technique for successful impurity profile determination is uniform removal of thin increments of silicon. Many methods have been investigated including lapping [Refs. 1, 14], and etching [Refs. 15, 16]. Lapping requires considerable technique and skill [Ref. 1]; it is better suited for sectioning deep 25-50 μm junctions than for shallow 1 μm junctions. Damage caused by the lapping operation can introduce mobility errors [Ref. 14].

Etching avoids mechanical damage and with care can be quite satisfactory for uniform layer removal [Ref. 15]. One extreme of control is the technique of boiling the silicon surface to be sectioned in water and then using hydrofluoric acid to remove the thin oxide grown by reaction of silicon with the water. This procedure has been calibrated to remove 3.4 nm (34 Å) of silicon per boiling [Ref. 16]. (This number reduces to 0.5 nm (5 Å) per boiling within 20 nm (200 Å of a junction).

The sectioning technique that put the incremental sheet resistivity method within the reach of everyone, however, is the anodic oxidation technique first used for this purpose by Tannenbaum [Ref. 2]. Anodic oxidation incorporates the control of the boiling water--HF method but makes possible the uniform, reproducible removal of silicon increments between 10 and 100 nm (100-1000 Å) or slightly greater. Even so, sectioning of surfaces to depths greater than the 2-4 μm depth is extremely time consuming and tedious.

Methods of anodic oxidation of silicon are well known [Refs. 17, 18]. Variation in the density of anodic oxide (and hence in quantity of silicon anodized per thickness of anodic oxide) exists but can easily be determined empirically by measuring both the thickness of the oxide and the depth of silicon used to form the oxide. Anodization in a hot (100°C) solution of 10 gm of ammonium nitrate in 3,000 gm of tetrahydrofurfuryl alcohol produces an oxide density substantially the same as that of thermal oxide. For thermal oxide the thickness of silicon removed corresponds to 0.44 the thickness of the anodic oxide.

3. Sheet Hall Coefficient Measurements

Including a measurement of sheet Hall coefficient in the sectioning analysis requires only the addition of a magnetic field when using a measuring geometry such as that of Van der Pauw or the square four-point probe array [Ref. 12]. Referring again to Fig. 7, the Van der Pauw theorem further states that if current is passed between any two alternate peripheral contacts on the edge of an arbitrarily shaped sample, the change in potential between the other two contacts brought about by the presence of a magnetic field normal to the sample surface is the Hall voltage. Consequently, the sheet Hall coefficient R_s is:

$$R_s = \frac{R_H}{x_j - x_i} = \frac{\Delta V_{a,c}}{I_{b,d} B} \tag{8}$$

where R_s is the sheet Hall coefficient; R_H is the bulk Hall coefficient; $\Delta V_{a,c}$ is the change in V_{ac} upon turning on the magnetic field; $I_{b,d}$ is the current between b and d; B is the magnetic induction; and x_j and x_i are defined in Fig. 1. Equation 8 is the same as that describing a rectangular sample with all fields uniform--the most simple boundary condition to solve.

Having an independent measure of carrier concentration now makes possible a measure of mobility along with carrier concentration when the sectioning analysis includes both sheet resistivity and sheet Hall coefficient measurements. The following expressions are used to calculate mobility and carrier concentration from measured values of sheet resistivity and Hall coefficient [Refs. 19 and 20]:

$$n(x) = \left(\frac{d\sigma_s}{dx}\right)^2 \Big/ q \left[\frac{d(R_s\sigma_s^2)}{dx}\right] = \frac{d\sigma_s}{dx} \Big/ q\,\mu_H(x) \tag{9}$$

$$\mu_H(x) = \frac{d(R_s\sigma_s^2)}{dx} \Big/ \frac{d\sigma_s}{dx} \tag{10}$$

These relations follow from the definitions of sheet conductivity and sheet Hall coefficient developed by Petritz [Ref. 21]:

$$R_s = \int n(x) < \mu(x)^2 > dx \Big/ q\left[\int n(x) < \mu(x) > dx\right]^2 \tag{11}$$

$$\sigma_s = q \int n(x) < \mu(x) > dx. \tag{12}$$

By combining 11 and 12,

$$R_s\sigma_s^2 = q \int n(x) < \mu(x)^2 > dx. \tag{13}$$

Differentiating eqs. 12 and 13 and solving for n(x) and μ(x) yields eqs. 9 and 10.

Again, the slopes required in equations 9 and 10 may be determined by plotting the data points on semilog scales so that the measured values of the quantities themselves are the most important variables of eqs. 9 and 10 rather than their slopes:

$$\frac{d\sigma_s}{dx} = \sigma_s \frac{d\ell n\sigma_s}{dx} \simeq (\sigma_s)_i \frac{\Delta\ell n(\sigma_s)_i}{\Delta x_i} \tag{14}$$

$$\frac{d(R_s\sigma_s^2)}{dx} = (R_s\sigma_s^2) \frac{d\ell n(R_s\sigma_s^2)}{dx} \simeq (R_s\sigma_s^2)_i \frac{\Delta\ell n(R_s\sigma_s^2)_i}{\Delta x_i} \tag{15}$$

In these expressions the difference between Hall mobility and conductivity mobility has been neglected.

In addition to determining a measured value of mobility for each increment removed, and hence, removing an uncertainty from the incremental sheet resistivity method, the values of n and μ can be used to estimate compensation. The method assumes that $n = |N_D - N_A|$ and that the impurity concentration dependence of μ is a measure of $N_D + N_A$. The specific value of $N_D + N_A$ is derived from a slot of Hall mobility vs. impurity concentration in which an impurity concentration corresponding to the measured Hall mobility value is read out. This concentration is set equal to $N_D + N_A$. Having both $N_D + N_A$ and $|N_D - N_A|$ allows calculation of both N_D and N_A separately.

The uncertainty in the relationship between Hall mobility and impurity concentration makes the measurement of $(N_D + N_A)$ by mobility measurements subject to large error. The method does satisfactorily explain certain data, however, as shown in Fig. 9. The profile (of an ion implanted bismuth layer) determined from resistivity measurements is that shown in circles. It differs considerably from what results by plotting the $N_D + N_A$ profile (shown by the dashed curve). By combining these two distributions, the separate plots of both N_D and N_A can be constructed as shown in Fig. 9. Making both sheet Hall coefficient and resistivity measurements relieves the need for assuming negligible compensation in the increments removed.

4. Conclusions

Incremental sheet resistivity measurements constitute a method for experimentally constructing a profile of the impurities in a given region. Plotting data on semilog plots rather than the conventional linear plots reduces the dependence of the results upon the procedures used to determine a slope. The procedures for carrying incremental sectioning out are well developed and relatively independent of operator technique. The basic method assumes that the layer removed has negligible compensation and that the carriers in the thin-surface increment drift with the same mobility as carriers in a similarly doped bulk specimen. For most analyses these assumptions are valid. When uncertainty exists as to their validity, supplementing measurements of sheet Hall coefficient, made concurrently with the sheet resistivity measurements, can clarify the uncertainty.

5. References

[1] C. S. Fuller and J. A. Ditzenberger, "Diffusion of Donor and Acceptor Elements in Silicon," J. Appl. Phys. 27, May 1956,pp.544-553.

[2] E. Tannenbaum, "Detailed Analysis of Thin Phosphorus-Diffused in p-type Silicon," Solid State Electronics 2, Mar/Apr. 1961, pp. 123-132.

[3] R. A. Evans, "Integrated Silicon Device Technology, Vol. 5, Physical/Electrical Properties"

[4] K. H. Nicholas, "Studies of Anomalous Diffusion of Impurities in Silicon," Solid State Electronics 9, Jan. 1966, pp. 35-47.

[5] R. A. Evans and R. P. Donovan, "Alternative Relationship for Converting Incremental Sheet Resistivity Measurements into Profiles of Impurity Concentration," Solid State Electronics 10, Jan. 1967, pp. 155-157.

[6] J. C. Irvin, "Resistivity of Bulk Silicon and of Diffused Layers in Silicon," Bell System Tech. J., March 1962, pp. 387-410.

[7] O. Leistiko, A. S. Grove and C. T. Sah, "Electron and Hole Mobilities in Inversion Layers on Thermally Oxidized Silicon Surfaces," IEEE Trans. Electron Devices ED-12, May 1965.

[8] J. W. Hauser, "Surface Field Effect Transistors" in Fundamentals of Silicon Integrated Device Technology, Vol. II, Bipolar and Unipolar Transistors, edited by R. M. Burger and R. P. Donovan, Prentice-Hall, Englewood Cliffs, N.W. (1968).

[9] 1968 Book of ASTM Standards (Part 8), Method F84-68T, pp. 729-743, ASTM, Philadelphia, Pa.

[10] F. M. Smits, "Measurement of Sheet Resistivity with the Four-Point Probe," Bell System Tech. J. 37, May 1958, pp. 711-718.

[11] L. J. Van der Pauw, "A Method of Measuring Specific Resistivity and Hall Effect of Diacs of Arbitrary Shape," Philips Research Repts. 13, Jan. 1958, pp. 1-9.

[12] L. J. Van der Pauw, "A Method of Measuring the Resistivity and Hall Coefficient of Lamellae of Arbitrary Shape," Philips Tech. Rev. 20 1958/59, pp. 220-224.

[13] M. G. Buehler and G. L. Pearson, "Magnetoconductive Correction Factors for An Isotropic Hall Plate with Point Sources," Solid State Electronics 9, 1966, pp. 395-407.

[14] H. Frank, "Influence of Damaged Surface Layer on Resistivity and Mobility of Thin Semiconductor Sheets," Solid State Electronics, 9, 1966, pp. 609-614.

[15] P. A. Iles and B. Leibenhaut, "Diffusant Impurity Concentration Profiles in Thin Layers on Silicon," Solid State Electronics 5, 1962, pp. 331-339.

[16] J. L. Moll, N. I. Meyer and D. J. Bartelink, "Hot Electron Emission from Silicon p-n Junctions Parallel to the Surface,", Phys. Rev. Letts. 7, Aug. 1, 1961, pp. 87-90.

[17] E. F. Duffek, C. Mylroie and E. A. Benjamini, "Electrode Reactions and Mechanism of Silicon Anodization in n-Methylacetamide", J. Electrochem. Soc. 111, Sept., 1964, pp. 1042-1046.

[18] K. M. Busen and R. Linzey, "Removal of Thin Layers of n-type Silicon by Anodic Oxidation", Trans. AIME 236, March, 1966, pp. 306-309.

[19] R. Baron, G. A. Shifrin, O. J. Marsh and J. W. Mayer,J. Appl. Phys. 40, Aug., 1969, 3702-3719.

[20] M.G. Buehler, Stanford Res. Rpt. SEL-66-064,1966.

[21] R. L. Petritz, "Theory of an Experiment for Measuring the Mobility and Density of Carriers in the Space Charge Region of a Semiconductor Surface", Phys. Rev., 110, June 15, 1958, pp. 1254-1262.

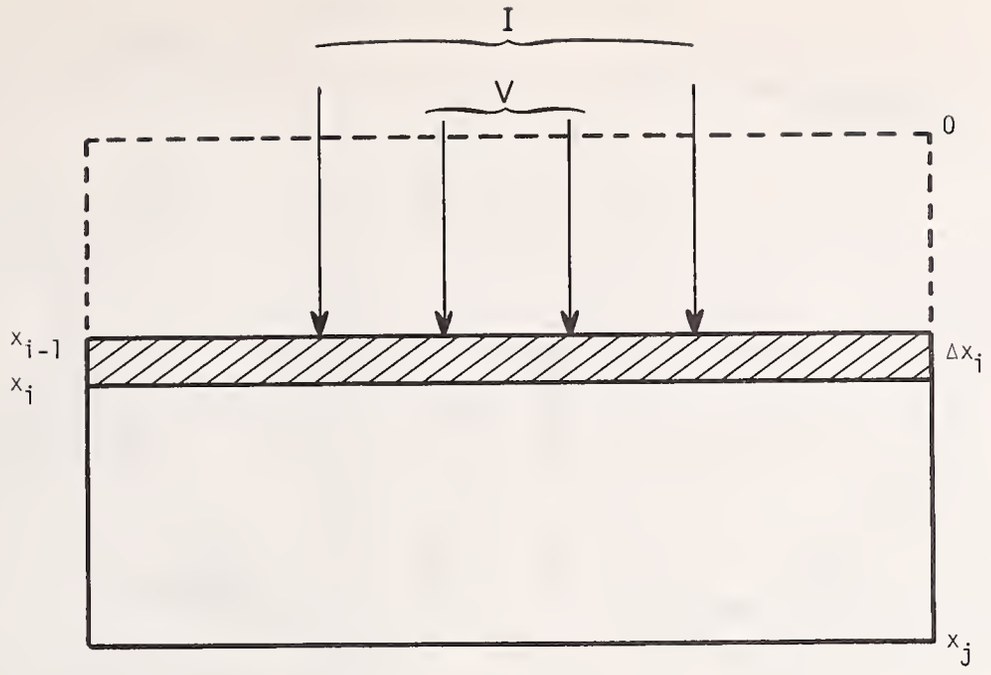


Fig. 1. Incremental Sheet Resistivity Method [Ref. 1]

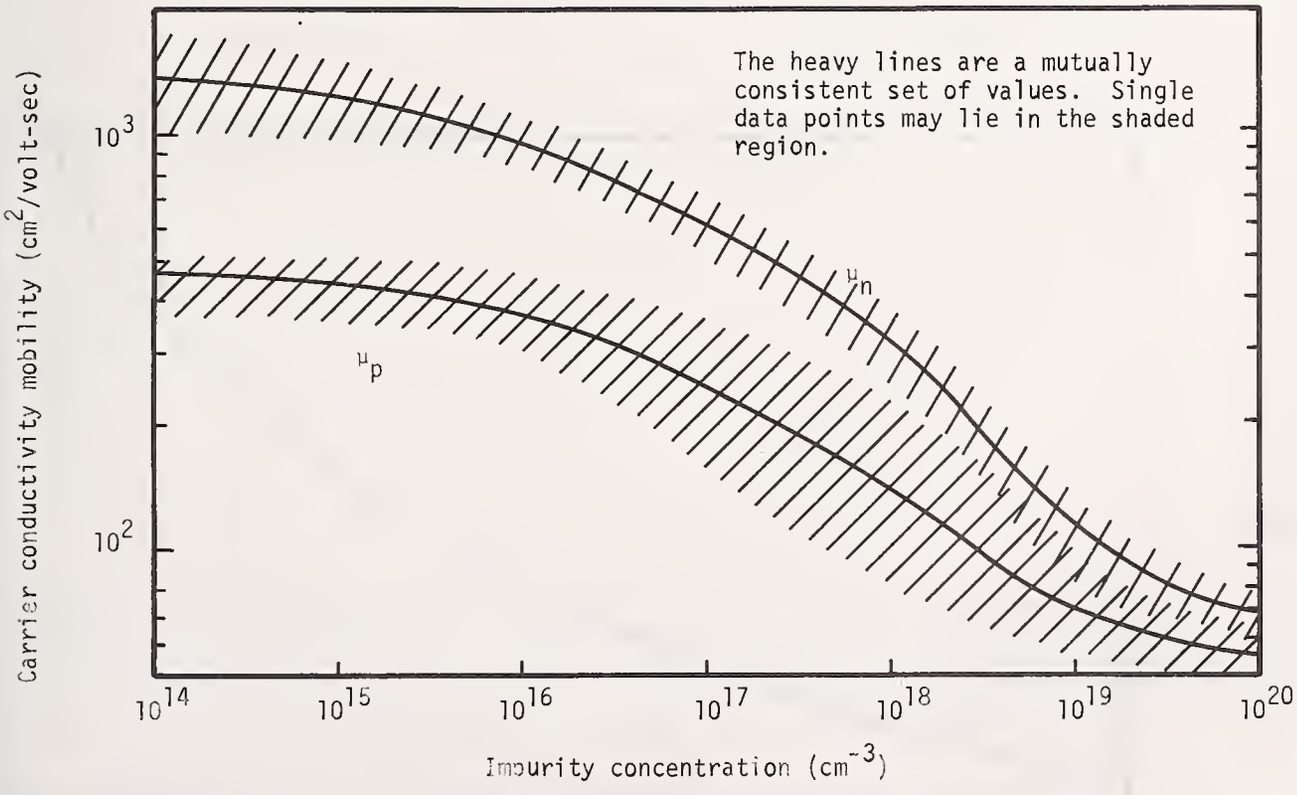


Fig. 2. Conductivity mobility of carriers at 300°K as a function of impurity concentration [Ref. 3]

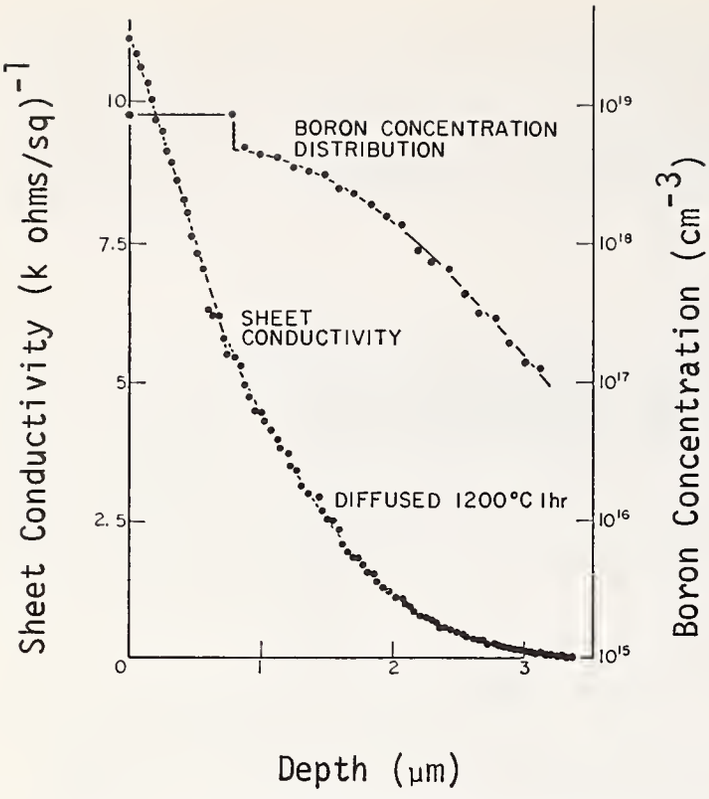


Fig. 3. Sheet conductivity profile with impurity profile determined from eq. 4. [Ref. 4]

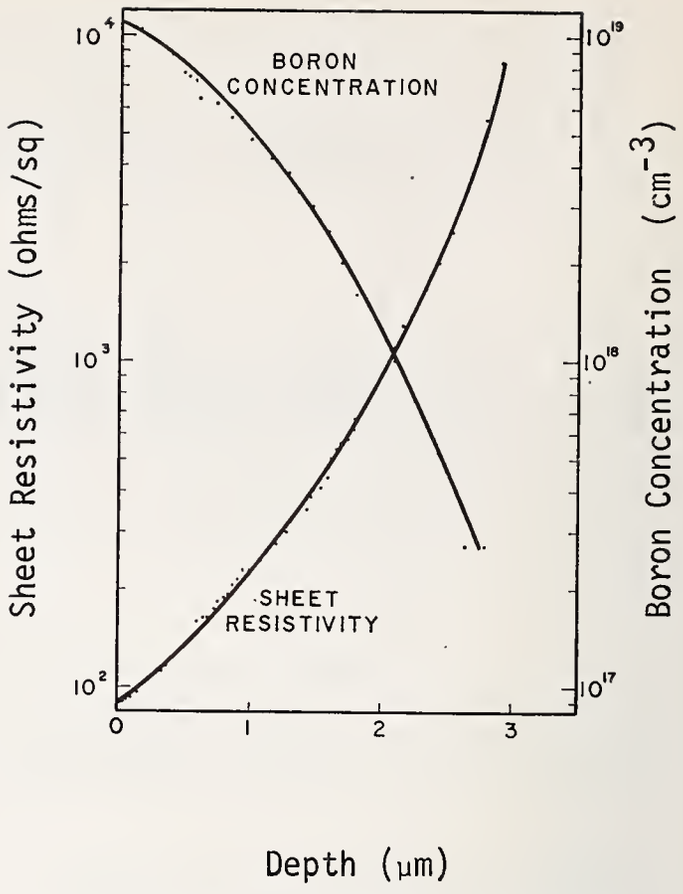


Fig. 4. Impurity profile constructed from the data points of Fig. 3 but using eq. 5 [Ref. 5]

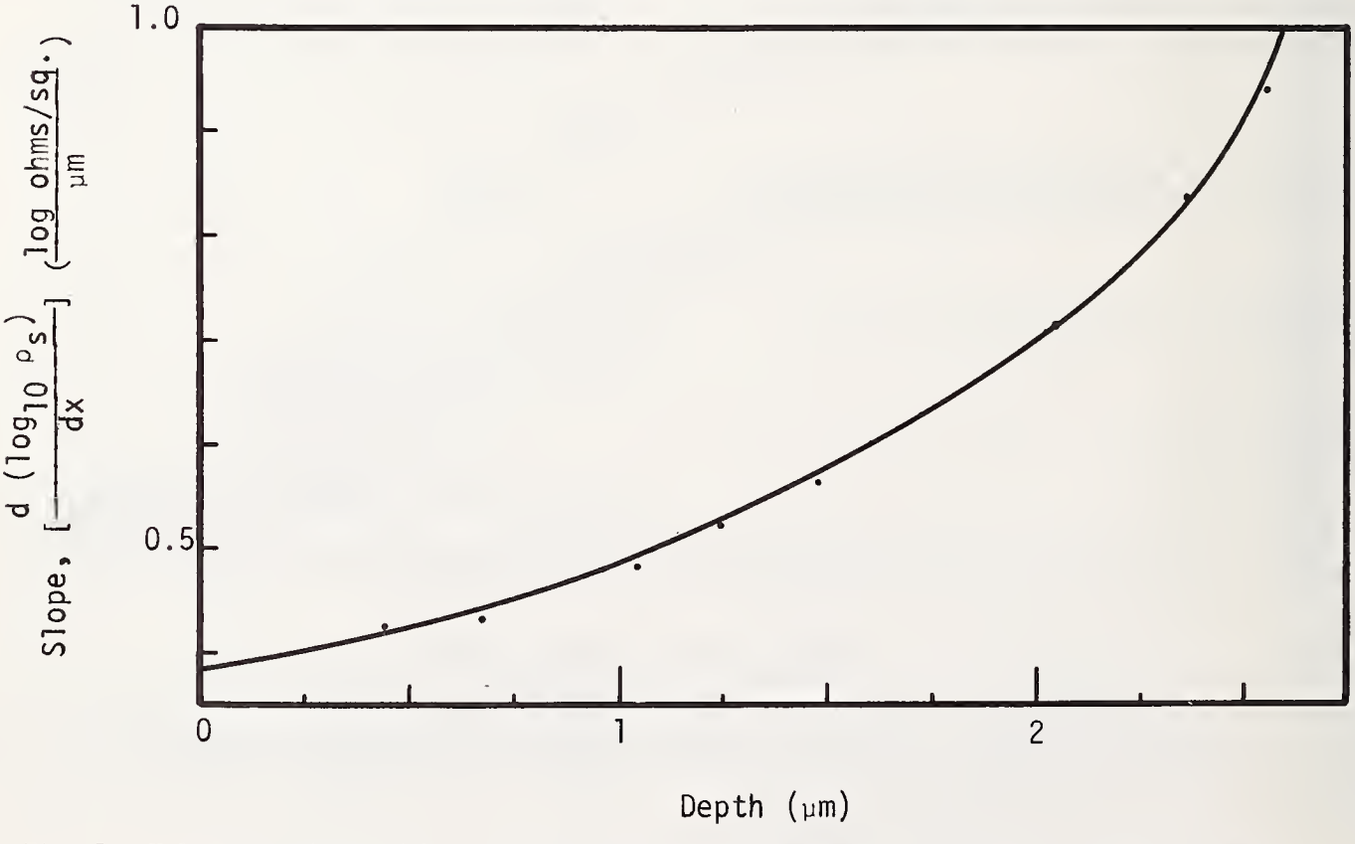


Fig. 5. Values of d(log ρ_s)/dx used in constructing the impurity profile of Fig. 4. [Ref. 3]

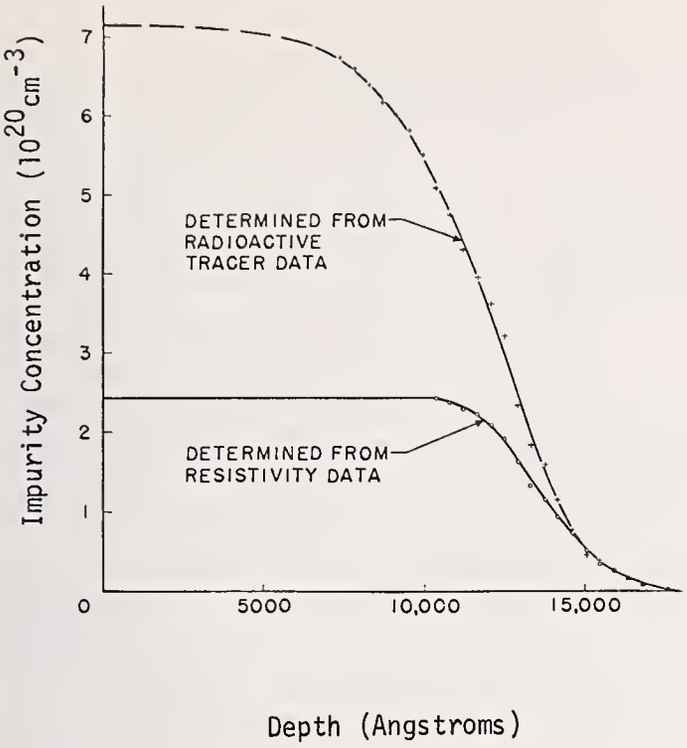


Fig. 6. Comparison of radioactively determined profile with that of the sample profile constructed from incremental sheet resistivity measurements. [Ref. 2]

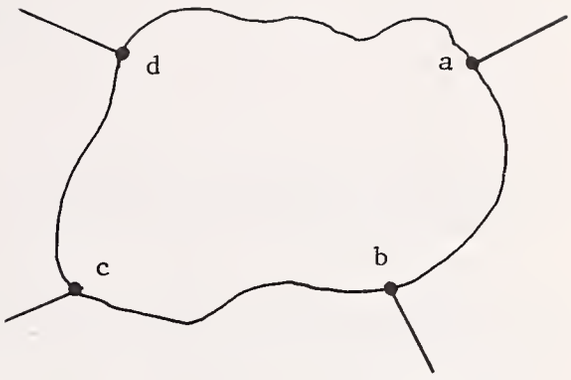


Fig. 7. Van der Pauw geometry. [Ref. 11]

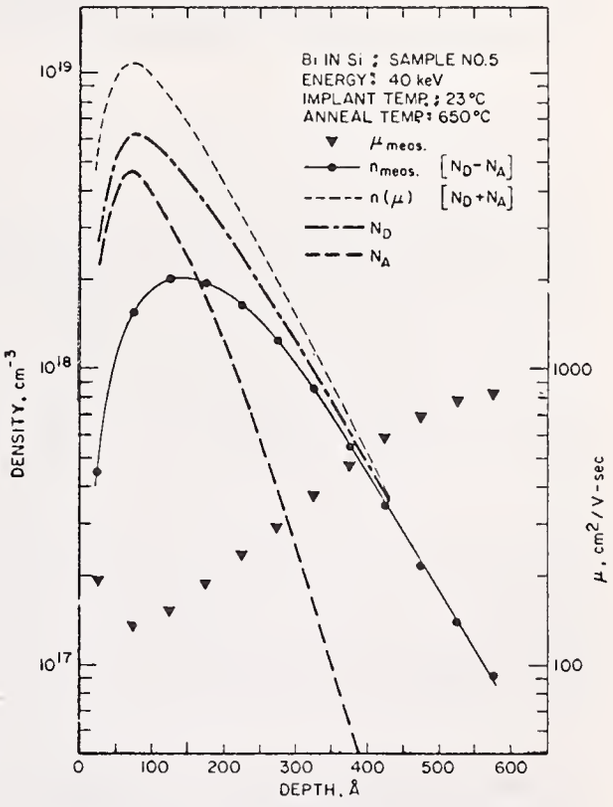
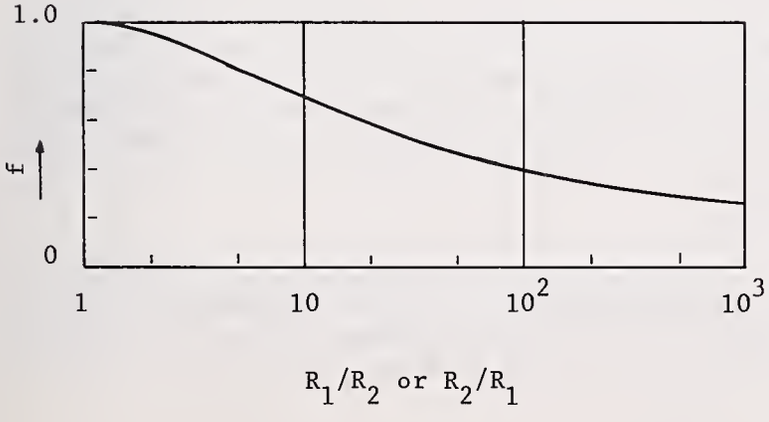


Fig. 9. Data and calculated curves from the incremental sectioning analysis (ρ_s and R_s) of an ion implanted bismuth layer in silicon. [Ref. 19]

Fig. 8. The function f. [Ref. 11]



Nuclear Methods for the Determination of Diffusion Profiles

B. J. Masters

IBM Components Division
East Fishkill Facility
Hopewell Junction, N. Y. 12533

Owing to the relative ease of detecting radioactive material in very small amounts, both radioactivation and radiotracer techniques are finding considerable application as means for obtaining impurity profiles in silicon. These radiochemical methods yield knowledge of the total concentration of specific impurities as a function of depth, regardless of the physical or chemical form in which they may be present. The application of these techniques to various device structures, and the uncertainties associated with radiochemical profiling, are discussed, together with the potential use of scattering techniques for the analysis of diffused layers.

Key Words: Charged-particle activation, diffusion, impurity profiling, ion implantation, neutron activation, radiation damage, radioactivation, radioisotope, radiotracer, silicon, silicon dioxide, silicon nitride.

1. Introduction

Silicon electronic devices depend for their operation upon the presence of impurities which are introduced into the device structure by processes such as melt growth, epitaxial vapor growth, diffusion, and ion implantation. A knowledge of the impurity distributions attained in these processes, and the redistributions which occur during subsequent processing steps, is a prime requirement in developing devices structures of optimum performance.

Usually, one of various electrical profiling methods for determinimg the distribution of excess charge carriers in silicon is employed to provide the needed information. Under certain circumstances, however, only chemical profiling for a specific impurity will suffice. For example, the impurity may be electrically inactive in the silicon matrix; or it may normally be electrically active, but present at such high concentration that it is clustered, precipitated, or otherwise inactive; or it may be compensated for by the presence of an electrically active impurity of the opposite dopant type; or the matrix itself may be one which is not amenable to electrical profiling, such as silicon dioxide or silicon nitride, or even silicon in an alloyed, displacement-damaged, amorphous, or polycrystalline state. In such cases a chemical profile provides a knowledge of the total concentration profile of the impurity of interest. Occasionally, both the total chemical profile and the net electrically active profile of a device structure must be acquired in order to determine the degree of impurity ionization, compensation, pairing or precipitation which accompanies a given fabrication process.

Because of the relatively small amounts of impurity contained in device structures, chemical profiles are usually obtained by employing radiochemical techniques. Very high analytical sensitivities may be attained if the material of interest can be converted into a radioactive form, since when the resulting nuclei undergo radioactive decay, relatively large amounts of energy are given off from small amounts of material. In the decay of a single nucleus, several million electron volts may be released, sufficient to trigger even the simplest nuclear detector. However, because only a very small fraction of impurity atoms can be made radioactive, and because the overall detection efficiency is further influenced by such factors as decay half-life and counting geometry, practical sensitivities for most impurities are on the order of about 10^{10} atoms. Thus, for a typical profiling application, in which a 5-cm^2 blanket-diffused area is sectioned by removal of 200Å-thick layers of silicon, a detection limit of 10^{15} atoms/cm^3 may be attained through direct counting of the removed layers.

The radiochemical profiling methods are applicable, of course, only to those elements for which a suitable radioactive tracer isotope is available, or which, upon nuclear activation, form a radioactive product of suitable half-life. Most of the dopant impurities commonly employed in silicon device technology, with the notable exception of boron, are amenable to profiling by one or both of the two most well-known radiochemical methods, the neutron activation technique and the radiotracer technique.

2. Radiochemical Profiling Methods

In the neutron activation technique, the fabrication cycle of interest is usually carried out without photolithographic masking, so that a large area ($\sim$1cm^2 or more) structure containing the appropriate impurity profile is obtained. The resulting structure is irradiated with thermal neutrons to convert a small fraction of the contained impurity into a radioactive product. The (n, γ) nuclear reaction usually predominates for thermal neutron activations, e.g., the reaction $^{75}As(n,\gamma)^{76}As$ which produces 26.5-hour radioarsenic. Usually, the specimen is annealed at a temperature of about 600^{o}C, to remove radiation-induced lattice damage, after which profiling is accomplished in a fairly straightforward manner. Successive layers of the structure are removed by anodization and stripping [1][1] or by other chemical or mechanical processes, employing non-radioactive isotopic carriers for the activation product to be measured and retaining all solutions or suspensions involved in the sectioning process for radioassay. Generally, these counting samples will contain radioactive products from the activation of silicon and other impurities present in the original structure, so that it may be necessary to separate the induced radioactivities. Either chemical separations or instrumental techniques, such as decay-curve analysis and gamma-ray spectroscopy, may be employed for this purpose. A known amount of the impurity element of interest is usually carried through the same activation and counting procedure, to provide the necessary standardization. Typical profiles obtained by the neutron activation technique are shown in Figs. 1-4.

In the radiotracer technique, the fabrication cycle of interest is carried out using a radioisotope-labeled dopant. The method is conveniently adaptable to closed-system processes such as capsule diffusions; for flow tube systems, suitable precautions must be taken to entrap the relatively large amounts of radioactive material which are not incorporated into the device structure. As in the case of neutron activation profiling, a blanket-doped sample, usually 1cm^2 or more in area, is employed. Successive sectioning and counting of the removed layers is performed as before; however, the absence of interfering radioactive species permits the use of simpler radioassay procedures than are required for the neutron activation technique. A specific activity measurement of the labeled starting material is necessary for standardization. Differences in isotopic diffusivities are usually neglected, except in studies of hydrogen impurity migration.

In addition to its usefulness for general profiling purposes, the radiotracer method is uniquely capable of providing measurements of atomic motions under equilibrium conditions, i.e., in the absence of a net migration of the species being traced. Thus, fundamental studies of self-diffusion [2] and of isoconcentration diffusion [3] may be carried out in such an experimentally simplified manner that interpretation of the results is greatly facilitated. As an example, the profiles resulting from identical diffusions of radioarsenic tracer into an initially intrinsic wafer and into a wafer homogenously doped with 5.5×10^{19} atoms/cm^3 of non-radioactive arsenic are shown in Fig. 5. Note that the local diffusion coefficient of arsenic near the surface of the wafer containing an arsenic concentration gradient is greater than the diffusion coefficient prevailing throughout the wafer in which concentration gradients are absent, owing to the so-called field-aided diffusion effect [4]. At greater depths in the non-homogeneous wafer, the local arsenic diffusivity becomes much smaller because of the concentration dependence of this parameter [3]. Neither of these complicating features are evident in the homogeneously doped or isoconcentration diffusion wafer.

Both the neutron activation and radiotracer techniques have been extensively applied to the acquistion of impurity profiles in silicon [5]. Less well-known are the applications of these radiochemical profiling methods to dielectrics [6, 7, 8] and to metals employed for fabricating silicon devices. Occasionally, an indirect profiling application may be encountered, as, for example, in obtaining the damage distribution profile shown in Fig. 6. This profile was obtained from an ion-implanted wafer after first "decorating" the vacancy-rich bombardment-damaged region with copper. Copper is a rapid interstitial diffuser in silicon, which seeks out vacancies and reacts with them to become substitutional and relatively immobile. The resulting radiochemical profile provides valuable information about the damage distribution which accompanies the implantation process [9].

Radiochemical profiling may also be carried out following impurity activation with fast neutrons, charged particles, or photons. Maekawa [10] has successfully employed 3.1 MeV protons in the nuclear reaction $^{11}B(p,n)^{11}C$ to obtain diffusion profiles of boron in silicon. However, the cross-sections for most nuclear reactions are strongly dependent upon energy; and corrections for energy loss as the beam penetrates the sample can be appreciable. Particular care must be taken to avoid complications arising from the effects of channeling, or the preferential penetration of ion beams along certain crystallographic orientations.

---

[1] Figures in brackets indicate the literature references at the end of this paper.

2.1 Discussion of Errors

All of the above radiochemical profiling techniques are subject to small uncertainties arising from variations in section removal, radiochemical manipulation, standard preparation, and from counting statistics. However, such uncertainites in individual concentration vs. depth data points may be held within limits of 5% relative and 10% absolute error by observing suitable precautions. Much of the precision and detail obtainable by radiochemical profiling may be attributed to the direct manner in which the impurity content of each layer is measured in the procedure outlined above. Alternatively, the residual wafer activity may be measured before and after the removal of each section, and the impurity content of the removed layer obtained by difference. However, since the differential measurement technique leads to a considerably greater scatter of the individual data points, the direct measurement technique is favored except under certain circumstances; e.g., when a possibility exists that radioactive material may be lost during dissolution of the removed section. The greatest difficulty and potential source of error in radiochemical profiling is the carry-over of radioactive contamination from one sample to the next. In steeply descending profiles, there may be an order of magnitude difference between the activity of two consecutively removed samples, and six orders of magnitude or more decrease over the portion of the profile for which reliable measurements are required. A miniscule portion of that activity which ideally would be completely removed in an early section, if carried over to a subsequent counting sample, may lead to a very large error in the latter. By the use of carefully developed radiochemical procedures, with scrupulous attention to radioactivity cleanliness, this source of error may be minimized. The net effect of the significant sources of error, such as contamination carry-over, non-planar sectioning, non-uniformity or parallelicity of the specimen over the area being sectioned, or displacement of the impurity profile during the activation process, is that the resulting radiochemical profiles may appear less steep than the true profile. In general, radiochemical profiles acquired by an experienced technician may be viewed as providing knowledge of total impurity profiles with at least the same degree of accuracy with which net electrically active impurity profiles are regarded, and with the direction of possible small deviations from true slopes known.

The possible displacement of impurity profiles during radioactivation deserves some further amplification. Let us consider the activation of arsenic impurity with thermal neutrons, which may be represented by the nuclear reaction

$$^{75}_{33}\mathrm{As} + ^{1}_{0}\mathrm{n} \rightarrow ^{76}_{33}\mathrm{As} + ^{0}_{0}\gamma. \qquad (1)$$

In a typical profiling application, neutron irradiation will be allowed to proceed until approximately a 10^{-7} fraction of the arsenic impurity atoms present in the sample has been coverted to the radioactive species ^{76}As. By comparing the atomic masses of ^{75}As, ^{1}n, and ^{76}As, it may be seen that about 0.008 atomic mass units are lost in reaction (1), corresponding to an energy release of about 7 MeV per event. This energy is carried off in the form of kinetic energy of the products, and, by application of the principle of momentum conservation, it may be estimated that the resulting ^{76}As recoil energy is several hundred electron volts. Since the recoil energy is several times larger then the energy required to displace a substitutional atom from the silicon lattice, the ^{76}As atom formed in reaction (1) will usually come to rest in an interstitial site, perhaps a few lattice spacings removed from the point of origin. At the temperature prevailing during the nuclear activation, or during subsequent heat-treatments of the specimen prior to sectioning, this interstitial atom may migrate an appreciable distance before finally combining with a silicon vacancy and reverting to the immobile, substitutional form. Thus, a possibility always exists that the radioactive product distribution acquired by radiochemical profiling may be displaced somewhat from the impurity distribution initially present in the sample. This effect becomes especially important in the case of charged-particle activations in which relatively much larger amounts of kinetic energy, often several thousand electron volts, may be imparted to the product nuclide. In addition to this interstitial migration mode, in which only the relatively small amounts of impurity which undergo activation are displaced, a shifting of the total impurity profile, owing to vacancy-enhanced substitutional diffusion, may occur if the activation process is carried out at moderately elevated temperature. For example, Pfister [11] has observed p-n junction movement at 600°C in silicon undergoing proton bombardment, and attributes this effect to impurity diffusion enhancement by vacancies produced during the irradiation.

Schmidt [12] has observed what appears to be deep radiation-enhanced diffusion tails in radiochemical profiles obtained from phosphorus-diffused Si/SiO_2 structures following neutron activation with a predominantly thermal neutron source at temperatures up to 250°C. It appears likely that his observations may be at least partially attributed to the previously discussed vacancy-enhancement effect. In our own work, however, activations have always been performed in the well-thermalized regions of a "swimming pool" reactor, with precautions taken to minimize gamma-heating of the specimens, and no such demonstrable long range migrations of activated species

have been observed. We have occasionally found tails on profiles which have been acquired over several orders of magnitude of decreasing concentration, and these are attributed primarily to contamination carry-over during the sectioning process. The possible interstitial migration mode has been further examined in an experiment in which arsenic containing a small fraction of ^{74}As radiotracer was diffused into silicon to form a steeply descending shallow profile. Following activation with $\sim 10^{18}$ thermal neutrons/cm^2 at $\sim 100^{o}C$, and subsequent annealing for 30 min. at 625^{o}C, the wafer was sectioned and the $^{74}As/^{75}As$ ratio was determined for each section. The results are shown in Fig. 7, in which the undisturbed profile, measured by the ^{74}As distribution, and the ^{75}As profile acquired by the neutron activation technique are seen to coincide within the normally anticipated error limits. It is concluded that impurity profile displacements may be made negligibly small for thermal neutron activations, provided suitable precautions are observed. However, activation with charged particles may lead to difficulties in this regard.

3. Scattering Methods

Not all nuclear methods for impurity detection are based upon radiochemical measurements of a tracer isotope or of an activated product. The prompt particles emitted during nuclear excitation afford several possibilities for the measurement of impurities which cannot be detected by more conventional techniques. For example, measurement of the prompt alpha particles or gamma rays produced during bombardment with low-energy neutrons or protons could be employed for the determination of boron, although such methods have not yet been applied to profiling studies.

Finally, brief mention should be made of the use of ion beams of energy insufficient to induce nuclear excitation, but which undergo elastic interactions with nuclei contained in thin layers of silicon. This Rutherford scattering technique, which is discussed in more detail in the following paper [13], enables a determination of impurity concentration and of concentration gradients to be made from the measured energy difference between incident and backscattered charged particles. Although the attainable sensitivity is generally less than that achieved in the radioisotope techniques, the scattering method may be combined with channeling measurements to differentiate between impurity atoms located on substitutional and those on interstitial lattice sites.

4. Acknowledgement

The author is indebted to Dr. R. H. Kastl for assistance with the development of profiling techniques, and to E. J. Schneider, Jr., F. V. Keicher, and G. Jung, for acquiring the profiles which accompany this manuscript.

5. References

[1] E. Tannenbaum, Solid State Electron. 2, 123 (1961).

[2] J. M. Fairfield and B. J. Masters, J. Appl. Phys. 38, 3148 (1967).

[3] B. J. Masters and J. M. Fairfield, J. Appl. Phys. 40, 2390 (1969).

[4] F. M. Smits, Proc. IRE 46, 1049 (1958).

[5] D. L. Kendall and D. B. DeVries, Semiconductor Silicon (The Electrochemical Society, Inc. N. Y. 1969), Tables I-IV.

[6] J. F. Osborne, G. B. Larrabee, and V. Harrap, Analytical Chemistry 39, 1144 (1967).

[7] P. J. Burkhardt, J. Electrochem. Soc. 114, 196 (1967).

[8] D. R. Kerr, IEEE Trans. Electron. Devices 14, 630 (1967).

[9] B. J. Masters, B. L. Crowder and J. M. Fairfield, Proc. Int. Conf. Ion Implantation in Semiconductors, Thousand Oaks, California (Gordon and Breach, N. Y., 1970).

[10] S. Maekawa and T. Oshida, J. Phys. Soc. Japan 19, 253 (1964).

[11] P. Baruch and J. C Pfister, Nuclear Instruments and Methods 38, 197 (1965).

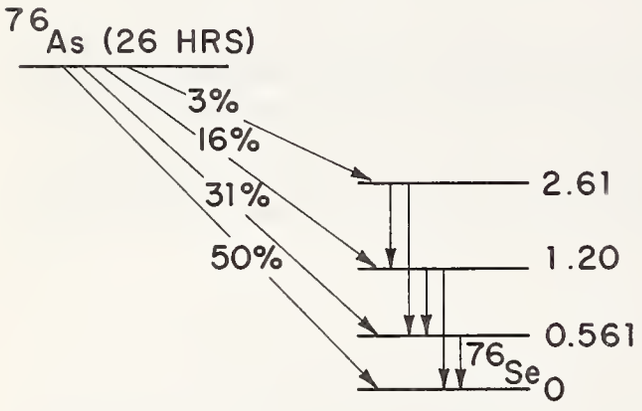
[12] P. F. Schmidt, Appl. Phys. Letters 8, 264 (1966).

[13] S. L. Chou. L. A. Davidson, and J. F. Gibbons, Symposium on Silicon Device Processing, Gaithersburg, Maryland (U. S. G. P. O., Washington, D.C. 1970).

RADIOCHEMICAL PROFILING PROCEDURES

NEUTRON ACTIVATION

1. DIFFUSE WITH ^{75}As (NORMAL ARSENIC)
2. ACTIVATE WITH THERMAL NEUTRONS ($^{75}As + {}^{0}n \longrightarrow {}^{76}As + \gamma$, AND OTHER REACTIONS)
3. SECTION AND RADIOASSAY (CHEMICAL OR INSTRUMENTAL SEP'N REQ'D)



RADIOISOTOPE TRACER

1. DIFFUSE WITH ^{74}As-LABELED ARSENIC
2. SECTION AND RADIOASSAY (SIMPLE COUNTING SUFFICES)

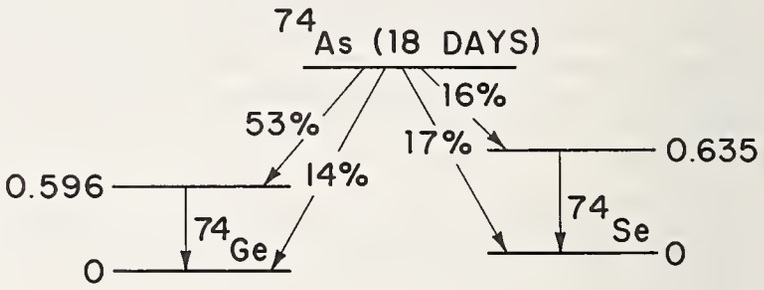
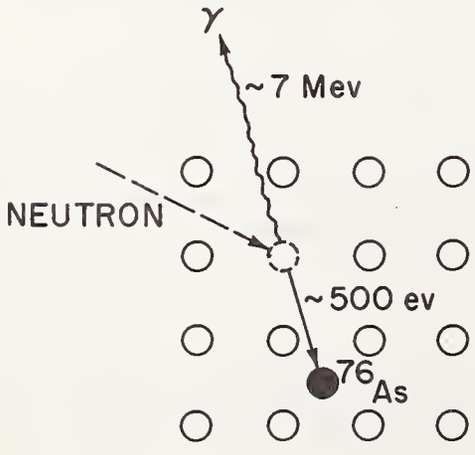


Fig. A

$$^{75}_{33}As + {}^{1}_{0}n \longrightarrow {}^{76}_{33}As + \gamma \qquad Q \approx +7 \text{ MeV}$$

(ONE IN $\sim 10^7$ ^{75}As-ATOMS CONVERTED TO ^{76}As, $t_{1/2}$ = 26HRS)



^{76}As RECOIL ENERGY $\approx$ 500eV

DISPLACEMENT ENERGY $\approx$ 20eV

$\therefore {}^{75}As_{(SUB'L)} \xrightarrow[ACTIVATION]{NEUTRON} {}^{76}As_{(INT'L)}$

$D_{(INTERSTITIAL)} > 10^{10} cm^2/sec$ AT 600°C

OR $2\sqrt{Dt} > 10$ MICRONS

Fig. B

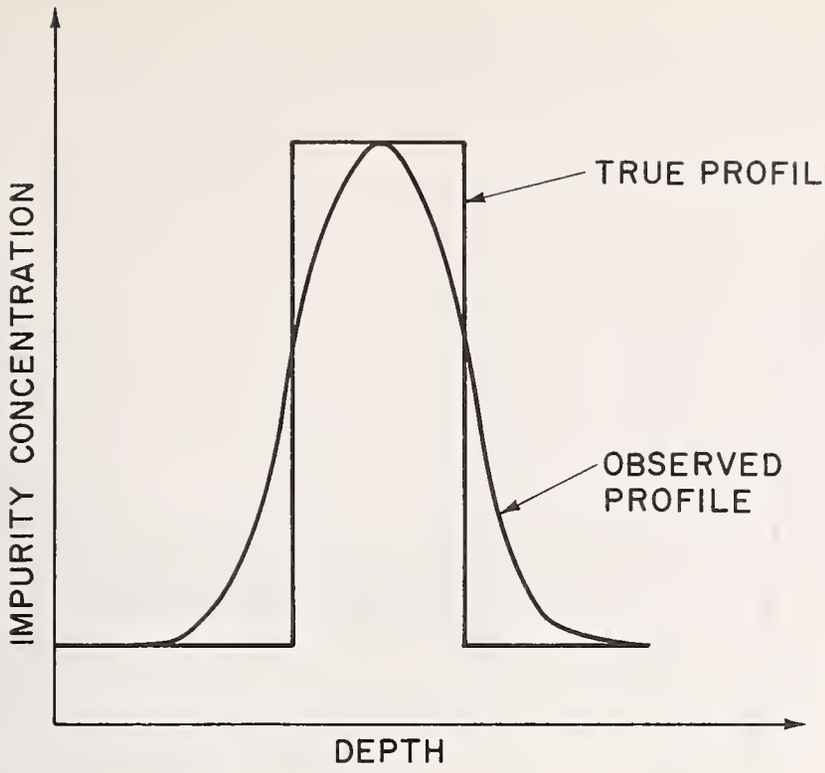


Fig. C

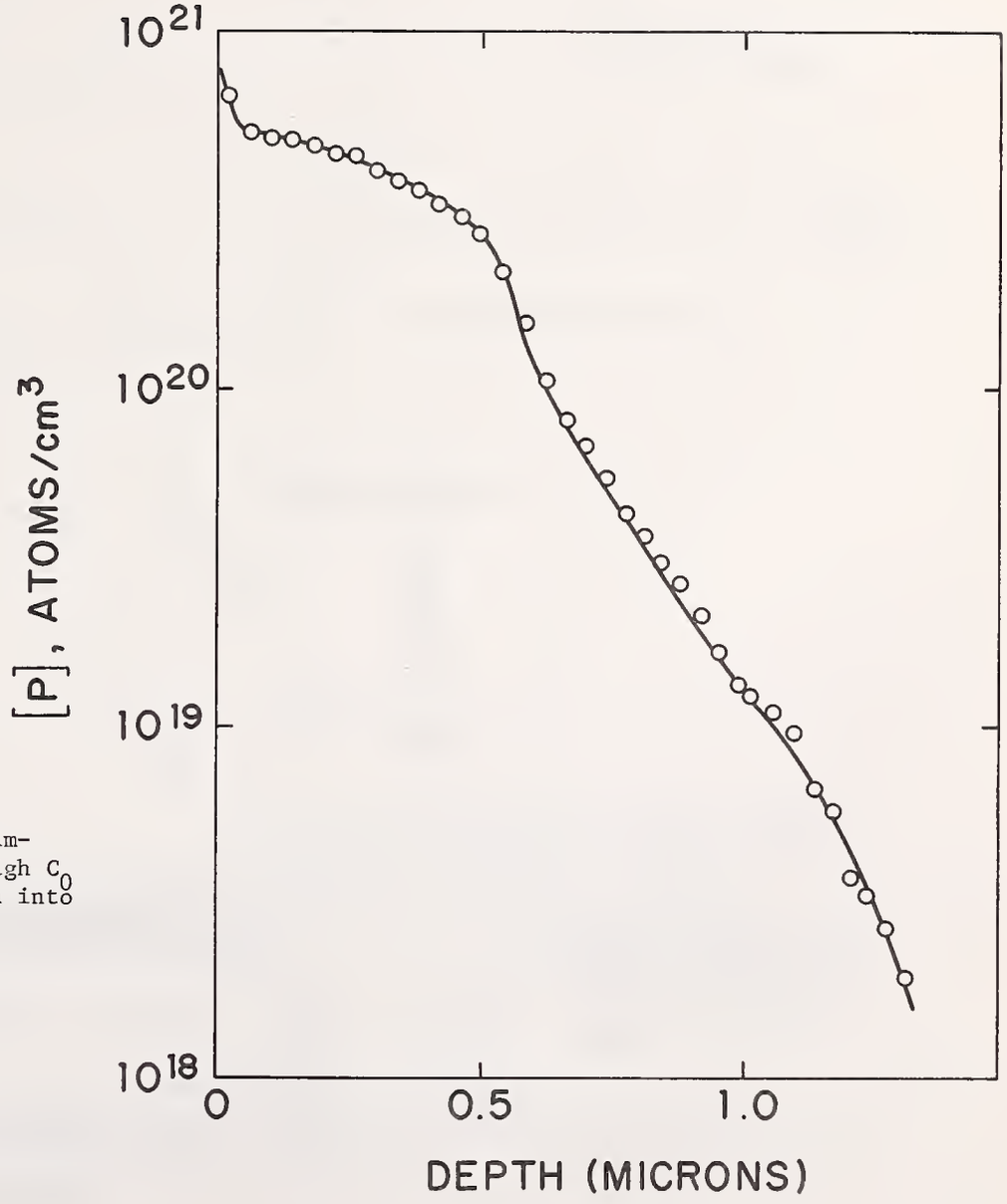


Fig. 1. Neutron activation impurity profile of high C_0 phosphorus diffusion into silicon.

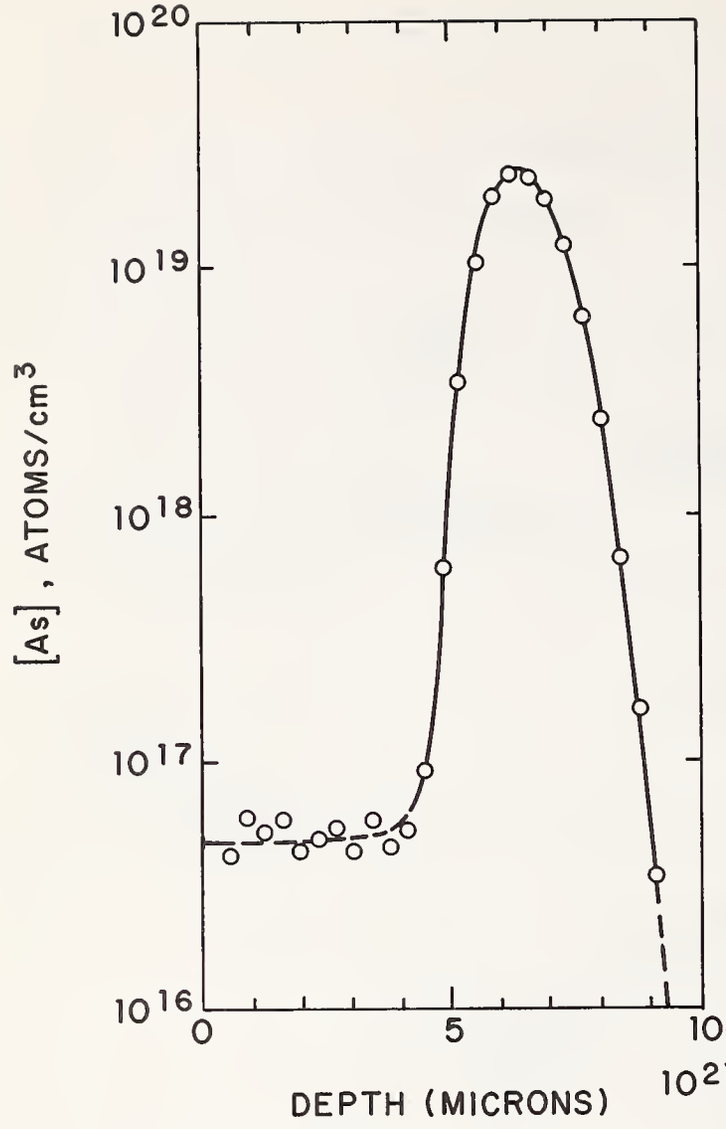


Fig. 2. Neutron activation profile of arsenic buried layer which was formed by diffusion and epitaxial vapor growth.

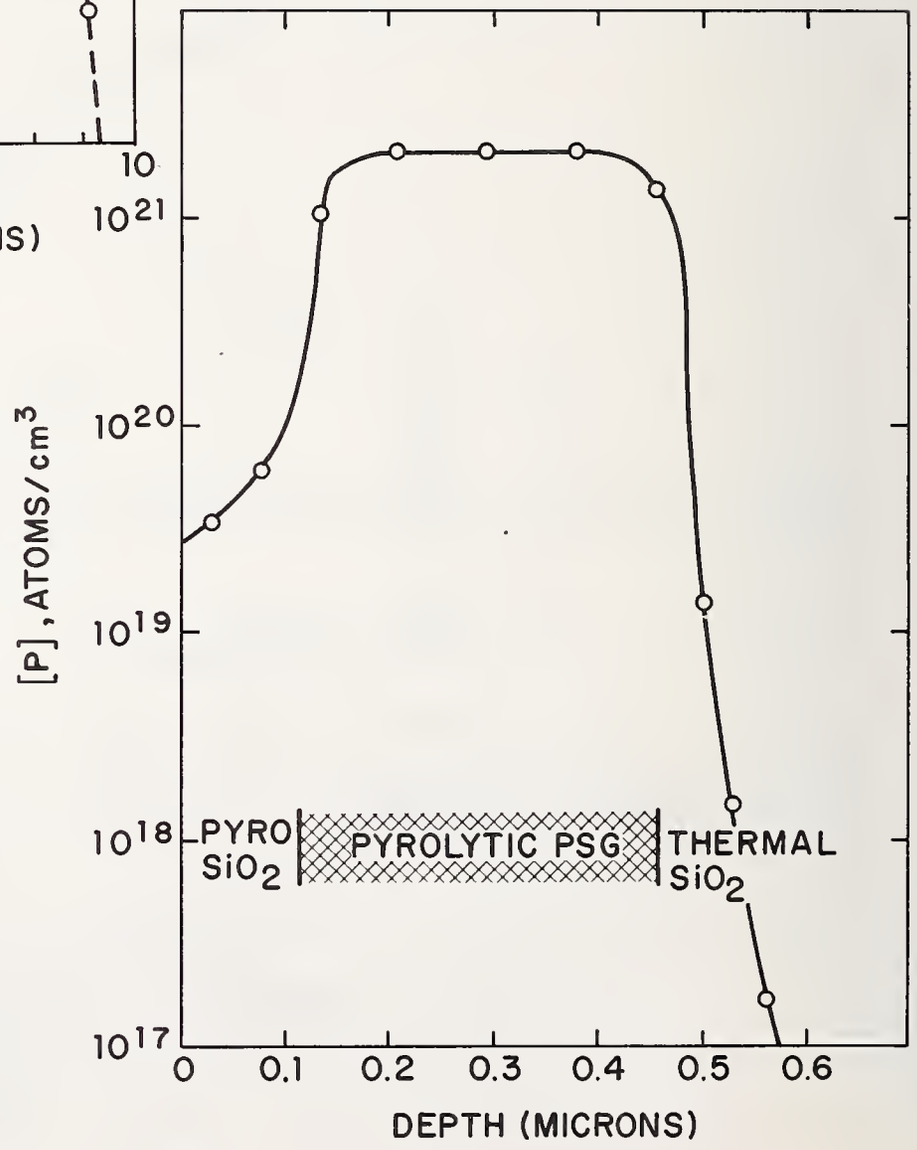


Fig. 3. Neutron activation profile of phosphorus impurity in multilayered SiO_2 film.

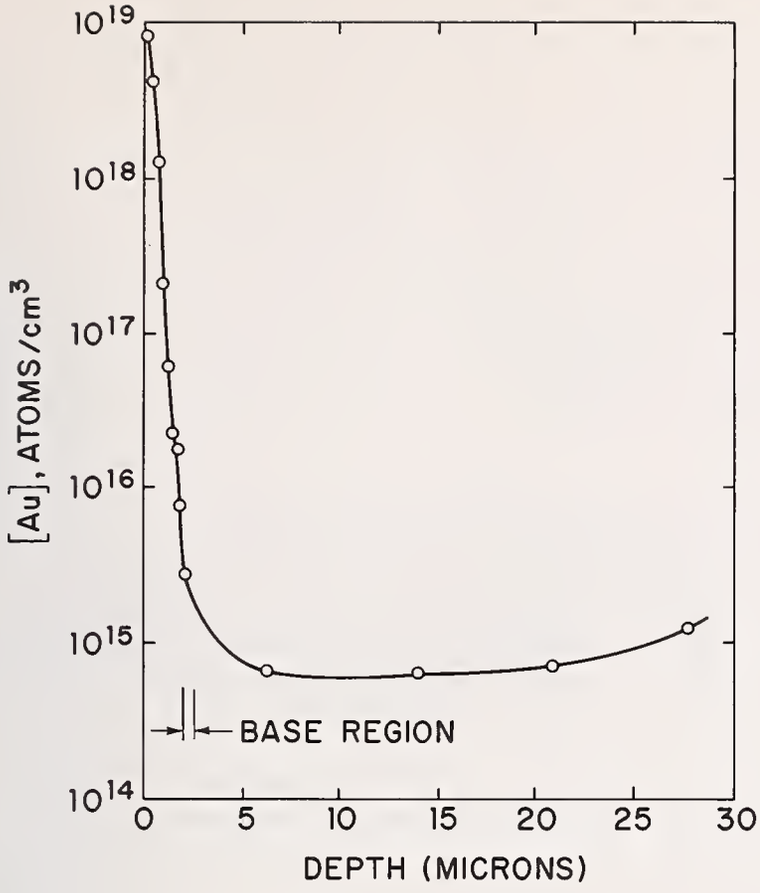


Fig. 4. Neutron activation profile of gold distribution in NPN transistor structure.

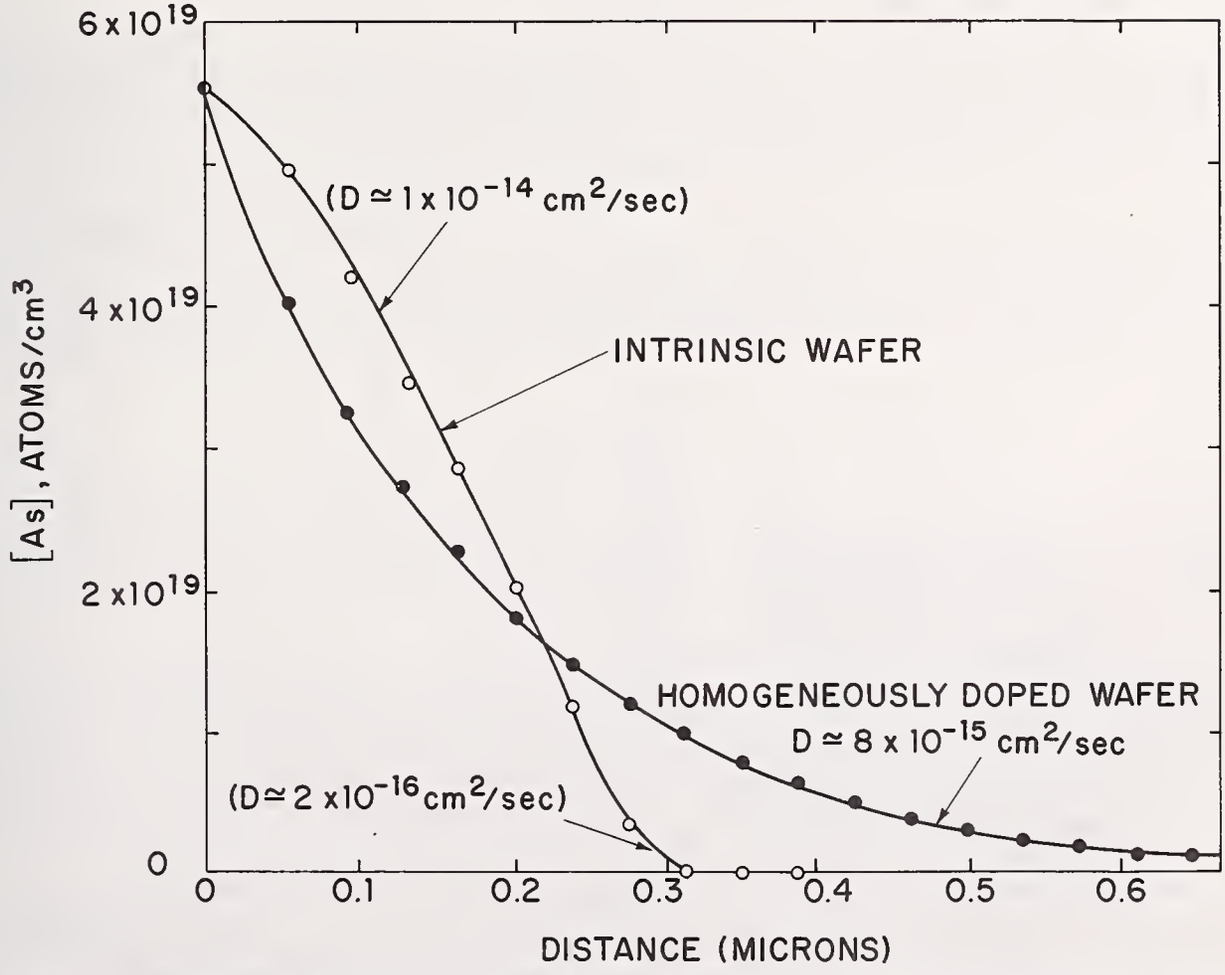


Fig. 5. Radiotracer diffusion of ^{76}As-labeled arsenic into initially intrinsic silicon (open circles) and into silicon homogeneously doped with 5.5x10^{19}As-atoms/cm^3 (closed circles), both 19 1/2 hrs. @ 956°C.

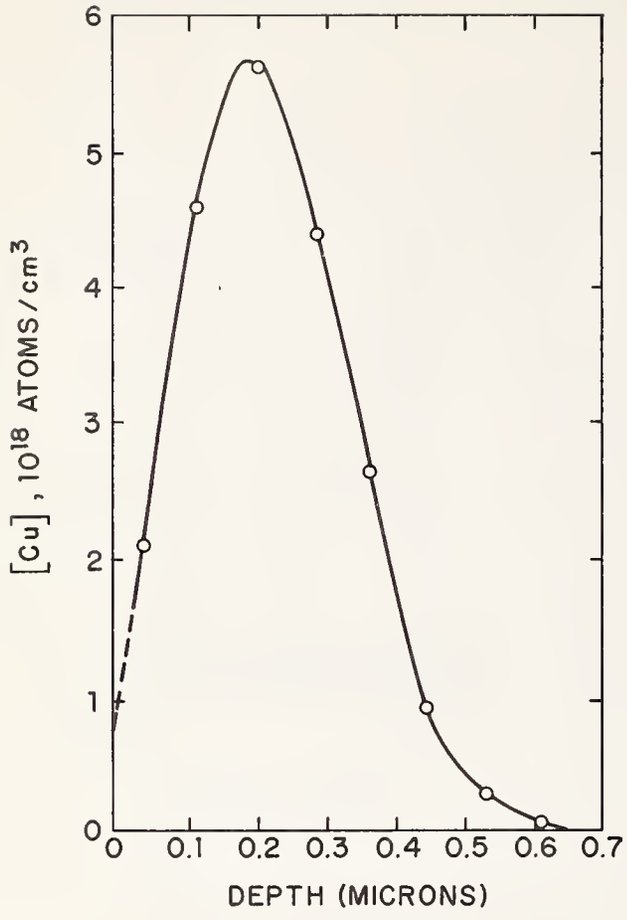


Fig. 6. Copper-decorated damage distribution profile resulting from 10^{16} ion/cm^2 argon implant at 250 KeV into silicon.

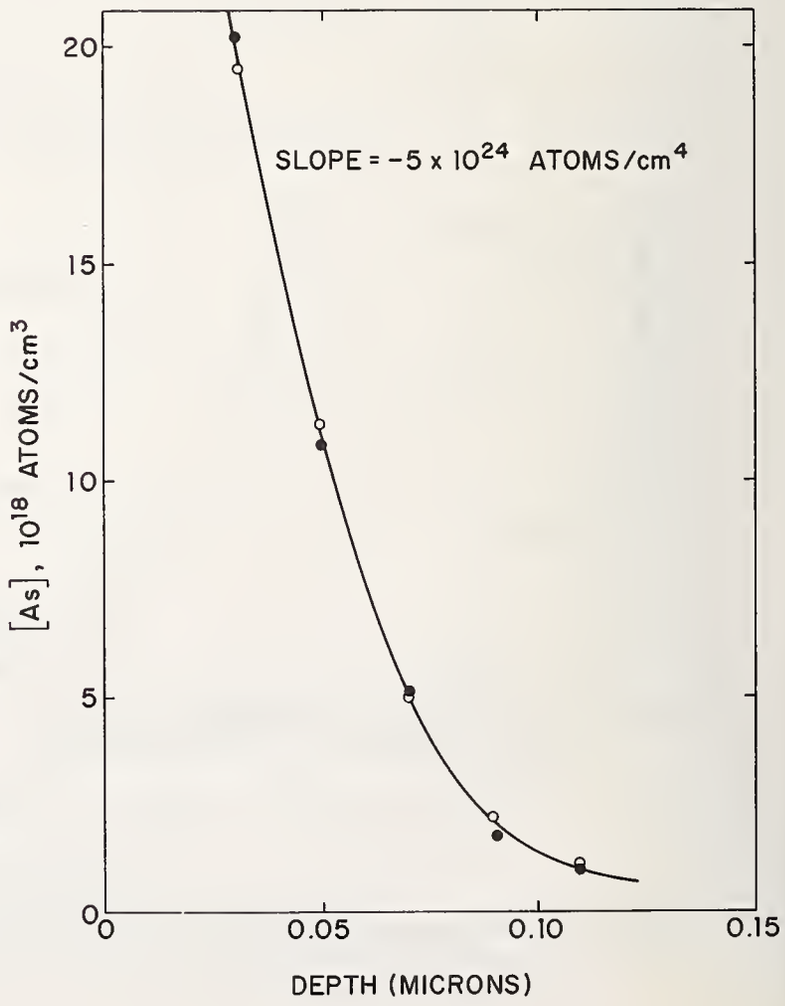


Fig. 7. Radiochemical profiles obtained from wafer diffused with ^{74}AS-labeled arsenic, following 24 hr. activation with 1×10^{13} thermal neutrons/cm^2 -sec. and post-activation anneal for 30 min. at 625 °C. Profile determined from ^{74}AS radiotracer (closed circles) and from ^{76}AS activation (open circles).

Use of High Energy Ion Beams for the Analysis of Doped Surface Layers

S. Chou,* L. A. Davidson,† and J. F. Gibbons

Stanford Electronics Laboratories
Stanford University
Stanford, California 94305

The purpose of this paper is to describe the use of MeV ion beams for the analysis of doped semiconductor surface layers. By monitoring the backscattered ions, X-rays, or other radiation emitted by the semiconductor during ion bombardment, it is possible to determine the impurity concentration. The dependence of radiation yield on the direction of incidence of the ion beam gives information regarding the crystalline order of the sample, as well as the lattice locations of impurities. This paper concentrates on the interpretation of elastic ion-backscattering data. Experimental results on arsenic-doped silicon are presented as an illustration of the technique.

Key words: Channeling, diffusion in silicon, evaluation of doped semiconductors, impurity lattice location, impurity profiles, ion-backscattering.

1. Introduction

High energy (MeV) ion beams are useful tools for studying the chemical composition and physical structure of thin material layers. Early applications of the technique included the investigation of lattice locations of impurities implanted into various materials, and the effects of implantation on the structure of the host material [1].‡ More recently, ion beams have been employed in the study of dielectric layers [2] and impurities diffused into silicon [3]. A comprehensive review of the technique and its applications (up to 1968) has been given by Gibson [4].

The purpose of this paper is to acquaint the reader with the basic physical and experimental aspects of using ion beams for material analysis. Section 2 gives a qualitative description of the physical principles behind the technique. Subsequent sections discuss in detail how the technique may be used (a) to determine impurity profiles (Sec. 3), (b) to find lattice locations of impurity elements (Sec. 4.2), and (c) to investigate crystal order (Sec. 4.3).

2. Physical Basis for the Utility of Ion Beams

The technique is based on the interactions of ion beams with matter. These interactions include: (a) the elastic scattering of ions from target atoms, (b) excitation of electrons resulting in characteristic X-ray emissions, and (c) nuclear reactions between energetic ions and target nuclei. The products of these interactions provide information on the identities of the target atoms. For example, the energy lost by an ion in a large-angle elastic scattering event depends on the mass of the target atom involved.

Besides identifying the target atoms, ion beams can also reveal the spatial distributions of, say, impurities in a host material. This is because an ion loses energy with distance of penetration into the target material. The average rate of spatial energy loss depends on the ion, the ion energy, and the target material. An ion entering the target surface at some known initial energy (E_i) would therefore have a lower energy by the time it interacts with an atom below the surface. Often, one can deduce from the observed interaction product the energy of the ion at the time of interaction (E_i'). The difference $E_i - E_i'$ gives a measure of the depth of the target atom involved in the interaction.

In a target with configurational order, the intensities of the various interaction products also depend on the *direction* of the ion beam. In particular, there is usually a pronounced decrease in interaction products when the beam is aligned with a major crystallographic axis or along a low-index plane. By

---

* Also with Fairchild Research and Development, Palo Alto, California.
† Now with IBM Components Division, Hopewell Junction, New York.
‡ Figures in brackets refer to literature references at the end of this paper.

observing the intensities of interaction products as a function of beam orientation relative to the target, one can deduce the crystal structure of the sample, or determine the absence of it. This method can also yield information on the lattice locations of impurities in a host crystal. For example, if the impurities were entirely in substitutional sites, one would expect the interaction yields from these impurities to vary with beam orientation in the same manner as the interaction yield from the atoms of the host crystal. On the other hand, if the impurities were distributed in random, off-lattice sites, their interaction yields should be more isotropic with respect to beam orientation.

Of the various interaction products produced by ion bombardment, backscattered ions are usually the easiest to detect and analyze. High resolution, solid-state charged-particle detectors are readily available, and the energy spectrum of the backscattered ions can be analyzed using simple scattering theory. In this paper, we shall concentrate on the interpretation of data derived from backscattered ions. Details regarding the observation of X-ray emission and nuclear interaction products may be found elsewhere [5,6]. However, many apsects of the technique do not depend on which particular interaction product is observed.

3. Use of Ion Scattering to Determine Impurity Profiles

3.1. Experimental Equipment

Figure 1 is a schematic diagram of the experimental arrangement used in our work. A proton or $^4He^+$ ion beam from a 3-MeV Van de Graaff accelerator is collimated so as to have a total angular divergence of less than 0.15°. The ion beam is directed toward the center of the sample, which is mounted on a goniometer. The goniometer can be used to tilt the sample about a vertical axis, or to rotate the sample about its own axis. The angles associated with the tilt and rotation are designated θ_1 and ϕ, respectively (see Fig. 1). The resolution for each angle is 0.025°. A solid-state (surface barrier) detector is mounted in the horizontal plane containing the incident ion beam and the sample axis, in such a way as to receive ions scattered through a mean angle $\theta_S = 120°$. The solid angle subtended by the detector at the center of the sample is approximately 7×10^{-3} sr. The beam spot size at the target is about 2 mm^2 and typical beam currents are in the range 1—10 nA. Not shown in Fig. 1 is a secondary electron suppression grid mounted in front of the sample holder and held at a negative potential relative to the sample.

A block diagram of the basic detector electronics appears in Fig. 2. The detector gives a pulse of charge proportional to the energy of the incident ion. This pulse is integrated by a charge-sensitive preamplifier at the target chamber, and the signal is transmitted to the main amplifier, where pulse-shaping is performed. The output pulses from the main amplifier are fed into a multichannel analyzer, which gives the energy spectrum of the backscattered ions. The output of the main amplifier is also fed into a pulse-height discriminator, which is followed by a frequency meter. The analog output of the frequency meter drives a recorder, which can thus be used to plot the ion backscattering rate as a function of beam orientation. This step is performed at the beginning of each experiment in order to determine the orientation of the sample, and will be discussed further in Sec. 4.3.

To obtain the impurity profiles in a crystalline substrate, the ion beam is aimed at a direction of *low* symmetry (i.e., away from all major crystallographic directions). Along low symmetry directions, the ions make uncorrelated collisions with both impurities and lattice atoms, and the results can be interpreted as if the sample was noncrystalline.

The impurity concentrations and profiles are obtained from the energy spectrum of the backscattered ions as recorded on the multichannel analyzer. Before presenting and interpreting the results, a brief discussion of the scattering process is given.

3.2. Elastic Ion Scattering in Unordered Material

Figure 3 is an idealized representation of a large-angle ion scattering event. The ion enters the surface of the sample with an initial energy E_i. As the ion penetrates the sample, it loses energy, for example, through excitation of electrons or successive small-angle elastic collisions with atoms. After penetrating to a certain vertical depth x, the ion is shown to be elastically scattered through a large angle ϕ_S. E_i' and E_S' denote the energies of the ion immediately before and after the encounter. The ion again loses energy on its outgoing trajectory, reemerges from the surface of the samples with an energy E_S, and is observed by a properly located detector.[1] For most ions with energies of the order of 1 MeV, elastic scattering of ions by atoms or nuclei can be treated as classical, two-body problems. By applying the principles of conservation of momentum and energy, one can readily show that an ion of mass

[1]The track of an ion in a solid is in general not straight, but consists of numerous, usually small deflections arising from collisions with atoms. The total deflection suffered by the ion while inside the solid is thus the sum of numerous small deflections, and possibly one or more large-angle deflections. In this treatment, the major deflection of the ion is assumed to be the result of a *single* close nuclear encounter.

m_i, scattered by an atom of mass m_t through an angle θ_s, is left with a fraction α of its original energy, where

$$\alpha = \frac{E_s'}{E_i'} = \left[\frac{\cos\theta_s + \sqrt{(m_t/m_i)^2 - \sin^2\theta_s}}{1 + \frac{m_t}{m_i}} \right]^2 \tag{1}$$

Plots of α as a function of m_t/m_i for various scattering angles are shown in Fig. 4. It is evident that the energy separation of ions backscattered from atoms of slightly differing mass is maximized by using large scattering angles and by making the mass of the ion comparable to the masses of the target atoms.

The scattering *probability* depends on the nature of the forces acting between the ion and target atom during the encounter [7]. In order to suffer large-angle deflections, protons and helium ions with energies of the order of 1 MeV must approach within 10^{-12} to 10^{-11} cm of the target nuclei. Since this distance is small compared to the electron screening radius (typically 10^{-9} cm), the shielding of the nuclear charge may be neglected, and one may consider the large-angle deflection to result from a coulomb interaction between the nuclei of the ion and the target atom. The differential cross section for this type of scattering is given by the well-known Rutherford formula,

$$\frac{d\sigma}{d\Omega} = 1.3\times10^{-27} \left(\frac{Z_i Z_t}{E_i'}\right)^2 \left(1 + \frac{m_i}{m_t}\right)^2 \frac{1}{\sin^4\left(\frac{\theta_s^*}{2}\right)} \frac{d\Omega^*}{d\Omega} \quad \text{cm}^2/\text{sr} \tag{2}$$

where Z_i and Z_t are the atomic numbers of the ion and target atom, respectively, and the initial ion energy E_i' is in MeV. The differential solid angle $d\Omega^*$ and the scattering angle θ_s^* are those pertaining to the center-of-mass system comprising the ion and target atom. However, distinction between center-of-mass and laboratory coordinates is important only for target atoms which are not very heavy compared to the incident ions [8].

The energy of the ion, as it emerges from the surface of the sample, can be written in the form

$$\begin{aligned} E_s &= \alpha E_i' - \int_{\ell_2} S(E)\,d\ell \\ &= \alpha E_i - \alpha \int_{\ell_1} S(E)\,d\ell_1 - \int_{\ell_2} S(E)\,d\ell_2 \end{aligned} \tag{3}$$

where $S(E)$ is the stopping power [9] of the solid material, and the integrals are carried out over the ingoing and outgoing paths (ℓ_1 and ℓ_2, respectively) of the ion. In general, the integrals are difficult to evaluate because the energy of the ion along its track is not known in advance. However, if $S(E)$ does not vary appreciably along each of the paths ℓ_1 and ℓ_2, one can approximate $S(E)$ by constant values $\bar{S}_1$ and $\bar{S}_2$ along the respective paths. Then

$$\begin{aligned} \int_{\ell_1} S(E)\,d\ell_1 &= \bar{S}_1 \ell_1 = \bar{S}_1 \frac{x}{\cos\theta_1} \\ \int_{\ell_2} S(E)\,d\ell_2 &= \bar{S}_2 \ell_2 = \bar{S}_2 \frac{x}{\cos\theta_2} \end{aligned} \tag{4}$$

Substituting these approximations into (3) gives an expression relating the depth (x) of the scattering atom to the observed energy (E_s) of the scattered ion:

$$x = \frac{\alpha E_i - E_s}{\alpha \frac{\bar{S}_1}{\cos\theta_1} + \frac{\bar{S}_2}{\cos\theta_2}} \tag{5}$$

Consider a detector placed so as to receive ions backscattered through an angle θ_s into a small solid angle $\Delta\Omega$. The number of ions (dn_s) scattered into this solid angle by a given species of atoms located in a thin layer between x and x + dx from the sample surface is

$$dn_s = N(x)\,\frac{I}{q}\,\frac{d\sigma(E_i')}{d\Omega}\,\Delta\Omega\,\frac{d\ell_1}{dx}\,dx$$

$$= N(x)\,\frac{I}{q}\,\frac{d\sigma\left(E_i - \frac{x}{\cos\theta_1}\bar{S}_1\right)}{d\Omega}\,\Delta\Omega\,\frac{dx}{\cos\theta_1} \qquad (6)$$

where N(x) is the volume density of the given species of atoms, I is the total beam current,[2] and $d\sigma/d\Omega$ is the Rutherford differential scattering cross section given by (2).

From (2), (5), and (6), one obtains the following expression for the number of backscattered ions reaching the detector with energy between E_s and $E_s + dE_s$:

$$dn_s = N[x(E_s)]\,\frac{I}{q}\,\frac{d\sigma(E_i)}{d\Omega}\,\frac{1}{\cos\theta_1}\,\frac{\alpha\,\frac{\bar{S}_1}{\cos\theta_1} + \frac{\bar{S}_2}{\cos\theta_2}}{\left(\frac{E_s}{E_i}\,\frac{\bar{S}_1}{\cos\theta_1} + \frac{\bar{S}_2}{\cos\theta_2}\right)^2}\,\Delta\Omega dE_s \qquad (7)$$

In the above equation, $x(E_s)$ represents the expression on the right side of (5). Note that the scattering cross section in (7) is for an ion energy E_i, not E_i' as in (6). Together, (5) and (7) can be used to predict the energy spectrum of backscattered ions reaching the detector for a given spatial distribution of identical atoms, and vice versa.

In the case of doped semiconductor layers, the impurity concentration can be calculated more accurately by comparing the scattering yields from the impurity atoms and the lattice atoms, instead of using the *absolute* scattering yield of the impurity atoms. In other words, (7) can be written for both the impurity and lattice atoms and the ratio taken. This procedure eliminates errors arising from some measured quantities (e.g., the ion current) and minimizes errors due to uncertainties in other parameters (such as stopping power).

Other than atomic concentration, the most important parameter that affects the ratio of scattering yields from two different species of atoms is the atomic number. This parameter appears in the Rutherford scattering cross section as Z_t^2 (see (2)). For target atoms heavy compared to the incident ions, a good initial estimate of the impurity volume density $N_{imp}(x)$ can be obtained by neglecting the factors involving the mass ratio γ, resulting in the simple expression

$$\frac{N_{imp}(x)}{N_S} = \frac{Z_S^2}{Z_{imp}^2}\,\frac{(dn_s/dE_s)_{imp}}{(dn_s/dE_s)_S} \qquad (8)$$

where the subscripts "imp" and "S" refer to the impurity atoms and semiconductor lattice atoms, respectively, and dn_s/dE_s is the scattering yield per unit energy, evaluated at the energy $E_s(x)$ (i.e., the energy of ions reaching the detector after scattering from an atom at depth x).

3.3. Experimental Results

Results for an arsenic-doped layer in silicon are presented here for purposes of illustration. The sample was obtained by diffusing arsenic at 1250°C into a 5 Ω-cm, p-type silicon substrate, which was cut a few degrees off the (111) plane and mechanically polished. The depth of the n^+-p junction so formed was 4.2 μm, and the sheet resistance of the arsenic layer was 9 Ω/□.

Figure 5 shows the energy spectra of $^4He^+$ ions backscattered from the arsenic-doped layer [3]. The channel number is proportional to energy; channel 100 corresponds approximately to the incident ion energy of 1.1 MeV. Three spectra are shown, for three different beam directions. In this section only the uppermost spectrum, for a low symmetry ("random") beam direction, will be considered. The other two spectra, for beams aligned along the <110> and <111> crystallographic axes, will be discussed in Sec. 4.

[2] Actually, I should represent the total beam current at the depth x, but in most cases of interest the range of the ion is much greater than the depth of observation; hence I may be regarded as approximately constant.

The spectrum was obtained after about two hours of counting time, during which approximately 3×10^{14} helium ions were placed on target. The maximum counting rate is limited not by the available beam current, but by "pile-up" (i.e., overlapping) of pulses at high counting rates. Shorter counting times and smaller doses of helium ions would have resulted in greater statistical fluctuations in the data.

Turning now to the interpretation of the spectrum for a low symmetry beam direction, two edges may be observed, around channels 65 and 85 (see Fig. 5). These edges define the energies of ions backscattered from silicon and arsenic atoms, respectively, on the *surface* of the sample. Counts to the left of each edge arise from ions backscattered from atoms located below the surface of the sample. The energy difference between an ion scattered from a surface atom and a buried atom can be calculated using (5). A depth scale can be constructed, as shown in Fig. 5 for the arsenic portion of the spectrum.[3] (The depth scale will differ slightly for the silicon portion of the spectrum, in accordance with (5)).

The arsenic portion of the spectrum below channel 65 is masked by the silicon portion of the spectrum, so that the limit of observation is about 2500 Å. Calculations based on the ratio of the backscatter yield show that the sample of Fig. 5 has an approximately uniform arsenic concentration of 5×10^{19} cm^{-3} from the surface down to the limit of observation. This result agrees well with the electron density profile obtained by the incremental sheet resistivity technique. The technique of stripping off thin layers of silicon can also be used in conjunction with ion backscattering to obtain an impurity profile to any desired depth. Note, however, that the ion backscattering method gives the actual impurity concentration and not just the concentration of electrically active impurities, as one would obtain from sheet resistivity measurements.

4. Use of Ion Channeling to Determine Impurity Lattice Locations and Crystal Order

In Sec. 4.1 we first present a simplified discussion on the phenomenon of channeling, which is responsible for the orientation dependence of the backscatter yield from crystalline targets. The aligned spectra of Fig. 5 are then interpreted in Sec. 4.2, where a few complicated effects associated with channeling are also pointed out. Finally, Sec. 4.3 will discuss briefly the application of channeling to the determination of lattice order and crystallographic orientation of a sample.

4.1. Channeling in Crystal Structures

The concept of channeling can be grasped intuitively by looking at Fig. 6, which shows a diamond-type lattice viewed along the <110> axis (Fig. 6a) and viewed along a low symmetry direction several degrees off the <110> axis (Fig. 6b). Ions incident in the <110> direction are confronted with a fairly "open" structure, allowing most of the ions to travel unobstructed through the "channels" between aligned rows of atoms, without hard collisions with lattice atoms. On the other hand, no such open channels exist for an ion incident in a low symmetry direction, and hard collisions are bound to be much more frequent.

The trajectory of a channeled ion is influenced by the periodic screened coulomb potentials of the lattice atoms. Since the trajectory of a channeled ion will not in general be exactly coincident with the channel axis, it is pertinent to inquire about the nature and stability of a channeled trajectory. A detailed theoretical treatment of this question (as well as the phenomenon of channeling in general) has been given by Linhard [10].

Figure 7 illustrates schematically the trajectory of a channeled ion. The energy of the ion as it crosses the channel axis may be divided into two parts, $E_{\parallel}$ and $E_{\perp}$, associated with motion parallel and perpendicular to the atomic rows. As the ion moves closer to either row of atoms, it experiences repulsive coulomb forces. If $E_{\perp}$ is sufficiently small, the ion will be gently deflected back into the center of the channel by the cooperative action of several successive atoms. As a result, the ion will execute a stable oscillating trajectory in the channel with a spatial period long compared to the interatomic separation of the lattice atoms. In such a trajectory the minimum distance of approach between the ion and lattice atoms is of the order of a, where a is the Thomas-Fermi screening distance ($a \simeq 0.2$ Å for silicon).

Linhard [10] has calculated the maximum angle ψ_c that the trajectory of a channeled ion can make with the channel axis. For protons and helium ions in silicon with energies of the order of 100 keV or higher, the critical angle is given by

$$\psi_c \simeq \sqrt{E_1/E} \; ; \; E_1 = \frac{q^2 Z_i Z_t}{2\pi\varepsilon_o d} \tag{9}$$

where E_i is the energy of the ion and E_1 may be interpreted as a "barrier energy" for escape from the channel. For a silicon lattice, E_1 is approximately 100 eV for protons and 200 eV for helium ions. A 1-MeV helium ion would thus have a critical angle for channeling of about 0.8°.

[3] Whaling [9] gives a stopping power of 300 keV/μm for unchanneled 1 MeV helium ions in silicon. This stopping power has been assumed for all calculations in this paper. At ion energies much below 1 MeV, no extensive data on stopping power in silicon are available yet.

Even for beams well aligned along channel directions, however, a small fraction of the incident ions (typically a few percent) will make close encounters with *surface* atoms and be deflected out of channeled orbits. These ions form a dechanneled or "random" component of the beam. The dechanneled ions make uncorrelated collisions with lattice atoms and contribute to backscatter yield as if they were initially incident in a low symmetry direction. As it is unlikely for a dechanneled ion to be subsequently scattered to within the critical angle of a channeled direction, the "random" component of the beam is not likely to become channeled.

The rate of energy loss of an ion in a channeled direction is less than that in a low symmetry direction. This must be kept in mind when comparing energy spectra for low symmetry and channeled beam directions, as the depth scales for the two spectra will not be the same. No accurate theoretical predictions for the rate of energy loss in channeled directions are yet available, but a few experimental measurements have been made [11,12]. For the energies of interest, the spatial rate of energy loss of a channeled helium ion may be assumed to be between 50% and 100% of the spatial rate of energy loss of an unchanneled helium ion.

4.2. Use of Ion Channeling to Determine Impurity Lattice Locations

Consider the case where impurities are present in the lattice. Figure 8 is a schematic diagram of the silicon lattice in the ($1\bar{1}0$) plane, showing the substitutional sites (where silicon atoms normally lie), and tetrahedral interstitial sites (which are unoccupied in a perfect crystal). An impurity atom can lie on either of these two sites, or it can lie on some other interstitial site. Suppose that all the impurities are on substitutional sites. Then ions can channel along both <110> and <111> directions (as well as other directions of high symmetry) without suffering large angle deflections from impurity atoms.[4] The backscatter yield from *impurity* atoms can thus be expected to be very low for beams aligned along channeling directions, the yield being determined, as before, by the dechanneled portion of the beam.

Now suppose that all the impurities occupy tetrahedral interstitial sites. (Note that only a small fraction of such sites would be occupied for impurity concentrations normally encountered.) Then (as Fig. 8 shows) the impurity atoms would obstruct the <110> channels, but not the <111> channels. The impurity atoms could therefore backscatter ions channeled in the <110> direction but not ions channeled in the <111> direction. One would expect to observe a large attenuation in the backscatter yield from the impurity atoms when a beam is oriented from a low symmetry direction to the <111> direction, but no attenuation in the <110> direction.

Lastly, consider the case where all the impurities are in interstitial sites that do not lie along lattice rows in either the <110> or <111> directions. These impurities can backscatter ions regardless of whether they are channeled in the <110> or <111> directions, or unchanneled. In this case no attenuation in the scattering yield should be observed as the ion beam is oriented from a direction of low symmetry to either the <110> or <111> directions.

From the foregoing discussion it is evident that the scattering yield from impurity atoms as a function of beam orientation can provide information on the lattice locations of the impurities. We shall illustrate its application by returning to Fig. 5, where all three spectra were taken for the same dose of helium ions on target. The backscatter yield from the silicon substrate atoms for <110> and <111> beam directions is seen to be a factor of 20–30 times lower than the yield for a low symmetry direction.[5] The attenuation becomes less pronounced at lower energies (corresponding to more deeply penetrating ions). This is probably due to the gradual dechanneling of the ions in the <110> and <111> directions, as a result of lattice defects and thermal vibrations. Note also a peak just prior to the edge of the silicon portion of the spectrum for both <110> and <111> beam directions. This peak was consistently observed, even on undoped substrates. It suggests a thin disordered layer (possibly oxide) at the surface of the silicon. The width of the peak is comparable to the detector resolution (about 20 keV FWHM, equivalent to approximately 300 Å of depth) and so the thickness of the disordered layer may be much smaller than the width of the peak suggests.

Consider next the arsenic portion of the spectrum (channels 65–85). The backscatter yield for the <110> and <111> beam directions is seen to be only about a factor of 3 lower than for the low symmetry beam direction. This indicates that there is a significant fraction of arsenic atoms in nonsubstitutional sites. These nonsubstitutional arsenic atoms do not appear to be on tetrahedral interstitial sites, for if they were, the scattering yield for the <110> beam direction would exceed the yield for the <111>

---

[4]Actually, this will take place as long as the impurity nucleus is located within a distance of the order of a from the atomic rows. Thus an impurity which appears "substitutional" may actually be located a few tenths of an angstrom away from a true substitutional site.

[5]The backscatter yield for a low symmetry direction depends on the angle of incidence (θ_1) of the ion beam (see (7)). Therefore, the backscatter yield for a channeling direction should be compared with the yield for a low symmetry direction close to the channeling direction. In Fig. 5 the uppermost spectrum was taken with the beam a few degrees off the <111> axis. The backscatter yield for a beam direction a few degrees off the <110> axis was slightly higher. The <110> spectrum in Fig. 5 has been adjusted (i.e., lowered slightly) to take this into account.

beam direction. Note also the presence of a peak prior to the edge of the arsenic portion of the spectrum, for both <110> and <111> beam directions. This suggests a layer at the silicon surface where the arsenic is primarily in nonsubstitutional sites, and could be related to the disordered surface layer indicated by the peak at the edge of the silicon portion of the spectrum. However, the magnitude of the peak prior to the edge of the impurity spectrum has been found to vary with processing conditions, and its significance, if any, has not yet been established.

From the data of Fig. 5, it appears on first sight that about two-thirds of the arsenic atoms are in substitutional sites, and one-third in (nontetrahedral) interstitial sites. However, several complicating factors should be kept in mind when trying to interpret the data quantitatively. First, even for a beam well aligned along the <110> and <111> directions, a small portion (3–5%) of the beam is not channeled, and thus would backscatter from impurities regardless of location. In the absence of the unchanneled portion of the beam, the scattering yield for the <110> and <111> directions would be a few percent lower.

The second complicating factor is that for a given channeling direction, not all nonsubstitutional impurities necessarily obstruct the channel. For the sake of argument, consider an impurity atom lying in line with a row of silicon atoms in the <110> direction but displaced from a substitutional site. Such an impurity atom would not backscatter ions channeled in the <110> direction. However, there are five other axes equivalent to the <110> axis, and inspection of the silicon lattice will show that such a nonsubstitutional impurity atom could backscatter an ion channeled in any of the other five equivalent directions. Since by symmetry the same backscatter yields should be observed for all equivalent beam directions, one concludes that *at most* one-sixth of the nonsubstitutional atoms can escape detection by not backscattering ions channeled in the <110> direction (or any other single equivalent direction). In the example of Fig. 5, this effect could involve up to 7% of the arsenic atoms.

It is only meaningful to compare yields of ions backscattered from the same incremental layer dx. Thus, another factor that should be taken into account is the dependence of the depth-energy scale (i.e., the incremental depth represented by a given channel) on beam direction. This arises from differences in angle of incidence, as well as differences in stopping power (i.e., rate of energy loss of the ion). Effects on the scattering yield due solely to changes in angle of incidence can be predicted (see Eq. (7)) or measured (by taking spectra in different low symmetry directions). After this has been taken into account (see footnote 5), there still remains an uncertainty because the rate of energy loss of a channeled ion is not well known. The maximum error can be estimated by assuming the stopping power of a channeled ion in its ingoing track to be one-half the stopping power for a low symmetry direction. For the example under discussion, the outgoing track of the ion after the scattering event was never in a channeling direction, and so there was no uncertainty regarding the stopping power in the outgoing portion of the ion's trajectory. For the example of Fig. 5, the maximum error in the depth-energy scale due solely to uncertainties in stopping power was about 20% (i.e., the depth corresponding to one channel in the <110> or <111> direction was at most 20% larger than for the low symmetry direction). This could at most lead to a 20% overestimation of the fraction of nonsubstitutional impurities.

Finally, it should be mentioned that there can be a significant focusing of channeled ions toward the channel axis [13]. This tends to enhance the backscatter yield from impurity atoms lying close to the center of the channel. Recent experimental evidence of this effect has been reported [13,14] for interstitial impurities in silicon, where the backscatter yield from the impurities was seen to *increase* appreciably in the <110> direction relative to a low symmetry direction. At the time of writing, the implications of this effect have not been fully determined, but it is suggested that results be interpreted with caution, especially when a large fraction of the impurity atoms are in interstitial sites.

Neglecting the focusing effect just mentioned, the uncertainties discussed earlier would give a figure of between 60% and 75% for the fraction of arsenic atoms in substitutional sites, the rest being in (nontetrahedral) interstitial sites.

Many examples of the use of ion channeling to determine lattice locations of impurities may be found in the literature (e.g., Refs. 1, 3, 4; it is not possible to list all the work done in this area).

4.3. Use of Ion Channeling to Investigate Crystal Structure

Since channeling occurs only in well-ordered material, the crystallinity of a sample can be investigated qualitatively by observing whether channeling exists. For the sample of Fig. 5, good crystalline order was indicated by the large attenuation in the scattering yield from the silicon lattice atoms as the beam was oriented from a low symmetry direction to one of the major crystallographic directions. If the surface of the sample had been damaged (as, for example, by ion implantation), the attenuation of the scattering yield would have been diminished or absent, over that portion of the spectrum corresponding to the damaged layer. Many examples of the use of channeling to investigate crystal order may be found in the literature [1,15].

By observing the beam directions for which channeling occurs, one can obtain information regarding the lattice structure and crystallographic orientation of a crystalline sample. For example, Fig. 9 is a recorder chart showing the orientation dependence of scattering yield for a typical silicon substrate cut close to the (111) plane. The sample was rotated about its own axis from $\phi = 0°$ to $\phi = 360°$ (i.e., one

rotation), with the axis of the sample tilted 10° from the beam direction. The discriminator (see Fig. 2) was set so that the frequency meter counted all pulses above channel 50. (The silicon spectral edge was around channel 65.) Figure 9 shows the analog output from the frequency meter.

Many dips are evident in the recorder trace. In particular, the major dips labeled A through F can be identified as arising from channeling along the {1$\bar{1}$0} family of planes. (The dips are not as pronounced as those obtained by swinging the beam through a major crystallographic axis.) The crystal symmetry and orientation become more evident if the angular coordinates of the dips A through F are plotted on polar graph paper with the tilt angle θ_1 as radial coordinate, as shown in Fig. 10. It is seen that the chords AD, BE, CF intersect at a common point, which gives the angular coordinates of the <111> axis ($\theta_1 = 4.2°$, $\phi = 349°$). The chords themselves give the angular positions of the {1$\bar{1}$0} family of planes.[6] Once the above information has been obtained, the alignment of the sample is precisely known.

5. Concluding Remarks

In concluding, it is pertinent to make a few remarks about the sensitivity and limitations of the technique.

Using our relatively simple experimental arrangement and rudimentary electronics, and allowing 4–5 hours for taking each set of three spectra, we have detected heavy impurities in silicon at concentrations down to $10^{18}/cm^3$. One of the main factors limiting the sensitivity of the technique is the overlapping ("pile-up") of pulses reaching the multichannel analyzer. Thus two pulses at channel 40, if they coincided exactly, would give an erroneous count at channel 80. This effect gives rise to an extended tail beyond the silicon spectral edge which tends to submerge the impurity spectrum, especially for low impurity concentrations. The probability of pulse overlap can be reduced by using lower counting rates, but the time needed to accumulate meaningful statistics quickly becomes impractically long.

The overlapping of pulses can also be reduced, and the sensitivity somewhat improved, by using faster and more sophisticated electronics. A vast improvement in sensitivity would probably result from energy selection of backscattered ions before they reached the detector. If only the relatively high energy ions (scattered from impurity atoms) were allowed to reach the detector, there would be no interference from overlapping pulses due to scattering from the silicon substrate.

The technique, as we have seen, is essentially nondestructive if only thin surface layers are of interest. Some damage will be caused by the ion bombardment, but most of this damage occurs toward the end of the ion's range (about 4 μm for 1 MeV $^4He^+$ on Si) and so does not appreciably affect the surface layer under observation.[7] On the other hand, one has the option of profiling to much greater depths by combining the technique with incremental stripping.

Perhaps the most serious limitation of the ion backscattering technique as described here is that the impurity atom to be investigated must have an appreciably higher atomic mass than the atoms of the host material. When this condition is not satisfied, other interaction products should be sought. For example, the lattice location of boron atoms in silicon has been investigated by detecting the alpha particles produced in the nuclear interaction $^{11}B\ (p,\alpha)\ ^8Be$ [6]. Characteristic X-ray emissions are also useful, especially when a high-resolution spectrometer is available.

As mentioned earlier, some caution must be exercised in interpreting the results, especially with regard to lattice locations of impurities. It is also important to realize that the electrical activity of an impurity is not necessarily related to the atomic fraction of the impurity in substitutional sites. This has been demonstrated by studies on the annealing characteristics of ion-implanted layers [1].

Despite its limitations (some of which may be removed after further work), the ion-backscattering technique promises to be a very useful complement to other techniques for investigating impurities in semiconductors.

6. References

[1] Mayer, J. W., Eriksson, L., and Davies, J. A., *Ion Implantation in Semiconductors* (Academic Press, 1970).

[2] Gyulai, J., Meyer, O., Mayer, J. W., and Rodriguez, V., *Appl. Phys. Letters* 16, 232 (1970).

[3] Chou, S., Davidson, L. A., and Gibbons, J. F., *Appl. Phys. Letters*, to be published July 1970.

[4] Gibson, W. M., in *Proc. Conf. on the Use of Small Accelerators for Teaching and Research*, J. L. Duggan, ed. (1968).

[6] The directions of symmetry of a crystal can be conveniently mapped using stereographic projections [16, 17]. Figure 10 is actually an enlargement of the central portion of a stereographic projection. The radial scale of such a projection is proportional to θ_1 only when θ_1 is small.

[7] However, heavy doses of ions may affect the lattice locations of certain impurities. After taking all the necessary data, it is wise to check on the reproducibility of the spectra for the channeled beam directions.

[5] Khan, J. M., Potter, D. L., and Worley, R. D., *Phys. Rev.* 139, A1735 (1965).

[6] Fladda, G., Bjorkqvist, K., Eriksson, L., and Sigurd, D., *Appl. Phys. Letters* 16, 313 (1970).

[7] Bohr, N., *Kgl. Danske Videnskab. Selskab.,Mat.-Fys. Medd.* 18, No. 8 (1948).

[8] Any text on theoretical mechanics, e.g., Bradbury, *Theoretical Mechanics* (Wiley, 1968).

[9] Whaling, W., *Handbuch der Physik* 34, 193 (1957).

[10] Linhard, J., *Kgl. Danske Videnskab. Selskab., Mat.-Fys. Medd.* 34, No. 14 (1965).

[11] Erginsoy, C., Wegner, H. E., and Gibson, W. M., *Phys. Rev. Letters* 13, 530 (1964).

[12] Westmoreland, J. E., Mayer, J. W., Eisen, F. H., and Welch, B. M., "Evaluation of Disorder in Boron Implanted Silicon," presented at the Int. Conf. on Ion Implantation in Semiconductors, Thousand Oaks, California, May 4–7, 1970. (Proceedings to be published in *Radiation Effects*.)

[13] Andreasen, O., Andersen, J. U., and Davies, J. A., "The Use of Channeling Effect Techniques in Locating Interstitial Atoms in Silicon," presented at the Int. Conf. on Ion Implantation in Semiconductors, Thousand Oaks, California, May 4–7, 1970. (Proceedings to be published in *Radiation Effects*.)

[14] Johansson, N.G.E., Fladda, G., and Domeij, B., "Rutherford Scattering Yields of More than 100% of Normal in the <110> Scattering of ^{12}C-Ions from Zr, Tl and Hg Implanted with Silicon," presented at the Int. Conf. on Ion Implantation in Semiconductors, Thousand Oaks, California, May 4–7, 1970. (Proceedings to be published in *Radiation Effects*.)

[15] See the Proc. Int. Conf. on Ion Implantation in Semiconductors, Thousand Oaks, California, May 4–7, 1970 (to be published in *Radiation Effects*).

[16] Cullity, B. D., *Elements of X-ray Diffraction* (Addison-Wesley, 1956).

[17] Hilton, H., *Mathematical Crystallography and the Theory of Groups of Movements* (Dover Publications, 1963).

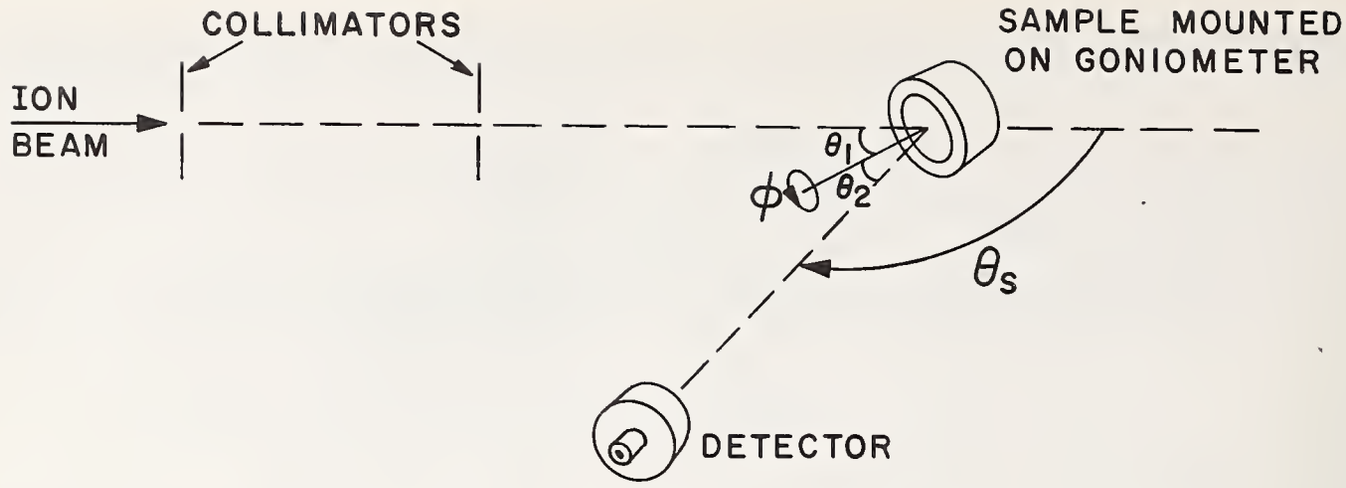


Fig. 1. Schematic diagram of experimental arrangement for ion-backscattering work.

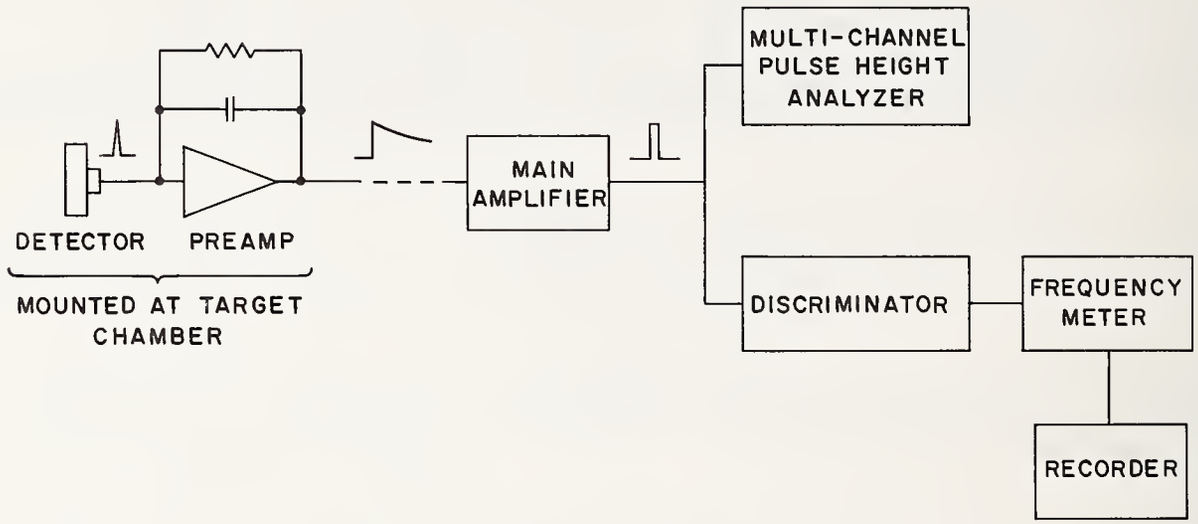


Fig. 2. Simplified block diagram of detector electronics.

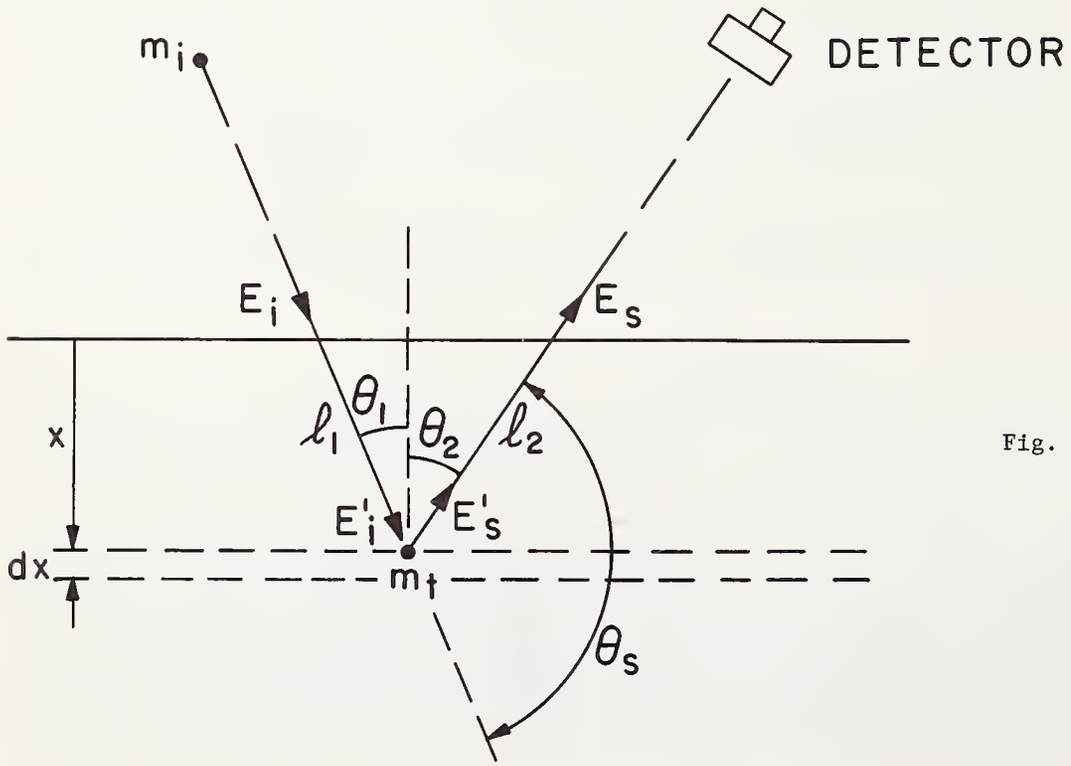


Fig. 3. Idealized representation of a large-angle ion scattering event.

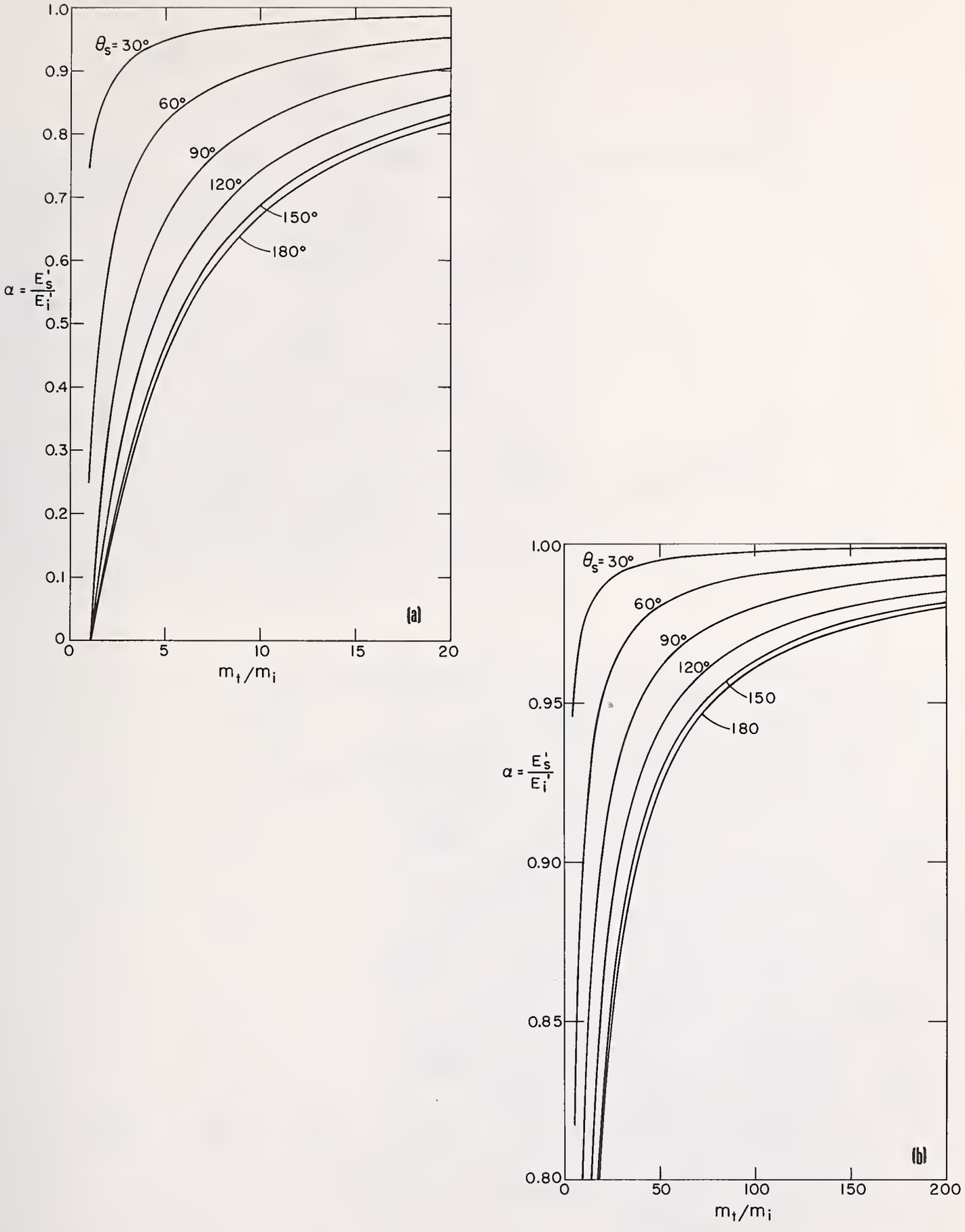


Fig. 4. Plots showing the fraction of energy retained by an ion in an elastic scattering event, as a function of scattering angle and target-to-ion mass ratio.

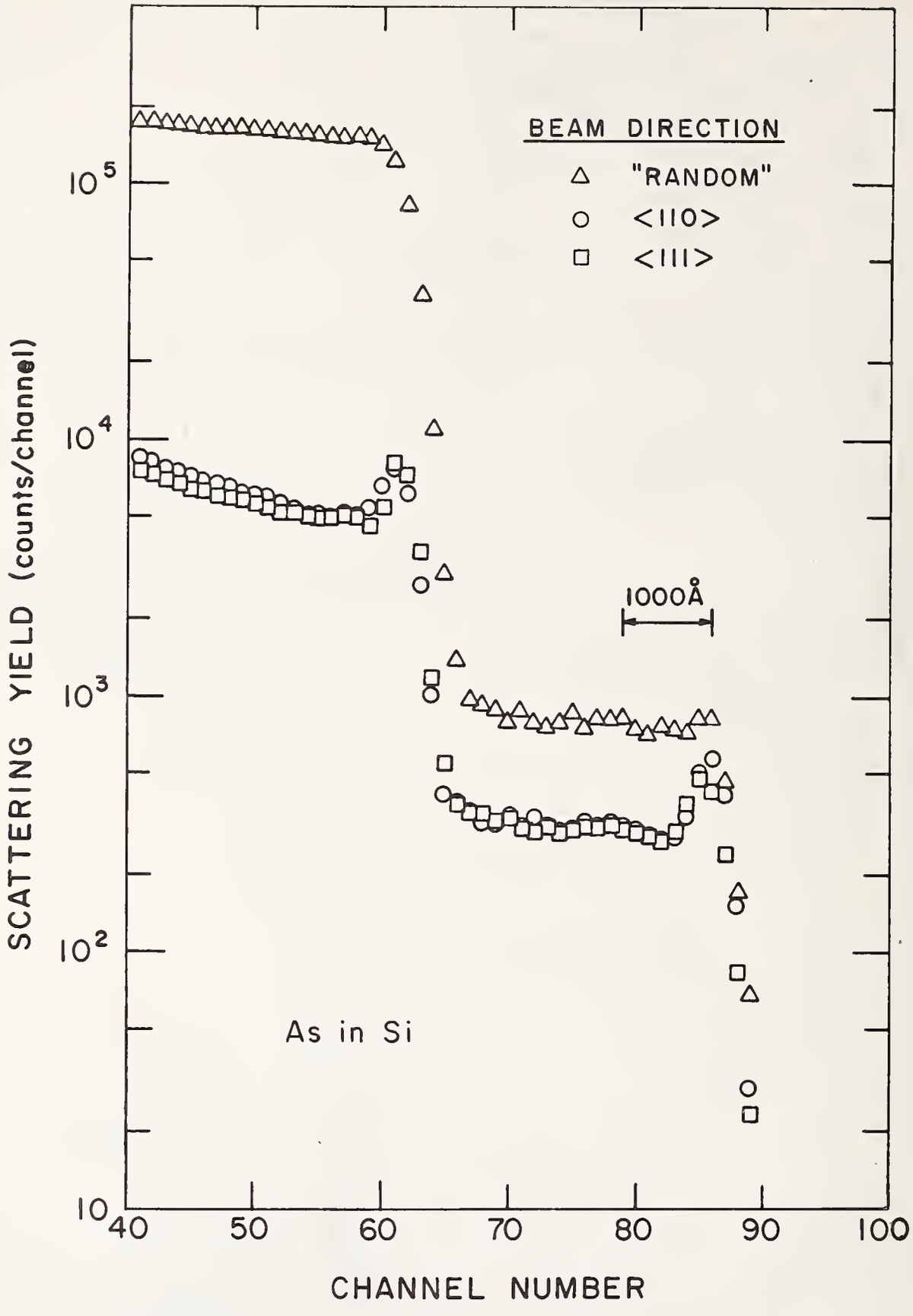


Fig. 5. Energy spectra of helium ions scattered from an arsenic doped silicon layer.

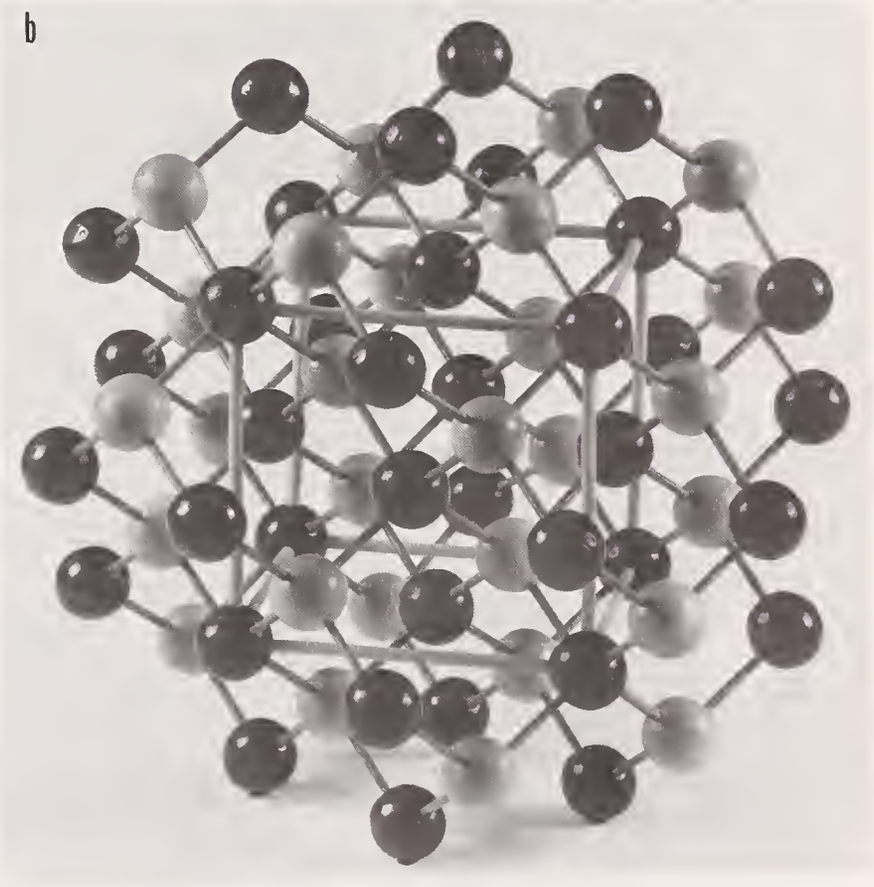


Fig. 6. Views of a diamond-type lattice from (a) a <110> direction, and (b) 10° off a <110> direction.

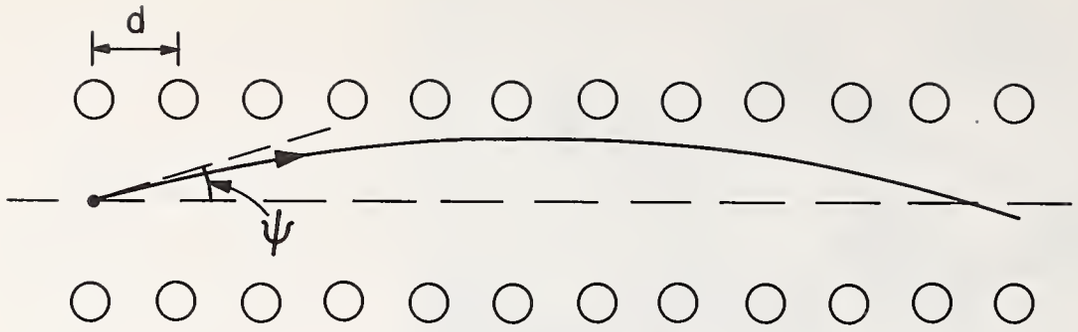


Fig. 7. Schematic diagram of the trajectory of a channeled ion.

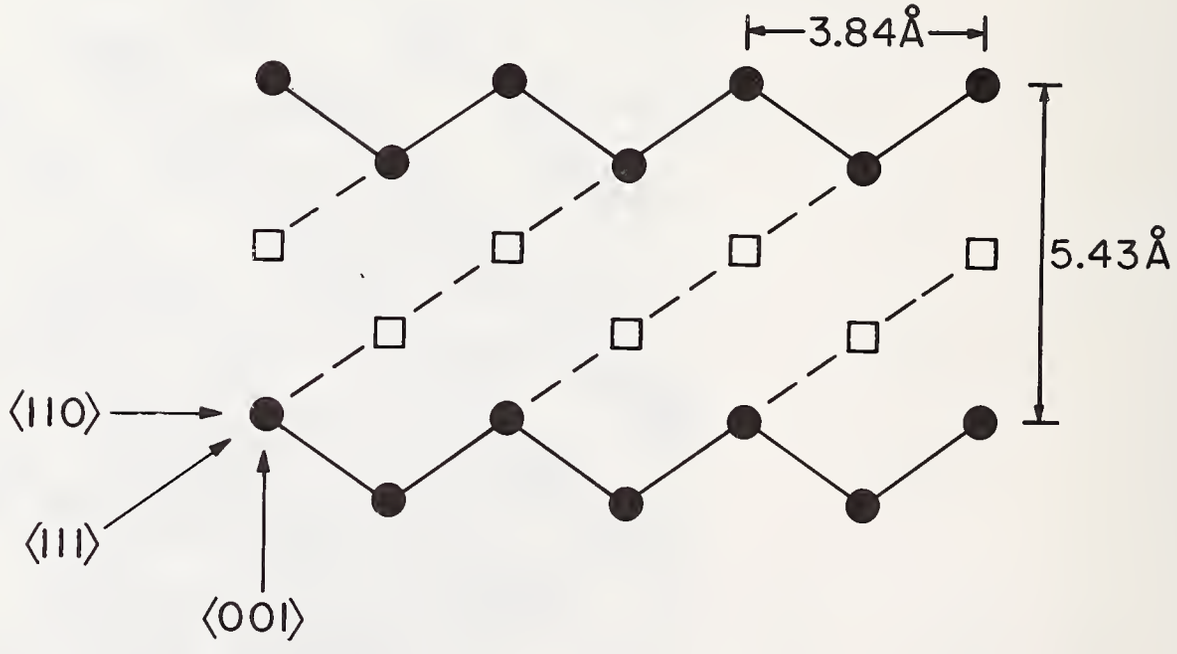


Fig. 8. Cross section of the silicon lattice in the $(1\bar{1}0)$ plane, showing substitutional sites (solid circles) and tetrahedral interstitial sites (open squares).

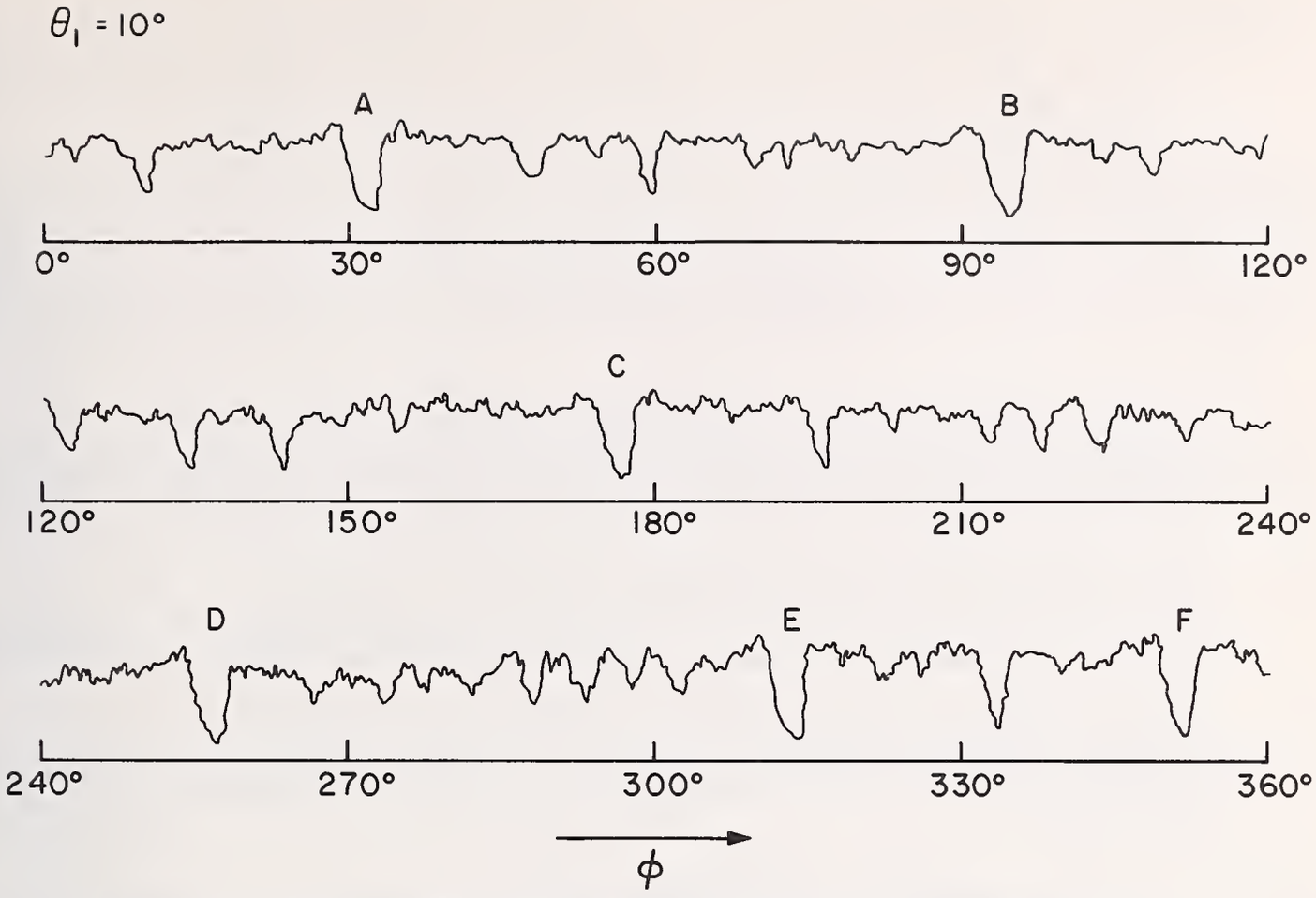


Fig. 9. Recorder output showing the dependence of helium ion backscattering rate on beam orientation, for a silicon sample

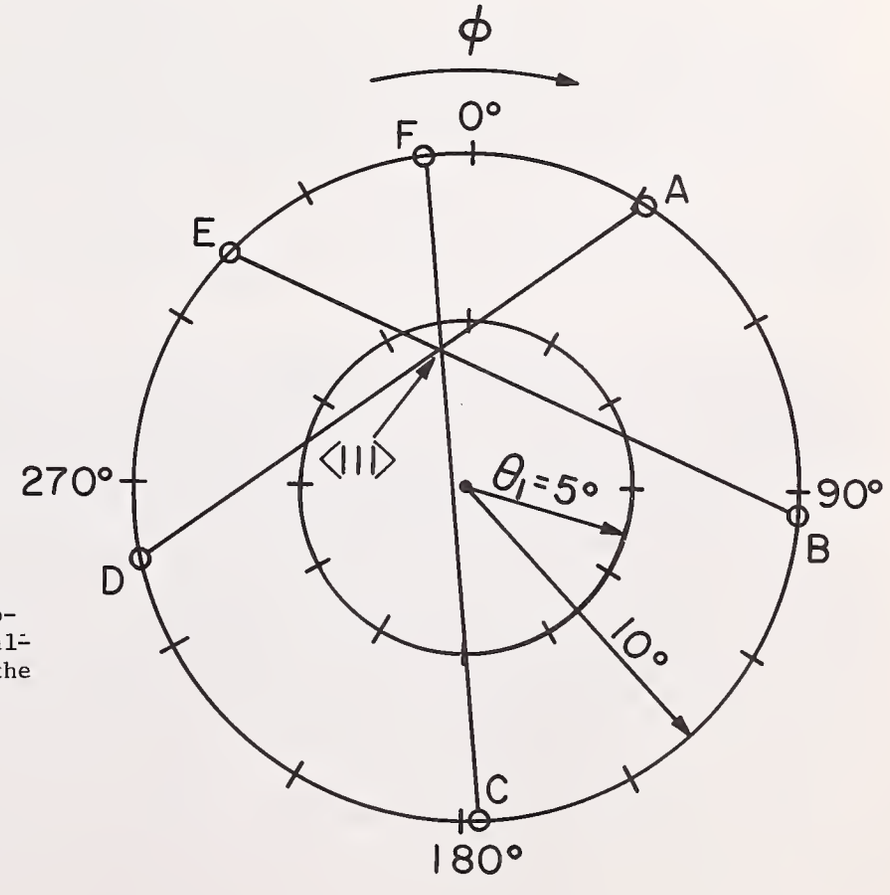


Fig. 10. Enlarged stereographic projection showing the crystallographic orientation of the silicon sample of Fig. 9.

Determination of Diffusion Coefficients in Silicon and Accepted Values

M. F. Millea

Aerospace Corporation
El Segundo, California 90245

The major emphasis of this review is directed towards a critical evaluation of experimentally determined diffusion coefficients of substitutional elements in silicon. A generalized definition of atomic migration is developed using the Onsager-Fuoss phenomenological approach and the conditions which must be satisfied when using Fick's first and second laws discussed. Various diffusion mechanisms are considered to illustrate possible interactions between crystal properties and atomic migration with special attention given to the effect of heavy doping and induced space-charge regions on diffusion. The effect of short-range solute-enhanced migration is also considered in discussing the effect of heavy doping on self-diffusion in silicon. The techniques commonly employed to determine atomic movement in silicon are presented; random and possible systematic errors are considered. The experimentally determined diffusion coefficients for Group IIIA, IVA and VA elements are examined and "accepted" diffusion coefficients for these elements in intrinsic silicon are proposed. Experimental studies of the effects of heavy doping on the self-diffusion and migration of tin, phosphorus, arsenic, antimony, and indium are then discussed.

Key Words: Diffusion, silicon, diffusion coefficients, accepted values, substitutional impurities, self-diffusion, heavy doping, induced field, chemical potential, driving force, activity coefficient.

1. Introduction

Atomic motion of Group IIIA, IVA and VA elements in silicon has received a large amount of attention for the past fifteen years. During this short span the study of atomic migration in silicon has become by far the most widely investigated diffusion system. This fundamental study of atomic migration in silicon has, of course, been due to the usefulness of solid state diffusion for fabricating semiconductor devices. Not only has the practical application of diffusion technology encouraged a detailed knowledge of atomic migration, but it has also revealed new aspects of diffusion in silicon which in the normal course of research would not have been so rapidly uncovered. In addition, the degradation of silicon devices by space radiation has encouraged fundamental studies of lattice defects in irradiated silicon. The results have been of significant value in understanding atomic motion.

Although the experimental investigation of diffusion of substitutional elements in silicon effectively started with Fuller and Ditzenberger's [1][1] classical paper, it is best to begin the history of atomic movement in silicon by reviewing the earilier work on germanium. The behavior of substitional (i. e. Group III, IV and V) elements in germanium indicated that the diffusion could be explained by a simple acceptor-type vacancy mechanism [2]. Group V (donors) atoms were found to diffuse faster than Group III (acceptors) or Group IV (neutrals) elements. This can easily be explained by a vacancy mechanism in which the vacancy acts as an acceptor. The negatively charged vacancy would be attracted to the positively charged donors. The neutral Group IV elements and negatively charged Group III atoms would not attract negative vacancies and, therefore, interact less frequently with vacancies and migrate slower. At about this time Reiss [3] showed that the solubility of an electrically active defect depends on the free electron concentration. Longini and Green [4] extended this work to vacancies and showed that if vacancies were acceptors their solubility would increase with n-type doping and decrease with p-type doping. Following Bardeen's suggestion that this behavior could be utilized for the study of atomic migration in semiconductors, Valenta and Ramasastry [5] investigated the effect of doping on the self-diffusion of germanium and found an increased diffusion coefficient in n-type

[1] Figures in brackets indicate the literature reference at the end of this paper.

material and a lower value in p-type as expected from a vacancy diffusion mechanism. Valenta extended these measurements to Group IIIA and VA elements and found a similar dependence [6]. At about this time the migration of fast diffusing elements was also being investigated. Lithium was shown to exist and diffuse as an interstitial donor [7]. The apparently erratic behavior of copper migration in germanium was brilliantly explained by Frank and Turnbull's "dissociation mechanism" [8].

The early hope that the diffusion mechanisms in silicon would closely parallel the explanation employed for germanium has not been completely fulfilled. The interstitial and dissociation mechanisms have accounted for the fast diffusers (e.g. lithium, [7] copper [9] and gold [10]). However, the fact that the migration of substitutionals in silicon is more complex than in germanium was indicated by the work of Fuller and Ditzenberger [1], who found that with the exception of phosphorus, the substitutional donors have smaller diffusion coefficients than acceptors. Although approximately a hundred papers have been published on atomic diffusion in silicon since Fuller and Ditzenberger's 1956 article, the question of the substitutional diffusion mechanism remains unresolved [2], [11].

2. General Theory

It is customary in discussing atomic migration to define a diffusion coefficient using Fick's first law which states that the particle flux density is proportional to the concentration gradient. Since silicon has a cubic crystal structure, the diffusion coefficient is a scalar. For atomic motion in one dimension

$$J = - D \, (dN/dx). \tag{1}$$

Application of the continuity equation along with the assumption that the diffusion coefficient is independent of concentration yields Fick's second law

$$dN/dt = D(d^2N/dx^2). \tag{2}$$

The techniques for solving these equations subject to the appropriate boundary conditions are readily available [12]. The temperature dependence of the diffusion coefficient experimentally determined in the above manner is found to obey the Arrhenius relation

$$D = D_o \exp (-Q/kT) \tag{3}$$

where k is Boltzmann's constant, T is the absolute temperature, and Q is the diffusion activation energy. The term D_o is usually observed to be temperature independent. Measured values of D_o and Q are expressed in units of cm^2/sec and electron-volts (23.06 kcal = 1eV), respectively.

Extreme caution must be exercised in using the above approach to determine diffusion coefficients since this procedure is only valid for very limiting conditions. First, the atomic motion is not independent of concentration or concentration gradients, and therefore, the physical interpretation of Eq. (1) and (2) becomes obscure. Secondly, this procedure assumes that the driving force for atomic motion is a concentration gradient. This assumption is only valid under very restrictive conditions. It is now generally recognized that the driving force for atomic motion is the chemical potential gradient and not the chemical gradient alone [13]. The flux in the x direction of the i species in a system of m constituents is

$$J_i = \sum_{j=1}^{m} M_{ij} \, \partial\mu_j/\partial x \tag{4}$$

where M_{ij} is the appropriate mobility and μ_j the chemical potential. A similar set of equations exist for each of the remaining m-1 constituents. The formal solution of these equations is, of course, complex. Therefore, only certain limiting cases will be considered here. Consider first the migration of an impurity diffusing in pure silicon by an interstitial mechanism (e.g. lithium which exists in silicon as an interstitial donor). To a first approximation the atomic motion of this species will respond only to its own chemical potential gradient, and we may simplify Eq. (4) to

$$J = - M \, \partial\mu/\partial x. \tag{5}$$

The chemical potential can be written as

$$\mu = \mu_o + kT \ln \gamma X \tag{6}$$

where μ_o is the chemical potential of the standard state, γ is the activity coefficient (i.e. $\gamma = \exp^{o}(\Delta H/kT)$ where ΔH is the differential heat of solution) and X is the mole fraction of the i species. The mole fraction term results from the entropy of mixing, and the activity coefficient represents all other energies. A diffusion coefficient is defined by an expression similar to the Nerst-Einstein relation as

$$D = kTM/N. \tag{7}$$

Assuming an electric field does not exist, Eq. (5) becomes

$$J = - D \left[1 + d(\ln \gamma)/d(\ln N\right] dN/dx \tag{8}$$

When the impurity is diffused into silicon at a low concentration, the activity coefficient and D are independent of concentration. Under these limiting conditions Eq. (10) reduces to Fick's first law.

When the diffusing species are electrically active and an electric field is present, one must use the electrochemical potential. One could explicitly include the electrostatic potential in Eq. (6) as

$$\mu = \mu_o + kT \ln \gamma X + qZ\Psi \tag{9}$$

or incorporate the electrostatic potential into the activity coefficient as

$$\gamma = \gamma_o \exp (qZ\Psi/kT). \tag{10}$$

In these equations qZ is the electronic charge and Ψ is the electrostatic potential. The effect of an electric field on the migration of an electrically active impurity is usually described by assuming that the atomic motion results from two forces: the concentration gradient and the electric field. For this case

$$J = - D\, dN/dx + MqZNE. \tag{11}$$

It has become standard practice to define an effective diffusion coefficient as

$$D_{eff} = D_I \left[1 + qZNE/(kT\, dN/dx)\right] \tag{12}$$

where D_I is the diffusion coefficient in intrinsic silicon. If the electric field is due to a nonuniform concentration of the diffusing specie, Smits [14], has shown assuming quasi-neutrality [15], that this equation becomes

$$D = D_I \left[1 + N/(4n_I^2 + N^2)^{\frac{1}{2}}\right] \tag{13}$$

where n_I is the intrinsic electron concentration. When $n_I >> N$, the induced electrical field is unimportant. According to this equation, the maximum increase in the effective diffusion coefficient which can occur as a result of an induced field is a factor of two. However, Shockley has pointed out that this increase can be even larger [16]. For degenerate material he evaluated Eq. (11) as

$$D = D_I \left[1 + (\pi/6)^{1/3}(N/N_c)^{2/3}\right] \tag{14}$$

where N_c is the density of states in the conduction band.

If in place of considering the combined action of a concentration gradient and an electric field on the particle flux one uses the chemical potential defined in Eq. (8), an equation identical to Eq. (11) is obtained.

When interstitial migration occurs in silicon doped with a second impurity a more complicated situation exists. The interstitial migration driving force depends not only on its own chemical potential gradient but also on the gradient of the chemical potential of the second impurity. In addition, the presence of the second impurity might change the effective mobility of the interstitial. If immobile complexes are formed between the two impurities, the effective mobility of the diffusing species will be reduced in proportion to the fraction of immobile pairs. The influence of the second impurity might also be extensive because of effects such as induced lattice distortion and solubility enhancement.

Substitutional migration is normally more complex than pure interstitial diffusion since several different species are usually responsible for the atomic movement. The thermodynamical formulation used above can yield the driving force for atomic motion. However, to calculate the atomic mobility one is forced to employ a microscopic kinetic approach since the atomic mobility depends on the mechanism by which diffusion occurs. Various attempts have been undertaken to calculate the atomic mobilities of substitutional elements [2], [11], [17]. It would be inappropriate to consider these theoretical estimates in detail here since the emphasis in this review is placed on the experimental aspects of diffusion; the various diffusion mechanisms will therefore only be briefly described.

Vacancy:

In a vacancy mechanism a substitutional atom adjacent to a vacant lattice site migrates by interchanging with the vacancy. The possibility that a divacancy is responsible for substitutional motion and not a monovacancy has also been proposed [11], [18].

Dissociation:

In this mechanism the diffusing constituent exists in both substitutional and interstitial positions. Atomic motion occurs by a substitutional atom moving into an interstitial position where it rapidly diffuses until it recombines back to a substitutional either by reacting with a vacant lattice site [8] or by pushing a silicon atom into an interstitial position [2]. This model has been very successful in explaining the migration of copper [9] and gold [10] in silicon.

Interstitialcy [19], [20]:

In this mechanism an interstitial atom pushes a substitutional atom into an interstitial position and occupies the vacated site.

Direct Interchange [19], [20]:

Direct interchange consists of two substitutional atoms exchanging their lattice positions. A variation of this mechanism is ring interchange where several atoms participate in the interchange.

As an illustration of the different type of species which might have to be allowed for in setting-up Eq. (4), consider phosphorus diffusing via vacancies. From conclusions based on radiation damage studies on silicon, a lattice vacancy is believed to exist in one of four possible charge states (i.e. V^{+}, V^{0}, V^{-} and V^{-2}) [21]. Strong pairing between phosphorus and vacancies has also been observed in irradiated silicon [22]. These complexes are thought to exist as neutral or negatively charged pairs. It has been suggested that phosphorus diffuses mainly as E^{0} centers (i.e. an uncharged phosphorus-vacancy pair [22]). A complete set of equations for phosphorus diffusion via vacancies would require one to consider 10 different species (i.e. charged and uncharged phosphorus, four different vacancy charged states, two complex charged states, and also the densities of electrons and holes). A general solution to Eq. (4) under the above conditions is formidable. For the present discussion, it will therefore be assumed that the chemical potential of free carriers and vacancies is constant. This assumption might not be valid for all diffusion processes employed in fabricating semiconductor devices, because the diffusion times used might not allow for the establishment of a sufficiently constant vacancy chemical potential. For the present, it will be assumed that the free carrier densities are constant throughout the specimen so that no electric field is present. We are interested in calculating how the diffusion via vacancies depends on the electron concentration. Pair diffusion will be considered later.

Consider first the migration via single negatively charged vacancies. Following Reiss [3a], [23], [24] we write the transition from an uncharged to a charged state as

$$V^{\circ} + e^{-} \rightleftarrows V^{-} \tag{15}$$

Since the Gibbs function is a minimum at equilibrium, one can combine the chemical potential of these three species as

$$K(T) = [V^{\circ}]\,[e^{-}]/[V^{-}] \tag{16}$$

where concentrations have been written for mole fractions. For regular solution behavior (i.e., γ_{Ω} = constant) the equilibrium constant K(T) is independent of concentrations. Since the concentration of neutral vacancies, $[V^{\circ}]$, is independent of electron concentration, the ratio of the concentration of $[V^{-}]$ in doped and intrinsic silicon is

$$[V^{-}]_D/[V^{-}]_I = n_D/n_I = \exp\,(q\Psi/kT) \tag{17}$$

where n_D is the electron concentration in the doped crystal and Ψ is the electrostatic potential measured relative to the Fermi level in intrinsic silicon. This equation can be generalized to any vacancy charged state Z

$$[V^{Z}]_D/[V^{Z}]_I = (n_I/n_D)^{Z} = \exp\,(-qZ\Psi/kT) \tag{18}$$

If the diffusion coefficient is proportional to the concentration of vacancies of charge Z, we can write

$$D_D/D_I = (n_I/n_D)^{Z} = \exp\,(-qZ\Psi/kT) \tag{19}$$

For donor-type vacancies, (Z = + 1), the migration increases in p-type material and decreases in n-type. Atomic movement through neutral vacancies is independent of the electron concentration. The effect of doping on a negatively charged vacancy is opposite that on a donor type: higher in n-type material and lower in p-type. A similar dependence would exist for divacancies which also have several charge states.

It does not appear reasonable to assume that the atomic motion is proportional to the total vacancy (or divacancy) density and thereby estimate the vacancy energy levels from the experimentally determined dependence of the diffusion on electron concentration. This procedure tacitly assumes that the diffusivity of a given vacancy is independent of its charge state.

The case where diffusion occurs through phosphorus-vacancy complexes is similar. The formation of a neutral pair, E°, can be written as

$$P^{\circ} + V^{\circ} \rightleftarrows E^{\circ} \tag{20}$$

Since the flux of phosphorus atoms is equal to the flux of E° pairs, one can write assuming regular solution behavior

$$J_P = J_{E^{\circ}} = -D_{E^{\circ}}\,(d[E^{\circ}]/dx) \tag{21}$$

where

$$[E^{\circ}] = K'\,[P^{\circ}]\,[V^{\circ}] \tag{22}$$

As before, K' and $[V^{\circ}]$ are independent of the electron concentration, but $[P^{\circ}]$ is not since it is an extrinsic defect. The concentration of neutral phosphorus depends on the total phosphorus concentration, P_T, as

$$[P^{\circ}] = P_T\,[1 + \exp(E_p - q\Psi)/kT]^{-1} \equiv P_T F \tag{23}$$

where $q\psi$ is electrostatic potential measured relative to the Fermi level which is assumed constant and is arbitrarily set equal to zero. The phosphorus donor level E_p, is assumed to be constant. For convenience, the degeneracy associated with the phosphorus level has been absorbed into E_p. Combining the above equations,

$$J_p = -D_{E^o} K' [V^o] F (dp_T/dx) \tag{24}$$

From this we may write a diffusion coefficient for phosphorus as

$$D = D_{E^o} K' [V^o] F . \tag{25}$$

The phosphorus diffusion coefficient relative to intrinsic silicon is

$$D_D/D_I = \left[1 + \exp (E_p/k_T)\right] \left[1 + (n_I/n_D) \exp (E_p/kT) \right]^{-1} . \tag{26}$$

This equation differs from Eq. (16) in that only neutral phosphorus form complexes with neutral vacancies (or conversely only positively charged phosphorus combine with negative charge vacancies). For $E_p >> kT$, Eq. (25) and (16) are identical. If migration via E^Z is considered, an equation similar to Eq. (25) is obtained except multiplied by a factor of $(n_I/n_D)^Z$. This dependence is also the same as when diffusion occurs by $[V^Z]$ if $E_p >> kT$.

An interesting situation develops when one considers the migration of phosphorus by E^o when a nonuniform electrostatic potential exists. An example of this is the diffusion of a high phosphorus concentration into intrinsic silicon. Since it has been assumed that phosphorus can only diffuse as a neutral pair, the induced field will not directly affect atomic migration. As before, the particle flux of phosphorus is equal to the flux of E^o. The E^o flux is proportional to the concentration gradient of E^o as in Eq. (21) since the activity coefficient of a neutral complex is independent of the electrostatic potential. Although the activity coefficient is a constant, the gradient of the pair concentration depends on the electrostatic potential because the concentration of neutral phosphorus depends on the electrostatic potential. That is

$$J_P = - D_{E^o} K'[V^o] F \left[dP_T/dx + (qP_TF/kT) (d\psi/dx) \exp (E_p - q\psi)/kT\right] \tag{27}$$

The electrical field does not directly increase the mass velocity of $[E^o]$, but indirectly by increasing the concentration gradient of $[E^o]$. It is easy to generalize this situation to migration through a phosphorus complex with net charge Z. The concentration of complexes is

$$[C^Z] = K''[P^o] [V^o] [e^-]^{-Z} \tag{28}$$

or

$$[C^Z] = K''[V^o] P_TF \exp (-qZ\psi/kT) \tag{29}$$

and the activity coefficient is

$$\nu = \nu_o \exp (qZ\psi/kT) . \tag{30}$$

The migration driving force is identical to the case of $[E^o]$ diffusion since the velocity dependence due to the induced electric field is balanced out by the dependence of concentration gradient on electrostatic potential. The particle flux is

$$J_P = - DK'' [V^o]F \exp (-qZ\psi/kT) \left[dP_T/dx + (qP_TF/kT) (d\psi/dx) \exp (E_p - q\psi)/kT\right] \tag{31}$$

A similar procedure can be used when an element diffuses by other than a vacancy-impurity pairs.

Returning now to diffusion in silicon with a uniform free carrier concentration: in practice it would not be expected that the migration of a given element would just proceed by one charged state of a given defect, or even perhaps by one type of defect. For sufficiently low defect densities the diffusion mechanisms are additive,

$$D = \sum_j D_j \quad . \tag{32}$$

The effect of electron concentration on the total diffusion coefficient can be approximately represented as a Laurent's series

$$D(n) = D_I \sum_{j=-\infty}^{\infty} d_j (n_I/n)^j \tag{33}$$

where D_I is the diffusion coefficient in intrinsic silicon, and j is defined as the diffusion charge state. A Laurent's series is preferred to a Taylor's because it has a direct physical interpretation. If the effect of heavy doping only influences migration by the long range electron concentration effect just discussed, the d_j coefficient represents the fraction of atomic motion in intrinsic silicon which occurs by a diffusion charge state j. The diffusion charge state is not necessarily equal to the charge state of the defect by which atomic motion proceeds. For example, the charge state for atomic motion by E^o pairs has j = - 1 whereas Z = 0. The d_j coefficients are expected to depend on temperature since the migration of the defect associated with the various d_j is a function of temperature. In applying Eq. (32) one must therefore experimentally determine these coefficients as a function of temperature.

The d_j coefficients will also depend on the type of impurity used to form the heavily doped region. Not only can doping influence the diffusion of a given species through its effect on the chemical potential of the defect responsible for migration (long range effect), the the presence of an impurity can enhance the migration by short range solute interactions. An example of this short range effect can be seen by considering self-diffusion in heavily doped crystals. Hoffman et al [25] have shown that self-diffusion is enhanced by the presence of a mobile impurity migrating by a vacancy mechanism as

$$D_D^o = (1 - \alpha X) D_I^o + \alpha X D \tag{34}$$

where D_D^o and D_I^o are the self-diffusion coefficients in doped and undoped crystals, α is the effective number of solvent-vacancy exchanges in the region near the impurity during the period required for the impurity to jump one lattice spacing, and X and D are the atom fraction and diffusion coefficients of the impurity for a vacancy mechanism α is nearly unity, but it can be greater [26]. If an impurity is migrating in silicon by forming a complex with a vacancy, at least four silicon atoms must move for the impurity to move one lattice spacing [22]. Solute enhancement of self-diffusion has been experimentally observed in metals although the theoretical explanation is still unresolved[27]. One might expect this effect to be more pronounced in silicon than in a metal. Substitutional impurity diffusivities relative to self-diffusion is much greater in silicon than in metals, and the impurity diffusivity can be greatly increased with doping through the long range electron concentration effect.

The effect of the migration of a mobile substitutional impurity on self-diffusion depends on the mechanism responsible for the impurity migration. For direct interchange α in Eq. (33) would be unity. It is possible under certain conditions for a substitutional impurity which diffuses interstitially not to enhance self-diffusion. (e. g. dissociation mechanism), whereas an interstitialcy mechanism would enhance self-diffusion. Short range solute self-diffusion enhancement should therefore yield valuable information on the detailed atomic migration motion.

3. Experimental Methods

The prime motivation in the determination of diffusion coefficients in silicon is to obtain significant veritable results. The judgement of "significant" in the investigation of atomic migration in silicon is relatively nebulous since one must balance the results both for their practical device utilization and their contribution to fundamental knowledge. The criterion for the evaluation of "veritable" is not nebulous. The normally accepted method of presenting experimental results is to report the average value of a set of measurements and the standard deviation of the mean. It is also customary to provide an error analysis indicating the source of the reported standard deviation of the mean and a detailed discussion of possible systematic errors. It becomes apparent after reviewing the large amount of published material devoted to the atomic motion in silicon over the past fifteen years that there is a great deal of uncertainity in the published diffusion coefficients. On the average there is a standard deviation of ± 0.5eV in the value of measured diffusion activation energy reported and of course a much wider variance in the pre-exponential factor, D_o. Considering the small temperature range over which substitutional diffusion coefficients are measured, this uncertainty in the values of Q and D_o are expected when the measured diffusion coefficients can range within approximately fifty per cent of the true value.

Systematic errors, which are seldom if ever discussed in the literature, appear to be more harmful in measuring diffusion coefficients than random errors. Random errors are easily detected and can be minimized by increasing the sample size or by improving the experimental technique. Systematic errors

are more insidious in that they can go undetected. The importance of systematic errors can be clearly seen by comparing the average values and the standard deviations of a number of measurements of the diffusion coefficient of a given element.

The following techniques have been employed to investigate atomic migration in silicon: radio tracer, radio activation, differential resistance, junction depth-sheet resistance, total conductance, drift mobility, ion-pairing, precipation, thermal quenching, radiation damage, permeation, and gaseous evolution [11], [28]. Aside from the first and last two techniques, atomic movement is detected electrically. As expected the nuclear techniques have yielded the most reliable measurements because this procedure is the least affected by systematic errors. The differential resistance method is perhaps the most widely employed of the electrical techniques for determining diffusivity, while the junction depth-sheet resistance procedure is used for monitoring diffused device fabrication. The drift mobility technique is of fundamental importance because it measures the mobility of the species under investigation and also determines the electronic charge. Ion-pair and precipation are useful in studying atomic motion of lower temperature where other procedures are not suitable. EPR studies of irradiated silicon have elucidated the electronic behavior of various defects and in some cases measured the motional activation energy [21], [22].

Although the radio tracer and radio activation techniques are possibly the best procedures for determining diffusion coefficients, the following potential sources of systematic errors must be considered: incorrect boundary conditions, non-parallel sectioning, incorrect diffusion time, induced space-charge, concentration effects, surface roughness, surface effects, non-equilibrium effects and contamination.

The occurance of these systematic errors can usually be detected by determining the effect of diffusion time on the measured diffusion coefficient at a constant temperature. One would hope that this will become the standard practice in the future and systematic errors which are so prevalent in published results will be greatly reduced.

4. Measured Diffusion Coefficients in Silicon

One would like to write an equation for the atomic motion of substitutional atoms in silicon in a manner which describes all diffusion situations of interest. In order to do this one would have to know the effect of temperature, electron concentration, induced field, induced lattice distortion, activity coefficient, short range solute enhancement, and the influence of non-equilibrium defect concentrations. If this knowledge existed, one would be in a position to explain the impurity profiles observed for multiple diffusions. Unfortunately, this is not possible at our present level of knowledge. Published results only allow one to estimate the diffusivity of substitutional elements in intrinsic silicon and to a limited extent estimate the effect of heavy doping on atomic motion. Our understanding of an induced field, non-constant activity coefficient, non-equilibrium defect concentrations, induced lattice distortion and short range solute enhanced diffusion is at best primitive.

In Table I the reported diffusion coefficients in near intrinsic silicon are presented. These results are evaluated from the standpoint of theory presented in a previous section to select a reasonable "accepted" diffusion coefficient. A probable error has been assigned whenever it appeared appropriate to indicate the diffusion coefficient one should expect in the temperature range from 1100 to 1300°C. In general, the diffusivities of the various substitutional elements should be within a factor of two of the "accepted" values listed in Table I. No accepted value is listed for germanium because the experimental results are questionable in the author's judgement.

The effect of heavy doping on the diffusivity of various elements [24], [43b], [45], [49b] is presented in Fig. (1) and (2) by plotting log (D_D/D_I) versus log (n_D/n_I). The effect of heavy doping on self-diffusion plotted in Fig. (1) does not include Ghostagora's [41] results. As suggested previously it is desirable to measure the electron concentration influence on diffusion as a function of temperature. The limited amount of published information does not allow this luxury. Since one does not expect the d_j coefficient in Eq. (32) to be a strong function of temperature, combining results obtained at different temperatures should not introduce a serious distortion. We have also not distinguished in these figures between the type of doping used. Only three of the impurities have been employed: phorphorus and arsenic for n-type and boron for p-type. Only one experiment has been reported where two different impurities were used to produce a large electron density (i.e. self-diffusion enhancement in which both phosphorus and arsenic were employed) and no observable difference was reported. The results presented in Figs. (1) and (2) will first be analyzed assuming no short range solute enhancement occurs. The effect of short range solute migration enhancement on self-diffusion will then be considered.

If one eliminated Ghostagore's [41] p-type results, one can approximately fit the self-diffusion coefficient dependence on electron concentration to a Laurent's series with $d_{+1} = 0.06$, $d_o = 0.78$ and $d_{-1} = 0.16$. The self-diffusion mainly occurs by a diffusion mechanism which has a diffusion charge state of zero. Fairfield and Masters [43b] suggested that the enhancement of self-diffusion in heavily doped n-type crystals is due to vacancies which behaved as acceptors. They further proposed that the self-diffusion is proportional to the total vacancy concentration and calculated that the vacancy acceptor level is 0.18eV above the mid-band position at 1100°C. This analysis of the electronic level assumed that the diffusivity of neutral and charge vacancies are equal. There has so far been no experimental

evidence presented to support this assumption. Ghostagore [18] and Kendall and DeVries [11] have suggested a divacancy mechanism for silicon self-diffusion whereas Seeger and Chik [2], [54] have proposed a modified interstitial mechanism. Fairfield and Masters [43b] suggested that the increased self-diffusion in heavily doped p-type (boron) crystals results from short range solute enhancement. This point will be discussed shortly.

Table I. Measured And "Accepted" diffusion coefficients in intrinsic silicon.

Element	Experimental			Accepted	
	D_o (cm^2/sec)	Q (eV)	Ref.	D_o (cm^2/sec)	Q (eV)
Boron	10.5	3.69	(1)	5	3.7 ± 0.1
	1.4	3.51	(29)		
	17.1	3.68	(30)		
	16.0	3.69	(31) (11)		
	2.02	3.52	(32)		
	106	4.25	(33) (11)		
	5.1	3.69	(34)		
	6×10^{-7}	1.67	(35)		
Aluminum	2800	3.9	(36)	7	3.5 ± 0.2
	8.0	3.49	(1)		
	4.8	3.36	(37)		
	0.5	3.0	(38)		
Gallium	3.6	3.51	(1)	2×10^2	4.1
	225	4.12	(39)		
Indium	16.5	3.90	(1)	16.5	3.9
Thallium	16.5	3.90	(1)	16.5	3.9
Carbon	1.9	3.1	(40)	1.9	3.1 ± 0.2
Silicon	1200	4.73	(41)	10^3	4.8 ± 0.3
	900	4.77	(42)		
	9000	5.13	(43)		
Germanium	6.26×10^5	5.28	(44)		
Tin	32	4.24	(45)	32	4.24
Phosphorus	1500	4.24	(46)	3	3.6 ± 0.1
	10.5	3.69	(1)		
	4.9	3.7	(47)		
	2.73	3.58	(48) (11)		
Arsenic	0.30	3.6	(1)	3	3.9 ± 0.3
	69	4.2	(49)		
	2.6	3.9	(50)		
	8.3×10^4	5.2	(51)		
Antimony	5.6	3.9	(1)	8	4.0 ± 0.1
	13	4.0	(52)		
Bismuth	1030	4.64	(1)	10^3	4.6 ± 0.2
	896	4.12	(53)		

The results presented in Fig. 1 on the effect of heavy doping on the diffusivity of tin in silicon are from Millea [24] and Yeh et al [45]. These two results are in reasonable agreement. One could fit these combined results to a Laurent's series using two coefficients, $d_o = 0.55$ and $d_{-1} = 0.45$. As for self-diffusion, these results suggest that tin migrates mainly by a mechanism with a diffusion charge state of zero.

The effect of heavy doping on the diffusivity of phosphorus [24], arsenic [49b] and antimony [24] have been combined and presented in Fig. (2). The enhanced diffusion of all three impurities in n-type material suggest $(D_D/D_I) = (n_D/n_I)$; the atomic motion of these substitutional donors proceeds via a -1 diffusion charge state. These results are consistent with the proposals that the migration of Group VA elements diffuse by negative vacancies [24], [49b], by a neutral phosphorus pair, [22], or a negatively charged divacancy. A modified interstitial mechanism has also been proposed to explain this behavior in n-type material [2], [54] The fact that the decrease of the diffusivity of phosphorus and antimony in p-type silicon does not follow (n_D/n_I) dependence indicates that there might be another diffusion charge state involved; perhaps a neutral vacancy. The diffusion coefficient of arsenic in heavily p-type specimens is not currently available.

The effect of heavy doping on the diffusivity of substitutional acceptors appears to be opposite to the behavior of donors [24]. The results of indium presented in Fig. (2b) suggest that this acceptor diffuses predominately with a diffusion charge state of +1. Seeger and Chik [2], [54] suggested a modified interstitial mechanism to explain this behavior where as Kendall and DeVries [11] favor a positive charged divacancy. A modified direct interchange has also been considered [24]. The fact that the diffusivity of indium in p-type material does not continuously decrease as (n_D/n_I) possibly suggests a component of diffusion charge state $j = 0$.

The effect of short range solute enhancement on the self-diffusion will now be considered. Fairfield and Masters [43b] proposed that the increase in the self-diffusion in heavily boron doped silicon might be accounted for by assuming that boron diffused by a vacancy mechanism and applied Eq. (33). They used Kurtz and Yee [29] boron diffusion coefficient measurements and assumed $\alpha = 1$. With this model they calculated a self-diffusion enhancement at a boron concentration of $2.2 \times 10^{20} cm^{-3}$ at $\sim 1100^{o}C$ of 1.6 relative to the intrinsic self-diffusion value. One might question the correctness of using the Kurtz and Yee boron diffusion coefficient since this value is appropriate for near intrinsic silicon and perhaps not heavily doped material. It appears reasonable to propose that the diffusion coefficient of boron follows a similar behavior with electron concentration as observed for indium [11], [34], [55]. With this assumption one would increase Kurtz and Yee's value by a factor of sixteen, but still using $\alpha = 1$, one would calculate a self-diffusion enhancement of a factor of 10; much larger than the experimentally reported enhancement (i.e. 1.7). Applying a similar analysis to Fairfield and Masters' heavily doped phosphorus specimen yields an enhancement of a factor of approximately twice the reported value. The fact that equal concentrations of phosphorus and arsenic yield approximately equal self-diffusion values would require a much higher α value for arsenic relative to phosphorus because the intrinsic diffusion coefficient of arsenic is approximately a factor of ten lower than the value for phosphorus. There are three possible explanations for the above discrepancy between analysis and experience: 1) incorrect boron and phosphorus diffusion coefficients 2) inaccurate experimental values, and, 3) boron and phosphorus does not diffuse by a mechanism which requires the displacement of host atoms. Additional measurements will hopefully resolve this discrepancy.

Various interesting observations have been reported on the diffusion of a high concentration of substitutional impurities and the interaction between diffusing impurities during multi-diffusion operations [56], [57]. Although it seems reasonable to assume these experimental observations can be explained in terms of induced fields [14b], lattice distortions [58], non-constant activity coefficients [59], short range solute interactions, and non-equilibrium effects along with the long range electron densities effects, the experimental observations appear too meager to warrant a detailed examination in this paper.

5. REFERENCES

[1] C. S. Fuller and J. A. Ditzenberger, J. Appl. Phys. 27, 544 (1956)

[2] A. Seeger and K. P. Chik, Phys. Stat. Sol. 29, 455 (1969)

[3a] H. Reiss, J. Chem. Phys. 21, 1209 (1953)

[3b] H. Reiss, C. S. Fuller and F. J. Morin, Bell System Tech. J. 35, 535 (1956)

[3c] F. A. Kröger, The Chemistry of Imperfect Crystals, (John Wiley & Sons, Inc., New York, 1964)

[4] R. L. Longini and R. F. Green, Phys. Rev. 120 992 (1952)

[5] M. W. Valenta and C. Ramasastry, Phys. Rev. 106, 73 (1957)

[6] M. W. Valenta, Ph.D. Dissertation, University of Illinois (1958)

[7] C. S. Fuller and J. C. Severiens, Phys. Rev. 96, 21 (1954)

[8] F. C. Frank and D. Turnbull, Phys. Rev. 104, 617 (1956)

[9] R. H. Hall and J. H. Racette, J. Appl. Phys. 35, 379 (1964)

[10] W. R. Wilcox, I. J. La Chappelle, J. Appl. Phys. 35, 240 (1964)

[11] D. L. Kendall and D. B. DeVries, Semiconductor Silicon (The Electrochemical Society, Inc. New York, 1969)

[12a] H. S. Carslaw and J. C. Jaeger, Conduction of Heat in Solids, (Clarendon Press, Oxford, England 1947)

[12b] J. Crank, Mathematics of Diffusion (Claredon Press, Oxford, England, 1956)

[12c] W. Jost, Diffusion in and Through Solids (Cambridge University Press, Cambridge,

[13] J. Bardeen and C. Herring, in Atom Movements p. 87 (Am. Soc. for Metals, Cleveland, Ohio 1951)

[14a] F. M. Smits, Proc. IRE 46, 1049 (1958)

[14b] S. M. Hu and S. Schmidt, J. Appl. Phys. 39 4272 (1968)

[14c] D. P. Kennedy, Proc IEEE, (Letters) 57, 1202 (1969)

[15] A. K. Jonscher, Principles of Semiconductor Device Operation (John Wiley and Sons, New York, 1960)

[16] W. Shockley, J. Appl. Phys. (Letters) 32, 1402 (1961)

[17] R. A. Swalin, J. Phys. Chem. Solids 18, 290 (1961)

[18] R. N. Ghostagore, Phys. Stat. Sol. 20, K89 (1967)

[19] J. R. Manning, Diffusion Kinetics for Atoms in Crystals, (D. Van Nostrand Co., Inc. Princeton, New Jersey, 1968)

[20] P. G. Shewmon, Diffusion in Solids (McGraw-Hill Book Company, New York, 1963)

[21] G. D. Watkins in F. L. Vook (ed.) Radiation Effects in Semiconductors (Plenum Press, New York, 1968)

[22a] G. D. Watkins and J. W. Corbett, Phys. Rev. 134, A1359 (1964)

[22b] E. L. Elkin and G. D. Watkins, Phys. Rev. 174, 881 (1968)

[23] W. Shockley and J. L. Moll, Phys. Rev. 119, 1480 (1960)

[24] M. F. Millea, J. Phys. Chem. Sol. 27, 315 (1966)

[25] R. E. Hoffman, D. Turnbull and E. W. Hart, Acta Met. 3, 417 (1955)

[26] H. Reiss, Phys. Rev. 113, 1445 (1959)

[27] D. Lazarus in F. Seitz and D. Turnbull (eds.) Solid State Phys. Vol. 10 (Academic Press, Inc. New York 1960)

[28] H. Reiss and C. S. Fuller in N. B. Hannay (ed) Semiconductors (Reinhold Publishers Corporation New York, 1959)

[29] A. D Kurtz and R. Yee, J. Appl. Phys. 31, 303 (1960)

[30] J. Yamaguchi, S. Horiuchi and K. Matsuwara, J. Phys. Soc. Japan 15, 1541 (1960)

[31] E. L. Williams, J. Electrochem. Soc. 108, 795 (1961)

[32] T. Kato and Y. Nishi, Japanese J. Appl. Phys. 3, 377 (1964)

[33] K. Nagano, S. Iwauchi and T. Tanaka, Japanese J. Appl. Phys. 7, 1361 (1968)

[34] M. Okamura, Japanese J. Appl. Phys. 8, 1440 (1969)

[35] G. L. Vick and K. M. Whittle, J. Electrochem. Soc. 116, 1142 (1969)

[36] B. Goldstein, Bull. Am. Phys. Soc., Ser. II, 1, 145 (1956)

[37] R. C. Miller and A. Savage, J. Appl. Phys. 27, 1430 (1956)

[38] Y. C. Kao, Electrochemical Technology 5, 90 (1967)

[39] A. D. Kurtz and C. L. Gravel, J. Appl. Phys. 29, 1456 (1958)

[40a] R. C. Newman and J. Wakefield, J. Phys. Chem. Solids 19, 230 (1961)

[40b] R. C. Newman and J. Wakefield, in J. B. Schroeder (ed.) Metallurgy of Semiconductor Materials, Vol. 15 (Interscience Publishers, New York, 1962)

[41] R. N. Ghostagore, Phys. Rev. Letters 16, 890 (1966)

[42] R. F. Peart, Phys. Stat. Sol. 15, K119 (1966)

[43a] J. M. Fairfield and B. J. Masters, Appl. Phys. Letters 8, 280 (1966)

[43b] J. M. Fairfield and B. J. Masters, J. Appl. Phys. 38, 3148 (1967)

[44] D. A. Petrov, Yu. M. Shaskov and I. P. Akimchenko, Voprosy Met. i. Fiz. Poluprovodikov (Moscow, Akad. Nauk SSSR) 130 Sbornik 1957 (Chem. Abst. 54, 1960 p. 17190e)

[45] T. H. Yeh, S. M. Hu and R. H. Kastl, J. Appl. Phys. 39, 4266 (1968)

[46] B. T. Howard, Fall Meeting Electrochemical Soc. (Oct. 1957)

[47] I. M. Mackintosh, J. Electrochem. Soc. 109, 392 (1962)

[48] S. Maekawa, J. Phys. Soc. Japan 17, 1592 (1962)

[49a] W. J. Armstrong, J. Electrochem. Soc. 109, 1065 (1962)

[49b] B. J. Masters and J. M. Fairfield, J. Appl. Phys. 40, 2690 (1969)

[50] P. S. Raju, N. K. Rao and E. V. K. Rao, Indian J. Pure Appl. Phys. 2, 353 (1964)

[51] Y. W. Hsueh, Electrochemical Technology 6, 361 (1968)

[52] J. J. Rohan, N. E. Pickering and J. Kennedy, J. Electrochem. Soc. 106, 705 (1959)

[53] D. Pomerrenig, Acta. Phys. Austriaca 20, 338 (1965)

[54] A. Seeger, Solid State Phys. (Comments) 2, 55 (1969)

[55] M. L. Barry and P. Olofsen, J. Electrochem. Soc. 116, 854 (1969)

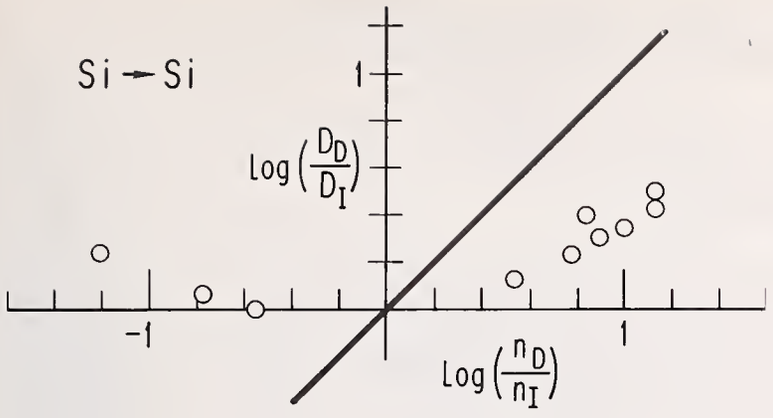
[56] E. Tannenbaum, Solid-State Electronics, 2, 123 (1961)

[57] A. F. W. Willoughby, J. Material Sci. 3, 89 (1968)

[58a] K. H. Nicholas, Solid-State Electronics 9, 35 (1966)

[58b] N. D. Thai, Solid-State Electronics 13, 165 (1970)

[59] P. E. Bakeman, Jr., and J. M. Borrego, J. Electrochem. Soc. 117, 688 (1970)



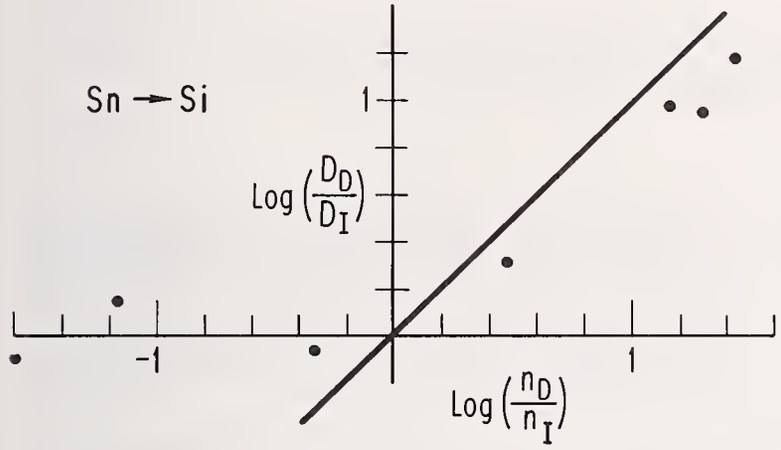
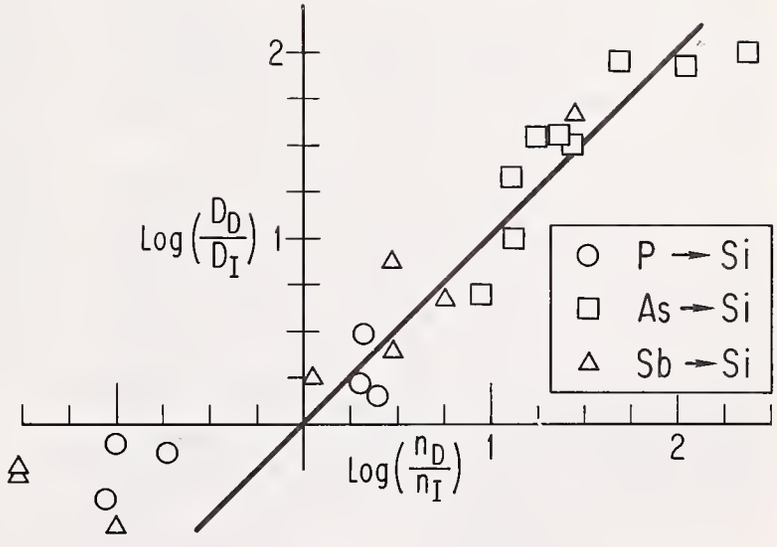


Fig. 1. The effect of heavy doping on the diffusion of silicon and tin in silicon.



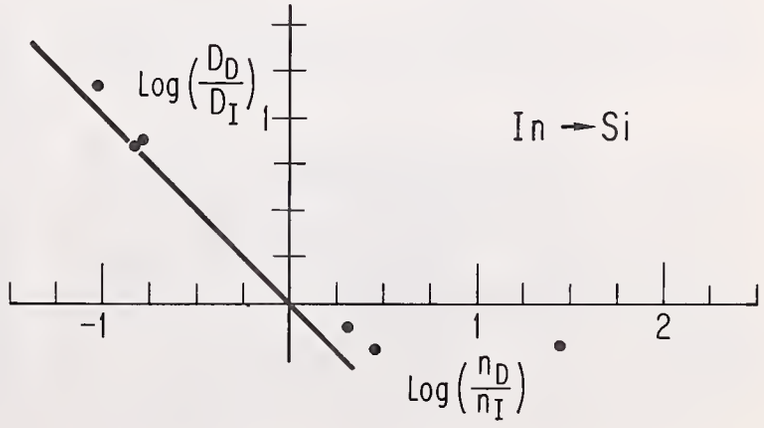


Fig. 2. The effect of heavy doping on the diffusion of phosphorus, arsenic, antimony, and indium in silicon.

DIFFUSION TECHNOLOGY FOR ADVANCED MICROELECTRONICS PROCESSING

by

W. Greig, K. Cunniff, H. Hymon, ond S. Muller
RCA Solid State Division
Somerville, N.J.

The so-called planar technology comprising solid-state diffusion, oxidation, and photoengraving techniques has matured at a rapid rate. This maturity has been marked by a series of modifications and improvements brought on by the demands for more precise controls required for fabrication of advanced-type devices. The diffusion technology of today reflects these changes. The development of phosphorus diffusion for forming emitters in n-p-n structures is reviewed chronologically to illustrate the process optimization that has occurred not only as a result of process evaluation but from a consideration of device requirements such as electrical properties, performance, and yield. Both the phorphorus pentoxide system and the phosphorus oxychloride system are discussed in detail.

Key Words

Solid-state diffusion; Phosphorus diffusion in silicon; Phosphorus pentoxide; Phosphorus oxychloride; Process evaluation; High-frequency solid-state devices; Integrated circuits; Oxide diffusion masking; Device performance; Diffusion-induced damage.

The ever-expanding state of the art in microelectronic device technology can be characterized by a continuing trend toward increased high-frequency performance, faster switching speeds, and increasing complexity and packing density of components. These requirements demand development of an equally challenging and highly sophisticated process technology which of necessity must complement the device technology.

The so-called planar technology comprising solid-state diffusion, oxidation, and photoengraving techniques has matured at a rapid rate. This maturity has been marked by a series of modifications and improvements brought on by the demands for more precise controls required for fabrication of advanced-type devices. The diffusion technology of today reflects the changes that have occurred. For example, phosphorus diffusion for forming emitters in n-p-n structures has undergone significant changes in terms of procedures and techniques. Therefore, a chronological review of phosphorus diffusion developments can serve as a review of what has transpired in establishing the present-day technology.

Although there are several techniques which can be employed to deposit phosphorus impurity atoms, and although the source of these impurity atoms comes in many forms (liquid, solid, or gas), this paper is restricted to those techniques which use solid phosphorus pentoxide and liquid phosphorus oxychloride as the source materials. The process development aspects can best be illustrated by a discussion of both systems in detail. In this manner, the advantages of a liquid system over a solid system with regard to reproducibility, control, and process flexibility can be made apparent. Process flexibility also emerges as an important factor in process optimization, particularly in relation to device yield and performance. This latter aspect of optimization, based upon not only process evaluation but also product requirements, is emphasized.

Phosphorus Pentoxide (P_2O_5)

Phosphorus pentoxide was perhaps the earliest and most widely used source material. Fig.1 shows a P_2O_5 deposition system consisting of a single quartz tube which passes through two furnaces that control the temperatures at which the P_2O_5 and the silicon wafers are maintained. The temperature of the P_2O_5 source determines its vapor pressure and, in turn, the number of impurity molecules available. The silicon temperature controls the diffusion rate, i.e., the rate at which impurity atoms diffuse into the silicon. A carrier gas is employed to transport the volatile impurity molecules from the source region downstream to the silicon.

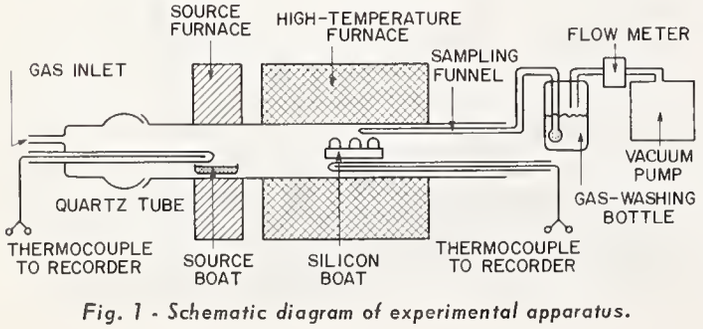


Fig. 1 - Schematic diagram of experimental apparatus.

In theory, control of the number of atoms available and the temperature and time of diffusion is sufficient for accurate control of the distribution of the impurity atoms in the silicon. In practice, it is not quite that simple. Fig. 1 also shows a sampling tube inserted in the quartz diffusion tube and positioned in close proximity to the silicon wafers. This tube can be used to sample the gas ambient and thus permit an analysis of the phosphorus content in the gas. This analysis, together with standard diffusion measurements such as sheet resistance and junction pentrations, can be used to evaluate the process performance. All sheet-resistance measurements given in this paper are made by use of a four-point probe. Junction measurements are from angle-lapped samples. Surface concentrations are calculated by use of the curves of J.C. Irvin and assume a complementary error function or a Gaussian-type impurity distribution.

Three system variables can be identified as follows: (a) source temperature, (b) source age or stability, and (c) carrier-gas composition and flow rate.

Source Temperature

Fig. 2 shows surface concentration as a function of source temperature with the silicon temperature held constant at 1100°C. The carrier gas, carrier-gas flow rate, and deposition time are also held constant, as indicated. For a range of source temperatures from approximately 150°C to 325°C, the surface concentration varies over two orders of magnitude, with saturation conditions being reached at about 250°C. From a control point of view, therefore, a source temperature in excess of 250°C is preferred.

Fig. 3 shows similar data for a silicon temperature of 1200°C. Of particular interest is the fact that the saturation condition (i.e., the point at which no further increase in surface concentration occurs with increasing source temperature) occurs at approximately 175°C. Under certain conditions, therefore, a change in silicon temperature results in a significant change in surface concentration with no change in P_2O_5 source temperature. This change is much in excess of what might be expected from a consideration of increased solid solubility with increased temperature.

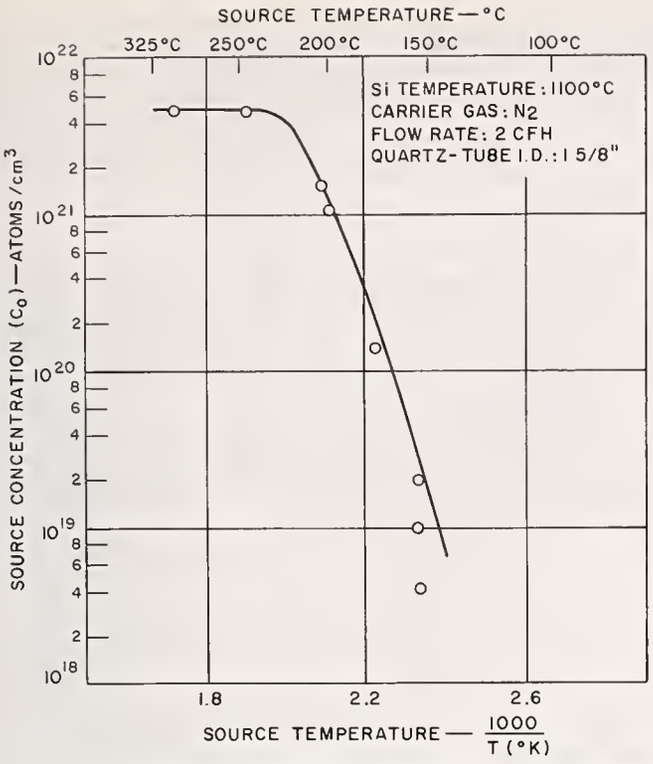


Fig. 2 - Surface concentration as a function of source temperature for a silicon temperature of 1100°C.

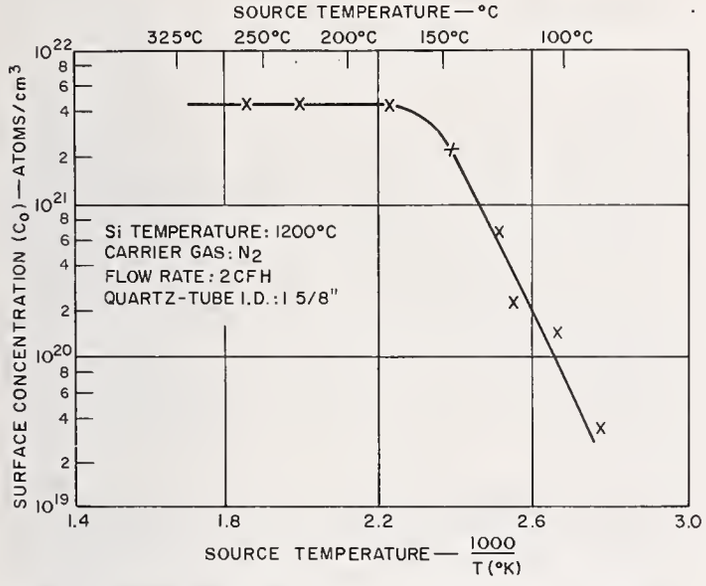


Fig. 3 - Surface concentration as a function of source temperature for a silicon temperature of 1200°C.

Source Age or Source Stability

Anhydrous P_2O_5 absorbs moisture from the air upon standing and can become useless for all practical purposes. Consequently, careful handling procedures must be employed to prevent moisture contamination.

Fig. 4 shows surface concentration as a function of source age, i.e., the time that the P_2O_5 is held at the temperature indicated. Curve A indicates a series of depositions in which silicon wafers were placed in the deposition system and removed at 30-minute intervals. Curve B represents wafers inserted at zero time, with one sample removed every 30 minutes. The slopes of the curves show how difficult reproducibility of results becomes after very short times at temperature.

Fig. 5 shows phosphorus content as a function of source age for three values of source temperature. Although previous data indicated a preference for higher source temperatures, at these conditions the aging effect is much more of a problem and must be taken into account from a process-control point of view.

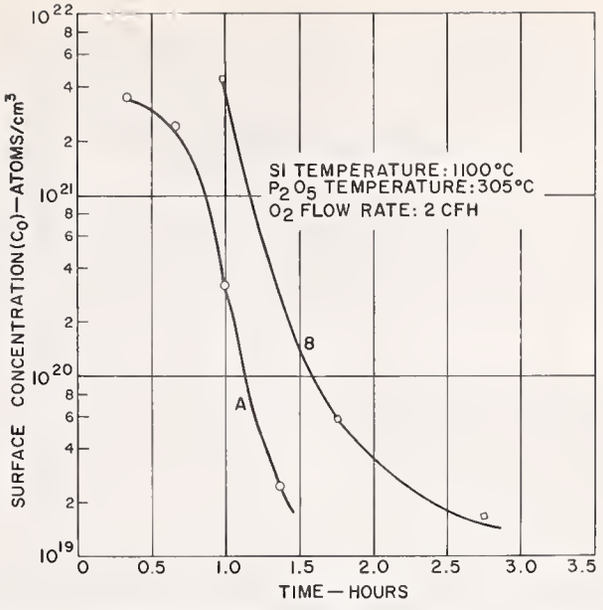


Fig. 4 - Surface concentration as a function of source age.

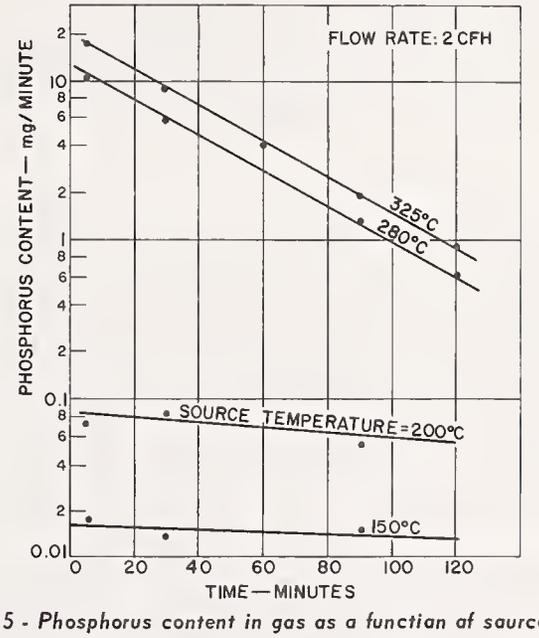


Fig. 5 - Phosphorus content in gas as a function of source age.

Carrier-Gas Composition and Flow Rate

The carrier gas also affects the diffusion parameters, as illustrated in Fig. 6, which shows surface concentration as a function of source temperature for carrier gases of nitrogen and oxygen. The use of oxygen suppresses the surface concentration, but does not appear to affect the saturation temperature drastically. Also, the suppression is much more pronounced at the lower source temperature.

Fig. 7 illustrates the effect of flow rate. Again, chemical analysis of the phosphorus content in the gas indicates that the source aging condition is also affected by the flow rate.

Need for Improved Process

In summary, the P_2O_5 system presents many problems, although it was satisfactory for earlier, less critical needs. Mesa devices, for example, represented the state of the art. Base widths on these devices were in the order of tenths of mils, and the variability of the emitter diffusion processing, although not totally acceptable from a yield point of view, was at least tenable for the device requirements. However, the advent of planar technology and the advances offered by this technology in terms of device performance put further demands upon the diffusion processes. The very heart of the planar technology, i.e., diffusion masking by use of silicon dioxide, was an added consideration requiring additional investigation.

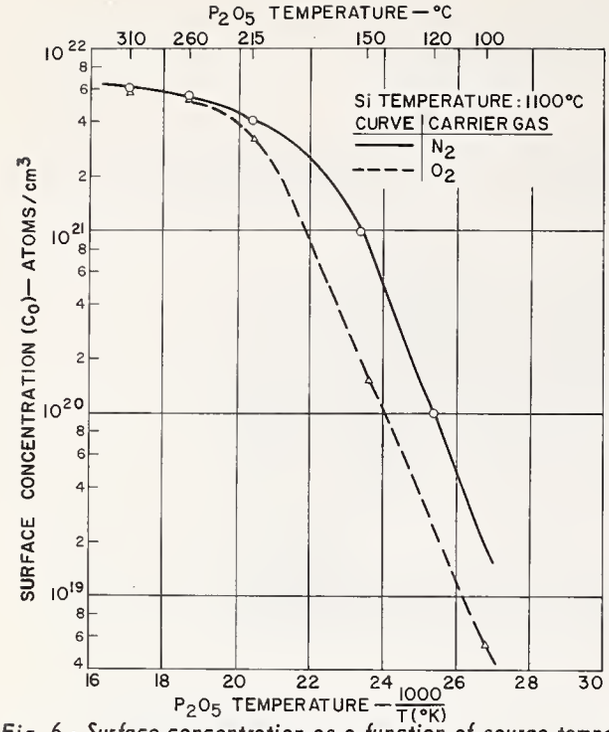


Fig. 6 - Surface concentration as a function of source temperature for carrier gases of nitrogen and oxygen.

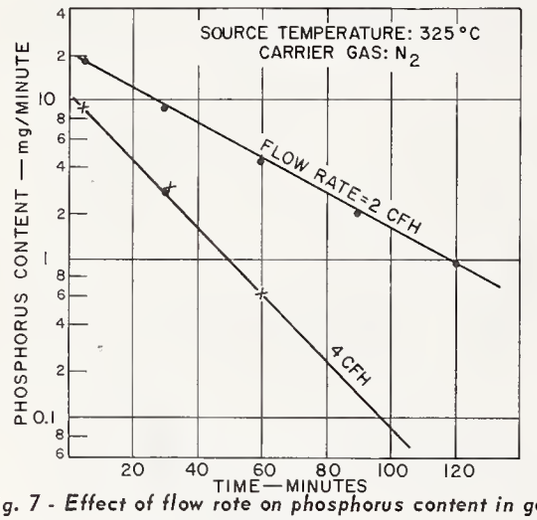


Fig. 7 - Effect of flow rate on phosphorus content in gas.

Process development on the P_2O_5 system proceeded, then, in a direction aimed at meeting the needs of this new technology. With the planar technology, the SiO_2 must function as a diffusion mask; therefore, it is important that the amount of SiO_2 required to mask against a particular diffusion be known. Fig. 8 shows a curve of failure time as a function of oxide thickness for three values of source temperature. The source temperature strongly affects the SiO_2 thickness required, i.e., higher source temperatures require more oxide. Thus, although source temperatures above the saturation condition produced no change in the diffusion, there were significant effects on the oxide. In effect, planar technology dictated the process to be followed. It also provided the impetus for further diffusion-process refinements and investigation of other materials and techniques.

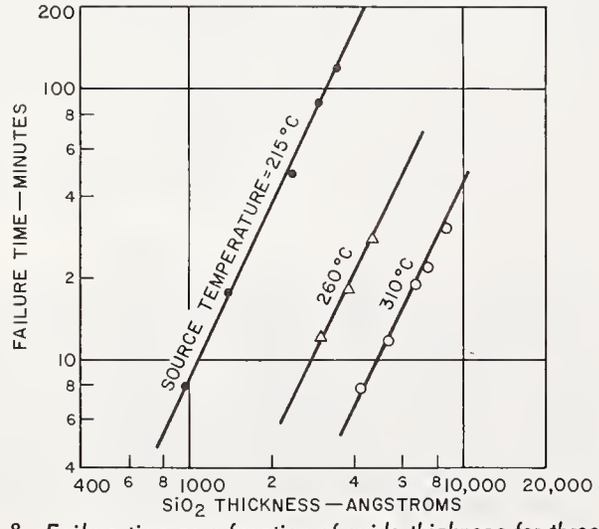


Fig. 8 - Failure time as a function of oxide thickness for three values of source temperature.

Phosphorus Oxychloride ($POCl_3$)

$POCl_3$, a liquid source material, seemed to offer immediate advantages over the P_2O_5 system. Fig. 9 shows a schematic diagram of an early $POCl_3$ system. With a single carrier gas, a bypass line was provided to add flexibility and to permit a portion of the total carrier-gas flow to pass over the source. In the earliest source container or bubbler, the gas passed over the liquid source.

Based upon the results obtained from the work with P_2O_5, a nitrogen carrier gas was first tried on the $POCl_3$ system with erratic results. A change to oxygen as the carrier produced the desired results described below.

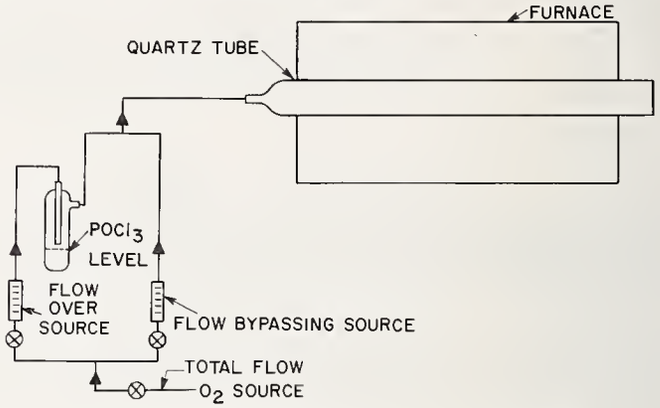


Fig. 9 - Schematic diagram of early $POCl_3$ system.

Fig. 10 shows surface concentration as a function of oxygen flow over the source. The data indicate saturation

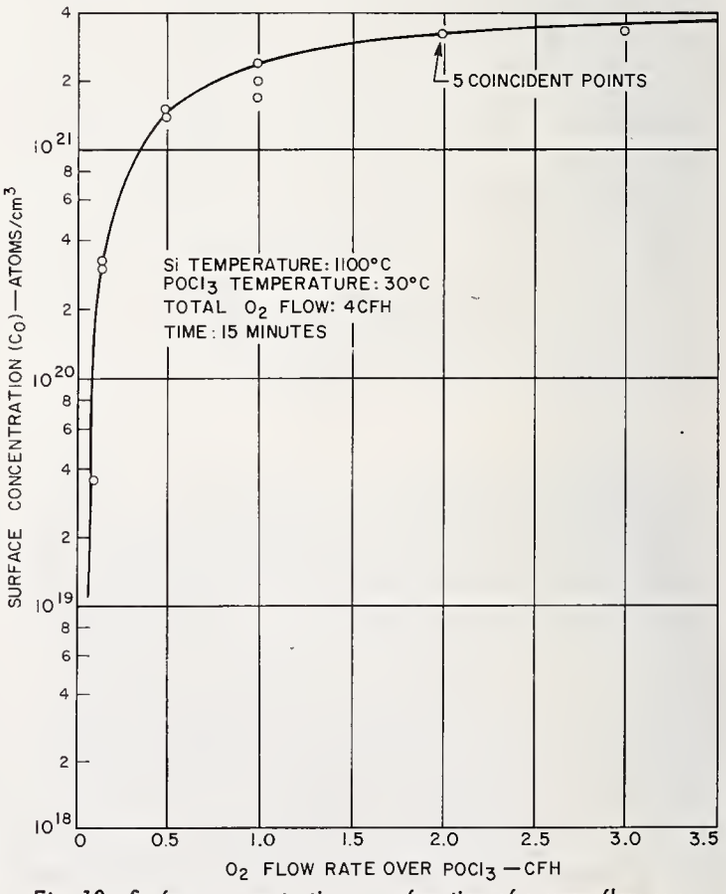


Fig. 10 - Surface concentration as a function of oxygen flow over the source.

conditions at a flow rate in excess of 2 cubic feet per hour over the source and some control of surface concentration at lower flows. Fig. 11 shows similar data, and also includes data for silicon temperatures of 1050°C and 1100°C and source temperatures of 30°C and 0°C.

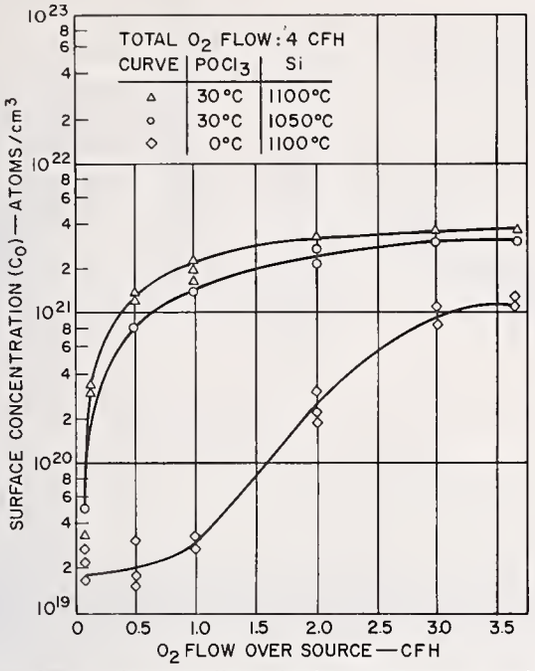


Fig. 11 - Curves showing effect of silicon temperoture ond source temperature on surfoce concentrotion os o function of oxygen flow over the source.

Fig. 12 shows oxide-mask failure time as a function of oxide thickness for two conditions of flow over the source. The results of this initial phase of development of the $POCl_3$ system indicated that it was superior to P_2O_5 as a process from the point of view of reproducibility, control, ease of operation, and maintainability. Additional testing and production runs on factory product also indicated yield improvement.

Analysis of both process and yield data led to further process development work. For example, the data relating oxide-mask failure time showed that, for equivalent phosphorus doping levels (in the silicon), the same oxide thickness afforded improved masking capability when the source of the phosphorus was $POCl_3$ rather than P_2O_5. Further investigation indicated that this improvement was directly related to the amount or thickness of phosphorus glass formed on the oxide during deposition.

The data from production runs on planar devices showed the yield improvement to be the result of a significant decrease in collector-to-base voltage (V_{CBO}) shrinkage. Device analysis further indicated that the V_{CBO} shrinkage was caused by localized phosphorus diffusion occuring along the base-collector junction periphery. These spurious diffusions were found to be the result of (1) pinholes or partial pinholes introduced into the oxide during the photoengraving operation for opening the emitter region, and (2) localized "attack" of the oxide by the phosphorus glass and consequent failure of the oxide as a mask at these sites.

Catefully controlled tests involving a common photoengraving step which eliminated the first cause above definitely showed that the $POCl_3$ system produced less oxide "attack" and therefore resulted in improved yield on planar-processed devices. The data clearly indicated that further optimization of the process should be directed toward minimization of the of the phosphorus glass.

The $POCl_3$ system offered several possible approaches and variations. One immediate innovation involved gas sequencing. For example, the system readily allowed for not only partial or total flow of carrier gas over the source, but also total bypass of the source if desired. This approach permitted a multi-cycle timing sequence during which the source would be essentially "on" or "off". The insertion of cold silicon wafers into a furnace with a flow of source vapors present was a problem with P_2O_5, but could be readily controlled with $POCl_3$. The $POCl_3$ system permitted wafers to be placed into the furnace with only carrier gas flowing; after some predetermined time, the appropriate valves were either opened or closed to pass the carrier gas over the source and into the tube. The sequence could then be reversed to shut off the source, and the wafers could be removed with only carrier gas flowing. Thus, a three-cycle process was evolved which consisted of a warm-up cycle during which the carrier gas bypassed the source, a deposition cycle in which part or all of the flow was passed over the source, and finally a purge cycle in which the gas again bypassed the source.

Process Control

The advent of uhf transistors and the increasing production of integrated circuits clearly indicated the need for development of a process which would provide for significant improvements in reproducibility and control. The control aspect is of prime importance in fabrication of high-frequency devices, which require accurate control of the depth of penetration of the emitter junction with respect to the base-collector junction to yield a base-width separation on the order of 1 micron and less. Integrated circuits at present require a lesser degree of control, but demand a more defect-free process because they are basically large-area devices. Process-induced defects such as the localized diffusions mentioned previously adversely affect the yield of IC's more than that of smaller-area discrete devices.

In view of previous data relating the amount of phosphorus glass deposited, the amount of oxide needed for masking, and device yield, an improved process was pursued to optimize the three-cycle process relative to these factors. Fig. 13 shows the results of an investigation to determine the effects of the time the source is actually "on" with respect to the total time. Fig. 14 presents similar data for different times.

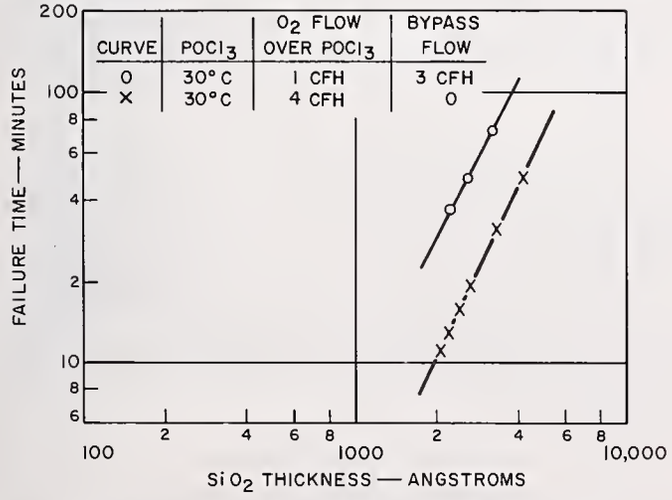


Fig. 12 - Failure time as a function of oxide thickness for two conditions of flow over the source.

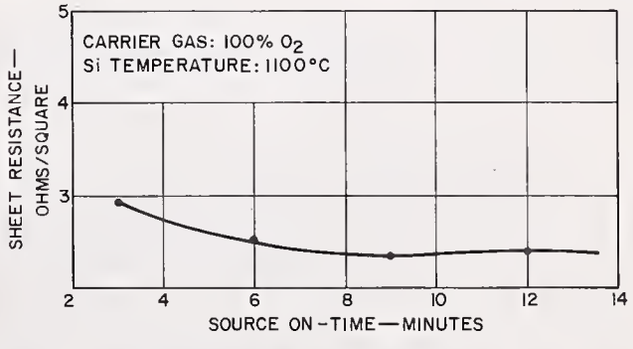


Fig. 13 - Sheet resistance os o function of source "on" time.

Both sets of curves indicate that a continuous flow of dopant is not necessary to provide saturated or infinite-source conditions. However, there was some indication that short on-times produce slightly higher sheet-resistance values, particularly if the total time is also short.

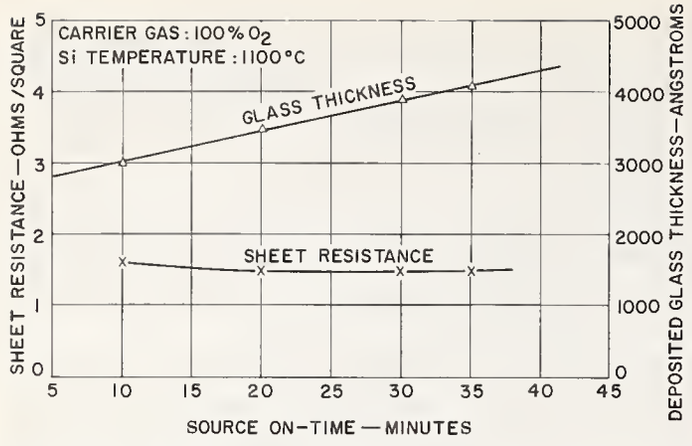


Fig. 14 - Sheet resistance and deposited glass thickness as functions of source "on" time.

Fig. 14 also shows the thickness of phosphorus glass deposited as a function of source on-time; as expected, there is a linear relationship, i.e., the longer the on-time, the more glass deposited. This evaluation led to a procedure in which the shortest on-time for saturation conditions (for a given total time) was determined for each deposition process.

Concurrent with these evaluations, the need for equipment refinement was evident and was implemented. An automatic gas-sequencing system was designed and installed which utilized electric timers and solenoid valves and permitted more accurate cycle changes. In addition, the pass-over source generator proved to be a problem relative to reproducibility over a period of time, and was replaced by a flow-through bubbler in which the carrier gas actually flowed through the liquid $POCl_3$. This design completely eliminated the reproducibility problem.

Requirements of High-Frequency Devices

The continuing development of high-frequency devices generated further improvements. Measurement of the beta of a transistor as a process control, permitted irregularity in the level and uniformity across a wafer to be noted. A very strong dependence on lot size (i.e., the number of wafers processed at one time) was evident. Analysis of various devices indicated that devices processed through an emitter cycle in which the total time was in the order of 15 minutes were most susceptible to this variation. The effect of the warm-up cycle, which was previously thought to be relatively unimportant, quickly became significant. Subsequent tests clearly established that the warm-up cycle should cover the time required for a specified lot size to reach the desired operating temperature. Because this warm-up cycle was equipment-oriented, the actual time required for return to temperature was determined. A fifteen-minute time cycle was found to be sufficient to cover most equipments, and was therefore standardized.

While improvements in phosphorus diffusion were being pursued, similar investigations were being conducted with boron diffusions. The significance and consequence of prior processing were continually recognized as factors to be considered in evaluating the phosphorus diffusion process.

High-frequency performance required reduced geometry. This feature, in turn, required the use of a dip-opened emitter, a non-re-oxidized emitter, or a washed-out emitter. In this approach, the opening in the oxide for the emitter diffusion was also used for the contact opening. This structure suffered from extremely low yields, and analysis of the shrinkage indicated that a significant percentage of the failures were caused by emitter-base shorts. Further investigations disclosed that the phosphorus glass removal step following the emitter deposition was extremely critical and directly affected the yield. Continuing efforts to optimize this process failed to produce the desired improvements, and attention was directed to the phosphorus deposition system. The investigation that followed resulted in the establishment of a process based entirely on the device in question.

On the basis of previous data from process evaluations and and device results, optimization was aimed at reducing the amount of phosphorus glass formed during the deposition cycle. It was obvious that minimum "on" time was not sufficient and alternate procedures would be required. The effect of the use of 100-per-cent oxygen as opposed to lesser amount in a two-gas system was therefore investigated.

Fig. 15 shows a schematic of the $POCl_3$ system. Nitrogen is passed through the $POCl_3$, and oxygen is introduced into

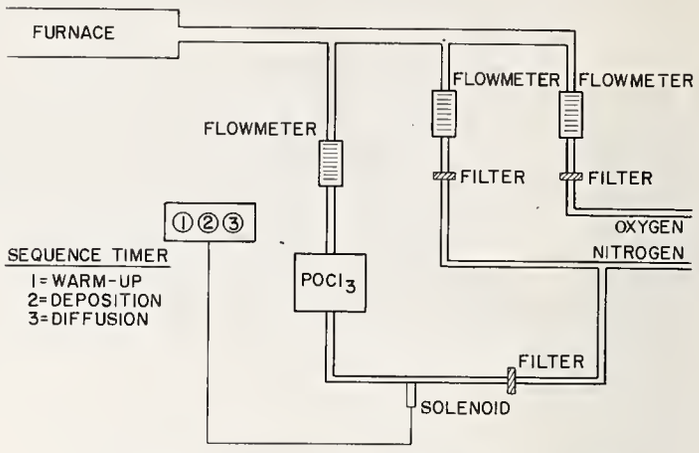


Fig. 15 - Schematic diagram of present $POCl_3$ system.

the system in the carrier-gas leg. Fig. 16 shows phosphorus glass thickness as a function of the percentage of oxygen in the carrier gas. The relationship between oxygen content and

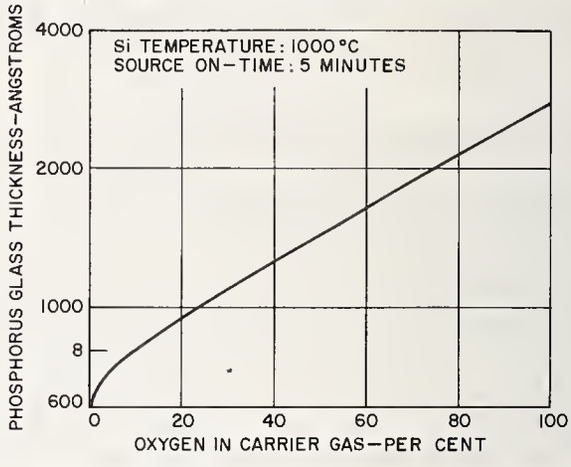


Fig. 16 - Phospharus glass thickness as a functian af axygen cantent in carrier gas.

phosphorus glass thickness is evident. Fig. 17 shows the effect of the percentage of oxygen in the carrier gas on the amount of original oxide removed by a phosphorus glass removal operation using the Pliskin etch. These curves show surprisingly little effect from 1 to 99 per cent oxygen, in view of the amount of glass deposited.

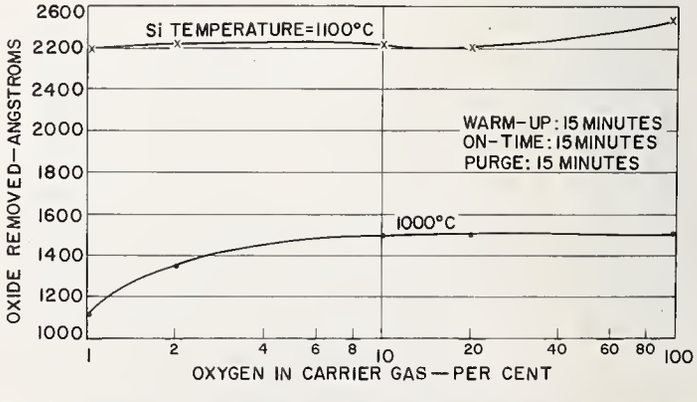


Fig. 17 - Amaunt af axide remaved as a functian af axygen in carrier gas.

Fig. 18 shows the relationship between oxygen content and sheet resistance. Of particular interest is the decrease in sheet resistance observed between oxygen concentrations of 1 and 2 per cent at a silicon temperature of 1000°C. A condition of saturation is reached beyond 2 per cent during which no change is observed until the concentration increases above approximately 20 per cent.

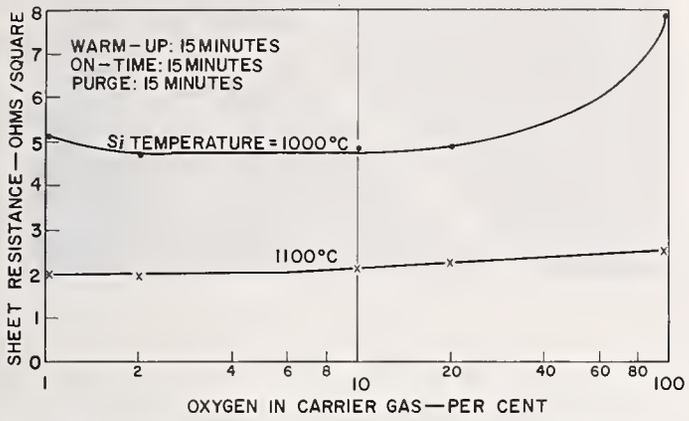


Fig. 18 - Sheet resistonce os o function of oxygen content in carrier gos.

Fig. 19 shows the effect of phosphorus on-time on the sheet resistance. The saturation time is readily apparent. Fig. 20 shows the amount of oxide removed by a Pliskin etch as a function of phosphorus on-time for a 1-per-cent oxygen content in the carrier gas.

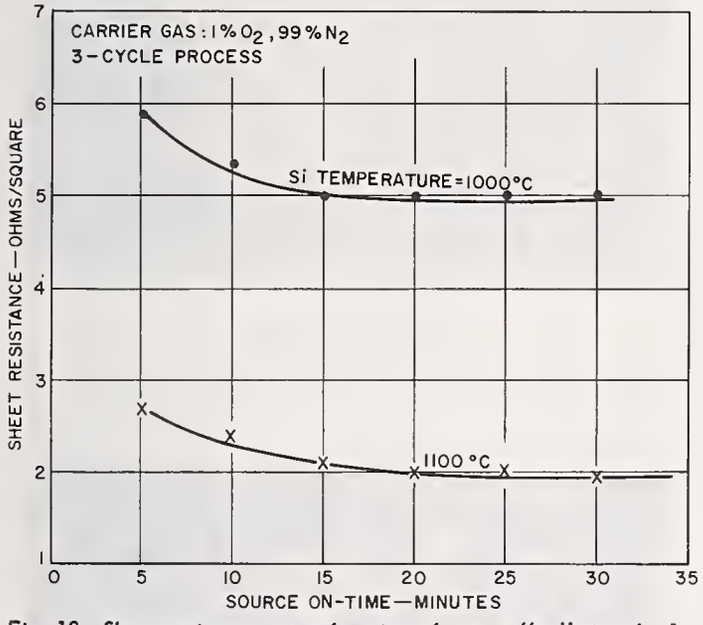


Fig. 19 - Sheet resistance os o function of source "on" time for 1-per-cent oxygen content in corrier gos.

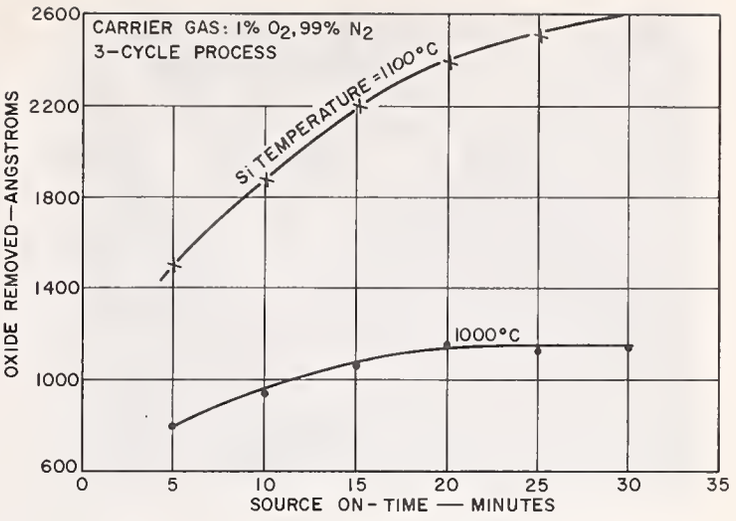


Fig. 20 - Amount of oxide removed os o function of source "on" time for 1-per-cent oxygen content in corrier gas.

Conclusion

As a result of the data obtained, a process was evolved which was optimized in terms of device parameters and device yields. Subsequent process optimization on other types of devices revealed strong dependence of several device parameters on the emitter deposition process. Although this dependence was expected in some instances (e.g., increased current gain with increased deposition times), some unexpected results included an observed change in beta with the phosphorus on-time at a constant total time at temperature, and a dependence of switching speed on the phosphorus on-time. Although no model is offered to explain these observed results, it is strongly suggested that major factors may be the gettering properties of the phosphorus glass and the diffusion-induced damage resulting from high concentrations of phosatoms into the lattice. For example, Fig. 21 shows photomicrographs (using phase contrast) of silicon wafers subjected to Sirtl etch following phosphorus depositions in which the on-times were varied. The relationship between induced damage and on-time is quite evident.

Optimization of the process in terms of the device indicated that minimum diffusion-induced damage (by reduction of source on-time) is desirable for switching transistors or integrated circuits. Reduction in switching speeds of 50 to 100 per cent have been observed on some types as a result of such damage. On the other hand, amplifier types having high beta and frequency response require longer source on-times and, therefore, more damage results. As mentioned previously, a low-oxygen-content carrier gas results in reduced "attack" of the oxide and consequently improves yield on dip-opened emitter types, integrated circuits, and other large-area devices.

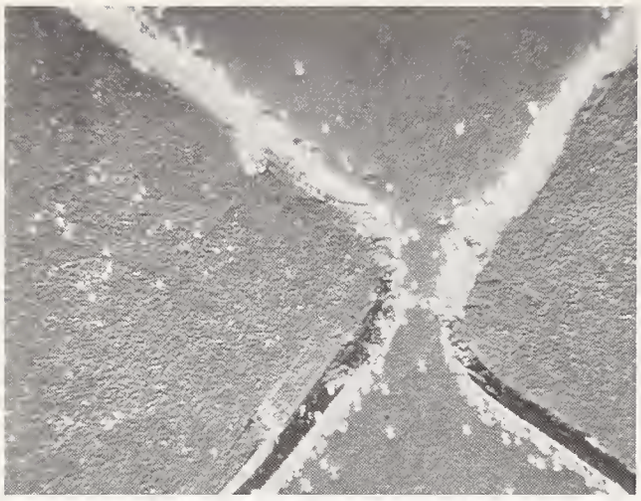
In summary, the diffusion development which resulted in the technology to today evolved from an extensive investigation of not only the processes and procedures but also the devices, their properties, and their structures. As a result, diffusion processes can be optimized to provide both increased yield and improved device performance.

Acknowledgements:

The authors wish to acknowledge the contributions, to the data reported herein, of the following, J.C. Sarace, L. Pomonte and S. Policastro.

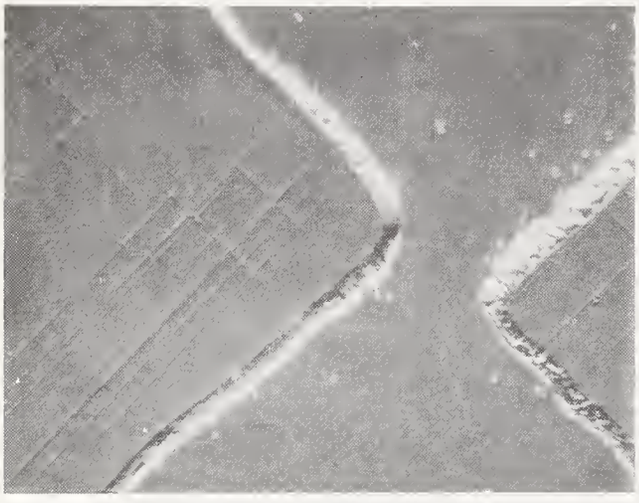
Bibliography

Integrated Silicon Device Technology Vol. 4 Reasearch Triangle Institute

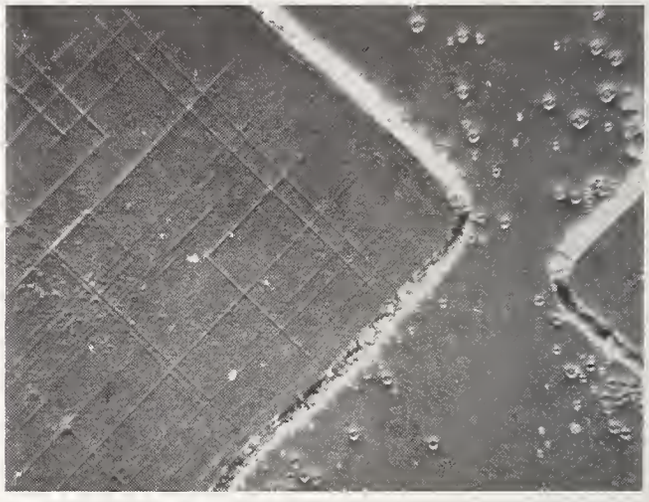


(a) CYCLE: 15-5-25
TEMP: 1000°C

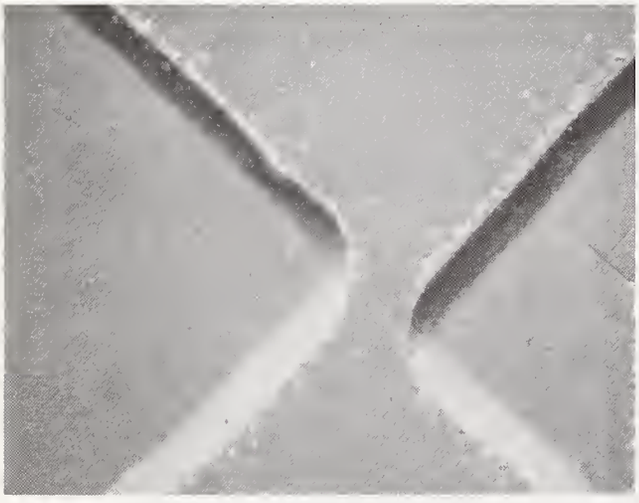
(b) CYCLE: 15-10-20
TEMP: 1000°C



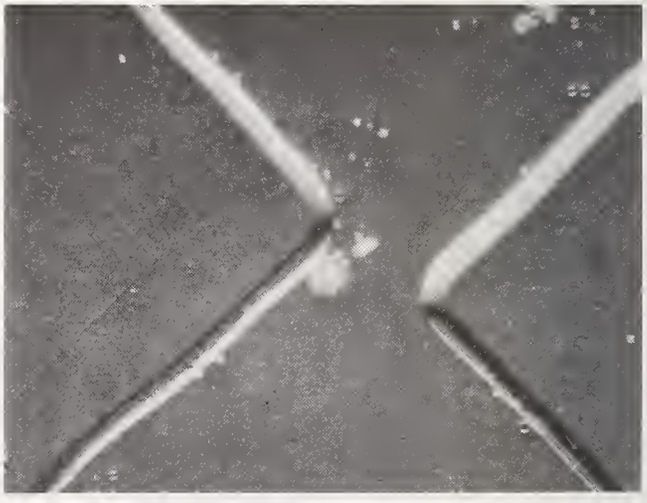
(c) CYCLE: 15-15-15
TEMP: 1000°C



(d) CYCLE: 15-20-10
TEMP: 1000°C



(e) CYCLE: 15-25-5
TEMP: 1000°C



(f) CYCLE: 15-30-0
TEMP: 1000°C

Fig. 21 - Silicon wafers subjected to Sirtl etch following phosphorus depositions for "on" times ranging from 5 to 30 minutes.

Diffusion from Doped-Oxide Sources

M. L. Barry

Research and Development Laboratory
Fairchild Camera and Instrument Corporation
Palo Alto, California 94304

Diffusion into silicon from low-temperature deposited doped oxides appears to offer several advantages over other diffusion technologies, but utilization of the technique has not been rapid. This paper reviews in some detail various processes used for deposition of doped oxides and their diffusion characteristics. The advantages claimed for doped oxide sources are compared to those of standard gas-phase predeposition–diffusion techniques.

Key words: Diffusion sources, doped oxides, semiconductor technology, solid-state diffusion.

1. Introduction

The use of doped oxides as diffusion sources has been reported for several years [1-4],* but it has only been recently that the technique has begun to be generally used in silicon device processing. Diffusion from deposited doped oxides offers in theory many advantages over the gas-phase predeposition–diffusion cycles commonly used. The primary advantages include independent control over concentrations and flexibility in processing; other advantages are improvement in uniformity and in-process control, reduction of diffusion-related defects, the ability to do simultaneous complementary diffusions, and an adaptability to continuous processing. The fact that there has been a five-year lag in exploiting this type of diffusion implies that these advantages either have not been realized in practice or have not been of major importance. It is the purpose of this paper to review in some detail the technology of diffusion into silicon from doped oxides, comparing it where possible with standard gas-phase predeposition–diffusion technology, and to reevaluate the advantages listed above.

The following sections will first describe the general behavior of diffusions from doped oxides and the various techniques of depositing these films. Next, a comparison between the control of dopant profiles from this technique and the usual gas-phase doping techniques will be made. Because of the proprietary restrictions necessary to most organizations in the semiconductor electronics industry, it is obviously not possible to gather and present detailed analyses of these various processes in terms of yields, internal control specifications, or device characterizations; however, some of the techniques and criteria which can be used to make valid comparisons will be described.

2. The Diffusion Model

In all techniques using this type of diffusion source, the oxide with a controlled concentration of dopant is deposited or grown directly on the substrate at temperatures below those where any appreciable diffusion takes place. The substrate is then heated to the diffusion temperature, where the diffusion takes place for a controlled length of time. The oxide may then be removed, or it may be left on as a passivating dielectric or as a barrier to subsequent diffusions. In the latter case, of course, it continues to act as a diffusion source during all subsequent high-temperature treatments.

Solutions to the diffusion equations for this system, with appropriate approximations,† have been developed for both a non-blocking boundary condition at the outer surface of the oxide [5-7] and a blocking boundary condition [8]. Both cases reduce to the same solution for the most commonly used case, i.e., where: (a) the thickness of the oxide is large compared to the diffusion length of the dopant in the oxide; (b) the diffusion takes place in an inert ambient, or the thickness of the oxide is large compared to the diffusion length in the oxide of any oxidizing species from the ambient; (c) the dopant is initially uniformly distributed in the oxide.

---

* Figures in brackets identify the literature references at the end of this paper.

† The most serious approximation is that the diffusion coefficients are independent of concentration, which implies that both field-aided diffusion and ion-vacancy interactions can be ignored.

This solution is that for diffusion from a semi-infinite source into a semi-infinite sink, or

$$C_2(x,t) = C_s \operatorname{erfc} \frac{x}{2\sqrt{D_2 t}} \tag{1}$$

where $C_2(x,t)$ is the concentration in the silicon at a distance x from the oxide-silicon interface, t is the diffusion time, D_2 is the diffusivity of the dopant in silicon, and C_s is the surface concentration. The surface concentration is independent of time and is related to C_o, the initial concentration of dopant in the oxide, by

$$C_s = \frac{C_o \sqrt{D_1/D_2}}{1 + \frac{\sqrt{D_1/D_2}}{m}} \tag{2}$$

where D_1 is the diffusivity of the dopant in the oxide, and m is the segregation coefficient, the ratio of the concentration of the dopant in the silicon to the concentration in the oxide at the oxide-silicon interface. Thus, for this case the surface concentration is independent of diffusion time and is directly proportional to the initial concentration of the dopant in the oxide, and the interface between the silicon and oxide is stationary. Furthermore, because of the values of D_1, D_2, and m for most commonly-used dopants, the surface concentration depends to first order on the square root of the ratio of the diffusion coefficients and only to second order on the segregation coefficient. It can be shown that the junction depth and sheet conductivity both depend linearly on the square root of the diffusion time, with slopes which are functions of C_s, D_2, and the bulk concentration, C_B.

If condition (a) above does not hold (that is, if there is depletion of the doped oxide source), the diffusion profile approaches a Gaussian curve as shown by Owen and Schmidt [8]. In this case the distinction between a blocking and a non-blocking boundary condition at the external surface of the doped oxide becomes important, and the rate of mass transfer of the dopant from the surface into the gas phase must be considered.

In certain cases, a phase change can occur in the doped oxide at the diffusion temperature. Such a change occurs with heavily phosphorus-doped films in nonoxidizing ambients, and perhaps also with arsenic doped films. In the case of phosphorus, the film has the appearance of actually turning liquid, although it does not appear to crystallize upon solidification. Liquefaction of course has the effect of greatly increasing D_1 and, hence, C_s as shown in eq (2). This phase change seems to correspond quite well to the phase diagram for SiO_2-P_2O_5 as proposed by Tien and Hummel [9]. The phase change does not appear to occur when the diffusion is performed in an oxidizing ambient.

The case where oxidation at the doped oxide-silicon interface occurs during the diffusion (that is, when condition (b) does not hold) is much more complicated. Qualitatively, the diffusion into the silicon is retarded significantly because of the barrier oxide growing concurrently at the interface. The resultant of the two competing processes, oxidation and diffusion, depends very strongly on the thickness and initial concentration of dopant of the doped oxide film, the oxidation ambient, the temperature, and the initial distribution of oxidizing species in the deposited film. Since diffusion in silicon and silicon oxide diminishes much more rapidly with decreasing temperature than does oxidation [10], the effects become especially pronounced at temperatures of 1000°C and below. The problem has been treated by Mecs [11], but, because of the approximations involved, his solution does not appear to predict the proper dependence on the initial oxide thickness [12]. Since several interesting applications of doped oxide diffusions utilize oxidizing ambients, it would appear that a quantitative model for this problem should be derived.

3. Deposition Techniques

Techniques of depositing doped oxides may be roughly classified as those involving chemical vapor deposition and those using deposition or growth from a liquid phase. The sources normally are silicon oxides which have oxides of the dopants incorporated within them. The major criteria for choice among the various techniques are that negligible diffusion into the silicon takes place during the deposition, that the concentration of dopant be uniform and reproducible, and that the oxide be relatively stable and clean under the diffusion conditions. Other criteria are the ability to deposit or grow thick oxide films, to control the concentration of dopant over a wide range, and of course, the speed and cost of the deposition process.

3.1. Chemical Vapor Deposition

Much of the earlier work in this area was directed toward diffusing elements such as zinc, tin, tellurium, and selenium into gallium arsenide or other III-V compound semiconductors [1,2,13]. Since the present paper is restricted to silicon technology, only those techniques and dopants which are applied to silicon will be discussed in detail.

a. Pyrolysis

One of the earliest reported techniques was that of Scott and Olmstead [4], who deposited boron- and phosphorus-doped silica films by the pyrolytic decomposition of alkoxysilanes (usually tetraethyl orthosilicate) and alkyl borates or alkyl phosphates, respectively. This technique has also been reported by Cuccia *et al.* [14] for deposition of arsenic-doped silica using arsenic trichloride. The deposition is usually made at temperatures in the range 700-800°C, and the dopant concentration in the silica is controlled by the relative partial pressures of the uncracked dopant compounds and the alkoxysilane in the deposition reactor. This deposition is usually made in a quartz tube, using either resistance or RF-induction heating.

The major advantages of the pyrolysis deposition are the thickness of the oxide which can be deposited and the range of concentrations which can be achieved. Scott and Olmstead [4] reported surface concentrations of boron from 2.5×10^{17} to 1.5×10^{20} atoms/cm^3; of phosphorus from 1×10^{19} to 2×10^{20} atoms/cm^3. Surface concentrations of arsenic have not been reported.

Disadvantages of this technique include the possibility of growing, prior to deposition, a thin barrier oxide on the silicon at the relatively high deposition temperature. Also, since the rates of pyrolysis of the alkoxysilane and the dopant compounds generally have different temperature and concentration dependences, excellent control of flow rates and temperature is required. Furthermore, since some hydrocarbon fragments are produced by the pyrolytic cracking of alkoxy compounds, the resulting oxides may be somewhat contaminated. Nevertheless, Cuccia *et al.* [14] reported reproducibility of within ±10% for the process, and uniformity within a single run (twenty 2-inch diameter wafers) of ±7% for the case of boron. Whittle and Vick [15] claim about ±7% reproducibility for boron, using a two-stage diffusion. Reproducibility and uniformity data for the other dopants, as well as deposition rate data, are not generally available in the literature.

b. Low-Temperature Oxidation

A second technique used to deposit doped oxides is the low-temperature oxidation of volatile compounds of silicon and the various dopants. Normally the hydrides (silane, diborane, phosphine, arsine) are used [7,11,16], but Lee has reported the use of tetraethyl orthosilicate and arsenic trichloride[17] and Gittler the use of trimethyl stibine [18]. The oxidation typically takes place at below 500°C, and the dopant concentration in the oxide is controlled by the relative concentrations of the dopant and silicon compounds in the reactor. Deposition of doped oxide from the oxidation of silane is usually performed in a reactor using vertical nozzles [19] or horizontal nozzles [20] or in a quartz-tube reactor with a resistance-heated substrate holder [21]. Typical deposition rates in all systems are 1 μ of oxide per wafer per minute.

The major advantages of this technique are, again, the thickness of the deposited film and the range of possible dopant concentrations. Surface concentrations over the range 2×10^{17} to 1.5×10^{20} atoms/cm^3 for boron and 6×10^{17} to 2.5×10^{20} atoms/cm^3 for phosphorus have been reported [7,22], and there are no indications that this range cannot be extended downward by at least one order of magnitude. Contamination levels in the deposited oxides are generally low when the hydrides are used. The deposited oxides are usually easy to mask and etch, although Abe *et al.* [23] found the incorporation of germanium oxide in high-concentration boron- and arsenic-doped films greatly improved their etching characteristics.

The major disadvantage of low-temperature oxidation of hydrides as a deposition process is the difficulty in achieving uniform concentrations. The oxidation kinetics of most of the dopant hydrides appear to be different from those of silane, and it is thus difficult to maintain uniform concentrations (or uniform ratios of concentrations) within the deposition zone. This appears to be more easily achieved with the nozzle reactors than with the long tube reactors. Uniformities of boron concentration that have a relative standard deviation of about 2% have been reported [24], but reproducibility data are not available.

c. Reactive Sputtering and Vapor Transport

Nagano *et al.* [6] have deposited boron-doped silica films by reactively sputtering a boron-doped silicon target in an oxygen ambient. After diffusion at 1200°C in nitrogen, surface concentrations ranged from 1×10^{15} to 1×10^{17} atoms/cm^3; the ratio of surface concentration after diffusion to the boron concentration in the target was 5×10^{-4}. The deposition rate of the doped oxide was approximately 40 Å/min. Uniformity across a wafer was apparently within ±5%, but reproducibility appeared to be only within 100%.

Chu and Gruber [25] have reported the deposition of doped silica films by vapor-phase transport of the fluorides. Hydrogen fluoride and water vapor were used to transport boron and silicon oxides from a Pyrex source at 150°C to the deposition zone at 500°C. Other dopants, such as gallium, phosphorus, arsenic and antimony, can be used by introducing them in the form of volatile chlorides while transporting silica from a quartz source. After diffusion, surface concentrations were measured of 3×10^{17} to 2×10^{19} atoms/cm^3 for boron; 10^{17} atoms/cm^3 for gallium; 3×10^{17} to 1×10^{19} atoms/cm^3 for phosphorus; 10^{19} atoms/cm^3 for arsenic; and 1×10^{17} to 5×10^{19} atoms/cm^3 for antimony. Metallic contaminants were also transported from the Pyrex source and included in the deposited film. Typical rates of deposition were 0.2—1 μ/hr. No uniformity or reproducibility data were reported.

3.2. Liquid-Phase Deposition

Liquid-phase deposition techniques include anodically forming a doped silicon oxide film, depositing a film by electrophoresis, and spinning on either a suspension of doped silica particles or a chemically-reactive liquid film.

a. Anodic Oxidation

The technique of growing doped oxide films on silicon by anodization was first reported by Schmidt and Owen [3], and diffusion from these films was elaborated on by Schmidt *et al.* [26]. They used pyrophosphoric acid in tetrahydrofurfuryl alchohol as the electrolyte to form very heavily phosphorus-doped oxides for emitter diffusions and diethylphosphate—potassium nitrite—tetrahydrofurfuryl alcohol for more lightly doped oxides for base diffusions. By varying the phosphorus concentration in the latter electrolyte, they were able to obtain surface concentrations ranging from 5×10^{16} to 1×10^{20} atoms/cm^3 after diffusion at 1200°C in nitrogen. Kraitchman and Oroshnik [27] have reported diffusing from boron-doped oxides grown by anodization of silicon in a polyhedral borane acid—tetrahydrofurfuryl alcohol electrolyte, but no diffusion data were given.

A major disadvantage of doped anodic oxide films is the rapid depletion of the source by out-diffusion. The maximum thickness of anodic silicon dioxide is about 0.2 μ; unless this thin oxide is covered by an undoped silica film, a large portion of the dopant escapes to the ambient during the first few minutes of diffusion [26,27]. The rate of out-diffusion is considerably faster than would be calculated from normal diffusivities of the dopants in silicon dioxide, implying a rather open structure of the anodic films.

Probably the most serious disadvantage of the anodization technique is the incorporation of ionic species in the oxide. Schmidt *et al.* [26] found by radio tracer studies approximately 10 ppm of irremovable potassium oxide in their films after anodization, and they reported that the levels of sodium incorporated were much higher if a sodium ion electrolyte was used. Alkali ion contamination at even the 10 ppm level is deleterious to device performance.

Data reported by Schmidt *et al.* [26] showed uniformity of about ±20% and reproducibility within the same figure. The reproducibility was taken from anodizations made in the same batch of electrolyte over a period of 36 days; the authors claim this is a worst case example. The main advantage of the anodic oxidation technique is that the doped oxide grows only in the areas where silicon is exposed to the electrolyte, and this can reduce in some cases the number of masking and etching steps required.

b. Electrophoresis

The deposition of boron diffusion sources by the electrophoresis of boric oxide particles in suspension in nitroethane has been reported by Scala and Sandor [28]. Surface concentrations varying between 1×10^{17} and 1×10^{21} atoms/cm^3 (calculated from sheet resistivity—junction depth data, assuming an erfc distribution) were obtained after diffusing at 1200°C in nitrogen. Reproducibility appeared to be only within an order of magnitude; uniformity data have not been reported.

c. Spinning—Centrifugation

Two techniques have been described which utilize spinning techniques to deposit doped oxide films, one from a suspension of doped silica particles, the other from a chemically-reacting liquid. Kudrak [29] has deposited phosphorus- and boron-doped silica by sedimentation from a suspension of 70—100 Å silica particles in dilute boric acid or ammonium phosphate solutions. Surface concentrations of boron from 2×10^{17} to 5×10^{18} atoms/cm^3 and of phosphorus from 3×10^{17} to 2×10^{19} atoms/cm^3 were obtained. Reproducibility appeared generally to be about ±10%, although lightly doped films tended to show much more variation. No uniformity data were presented, nor was any attempt made to measure levels of contamination.

A commercially-available "spin-on" film [30] apparently uses the chemical condensation of alkoxysilanes with acetic anhydride to obtain a low-temperature silica film. This type of reaction is usually acid catalyzed, so that acidic forms of the dopants (e.g., phosphoric or boric acids) may be employed for

a dual purpose. The films are normally applied as a liquid to the rapidly rotating substrate; as the solvent diluent evaporates, the condensation is initiated. A wide range of surface concentrations of phosphorus, boron, arsenic, gallium, and antimony appear to be available. The main advantage of this technique is the ease of application, especially for use in high-volume production. Reproducibility and uniformity data are not presently available in the literature. Contamination from alkali metal ions can be severe unless special precautions are observed in the formulation and packaging of these materials.

4. Comparison with Standard Predeposition–Diffusion Techniques

As has been shown in the preceding sections, it is possible to control the surface concentration for doped oxide diffusions over a very large range: from less than 1×10^{16} to well over 1×10^{20} atoms/cm^3 in most instances. It is also possible to vary the concentration profile from one approaching an erfc relation to one approaching a Gaussian distribution simply by diminishing the thickness of the deposited film while diffusing in an inert atmosphere. An additional number of diffusion profiles can be achieved by diffusing in an oxidizing ambient; these will depend on the relative rates of diffusion and oxidation and may show either depletion or accumulation of dopant at the oxide-silicon interface, depending on the segregation coefficient and the relative diffusivities of the dopant in the oxide and the silicon. Obviously, combinations of sequential diffusions in inert and oxidizing ambients will give even more varied diffusion profiles. The important fact is that it is possible to vary the surface concentration over a wide range and, if desired, at the same time achieve fairly predictable concentration profiles.

Standard predeposition–diffusion techniques utilize a very shallow diffusion of a very high concentration of dopant into the silicon from the gas phase, usually with concurrent oxidation. This shallow layer of dopant is then redistributed into the silicon in subsequent diffusion steps, which also normally involve concurrent oxidation. For reasonable control of the predeposition process, it is usually necessary for the surface concentration to be near the solubility limit of the dopant in silicon; thus lower surface concentrations can be achieved only by long diffusions or by oxidizing the surface to deplete the dopant concentration.

Control of the predeposition step is a difficult matter. Two or more competing processes are generally occurring: decomposition of the dopant compound and diffusion into the substrate; and oxidation of the substrate, tending to block further diffusion. Small changes in the rate of either of these processes can greatly affect the quantity and distribution of the dopant in the silicon. Temperature control and temperature recovery of the furnace may be very important. For dopant compounds such as the halides, oxyhalides, and hydrides of phosphorus, boron and arsenic, or alkyl borates, which probably involve some gas-phase oxidation or decomposition, the concentration of oxygen or the partial pressure of the species itself may be critical. This involves careful control of flow rates and source temperatures if liquid sources are used. If sublimation from solids such as phosphorus pentoxide, antimony trioxide, or arsenic trioxide is used, then control of the sublimation furnace temperature profile may be critical. Many solid sources form a glassy phase which can sharply reduce the rate of sublimation; in these cases source aging time may be of importance.

Redistribution of the dopant during the diffusion cycle may also require close control. For dopants such as boron that tend to segregate into the oxide, the relative rates of oxidation and diffusion are again critical. Any traces of moisture in nominally dry oxygen can cause serious depletion of the boron. Also, for diffusions which use sequential dry-wet oxidation cycles, the time of each period of the sequence can be critical. Another serious problem is that the diffusivities of most dopants are dependent on concentration at high concentration levels. Thus a slightly enhanced surface concentration from the predeposition cycle can cause increased diffusion during the redistribution cycle, and this can lead to significant changes in diffusion depths and profiles.

In spite of this, standard predeposition–diffusion processes have a history of very successful use in silicon device manufacture. An empirical approach to these problems has resulted in processing which is remarkably reproducible and uniform, in terms of the characteristics of the devices produced.

Several different types of measurements are usually made for controlling the predeposition–diffusion process. Normally, sheet resistivity (or V/I measured with a 4-point probe) is used as an indicator for the predeposition process, and V/I, junction depth and junction breakdown voltages are used as control measurements for the diffusion cycle. Other measurements such as IR plasma frequency [31] and capacitance-voltage characteristics can be used in certain cases to determine surface concentrations and concentration profiles.

In theory, these measurements should be sufficient to describe diffusion profiles and thus device behavior. For proper process control the meaning of these measurements should be unambiguous: Ideally, sheet conductivity measurements after predeposition should accurately represent the total quantity of dopant material available for redistribution during diffusion; the junction depth and sheet conductivity after diffusion should represent a given concentration profile. In actuality, however, there can be a great deal of ambiguity in these measurements. For example, after most phosphorus predepositions there is an appreciable quantity of phosphorus in the silicon which is not electrically active and thus not measurable by sheet conductivity. The redistribution of boron during a dry-wet oxidation cycle is very

difficult to predict because of the concentration dependence of both the diffusivity of the boron and the oxidation kinetics of the silicon. Since the contribution to sheet conductivity comes predominantly from the high concentration regions, and these lie close to the surface, $V/I-x_j$ measurements give very little actual information about doping profiles in regions of lower concentration, e.g., in the active base.

From the foregoing, it would appear that the use of doped-oxide diffusion sources would be nearly universal, based solely on the independent control of surface concentrations and the predictability of diffusion profiles. However, the use of doped oxide sources is very rarely justified on this basis. The reason for this is that our present knowledge of both diffusion behavior and device design is not accurate enough to predict exact device characteristics from diffusion measurements. For example, for many high-gain, narrow-base-width NPN transistors using phosphorus emitters, the base width is affected by the amount of anomalous diffusion of the base under the emitter region, that is, by the emitter push effect. It appears impossible to predict from basic diffusion theory the exact concentration profile in such a base region. Thus, for doped oxide sources as well as for standard gas-phase sources, diffusion measurements must still act simply as indicators with a more-or-less successful correlation with final device characteristics.

There may still be other important advantages of doped oxide diffusions. There is no good basis of comparison of uniformity and reproducibility because most of these data for standard predeposition—diffusion cycles are considered proprietary. However, there is reason to believe that the uniformity and reproducibility quoted earlier for some doped oxide diffusions compare well with that for standard gas-phase processes. One advantage of doped oxide diffusions which is real is that of in-process control. It is possible to prediffuse a test wafer after deposition to gain precise information about concentration levels. It is then possible to compensate for variations either in concentration in the deposited film or in the previous diffusion steps by making modifications in the diffusion cycle. If it is not possible to compensate by means of a changed diffusion schedule, it is nearly always possible to etch off the doped oxide and redeposit a new source. Such corrective action is usually not possible with regular predeposition—diffusion techniques. If a sizable variation is encountered after predeposition, it is usually too late to compensate for it by modifying the diffusion cycle.

Flexibility in processing is also another real advantage of doped oxide diffusions. One example of this is the ability to make the thickness of the oxide (which may be used for subsequent diffusion masking, passivation, etc.) independent of the diffusion cycle. For shallow devices where diffusion times are not adequate to grow sufficient silicon dioxide for passivation, this is an advantage. Other examples of flexibility are the ability to do simultaneous diffusions of different dopants or the same dopant at different concentrations. This has important applications in making circuits which use complementary MOS or bipolar devices or diffused resistors with different sheet resistivities. Another example of this flexibility is the ability to do simultaneous source and drain diffusion and gate oxidation for self-aligned MOS transistors.

The use of doped oxide sources together with redundant masking techniques can result in a dramatic reduction in diffused pipes through pinholes or defects in the field oxide. Evaluation of this technique, however, requires careful examination of extensive yield data of specific devices before advantages can be claimed over standard diffusion techniques.

5. Summary

There are a large number of techniques that are useful in depositing doped oxides which can be used as diffusion sources. In general, these sources have the characteristics of providing independent control over surface concentration and of resulting in relatively unambiguous concentration profiles. Certain advantages of doped oxide diffusions over standard predeposition—diffusion techniques are immediately obvious due to their suitability for certain types of devices; examples of this are complementary and self-aligned gate MOS devices. Other suggested advantages, such as uniformity, reproducibility, and reduction in diffusion-related defects, can be tested only by thorough and detailed comparison with existing processes. A third class of advantages, including improved in-process control and adaptability to continuous high-volume processing, can be realized by careful integration with the preceding and subsequent processes. The major disadvantages of doped oxide diffusions are the positive control required over the dopant concentrations and the relative complexity of the deposition process.

6. References

[1] Shortes, S. R. and Wurst, E. C., Diffusion of N- and P-type Impurities into GaAs from Doped SiO_2 Films, Abstract #51, New York Meeting of The Electrochemical Society, Oct. 1963.

[2] Becke, H., Flatley, D., Kern, W., and Stolnitz, D., The Diffusion of Zinc into Gallium Arsenide to Achieve Low Surface Concentrations, *Trans. Met. Soc. AIME* 230, 307 (1964).

[3] Schmidt, P. F. and Owen, A. E., Anodic Oxide Films for Device Fabrication in Silicon, *J. Electrochem. Soc.* 111, 682 (1964).

[4] Scott, J. and Olmstead, J., A Solid-to-Solid-Diffusion Technique, *RCA Review* 26, 357 (1965).

[5] Barry, M. L. and Olofsen, P., Doped Oxides as Diffusion Sources, Abstract #169, Chicago Meeting of The Electrochemical Society, Oct. 1967.

[6] Nagano, K., Iwauchi, S., and Tanaka, T., Solid-Solid Diffusion of Boron in Silicon using Reactive Sputtering, *Japan. J. Appl. Phys.* 7, 1361 (1968).

[7] Barry, M. L. and Olofsen, P., Doped Oxides as Diffusion Sources: I. Boron into Silicon, *J. Electrochem. Soc.* 116, 854 (1969).

[8] Owen, A. E. and Schmidt, P. F., Diffusion from a Plane Finite Source into a Second Phase with Special Reference to Oxide-Film Diffusion Sources on Silicon, *J. Electrochem. Soc.* 115, 548 (1968).

[9] Tien, T.-Y. and Hummel, F. A., The System SiO_2-P_2O_5, *J. Am. Ceram. Soc.* 45, 422 (1962).

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

[11] Mecs, B., Diffusion from Doped SiO_2 in Various Ambiences, *Bull. Am. Phys. Soc.* 13, 1476 (1968).

[12] Mecs, B., private communication.

[13] Von Muench, W., Tin and Zinc Diffusion into Gallium Arsenide from Doped Silicon Dioxide Layers, *Solid-State Electron.* 9, 619 (1966).

[14] Cuccia, A., Shrank, G., and Queirolo, G., Diffusion from Doped Oxides, Abstract #305, New York Meeting of The Electrochemical Society, May 1969.

[15] Whittle, K. M. and Vick, G. L., Control of Boron Diffusion from a Pyrolytic Borosilicate Glass Source, *J. Electrochem. Soc.* 116, 645 (1969).

[16] Toliver, D. L. and Elgan, D. L., Investigation of Deposited Phosphorus Oxides and Their Use as a Diffusion Source for Silicon, Abstract #183, Philadelphia Meeting of The Electrochemical Society, Oct. 1966.

[17] Lee, D. B., Diffusion into Silicon from an Arsenic-Doped Oxide, *Solid-State Electron.* 10, 623 (1967).

[18] Gittler, F. L., A Doped Oxide Deposition System for Antimony Diffusions, Abstract #182, Detroit Meeting of The Electrochemical Society, Oct. 1969.

[19] Flaschen, S. S., Evitts, H. C., and Black, J. R., Large Area Semiconductor Surface Passivation Production Program, AD439885, Jan. 1964.

[20] Barry, M. L., A Continuous Process for the Deposition of Silicon Dioxide from the Oxidation of Silane, 2nd Int'l Conf. on Chemical Vapor Deposition, Los Angeles Meeting of The Electrochemical Society, May 1970.

[21] Swanson, T. B. and Hull, R., Chemical Vapor Deposition of Doped and Undoped Silicon Dioxide Films using a Horizontal System, Abstract #87, Detroit Meeting of The Electrochemical Society, Oct. 1969.

[22] Barry, M. L., Doped Oxides as Diffusion Sources: II. Phosphorus into Silicon, submitted to *J. Electrochem. Soc.* (1970).

[23] Abe, T., Sato, K., Konaka, M., and Miyazaki, A., Impurity Diffusion into Si using Ge-Doped SiO_2 Films, Abstract #184, Detroit Meeting of The Electrochemical Society, Oct. 1969.

[24] Barry, M. L. and Olofsen, P., Advances in Doped Oxides as Diffusion Sources, *Solid-State Tech.* 11, No. 10, 39 (1968).

[25] Chu, T. L. and Gruber, G. A., Doped Silica Films, *Electrochem. Tech.* 5, 44 (1967).

[26] Schmidt, P. F., O'Keeffe, T. W., Oroshnik, J., and Owen, A. E., Doped Anodic Films for Device Fabrication in Silicon, *J. Electrochem. Soc.* 112, 800 (1965).

[27] Kraitchman, J. and Oroshnik, J., Anodic Oxide Growth Behavior on Silicon, *J. Electrochem. Soc.* 114, 405 (1967).

[28] Scala, L. C. and Sandor, J. E., Electrophoretic Process for the Controlled Deposition of Boron Diffusion Sources on Silicon, *Electrochem. Tech.* 6, 435 (1968).

[29] Kudrak, D. R., Diffusion from Doped Oxides, RADC-TR 68-239 (1968).

[30] Emulsitone Company, Livingston, New Jersey.

[31] Gardner, E. E., Kappallo, W., and Gordon, C. R., Measurement of Diffused Semiconductors Surface Concentrations by Infrared Plasma Reflection, *Appl. Phys. Letters* 9, 432 (1966).

Capacitance - A Device Parameter and Tool for Measuring Doping Profiles

B. R. Chawla

Bell Telephone Laboratories, Incorporated
Murray Hill, New Jersey 07974

The performance of semiconductor junction devices depends greatly upon their geometry and the doping profiles. One of the important techniques for measuring the doping profile nondestructively, has been the use of capacitance-voltage measurements. Capacitance also forms an important parameter for varactors and some of the transistor characteristics. The technique for measuring doping profiles using capacitance-voltage measurements is based on the classical C square-root V law for a step junction. An automatic implementation of this technique for plotting inverse of doping versus distance has been recently reported by J. A. Copeland [1].[1] However, the use of this technique has certain limitations. In this paper we shall discuss these limitations and present results which increase the range of application of the capacitance-voltage technique. The applicability of the square-root law is limited to junctions where the doping on one side of the junction is much greater than that on the other side. The normal usage of this technique lies in the range of reverse bias which restricts the knowledge of the doping profile away from the junction. To obtain information about the doping profile, in the vicinity of the junction, one needs to forward bias the junction. It is in this range that the concepts of "built-in voltage" and of a narrowing parallel plate capacitor lead to a nonphysical result, viz., infinite capacitance. Using an alternative definition for the transition region capacitance (not including the diffusion capacitance) our calculation shows that it remains finite and that one can still think in terms of an equivalent parallel plate capacitor in which the position of the plate is well marked. Consequently, we observe that the quantity "built-in voltage" is rather ill defined. Based on the analysis of linearly graded junctions, a new quantity "offset voltage" will be proposed. Further limitations of the C square-root V law arise when a region of higher doping density than the one near the surface, e.g., the interface between a lightly doped epitaxial layer on a highly doped surface, is probed by the C-V technique. Kennedy et al. [2] have demonstrated that the quantity obtained by this technique is the equilibrium majority carrier concentration and not the impurity doping profile. This conclusion is supported by our calculations for more practical doping profiles than the ones considered in the above paper.

Key Words: Built-in voltage, C square-root V law, C-V technique, capacitance, carrier concentration, doping profile, forward bias, impurity doping, junction devices, transition region capacitance.

---

[1] Figures in brackets indicate the literature references at the end of this abstract.

References

[1] J. A. Copeland, A technique for directly plotting the inverse coping profile of semiconductor wafers, IEEE Trans. Elec. Dev., Vol. ED-16, pp. 445-449 (May 1969).

[2] D. P. Kennedy, P. C. Murley and W. Kleinfelder, On the measurement of impurity atom distributions in silicon by the differential capacitance technique, IBM J. Res. Dev., pp. 399-409 (Sept. 1968).

Concentration Dependent Diffusion Phenomena

P. E. Bakeman, Jr.

Rensselaer Research Corporation
Troy, New York

J. M. Borrego

Rensselaer Polytechnic Institute
Troy, New York

At low impurity concentrations, the diffusion of impurities in semiconducting materials obeys Fick's first and second laws. The impurity atom flux is equal to the product of the concentration gradient and the diffusion coefficient; for a given materials system, the latter is a function of temperature only. At high impurity concentrations, however, many workers have experimentally observed that this relationship is no longer valid. In these cases experimental diffusion results can be expressed in terms of a concentration as well as temperature dependent diffusion coefficient.

In this presentation we will review the experimental work and associated theoretical explainations for concentration dependent diffusion phenomena. Topics to be covered include field aided diffusion, strain enhanced diffusion, electronic vacancy diffusion effects and experimental measurement techniques.

Also to be presented is a theory developed by the authors which treats concentration dependent diffusion phenomena from a fundamental point of view. The theory is based on a simple assumption for impurity atom interaction in diffusion systems which exhibit solid solubility saturation. The theory yields a self consistent result for the diffusion coefficient and chemical potential as a function of impurity concentration. The theoretically predicted diffusion coefficient is : $D' = D (C_{max} + C) / (C_{max} - C)$, where D is the low level diffusion coefficient, C is the impurity concentration, and C_{max} is the saturation impurity concentration. The theory is compared with available experimental data on the diffusion of phosphorus in silicon. The theory and experiments are in excellent agreement.

Key Words: Anamolous diffusion, concentration dependent diffusion diffusion review, diffusion theory, semiconductor impurity diffusion, solid solubility saturation.

1. Introduction

The diffusion of doping impurities into semiconducting materials is generally regarded to obey Fick's first and second laws of diffusion. That is - - the atomic flux is proportional to the concentration gradient.

$$J_x = - D \frac{dC(x)}{dx} \tag{1}$$

D, the diffusion constant, is the constant of proportionality. The second law is the result of applying conservation of matter to the first law and is mathematically stated:

$$\frac{dC(x)}{dt} = \frac{d}{dx} \left(D \frac{dC(x)}{dx}\right) \tag{2}$$

or

$$\frac{dC(x)}{dt} = +D \frac{d^2}{dx^2} (C(x)) \tag{3}$$

if the diffusion constant is not a function of position.

The well known complementary error function diffusion profile results when equation 3 is solved with the boundary condition of constant surface concentration. If a source of a fixed number of impurity atoms is assumed to be on the surface at the start of diffusion, a gaussian diffusion profile results.

Experimental deviations of measured diffusion profiles from the theoretical ones predicted using equation 3 are termed anomalous. The anomalous behavior to be treated in this paper is the result of the diffusion "constant" being a function of concentration as well as temperature. The variation of the diffusion constant as a function of concentration comes from the analysis of constant surface concentration diffusion data as well as radiotracer isoconcentration diffusion studies.

The classical approach to deriving the diffusion constant from the atomic model of an impurity atom diffusing in a host lattice involves the application of random-walk theory to the motion of impurity atoms. If the diffusing impurity moves by a substitutional mechanism, the diffusion constant is proportional to the density of vacancies times the probability that the impurity atom will jump into an adjacent vacant site. The foregoing assumes that each impurity atom moves entirely independent of other impurity atoms and thus results in a diffusion constant which is a function of temperature only.

Interactions between diffusing impurity atoms are not considered in classical diffusion theory. Consequently, concentration dependent diffusion results are considered anamolous. The authors have recently shown that by including impurity atom interactions, a concentration dependent diffusion coefficient can be derived which is in excellent agreement with experimentally determined data.

2. Experimental Review

2.1 Experimental Procedures

A great deal of effort has been expended in studying the diffusion of doping impurities into semiconducting materials because of the importance of the process in fabricating semiconducting devices. The experimental techniques can be divided into two classes:

1) The evaluation of the impurity profile after a standard diffusion process

2) Performing a special diffusion process for the purpose of profile determination

The measurement of a diffusion profile after a standard diffusion process can be performed by removing thin layers of semiconductor and measuring either the amount of impurity removed or the amount remaining in the wafer. The amount of impurity in the removed portion can be determined by chemical analysis or by radioactive decay measurements. The distribution of dopant in the diffused wafer can also be calculated from measurements of sheet conductivity as a function of the amount

of material removed. This latter method requires knowledge of carrier mobility and impurity ionization as a function of impurity concentration, which are not always known. In particular, impurity precipitation within the semiconductor will yield calculated values of impurity concentration which are below the true value especially at high impurity concentrations. In order to measure the total impurity concentration (active and inactive), Tannenbaum (1)* neutron activated phosphorus diffused silicon, after it had been diffused, and measured the concentration of the radioactive phosphorus by sectioning techniques. This measurement can be performed in conjunction with sheet conductivity profiles to determine the electrically active and total impurity concentration profiles in the same sample.

All the above mentioned methods of impurity profile determination require that thin sections of known thickness be removed from the sample as part of the measuring process. This is usually performed in silicon by anodic oxidation and subsequent oxide removal. This technique provides accurate removal (1) of thin layers which are parallel to the original source-semiconductor interface.

Special diffusion processes for measuring diffusion profiles include diode fabrication for depletion capacitance measurements and radioactive tracer diffusion studies. The capacitance measurements are best suited to low impurity concentrations where diffusion is concentration independent. Radioactive tracer diffusion studies can be performed in numerous ways which are dependent upon the materials system to be studied. Masters and Fairfield (2) studied the diffusion of radioactive arsenic 76 in arsenic doped silicon and showed the concentration dependance of the diffusion coefficient of arsenic in silicon in the absence of concentration gradients and dislocation networks. Similiarly they (3) studied the diffusion of radioactive silicon 31, half-life 2.62 hrs, in silicon as a function of impurity concentration.

Other measurements while they will not provide detailed diffusion profiles, are useful in detecting anomalous diffusion results. Mackintosh (4) evaluated the diffusion of phosphorus in silicon using the sheet conductivity and the junction depth of the diffused layer. His data indicate enhanced diffusion at high phosphorus concentration.

2.2 Experimental Results

There are numerous puplished reports of anomalous high concentration diffusion results. There is general agreement in the experimental results of various workers where the experimental diffusion conditions were similar. We shall list pertinent experimental observations and results along with references to their sources.

Observed and measured enhanced high concentration diffusion coefficients of phosphorus in silicon (1,4,5,6,7,8,9)

Observed and measured enhanced high concentration diffusion coefficients of boron in silicon (7,10,11)

Found the diffusion constant of arsenic in intrinsic silicon to be $D_i = 60 \exp(-4.20 \text{ ev/kT})$; and $D = D_i\left(\frac{n}{n_i}\right)$ in intrinsic silicon (2)

Found the self diffusion coefficient of silicon to be: $D = 9{,}000 \exp(-5.13 \text{ ev/kt})$ in intrinsic silicon. High doping levels of either phosphorus, boron, or arsenic increased this value (3)

---

* Figures in brackets indicate the literature references at the end of this paper

Observed the formation of high dislocation density areas due to high concentration phosphorus diffusions (5,6,12). Only slight dislocation formation found due to arsenic diffusion(5)

Observed precipitation of phosphorus on dislocations formed by high concentration diffusion (5,6,12)

Noted diffusion retardation effect for phosphorus in silicon just before dislocation network formations (8)

3. Theoretical Review

Several authors have advanced theories of concentration dependent diffusion phenomena. The theories include field aided diffusion, strain enhanced diffusion, electronic vacancy generation and dislocation precipitation. In addition to these, S, M. Hu (13) has developed a general thermodynamic model to describe diffusion.

The field aided diffusion is the result of an electric field present in the semiconductor at diffusion temperatures. The field is the result of the concentration gradient of ionized doping impurity. The ion current (impurity flux) is represented by the equation:

$$J_+ = -D_+ \frac{dC}{dx} + EC \frac{qD+}{kT} \tag{4}$$

The electron current is:

$$J_- = +qnE\mu_- + qD_- \frac{dn}{dx} \tag{5}$$

and

$$D = \frac{kT}{q} \mu \tag{6}$$

but no electron current is allowed to flow, therefore

$$E = - \frac{kT}{q} \frac{dn}{dx} \frac{1}{n} \tag{7}$$

Thus if the material is extrinsic and n=C the ion current is

$$J_+ = -2D_+ \frac{dC}{dx} \tag{8}$$

and the diffusion constant is increased by a factor of two. This model is generally accepted as valid, but does not produce a sufficiently great enhancement of the diffusion constant to explain the observed results.

Strain enhanced diffusion and electronic vacancy generation mechanisms have been proposed. by Thai (14) and Masters and Fairfield (2), respectively. Both theories rely on the production of excess vacancies to increase the impurity diffusion constant. Thai's theory predicts the generation of vacancies due to plastic flow in silicon with high concentration phosphorus diffusions. His theory is based on Prussin's theory (15) of dislocation stress relief due to the lattice concentraction of boron and phosphorus in silicon. Thai's theory does not account for the rapid diffusion and surface recombination of excess non-equilibrum vacancies; this consideration could lower the theoretically predicted diffusion enhancement considerably.

Fairfield and Masters (3,2) attribute the concentration dependent diffusion phenomena that they observed to a vacancy acceptor level 0.34ev below the conduction band. While this explains

concentration dependent diffusion in N type silicon, they state that an alternate explanation is required for the results observed in P-type silicon.

Joshi and Dash (16) attribute the deviation of measured diffusion profiles from the ideal complementary error function diffusion profile to the precipitation of phosphorus on stress induced dislocations. The precipitated phosphorus is electrically inactive and therefore cannot be measured by sheet conductivity measurements. While this model may explain some of the observed results, it does not explain radioactive tracer diffusion profiles or isocentration diffusion results.

4. Impurity Atom Interaction

Fick's first law is based on an ideal model in which the impurity atoms move at random in the host lattice. The motion of each impurity atom is random because there are no net forces acting on the diffusing atoms. An atomic flux results from a concentration gradient simply because there are more impurity atoms in high concentration regions which randomly diffuse into lower concentration regions than visa versa. Thus, if we could tag a given impurity atom, there would be an equal probability that it would move in the plux x and the minus x directions regardless of the presence of a concentration gradient. This model is very simple and usually describes observed diffusion phenomena very well.

However, impurities which are used to dope semiconductors usually exhibit a phenomenon which is inconsistent with the assumption used to drive Fick's first law. This phenomenon is solid solubility saturation. It is experimentally found that the amount of impurity that will dissolve in a semiconductor, at an elevated temperature, saturates at a specific value as the concentration of the impurity at the surface of the semiconductor is increased. The concentration at which saturation occurs is temperature dependent and is referred to as the solid solubility concentration. This saturation effect has largely been ignored in explaining the reasons for concentration dependent diffusion phenomena.

The authors recently published a paper which includes impurity atom interaction effects in the diffusion of impurities which exhibit solid solubility saturation (17). The derivation will not be presented in detail here, but the approach and results will be discussed. The derivation was performed in the same manner as that of the classical derivation of Fick's first law (18,19) except that interactions between diffusing impurity atoms were included in the calculations. This was done by assuming that a maximum density of impurity could exist in the host lattice. The impurity atoms were then assumed to move at random to available locations within the lattice. The density of available locations is equal to the maximum concentration of impurity minus the actual density of impurity.

The effect of the interaction is to increase the probability that an impurity atom will move into low concentration rather than into high concentration regions of the host lattice. When impurity atom interaction is included, it is no longer thrue that a "tagged" impurity atom has an equal probability of moving in any direction regardless of the concentration gradient. There are more available impurity atom locations in low concentration regions; thus, an impurity atom preferentially moves in the direction of lower concentration.

The impurity atom flux which results from the above interaction assumption is:

$$J = -1/2 \Gamma \alpha^2 \frac{dC(x)}{dx} \frac{C_{max} + C(x)}{C_{max} - C(x)}$$

where, $D = 1/2 \Gamma \alpha^2$ is the impurity diffusion constant without impurity atom interaction. Thus, the concentration dependent diffusion coefficient is:

$$D' = D \frac{C_{max} + C(x)}{C_{max} - C(x)}$$

where C_{max} is the solid solubility concentration of the impurity and C(x) is the impurity concentration.

This theoretical result is compared with the experimental results of Tannebaum (1) in figure 1. The data points are from experimental radio tracer measurements of two phosphorus diffused silicon samples, 1050° C -- 30 minutes each (1). The curves are calculated from equation 10 using $D = 7.4 \times 10^{-13}$ cm^2/sec and $C_{max} = 7.4 \times 10^{20}$ cm^{-3} (upper) and $D = 4.8 \times 10^{-13}$ cm^2/sec and $C_{max} = 7.7 \times 10^{20}$ cm^{-3} (lower). The agreement between theoretical and experimental data is excellent.

Figure 2 shows the chemical potential of an impurity in a host lattice as a function of concentration. The dotted line is for an impurity which does not exhibit solid solubility saturation. The solid curve represents an impurity that shows solid solubility saturation at C_{max}. This curve was calculated by equating the impurity atom flux obtained from equation 10 to the same flux as a function of chemical potential gradient. It is of interest to note that the rapid rise in chemical potential at C_{max} would be required to cause the experimentally observed and theoretically assumed phenomenon of solid solubility saturation. Thus, the derivation which we have outlined in this section is a self-consistent theory of concentration dependent diffusion phenomena which includes the effects of impurity solid solubility saturation.

5. Conclusions

Experimental work in the field of concentration dependent diffusion has been briefly reviewed. The theoretical models which have been used to describe the observed experimental phenomena have been presented and discussed. A new theory which included the effects of solid solubility saturation is presented and found to agree well with previous experimental results.

6. References

(1) Tannenbaum, Eileen, - Detailed Analysis of Thin Phosphorus-Diffused Layers in p-Type Silicon, Solid State Electronics 2 pp. 123-132, (Oct. 1960).

(2) Masters, B. J., and Fairfield, J. M. - Arsenic Isoconcentration Diffusion Studies in Silicon, J. Appl. Phys. 40, pp. 2390-2394, (May 1969).

(3) Fairfield, J. M. and Masters, B. J., - Self-Diffusion in Intrinsic and Extrinsic Silicon Journal of Applied Phys. 38, pp. 3184-3154, (July 1967).

(4) Mackintosh, I. M., - The Diffusion of Phosphorus in Silicon, J. of Electrochemical Society, 109 , pp. 392-401, (May 1962).

(5) Joshi, M. L. and Wilhelm, F., - Diffusion-Induced Imperfections in Silicon, Journal of Electrochemical Society, pp. 185-188, (Feb. 1965).

(6) McDonald, R. A., Ehlenberger, G. G. and Huffman, T. R., - Control of Diffusion Induced Dislocations in Phosphorus Diffused Silicon, Solid-State Elec., 9, pp. 807-812, (1966).

(7) Nicholas, K. H., - Studies of Anomalous Diffusion of Impurities in Silicon, Solid-State Electronics, 9, pp. 35-47, pp. 35-47, (1966).

(8) Duffy, M. C., Barson, F., Fairfield, J. M. and Schwuttke, G. H., - Effects of High Phosphorus Concentration on Diffusion into Silicon, J. Elec. Soc., 115, pp. 84-88, (Jan. 1968).

(9) Tsai, Joseph C.C., - Shallow Phosphorus Diffusion Profiles in Silicon, Proceedings of the IEEE, 9, pp. 1499-1506, (Sept. 1969).

(10) Barry, M. L. and Olofsen, P., - Doped Oxides as Diffusion Sources, J. of Elec. Society, 116, pp. 854-869, (June 1969).

(11) Vick, G. L. and Whittle, K. M., - Solid Solubility and Diffusion Coefficients of Boron in Silicon, J. of Electrochemical Society, 116, pp. 1142-1144, (Aug. 1969).

(12) Jaccodine, R. J., - Precipitates Formed by High-Concentration Phosphorus Diffusion in Silicon, J. of Appl. Phys. 39, pp. 3105-3108, (June 1968).

(13) Hu, S. M. - General Theory of Impurity Diffusion in Semiconductors via the Vacancy Mechanism, Physical Review, 180, pp. 773-784, (April 1969).

(14) Thai, N. D., Anomalous Diffusion in Semiconductors-A Quantitative Analysis, Solid-State Electronics, 13, pp. 165-172, (Feb. 1969).

(15) Prussin, S., - Generation and Distribution of Dislocations by Solute Diffusion, J. of Appl. Phys., 32, pp. 1876-1881, (Oct. 1961).

(16) Joshi, M. L. and Dash, S., - Dislocation-Deviation of Phosphorus-Diffusion Profiles in Silicon, IBM Journal, pp. 446-454, (Nov. 1966).

(17) Bakeman, P. E., Jr., and Borrego, J. M. - The Theory of Anomalous Diffusion in Solids Near Diffusant Saturation Concentrations: Example-Phosphorus in Silicon, Journal of Electrochemical Society, 117, pp. 688-690, (May 1970).

(18) Shewman, P. G., - Diffusion in Solids, McGraw-Hill Book Co., pp. 41-42, (1963).

(19) Manning, J. R. - Diffusion Kinetics for Atoms In Crystals, D. Van Nostrand Co., Inc., (1968).

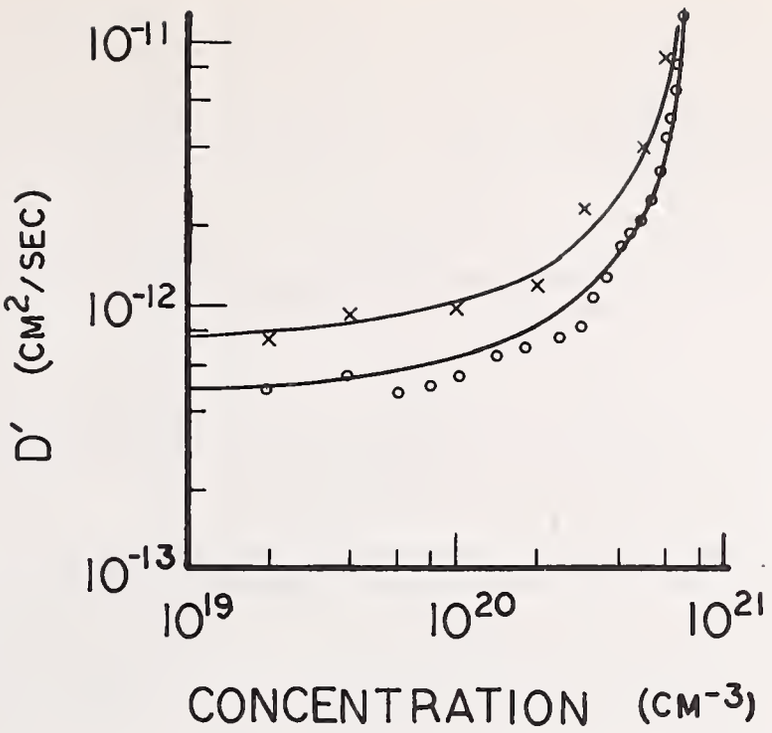


Figure 1. Experimental data points (from Tannenbaum (1)) and theoretical curves (equation 10) for the concentration dependent diffusion coefficient of phosphorus in silicon 1050° C.

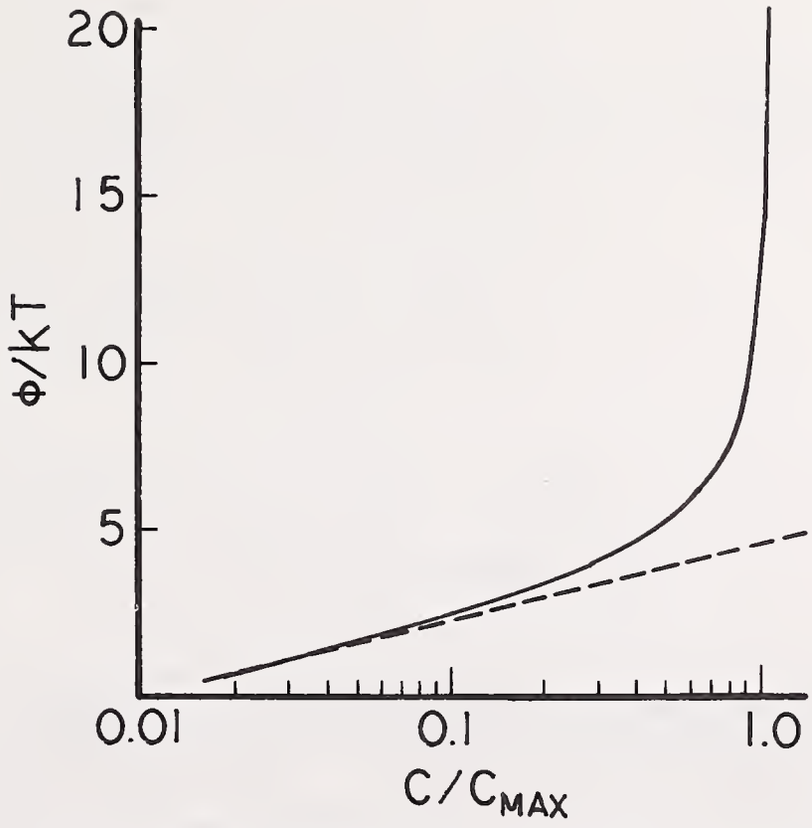


Figure 2. Impurity chemical potential versus concentration. The dotted curve is without impurity interaction effects --solid curve includes impurity interaction and corresponds to the concentration dependent diffusion coefficient of equation 10.

Orientation Dependent Diffusion Phenomena

L. E. Katz

Bell Telephone Laboratories, Inc.
Allentown, Pennsylvania

This paper is devoted to reviewing the present knowledge and analyzing those parameters associated with orientation dependent diffusion in silicon. Particular emphasis is placed on the orientation dependence of boron in silicon.

It has been experimentally observed that the diffusion of boron in (100) is greater than in (111) under oxidizing and perhaps other ambients. Concentration profiles of the diffusion of boron in silicon have been calculated in an effort to determine the parameters which are responsible for the enhanced diffusion. The parameters investigated were segregation coefficient, oxide growth rate constant and diffusion coefficient. These parameters were individually varied in the diffusion equation in order to observe their effect on the concentration profile and hence the junction depth.

The computer generated results suggest that:

A) The increased growth rate of the oxide in (111) in comparison with (100) is insufficient to produce the increased junction depth in (100) which is experimentally observed (30%).

B) A variation of the segregation coefficient from 3 to 100 changes the junction depth less than 5%. It is unlikely that the segregation coefficient would change by a sufficient amount to alter the junction depth to the observed value.

C) A solute-dislocation interaction is discussed which could increase the diffusion coefficient in (100) with respect to (111) due to an increased rate of vacancy generation. If boron is considered to diffuse by a monovacancy mechanism, an increased junction depth of about 15% is calculated. If a divacancy mechanism is considered, for which there is strong evidence, the junction depth is calculated to increase by 30-35% in good agreement with observed results.

The problem of orientation dependence is also analyzed on the basis of work hardening theory considering vacancy generation via dislocation—"dislocation forest" interactions. It is concluded that any enhanced diffusion based on this mechanism would be expected to occur at high concentration (mid to upper $10^{19}/cm^3$ range) and would result in more enhancement for diffusion in the (111) direction in comparison with the (100) direction.

Key Words: Boron, diffusion, dislocation, orientation dependence, segregation coefficient, silicon, stress enhanced diffusion.

1. Introduction

Many experimenters [1-10] have observed that anomalous diffusion occurs when impurities such as boron and phosphorus are diffused into silicon. Generally, the attempts made to explain the observed results deal with strain and strain relief, dislocation generation, dislocation interaction and production of vacancies.

Aside from the anomalous diffusion, it has also been observed that under certain experimental conditions diffusion of boron in silicon is orientation dependent.[11-13] Boron diffuses more rapidly in (100) oriented material than that of (111) orientation. Benson and Paulnack[11] observed the junction depth to be about 30% deeper in (100) when compared to (111) following a steam oxidation. Wills[12] performed a drive-in at 1050°C in oxygen and observed an enhancement of the order of 50%. He did not observe the enhancement when a nitrogen atmosphere was used and suggested that the effect could be caused by an increase in the diffusion coefficient in the (100) direction in an oxygen ambient, or by an orientation dependent change in the impurity redistribution of the diffused layer during oxidation. Kovalev et al.[13] observed similar effects and offered an explanation based on the different degrees of depletion of the diffusion sources, due to segregation coefficient changes caused by the different "reticular" (lattice array) densities of the planes. Their analysis stated the ratio of the densities for the planes (100), (110) and (111) as 1:1.414:2.308 whereas the correct values are 1:1.414:1.16. Their subsequent analysis is based on the incorrect ratio. Use of the correct values have shown their arguments to be invalid.

The author has observed enhancement of (100) in contrast to (111) following steam oxidation. An enhancement, of lesser magnitude, has also been observed following drive-in in a non-oxidizing ambient (1 to 3 percent oxygen used to prevent nitride formation). Indications are that the effect may also occur during the pre-depth stage, although this is difficult to substantiate due to the extremely shallow junctions. It was deemed advisable to analyze the entire diffusion process in terms of the various parameters in an effort to determine the cause of the orientation dependent diffusion. A recent paper[14] has discussed anomalous diffusion in semiconductors on the basis of work-hardening theory. This work hardening mechanism has been analyzed here from the standpoint of its orientation dependence.

The initial portion of this paper is devoted to investigating the parameters in the diffusion equation which are orientation dependent and could alter the junction depth.

2. Diffusion Conditions, Parameters and Equation

The first step in the investigation was to obtain a solution to the diffusion equation with the particular set of boundary conditions pertinent to the experimental conditions. Generally the experimental setup involves a "pre-deposition" followed by a drive-in prior to the steam oxidation. Normally an error function distribution exists following the "pre-deposition" and a Gaussian after the drive-in. Most available solutions to the diffusion equation consider an initial uniform distribution (step-function); the solutions become complicated when other initial boundary conditions are considered.

Kato and Nishi[15] studied the redistribution of diffused boron in silicon by thermal oxidation considering as boundary conditions:

a) predeposition of boron onto the surface of a silicon slice and simultaneous diffusion into the silicon under the constant surface concentration of boron. The resulting distribution was an error function.

b) thermal oxidation of the silicon slice accompanied by redistribution of boron at the silicon-silicon dioxide interface.

This is a moving boundary problem taking into consideration the variables with which we are concerned; i.e., segregation coefficient, oxide growth constant and diffusion coefficient.

The solution to Kato and Nishi's[15] differential diffusion equation is given approximately by means of the method of Green's function. Cave[16,17] has simplified Kato and Nishi's[15] equations to arrive at mathematically convenient results. His assumptions which differ from those of Kato and Nishi[15] are that the oxide growth follows a linear law (parabolic growth is usually assumed although the growth is initially linear and subsequently parabolic[18]) and that the boron distribution prior to the oxidation is given by exp (- y/L), where L is obtained by matching the exponential profile to the erfc profile and y is the distance from the Si-SiO_2 interface. Cave[17] allows for the possibility of diffusion of boron across the oxide. In view of the approximations made, the simpler approach of Cave[17] yields very good results when compared with Kato and Nishi. This equation is quite convenient to work with and can readily be programmed. It is therefore possible to investigate theoretically changes in all the variables in the equation; namely, predeposition temperature and time, oxidation time, distance from the interface, diffusion coefficient, segregation

coefficient and oxide growth rate.

The possibility of the diffusion coefficient varying with orientation must be considered since defect movement and interaction may be orientation dependent. An interesting paper by Parker[6] theoretically studies the generation of excess vacancies at climbing diffusion-induced dislocations and dislocation enhanced diffusion in (001) crystals. His final expression shows that

$$\left(\frac{dN}{dx}\right)_{[001]} = 1.26 \left(\frac{dN}{dx}\right)_{[111]} \quad (1)$$

where $\frac{dN}{dx}$ is the rate of emission of vacancies from a band of dislocations as the band moves through the crystal occupying approximately the same position as the steepest part of the diffusion profile. Since the density of the diffusion-induced dislocations ($\sim 10^9$ cm^{-2}) is generally considerably higher than the initial dislocation density ($\sim 10^3$ cm^{-2}), the excess vacancies cannot be absorbed at nearby negative edge dislocations, but must diffuse to the crystal surface. This results in an enhancement of the rate of diffusion which is comparable in magnitude to the normal rate of diffusion[6] at 1000°C. Since boron diffuses substitutionally in silicon, with the rate of diffusion proportional to the total vacancy concentration, the above observations must be considered as a possible explanation; i.e., larger diffusion coefficient in (100) with respect to (111); for the observed enhanced diffusion in (100).

Prior to performing the computer analysis of the variables, the boundary conditions were chosen. Since, normally an erfc or Gaussian distribution is assumed prior to oxidation, the predeposition and drive-in were considered to be one combined operation giving the exponential (fitted to an erfc) required by exp (- y/L). The boundary conditions were as follows:

pre-depth and drive in	25 minutes @ 1015°C
oxidation	120 minutes @ 1100°C
K (linear oxide growth rate constant)	1.5×10^{-8} cm/sec for (111)
	1.3×10^{-8} cm/sec for (100)

K was determined approximately from experimental data.
k = segregation coefficient, defined as the equilibrium concentration of boron in the oxide divided by the equilibrium concentration of boron in silicon
k[15][17][19] = 3 or 10; the most commonly quoted values in the literature. D, the diffusion coefficient of boron in silicon @ 1100°C, was taken to be 3×10^{-13} cm^2/sec see Ref. [15].

It was pointed out, in a recent review article by Kendall,[20] that the diffusion coefficient of boron in silicon is concentration dependent. The diffusion coefficient increases by about an order of magnitude as the boron concentration increases from about 5×10^{19} to 6×10^{20}/cm^3.

3. Calculations and Analysis of Results

The computer analysis was carried out under the following conditions:

a) k and D were constant with orientation and only K was changed.

$$K = 1.30 \times 10^{-8} \text{ cm/sec for (100)}$$
$$K = 1.50 \times 10^{-8} \text{ cm/sec for (111)}$$

Figure (1) shows the resulting profiles; n/N_{SO} is the boron concentration divided by the initial surface concentration ($\sim 10^{20}$/cm^3). If the junction occurs at the 10^{16}/cm^3 level (corresponding to 10^{-4} on the ordinate of Figure 1, o and x), it is seen that there is virtually no difference in the junction depth between the two orientation. Certainly the difference is less than 5%, the (100) junction depth being slightly deeper. In this case k was set equal to 3. When k equals 10 a slightly larger difference is observed (Figure 2, Δ and □); however, this amounts to only about a 5% difference, the (100) again being deeper. These particular results suggest that the increased growth rate of the oxide in (111) in comparison with (100) is not sufficient to cause the observed increased junction depth of 30%.

b) The oxide growth rate constant was set equal to 1.30×10^{-8} cm/sec for both orientations and D was set equal to 3×10^{-13} cm^2/sec for both orientations in an effort to check the effect of variation of the segregation coefficient which

was set equal to 3, 10 and 100. Figure 3 shows that at the junction depth, 10^{-4} on the ordinate as above; once again the difference is less than 5%, showing that the change in the segregation coefficient is insufficient to produce the observed results. Changes in the segregation coefficient alter the concentration profile near the surface but produce little effect in the interior of the crystal.

c) The next set of boundary conditions undertaken was the combined variation of K and k. Since the (100) specimens have a larger diffused volume after steam oxidation, it may be that fewer boron atoms segregate to the oxide during the oxidation implying that the segregation coefficient would be lower for (100) than (111). The following conditions were computed:

$$(100)\ K = 1.3 \times 10^{-8}\ \text{cm/sec},\ k = 3$$

$$(111)\ K = 1.5 \times 10^{-8}\ \text{cm/sec},\ k = 10$$

Results are shown in Fig. 4 (Δ = (111), x = (100)). Once again it is apparent that the junction depth differs by only about 5%; the (100) is deeper.

d) Another consideration was the possibility of the diffusion coefficient, D, being larger in (100) than (111) for reasons discussed previously. The computation was carried out for $D = 3 \times 10^{-13}\ \text{cm}^2/\text{sec}$, an experimentally observed value, and $D = 3.78 \times 10^{-13}\ \text{cm}^2/\text{sec}$, a 26% increase as would be expected if the rate of vacancy generation was directly proportional to D (see eq 1 and subsequent discussion). This assumes boron to diffuse by a mono-vacancy mechanism. The following values were compared:

1) (100) $K = 1.30 \times 10^{-8}$ cm/sec $k = 3$ $D = 3.78 \times 10^{-13}\ \text{cm}^2/\text{sec}$
   (111) $K = 1.50 \times 10^{-8}$ cm/sec $k = 3$ $D = 3.00 \times 10^{-13}\ \text{cm}^2/\text{sec}$

2) (100) $K = 1.30 \times 10^{-8}$ cm/sec $k = 10$ $D = 3.78 \times 10^{-13}\ \text{cm}^2/\text{sec}$
   (111) $K = 1.50 \times 10^{-8}$ cm/sec $k = 10$ $D = 3.00 \times 10^{-13}\ \text{cm}^2/\text{sec}$

Figure 1 (∇ and 0) and Figure 2 (0 and Δ) show the results of both sets of conditions. It is immediately seen that larger differences are now generated. In both cases (k = 3 and k = 10) the junction depths are about 15 percent deeper in (100) for the 26% change in D. If a direct comparison of the effect of changing D is desired (exclusive of K) this can also be obtained from Figures 1 and 2 (i.e., Fig. 1 (∇ and X), Fig. 2 (0 and $\square$)). The junction depth difference is still 10 to 12%. Figure 4 (Δ and $\square$) shows the combined effect of varying K, k and D. The junction depth for the (100) is increased almost 20%, most of which is attributable to the change in D.

Kendall[20] presents strong evidence that boron may diffuse by a divacancy mechanism rather than by single vacancies. The question which naturally arises is, "How will the junction depth be altered if Parker's[6] excess vacancies are considered to interact with boron by a divacancy mechanism?"

The number of vacancies that are in divacancies is given by the number of vacancies times the probability that there will be a vacancy on any one of the vacancy's four nearest neighbor sites. If there is no interaction between the vacancies, the atom fraction of divacancies in a diamond cubic lattice (N_{V2}) is given by twice the probability squared of a vacancy being on any specific site (N_V), where N_V equals the number of vacancies (n_V) divided by the number (#) of sites.

$$N_{V2} = 2\ N_V^2 \tag{2}$$

See Shewmon[21] for a more complete discussion.

If motion is by divacancy mechanism and a diffusion coefficient is measured, say D, what will be the effect of increased rate of vacancy formation on the divacancy concentration and hence D? Above we considered the diffusion coefficient as being directly proportional to the vacancy concentration. Now suppose it to be directly proportional to the divacancy concentration.

1) no excess generation of vacancies

$$N_{V2} = 2 \left(\frac{n_V}{\#_{\text{sites}}} \right)^2$$

2) excess generation of vacancies as a function of orientation

$$N_{V2} = 2\left[\frac{(n_v + 0.26\, n_v)}{\#_{sites}}\right]^2$$

Thus, the diffusion coefficient (3×10^{-13}) will be increased by a factor of $(1.26)^2$, i.e., to 4.77×10^{-13}. The following values were compared:

(100) $K = 1.30 \times 10^{-8}$ cm/sec $k = 3$ $D = 4.77 \times 10^{-13}$ cm^2/sec
(111) $K = 1.50 \times 10^{-8}$ cm/sec $k = 3$ $D = 3.00 \times 10^{-13}$ cm^2/sec
(100) $K = 1.30 \times 10^{-8}$ cm/sec $k = 10$ $D = 4.77 \times 10^{-13}$ cm^2/sec
(111) $K = 1.50 \times 10^{-8}$ cm/sec $k = 10$ $D = 3.00 \times 10^{-13}$ cm^2/sec

Figure 1 (□ and 0) and Figure 2 (x and Δ) show the results of both sets of conditions. It is evident that junction depths in (100) with respect to (111) are now 30 to 35 percent deeper, in good agreement with experimentally observed.

Since the values of K are experimentally known and do not alter the junction depths appreciably and since the junction depth is relatively insensitive to k (i.e., a change from 3 to 100 only produces about a 2.5% change in the junction depth) it appears, from these computer generated results, that variation of D with orientation is the most likely candidate in explaining the experimental observations. Since D is not expected to vary with orientation in a cubic crystal, the theory of Parker[6] concerning the rate of generation of excess vacancies at a function of orientation in silicon appears attractive as a possible explanation, especially when a divacancy mechanism of boron diffusion is considered.

4. Work Hardening Theory

Very recently, Thai[14] proposed an explanation of anomalous diffusion in semiconductors based on the theory of work-hardening. The analysis was found to be in good agreement with experiment. From the model it is possible to predict the impurity concentration above which anomalous diffusion sets in as well as the enhancement of the diffusion coefficient. He considers creation of vacancies by a dislocation moving in its slip plane passing through a dislocation "forest." The analysis is based on Saada's[22-24] theory of work hardening. If slip is not too difficult, the jog created by the passage of a "tree" will produce an appreciable number of vacancies when the "tree" is attractive. The equations presented are valid in the linear region of the plastic strain-stress curve. Equation (3) gives the number of defects formed per unit volume (v).

$$v = \beta^2 EC^2/6(1-\nu)\,\mu b^3 + v_o \tag{3}$$

where
β = lattice constriction term
E = Young's modulus
C = impurity concentration
ν = Poisson's ratio
μ = shear modulus
b = magnitude of Burgers vectors
v_o = thermal equilibrium vacancy concentration.

Thai[14] rewrites eq (3) and defines the term C_o.

$$v/v_o = (C/C_o)^2 + 1 \tag{4}$$

$$C_o^2 = [6(1-\nu)\,\mu b^3/\beta^2 E]\, v_o \tag{5}$$

The term C_o represents the impurity concentration at which the diffusion coefficient is enhanced by a factor of two, neglecting the effect of internal electric field.

Following Thai, for substitutional diffusion, where the diffusion coefficient (D*) is proportional to the vacancy concentration (K* is a proportional factor)

$$D^* = K^* v$$

$$D^* = (K^* v_o)\ (v/v_o) \tag{6}$$

$$D^* = D'(v/v_o)$$

If one includes field enhancement

$$D' = D\{1 + [1 + (2n_i/C)^2]^{-1/2}\} \tag{7}$$

giving

$$D^*/D = [1 + \{1 + (2n_i/C)^2\}^{-1/2}][(C/C_o)^2 + 1] \tag{8}$$

where n_i is the intrinsic carrier concentration. It is equal to $1.5 \times 10^{19}/cm^2$ for the temperature range under consideration.

Since equations (5) and (8) appear to describe the anomalous diffusion effects which occur during impurity diffusion in semiconductors, it seemed advisable to analyze Thai's[14] equations in an effort to explain the anomalous orientation effects which are observed. The orientation dependent parameters in the above equations are Poisson's ratio, Young's modulus and the shear modulus.

In applying the above theory one must begin with the elastic stiffness constants for silicon,[25] obtain values for these constants at the diffusion temperature (the elasticity temperature coefficient must be known),[26] relate the elastic stiffness compliance constants, transform these latter constants to the desired set of coordinates and finally relate these values to Poisson's ratio, Young's modulus and the shear modulus. The diffusion directions of interest are [111] and [100]. The results of such a calculation (for 1125°C) are:

For diffusion in a [001] direction for an (001) surface

$$E' = 1.16 \times 10^{12}\ \text{dynes/cm}^2$$

$$\nu' = 0.29$$

$$\mu' = 7.47 \times 10^{11}\ \text{dynes/cm}^2$$

For diffusion in a [111] direction for a plane containing this direction; (110):

$$E' = 1.56 \times 10^{12}\ \text{dynes/cm}^2$$

$$\nu' = 0.106$$

$$\mu' = 5.15 \times 10^{11}\ \text{dynes/cm}^2$$

where the prime refers to the high temperature value for the particular coordinate system.

Consider now the number of defects (vacancies) formed per unit volume (Eq 3) and the concentration (C_o) defined previously (Eq 5). The calculation is carried out for an impurity concentration (boron in silicon) of $10^{20}/cm^3$ and for a temperature of 1125°C. The value of β, the lattice constriction term is given as $5.6 \times 10^{-24} cm^3$/atom at room temperature by Prussin[7] who makes the reasonable assumption that it will hold for higher temperatures. In calculating the equilibrium vacancy concentration,

$$v_o = A \exp - E_f/kT \tag{9}$$

a value of 2.3 eV for E_f and 4.6×10^{23} for A is used.[12] The Burger's vector is taken as $\sqrt{2}a/2$ where a is the lattice parameter.

The results of the calculation are:

for (001) surface [001] direction.

$v = 3.54 \times 10^{15}/cm^3$ for a conc. of $10^{20}/cm^3$

$C_o = 8.68 \times 10^{19}/cm^3$

from equation (8)

$\frac{D^*}{D} = 4.57$ for a boron concentration of $10^{20}/cm^3$

$\frac{D^*}{D} = 1.33$ for a boron concentration of $10^{19}/cm^3$

for (110) [111]

$v = 4.64 \times 10^{15}/cm^3$ for a conc. of $10^{20}/cm^3$

$C_o = 6.96 \times 10^{19}/cm^3$

$\frac{D^*}{D} = 6.02$ for a boron concentration of $10^{20}/cm^3$

$\frac{D^*}{D} = 1.35$ for a boron concentration of $10^{19}/cm^3$

The contribution to enhanced diffusion by plastic deformation has virtually disappeared for a boron concentration of $10^{19}/cm^3$. The enhancement at this concentration is almost entirely contributed by the field effect term.

A comparison of the values obtained for the number of defects (vacancies) formed per unit volume as a function of orientation shows that more vacancies are formed in the case of [111] diffusion than [100], the values being $4.64 \times 10^{15}/cm^3$ and $3.54 \times 10^{15}/cm^3$ respectively for a boron concentration of $10^{20}/cm^3$. For a boron concentration of $10^{19}/cm^3$, the value is equal to the number of thermally generated vacancies for both orientations, indicating that there should be no enhancement of diffusion attributable to plastic deformation at this concentration. The values obtained for C_o, the concentration at which plastic deformation enhances the diffusion coefficient by a factor of two, are $6.96 \times 10^{19}/cm^3$ for [111] and $8.68 \times 10^{19}/cm$ for [100]. Thus, a higher solute concentration is necessary to cause the enhancement in the case of [100].

Carrying the calculations one step further it is seen that, at a concentration of $10^{20}/cm^3$, the diffusion coefficient is enhanced by a factor of 6.0 for [111] and 4.6 for [100]. If one eliminates the field effect contribution the values become 3.1 and 2.3 respectively. For a concentration of $10^{19}/cm^3$ the enhancement is attributed almost entirely to the field effect.

Generally, the boron surface concentration values obtained from a pre-depth carried out in the 870°C to 925°C range are of the order of 10^{19}-$10^{20}/cm^3$. Following a short high-temperature drive-in (i.e., 1150°C for 20 minutes) the surface concentration typically decreases somewhat (about half an order of magnitude). A steam oxidation performed at 1050°C for 120 minutes would lower the surface concentration even further. The pre-depth concentration is in the range in which plastic deformation effects begin to take place. However, little diffusion takes place during this step. The maximum boron concentration during drive-in and steam oxidation (for the temperature considered above) is below the level at which enhanced diffusion due to plastic deformation would be expected to occur. If the concentration levels were an order of magnitude higher, the above theory shows that diffusion would be enhanced, the enhancement being greater for [111] diffusion than [100] diffusion.

5. Conclusions

An analysis of the diffusion equations indicated that a change in the oxide growth rate and/or a segregation coefficient can result in a small shift in the diffusion profile. The change in junction depth is not sufficiently large to account for the experimentally observed value. If boron is considered to diffuse by a monovacancy mechanism, excess vacancies as generated by a solute dislocation interaction (which is a function of orientation) can account for about one-half of the observed shift. A divacancy mechanism, for which there is strong support, is shown to fully account for the shift.

It is concluded that enhanced boron diffusion by a divacancy mechanism is the most likely candidate of the parameters studied in explaining the orientation dependence of the junction depth.

Any enhancement due to plastic deformation based on the creation of vacancies by a dislocation-dislocation forest interaction would be expected to occur at high concentrations and would favor the [111] direction over the [100].

6. References

[1] A. F. Willoughby, J. Matl. Sci. 3, 89 (1968).

[2] S. Maekawa and T. Oshida, J. Phys. Soc. Japan 19, 253 (1964).

[3] K. H. Nicholas, Solid State Electron 9, 35 (1966).

[4] T. J. Parker, J. Appl. Phys. 38, 3471 (1967).

[5] T. J. Parker, J. Appl. Phys. 38, 3475 (1967).

[6] T. J. Parker, J. Appl. Phys. 39, 2043 (1968).

[7] S. Prussin, J. Appl. Phys. 32, 1876 (1961).

[8] R. W. Balluffi and A. L. Ruoff, J. Appl. Phys. 34, 1634 (1963).

[9] A. L. Ruoff and R. W. Balluffi, J. Appl. Phys. 34, 1848 (1963).

[10] A. L. Ruoff and R. W. Balluffi, J. Appl. Phys. 34, 2862 (1963).

[11] K. E. Benson, C. L. Paulnack, private communication.

[12] G. N. Wills, Solid State Electron 12, 133 (1969).

[13] R. A. Kovalev, V. B. Bernikov, Yu. I. Pashintsev and V. A. Marasanov, Soviet Physics - Solid State, vol 11, No. 7 1571 (1970).

[14] N. D. Thai, Solid State Electron, 13, 165 (1970).

[15] T. Kato and Y. Nishi, Japan J. Appl. Phys. 3, 377 (1964).

[16] K. S. Cave, Solid State Electron, 8, 991 (1965).

[17] K. S. Cave, Solid State Electron, 9, 910 (1966).

[18] B. E. Deal and A. S. Grove, J. Appl. Phys. 36, 3770 (1965).

[19] A. S. Grove, O. Leistiko and C. T. Sah, J. Appl. Phys. 35, 2695 (1964)

[20] D. Kendall, Article in "Semiconductor Silicon," Edited by R. Haberecht and E. L. Kern, The Electrochem. Soc., N. Y. (1969).

[21] P. G. Shewmon, Diffusion in Solids, p. 78, McGraw-Hill, New York (1963).

[22] G. Saada, Acta Met. 8, 200 (1960).

[23] G. Saada, Acta Met. 8, 841 (1960).

[24] G. Saada, Acta Met. 9, 965 (1961).

[25] A. Davies, Met. Rev., 10, 38 (1965).

[26] H. J. McSkimin, W. L. Bond, E. Buehler and G. K. Teal, Phys. Rev. 83, 1080 (1951).

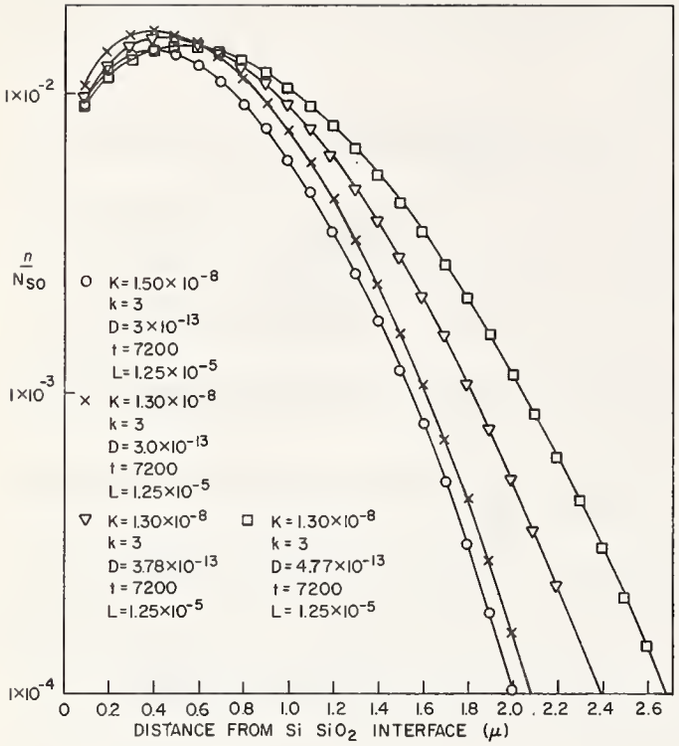


Fig. 1. Diffusion Profile; Segregation coefficient held constant at 3; Comparison of oxide growth rate constants; comparison of diffusion coefficients.

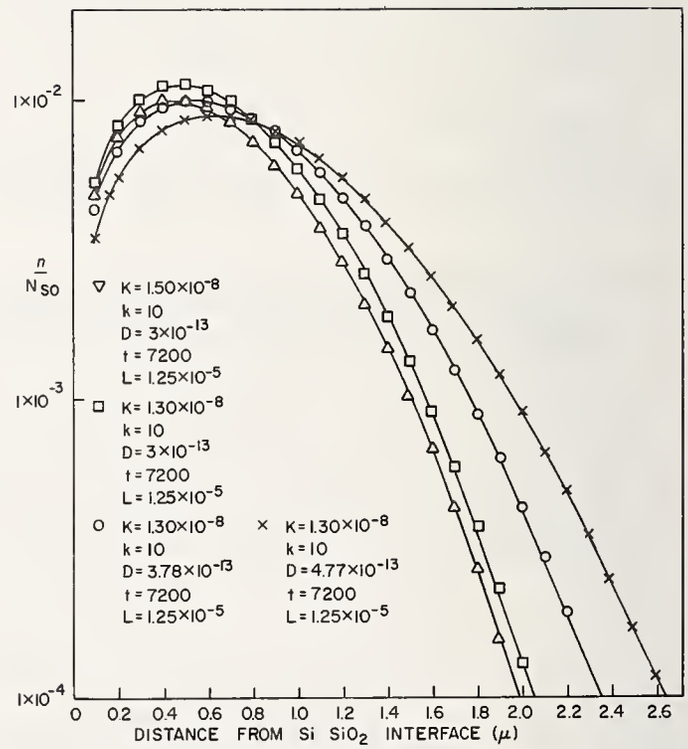


Fig. 2. Diffusion Profile; Segregation coefficient held constant at 10; Comparison of oxide growth rate constants; comparison of diffusion coefficients.

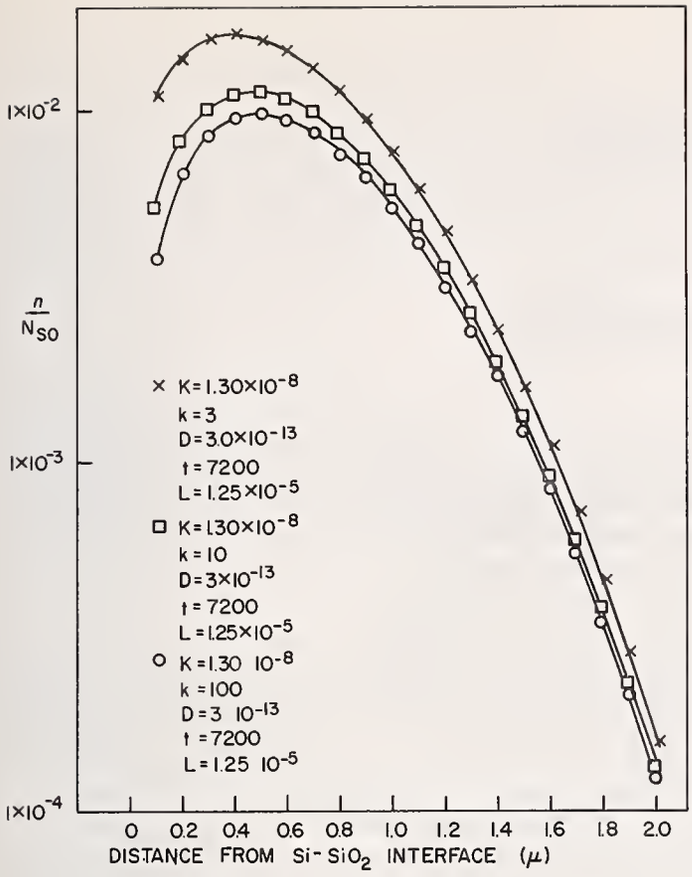


Fig. 3. Diffusion profile for segregation coefficient equal to 3, 10, and 100.

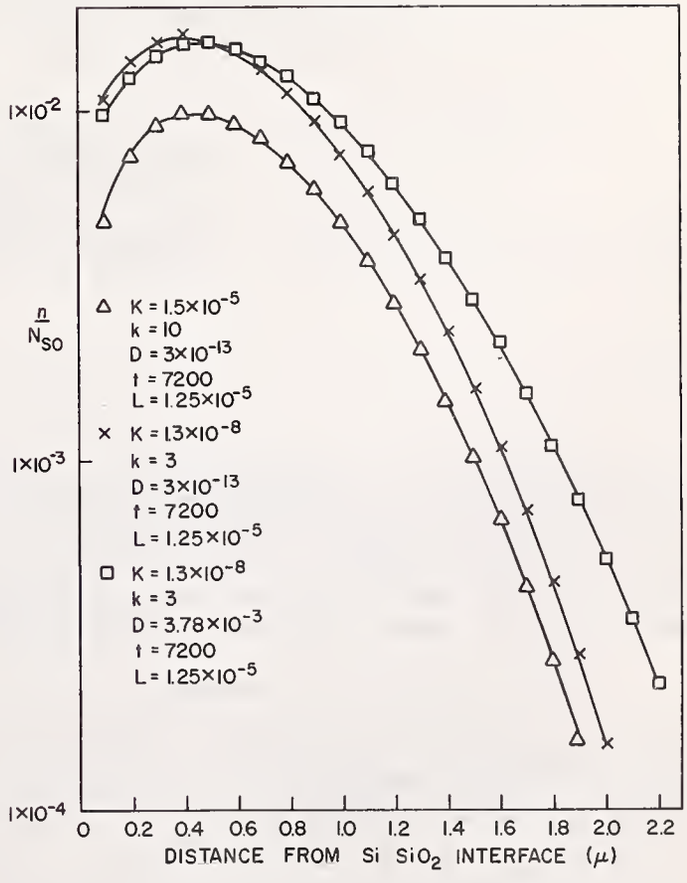


Fig. 4. Diffusion profile; Comparison of oxide growth rate constants, segregation coefficients and diffusion coefficients.

Diffusion-Induced Defects and Diffusion Kinetics in Silicon

/S. Dash* and M. L. Joshi

IBM Corporation, Components Division
Essex Junction, Vermont 05452

Diffusion of phosphorus and boron is known to introduce plastic deformation in the form of misfit dislocations in silicon. Our studies show that the initial stage of diffusion-induced plastic deformation is associated with generation of sessile dislocation loops of interstitial nature. Misfit dislocation nets are formed through interaction of these loops during diffusion. Phosphorus impurity profiles in silicon are observed to follow ideal distribution functions in the absence of dislocations, even when surface concentration C_s of phosphorus reaches solid solubility limit. Deviation of both impurity profiles and dislocation generation depend upon the critical value of integrated doping of phosphorus viz. 1.2×10^{15} atoms/cm^2, and not upon the value of C_s alone. Arsenic diffusion also introduces sessile dislocation loops of interstitial nature. These loops, however, do not interact to produce misfit dislocation nets; they produce dislocation tangles. Introduction of interstitial loops during diffusion lends support to the theory that diffusion in silicon takes place via extended interstitials and not via single or double vacancies.

Key Words: Arsenic diffusion, diffusion mechanism, impurity profiles, interstitial loops, phosphorus diffusion, plastic deformation, silicon dislocation, materials.

1. Introduction

The phenomena of diffusion-induced slip and precipitation in silicon have been extensively studied and discussed in literature. [1-9][1] A large number of facts have been obtained; a coherent picture accounting all these, however, has not yet emerged. Besides, there seems to be a lack of serious effort in correlating the diffusion-induced defects and the impurity profiles. Shape of an impurity distribution is expected to deviate from ideality when the impurities are trapped at the defects. [10-12] We have attempted to present a coherent picture relating defects induced by diffusion of phosphorus and its distribution profile.

Diffusions of impurities belonging to Groups III and V into silicon is usually assumed to take place by means of vacancy mechanism similar to that in metals. A recent comprehensive analysis by Seeger and Chik has cast serious doubts on the vacancy mechanism of diffusion in both Si and Ge. [13] An electron microscopic study of point defects induced through diffusion of arsenic and phosphorus in silicon is presented to strengthen the arguments of Seeger and Chik.

* Present address: Fairchild Semiconductor, Mountain View, California 94040

[1] Figures in brackets indicate the literature references at the end of this paper.

2. Theoretical

Diffusion of phosphorus and boron into silicon single-crystal wafers is known to cause generation of misfit dislocation networks. [2,4,7] Most of these dislocations lie on the diffusion plane and have edge character. Dislocation distribution studies in phosphorus-diffused silicon indicate that the density of these dislocations becomes negligibly small at depths greater than one-third the junction depth. [7] One of the early and the significant models relating the impurity distribution and the associated strain with the dislocation generation is due to Prussin. [2] For constant C_s, the strain ϵ_x according to Prussin is:

$$\epsilon_x = \frac{\beta C_s}{1-\nu}\left[\operatorname{erfc}\frac{d-z}{(4Dt)^{\frac{1}{2}}} - \frac{2(\pi Dt)^{\frac{1}{2}}}{d\pi}\right] \tag{1}$$

where the diffusion plane, with z-direction as the normal, contains x- and y-directions; β is the solute lattice contact coefficient; 2d is the thickness of the wafer; C_s is surface concentration; ν is the Poisson's ratio; D is the diffusion coefficient of the impurity and t is the time of distribution of impurities. In the Prussin model, the maximum stress, which is generated at t=0, is

$$\sigma_{max} = \frac{\beta C_s E}{1-\nu} \tag{2}$$

where E is the Young's modulus, and the dislocation density ρ at any time is distributed in the diffusion zone according to the relation:

$$\rho = \frac{\beta}{\alpha}\frac{dc}{dz} \tag{3}$$

where α is the x-component of the Burgers vector of the edge dislocation introduced.

The Prussin model assumes that the dislocation generation spontaneously starts at t=0. The total number of dislocations N per unit length of diffused layer is determined to be:

$$N = \int \rho d\alpha \simeq \left(\frac{\beta}{\alpha}\right) C_s \quad . \tag{4}$$

These dislocations are later dispersed into depth as diffusion of impurities proceeds. We will present experimental evidence in the case of phosphorus diffusion in silicon to show that the model is not valid.

Shockley has shown that the dislocation generation is dependent not only on the surface concentration but also on the total quantity of solute atoms. [1] The min number of atoms, Q_{min} necessary for slipping dislocations to a depth r in silicon, is given by:

$$Q_{min} \simeq \left\{\frac{N_{Si} b}{\delta}\frac{\ell n\,(2\sqrt{2r/r_o})}{4\pi\sqrt{2}(1+\nu)}\right\} \quad . \tag{5}$$

Where N_{Si} = number of silicon atoms/cc $\simeq 5 \times 10^{22}$ atoms/cc; b is the Burgers vector or individual dislocations $\simeq 3.84$ Å; δ is the fractional reduction in lattice parameter per atomic fraction of solute, ν is the Poisson's ratio = 0.27 and r_o is the effective core radius of dislocations $\simeq 1.92$ Å. Assuming for phosphorus diffusion

$$\delta = \frac{\text{difference in the atomic radii of Si and P}}{\text{atomic radius of Si}} = 6 \times 10^{-2}$$

and the constraint $r = r_o$, it is found that

$$Q_{min} \simeq 24 N_{Si} b \times 10^{-4} \simeq 5.53 \times 10^{12} \text{ atoms/cm}^2 \tag{6}$$

is the critical concentration below which no dislocation should be generated. (The assumption $r = r_o$ only means that no dislocations are introduced from the surface.) For the case of diffusion into intrinsic silicon from a source giving constant surface concentration (in such a case the ideal diffusion profile is a complementary error function), we can show that

$$Q \simeq \frac{1}{3\sqrt{\pi}} C_s x_j \tag{7}$$

where C_S is the surface concentration in atoms/cc and x_j is the junction depth. With $C_S = 1.2 \times 10^{21}$ atoms/cc (the phosphorus solubility limit at 1000 °C) by the use of eqs. (6) and (7), we get the critical value of $x_j \simeq 500$ Å for phosphorus diffusion. Hence, as soon as the junction depth becomes 500 Å, the edge dislocations are supposed to enter the crystal by slip. Shockley's model is obviously simplistic. The Burgers vectors of the misfit dislocations lie in the plane of the silicon wafer surface and the motion of the dislocations from the surface into the bulk is known to be not by the simple slip process assumed in Shockley's model. [4,5] Consequently, the critical value of x_j will be quite different than 500 Å.

A more rigorous model of interfacial misfit dislocations has been advanced by Van Der Merwe for epitaxy and the diffusion zone of a bicrystal system. In principle, the model is rigorously applicable even to a single crystal in which diffusion of odd-sized impurity creates a top layer with different lattice parameters than that of the substrate. Starting with a Lenard-Jones potential and assuming a definite bonding of the atoms of the top layer with the substrate atoms, calculations were made by Van Der Merwe for dislocation density, residual stress, interfacial energy, and the thickness dependence of the critical misfit before there is a spontaneous generation of a dislocation network. For a small misfit, Van Der Merwe derived an equation for a single-crystal top layer relating the ratio h/a (the thickness of the top layer in units of lattice parameters) with δ_c, the critical misfit. This equation, eq. (25) of ref. 14, reduces to:

$$\ell n \frac{2\pi e \delta_c}{(1-\nu)} + \frac{4\pi (1-\nu)^2 h \delta_c}{(1-2\nu)a} = 0 \tag{8}$$

for the case of diffusion-induced strain in single-crystal silicon. Assuming $\nu = 0.27$, we obtain

$$\ell n\, 23.4\delta_c + 14.6\frac{h}{a}\delta_c = 0 \ . \tag{9}$$

Below the critical misfit, according to the model, the diffusion layer should conform to the substrate crystal through elastic deformation. For misfits greater than δ_c interfacial dislocations are created between the substrate and the top layer. As will be seen later, our experimental observations are found to correlate very well with the Van Der Merwe model.

For large misfits, the energy is not totally relieved by interfacial dislocations. A small amount of residual strain is left behind. [15] When the top layer of uniform lattice parameter is replaced by a diffusion zone with varying lattice parameter across the thickness of the film, the entire thickness could be divided into subzones with subinterfaces instead of one discrete interface.

Total energy of the diffusion zone then is the sum of the dislocation grid energy and a linear elastic deformation energy. The sum of the dislocation densities in subinterfaces will be equal to the dislocation density expected on a discrete interface. The large misfits of phosphorus in silicon is easily detected. Diffusion of large amounts of phosphorus in Si is known to introduce dislocations in the subinterfaces with a distribution such that the dislocation density decreases sharply towards the diffusion front.

Van Der Merwe's calculations do not take into account the effect of climb forces due to point defects, which are expected to be generated inside a dislocation-free material as a result of unequal diffusivities of the two-atom species as in a Kirkendall couple. Hirth has considered this effect. [16] The net climb force due to solute atoms can be written as

$$F = \frac{bkT}{\Omega} \ell n \frac{C_v}{C_{ve}} + \sigma b \tag{10}$$

where Ω is the atomic volume, C_V and C_{Ve} are the real vacancy concentration and the equilibrium vacancy concentration. The net stress, due to the shift of the Matano interface at the dislocation interface is given by σ. We will not be able to verify this equation. We will give evidence of the existence of climb forces during diffusion. However, the climb forces will be shown in silicon (diffused with high amounts of P and As) to be due to excess interstitials and not due to excess vacancies.

Thus far, we have discussed various models that relate these parameters: atomic critical misfit, min thickness of diffusion zone, and the distribution of induced dislocations. We present experimental results that take the above parameters into account and a detailed account of the actual mechanisms of the initiation of the misfit-dislocation networks. We will also show that, for phosphorus diffusion, there is a critical junction depth below which no misfit dislocations are generated. We show that phosphorus impurity distributions obey expected ideal functions when the diffusion depth is below a critical value, which invalidates the widely accepted dependence of the diffusivity of phosphorus on the concentration for values above 5×10^{19} atoms/cc. [10,17] Finally, defect structure induced by diffusion of arsenic, the covalent atomic size of which is about the same as that of silicon, will be compared with that induced by diffusion of phosphorus. This comparison is made to illustrate the effect of small atomic misfits.

3. Experimental

High resistivity (111) and (100) p-type silicon wafers, $\approx$ 8 mil thick and free of damage, were diffused with phosphorus and arsenic by the standard open-tube gaseous diffusion and capsule diffusion processes. [17] The aim of the experiment was to obtain various phosphorus diffusion depths in silicon so that we could follow the earliest stages of generation of dislocations, point defects and their clustering, and precipitates. For phosphorus diffusion, $POCl_3$, and phosphorus-rich powder silicon sources were used as sources. In the case of arsenic diffusion, the capsule process was followed, using arsenic-rich powder silicon sources. The diffusion temperature ranged from 900 to 1200 ^{o}C.

Some open-tube phosphorus-diffused samples were quenched in mineral oil to retain the equilibrium impurity concentration existing at high temperature. Sheet resistances were measured to see the effect of quenching on the impurity concentration in solid solution. Annealing experiments were done on these samples and also normal-cooled diffused samples to follow the growth kinetics of the diffusion-induced defects. Diffused samples were annealed in N_2 atmosphere in the temperature range of 500 to 900 ^{o}C. Time of anneal was larger for lower annealing temperature.

Impurity profiles of samples were determined only in the case of diffusion of phosphorus by both the electrical and neutron activation techniques. [10] In the electrical technique, four-point probe sheet-resistance measurements were taken on silicon after removal of successive layers of silicon through anodization and HF etching. The data was analyzed in the same manner as that by Tannenbaum. [10] Specimens of appropriate sizes were cut, chemically thinned, and studied using a Philips EM-300 transmission electron microscope (TEM) for diffusion-induced defects.

4. Results

4.1. Phosphorus Open-Tube Diffusion

a. Diffusion Profiles

Phosphorus was diffused in p-type intrinsic silicon wafers at 900 ^{o}C from a $POCl_3/O_2$ source in an open-tube furnace using argon as the carrier gas. The low temperature of 900 ^{o}C was chosen to obtain a reasonable amount of process time for reproducible diffusions in the shallow junction range of 1000 to 5000 Å. In Fig. 1, we show four phosphorus impurity distributions: A, B, C, and D with junction depths of 2400 Å, 3750 Å, 4250 Å and 5150 Å; and diffusion time of 10 min, 25 min, 30 min, and 35 min, respectively.

Impurity distributions A and B were both confirmed to obey complementary error functions after observing the linear variation of the plot of $C/2C_s$ against the position coordinate x on a probability paper. Distribution C and D deviated considerably from the complementary error functions and showed the anomalous variation of the phenomological diffusion coefficient D with the concentration. This anomalous variation is in accordance with the well-established results of Tannenbaum.

The error-function distribution distribution of the curves A and B imply that the constant source condition at the phosphosilicate glass/silicon interface is reached almost immediately at the start of the diffusion process. The solution of the Fick's equations of diffusion for a semi-infinite body with the condition of a constant concentration C_s at the surface as given in standard textbooks is

$$C(x,t) = C_s \operatorname{erfc} x/2\sqrt{Dt} \quad . \tag{11}$$

A p-n junction is formed in a semiconductor at the point where the donor impurity becomes equal to the acceptor impurity. Using this condition, it is easy to derive from eq. (11)

$$x_j = 2\sqrt{Dt}\ \mathrm{erfc}^{-1}\frac{C_a}{C_s} \tag{12}$$

where C_a is the acceptor concentration. D is calculated to be $= 2.4 \times 10^{-10}$ cm^2/sec from eq. (12) for both curves A and B.

Phosphorus D vs 1/T data due to I. M. Mackintosh and also calculated diffusivity from our A and B erfc distributions are shown in Fig. 2. [18] The good agreement, depicted in Fig. 2, may be considered fortuitous on the basis that Mackintosh's data includes effective diffusivities calculated from some anomalous distributions, along with the true diffusivities obtained from ideal distributions and calculated from low temperature and relatively long-time diffusions. The effective diffusivities were calculated by Mackintosh from the data of sheet resistance and junction depth assuming ideal impurity distributions. Tannenbaum has shown that within experimental error, the diffusion coefficient is constant at phosphorus concentrations below 10^{20} atoms/cc. Above that concentration, she shows that the diffusion coefficient rises very rapidly. The low-concentration limits of the diffusion coefficients shown in Fig. 2 were calculated by Tannenbaum. These values agree with those of Mackintosh within diffusion control limits. In the light of these facts, the rapid increase of true diffusion coefficient above phosphorus concentration $\simeq 10^{20}$ atoms/cc should be held in doubt.

A comparison of distributions A and D in Fig. 1 (the four-point probe impurity distributions) is made with A_n and D_n in Fig. 3 (the neutron activation profiles) for the same diffusions. The difference between the impurity distributions A and A_n obtained is small. The difference, however, becomes larger with the diffusion penetration. The reason for wide differences between A and A_n in the junction region was found attributable to the limitations of the method of neutron activation. The neutron activation technique is expected to displace a certain small fraction of the impurity in the surface region towards the junction. A low-temperature annealing (300 °C) treatment is invariably used to restore the original resistivity of the neutron-activated samples. Therefore, in a shallow diffusion, a junction is expected to get smeared after annealing of the neutron-activated samples.

Tannenbaum has interpreted the significant differences between the neutron activation and the four-point probe profiles as the amount of impurity, which is precipitated or out of substitutional sites. We also favor the same interpretation. The four-point probe profile is expected to give the distribution of the impurities that contribute to electrical activity. The neutron activation profile, on the other hand, is expected to give the distribution of the impurities both electrically active and inactive.

b. Dislocation Generation

Silicon samples of (111) orientation after diffusions corresponding to the impurity distributions A, B, C, D given in Fig. 1 were cut, chemically thinned from one side, and examined through a transmission electron microscope (TEM) for diffusion-induced defects. No misfit-dislocation grids could be observed for samples of A-and B-type. Samples of C-type showed a few patches of dislocation nodes along with interacting dislocation loops, as shown in Fig. 4(a). Samples of D-type showed extensive generation of misfit-dislocation nets, as shown in Fig. 4(b) for (111) silicon and in Fig. 4(c) for (110) silicon. The dislocation nets are seen distributed and interconnected in 3-D in different layers of (110) silicon as seen in Fig. 4(c). Why (110) silicon gives rise to the 3-D networks is not clear. Many researchers have previously analyzed the dislocations to be mainly edge-type in character and to possess Burgers vectors of the individual dislocations to lie in the plane of the silicon surface as seen in Fig. 4(b). All these results indicate that for high C_s phosphorus diffusions at 900 °C, the dislocation generation begins at a critical depth of approximately 3750 Å. Silicon samples diffused at higher temperatures (970 and 1050 °C) also gave about the same value of critical x_j indicating that the temperature influence was not significant in determining the value of x_j, which is the point at which diffusion induces dislocations in silicon. Detailed dislocation morphologies are illustrated in Fig. 5(a), (b), and (c) for silicon samples having C-type diffusion. We observed in these figures nodes, half-loops, loops, interacting dislocations, stacking faults bounded by half-loops, and precipitates. Very thin Si-P precipitates are expected to give rise to fringe contrasts. This point is illustrated in Fig. 5(b) and is in accordance with the observations on thin precipitates by Levine et al. [5] A precipitate platelet causing dislocation generation has been illustrated in Fig. 5(a). This precipitate platelet is observed to intersect (111) silicon surface in a ⟨220⟩ direction. The platelet, therefore, is expected to lie on inclined {111} planes. Phosphorus precipitate platelets lying on inclined {111} planes have been extensively analyzed previously in thin-silicon foils by the present authors. [19] The association of phosphorus precipitate platelets and stacking faults has been well illustrated in that work. The stacking faults observed in Fig. 5(c) are expected to accumulate

phosphorus atoms and produce thin precipitates. Figures 4(a) and 5(a), (b), and (c) clearly illustrate that misfit dislocations do not start uniformly all over the surface. In localized areas, misfit dislocations are nucleated in stages.

Dislocation loops were observed (Fig. 6) in much larger densities in C-type samples gone through annealing at 800 °C for 2 h in nitrogen. This process of generation of dislocation loops is indicative of condensation of point defects, the nature of which will be discussed in a later section.

4.2. Phosphorus Capsule Diffusion

A representative phosphorus distribution in a (100) silicon processed through capsule diffusion at 1050 °C for 90 min, using a high-concentration phosphorus source, is shown in Fig. 7(a). In Fig. 7(b), we have a TEM view of the corresponding defect structure at magnification of 120,000. Very small dislocation loops in large densities are observed. After annealing the diffusion samples at 500 °C or higher, the dislocation loops grew. A typical view of the ⟨100⟩ diffused silicon after annealing at 700 °C for 2 h is shown in Fig. 8(a). Following the same annealing treatment, a (111) diffused silicon wafer is observed to produce dislocation loop structure as shown in Fig. 8(b). In both these pictures, we observe thin precipitates and loops lying along ⟨220⟩ directions and loop interactions producing dislocation network. The growth process of the dislocations has to be associated with condensation of point defects during the annealing treatment.

In Fig. 9(c), we have shown a stereographic projection of the loop planes corresponding to TEM view in Fig. 9(a) of dislocation loops observed in a (111) silicon. It will be seen that the most stable configuration of the dislocation loops tends to be confined to the (101) planes inclined 90° to the (111) diffusion surface. Many small dislocation loops seen in Fig. 8(a) and (b) have their long axes oriented along directions parallel to the traces of precipitate platelets on {111}, which are ⟨110⟩ type. This observation implies that the small loops originate on the {111} planes inclined to the (111) or the (100) diffusion surface and tend to move to the planes perpendicular to this surface during annealing.

The dislocation loops were identified to be interstitial in nature through the tilting method described by Amelinckx. [20] Later, we will illustrate the method in detail in Appendix I in the case of arsenic diffusion-induced loops, which will also be identified as interstitial in nature. The Burgers vector of the loops was determined to be a/2 ⟨110⟩ lying on the (111) foil plane as can be seen from Fig. 9(b) where the loops parallel to the ⟨224⟩ diffraction vector have vanished. The loops lying on the inclined {110} planes must, therefore, be perfect prismatic loops.

Physically, the loops lying on the inclined {110} planes amount to discs of extra silicon-atom planes inserted from the ⟨111⟩ diffusion surface. From the point of view of relief of strains, the misfit dislocation nets are equivalent to the dislocation loops with their Burgers vector in the plane of diffusion. The point defect density calculated from size and density of loops varied from 10^{19} to 10^{20}/cc. The loops were found within a max depth of 1μ from the surface for diffusion depth $x_j \simeq 3\mu$.

Two curves for dislocation-loop size (long-axis length) versus annealing time were obtained for 600 and 800 °C annealing temperatures. The activation energy responsible for the loop growth was calculated from the slope of the growth rate of the loops versus 1/T(°K) curve and was found to be ≈ 2.1 eV. In the following section, similar results on the kinetics of loop growth will be presented in considerable details.

4.3. Arsenic Capsule Diffusion

High-concentration diffusion of arsenic was obtained in the (111) silicon wafers via capsule diffusion technique using arsenic-rich powder silicon sources at 1100 °C.

Many as-diffused samples had very fine dislocation loops, as seen in Fig. 10(a). Upon annealing, these loops grew as the temperature and the time of anneal were raised. The proof that the loops are interstitial is given in Appendix I. Small dislocation loops in arsenic-diffused samples are seen hexagonal in shape, as shown in Fig. 10(b), with edges parallel to ⟨110⟩. The habit plane, therefore, for hexagonal loops are {111}. Long-time and high-temperature annealing causes the loops to enlarge and lose hexagonal shape through a combined process of climb and glide, and habit planes no longer remain {111}, except for those loops lying on the (111) wafer surface. After a certain annealing time, the dislocation loops break up, interact among themselves, and form tangles and nodes, as seen in Fig. 10(d). During annealing of the arsenic-diffused samples in the temperature range of 500 to 900 °C, no discreet precipitation of any second phase was observed. The effects of a typical 900 °C annealing sequence on loop growth is shown in Fig. 10(a), (b), (c), and (d).

Sometimes, as shown later in Fig. 18, many dislocation loops contained stacking faults with displacement vector 1/3 a ⟨111⟩.

The point defect density was roughly calculated to be 10^{19}/cc from the area of the loops in Fig. 10(a). The Burgers vectors of the inclined loops, which are prismatic, were found to lie in one of the three ⟨110⟩ directions corresponding to the inclined {111} diffusion planes. The dislocation loops lying on the (111) diffusion surface have their Burgers vectors on the three inclined {111} planes. Thus, arsenic diffusion induces dislocation loops whose slip vectors do not lie on the diffusion plane. This situation is contrary to the case of phosphorus diffusion-induced loops. A schematic of both situations is presented in Fig. 11.

Annealing of arsenic-induced loops caused the growth of the loops lying mostly near the diffusion surface. Those loops which lie deeper than 4000 Å grow extremely slowly as illustrated in Fig. 9. A detailed explanation is later given in the discussion section. The loops near the surface lie in the high concentration region of the diffusion gradient. The growth of the sessile extrinsic loops is through accumulation of self-diffusing silicon atoms. Consequently, we should be able to obtain from the growth kinetics of these loops significant information about the effects of dopant concentration on the activation energy of self-diffusion of silicon atoms.

Longer periods of annealing causes the planar loops to interact and produce dislocation tangles. Frank loops on inclined planes, however, grow until they intersect the free surface where they become extrinsic stacking faults bounded by a single partial below the surface. Annealing growth of prismatic dislocation loops in the vicinity of silicon surface is presented in Fig. 12, and a plot of growth rate vs. $1/T\,^{o}K$ is shown in Fig. 13. We get activation energy of 2.12 eV for loop growth from the slope of this curve. In another similar experiment with a wafer with lower surface concentration of arsenic, an activation energy of 2.29 eV was obtained. A rough estimate for the activation energy of the slow-growing loops lying deeper than 4000 Å was also made and was found to lie somewhere between the values 4.5 and 5.5 eV. We learn that higher the concentration of the ambient arsenic, lower the activation energy of self-diffusion of Si.

Dislocation loops in silicon have very complex morphological characteristics, depending on the composition, diffusion temperature, and annealing atmosphere used. Most of the arsenic-induced loops on {111} generally have their long axes along ⟨110⟩ directions. Multiple generation of dislocation loops inside already existing loops have been occasionally observed indicating a Herring-Bardeen source at the center. Steam oxidation at 1000 °C accentuates this situation as seen in Fig. 14. Here we observe various loop geometries: elongated perfect prismatic loops, hexagonal Frank loops, irregular prismatic loops [on (111) diffusion surface], and loops within loops.

In some arsenic-diffused samples, parallel moiré fringes with a fringe spacing of 22 Å were observed in hexagonal loops occasionally as depicted in Fig. 14(b). Similar moiré fringes were also sometimes observed in P-diffused samples. Since parallel moiré fringes are caused by two thin layers of material of slightly differing lattice parameter, we concluded that there was clustering of impurity atoms inside the loop.

5. Summary of Experimental Observations

Dislocation generation starts at x_j = 3750 Å and $C_S = 6 \times 10^{20}$/cc in silicon undergoing diffusion in an open-tube system where phosphosilicate glass provides a constant impurity concentration at the surface. Dislocation generation starts in localized areas in the form of half-loops, threefold nodes, and patchy misfit networks. Precipitates and dislocations are often associated with each other. In very shallow phosphorus diffusions, low-temperature annealing is observed to bring out interstitial dislocation loops.

Phosphorus diffusion coefficient is not dependent on concentration in high-concentration diffusion shallower than x_j = 4000 Å. It will depend on concentration as diffusion penetrates deeper than this value of x_j because of the generation of dislocation loops and networks.

Phosphorus impurity profiles measured through the four-point probe method and the neutron-activation method start diverging from each other considerably at the point when dislocations begin to be generated through diffusion. Faster cooling of diffused samples helps retain more electrically active concentration of phosphorus. Both observations indicate that the dislocations help inactivate electrically active impurities.

Phosphorus capsule diffusions produce interstitial dislocation loops on {111} planes inclined to the (111) or (100) diffusion surface. The Burgers vector of these loops is a/2 ⟨110⟩ and lies on the diffusion plane. After annealing for a certain time, the loops tend to orient themselves into a more stable configuration on the {110} planes perpendicular to the (111) diffusion surface and become perfect prismatic loops. The loops during annealing interact and produce misfit dislocation nets parallel to the diffusion surface. Thin precipitates of phosphorus are also observed to be generated during annealing.

Arsenic capsule diffusions also produce interstitial dislocation loops. Their Burgers vectors, however, do not lie on the diffusion plane. A considerable number of these loops contain stacking faults. After long-time annealing, the dislocation loops interact to form dislocation tangles. However, no misfit dislocation nets are formed. Occasionally, dislocation loops are observed to accumulate impurities. Discreet As-Si precipitates, however, are not observed.

From the growth kinetics of dislocations loops either in arsenic-diffused or phosphorus-diffused silicon, the activation energy for self-diffusion of Si is found to increase from 2.12 eV near the surface to the mean value of about 5 eV in the vicinity of the junction.

6. Discussion

6.1. Misfit Dislocations and the Various Theories

In the case of phosphorus open-tube diffusion, one of the major results of our investigation is that generation of misfit dislocations is associated with considerable deviation from the ideally expected distribution function. Furthermore, we also know that concentration dependence of diffusivity and also dislocations are not observed in diffusions shallower than 3750 Å. McDonald et al. [12] have previously shown that when the average sheet conductivity of silicon diffused in a $POCl_3$ open-tube furnace is below 1.6×10^{-3} (ohm-cm)$^{-1}$ density of slip patterns is that neither the high value of C_s nor that of ($x_j > 1\,\mu$) are ideally expected complementary functions. [12] Based upon these facts, we can conclude that neither the high value of C_s nor that of x_j alone is responsible for the deviations of impurity profiles from ideal ones. The important factor for the deviations is the dislocation generation, which depends upon the total amount of diffused impurity and its atomic misfit. Using eq. (7) with the knowledge of critical value of x_j for dislocation generation to be $\approx$ 4000 Å for $C_s = 1.5 \times 10^{21}$ atoms/cc, we obtained the total amount of impurity, $Q_{crit} \simeq 1.1 \times 10^{15}$ atoms/cm^2 for shallow phosphorus diffusions. Using the data of McDonald et al. for deep diffusions, $Q_{crit} \simeq 1.2 \times 10^{15}$ atoms/cm^2. In the light of process and measurement limits of reproducibility, the values of Q_{crit} are in excellent agreement. Thus, we see that neither C_s nor x_j alone as implied in the models by Prussin and Czaja is responsible for diffusion-induced dislocations. [8] It is Q that is important. However, the Queisser-Shockley model, as pointed out in our introduction, does indicate this; but it cannot predict the critical value of x_j.

We can calculate the max stress generated due to diffusion by using Prussin's model. From eq. (2), using $C_s = 1.5 \times 10^{21}$ atoms/cc, $E = 1.66 \times 10^{11}$ dynes/cm^2, $\nu = 2.7$ and $\beta \simeq 2 \times 10^{-24}$, as seen in refs. 2 and 18, we obtain $\sigma_{max} \simeq 2 \times 10^9$ dynes/cm^2, which cannot be enough to cause yield through slip. In a totally dislocation-free material, the usual expression for stress necessary to start yield is $\frac{\text{shear modulus}}{30} \simeq \frac{4.94}{30} \times 10^{11}$, $\simeq 10^{10}$ dynes/cm^2, which may never be reached through solute stress alone.

Limitations of the Prussin model for dislocation generation are obvious from the above discussion. According to this model: (1) dislocation generation occurs all at once right at the start of diffusion, (2) the dislocation thus created is dispersed inside silicon during diffusion, (3) surface concentration alone decides the amount of misfit dislocations generated, and (4) the misfit dislocation generation takes place via glide processes alone. Our experimental results are in total contradiction with all these points.

It is now worth exploring the Van Der Merwe theory for phosphorus diffusion-induced dislocations. By the use of eq. (9), we obtain, as given in Table I, a set of values of h (the thickness of the top-layer) for a set of values of δ_c. The values of δ_c for each value of C_s were in turn calculated using the Vegard's rule $\delta_c = \frac{\Delta a}{a} = \frac{\eta \Delta r}{r}$ (where a = lattice parameter, η = atomic fraction of phosphorus in silicon). Furthermore, we assumed that the impurity concentration C_s is uniformly held in substitutional sites throughout the thickness h. A corresponding plot in Fig. 14 shows a relationship between δ_s and h. Since in a diffused region corresponding to thickness h, uniform concentration across the thickness is absent, we must take an appropriate average concentration. Referring to the curve C in Fig. 1, we obtain average $C = 3.6 \times 10^{20}$ atoms/cc. Assuming this amount of phosphorus to be uniformly distributed, we obtain the critical values of h $\simeq$ 4000 Å, which is in good agreement with the value of depth obtained in our experiments. Van Der Merwe's theory essentially takes into account both the parameters Q_{crit} and $x_{j\,crit}$ in a sensitive manner. Because of the little covalent misfit, arsenic needs to be diffused in silicon to very large depths to produce misfit dislocation nets as seen in Fig. 14. The fact that dislocation loops observed in arsenic-diffused silicon do not have their Burgers vectors lying on the plane of diffusion is consistent with the consideration that misfit of arsenic in silicon lattice is negligible.

Table I. Data for Van Der Merwe eq. (9).

C_s	δ_c	h(Å)
1.5×10^{21}	1.3×10^{-3}	590
8.0×10^{20}	6.9×10^{-4}	1 280
6.0×10^{20}	5.2×10^{-4}	1 780
4.0×10^{20}	3.4×10^{-4}	2 950
2.0×10^{20}	1.7×10^{-4}	6 980
1.0×10^{20}	0.9×10^{-4}	14 730

6.2 Mechanism of Misfit Dislocation Generation

Formation, propagation, and interaction of diffusion-induced misfit dislocations in silicon crystals of various orientations have been studied in detail by G. Thomas and his associates. [4] They proposed for a (100) silicon creation of long 60^o dislocations, lying on planes parallel to diffusion surface, during early stages of diffusion. These dislocations by glide on {111} planes follow the traveling zone of misfit to reduce solute misfit. When suitably-oriented glide dislocations meet, a stable network is produced through interaction. Later on for (111) silicon, they proposed climb mechanism for the motion of the long-edge dislocations, which after suitable orientation are expected to be produced through interaction-dislocation nets parallel to the diffusion surface. [5]

Our observations point to the initial generation of interstitial prismatic loops with their Burgers vectors oriented in the plane of diffusion. These loops give rise to misfit network during their growth and interaction as shown sequentially for a (111) Si in Fig. 16 through stages (a) to (e). The final configuration after interaction of 60^o dislocations on the same plane is hexagonal net as shown in Fig. 16(e). Through interaction of two pure-edge dislocations or an edge dislocation and a 60^o dislocation, we can obtain networks on a (111) plane with mixed configurations. Since singular dislocations are also expected to be occasionally present, widely varying dislocation configurations are expected and also invariably occur. Sometimes precipitates alone are observed, as shown in Fig. 4, to give rise to patches of dislocation nets. These patches seem to be created for the purpose of relieving precipitation stresses.

In arsenic-diffused silicon, prismatic loops were not observed to possess Burgers vectors in the diffusion plane. Misfit dislocation nets were also unobserved. As indicated earlier, the absence of misfit dislocations is expected since solute stresses are only marginal in this case because of almost the same radius of arsenic in silicon lattice.

Generation of interstitial loops occurs in silicon during both arsenic and phosphorus diffusion. The source of these dislocation loops is a problem of considerable importance. This problem is discussed next.

6.3. Diffusion Mechanism

A. Seeger and K. P. Chik [13] have made, as pointed out in our introduction, an exhaustive review of diffusion mechanisms and point defects in silicon and germanium. They concluded that self-diffusion in silicon at temperatures higher than 800 ^oC is due to "extended" interstitial in Si instead of the usually assumed "simple" defect of vacancy type. This conclusion was partly based upon such facts as very low values of the self-diffusion coefficients D^{SD} and high values of the pre-exponential factors D_o^{SD} of the valence crystals in comparison to those of the metals. After reviewing the vast diffusion data of Groups III and V impurities, and Groups I and VIII impurities, vacancies and interstitials of single or double character were categorically ruled out by them as defects responsible for diffusion in Si and Ge. They gave considerable importance to the idea that these defects cannot account for the large experimental values ($\approx$ 15 k) of the entropy of self-diffusion. Only "spread-out" defects could account for these entropy values. The extended interstitial defect, as conceived by Seeger and Chik, is a locally-melted region of about 10 to 11 atoms in size and without appreciable long-range strains.

Observation of interstitial-type dislocation loops in both P and As-diffused silicon crystals give considerable credence to the concept of some sort of interstitial-type of defect being responsible for diffusion. Nevertheless,

we cannot establish, through our experiments, the extended nature of the defect. Information about the activation energy for self-diffusion of silicon in the presence of large amounts of P and As ($\simeq$ 0.2 atomic %), however, has been calculated to be $\approx$ 2.12 eV by observing the process of the climb of dislocation loops. The utility of the observation of the climb of loops has been demonstrated by Silcox and Whelan and by Thomas and Kannan. [21,22] The phenomenon is theoretically interpreted by Friedel in terms of climb. [23] The above value of activation energy for self-diffusion ΔH^{SD} is considerably less than the experimental tracer values, 4.86 eV and $\simeq$ 5.3 eV in intrinsic silicon. We observed a sharp drop in activation energy, which occurs at the solubility limits in supersaturated alloys where the defect concentration is expected to be much greater than in unsaturated alloys. [22] In the light of this consideration, the self-diffusion value of 2.12 eV obtained seems reasonable for max possible doping in silicon with P and As. Fairfield and Masters [25] have obtained self-diffusion coefficients of Si via tracer technique in extrinsic silicon doped with P and As in the concentration range of 7×10^{19} and 1.88×10^{20} atoms/cc. [24,25] This data shows reduction in the activation energy of self-diffusion by small amounts and the max reduction for 1.88×10^{20} atoms/cc of phosphorus is, however, only about 0.15 eV. The sharp drop in the self-activation energy of silicon observed in our samples in the vicinity of the surface seems to be the result of considerably much higher concentration. Assuming the solid solubility limit concentration at the diffusion temperature and combining our data with that of Fairfield and Masters, a rough-estimate plot of ΔH^{SD} for silicon self-diffusion versus arsenic (or phosphorus) concentration is given in Fig. 17. The origin of the large concentration dependence of diffusivity of either phosphorus or arsenic in deep diffusions is no doubt connected with the growth of the interstitial dislocation loops, which is the manifestation of condensation of nonequilibrium concentration of interstitial silicon.

How does supersaturation of the interstitial-type defects occur in diffused silicon wafers without any special precaution to freeze the defects is a difficult question to answer. A plausible explanation is that, like as in a Kirkendall couple, the "extended-interstitial" current condenses during diffusion to form flat discs in n-type region. These discs collapse into prismatic dislocations upon annealing. Various Kirkendall experiments have shown that microscopic spherical holes are formed, instead of prismatic dislocation loops, by condensation of vacancies on inclusions in the side of the diffusion couple containing initially all the fast-moving impurity. [26-28] Unlike the usual Kirkendall couples, silicon in our experiments is superpure, dislocation-free and single crystal. Consequently, prismatic dislocation loops, instead of spherical holes, should be the result of condensation of defect current during diffusion. Since the Kirkendall effect is valid in the case of both vacancy and interstitial diffusion mechanisms, the above explanation for generation of interstitial dislocation loops seems valid. In the strained lattice involving diffusion of phosphorus, the loops tend to orient their Burgers vectors in the plane of diffusion to reduce the solute lattice trains. In the almost solute strain-free arsenic-diffused silicon, the loops have their Burgers vectors randomly oriented. The so-called concentration dependence of diffusivity of arsenic or phosphorus ensues after generation of these dislocation loops.

7. Acknowledgments

The authors are indebted to Mr. S. Ancri, Dr. G. R. Gunther-Mohr, and Dr. E. S. Wajda for their continued interest, support, and appreciation of our work. Also, the authors wish to thank Mr. R. Shasteen for his excellent help in transmission electron microscopy.

8. References

[1] H. J. Queisser, J. Appl. Phys. 32, 1776 (1961).

[2] S. Prussin, J. Appl. Phys. 32, 1876 (1961).

[3] G. H. Schwuttke and H. J. Queisser, J. Appl. Phys. 33, 1540 (1962).

[4] J. Washburn, G. Thomas, and H. J. Queisser, J. Appl. Phys. 35, 1909 (1967).

[5] E. Levine, J. Washburn, and G. Thomas, J. Appl. Phys. 38, 81 (1967).

[6] M. L. Joshi and F. Wilheim, J. Electrochem. Soc. 112, 185 (1965).

[7] R. J. Jaccodine, Appl. Phys. Letters 4, 114 (1964).

[8] W. Czaja, J. Appl. Phys. 37, 3441 (1966).

[9] M. Yoshida, H. Arata, and Y. Terunuma, Japan. J. Appl. Phys. 7, 209 (1968)

[10] E. Tannenbaum, Solid State Electronics 9, 807 (1966).

[11] M. L. Joshi and S. Dash, IBM J. Res. Develop. 10, 446 (1966).

[12] R. A. McDonald, G. G. Ehlenberger, and T. R. Huffman, Solid State Electronics 9, 807 (1966).

[13] A. Seeger and K. P. Chik, Phys. Stat. Sol. 29, 455 (1968).

[14] J. H. Van Der Merwe, Single Crystal Films (Macmillan Company, New York, 1964), pp. 139-163.

[15] M. L. Joshi, C. H. Ma, and J. Makris, J. Appl. Phys. 38, 725 (1967).

[16] J. P. Hirth, Single Crystal Films (Macmillan Company, New York, 1964) pp. 173-176.

[17] Integrated Silicon Device Technology, AD603716 Diffusion (Research Triangle Institute, Durham, North Carolina, U.S.A., 1964), Vol. IV.

[18] I. M. Mackintosh, J. Electrochem. Soc. 109, 392 (1962).

[19] M. L. Joshi and S. Dash, IBM J. Res. Develop. 11, 271 (1967).

[20] S. Amelinckx, The Direct Observation of Dislocations (Academic Press, New York, 1964), pp. 361-363.

[21] M. J. Whelan and J. Silcox, Phil. Mag. 5, 1 (1960).

[22] V. C. Kannan and G. Thomas, J. Appl. Phys. 37, 2363 (1966).

[23] J. Friedel, Les Dislocations (Gauthier-Villars, Paris, France, 1956), p.72.

[24] B. J. Masters and J. M. Fairfield, Appl. Phys. Letters 8, 280 (1966)

[25] J. M. Fairfield and B. J. Masters, J. Appl. Phys. 38, 3148 (1967).

[26] F. Seitz, Phys. Rev. 74, 1513 (1950).

[27] F. Seitz, Acta Cryst. 3, 355 (1950).

[28] F. Seitz, Acta Met. 1, 355 (1953).

9. Appendix

9.1. Interstitial Nature of Arsenic Diffusion-Induced Loops

We had available a tilting device that allowed tilting over a large angle ($\pm 35^{o}$) in the Philips 300-EM transmission electron microscope (TEM). Therefore, the tilting method first devised by G. K. Williamson and B. Edmondson for determination of the nature of the prismatic loops could be used. [1] To determine whether the loop is vacancy or interstitial type, it is essential to maintain the two-beam condition and to obtain: (1) the sense of the slope of a loop plane, (2) the sense of the diffraction vector $\vec{g}$, and (3) the loop size under positive and negative deviations from the exact Bragg diffraction condition, i.e., under + and - values of $\vec{s}$.

In Figs. 18(a) and (b), we have a TEM view of arsenic-induced loops in a silicon foil under the same diffraction vector $\langle 220 \rangle$ but with different signs of $\vec{s}$ as indicated. Electron micrographs obtained on a set of loops under different tilts with the goniometer-tilt axis approximately parallel to their long axes are shown in Fig. 18(c). The size changes, occurring for purely geometric reasons, are obtained in these micrographs because of the large amount of tilting involved. The size changes established the sense of the loop plane as shown.

For the same sense of the strong operating diffraction vector, the loops are found to enhance in size due to dependency of strain contrast on the sign of $\vec{s}$, as seen in Figs. 18(a) and (b). The loop size increases as we rotate the crystal from $-\vec{s}$ to $+\vec{s}$, as expected for an interstitial loop lying on a plane with the sense of inclination shown in Fig. 18. The correct sense of $\vec{g}$ was obtained by taking into account both the electron-optical rotation and the relative inversion of the micrographs and the diffraction pattern. Sign of $\vec{s}$ is + ve when the Kikuchi line is further from the center than the diffracting spot on the diffraction pattern.

Many of the dislocation loops in Fig. 18(a) have stacking faults in them. They are Frank loops of extrinsic type and are geometrically disposed as shown in Fig. 11.

10. References

[1] G. K. Williamson and B. Edmondson, Proc. Joint Conf. on Inorganic and Intermetallic Crystals, U. of Birmingham, Eng. (1963).

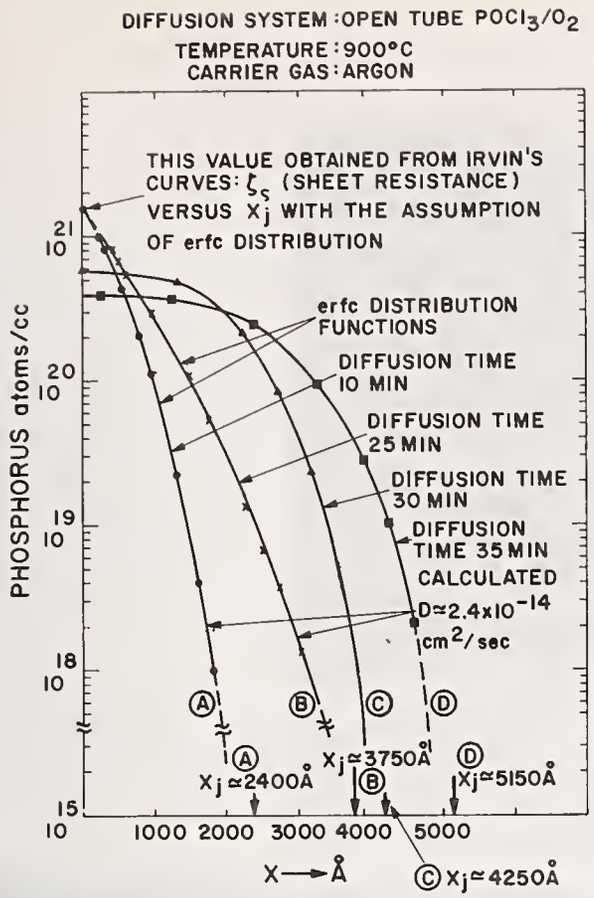


Fig. 1. Phosphorus distribution obtained through open-tube diffusion processes. Observe the reduction of C_s after exceeding $x_j \simeq 3750$ Å.

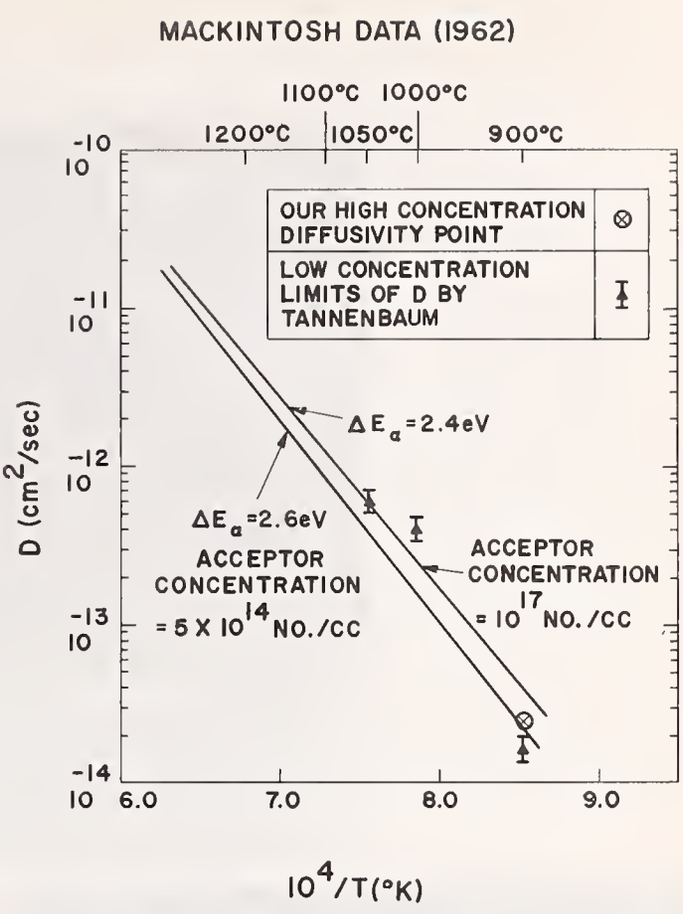


Fig. 2. Phosphorus diffusivity as a function of reciprocal absolute temperature.

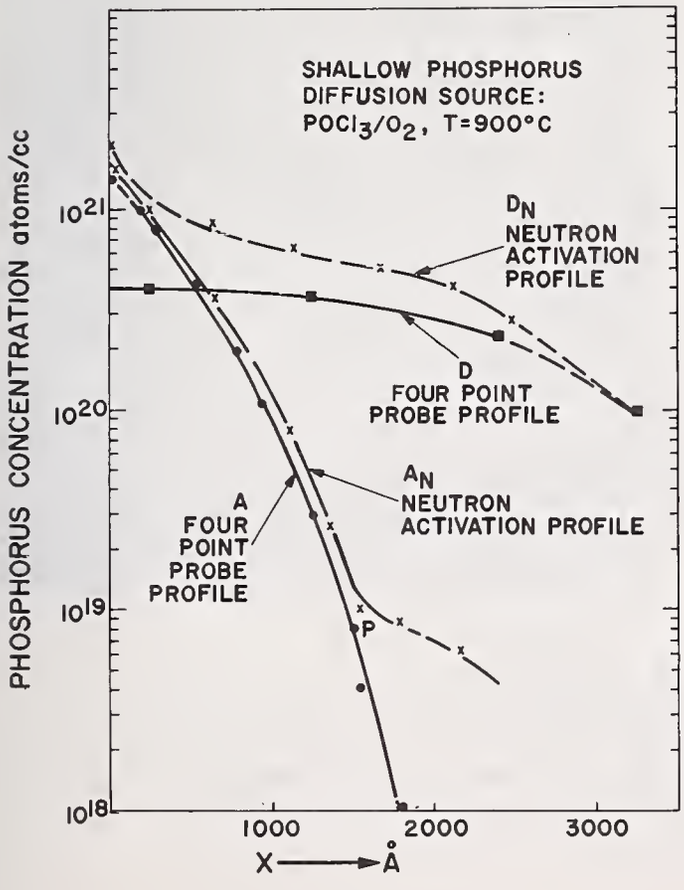
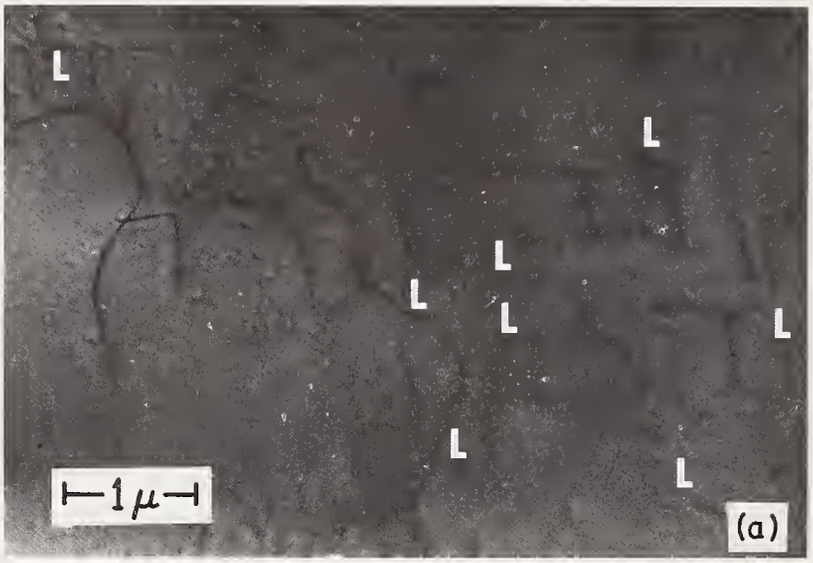
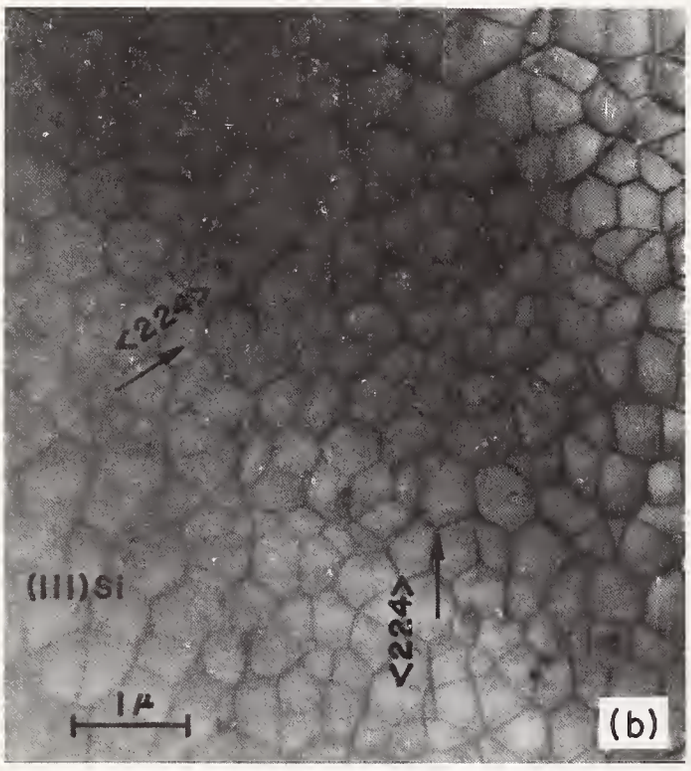


Fig. 3. Comparison of phosphorus distributions obtained by the techniques of neutron activation and four-point probe. Observe the increased difference in the distributions obtained by the two techniques for the deeper diffusion.





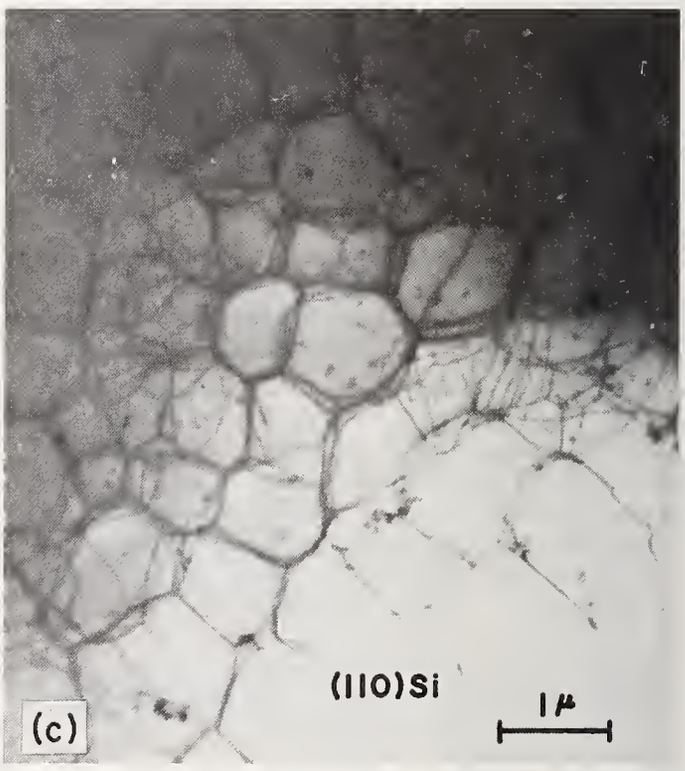


Fig. 4. Dislocations in silicon corresponding to shallow and deep diffusions.
(a) Dislocation patches and the interacting dislocation loops in a (111) silicon corresponding to C-type impurity distribution given in Fig. 1.
(b) Dense misfit dislocation nets in a (111) silicon corresponding to D-type impurity distribution in Fig. 1.
(c) Dense three-dimensional dislocation nets in a (110) silicon corresponding to D-type impurity distribution in Fig. 1.

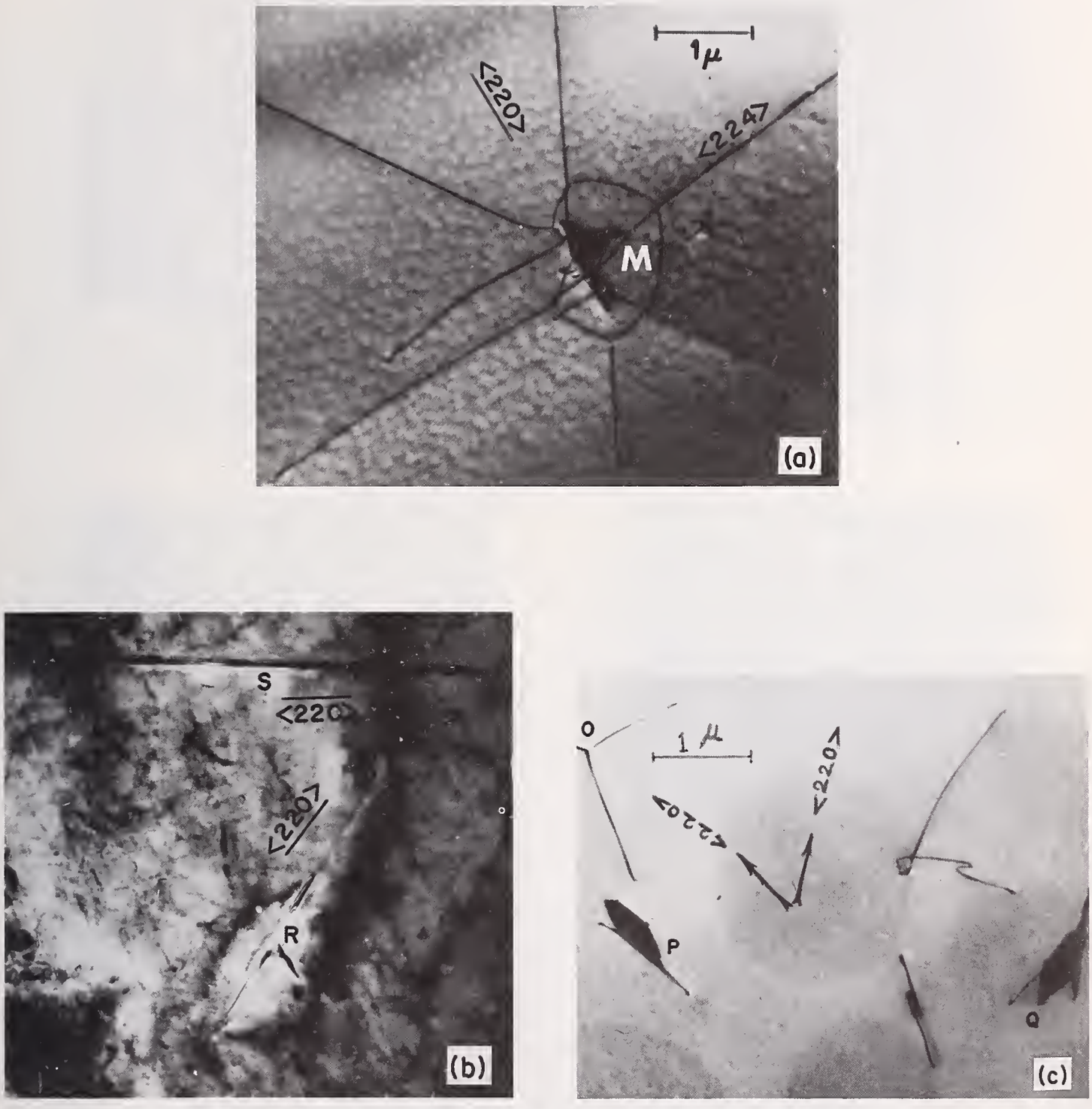


Fig. 5. Precipitates of phosphorus in silicon corresponding to shallow diffusion.
(a) Generation of a dislocation patch at the intersection of a precipitate platelet M in (111) silicon corresponding to C-type impurity distribution given in Fig. 1.
(b) Structure giving fringe contrast similar to that of a stacking fault. Contrast experiments indicated that these are most probably very thin P-Si precipitates corresponding to the same diffusion as in Fig. 5(a).
(c) Singular dislocation node O and clearly identifiable stacking faults P and Q are observed. Clearly identifiable stacking faults, however, were only occasionally observed.

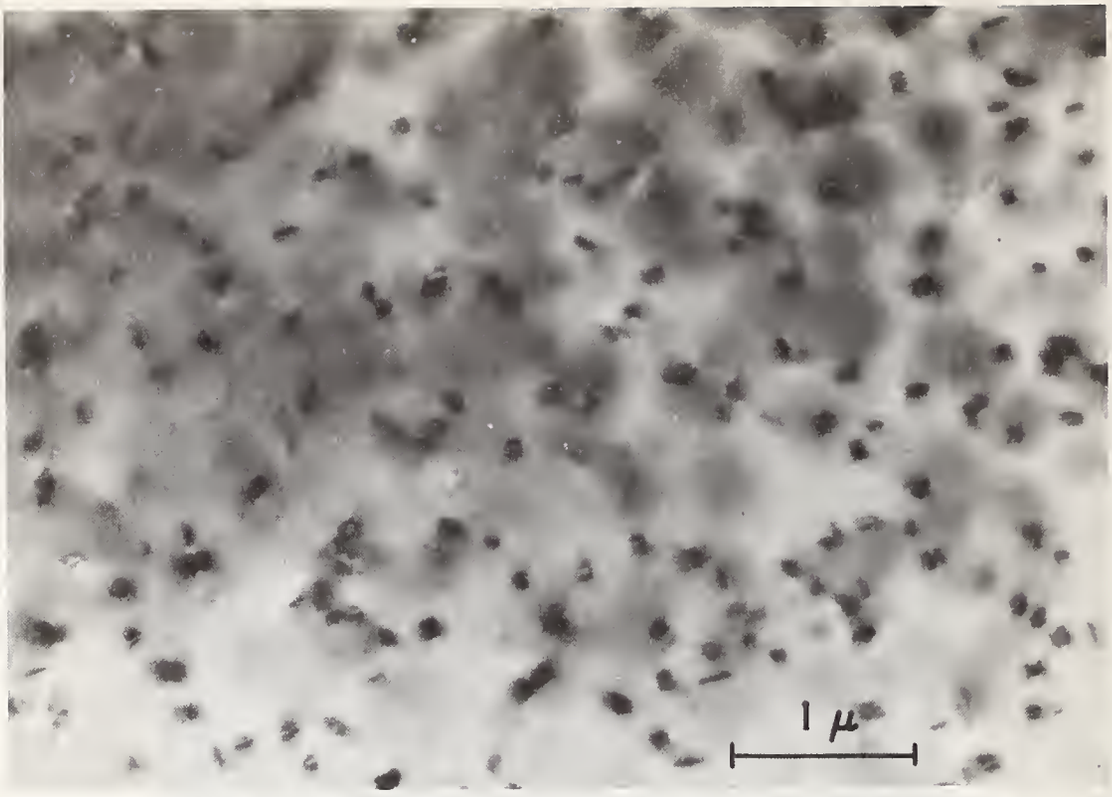


Fig. 6. Annealing-induced dislocation loops in (111) Si corresponding to C-type phosphorus distribution given in Fig. 1.

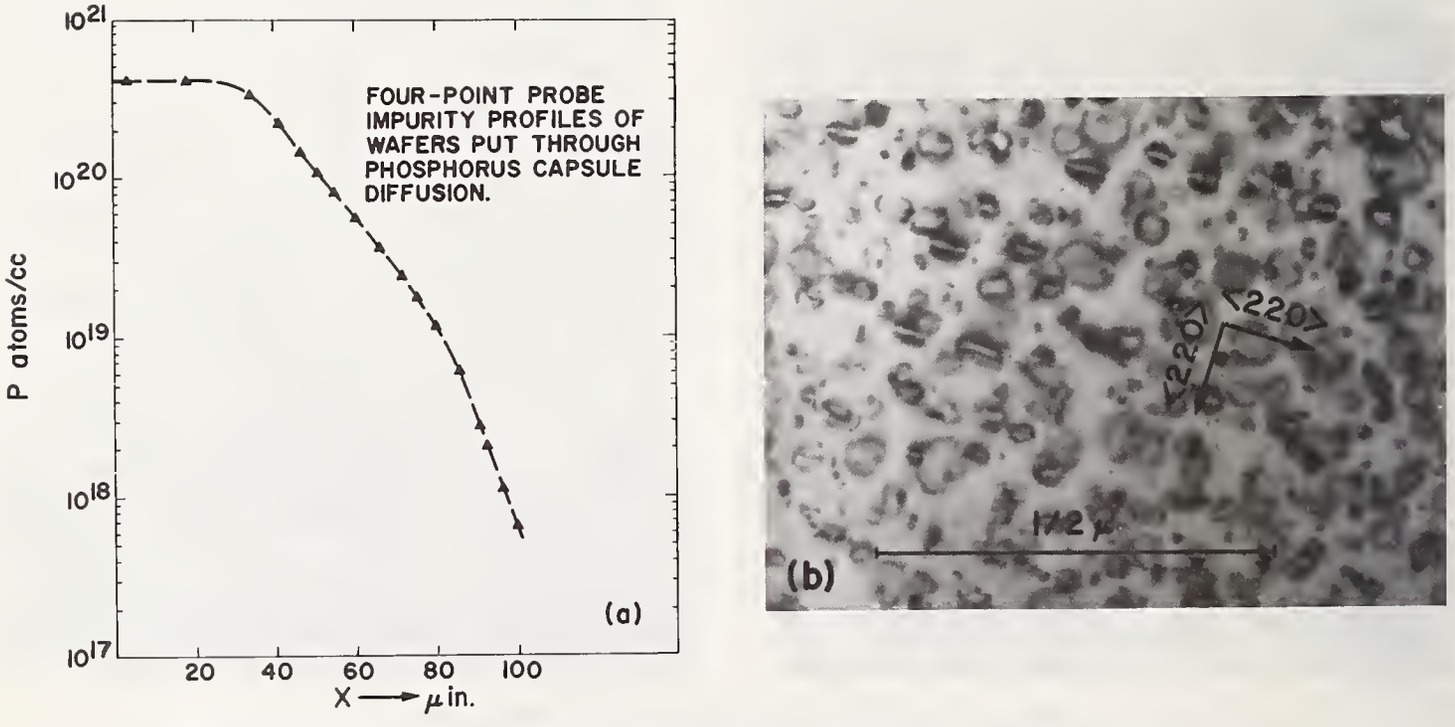


Fig. 7. Defects in silicon corresponding to deep diffusions obtained through capsule technique.
(a) Phosphorus impurity distribution in (100) Si obtained using a high concentration P-Si powder diffusion source in a quartz capsule.
(b) Very high density of dislocation loops oriented along <220> directions in (100) Si are observed corresponding to impurity distribution in Fig. 7(a).

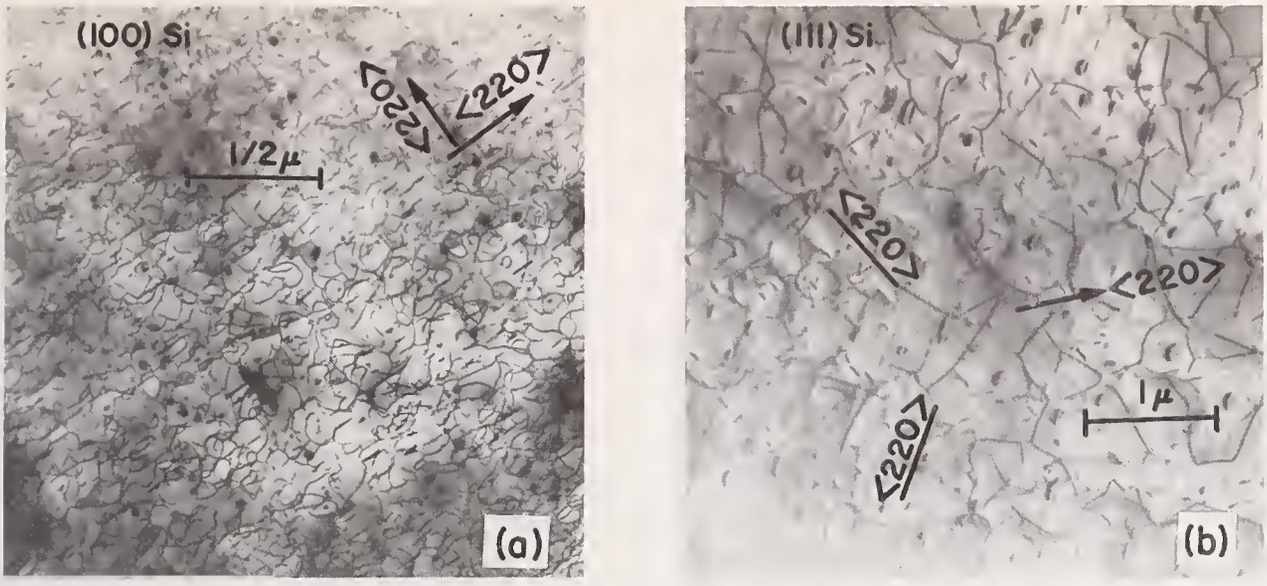


Fig. 8. Dislocation loop interaction induced through annealing treatment.
(a) Growing and interacting dislocation loops are in process of producing misfit dislocation nets during 700 °C 2 h annealing treatment of (100) Si corresponding to impurity distribution in Fig. 7(a).
(b) Same phenomenon as in Fig. 8(a) for (111) Si. Thin precipitates oriented along <220> directions are observed in addition.

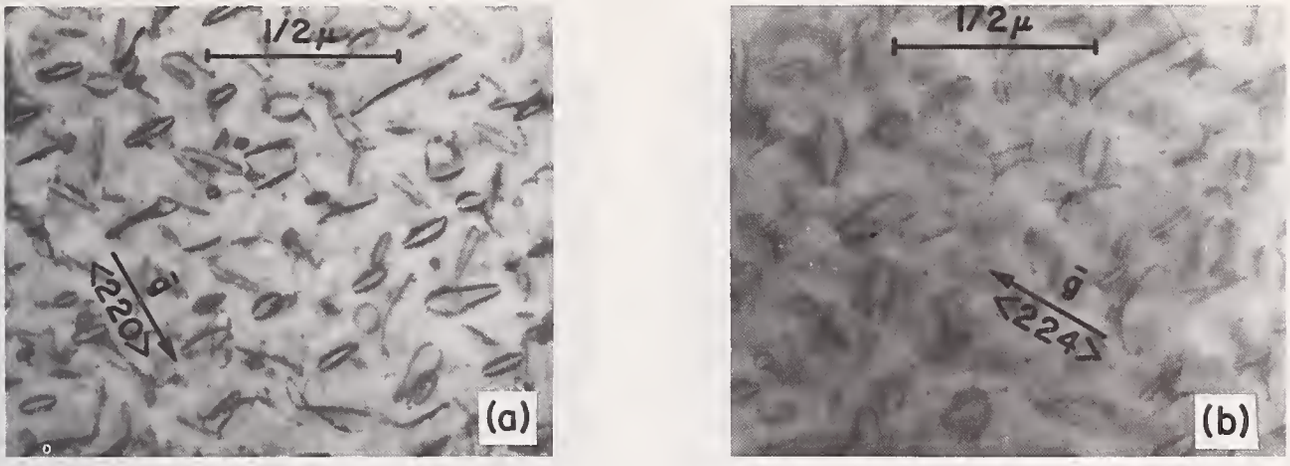
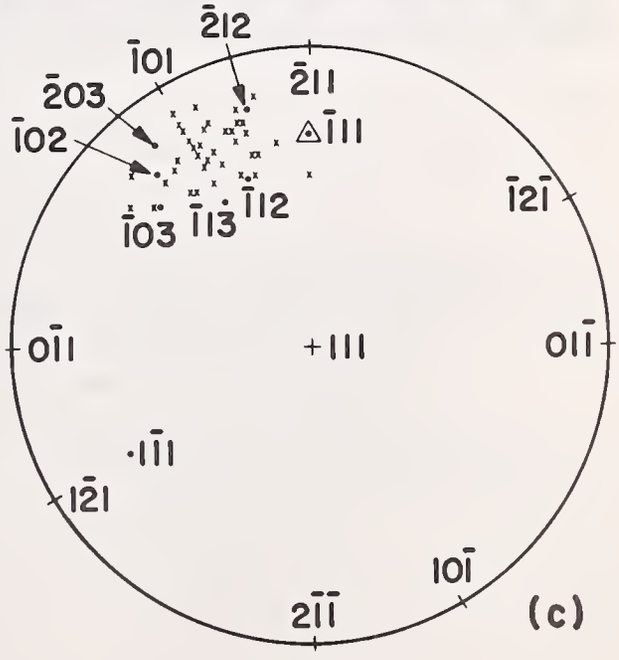


Fig. 9. Diffraction contrasts of loops induced through phosphorus diffusion for two different diffraction vectors.
(a) A view of phosphorus diffusion-induced dislocation loops in (111) silicon for the diffraction vector, $\vec{g}$ = <220>.
(b) Same view as in Fig. 9(a) for the diffraction vector, $\vec{g}$ = <224>. Note the vanishing of dislocation loops parallel to this vector. This observation, in addition to the information in Fig. 9(c), determined the Burgers vector of the loops.
(c) Stereographic projection of loop planes in Fig. 9(a). Multiplicity of indices of the loop planes in Fig. 9(a) are reduced for representation in one part of the stereograph for convenience.



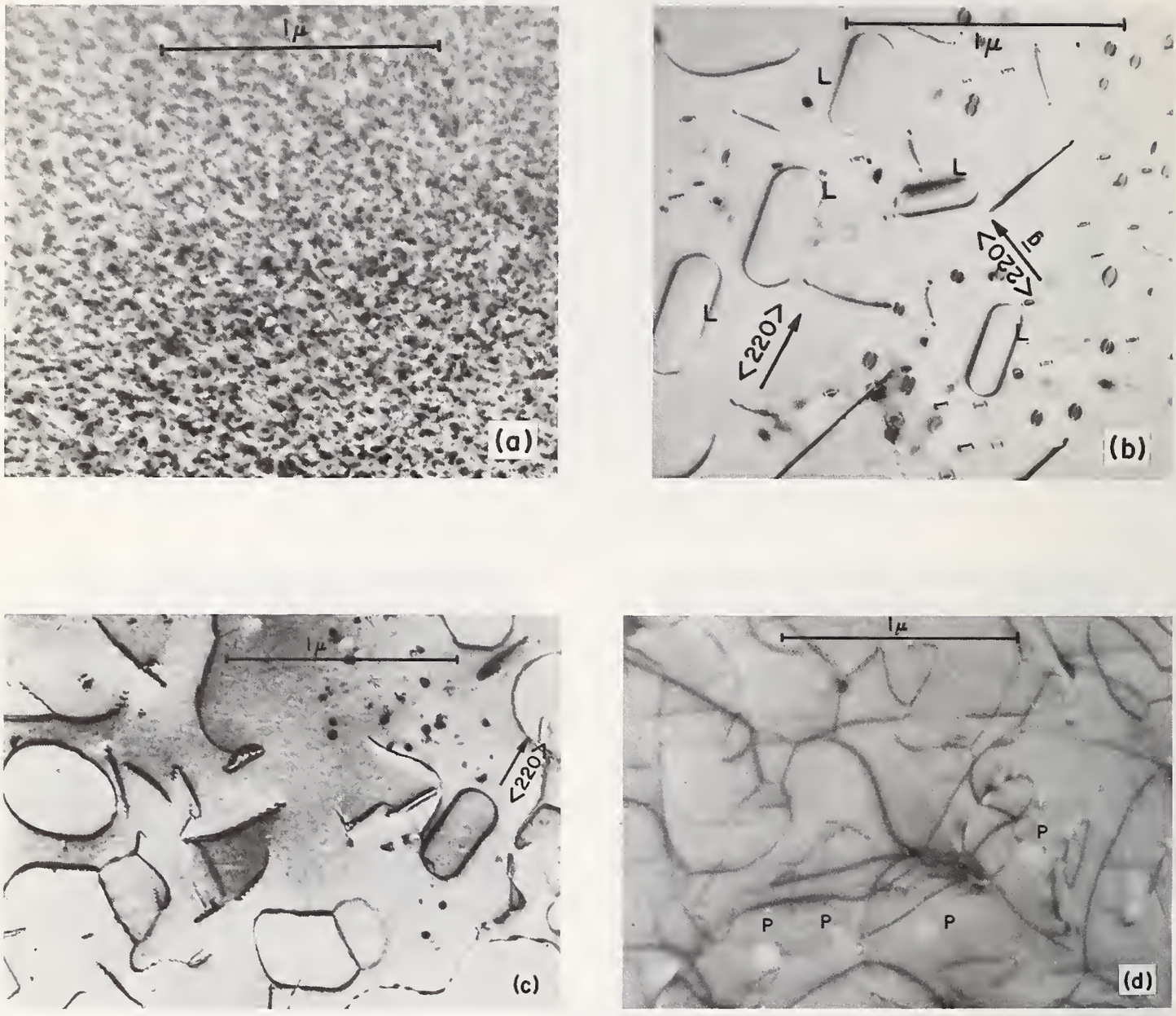


Fig. 10. Annealing growth of dislocation loop-induced arsenic diffusion.
(a) High density of dislocation loops as seen in As-diffused (111) samples.
(b) Samples corresponding to diffusion of Fig. 10(a) were annealed at 900°C for 2 h. Note the loops on inclined planes do not grow as rapidly as the loops on (111) Si samples surface.
(c) Annealing of samples in Fig. 10(a) for 3 h at 900 °C. Loops achieved almost the max size.
(d) Annealing of samples in Fig. 10(c) for 4 h at 900 °C. Loops are broken and formed dislocation tangles. P-type circular structures could not be definitely identified. They may be interpreted to be porous regions in the material since they are not bound by dislocations.

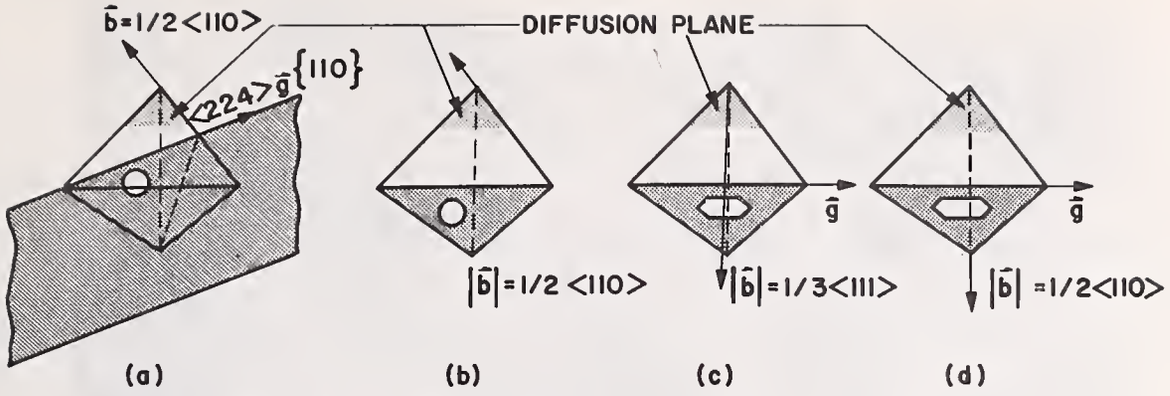


Fig. 11. Geometry of phosphorus-and arsenic-induced sessile loops in silicon.
(a) Phosphorus-diffused silicon-pure edge prismatic loops (stable configuration).
(b) Phosphorus-diffused silicon-prismatic loops with partial edge character.
(c) Arsenic-diffused silicon Frank-type loops.
(d) Arsenic-diffused silicon-prismatic loops with partial edge character.

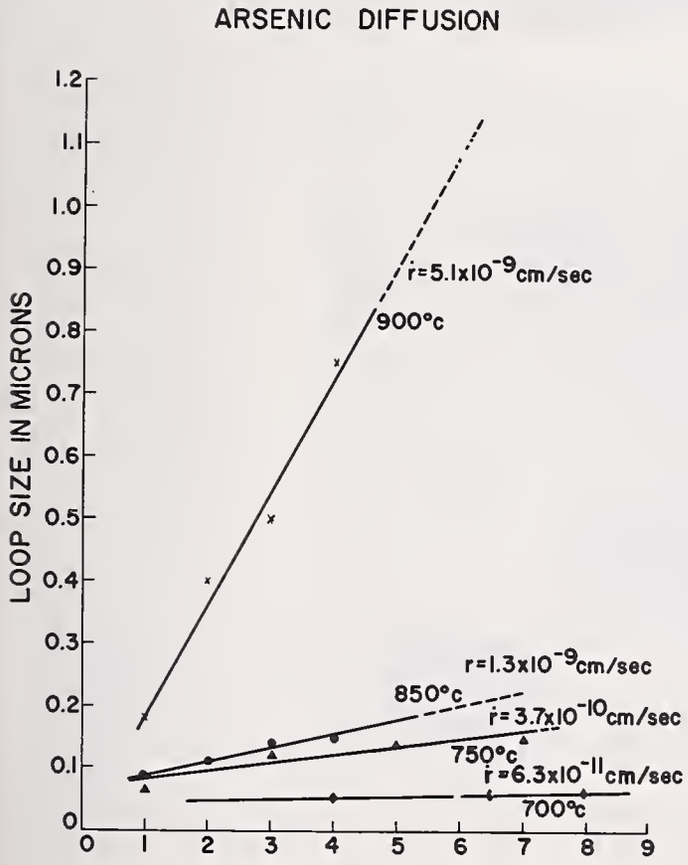


Fig. 12. Growth kinetics of arsenic-induced sessile loops in silicon.

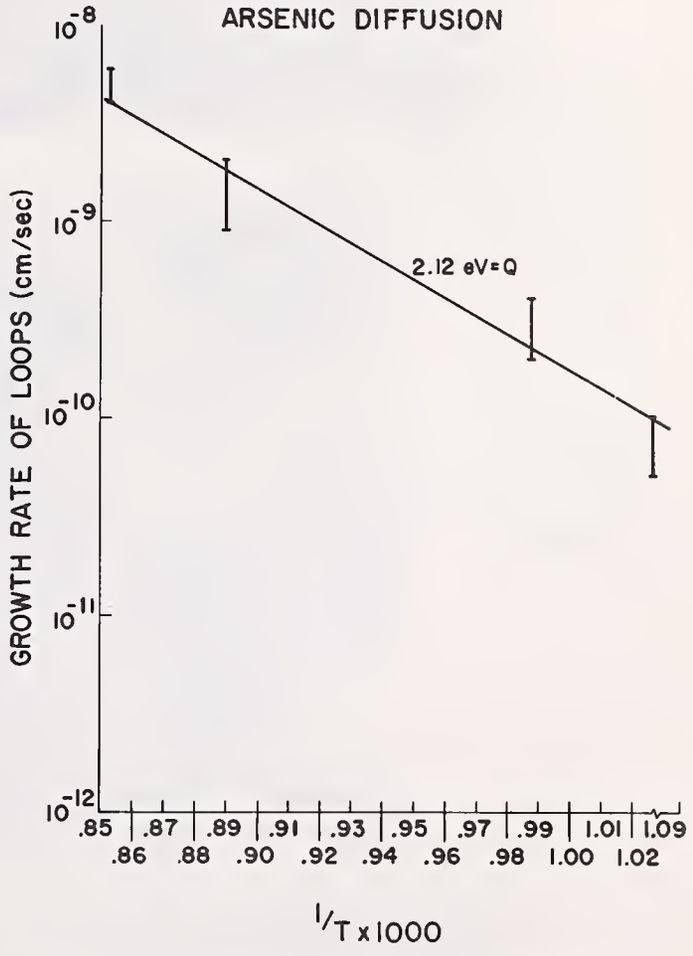


Fig. 13. Determination of activation energy of self-diffusion in arsenic-doped silicon.

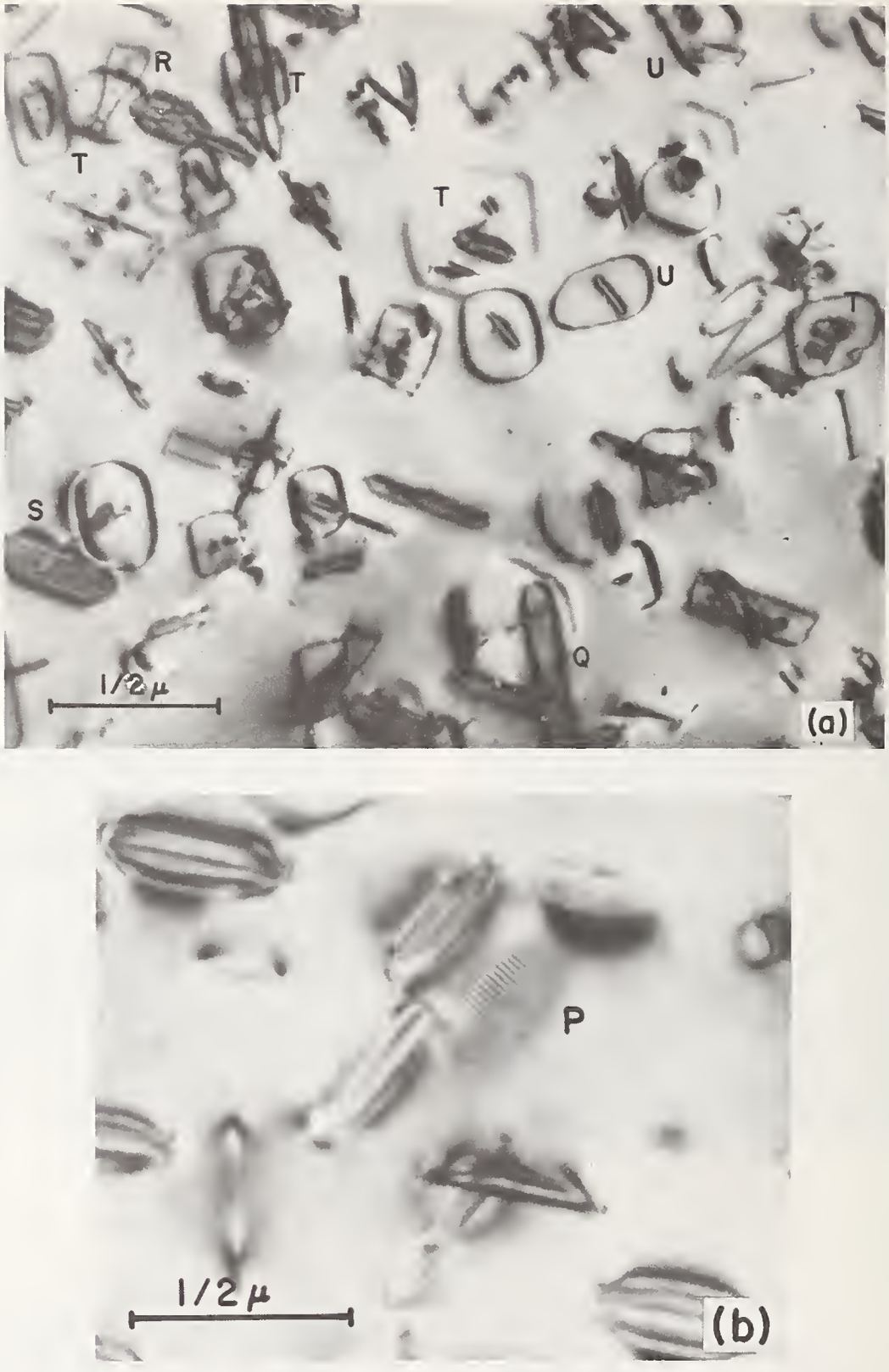


Fig. 14. Morphology of arsenic-induced loops after steam heat-treatment.
(a) Arsenic-diffused silicon heat treated at 1000 °C for .5 h in steam presents a complex morphology of loops.
Q - Elongated prismatic loops
R - Hexagonal Frank loops
S - Frank loop without a central fault at the center
T - Multiple loop generation corresponding to Bardeen-Herring type of source.
U - Irregular prismatic loops lying on the planes of the wafer
(b) Some sample regions corresponding to Fig. 14(a) show a loop P with moiré pattern.

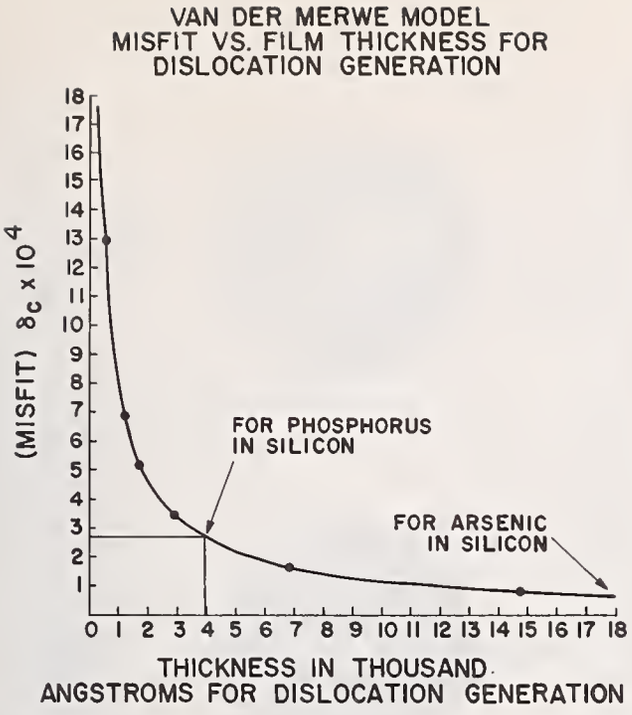


Fig. 15. Atomic misfit versus critical film thickness curve based upon the Van Der Merwe model.

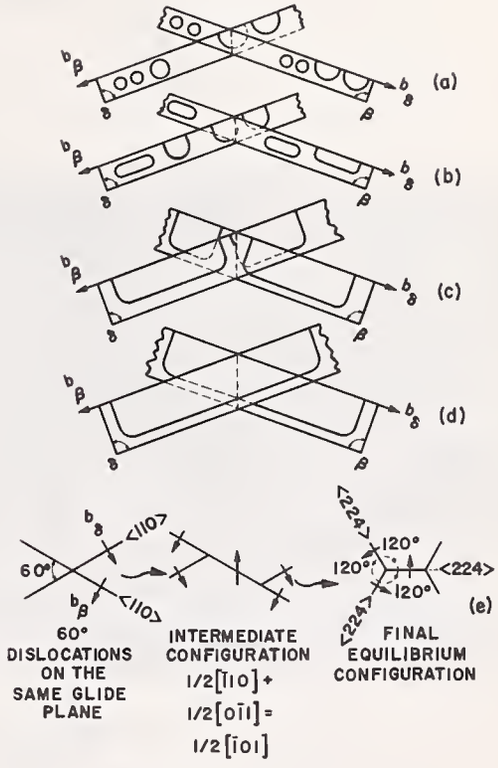


Fig. 16. Stages of sessile loop interaction leading to generation of misfit dislocations.

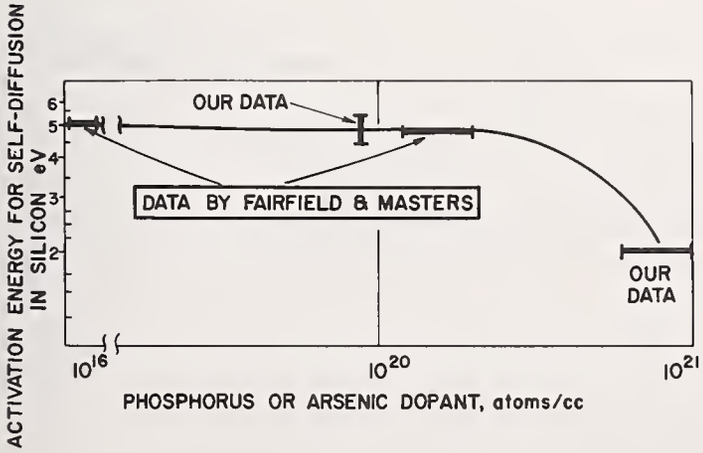


Fig. 17. A rough-estimate plot of self-activation energy of silicon as a function of either arsenic or phosphorus concentration.

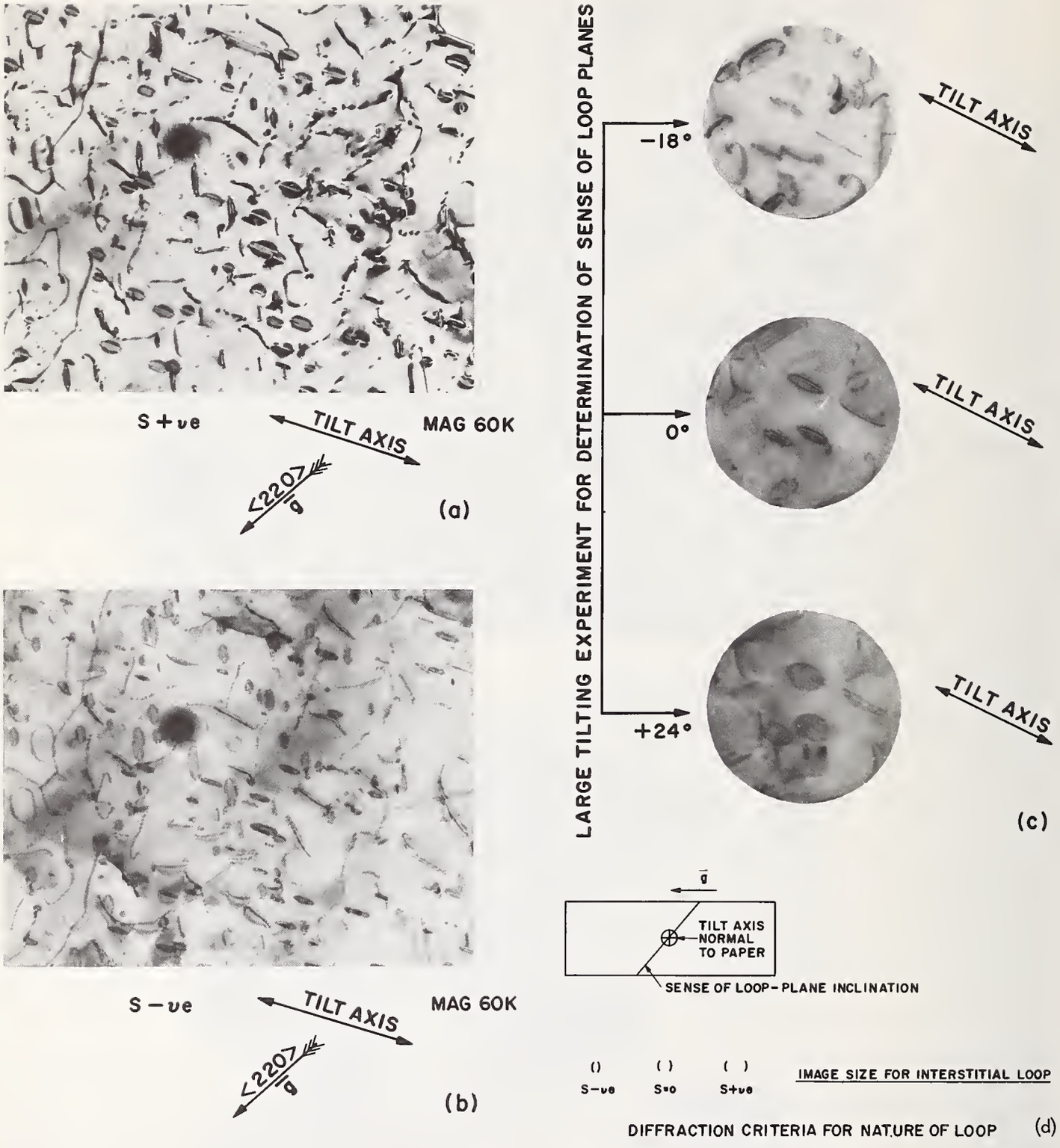


Fig. 18. Diffraction contrast experiments for demonstrating the interstitial nature of dislocation loops.

LIMITATIONS OF CURRENT EPITAXIAL EVALUATIONS

R. N. Tucker

Fairchild Semiconductor
Palo Alto, California

The characterization of epitaxial silicon films can be put in two categories: as an analytical procedure wherein nothing is presumed concerning growing conditions, type, resistivity, gradients or thickness, or as a quality control procedure wherein one verifies the presumed output of the equipment. The former can be a laborious and time-consuming effort. The latter can often be a simple rapid measurement with tolerances depending upon the use of which this material will be put. This is because most epitaxial material for semiconductor devices today is grown with approximately flat profiles; the chemistry for such epitaxial depositions is fairly well defined; and more sophisticated equipment has made possible better reproducibility in epitaxial growth. In such equipment the accuracy of control of flow meters, temperature and time are equivalent to the accuracy of the evaluation equipment. Because of this one expects thickness control to be a minor problem in such cases. More important from a quality control standpoint is the checking of resistivity (which may deviate from the expected due to surface contamination or anomalous autodoping effects), and excessive junction leakage (also due to such factors as improper cleaning).

Other areas of epitaxial deposition are not in such good control. Some examples are the cases of thin films and films having precisely controlled gradients. Reproducibility of $\pm$2000 angstroms would be considered excellent for 5 micron films but might be excessive for a submicron film. Yet by simple extrapolation of growth curves, and by using conventional growth/cleaning techniques better reproducibility is rarely achieved. The nucleation processes and those factors which can influence nucleation are not well understood today. Likewise the delay factors in the introduction of a dopant into the film compared to its introduction into the gas stream are not well-defined. Other cases exist in vapor phase depositions in which diffusivity or structure may be strongly influenced by substrate preparation or deposition conditions. A clear understanding of the controlling factors in all such reactions is imperative before those epitaxial materials can be widely utilized. Having this knowledge the materials scientist or engineer can then work with the equipment engineer and the device engineer to determine which evaluation measurements are essential for any given material after its growth to assure reasonable success in the fabrication of a particular device structure.

On The Interpretation of Some Measurement Methods for Epitaxially Grown Layers

P. J. Severin
Philips Research Laboratories
N.V. Philips' Gloeilamperfabrieken
Eindhoven - The Netherlands

The thickness of epitaxially grown layers is commonly measured by infrared multiple reflection. It has been proposed by ASTM to apply a correction for the phase shift at the layer-substrate interface as a function of wavelength.

By direct calculation from measured spectra, we have found a phase shift independent of the wavelength. The layer-substrate reflection coefficient, generally assumed to be constant, has been verified to vary exponentially with the wave number, yielding another constant phase shift. The measured constant phase shift cannot be explained from this dependence. Generally its reproducible value is different from slice to slice and over a slice but rather constant when the epitaxial layer is grown in a reactor designed for growing layers with constant thickness.

A large number of spectra were analysed and found to yield a very precise thickness and constant phase shift. The correlation of the measured values to a linear wave number versus integer relationship is typically 0.9998 using the method of least squares with a desk calculator.

Spreading resistance measurements are carried out with a blunt formed specially hardened steel pin from an ordinary Dumas four-point probe head. Microscopic observations and electrical measurements provide conclusive evidence that the pin is multi-tipped. This leads to an equivalent circuit consisting of about fifteen spreading-resistance controlled branches. Diode controlled branches are ineffective. A straight line calibration is found for silicon between 0.005 and 10 ohm cm. Depending on the silicon surface preparation,polished or lapped, the reproducibility on the bulk material is 1 or 5 percent. The resolving power in depth turns out to be about half a micron, horizontally it amounts to 25μm. The characteristics of the system regarding reproducibility and resolving power are both due to the multi-tipped aspect of the pin. This also modifies the interpretation of the measurements on (angle-lapped) thin layers: in the spreading resistance two different stages, shallow and deep, should be distinguished. They are about equally important and probably play a role with any probe material.

In the introduction various characteristics of a measurement method: accuracy, precision, reproducibility, destructiveness, speed and in-process moment have been elaboated and applied particularly to the four-point probe, spreading resistance and capacitance-voltage methods. These methods are competitive and complimentary in obtaining the resistivity ρ.

Key Words: Capacitance-voltage method, epitaxial layer,four-point probe method, infrared interference method, resistivity, silicon, spreading resistance method, thickness.

1. Introduction

It is a characteristic feature of advanced electronic technology that it is strongly material-dependent. Because of the tendency to design smaller components and to integrate to larger units,

further growth is conditioned largely by the quality of the material available.

Integrated circuit technology created a rapidly increasing demand for thin slices backed by a well-conducting or well-isolated supporting substrate. The rise of epitaxially grown Si on Si technology is due to and concomitant with this requirement: complex integrated circuits call for the characterization of a piece of silicon in sophisticated terms.

Any physical quantity can be used for such characterization, but not all of them are relevant. The circuit designer formulates his requirements specifying the relevant physical quantities and the precision and accuracy with which they should be measured. It may happen that the accuracy and precision of a particular parameter cannot be attained by any method suitable as a measurement method. Then his device is an ideal test vehicle for accurate and precise measurement of this parameter. Such a device should not be too hypersensitive to two or more parameters in the same way, because it may be difficult to disentangle the observed effects. Generally, it is economically most desirable to measure the relevant physical parameters in the very early stage of device processing. However, it should be verified then that this parameter is not subject to further changes during the ensuing stages of the manufacturing process. This may apply to sharp-profile epitaxy followed by diffusion stages.

The physical quantities we want to discuss more specifically in this contribution are the thickness d, the resistivity ρ of the epitaxial layer and the local variations of these quantities. The infrared interference, spreading resistance, four-point probe and capacitance-voltage methods will be treated.

Any measurement produces a number. Without understanding the real process by which this number is obtained it can be used for characterization of a slice, if the measurement result is influenced by the desired parameter in a known way and other parameters have negligible influence. When the value of the parameter is used not only for simply distinguishing but also for deciding on further slice handling, the theory of the measurement should be more carefully studied. This applies to infrared interference thickness measurement.

The terms accuracy and precision have been adequately defined in the ASTM practice[1]. The term reproducibility which is said to be rather ambiguous, nevertheless appears to be indispensable. A measurement method can very well be destructive in the sense that a certain type of measurement cannot be repeated exactly on the same spot of the same piece of material. A destructive measurement method is also an irreproducible one. In case the material were perfectly homogeneous, the measurement could safely be repeated on a neighboring spot, thus determining the degree to which it reproduces, which is basically the precision if it were a non-destructive process. However, the material is not perfectly homogenous and the procedure to determine both the reproducibility and the homogeneity of the material is outlined with the discussion of the spreading resistance method. In general, the reproducibility should be defined operationally being determined in a process-dependent way. With a destructive method usually the specimen is being destructed but of course the probe may be affected too. If a new probe were used for every measurement, the reproducibility simply reflects the spread in the relevant property of the probe. If the probe is not replaced then the reproducibility is determined by the previous history of the probe. The measurements then are not entirely independent and subjected to true statistics, but show a secular variation which should be accounted for. This will also be discussed with the discussion of the spreading resistance method.

By the spatial resolving power $\bar{x}$ of a measurement method is meant the surface area $\bar{x}^2$ covered by one measurement. If the physical quantity measured shows a dependence on position, a weighted average results. When a slice is characterized by a few such measurements, a distance $>\bar{x}$ apart, the local value of the phsyical quantity should not be specified more precisely than the relative variation over the surface area covered. When a slice is measured at many points a distance $<\bar{x}$ apart, it makes sense to step up the precision of the measurement. When a slice is characterized by one measurement it would be wise to use an averaging length about equal to the slice radius. This sort of consideration is relevant to the four-point probe measurement method.

Silicon device technology being strongly related to industrial activity speed and easiness of measurement may be of paramount importance. This seems to be the intention of the various development activities in the field of automated capacitance-voltage measurements. It is evident that where a method is a vital procedure for silicon device-product selection it should be applied at the earliest stage possible on the virginal slice. Speed of measurement is increased much more by simple and quick diode creation than by automating the bridge measurement. Therefore we measure the capacitance-voltage relationship of an ad hoc and instantaneously produced Schottky barrier diode, with an automated instrument for capacitance-voltage measurement.

This introduction should serve as an explicit justification for studying the cognitive value of the parameters concerned and how to increase it.

---

[1] Figures in brackets indicate the literature references at the end of this paper.

The various topics discussed above in general measurement terms will be treated in greater detail in the next sections.

2. The Infrared Thickness Measurement of Epitaxially Grown Silicon Layers

2.1 Introduction

The thickness of a silicon layer epitaxially grown on a silicon substrate of higher conductivity is generally measured by infrared multiple interference. For the wavelength range 2.5 to 50 μm the complex refractive index can be calculated. In this way the phase change δ upon reflection at the high-low interface can be calculated and computed.

The author has experimentally observed that this phase change does not show the theoretical wavelength-dependence, but that a wavelength-independent value δ is found.

It can be shown that an exponential decrease of the interface reflection coefficient $\hat{r}_{12}$ with the wave number $k = 1/\lambda$ is formally identical with a constant phase shift. The reflection coefficient $\hat{r}_{12}$ in fact decreases in accordance with this law, but the constant phase shift is not due to this effect only.

2.2. Theory of Multiple Reflection for Thickness Measurement

The infrared interference method was introduced for industrial application by Albert and Comba[2]. It is shown in any textbook on optics that the rays directly reflected at the (01) interface (r_{01}) combine with the rays transmitted through layer (1) and reflected at the (12) interface (r_{12}) to an effective intensity reflection coefficient R given by

$$R = 1 - \frac{(1 - \hat{r}_{01}^2)(1 - \hat{r}_{12}^2)}{(1 - \hat{r}_{01}\hat{r}_{12})^2 + 4\hat{r}_{01}\hat{r}_{12}\sin^2[(\phi - \delta)/2]}, \tag{1}$$

where $\phi = 4\pi d n_1, k \cos\Theta$, $r = \hat{r}\exp(j\delta)$, $\delta_{01} = \pi$, $\delta_{12} = \delta$ and

the geometry is defined in Fig. 1.
This expression can be approximated with $\hat{r}_{12}^2 \ll 1$ and $\hat{r}_{01}\,\hat{r}_{12} \ll 1$ to give

$$R = \hat{r}_{01}^2 - 2(1 - \hat{r}_{01}^2)\,\hat{r}_{01}\,\hat{r}_{12}\cos(\phi - \delta). \tag{1a}$$

Extreme values for both equations (1) and (1a) are found when

$$\phi - \delta = \ell 2\pi \quad \text{for minima and}$$

$$\phi - \delta = (\ell - \tfrac{1}{2})\,2\pi \quad \text{for maxima,} \tag{2}$$

where the order is indicated by an integer ℓ.

However, when it is assumed that the reflection coefficient $\hat{r}_{12}$ depends on the wave number k, the extreme values are different. Determining again the extreme values of the denominator in equation (1) the relation

$$\hat{r}_{01}\hat{r}_{12} = \cos(\phi - \delta) - \hat{r}_{12}\left(\frac{\delta\bar{r}_{12}}{dk}\right)^{-1}\left(\frac{d\phi}{dk} - \frac{d\delta}{dk}\right)\sin(\phi - \delta) \tag{3}$$

is found on writing $\tan\delta^* = \hat{r}_{12}\left(\frac{d\hat{r}_{12}}{dk}\right)^{-1}\left(\frac{d\phi}{dk} - \frac{d\delta}{dk}\right)$ (4)

$$\hat{r}_{01}\hat{r}_{12} \cos \delta^* = \cos \{\phi - (\delta - \delta^*)\} \quad , \tag{3a}$$

and assuming $\hat{r}_{01}\ \hat{r}_{12} << 1$, as found from equation (2b),

$$\cos \{\Phi - (\delta - \delta^*)\} = 0 \ . \tag{3b}$$

The procedure for measuring the thickness of an epitaxial layer recommended by Albert and Combs[2] is based on equation (2), where, in addition to a wavelength-independent $\hat{r}_{12}$, it is assumed that $\delta = 0$ or, since only perpendicular incidence is considered, 2 dnk = ℓ for minima and 2 dnk = $(\ell - 1/2)$ for maxima (2a). It is numerically obvious, however, that for reflection from a substrate of 10^{-2} or 10^{-3} Ωcm N-type silicon the phase shift δ should have a finite k-dependent value. Schumann, Philips and Olshefski[3,4] present expressions from which δ as a function of λ is computed for N on N+ and P on P+ silicon, with substrate resistivities between 0.001 and 0.2Ωcm. The method they adopt for the analysis of infrared spectra is as follows. From the key formula

$$2 \text{ dnk} = \ell - \frac{1}{2} + \frac{\delta}{2\pi} \tag{2}$$

at two different, not too close, maxima λ_1, ℓ_1 and λ_2, $\ell_2 = m + \ell_1$ the order is found

$$\ell_2 = \frac{mk_2}{k_2 - k_1} + \frac{1}{2} - \frac{k_2\delta(k_1) - k_1\delta(k_2)}{k_2 - k_1} \tag{5}$$

Using then the appropriate computed δ (k) relation in equation (2), for every maximum a value of d is found which, after averaging, is supposed to give the true thickness. This procedure has been accepted by the American Society for Testing and Materials[5] as a tentative method pending adoption as a standard.

2.3. Experimental Results on the Phase Shift δ and the Reflection Coefficient $\hat{r}_{12}$

If the simple theory discussed in the preceding section and the ASTM procedure based on this theory hold true, it should be possible to find the relation δ (k) experimentally. According to equation (2) maxima plotted in an ℓ versus k diagram should be on a straight line which intersects the ℓ-axis at $\ell = \frac{1}{2}$, if δ (k) = 0. Since according to theory δ should increase from zero to π with increasing wavelength, δ (k) = 0 holds only at high values of k. The difference between this straight line and the experimental ℓ versus k plot yields $\delta \equiv \delta$ (k).

The procedure described above has been applied to hundreds of spectra obtained from slices grown under a variety of circumstances by many different manufacturers. It has always been found that ℓ versus k for the extreme values forms a straight line which intersects the ℓ-axis at any value $+0.5 > l > -0.5$, in other words $0 < \delta_0 < 2\Pi$ and $\delta(k) \equiv \delta_0$ in equation (2). This results in in contrast to what workers have reported.

The value δ_0, independent of k, is generally found to be different from slice to slice, even when grown in the same batch, and at different spots on the slice. However, on the basis of a theory recently developed by Eversteyn et al[6] the growth of epitaxial layers of outstandingly constant thickness along the reactor has been achieved. In slices produced in this way δ_0 has been found to be fairly uniform over the slice and from slice to slice, as shown in Fig. 2.

It is surprising that although much attention has been paid in the literature to δ (k), the more conspicuous $\hat{r}_{12}$ (k) dependence has been neglected. In all interference spectra of epitaxial layer systems produced in the usual way the fringes disappear at wavelengths below about 10 μ. Since $\hat{r}_{01}^2 = 0.30$, this implies that $\hat{r}_{12}$ gets below about 0.01 as can be seen from equation (1a). With this equation $\hat{r}_{12}$ has been calculated as a function of the wave number k from spectra of a large number of slices produced under a variety of conditions. They all show an exponential dependence on the wave number according to

$$\hat{r}_{12} = \hat{r}_{12} \exp (-k d_0) \quad , \tag{6}$$

where the characteristic length d_o assumes any value between 40 and 100 μm and the long-wavelength value

r_{12}^0 between 0.1 and 0.3. Just like δ_0, this dependence cannot be explained by the classical expressions Their understanding should be based on specific properties of the epitaxial layer system.

It can easily be shown that equations (4) and (6) can be combined to yield a constant δ^* such that with $\delta(k)$ δ_0

$$\tan \delta^* = \frac{4\pi dn}{d_0} \tag{7}$$

It would be tempting to identify in equation (3b) the angle δ^* with the wavelength-independent phase shift $\delta = \delta_0$. However, $\delta_0/2\pi$ has been found to cover the range 0 to 1, whereas $\delta^*/2\pi$ can cover the range 0 to 0.25 only. In other words, δ^* whould be added to the measured δ_0 in order to get δ.

2.4. Conclusions and Discussion

On the basis of the experimental observations described above an alternative procedure for epitaxial layer thickness measurement is suggested.

The k-values of maxima and minima are read from a spectrum. Then it turns out to be possible to draw a straight line through the extrema plotted on ℓ versus k paper (in fact the maxima are plotted and the minima are inserted). The thickness can be determined by combining two, not too close, values k_1, ℓ_1 and k_2, $\ell_2 = \ell_1 + m$ to give

$$d = \frac{m}{2n(k_2 - k_1)} \quad . \tag{8}$$

All extrema can be combined with one of them and the values of d obtained in this way are averaged.

In order to obtain the highest precision available the thickness d may be calculated using the method of least squares with a Hewlett-Packard 9100 calculator. The high correlation of the data, typically 0.9998, shows that the straight-line approximation is correct. The value of 2 dn k yields for maxima $\ell - \frac{1}{2} + \delta_0/2\pi$, and for minima $\ell + \delta_0/2\pi$, from which δ_0 can be found. If a mistake had been made in assessing d, this should be manifested by a monotonic increase or decrease in δ_0 with k, as can be seen from equation (2).

In this contribution facts have been interpreted. Detailed experimental evidence on a number of published spectra and a discussion of various physical models which may explain the observations consistently will be published elsewhere.

3. Measurement of Resistivity of Silicon by the Spreading Resistance Method

3.1. Introduction

It appears that the spreading resistance method for the resistivity evaluation of semiconductor material, in particular silicon, is becoming increasingly popular[7,9]. In the following two sections our own experiments are presented which are characterized by the use of a multi-tipped probe as an electrical contact. This necessarily implies that the spreading resistance is composed of two series resistances with shallow and with deep information sampling depths.

3.2. Experimental Results on the Probe

The experiments described in this section were performed with a probing apparatus of the following description. Two specially hardened steel probes obtained from an ordinary Dumas four-point probe head were mounted on long arms and loaded with about 200 gm. Designating one probe as common and the other as the current probe, data were taken by recording the current flow through the probes required for a potential drop of 10 mV between the common probe and the bottom contact. The current flow can be recorded on a linear or on a four-decade logarithmic scale.

Microscopic inspection of the marks made by the probe on a finely ground surface revealed that it chips off only part of the surface in a 25μm diameter print. On a lapped and subsequently etched surface which was not exactly perpendicular to the probe axis, a number of fine scratches could be seen about 10μm long and 1 μm wide. Apparently they are caused by protrusions on the probe which produced many microcontacts of a radius smaller than 1 μm.

It has been found that when the pressure is not high enough, the contact resistance increase and is dominated by the Schottky barrier diode. It is clear that a number of small etch pits must be due to low-pressure and hence small-diameter contacts. These branches should consist of a zero bias barrier resistance and large spreading resistance in series. They can be neglected when compared with the high-pressure, large-area spots which apparently yield spreading resistance-dominated branches. For simplicity these are supposed to be equal and caused by contact spots πa^2 in surface area. The number n is determined by counting the number of scratches or the number of large etch pits: $n \approx 15$ and $a \approx 1$ μm. The total spreading resistance is now given by:

$$R_s = \frac{\rho}{4na} \tag{9}$$

In this expression it is assumed that the microcontacts act independently. If they do not, the relation[10]

$$R_s = \frac{\rho}{4na} \frac{2}{\pi} \tan^{-1} \frac{\sqrt{\ell^2 - a^2}}{a} \tag{10}$$

is valid, where 2ℓ is the mean distance between contact centers. The ratio ℓ/a is restricted to the range 1 to ∞; this correction amounts to 1% or 25% at $2\ell = 100\,a$ or $2\ell = 6a$, respectively. As can be seen from Figure 3, the contacts are rarely very close, and on the average $2\ell \approx 5$ μm $\approx 5a$.

The prints of the probe shown in Figure 3, show that the detailed pattern is reproduced apart from minute differences, probably due to statistical fluctuations, e.g., locally different surface conditions. These prints also show manifestly that the structure is in the probe and not in the surface. It is suggested that the high reproducibility of this system is due to the large number of microcontacts averaging out statistical fluctuations. That is also why the reproducibility on lapped material is smaller than on polished.

At a depth equal to about 2a the equipotential lines are horizontal and the individual microcontacts cannot be distinguished anymore. It is evident that this macrocontact is again subjected to the spreading resistance phenomenon. The contributions of the two components, to be referred to as the shallow and the deep stage, can be written with equation (10) as

$$R_s = \frac{\rho}{4A} \left[\frac{1}{\sqrt{n}} \frac{\ell}{a} \frac{4}{\pi^{3/2}} \tan^{-1} \sqrt{\frac{\ell^2}{a^2} - 1} + 1 \right] , \tag{11}$$

where by definition $\pi A^2 = 4n\ell^2$.
For independent microcontacts, $1 >> a$, equation (11) can be simplified to

$$R_s = \frac{\rho}{4na} + \frac{\rho}{4A} . \tag{12}$$

From equation (11) it can be seen that a and ℓ being constants, the shallow contribution becomes more important with smaller number n.

In order to assess the range of validity of the fundamental assumption $R_s \sim \rho$ and to find out the proportionality constant ρ/R_s for different materials and probes, a number of N-type Si slices between 2×10^{-3} and 20 Ω cm were used for calibration. The resistivity was measured by the four-point probe and via the capacitance-voltage method. It was found that for ρ between 2×10^{-3} and 10 Ω cm both N- and P-type Si yield $R_s \sim \rho$ with proportionality constants ρ/R_s equal to 40 and 70 μm, respectively. Since A is about 15 μm, na is of the same order and $n\ell^2 = 180$ μm^2. Numerical evaluation with $a = 1$ μm, $\ell = 3$ μm and $n = 20$ yields for the shallow contact term 0.55. Thus the value n, ℓ and a found microscopically are substantiated electrically. This calibration, shown in Figure 5, is valid for a surface preparation procedure consisting of lapping and HF/HNO_3 etching.

3.3. Experimental Results on the Semiconductor

A typical example of horizontal variation in resistivity obtained from a 10^{-2} Ω cm N-type Si slice as offered for epitaxy, is shown in Figure 6. Two parallel tracks at 100 μm distance are seen to be almost identical. Assuming that a smooth line should really be drawn through the measurement points, it is clear that the erratic signal representing the reproducibility amounts to about 1%. This is typical for chem-mechanically and prior to measurement HF/HNO_3 etched slices. It is evident that, since the four-point probe can hardly claim 5% reproducibility, within one four-point probe measurement area, several values for the spreading resistance are found, as shown in Figure 5.

The spreading resistance method is most powerful for investigating profiles in angle-lapped layered systems, in particular isotype and heterotype epitaxial layers. The bevel used in the angle-lapping procedure is 1:800 and hence the widths of the print (25 μm) and of a horizontal step (100 μm) correspond in depth to 0.03 and 0.12 μm, respectively. For correct conversion of horizontal scale to depth the slope obtained should be independently checked.

This kind of measurement has been done on a large variety of slices and some results are given in Figure 7. The reproducibility, as can be seen on the homogeneous part, is about 5%. In most isotype (N on N+) systems the profile turns out to extend over 1 and 2 μm, covered by about 10 readings. It is easy to understand, and it can be seen from Figure 8, that when the measurement starts at the substrate, the foot of the profile can be determined to within one step. The information sampling depth for the shallow stage is between one half and one micron and hence an abrupt transition would be plotted as a profile extending over this length. When several readings are taken within this length a gradual disappearance of the influence of the substrate will be registered. The resolving power in this arrangement is the length of the smallest profile that can be distinguished and is experimentally found to be about half a micron. This again substantiates the concept of a number of independent small microcontact spots of about that size. The double stage aspect of the spreading resistance probe can be verified when the thickness of the epitaxial layer d < A. Due to the logarithmic representation the deep stage can hardly be seen with N on N+ profiles, but for the same reason it is very conspicuous with N+ on N or N on P structures. In an angle-lapped PN system a steep rise in resistivity is found from the substrate upwards, peaking three or four orders of magnitude in excess of the substrate value. The resistance then decays slowly through the layer. The thin layer series resistance can be separated from the deep contact contribution by varying the distance between the two probes.

3.4. Discussion

The double stage aspect is not new: it has been described for thermal contacts on a rough surface in a similar way.

The complicated physical nature of the probe makes calibration of each new probe and check slice tests at regular intervals absolutely necessary. Generally, the density of microcontacts at the edge of the contact is greater than in the center initially. During aging of the probe the density becomes more uniform. The accuracy of the measurement may be affected by secular variation of R_s/ρ during the measurement also. This effect can be offset again by etching the slice. The mechanical interaction probe surface-silicon surface appears to be of decisive importance for the accuracy of the measurement. This is being investigated.

4. References

[1] American Society for Testing and Materials, Use of the terms precision and accuracy as applied to measurement of a property of a material, E177-61T.

[2] Albert, M.P. and Combs,J. F., Thickness of epitaxial films by the infrared interference method, J. Electrochem. Soc. 109 709 (1962).

[3] Schumann, P. A., Phillips, R. P. and Olshefski, P. J., Phase shift corrections for infrared interference measurement of epitaxial layer thickness, J. Electrochem. Soc. 113 368 (1966).

[4] Schumann, P. A., The infrared interference method of measuring epitaxial layer thickness, J. Electrochem. Soc., 116 409 (1969).

[5] American Society for Testing and Materials, Thickness of epitaxial layers of silicon on substrates of the same type by infrared reflectance, F95-68T.

[6] Eversteyn, F. C., Severin, P.J., Brekel, C.H.J. van der, and Peek, H.L., A stagnant layer model for the epitaxial growth of silicon from silane in a horizontal reactor, J. Electrochem.Soc., submitted for publication.

[7] Mazur, R. G. and Dickey, D. H., A pareading resistance technique for resistivity measurements on silicon, J. Electrochem. Soc. 113 255 (1966).

[8] Keenan, W. W., Schumann, P. A., Phillips, R. P. and Tong, H. H., A model for the metal-semiconductor contact in the spreading resistance probe, Ohmic contacts to silicon (Ed. Schwartz, B.), The Electrochemical Society, 1969, p. 263.

[9] Gupta, D. C., Determination of mobility and its profile in n/n+ silicon epitaxial layers, J. Electrochem. Soc. 116 670 (1969).

(10] Holm, R., Electric Contacts, Springer, Berlin, 1967.

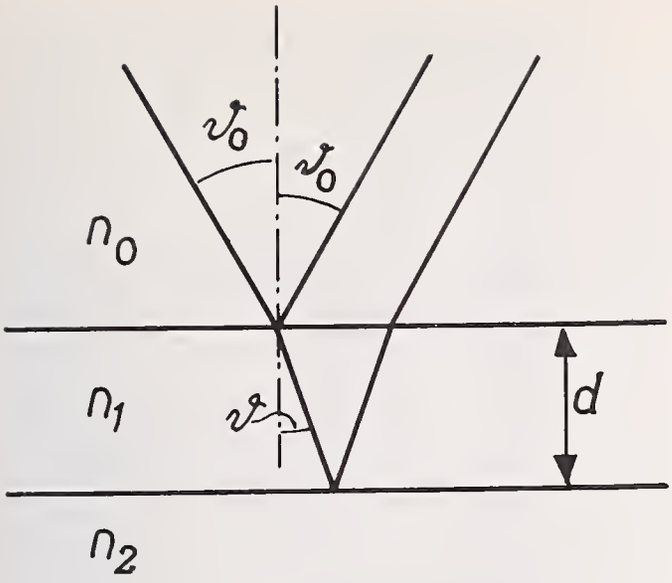


Fig. 1. Definition of the geometry used.

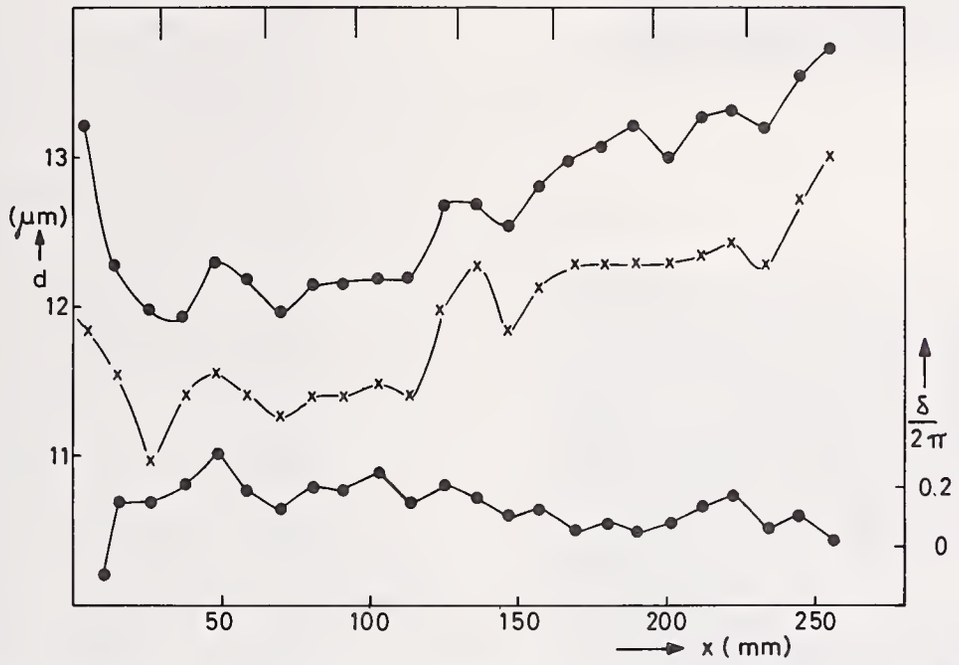


Fig. 2. Thickness d and wavelength-independent phase shift δ_0 as a function of position along the susceptor (top and lower curve). Middle curve is thickness measured by bevel and stain.

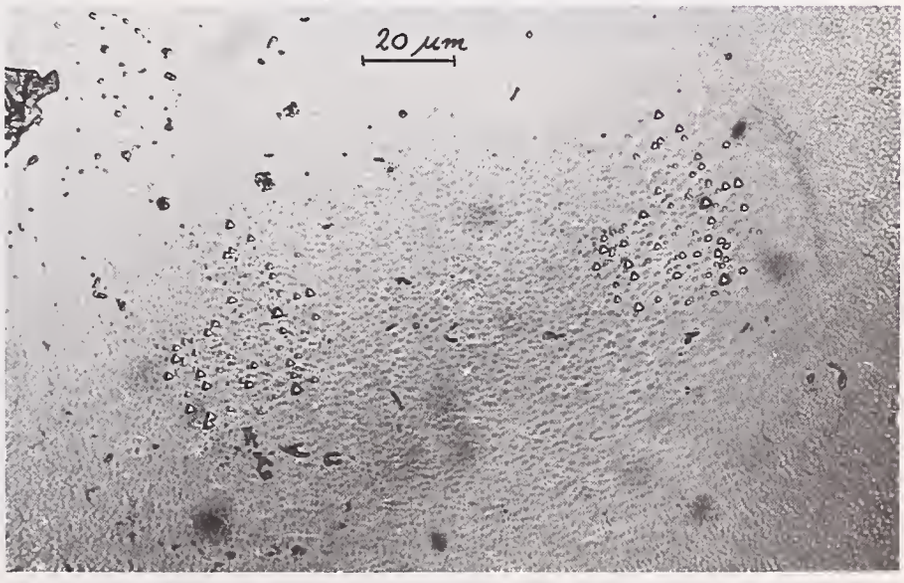


Fig. 3. Two prints of the same probe at 100 μm distance on a chem-mechanically polished and subsequently etched 10^{-2} Ω cm N-type Si slice.

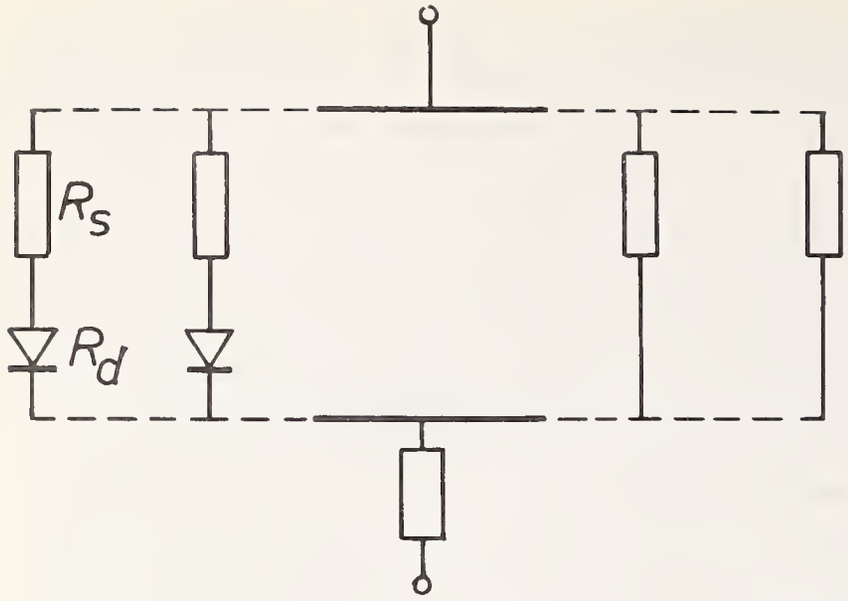


Fig. 4. The equivalent circuit, as deduced in the text. The diode-controlled branches are ineffective because the spreading resistance-controlled branches are more conductive. The series resistance stands for the deep component.

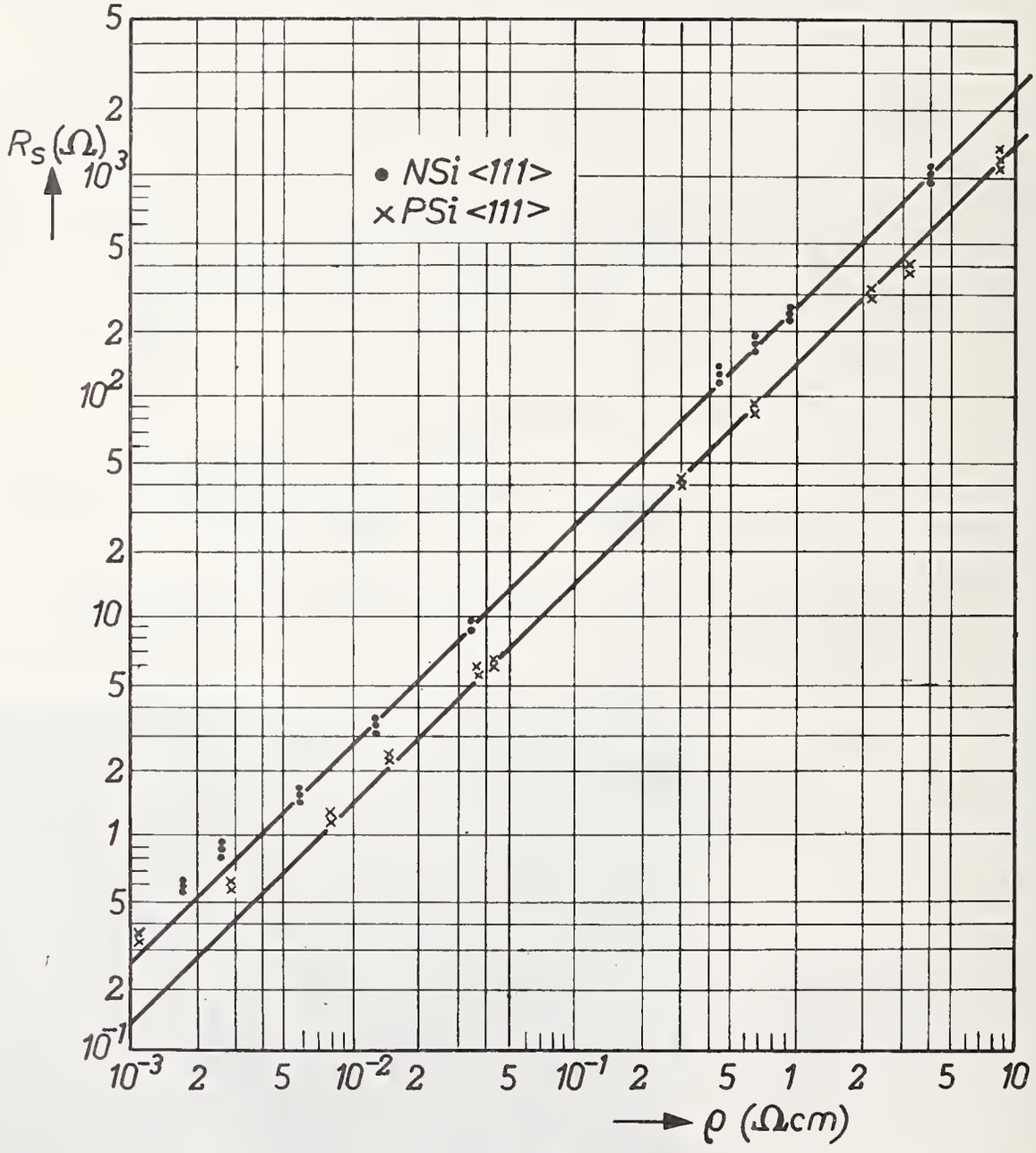


Fig. 5. Calibration curves for N- and P-type Si.

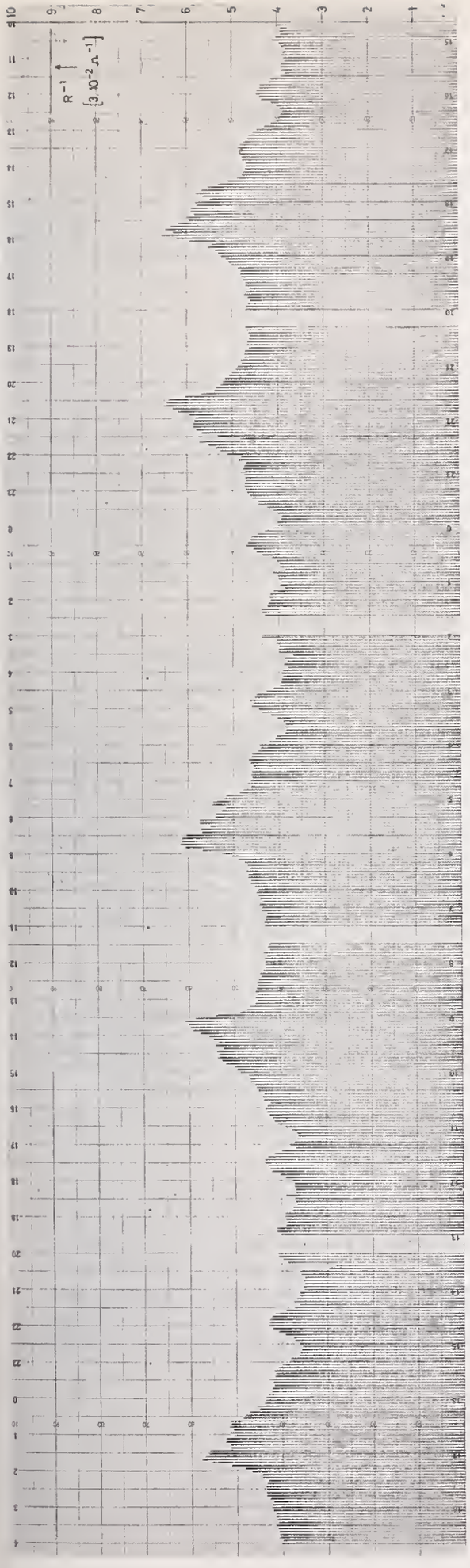
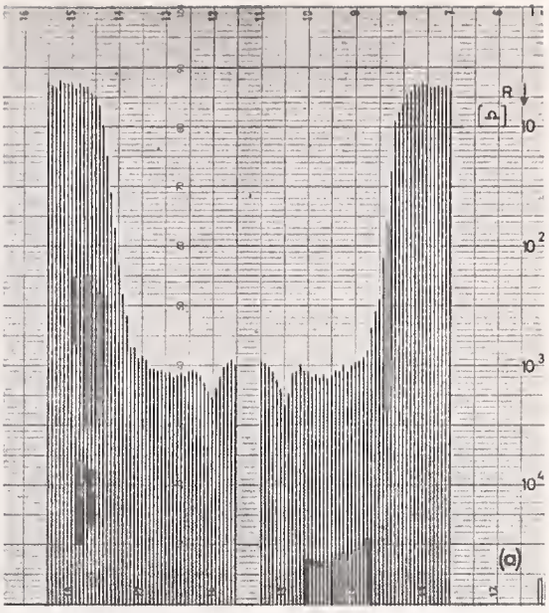


Fig. 6. Horizontal variation in resistivity on 0,02 Ω cm N-type Si slice, chem-mechanically polished and subsequently etched. The same track has been produced five times and the reproducibility is seen to be about ± 1%. Horizontal scale: 100 µm/ step, vertical scale 0,03 Ω^{-1}/maj. div.



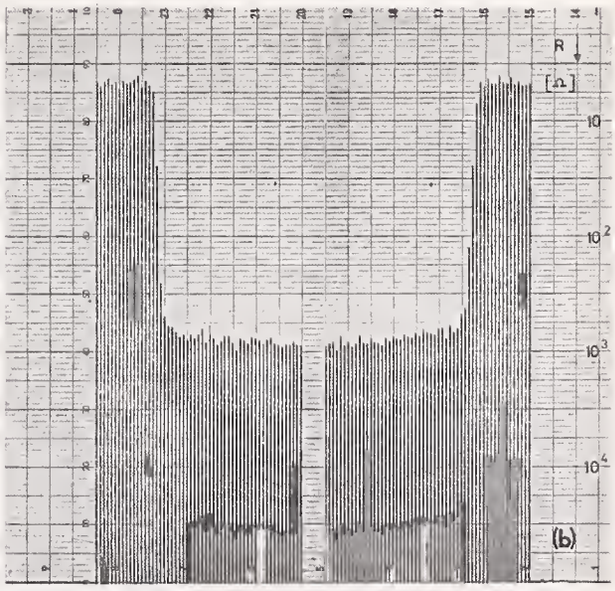


Fig. 7. Epitaxially grown N on N^+ layers, 1:800 angle-lapped, 1 Ω cm N on 10^{-2} Ω cm N^+,(a) with HCl etch, (b) without. The same track is produced twice on a logarithmic scale.

Thickness Measurement of Very Thin Epitaxial Layers by Infrared Reflectance

P. A. Schumann, Jr.

IBM Components Division
East Fishkill Laboratory
Hopewell Junction, N.Y. 12533

A review is given of the development of infrared reflectance for measuring the thickness of a semiconductor epitaxial layer deposited on a substrate of the same conductivity type but of higher impurity concentration. Various techniques of interpreting the interference spectra are discussed and their limits of applicability given. It is shown that with recent techniques, measurements can be made on layers as thin as 0.3μm. Experimental comparisons are presented with bevel and stain, where possible, and spreading resistance. A new method of interpreting the interference spectra is discussed which enables measurement of both the layer thickness and substrate carrier concentration. It is shown that this may be extended to provide a carrier concentration profile as a function of depth of a portion of the out-diffused region of the epitaxial layer.

Key Words: Bevel and stain, carrier concentration, epitaxial layer, infrared reflectance, interference spectra, spreading resistance, substrate carrier concentration, thickness measurement.

A lightly doped silicon epitaxial layer deposited on a highly doped single crystal silicon substrate of the same conductivity type is a common structure in the semiconductor industry for planar devices or integrated circuits. The change in carrier concentration from the epitaxial layer to the substrate produces a corresponding change in the infrared optical constants. The transition of the optical constants is rapid enough that a reflection can occur. First to make use of this information to measure the epitaxial layer thickness were Spitzer and Tannebaum in 1961.[1]

A simplified schematic of the geometry of the infrared interference measurement is shown in Fig. 1. The multireflected beam is omitted for clarity. Nonpolarized infrared radiation is incident on the epitaxial layer at a small angle--usually less than 30°. The high index of refraction of the lightly doped epitaxial layer makes the angle of incidence at the epitaxial layer-substrate interface very near normal. The infrared radiation transmitted through the layer and reflected back interferes with radiation reflected off the air - epitaxial layer interface. When the wavelength is varied and the total reflectivity measured, a series of interference fringes is observed. The information in this interference spectrum can be used to calculate the layer thickness.

With three simplifying assumptions, an approximate expression for the thickness can be derived. First, it is assumed that the phase change or reflection at the epitaxial layer - substrate interface is independent of wavelength. Second, it is assumed that the reflectivity at the epitaxial layer - substrate interface is not a strong function of wavelength. Both of these assumptions are necessary so that the maxima and minima observed can be described by a simple phase relationship between the wavelength positions of the extrema and an integer or half integer multiple of 2π. Third, it is assumed that the depth profile of impurities is sharp enough that the layer-substrate interface may be treated as though abrupt. For this case the thickness of the epitaxial layer h is given by

$$h = \frac{m(\lambda_1\lambda_2)}{2(n_1^2 - \sin^2\theta)^{1/2}(\lambda_1 - \lambda_2)} \qquad (1)$$

where $m = p_2 - p_1$, $\lambda_1 > \lambda_2$, λ_1 and λ_2 are wavelength positions of extrema, n_1 is the index of refraction of the epitaxial layer, and θ is the angle of incidence of the infrared radiation. The orders of the fringes p_1 and p_2 need not be known, as m is merely the number of fringes between λ_1 and λ_2. This expression, however simple, has limited application, primarily for epitaxial layers greater than 10μm thick. Its use is improved if the tangent points of the reflectivity envelope are used rather than the actual maxima or minima, as pointed out by Albert and Combs.[2]

However, it is still limited because the wavelength dependence of the phase shift at the epitaxial layer - substrate interface has been neglected. This was pointed out by Schumann, Phillips, and Olshefski[3] in 1966 and by Bilenko[4] et al. in 1969. If it is assumed that the epitaxial layer is lightly doped, so that the extinction coefficient k_1 of the epitaxial layer is zero, the phase shift ϕ at the abrupt epitaxial layer - substrate interface is

$$\phi = \tan^{-1} \frac{2n_1 k_2}{n_1^2 - n_2^2 - k_2^2} \tag{2}$$

where k_2 is the extinction coefficient of the substrate, and n_2 is the index of refraction of the substrate.

Two expressions, Eqs. (3) and (4), can be used to calculate the thickness,

$$h = \frac{m\lambda_1\lambda_2}{2(n_1^2 - \sin^2\theta)^{1/2}(\lambda_1 - \lambda_2)} \left[1 - \frac{\phi_1 - \phi_2}{2\pi m}\right], \tag{3}$$

where ϕ_j is the phase shift at the layer-substrate interface of the jth fringe. This expression is similar to Eq. (1) except for the correction factor involving the phase shifts. It is easy to see that if the phase is not a function of wavelength, then $\phi_1 - \phi_2 = 0$ and there is no correction. It is also true that if m is large, i.e., thickness large, the effects of the phase correction will be reduced. A complementary expression for calculating the thickness from each fringe or an individual fringe for thin layers, is

$$h = \left(p_j - \frac{1}{2} + \frac{\phi_j}{2\pi}\right) \frac{\lambda_j}{2(n_1^2 - \sin^2\theta)^{1/2}} \tag{4}$$

where p_j is the order of the jth fringe. The order can be calculated from

$$p_j = \frac{m\lambda_{j-m}}{\lambda_{j-m} - \lambda_j} + \frac{1}{2} - \frac{\phi_{j-m}\lambda_{j-m} - \phi_j\lambda_j}{2\pi(\lambda_{j-m} - \lambda_j)} \tag{5}$$

where $\lambda_{j-m} > \lambda_j$.

The optical constants, index of refraction, and extinction coefficient of the substrate must be calculated. This was done by using a semiclassical model of free carrier absorption and compared for accuracy by using the reflectivity minima associated with plasma resonance.[5] The resulting optical constants are

$$n = \frac{1}{2\sqrt{2}} \left[K_L - \frac{e^2\lambda^2 NJ(D)}{4\pi^2\epsilon_o C^2 m^*}\right]^{1/2} \left\{1 \pm \left[1 + \frac{e^8\lambda^6 N^4 \rho_o^2 g^2 [L(D)]^2}{64\pi^6\epsilon_o^2 C^6 m^{*4} \left[K_L - \frac{e^2\lambda^2 NJ(D)}{4\pi^2\epsilon_o C^2 m^*}\right]^2}\right]^{1/2}\right\}^{1/2} \tag{6}$$

$$k_2 = \frac{1}{\sqrt{2}} \left[K_L - \frac{e^2 \lambda^2 NJ(D)}{4\pi^2 \epsilon_o C^2 m^*} \right]^{1/2} \left\{ -1 \pm \left[1 + \frac{e^8 \lambda^6 N^4 \rho_o^2 g^2 [L(D)]^2}{64\pi^6 \epsilon_o^2 C^6 m^{*4} \left[K_L - \frac{e^2 \lambda^2 NJ(D)}{4\pi^2 \epsilon_o C^2 m^*} \right]^2} \right]^{1/2} \right\}^{1/2} \tag{7}$$

where

$$J(D) = \frac{1}{\Gamma(5/2)} \int_0^\infty \frac{x^{9/2} e^{-x} dx}{x^3 + D} \qquad D = \frac{N^2 e^4 \lambda^2 \rho_o^2 \Gamma(4) g}{4\pi^2 C^2 m^{*2}}$$

$$L(D) = \int_0^\infty \frac{x^3 e^{-x} dx}{x^3 + D} \qquad g = \frac{\Gamma(4)}{[\Gamma(5/2)]^2}$$

and N is the carrier concentration of the substrate, ρ_o is the dc resistivity of the substrate, m* is the conductivity effective mass, e is the charge on the electron, K_L is the dielectric constant of intrinsic silicon at long wavelengths, and C is the velocity of light. All are in MKS units. D is a convenient parameter in the integrals.

An example of the phase shift calculated from these expressions is shown in Fig. 2. Tabulations of the phase shift are also available.[6,7]

The infrared interference technique in which phase shift corrections are used was adopted by the American Society for Testing and Materials in 1968.[8] The multilaboratory three sigma precision utilizing this technique for silicon layers thicker than 2μm is $\pm$ (0.25μm + 0.025 h) for p-type samples and $\pm$ (0.25μm + 0.005 h) for n-type samples where h is the layer thickness in micrometers. Although the precision study was carried out to 2-μm thickness, additional errors can occur below 5μm because of the assumption concerning wavelength dependence of the layer-substrate interface reflectivity.

For thin epitaxial layers, the difference between the wavelength positions of the extrema and the tangent point of the reflectivity envelope can be significant. Since the tangent point is much more difficult to determine than the maximum or minium, calculations were made of the reflectivity of the abrupt epitaxial layer-substrate system and the positions of the extrema located numerically.[9]

The reflection from the abrupt system shown in Fig. 1 if near normal incidence, $\theta \leq 30^o$, is assumed, is

$$R = \frac{r_1^2 + r_2^2 - 2r_1 r_2 \cos(\delta - \phi)}{1 + r_1^2 r_2^2 - 2r_1 r_2 \cos(\delta - \phi)} \tag{8}$$

where

$$r_1^2 = \left(\frac{n_1 - 1}{n_1 + 1} \right), \quad r_2^2 = \frac{(n_2 - n_1)^2 + k_2^2}{(n_2 + n_1) + k_2^2}, \quad \text{and}$$

$$\delta = \frac{4\pi n_1 h}{\lambda} \left(1 - \frac{\sin^2 \theta}{n_1^2} \right)^{1/2}.$$

A normalized thickness h' is defined as

$$h' = h\left(1 - \frac{\sin^2\theta}{n_1^2}\right)^{1/2} .$$

An example of this type of calculation is given in Fig. 3. Similar charts for other orders and a tabulation are given elsewhere.[10] The precision of this thickness chart technique is good, as shown in Table I. The precision was obtained by running the sample daily for 16 days and calculating the thickness daily.

Table I. Single instrument precision of thickness-chart technique for interpreting infrared interference spectra.

Average Thickness (μm)	Standard Deviation (μm)	Percent Standard Deviation (%)
2.23	$\pm$ 0.015	$\pm$ 0.68
4.67	$\pm$ 0.018	$\pm$ 0.38
8.75	$\pm$ 0.037	$\pm$ 0.42

The extension of the measurement of thin epitaxial layers[11] produces an interesting connection with the plasma resonance measurement technique,[5,12] as seen in Fig. 4. The higher order fringes disappear as the thickness goes to zero. However, the order one-half minimum becomes the plasma resonance minimum of the substrate. Shown for comparison on this theorctical curve is the experimentally determined plasma resonance minimum for that carrier concentration which is in good agreement with the calculation. This means that very thin layers can be measured. Figure 5 is an example of a thickness chart for very thin epitaxial layers. The use of this chart precludes a knowledge of the substrate carrier concentration. And, as can be seen, the measurement is very sensitive to this variable. The substrate carrier concentration is not always known, especially in diffused n/n+/p structures. Berman[13] has proposed an iterative scheme for gallium arsenide whereby both the thickness of the epitaxial layer and the carrier concentration of the substrate can be determined from the wavelength position of the order 0.5 minimum. This technique starts out by assuming that the order 0.5 minimum is the plasma resonance minimum, which corresponds to a first approximation to the substrate carrier concentration. Use is then made of the shift of the wavelength position of the order 0.5 minimum, which yields an approximate layer thickness. Then this value is used for a second-order calculation of carrier concentration. The substrate carrier concentration and layer thickness given by the second approximation usually are adequate.

An alternate technique has been proposed[14] that makes use of both the wavelength positions of the extrema and the corresponding reflectivity values. Calculations of the reflectivity were made with Eq. (9). The wavelength positions as well as the corresponding reflectance of the extrema were determined numerically. These values were plotted in a two-parameter graph such as Fig. 6. A determination of reflectance and wavelength position of the maximum uniquely determines both the layer thickness and substrate carrier concentration. This is true for all whole order maxima. The curves are double valued for the half order minima, as shown in Fig. 7. Here the curve is folded about zero reflectance and plotted as though it were negative reflectance. The reflectance-wavelength pair determines two possible thickness - carrier concentrations pairs. This discrepancy is usually eliminated if other orders are present but requires some prior knowledge if only one order is present. Note that for lower substrate carrier concentrations, a minimum is observable for thin layers at long wavelengths.

Figure 8 is an example of the use of this measurement technique on a very thin epitaxial n/n+/p structure. The carrier concentration profile was determined by the multilayer corrected[15] spreading resistance technique.[16] Shown on the curve is the plasma resonance value of carrier concentration determined before epitaxial deposition. The stain process was described earlier.[9] A step formed by masking with silicon oxide during growth is also shown. The values of thickness and carrier concentration marked IR were determined by the infrared reflectance technique. Note that IR values determined a point on the profile. Thicknesses as low as 0.3μm have been measured by this technique.

As can be seen in Fig. 8, it is difficult to define an epitaxial layer thickness in thin layers because of the graded nature of the carrier concentration profile. At this stage of development, the use of the infrared reflectance technique has removed assumptions one and two but still requires the abrupt interface model of the reflectivity.

Abe and Kato[17] were the first to consider the effects of the profile on the infrared interference spectrum, in 1965. They developed[18,19] a multilayer approximation system for a complementary error function out-diffusion of impurities and discussed its effect on the wavelength positions of the reflectivity extrema. Sato[20] et al. pointed out that the profile could be taken into account by an effective index of refraction of the epitaxial layer, which was lower than the intrinsic value. All of these papers pointed out that the extrema become less pronounced and shift in wavelength in a complicated manner as the graded region of the profile is increased.

Somewhat contrary to these theoretical calculations are the experimental data created by heat-treating epitaxial layers and remeasuring by the thickness chart technique.[9] It was shown that the gradient could be neglected if errors up to a maximum of 4% in thickness could be tolerated.

Tsunoda and Mori[21] pointed out that with a proper theory or empirical relationship, the problem could be turned around to measure the profile from the infrared interference spectrum. So far no method has been published which would allow this.

The infrared reflectance technique makes it possible to generate a portion of the out-diffused region of the epitaxial layer. It was observed that different orders yield different values of substrate carrier concentration and layer thickness. When plotted, these values agreed with corrected spreading resistance profiles, in spite of having used an abrupt interface model of the reflectivity in the model.

Table II shows an example of an n/n+ silicon sample with a layer thickness about 8.2μm and a substrate carrier concentration of 8.4×10^{19} cm^{-3}. Note the double values at the half order minima. Also note that some of the double values fell off the chart which are obviously unusuable. The comparison of the infrared reflectance profile and the corrected spreading resistance profile for this sample is shown in Fig. 9.

Table II. Values of substrate carrier concentration and layer thickness obtained for sample shown in Fig. 9.

Order	Reflectance	Wavelength (μm)	Thickness (μm)	Carrier Concentration (cm^{-3})	Thickness (μm)	Carrier Concentration (cm^{-3})
1.5	0.060	45.2	9.23	3.19×10^{19}	8.12	4.20×10^{18}
2.0	0.580	34.4	8.64	7.8×10^{18}		
2.5	0.118	26.3	9.20	1.02×10^{20}	8.35	7.0×10^{18}
3.0	0.462	22.1	8.72	7.8×10^{18}		
3.5	0.198	18.5	9.30	4.2×10^{20}	8.45	7.9×10^{18}
4.0	0.368	16.3	8.56	5.7×10^{18}		
4.5	0.248	14.2	off chart		8.48	5.63×10^{18}
5.0	0.338	12.5	8.35	5.2×10^{18}		
5.5	0.280	11.2	off chart		8.29	3.2×10^{18}

Figure 10 is a much thinner epitaxial layer but again the agreement is good between the two techniques of profiling. Precision studies were made on a sample similar to Fig. 10, with a thickness of about 1.8μm and a peak concentration of 5.5×10^{19} cm^{-3}. The average precision in depth over 8 readings for 6 orders was $\pm$ 1.1% for one standard deviation. The standard deviation in carrier concentration was $\pm$ 11%. This variation in precision is easily understood, as the carrier concentration determination depends primarily on the value of reflectance, which is more difficult to determine.

While the infrared reflectance technique is not completely understood, it apparently yields information regarding a portion of the out-diffused profile. This information results in spite of having used an abrupt interface model. Whether the effect observed is real or an accident of the experiment will remain unsolved until an adequate model of the graded interface is developed and applied in a manner that will reduce the problem.

This method of determining the profile is not meant to be complete. It should be used only as an example of the type of information obtainable from the infrared interference spectrum. Theoretical models for nonuniform epitaxial layers must be reduced to useful forms and applied in reverse in order to accurately determine the profile.

References

[1] W. G. Spitzer and M. Tannebaum, J. Appl. Phys. 32, 744 (1961).

[2] M. P. Albert and J. F. Combs, J. Electrochem. Soc. 109, 709 (1962).

[3] P. A. Schumann, Jr., R. P. Phillips and P. J. Olshefski, J. Electrochem. Soc. 113, 368 (1966).

[4] D. I. Bilenko, E. M. Ornatskaya and Z. V. Shekter, Pribory i Teknika Eksperimenta 3, 198 (1969).

[5] P. A. Schumann, Jr. and R. P. Phillips, Solid-State Electron. 10, 943 (1967).

[6] P. A. Schumann, Jr., IBM TR 22.182 (1965).

[7] P. A. Sehumann, Jr., IBM TR 22.432 (1967).

[8] American Society for Testing and Materials, "Method of Test for Thickness of Epitaxial Layers of Silicon on Substrates of the Same Type of Infrared Reflectance," F95-68T (1968).

[9] P. A. Schumann, Jr., J. Electrochem. Soc. 116, 409 (1969).

[10] P. A. Schumann, Jr., and C. M. Sakkas, IBM TR 22.537 (1968).

[11] P. A. Schumann, Jr., and C. P. Schneider, to be published in J. Appl. Phys.

[12] E. E. Gardner, W. Kapallo and C. R. Gordon, Appl. Phys. Let. 9, 432 (1966).

[13] R. Berman, Electrochemical Society, Boston, May 1968, Abs. 64.

[14] P. A. Schumann, Jr., Semiconductor Silicon, Electrochemical Society, N.Y. (1969), p. 662.

[15] P. A. Schumann, Jr., and E. E. Gardner, Solid-State Electron. 12, 371 (1969).

[16] J. M. Adley, M. R. Poponiak, C. P. Schneider, P. A. Schumann, Jr. and A. H. Tong, Semiconductor Silicon, The Electrochemical Society, N.Y. (1969), p. 721.

[17] T. Abe and T. Kato, Japan. J. Appl. Phys. 4, 742 (1965).

[18] T. Abe, Y. Nishi, K. Goto and M. Konaka, Daushi Tsushin Gakkai (Electron Communications Society), Tohoku, August 1967, Reference Material, Transistor Research Committee, Reference Number SSD67-16 (1967-08).

[19] T. Abe and Y. Nishi, Oyo Butsuri (Applied Physics) 36, 562 (1967).

[20] K. Sato, Y. Ishikawa and K. Sugawara, Solid-State Electron. 9, 771 (1966).

[21] Y. Tsunoda and O. Mori, J. Phys. Soc. Japan 27, 1077 (1969).

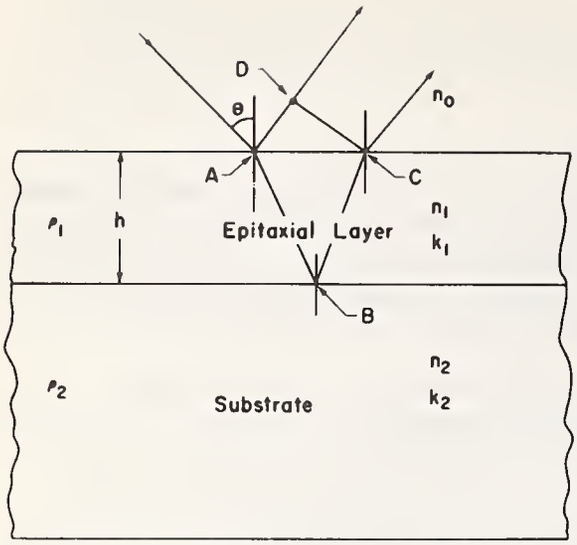


Fig. 1. Geometry of infrared interference technique.

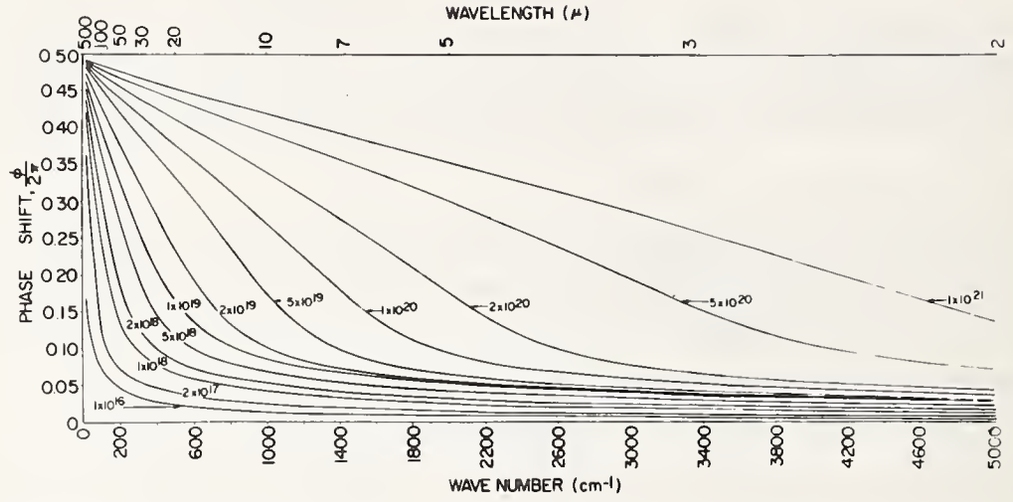


Fig. 2. Phase shift at epitaxial layer-substrate interface for n/n+ silicon structure.

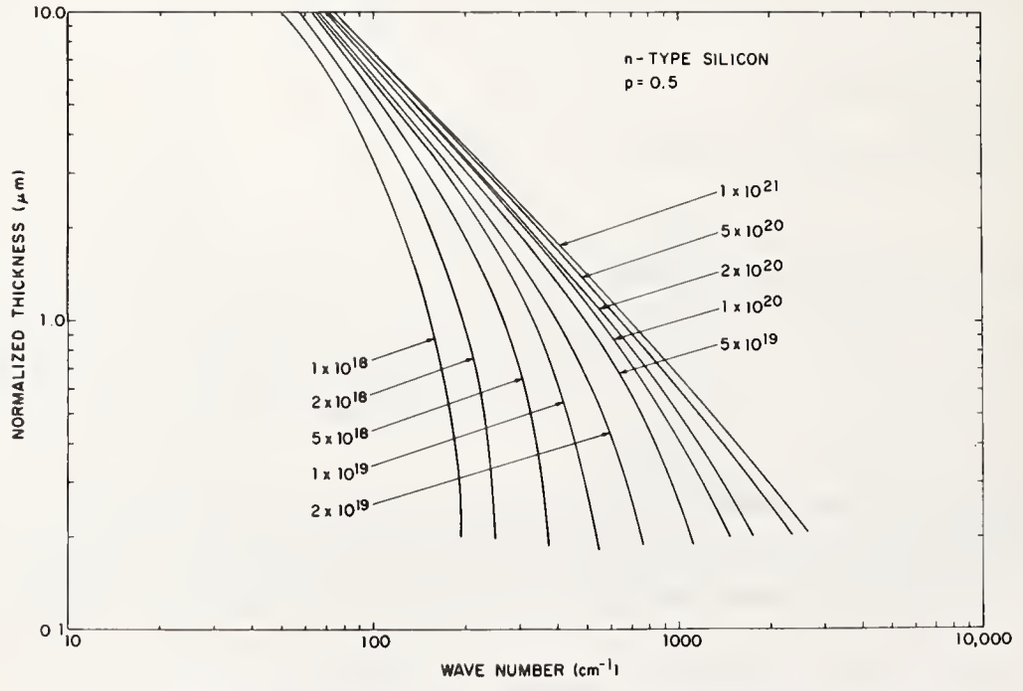


Fig. 3. Normalized thickness as a function of wavenumber, reciprocal of wavelength, for n/n+ silicon and order 0.5.

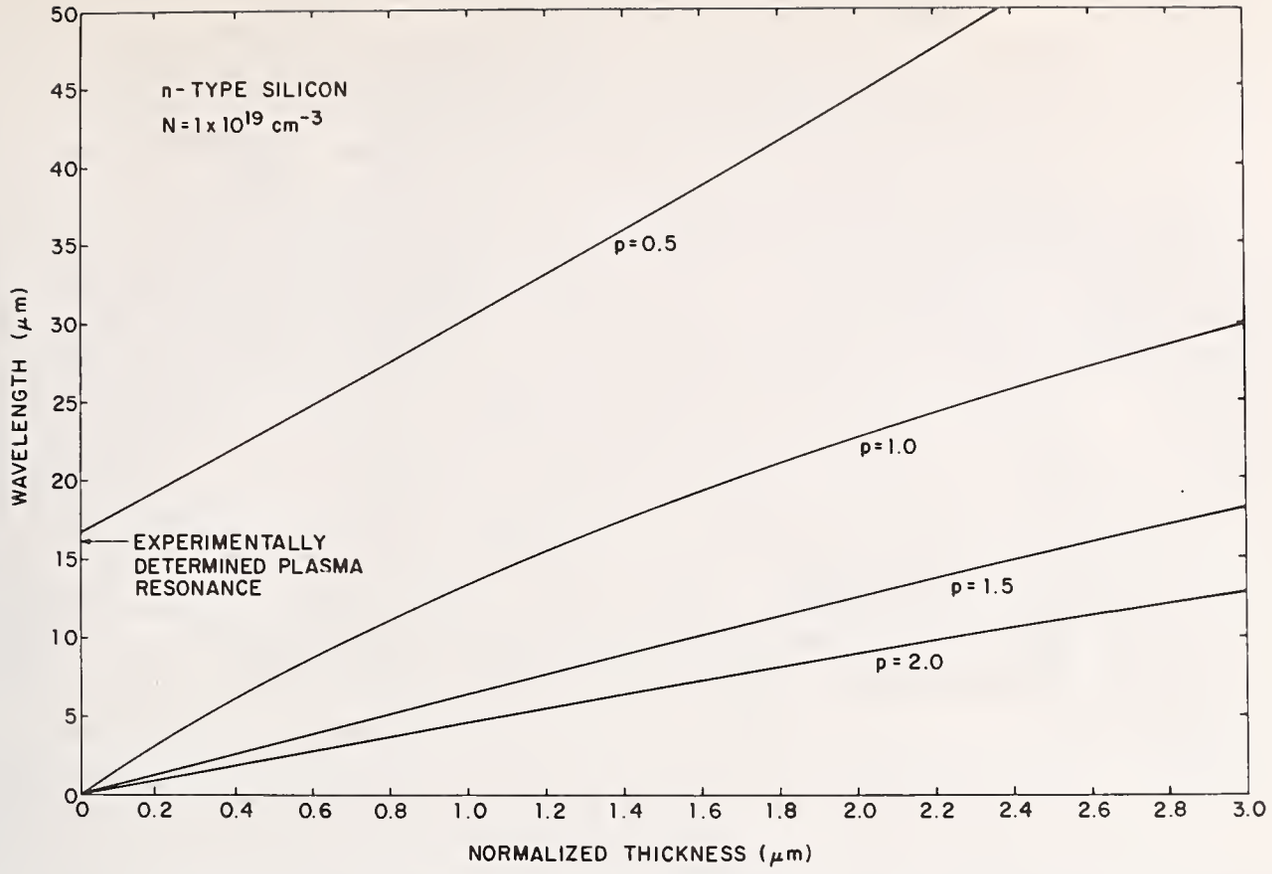


Fig. 4. Wavelength position of extrema as a function of normalized thickness for n-type silicon with a substrate carrier concentration of 1 x 10^{19} cm^{-3}; p is the order.

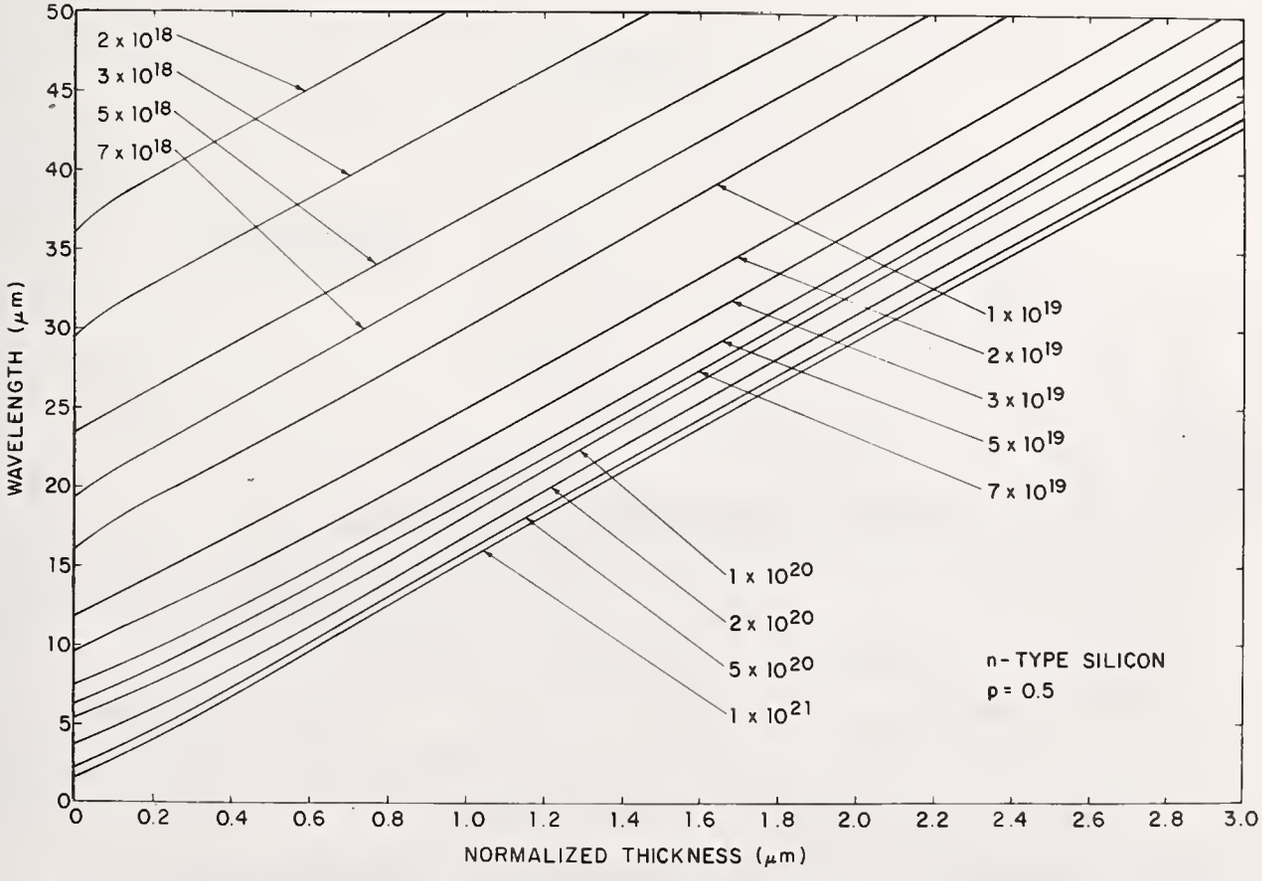


Fig. 5. Wavelength position of order 0.5 minimum as a function of normalized thickness for n/n+ silicon and various substrate carrier concentrations.

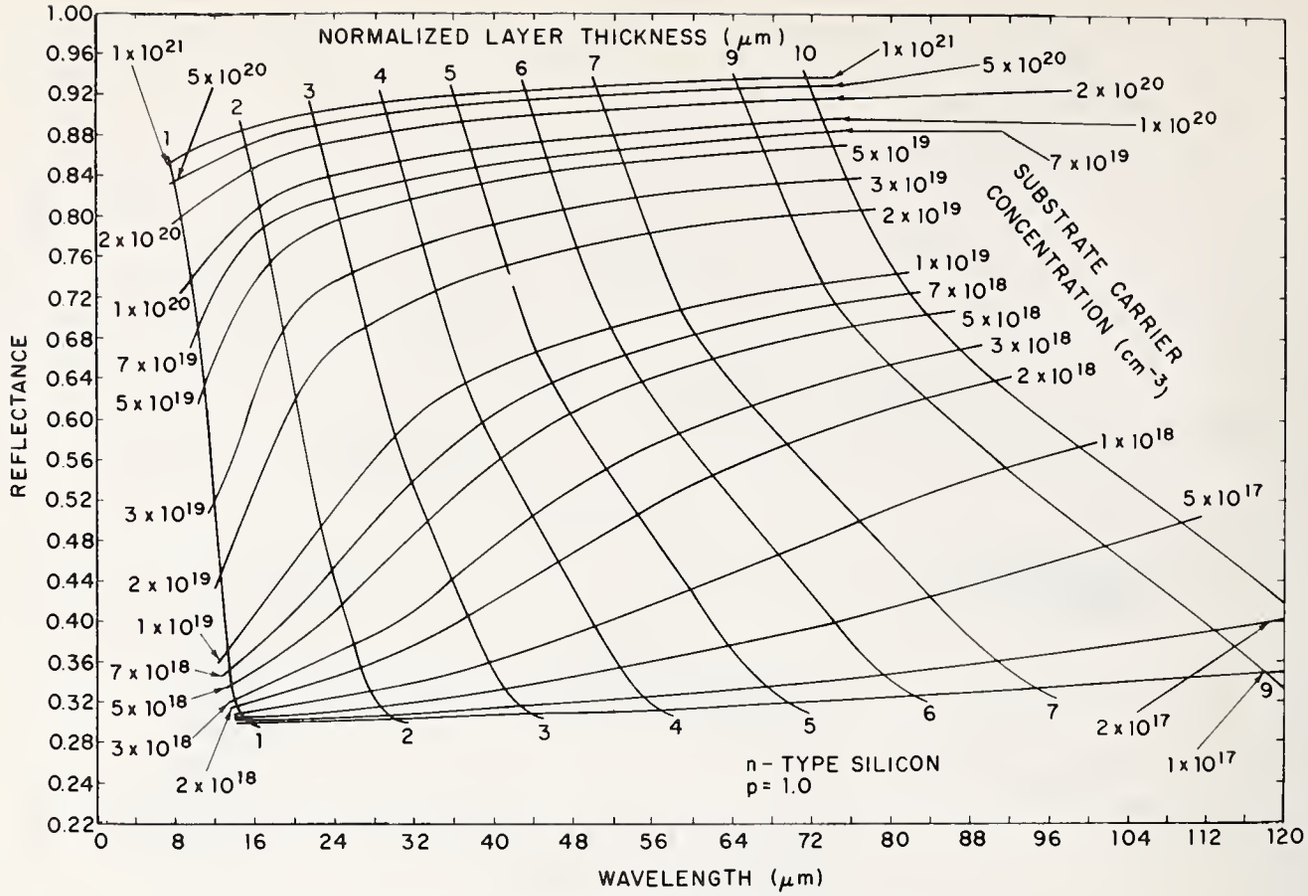


Fig. 6. Reflectance of order 1 maximum as a function of wavelength position of order 1 maximum for n-type silicon and various substrate carrier concentrations and normalized layer thicknesses.

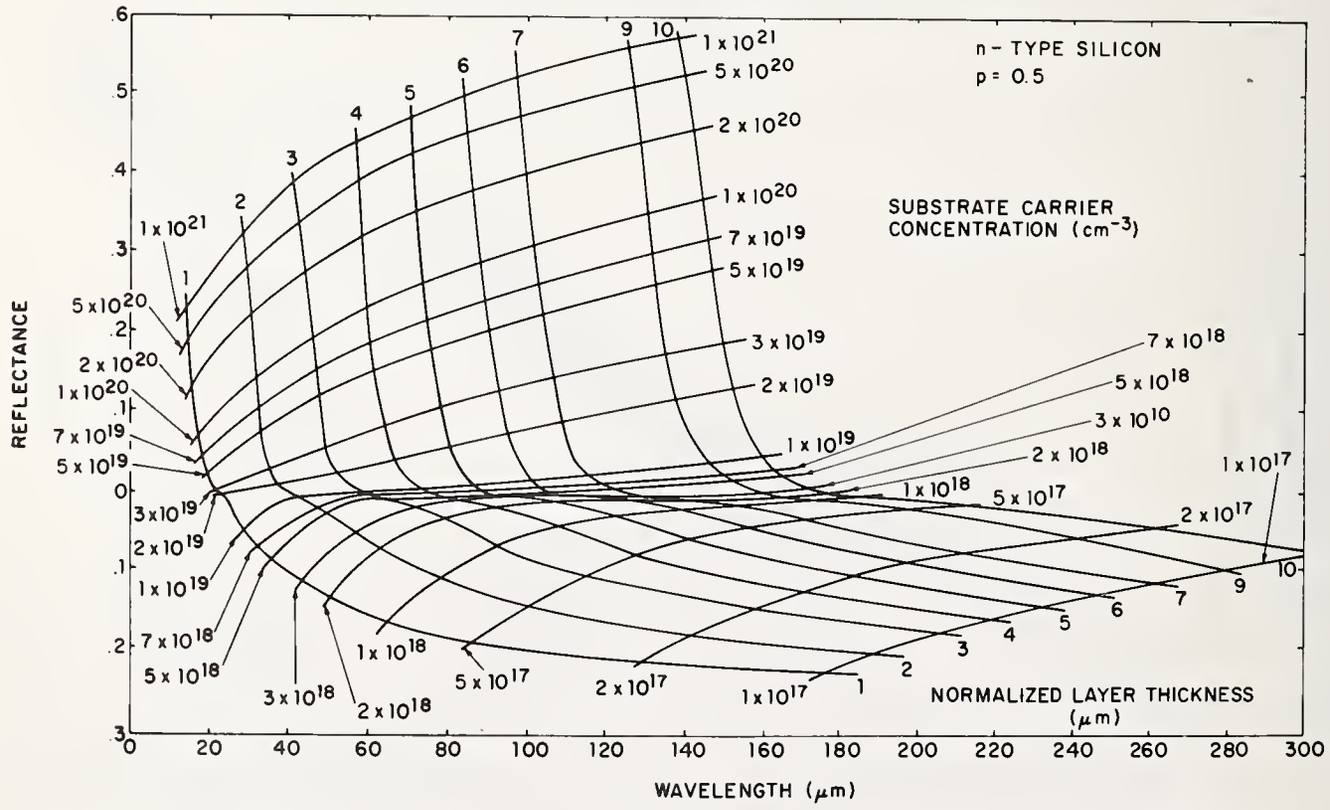


Fig. 7. Reflectance of order 0.5 minimum as a function of wavelength position of 0.5 minimum for n-type silicon and various substrate carrier concentrations and normalized layer thicknesses.

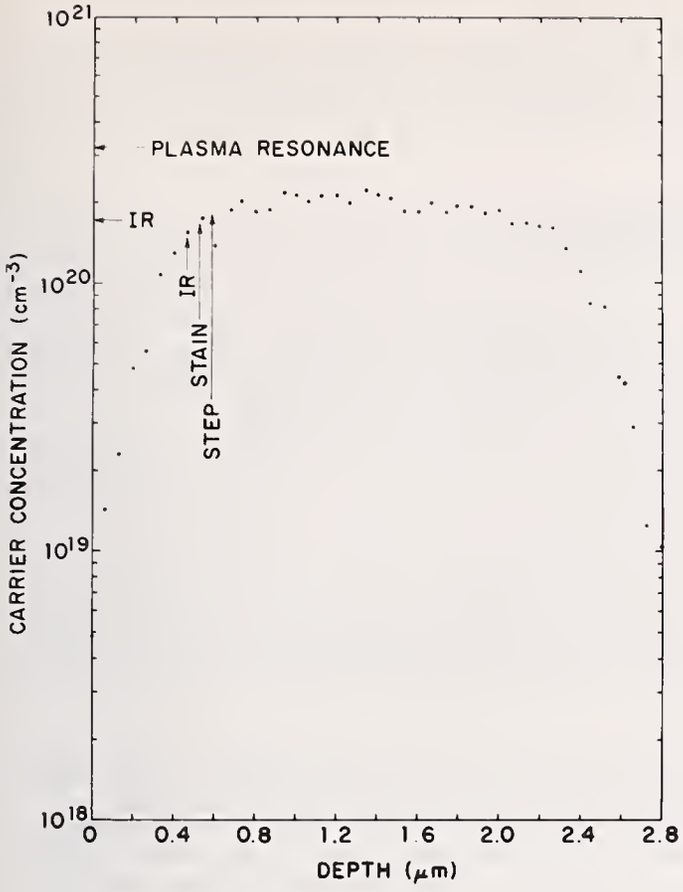


Fig. 8. Carrier concentration as a function of depth for a silicon n/n+/p structure showing the positions of various measurement techniques. The profile was determined by spreading resistance.

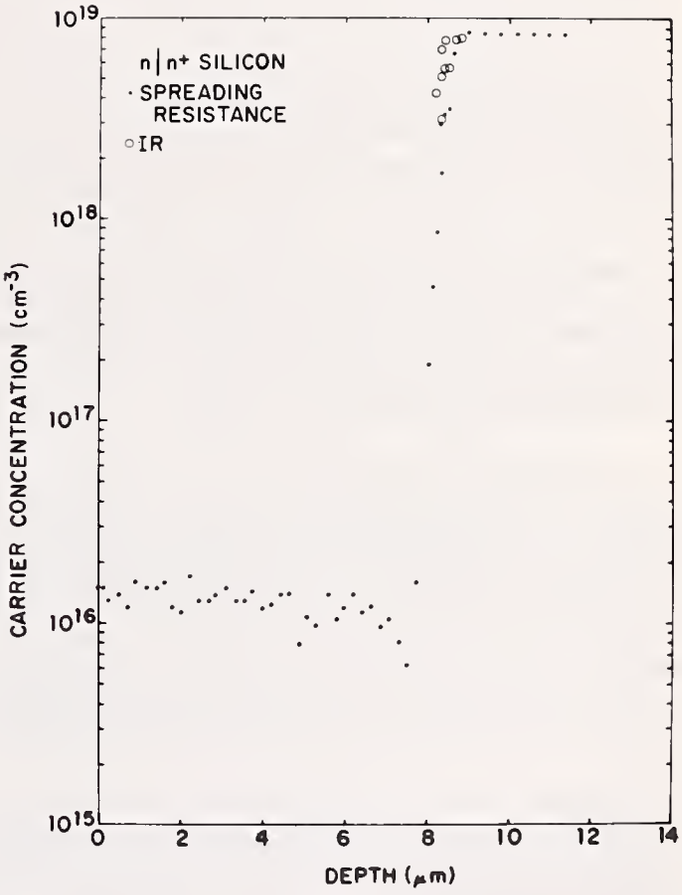


Fig. 9. Carrier concentration as a function of depth for silicon n/n+ structure showing the profile obtained by spreading resistance and that obtained by the infrared reflectance technique (IR)

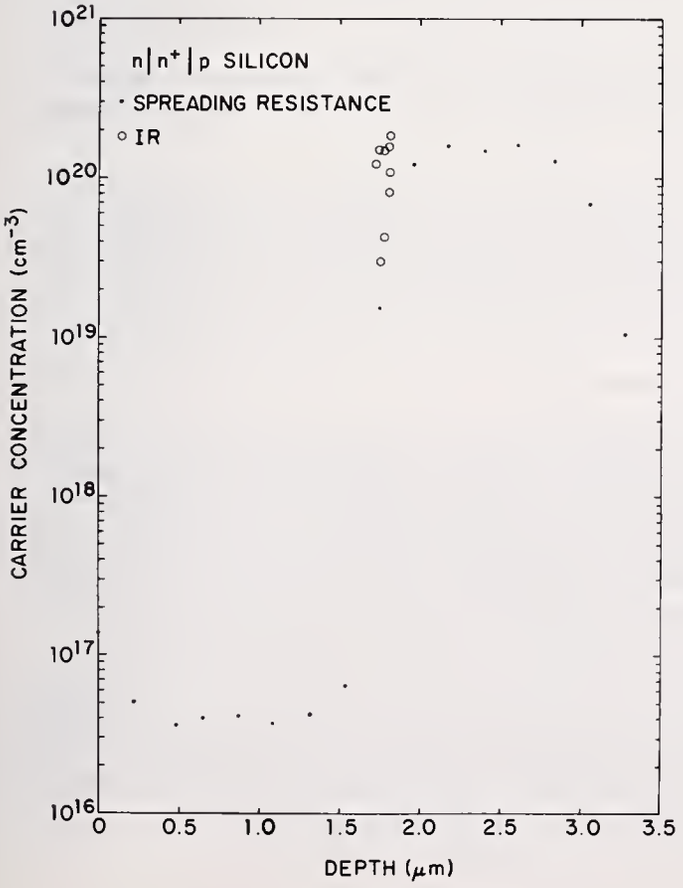


Fig. 10. Carrier concentration as a function of depth for silicon n/n+/p structure showing the profile obtained by spreading resistance and that obtained by the infrared reflectance technique.

Spreading Resistance Measurements On Buried Layers in Silicon Structures

R. G. Mazur
Westinghouse Research Laboratories
Pittsburgh, Pennsylvania 15235

The uses and limitations of the spreading resistance technique for the evaluation of buried collector layer structures in integrated circuits are discussed. Detailed results are presented on one particular type of structure to illustrate the applicability of spreading resistance measurements to buried layer problems.

1. Introduction

Buried layer structures have been extensively used in silicon integrated circuit technology for several years. These structures are used to lower collector resistance in integrated circuits, and combine locally-diffused regions (or "tubs") with epitaxial layer overgrowths of the same conductivity type. The most common arrangement embodies a p-type substrate with locally-diffused n^+ regions and an n-type epitaxial layer, as shown schematically in Fig. 1.

The evaluation of the all-important electrical properties of the sub-surface diffused regions and the epitaxial layer is very difficult. Even the limited information provided by the standard four-point-probe resistivity technique is not readily obtainable due to the effect of the sub-surface diffused layers on the overall sheet resistance of the n-type material. In addition, if the process parameters are to be properly controlled for device production, it is necessary to have the complete carrier concentration profile in the structure.

This buried layer measurement problem results from the limited spatial resolution of the standard methods for determining semiconductor resistivity or doping density. There is no practical routine method to overcome the lack of isolation between the heavily-doped sub-surface diffused regions and the epitaxial layer. Generally, attempts are made to control the electrical properties of these structures by monitoring the sheet resistance and junction depth of the n^+ diffused layer prior to epitaxy and by checking the resistivity of the epi-layer with control wafers (p-type substrates without integral n^+ diffused regions) placed in the reactor during the n-type epi-growth. These methods have their limitations. It is well-known that some dopant removal takes place during the slice cleaning step in the epi-reactor just before deposition, leading to "autodoping" of the epitaxial layer during its growth. The result is that often neither the dopant left in the sub-surface n^+ regions nor the dopant incorporated into the n-type epitaxial region above the buried layers is well-known.

Because of the high spatial resolution capability of the spreading resistance technique [1-6],[1] quantitative measurements of resistivity (or carrier concentration) on the standard nn^+ buried layer structures are possible. This paper will discuss these measurements and will present results obtained on several such structures. These results will show that spreading resistance measurements can be successfully used to control the important processes used in producing buried layer structures.

2. Experimental Results

All measurements were made with an automatic spreading resistance probe, using either a two-probe arrangement (with both probes set at the same depth in the sample during thickness profile measurements) or a single probe with a large-area ultrasonically-soldered current return contact. The data were analyzed using a digital computer and employing calibration procedures and theoretical corrections discussed elsewhere [3,4,5]. The orientation of sample and probes during thickness profiles, and the nomenclature used, are illustrated in Fig. 2. All thickness-profile measurements were made on lapped bevel surfaces. In this paper, measured spreading resistance values are referred to as "R_s"; the distance between successive measurement points (along a bevel surface or an as-grown surface) is indicated as "ΔX".

Figure 3 shows the results of a spreading resistance scan across the as-grown surface of a typical buried-layer slice, along the diameter parallel to the flow direction. For this run, two probes were used, set approximately 1 mm apart and arranged such that both probes remained out of the buried collector

[1] Figures in brackets indicate the literature references at the end of this paper.

windows at all times. The distance between successive measurement points is 0.25 mm. The plot includes only the center 25 mm of the 33 mm diameter slice. Note the increase of measured spreading resistance of 15-20% from left to right. The relationship between spreading resistance and resistivity over small ranges is approximately linear, so that the variation seen in Fig. 3 may be taken as reflecting a real 15-20% variation in the epi-layer resistivity. From later spreading resistance data on the lapped bevel surface of this sample, the resistivity near the center of the data in Fig. 3 was determined to be 0.35 ohm cm.

Figure 4 is a plot of spreading resistance vs position along the slice diameter perpendicular to the flow direction. The point-to-point probe spacing between successive measurements (ΔX) is again 0.25 mm. In this scan, the probes could not be restricted to the area outside of the buried collector windows (due to the device geometry). The result is a scatter in the data from an outside-window value of 1-1.1 x 10^3 Ω (similar to the data in Fig. 3) to an inside-window value of 820 Ω to 920 Ω. It will be shown later that this difference in measured spreading resistance is most likely due to the difference in n-layer thickness in and out of the buried layer window areas.

Figure 5a shows a spreading resistance plot with the scan running parallel to the flow direction and with ΔX = 0.025 mm. The scan covers about 7 mm and is approximately centered on the slice. Note that the outside-window R_s value is about 1.1 x 10^3 Ω and the inside-window value is 820 Ω. Figure 5b is a photomicrograph showing the probe marks made during this run. There are two sets of probe tracks--aligned so that both probes are simultaneously in and out of the buried-collector window areas. The difference in contrast of the two probe tracks is normal and is not considered to be significant. The point marked "A" on Fig. 5a corresponds to that marked "A" on the photograph of Fig. 5b and allows a point-by-point comparison of data from Fig. 5a to 5b. The set of larger-spaced probe tracks appearing near the center of the photograph are probe marks left during the scans made in acquiring the data in Fig. 3.

The data shown in Fig. 6 are similar to the 5a set, except that ΔX = 10μ in Fig. 6 and the probes are closer together--leading to a situation where the probes are either both inside or both outside the collector-window areas, or one-in, one-out. This results in three different levels in the spreading resistance data. (Note also that this scan is in an area where the top surface R_s value is 1 x 10^3 Ω rather than the 1.1 x 10^3 Ω value of Fig. 5a.) Of course, it is also possible to make a direct point-by-point comparison of these data with the probe-marked sample. By this one-to-one comparison of spreading resistance values and probe marks on a sample on a fine scale it is possible to determine the extent of lateral diffusion during high-temperature processing.

After completing the surface spreading resistance measurements, the sample used in Figs. 5 and 6 was scribed, broken and angle-lapped on a 3° bevel to obtain spreading resistance profiles in the thickness direction. The thickness profile was measured with two probes arranged parallel to the bevel edge and lying in complementary sections of adjoining patterns or, alternatively, with an ultrasonically-soldered contact applied to both p- and n-regions serving as the base contact for a spreading resistance measurement using a single probe.

Figure 7 shows a single probe spreading resistance profile on a solder-contacted sample, with the measurement scan made through an outside-window area. Thus, only the epi-layer is seen, with a total thickness of 9.5 μm.

The profile observed through the combined epi-layer and diffused buried collector via a single probe measurement scan is illustrated in Fig. 8, which is a plot of spreading resistance along the bevel surface of the sample (and hence in the thickness direction). For this run, ΔX = 10 μm, and the value of tan α (where α is the angle between the bevel surface and the top surface of the sample) was 0.0475. Therefore, the point-to-point separation in depth was $\Delta Z = \Delta X \sin\alpha$ = 0.475μm and the overall thickness of the epi-layer and the buried diffused layer was computed to be 13.3 μm, as shown on the figure. These thickness measurements are considered accurate to ± 0.5 μm. The average resistivity at various points in the profile is indicated in the figure.

Note that the 20-25% difference in spreading resistance on the slice surface inside and outside collector-window areas noted earlier is reduced to a difference of less than 10% in resistivity, when corrected for thickness effects. That is, after making a correction to the raw spreading resistance data according to our standard procedure,[4] the measured resistivity of the epi-layer near the top surface whether above a buried layer region or not can be stated as 0.33 ± 0.02 Ωcm. The resistivity minimum in the buried collector structure is seen to be approximately 0.027 Ωcm.

Figure 9 is shown to illustrate the capability of the spreading resistance technique in following a given structure through high-temperature process steps. Figure 9a gives the resistivity profile through a typical nn^+p transition, with the profile being run immediately after the epitaxial deposition step. Figure 9b is a spreading resistance resistivity profile taken on the other half of the slice shown in Fig. 9a, after the p-wall drive diffusion. Note that the shift in the interface position and slope between the n and n^+ regions is clearly detailed, in addition to other parameters of interest, such as the minimum in the n^+ resistivity, the actual n-layer resistivity, and the position of the junction between the n^+ buried layer and the p-type substrate. A quantitative calculation of the sheet resistance

of the buried collector structure after the p-wall diffusion can easily be made by a simple integration on the data shown. Figures 10a and 10b show the carrier concentration profiles obtained by conversion of the data in Figs. 9a and 9b respectively through the use of Irvin's curves[7].

The spreading resistance profile shown in Fig. 11a is typical of "normal" buried layer profiles. Figure 11b shows the resistivity profile produced by processing the raw data in Fig. 11a. Note that the minimum value in resistivity is $\sim$ 0.1 ohm cm.

Figure 12a shows a profile obtained on the same slice and bevel surface as the data in Fig. 11, but through another buried layer "tub". The positions of the interfaces in the structure are essentially the same as in Fig. 11a, but the buried layer shows a much higher resistance throughout its depth than the "normal" structure shown in Fig. 11a. The resistivity profile obtained from Fig. 12a is shown in Fig. 12 b, indicating that, although the minimum resistivity in the n^+ region is close to the 0.1 ohm cm value of Fig. 11b, the n^+ region is overall deficient in n-dopant, particularly near the nn^+ interface. One might speculate that n^+ dopant deficiencies could occur in certain "tub" regions either by incomplete photoresist removal prior to the n^+ diffusant deposition or by locally accelerated etching in the epitaxial reactor prior to the n-layer growth. This latter possibility is supported by the fact that the diffused n^+p junction occurs at the same depth in both tubs (a low surface concentration in a particular region prior to diffusion should affect the diffusion depth). Further support for a locally higher pre-deposition etch-rate comes from the data shown in Fig. 13, which is a spreading resistance and resistivity profile through the epitaxial layers of the slice profiled in Figs. 11 and 12, with the profile being taken through the area outside the buried layer tubs. The resistivity profile (Fig. 13b) gives clear evidence of excessive n^+ dopant in the reactor at the beginning of the epitaxial layer growth (some of which may have come from particular tubs, excessively etched prior to the epitaxial deposition).

The profile in Fig. 13 also illustrates one of the outstanding problems in buried layer technology--obtaining control of the properties of the epitaxially-grown n-layer in the presence of the heavily-doped n^+ regions. In fact, it is probable that the major problem in material control in buried layer structures is that involving the production of good n-epi-layers rather than the more obvious problem of producing n^+ regions with specified properties. It is, in general, necessary to trade off the desired low sheet resistance in the buried collector regions against the autodoping in the n-epi-layer resulting from the high n^+ surface concentration. Thus it is necessary to limit the surface concentration reached in the initial n^+ diffusion step.

The problem caused by too much n^+ diffusant is well illustrated in Fig. 14. This was intended to be a p^+nn^+p structure, with the localized n^+ diffusion into the p-substrate followed by a double epitaxial deposition, consisting of an n layer followed by a p^+ layer. The carrier concentration profile of Fig. 14b (which was, of course, produced by processing the raw spreading resistance data from Fig. 14a) clearly shows that, because of excessive n^+ diffusant in the epitaxial reactor (probably due to a combination of a too-high surface concentration and an unmasked n^+ back surface), the n-layer appears as almost a mirror-image of the n^+ diffusion profile. This structure would very likely fail to sustain the required collector voltage because of the relatively heavy doping throughout the n-collector region.

That it is, in fact, possible to produce well-controlled epitaxial layers on substrates containing n^+ diffused regions is illustrated in the profiles of Fig. 15, which show the measured spreading resistance profile ("a"), the resulting resistivity profile ("b") and the carrier concentration profile ("c") of an n-epi-layer, profiled through the epitaxial layer outside the n^+ diffused regions on the wafer. Note that the resistivity and carrier concentration profiles are flat after making the standard thickness correction to the raw data. The measured sheet resistance of this layer was 1820 Ω, which, together with a thickness of 8 μm taken from Fig. 15, would give an average resistivity of 1.5 ohm cm. This value agrees with the spreading resistance derived resistivity in the layer within the experimental error of these data.

3. Concluding Remarks

The values of absolute resistivity obtained by the spreading resistance technique are limited in accuracy only by the quantity and quality of calibration data used. With sufficiently detailed calibration curves and the use of homogeneous standard samples, the measurement accuracy will closely approach the reproducibility of the technique, which has been demonstrated to be $\pm$ 1% or better[2,8] for bulk samples.

For structures which are not thick compared to the spreading resistance probe contact diameter, the necessary theoretical boundary correction factor degrades the accuracy. Complete treatments of this situation are detailed elsewhere[4,5] but a demonstrated accuracy of $\pm$ 10% has been repeatedly achieved for layer thicknesses of 5 μm. With the use of smaller bevel angles and careful junction location, there is no fundamental impediment to obtaining a consistent $\pm$ 10% or better accuracy in these and thinner layers.

Finally, regardless of the accuracy obtained in the measurement of individual resistivity values in a given profile, it is in fact possible to reproduce precisely a particular device structure simply by reproducing the exact spreading resistance profile.

4. Acknowledgment

The author is grateful for the assistance of R. J. Pfeil during the experimental work and to T. L. Chu, Yung-Wei Hsueh and G. A. Gruber for technical discussions.

5. References

[1] Mazur, R. G. and Dickey, D. J., J. Electrochem. Soc. 113, No. 3, 255-259 (March 1966).

[2] Mazur, R. G., J. Electrochem. Soc. 114, No. 3, 255-259 (March 1967).

[3] Dickey, D. H., Abstract No. 57, 1963 Spring Meeting of the Electrochemical Society, April 15-18, 1963.

[4] Mazur, R. G., Abstract 164, 1966 Fall Meeting of the Electrochemical Society, October 9-14, 1966.

[5] Schumann, Jr., P. A. and Gardner, E. E., J. Electrochem. Soc. 116, No. 1, 87-91 (January 1969).

[6] Gupta, D. C., J. Electrochem. Soc. 116, No. 5, 670-1 (May 1970).

[7] Irvin, J. C., Bell Syst. Tech. J. 41, No. 2, 387-410 (March 1962).

[8] Gupta, D. C. and Chan, J. Y., Rev. Sci. Instr. 41, No. 2, 176-179 (February 1970).

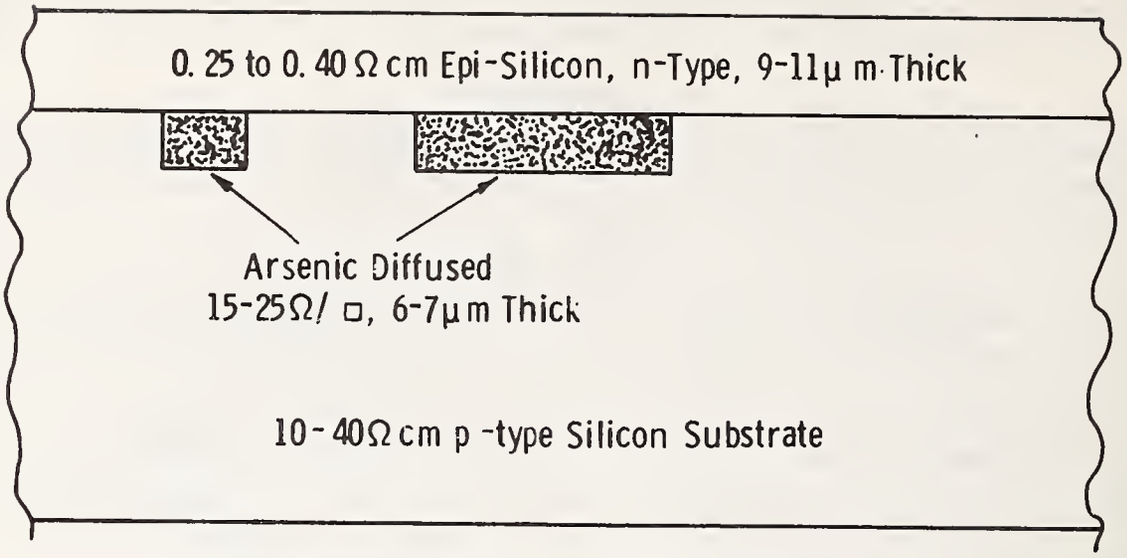


Fig. 1. Schematic cross-section view of an nn+p buried layer structure.

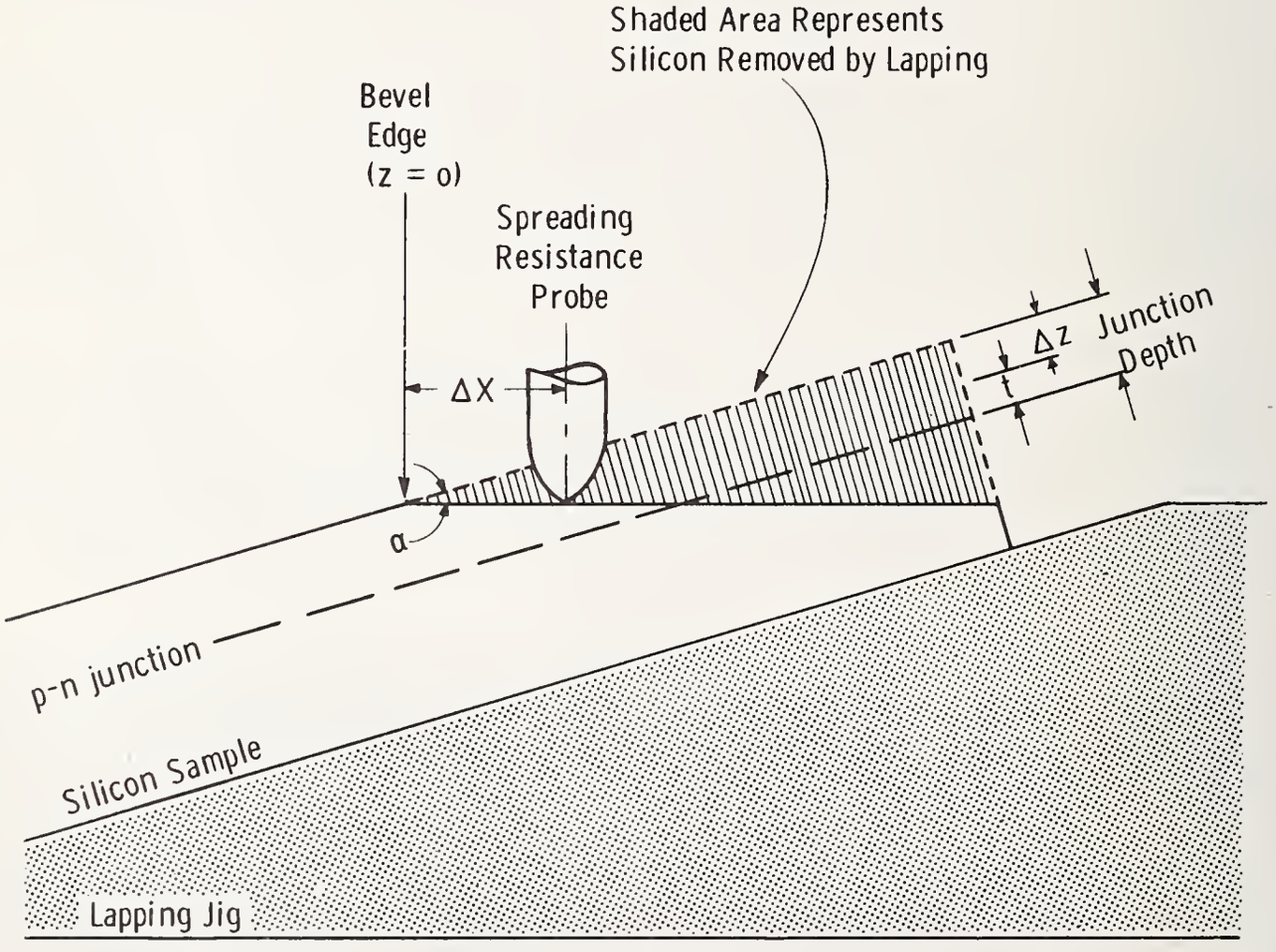


Fig. 2. Schematic side-view of a spreading resistance probe on an angle-lapped silicon sample containing a p-n junction during a thickness-direction resistivity profile.

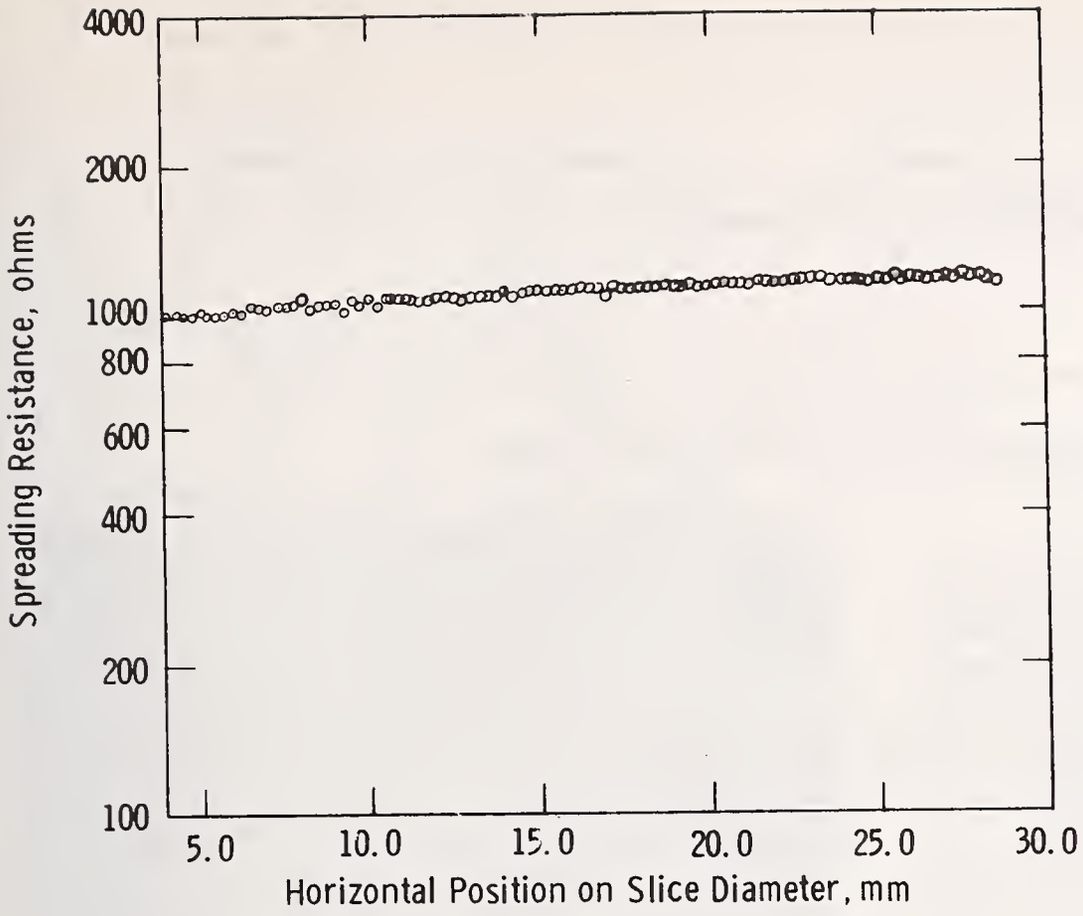
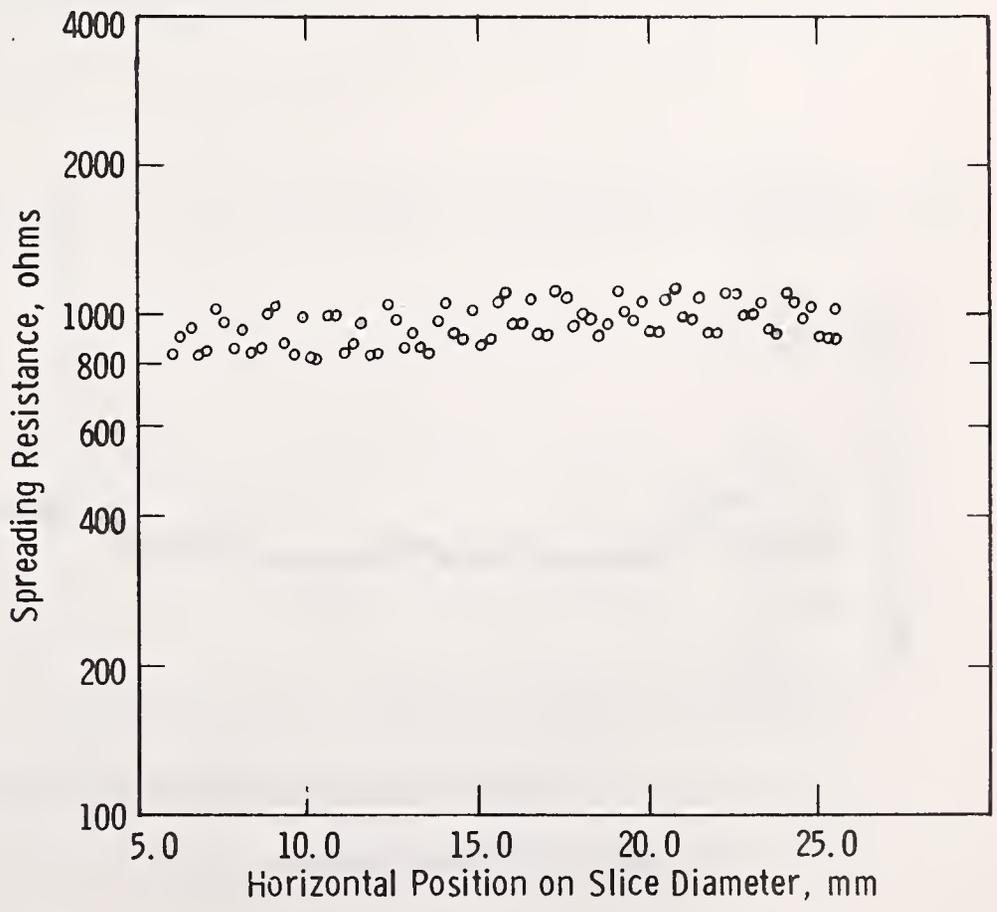


Fig. 3. Spreading resistances measured on an as-grown n-type epitaxial surface, parallel to the epiflow direction, with both probes always outside the burried-collector windows. $\Delta X = 0.25$ mm.

Fig. 4. Spreading resistance vs position perpendicular to the flow direction on an as-grown surface of an n epi-layer with buried collector regions; spreading resistance probes either both in or both outside of the buried layer regions or in a one-in, one-out configuration. $\Delta X = 0.25$ mm.



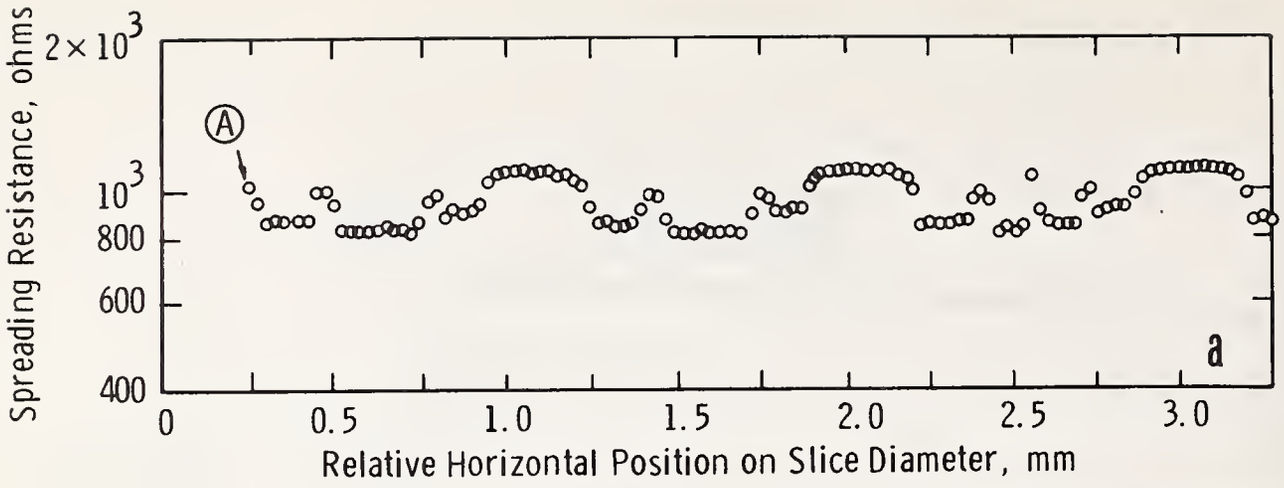
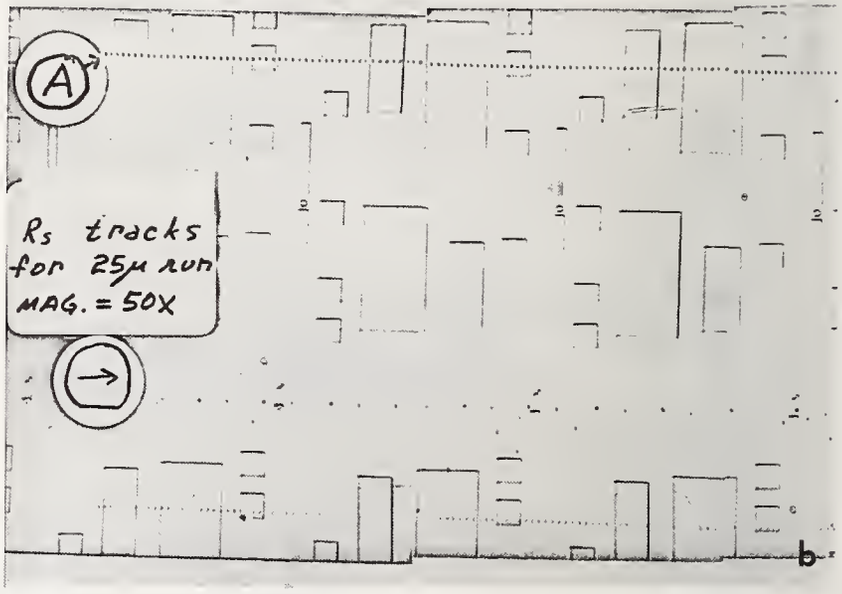


Fig. 5. a) Spreading resistance vs position on an as-grown, n-type, epitaxial layer, grown on a p-type substrate with n+ diffused layer regions. The two spreading resistance probes were incrementally displaced between measurements in 10 μm steps, parallel to the flow direction

b) Photomicrograph of probe masks left on the silicon sample during the run shown in Fig. 5a.



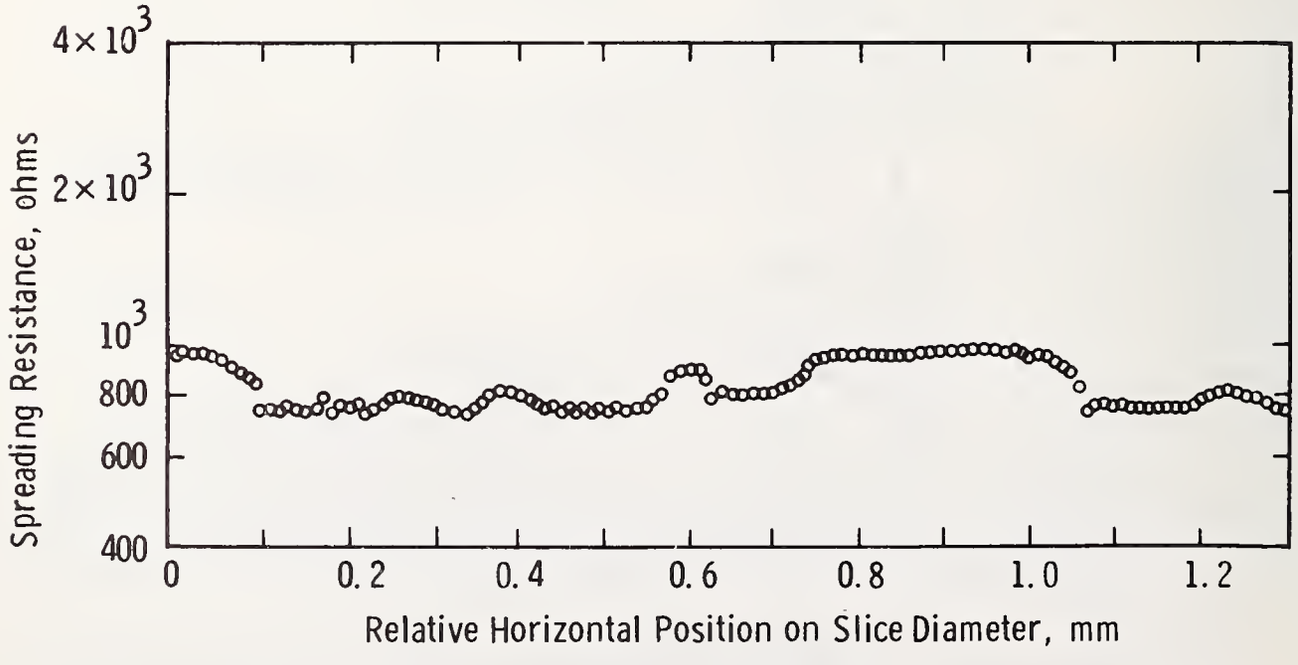


Fig. 6. Spreading resistance profile across part of the an as-grown surface of a wafer containing buried layers. Horizontal step increment (ΔX = 10 μm.

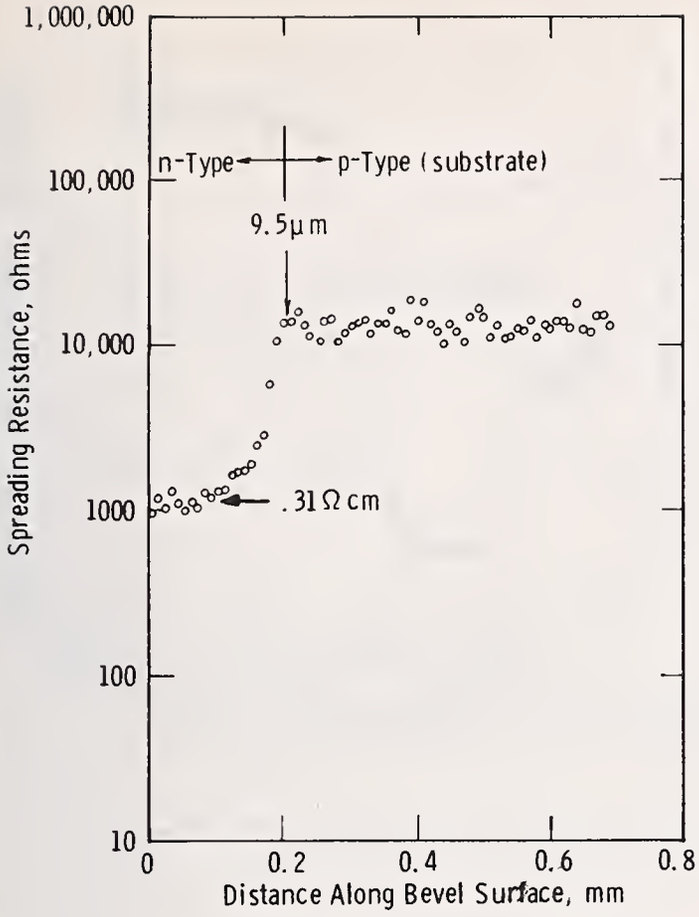
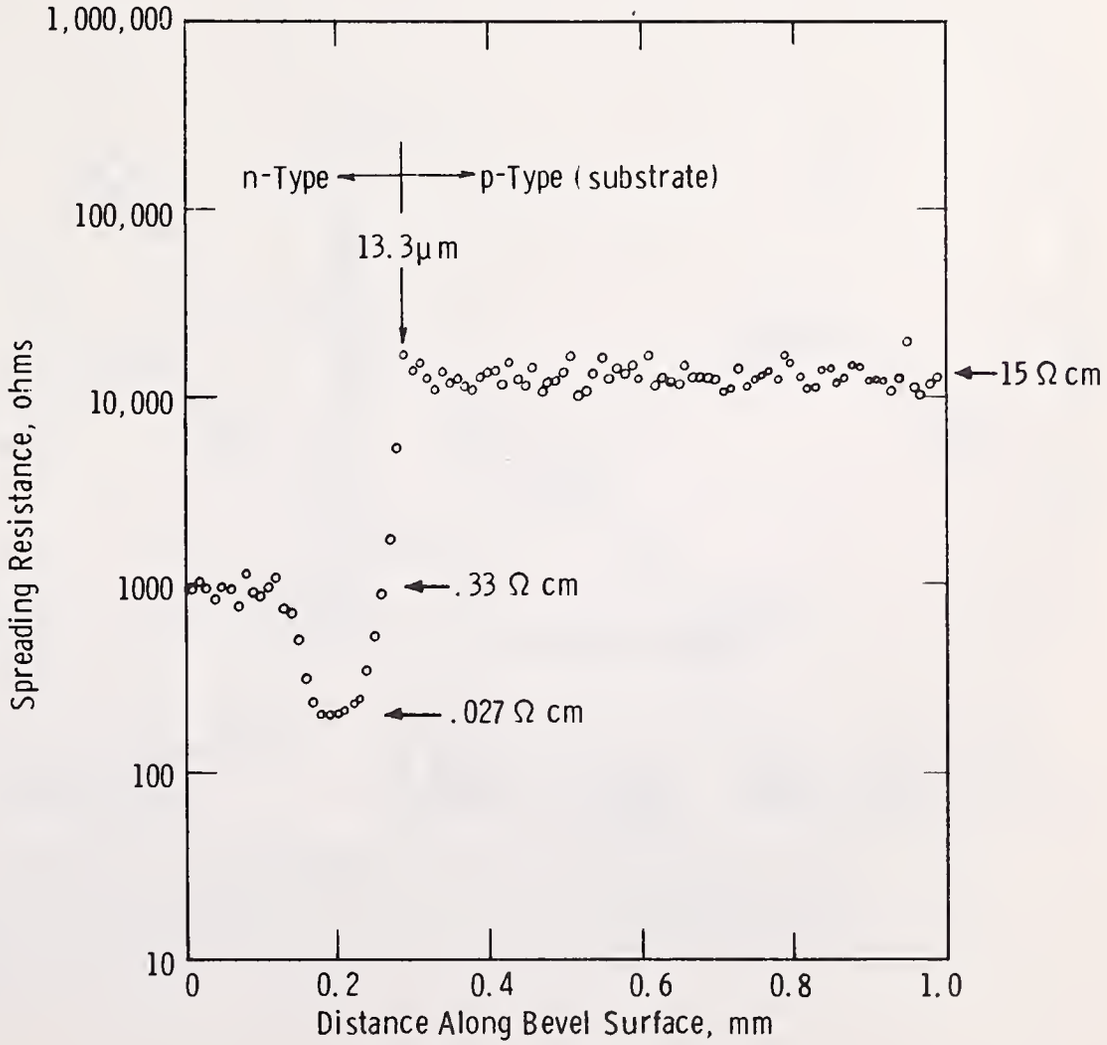


Fig. 7. Spreading resistance profile of an n on p epitaxial layer outside of the buried layer regions, using a single probe with a soldered large area current-return contact. Lapped bevel, tan $\alpha = 0.0475$. $\Delta X = 10$ μm.

Fig. 8. Spreading resistance profile through both the epitaxial n on p layer and the buried (diffused) n^+ layer: single spreading resistance probe, using a large-area current-return contact and a horizontal step of 10 μm between successive measurement points.



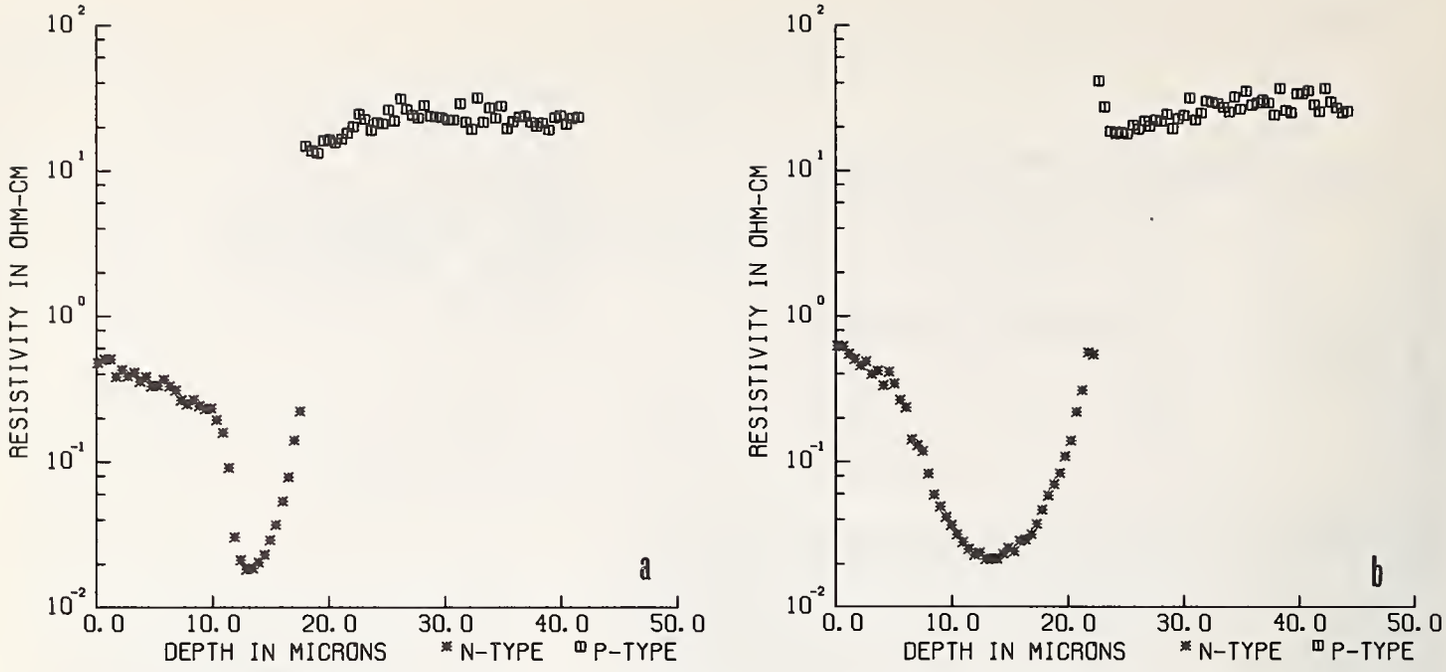


Fig. 9. Resistivity profiles of a buried layer structure from spreading resistance measurements. The profile is through the epitaxial n-layer and the buried n^{+} layer to the p substrate: a) immediately after epitaxial deposition; b) after p-wall drive diffusion.

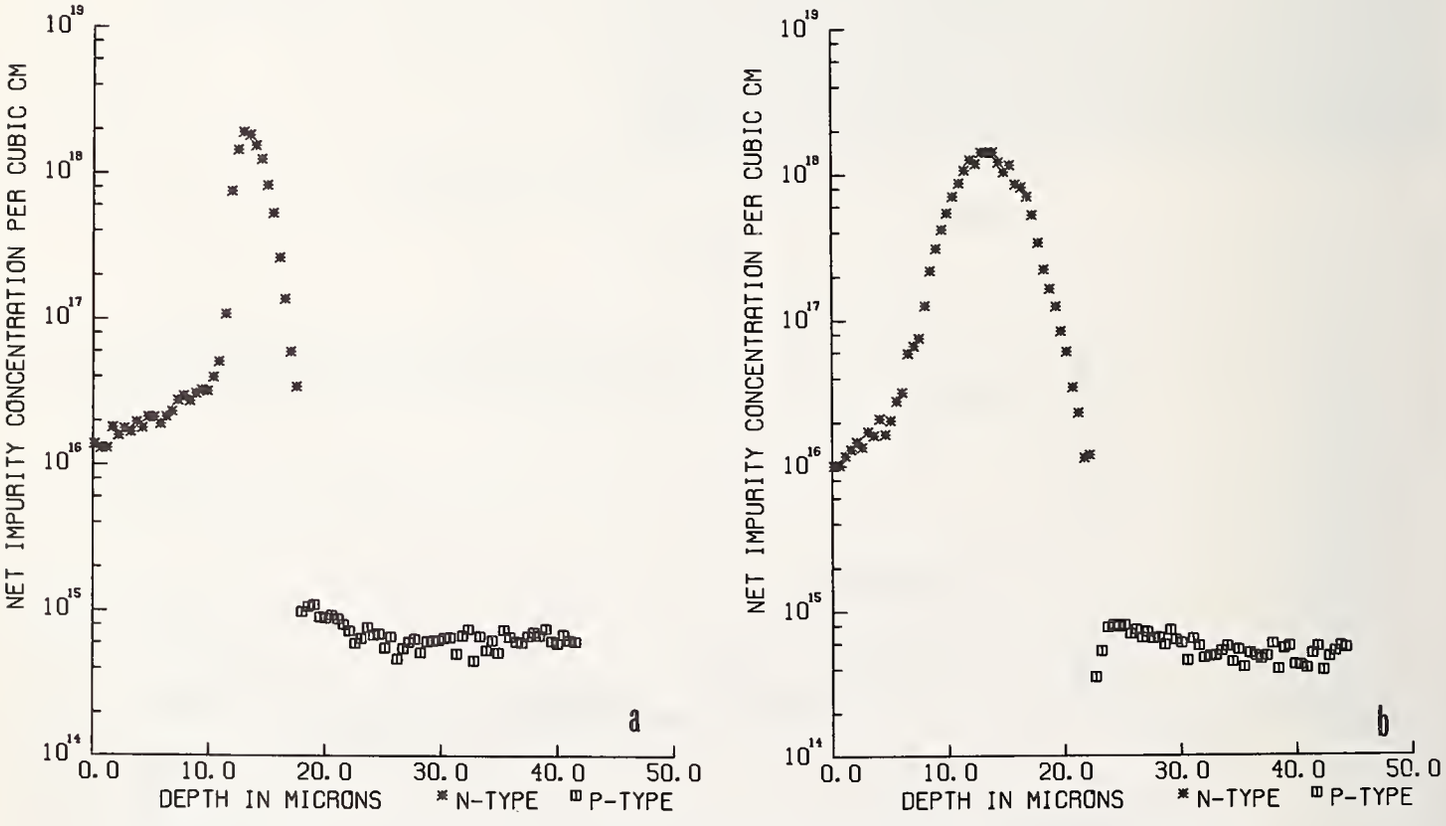


Fig. 10. Carrier concentration profiles corresponding to the resistivity profiles chown in Fig. 9: a) immediately after epitaxial deposition; b) after p-wall drive diffusion.

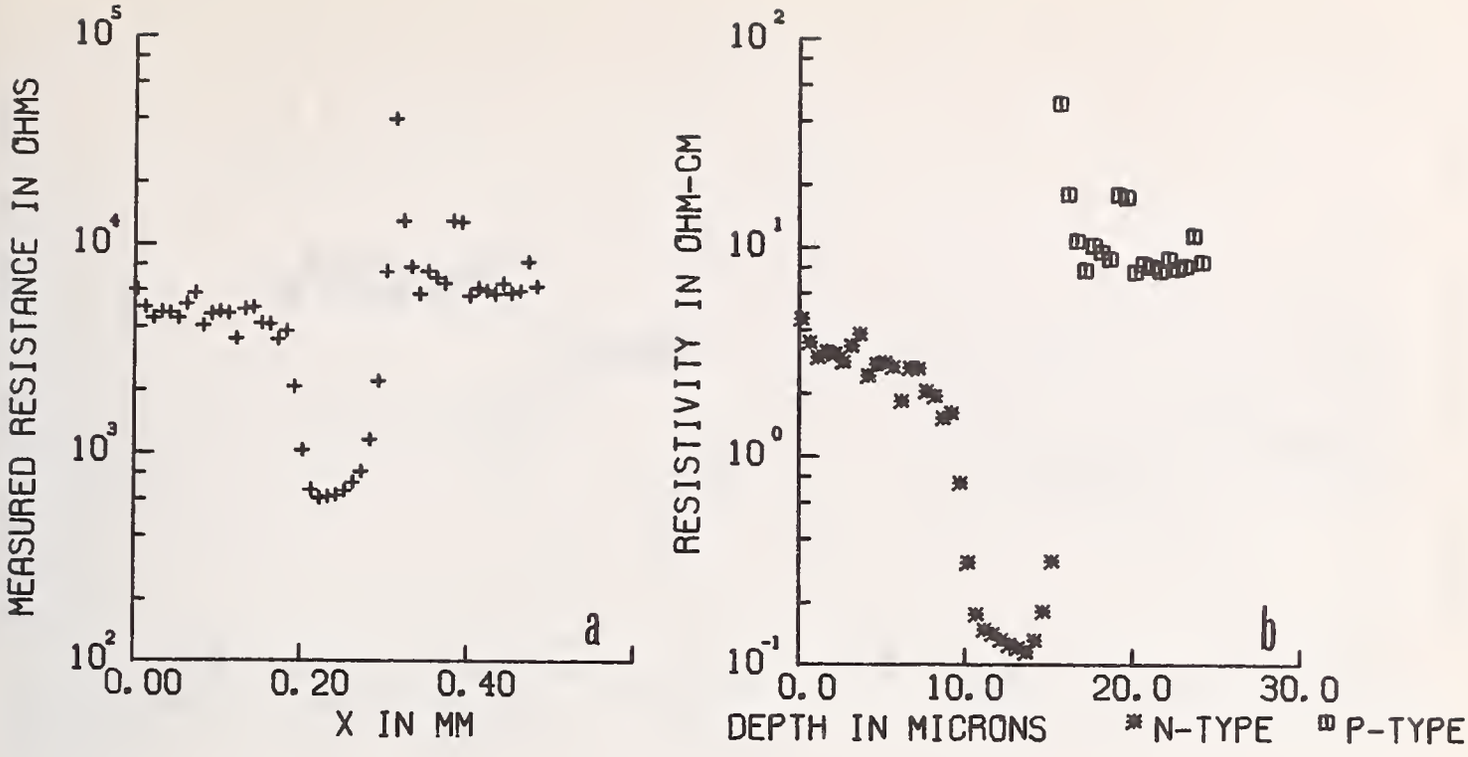


Fig. 11. Typical profile through a "normal" buried layer region: a) measured spreading resistance; b) resistivity profile obtained from the raw data in Fig. 11a.

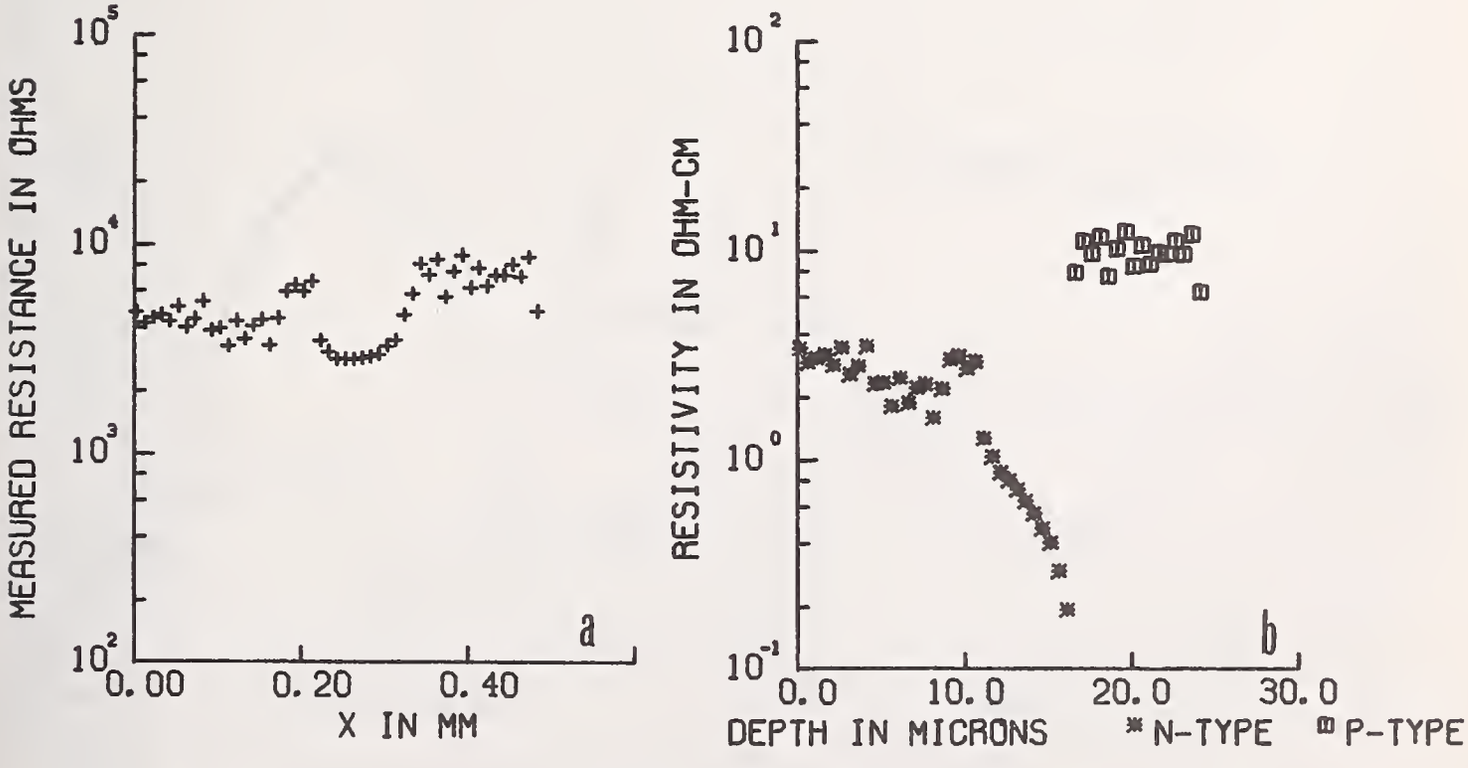


Fig. 12. Profile through an abnormal buried layer region on the same wafer as that used for the data in Fig. 11: a) measured spreading resistance profile; b) resistivity profile from the data in Fig. 12a.

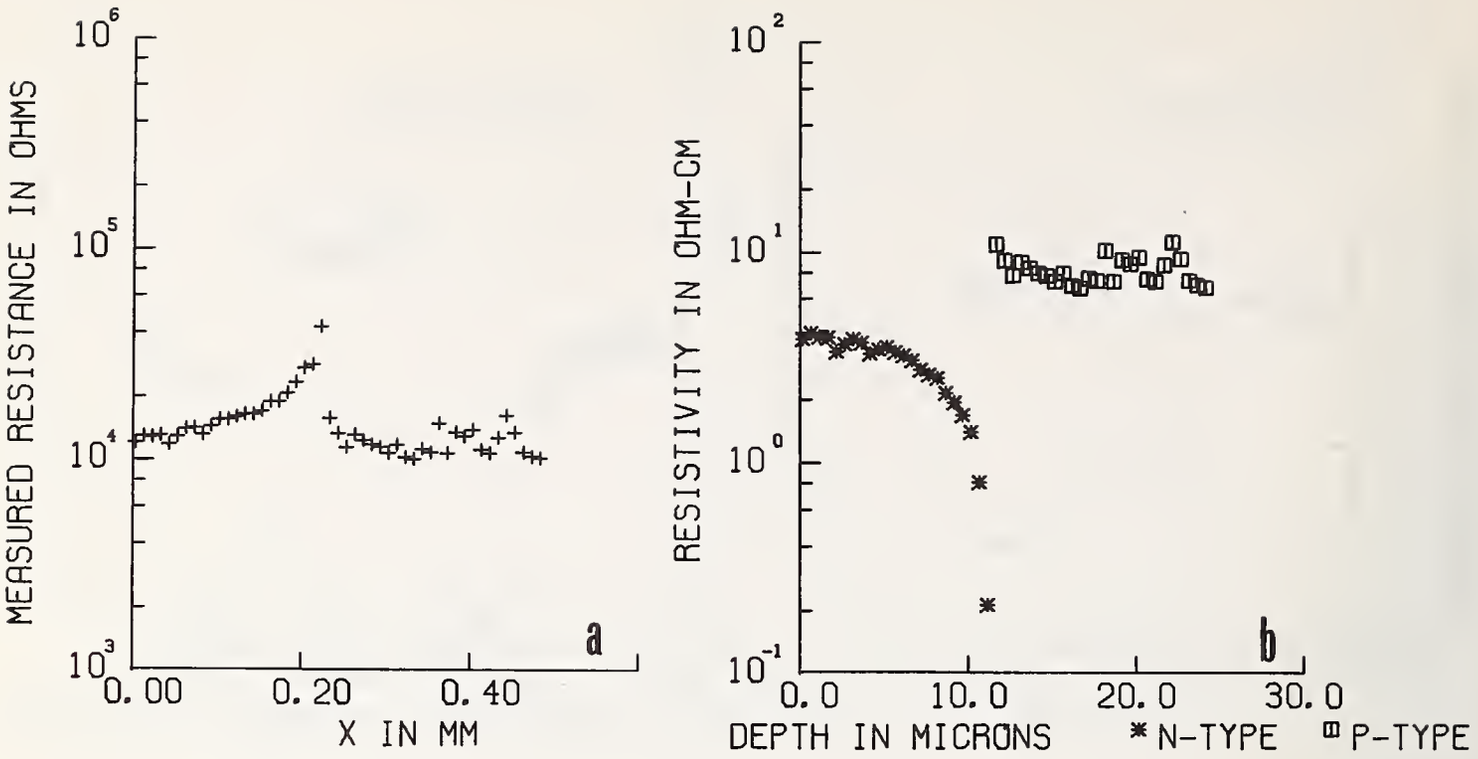


Fig. 13. Epitaxial layer profile on the same wafer as in Figs. 11 and 12: a) spreading resistance profile through the epi-layer only (outside the buried layer regions); b) resistivity profile of the epitaxial layer, obtained from the raw data in Fig. 13a.

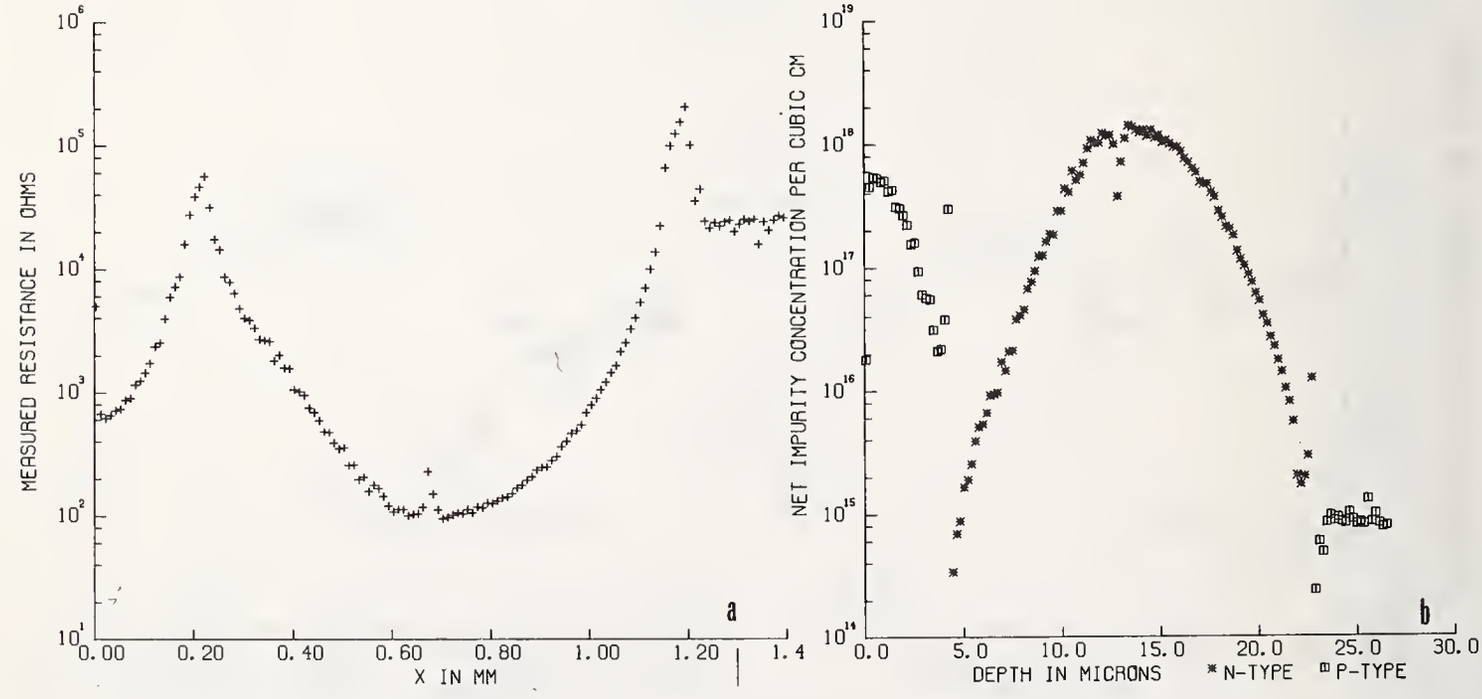


Fig. 14. Spreading resistance and carrier concentration profiles through a buried layer structure, showing extreme autodoping of the n-epitaxial layer: a) spreading resistance profile; b) carrier concentration profile.

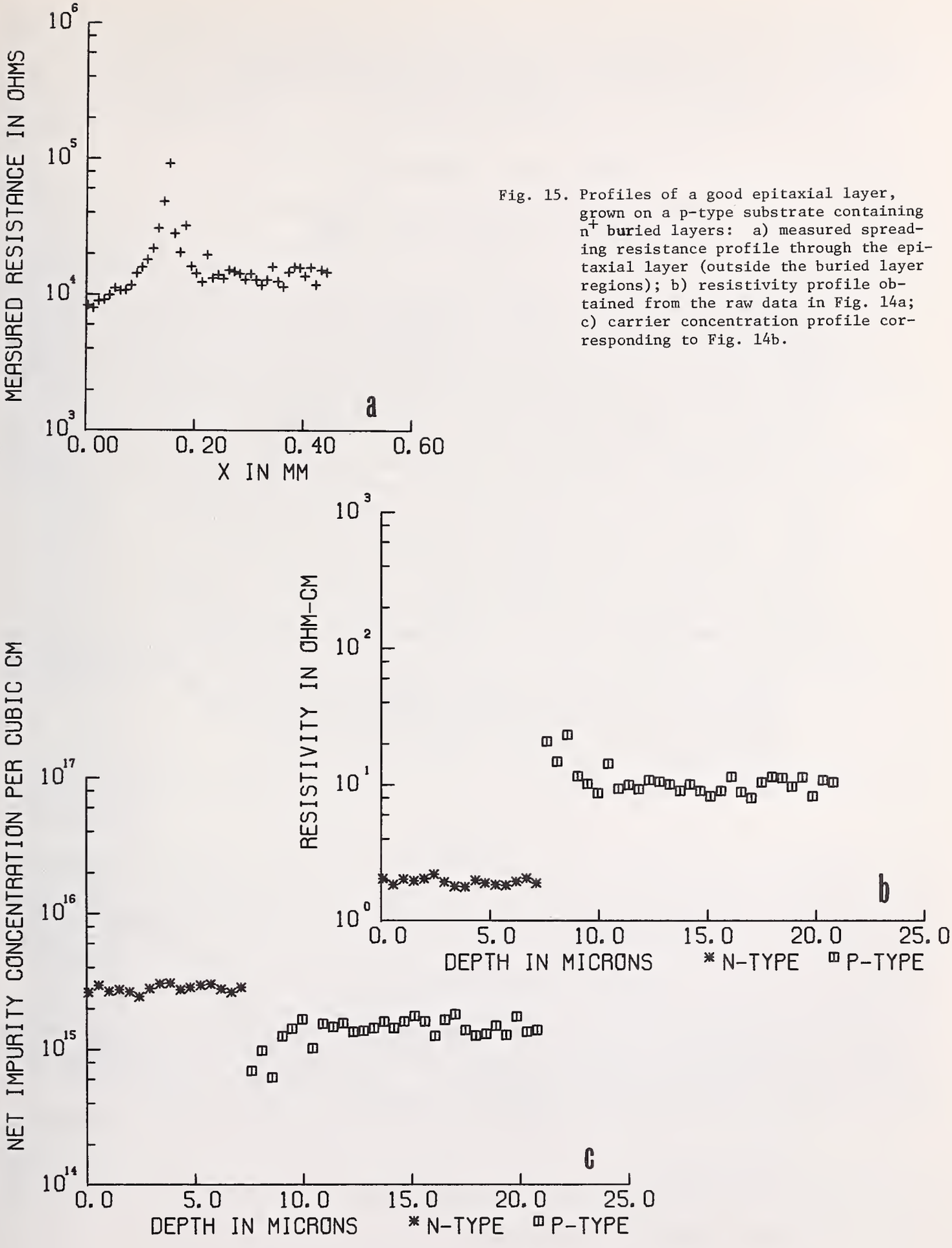


Fig. 15. Profiles of a good epitaxial layer, grown on a p-type substrate containing n^+ buried layers: a) measured spreading resistance profile through the epitaxial layer (outside the buried layer regions); b) resistivity profile obtained from the raw data in Fig. 14a; c) carrier concentration profile corresponding to Fig. 14b.

Variations of a Basic Capacitance-Voltage Technique for Determination of Impurity Profiles in Semiconductors

W. C. Niehaus, W. van Gelder,* T. D. Jones,* and P. Langer**

Bell Telephone Laboratories, Incorporated
Murray Hill, New Jersey 07974

Essential to the development of new devices with preselected electrical characteristics is a knowledge of and a control over the impurity profile of those devices. A simple relationship between reverse biased step junction capacitance, C_m, and width, W, of the depletion region provides this impurity profile, N(W). A review of two recently developed variations of this basic C-V technique is presented: 1) pulsed MIS C-V measurements which can be used to measure the impurity profile of a sample covered by a dielectric film, and 2) a second harmonic profiling technique which has the advantages of instantaneous readout and simplicity of operation. The third subject presented here involves corrections to both the second harmonic and standard C-V profiling techniques when the experimental measurements are made on graded p-n junctions. Finally, a modification of the standard C-V analysis method is presented which permits the use of capacitance interval sizes up to 95% of the zero bias value. This analysis method eliminates the need to make small capacitance intervals which has long been a hindrance to precise impurity profile determination. Specific examples of these four techniques will be presented.

Key words: Capacitance-voltage measurements, electrical conductivity, gallium arsenide, impurity profiles, MIS structures, p-n junctions, pulsed MIS measurements, Schottky barriers, silicon, thermal variations in capacitance.

1. Introduction

Development of new devices with preselected characteristics requires a knowledge of and a control over the impurity profile. The step junction dC/dV method is generally used to determine the profile, N(W) [1-6]. It is the purpose of this paper to review two recently developed variations of this basic C-V technique which offer advantages over the standard method in particular cases. First, consider the pulsed MIS C-V technique, which experiments have shown can be used to measure the impurity distribution below the surface of a silicon slice with an oxide layer on it. This method is particularly useful for nondestructive profiling of silicon since good, nondestructive Schottky barriers are difficult to produce on silicon, and the use of diffused p-n junctions will alter the doping profile to be measured as well as prevent measurements close to the surface. The second variation on the standard step junction C-V technique is a second harmonic profiling technique. The advantages of this method over the standard C-V technique are: high resolution, limited only by the material's Debye length; simplicity of operation; and instantaneous readout, without need for calculations to reduce the experimental data. In addition

*Bell Telephone Laboratories, Allentown, Pennsylvania.

**Western Electric, Allentown, Pennsylvania.

there will be a discussion of corrections to impurity profile data obtained on p-n junctions by either the standard C-V technique or the second harmonic profiling technique to take into account that the p-n junction is graded and not the ideal step junction usually assumed. Finally, for those who have access to only the standard C-V measurement technique, a discussion of a large capacitance interval technique to optimize the precision in the determination of the impurity concentration will be presented. A brief example of each of these techniques will be given.

2. Pulsed MIS C-V Measurements

2.1 Theory

The pulsed MIS C-V technique is a nondestructive method of measuring point by point impurity profiles in silicon. The principal advantage of this technique, over other C-V techniques which apply constant voltages to the junction, is that the millisecond pulses of voltage are of short enough duration that inversion layers of minority carriers cannot form at the surface. Once an inversion layer forms, the depletion layer width will no longer increase into the material. This MOS technique eliminates the need for a p-n junction in silicon to determine the profile. The diffusion process to form the p-n junction is inherently destructive to the surface of the slice being studied. The theory for profile determination of a sample covered by a dielectric film is remarkably similar to the one used for the step junction dC/dV method [1-6]. Consider an MOS capacitor which can be driven into deep depletion by the application of pulse bias [7]. Only the depletion region of the capacitor characteristics will be considered, and it will be assumed that interface states have a negligible influence.

A small increase in applied voltage dV is equal to:

$$dV = dV_o + d\varphi = dQ/C_o + d\varphi \tag{1}$$

where V_o is the voltage drop across the oxide, φ is the silicon surface potential, C_o is the oxide capacitance/unit area, and Q is the space charge density/unit area. The increase of charge in the depletion layer causes an increase in the electric field

$$dE = dQ/\varepsilon_s \tag{2}$$

in which ε_s is the dielectric permittivity of silicon. The corresponding increase in surface potential is approximately:

$$d\varphi = WdE \quad . \tag{3}$$

Substituting eqs (2) and (3) into eq (1):

$$dV = dQ(1/C_o+1/C_s) = dQ/C_m \tag{4}$$

where C_m is total measured capacitance, $C_s = \varepsilon_s/W$, and W is space charge width. Using the relation $dQ = qN(W)dW$ in eq (4), where $N(W)$ is the doping concentration, we get:

$$dV = qN(W)dW/C_m \quad . \tag{5}$$

But,

$$dW = \varepsilon_s d(1/C_s) = \varepsilon_s d(1/C_m) \tag{6}$$

as $1/C_s = 1/C_m - 1/C_o$. Substituting eq (6) into eq (5) and solving for N(W), we get:

$$N(W) = 2\left(q\varepsilon_s d(1/C_m^2)/dV\right)^{-1} \quad . \tag{7}$$

This equation shows that the doping concentration N(W) can be calculated from the slope of the $1/C^2$ versus V curve. Note that eq (7) is the same equation as used in the step junction dC/dV method given by Grove [6]. The only difference is that C_m in eq (7) includes C_s and C_o in series, while in the junction dC/dV method C_s is measured directly.

2.2 Apparatus and Measurement Technique

Apparatus and measurement technique used were identical to those outlined by Goetzberger and Nicollian [7]. Signal frequency was 500 kHz, pulse width 6 msec, and duty cycle 6%.

2.3 Sample Preparation

As a test of this point by point profiling technique on silicon, three slices were oxidized, each one in a different way, to produce a different "redistribution" of impurities and see if the technique could measure these changes. The slices were $\langle 100 \rangle$ oriented, cut from Sb doped pulled crystals with $N_D \approx 10^{15} cm^{-3}$, ~8 mil thick, lapped on both sides, and then polished on one side. The metal field plate was filament evaporated aluminum. Before oxidation the 4 point probe resistivity was measured. Prior to each application of the aluminum field plate the slices were baked in N_2 at 1100°C for 25 minutes to assure a low interface state density.

The first oxidation of the three slices took place in separate runs in a tube furnace at 1100°C in wet O_2. (Oxygen bubbled through 80°C water.) Oxide thickness was ~1200 Å. Table 1 shows the four point probe N_D before oxidation and the N_D at the surface as measured from MOS C_{max}/C_{min} after oxidation. (Read this table from top to bottom for the proper sequence of process steps and measurements.) Note that slice 3 was oxidized together (in one run) with six N+ phosphorous diffused bulk silicon slices. The N_D from MOS C_{max}/C_{min} was measured according to the method of Deal, et al. [8]. Apparently, the presence of the N+ slices during oxidation caused a large increase in N_D. Apart from slice 3 the difference in N_D in row 1 and row 3 could be explained by the "pile-up" effect or "redistribution" during oxidation [8].

Table 1. Sequence of process steps and measurements.

		Slice No. 1	Slice No. 2	Slice No. 3
1.	Initial 4 point Probe N_D ($cm^{-3} \times 10^{15}$)	0.75	0.50	0.89
2.	Process	Wet O_2 oxidize 1100°C 1200Å	Wet O_2 oxidize 1100°C 1200Å	Wet O_2 oxidize 1100°C 1200Å with 6 N+ slices
3.	MOS C_{max}/C_{min} N_D ($cm^{-3} \times 10^{15}$)	1.05	0.61	69.2
4.	Process	Strip oxide steam oxidize 1100°C 6000Å	-	-
5.	Process	Strip oxide wet O_2 oxidize 1100°C 1200Å	-	-
6.	MOS C_{max}/C_{min} N_D ($cm^{-3} \times 10^{15}$)	3.45	-	-
7.	Process	Si etch left half Strip oxide wet O_2 oxidize 1100°C 1200Å	-	-
8.	MOS C_{max}/C_{min} N_D ($cm^{-3} \times 10^{15}$)	Left 2.17 Right 4.32	-	-
9.	Pulsed MOS Measurements	See Figure 1		
10.	Final 4 point Probe N_D ($cm^{-3} \times 10^{15}$)	Left 0.80 Right 0.80	0.51	1.01

Slice No. 1 was further treated as follows. The Al dots and the SiO_2 were stripped, and the slice was oxidized in 100% steam at 1100°C to an SiO_2 thickness of 6000Å. This oxide layer was stripped and the slice was oxidized again in wet O_2 at 1100°C to a thickness of ~1200 Å SiO_2 (row 5). Al dots were reapplied and the N_D from MOS C_{max}/C_{min} was $3.45 \times 10^{15} cm^{-3}$. This amounted to a large increase in N_D compared to the previous MOS measurement of $1.05 \times 10^{15} cm^{-3}$ or the initial four point probe value of $.75 \times 10^{15} cm^{-3}$. The increase was too large to be explained by the "redistribution" effect. To investigate further the depth of this increase, the aluminum and the oxide were stripped again but only from the left half of the slice. The slice was then etched in a silicon etch and 1.1 μm of silicon was removed from the left half. The oxide on the right half was only slightly attacked by the Si etch. This oxide was removed and the whole slice was oxidized again in wet O_2 at 1100°C to form 1200 Å of oxide. Aluminum dots were again evaporated over the whole slice. At this point the 3 slices were subjected to the pulsed MOS C-V measurements. After these measurements the slices were stripped and 4 point probe measurements repeated. The final 4 point probe N_D is shown in row 10. It is clear from these values, especially in the case of slice 3, that the MOS measured increase in N_D was confined to a thin layer at the surface.

2.4 Results of Pulsed MOS C-V Measurements

The measurements are summarized in Fig. 1 where $(1/C_m)^2$ is plotted against $V_A - V_{FB}$ where V_A is the applied voltage and V_{FB} is the flat band voltage. A least square straight line was fitted as indicated. The slope of the straight line gave calculated values of N_D using eq (7). Depth of measurement can be read from the right hand ordinate.

The doping concentration at each depth, N(W) was calculated. Values of W were calculated by taking the average of each of two neighboring $(1/C_m)^2$ values and then correcting for C_o. Each corresponding N_D value was calculated from the slope of the straight line between the points in Fig. 1.

2.5 Discussion of Results

The straight line portions of the curves in Fig. 1 yield N_D values that correspond well with the initial and final 4 point probe values in Table 1 except for slice 3. Apparently the depth of measurement was large enough for slices 1 and 2 to reach into the undisturbed region of bulk doping. The largest value of W was ~3 μm for slice 1 and ~4 μm for slice 2. Slice 3, however, did not go deeper than 0.24 μm although the applied maximum voltage was the same. This points out a basic limitation in all C-V profile measurements. The higher doping levels require higher voltage to measure at the same depth, but the breakdown voltage, which limits the maximum applicable voltage, is lower for higher N_D. In the case of slice 3, the N_D value from Fig. 1 corresponds more to the MOS C_{max}/C_{min} value of Table 1 than to the initial or final 4 point probe. The measurement apparently did not reach deep enough to reveal the bulk N_D. The N_D value for slice 2 becomes equal to the bulk N_D at $W \geq 0.5$ micron. As expected, slice 1 unetched differs in profile from slice 1 when etched.

It appeared that the repeated process sequence of oxidation-oxide strip caused an increase in N_D at the surface after each oxidation and that this "build-up" can be removed by etching the silicon. The exact reason for and mechanism of this phenomenon is unknown. But, in any case, these experimental results prove that the pulsed MOS C-V technique can detect such variations in impurity profiles under thermally oxidized surfaces.

3. The Second Harmonic Profiling Technique

3.1 Theory

The second harmonic profiling technique employs a reverse biased junction driven by a 5 MHz constant current source of a few hundred milliamps. It will be shown that N(W) and W can be obtained by monitoring the voltage across the diode at 10 and 5 MHz respectively. The advantages over the standard C-V technique are direct readout, economy, high resolution, and simplicity of operation. The technique [9] is based on the electrical properties of a reverse biased Schottky barrier diode (or p-n junction). Consider the incremental change in voltage, dV, across the barrier when the diode is biased by a 5 MHz constant current source. The fundamental relationship between depletion layer voltage, V, and the charge per unit area, Q, applied at the diode surface is

$$V(Q) = \int_0^{W(Q)} E(x,Q)dx \tag{8}$$

where W is the space charge depth given by

$$Q = q\int_0^{W(Q)} N(x)dx \quad . \tag{9}$$

Here N(x) is the spacial dependence of the impurity concentration from the diode surface, x = 0, to the depletion depth W. The electric field at an arbitrary point, x, in the depletion layer is:

$$E(x,Q) = Q/\varepsilon_s - q/\varepsilon_s \int_0^x N(x)dx \tag{10}$$

Increasing the charge per unit area by dQ, causes the voltage to change by an amount.

$$dV = \int_0^{W(Q)} \left[E(x,Q+dQ) - E(x,Q)dx \right] + \int_{W(Q)}^{W(x,Q+dQ)} E(x,Q+dQ)dx \qquad (11)$$

inserting eq (10) into eq (11)

$$= \frac{dQ\ W(Q)}{\varepsilon_s} + \frac{q}{\varepsilon_s} \int_{W(Q)}^{W(Q+dQ)} \int_x^{W(Q+dQ)} N(x')dx'dx \quad . \qquad (12)$$

For small charge increments we can assume $N(x) \simeq N(W)$ between $W(Q)$ and $W(Q+dQ)$, so:

$$\simeq \frac{dQ\ W(Q)}{\varepsilon_s} + \frac{q}{2\varepsilon_s} N(W)[W(Q+dQ) - W(Q)]^2 \quad . \qquad (13)$$

From eq (9)

$$dQ = q \int_{W(Q)}^{W(Q+dQ)} N(x)dx \qquad (14)$$

$$= N(Q)q[W(Q+dQ) - W(Q)] \qquad (15)$$

so eq (12) becomes

$$dV = \frac{dQ\ W(Q)}{\varepsilon_s} + \frac{1}{2q\varepsilon_s} \frac{dQ^2}{N(Q)} \quad . \qquad (16)$$

Applying an alternating current per unit area, $J \sin \omega t$, to the diode results in an incremental space charge

$$dQ = \frac{J}{\omega} \cos \omega t \qquad (17)$$

where $\omega = 2\pi \cdot (5\ \text{MHz})$. Combining eq (16) and (17) yields

$$dV = \frac{JW}{\omega\varepsilon_s} \cos \omega t + \frac{J^2}{4\omega^2 q\varepsilon_s} \frac{1}{N(W)} [\cos(2\omega t) + 1] \qquad (18)$$

This equation shows that the voltage at the fundamental frequency, V_f, is proportional to W and the second harmonic voltage, V_{2f}, is proportional to $N^{-1}(W)$. To obtain a direct plot N(W) vs W, the 10 MHz voltage is fed to a logarithmic converter and the output is plotted on a semilog paper. As in the standard C-V measurement, the depletion layer depth is varied by the D.C. bias applied, but this is the only purpose of the bias in this technique. The voltage at the fundamental frequency in eqs (16) and (18) is just the standard relationship between the diode capacitance, incremental voltage, and incremental charge given in eq (4) where the diode capacitance per unit area equals ε_s/W and C_o is infinite. The voltage at the second harmonic which is proportional to $N(W)^{-1}$ has no direct analogy with the method of determining N(W) in the standard C-V technique. When the profiler is used to measure a p-n junction profile, V_{2f} is proportional to $(N(W_B)^{-1}+P(W_D)^{-1})$. $P(W_D)$ is the hole concentration at the depletion edge, W_D, in the diffused p-side of the junction, and $N(W_B)$ is the electron concentration at the depletion edge, W_B, in the bulk of the n-type material. Also in the case of a p-n junction, V_f is proportional to the total depletion width, W_T, which is the sum of W_B and W_D. In p^+-n junctions $P(W_D)^{-1} \simeq 0$, and $W_D \simeq 0$ so that the profile measured is essentially that of the lower doped n-layer if $P(W) >> N(W)$ and the junction is abrupt. If the inequality $P(W_B) >> N(W)$ is not satisfied and the junction is not abrupt, corrections to the profiles must be considered. These corrections will be discussed in Section 4. The princeipl advantages of this technique over the standard capacitance voltage method are: high resolution limited only the material's Debye length, instantaneous readouts, and simplicity of operation. Instrumentation consists of two R.F. voltmeters, a 300 mA RF current source and an X-Y recorder. The circuit is calibrated by scaling the 5 and 10 MHz voltage outputs to reproduce the profile curve of a calibration diode whose profile has been determined by the standard C-V technique. The calibration of this standard diode is then the limiting factor in the accuracy of the second harmonic profiling technique. But this diode can be one which approaches the ideal step junction diode assumed in the standard C-V technique.

3.2 Measurement Technique

This profiling technique has been used to monitor the electrical properties of ≈ 1 ohm-cm n-type GaAs epitaxially grown at Bell Laboratories. For comparison, the analysis to follow has also been applied to Monsanto 1 ohm-cm crystals grown by the horizontal Bridgeman (H.B.) technique. After growth of a five to ten micron epitaxial layer a matrix of 20 mil diameter, 300 Å thick gold Schottky diodes were deposited on the epitaxial surface. Profiles of N(W) versus W over the wafer surface were plotted to detect variation in doping concentration with depth as well as lateral position on the wafer. These profiles were measured in room light at room temperature as well as at near LN_2 temperature and 400°K.

As shown in Fig. 2, in certain crystals there were significant changes in these profiles with temperature. These changes were indicative of freeze-out of impurity levels as well as carrier trapping effects from impurity states with energies near the middle of the energy gap. In order to identify these charge trapping centers, an extension of the well known thermally stimulate current technique was employed [11]. Excess carriers were excited into the traps at near liquid nitrogen temperature either by shining light (from a tungsten lamp) [11,12] onto the diodes or by applying a forward bias [12]. A reverse bias was applied and the diode heated in the dark. Plots of depletion depth at a fixed bias versus continuously varying temperature were made. These are shown in Fig. 3. Thermally stimulated changes in capacitance can be related to the trap energy, E, by [11,12]

$$E \cong 0.002\ T_m \tag{19}$$

where T_m is the temperature where the maximum rate of change of capacitance occurs. The constant heating rate was varied from $0.04 < dT/dt < 8.0$°K/s and T_m was found to be essentially independent of dT/dt compared to the uncertainties in the parameters used to derive eq (19). Supplementary N(W) versus W data at fixed temperatures above and below the transition temperatures, T_m, can be taken to detect any spacial variation of the impurity state detected by the data of Fig. 3.

3.3 Results of Temperature Dependent Profiles

Table 2 gives the energy depth, concentration, and nature of impurity centers in the H.B. and epitaxial crystals studied. The variation of the H.B. profile with temperature indicated two impurity levels. The freeze-out of carriers below 150°K is due to a donor level E = 0.10±0.05 eV below the conduction band [11,13]. This level has been associated with a lattice defect state. The room temperature variation in capacitance are interpreted as the emptying of an oxygen donor level [11,13] which has trapped free carriers in the depletion region generated by the light or forward bias conditions. Others [14,15] have found, by time dependent capacitance techniques, a level ~0.7 eV below the conduction band which may be this same mid-gap level. Data in the literature on this oxygen donor level indicates it is negatively charged when filled [14]. In the Bridgeman growth technique oxygen is used to suppress silicon contamination from the quartz tubing used in the growth furnaces [16]. This room temperature capacitance change was independent of the type of excitation used.

Table 2. The activation energy, concentration and nature of the impurity levels in H.B. and vapour phase epitaxial GaAs.

Material	Energy Depth below conduction band	Concentration+	Nature
H.B. Crystals	0.1 ± 0.05 eV 0.65 ± 0.05 eV	$\sim 10^{15}cm^{-3}$ $\geq 2\times10^{14}cm^{-3}$	defect oxygen
General Vapour Phase Epitaxial layer	0.1 < E < .8 eV	$< 10^{13}cm^{-3}$	
E397	0.65 ± 0.05 eV	$\geq 3\times10^{15}cm^{-3}$	oxygen

+All impurities were uniformly distributed in material.

The vapour phase epitaxial GaAs was dominated by shallow donors which cannot be frozen out at 95°K. Deep level concentrations, except in special cases like E397 were more than two orders of magnitude below the 10^{15} doping carrier concentration at 300°K. The E397 wafer growth conditions included an unwanted incorporation of oxygen. The presence of oxygen in the growth process caused an unexpected reduction in growth rate and nonuniform N(W) as shown in Fig. 2. The change in W with temperature fixed the lower bound on the oxygen concentration in E397 to be $\sim3\times10^{15}cm^{-3}$. The ~ 0.1 eV and ~ 0.65 eV levels were uniformly distributed throughout the material in all these layers discussed.

3.4 Summary

This second harmonic technique has proved valuable as a rapid, inexpensive, and simple characterization method for a material such as GaAs when good Schottky barriers can be fabricated. Measuring these profiles at various temperatures has detected the presence of deep trapping states associated with oxygen when it is present in the GaAs growth process.

4. Corrections for Graded Profiles Determined by the C-V Technique

4.1 Introduction

Large errors can result when calculating the impurity concentration below a p-n junction from C-V data or by the second harmonic profiling technique when the exact nature of the junction is not taken into account.

Calculations have been made to correct for the effect of nonabrupt junctions when computing the impurity concentration from C-V data using this diode-capacitance method. Corrections of this type have already been developed by Lawrence and Warner [17], Hilibrand and Gold [2], and Decker [5]. Lawrence and Warner, and Hilibrand and Gold both assumed the concentration under study was uniform. Both techniques required skillful manipulation of charts and figures to obtain the corrected results and are cumbersome to use. Decker assumed a variable profile and neglected effects very close to the junction, but his approach required a sequential knowledge of the profile. This technique has specific application in the area of epitaxial structures, i.e., a thin layer grown upon a thicker substrate of the same material where a diffused mesa or planar structure has been constructed to study the films concentration profile.

Two corrections must be made to extract from C-V data the profile that existed before the p-n junction was diffused into the material. The first correction takes into account the depletion layer widening into the more highly doped diffused area, while the second correction subtracts out the compensation near the junction due to the impurity diffusion tail itself. The correction factors have been used on layers with concentrations ranging up to 2×10^{18} atoms/cm^3 over a range of junction depths from 0.3 to 5.0 microns.

4.2 Corrections for Depletion Layer Broadening Into The Diffused Junction

The model employed for studying the diffusion junction structure is shown in Figure 4. $N_B(W)$ is the initial concentration in the material under study, $N_D(W)$ is the diffusion profile which forms the p-n junction and N(W) is the net concentration at any point W in the material.

$$N(W) = N_B(W) - N_D(W) \tag{20}$$

$N_B(W)$ is assumed constant in this analysis. W_D and W_B are the depletion layer widths relative to x = 0 for the diffusion and bulk sides of the junction respectively. The total junction depletion width, W_T, is then

$$W_T = W_B + W_D \tag{21}$$

The approach to the problem is that any real diffusion profile may be fitted over the range of the depletion layer widening by an exponential function of the form

$$N_D(W) = Ae^{-B(W+J)} \tag{22}$$

where A is the surface concentration after diffusion, J is the junction depth below the surface and B is a constant which depends on the specific diffusion profile as well as on the ratio of the surface concentration to the bulk concentration. The surface concentration junction depth and diffusion profile shape must be known experimentally in order to determine $N_B(W)$.

Inserting eq (20) and eq (22) into Poisson's equation

$$\frac{d^2V}{dW^2} = \frac{qN(W)}{\varepsilon_s} \tag{23}$$

and integrating with respect to W from the junction to the depletion layer edge in each direction, we obtain the electric fields from $W = -W_D$ to $W = 0$ and from $W = 0$ to $W = W_B$. Since $N(W) = 0$ at $W = 0$ and the electric field is continuous through $W = 0$, we obtain a relationship between W_B and W_T

$$W_B = B^{-1} \ell n \left[\frac{e^{BW_T} - 1}{BW_T} \right] \tag{24}$$

where W_T is related to the measured small signal capacitance per unit area, C, by

$$C = \frac{\varepsilon_s}{W_T} \quad . \tag{25}$$

The second integration of eq (23) for the depletion layer side of the junction yields the voltage. Differentiating with respect to W_B gives

$$N(W_B) = \frac{\varepsilon_s}{qW_T} \frac{dV}{dW_B} \quad . \tag{26}$$

Using the chain rule of integration eq (26) is transformed into an equation where all the variables can be evaluated from the measured values of capacitance, voltage, and diode area.

$$N(W_B) = \frac{\varepsilon_s}{qW_T} \frac{dV}{dW_T} \frac{dW_T}{dW_B} \tag{27}$$

dW_T/dW_B can be evaluated by differentiating the electric field expression at W = 0

$$\frac{dW_T}{dW_B} = \frac{BW_T(e^{BW_T}-1)}{e^{BW_T}(BW_T-1) + 1} \quad . \tag{28}$$

The net impurity concentration on the bulk side of the junction $N(W_B)$ given by eq (27) can now be substituted into eq (20) to yield the concentration in the bulk prior to the diffusion process, $N_B(W_B)$. This gives

$$N_B(W_B) = N(W_B) \cdot \left[\frac{e^{BW_T} - 1}{e^{BW_T} - BW_T - 1} \right] \quad . \tag{29}$$

The term in brackets in eq (29) is only important close to the junction and approaches unity rapidly at values of $BW_T >> 1$. In summary then, to determine the impurity concentration in the bulk semiconductor before diffusion by eq (24) and eq (29) one needs C-V data, the junction depth, J, surface concentration, A, and shape of the diffusion profile, B.

To take these two corrections into account when profiling p-n junctions with the second harmonic profiler, eqs (24), (27), (28), and (29) are used. The profiler output at V_{2f} gives $(N(W)^{-1} + N_D(W)^{-1})$ for p-n junctions which is just $(\varepsilon_s(qW_T)^{-1}dV/dW_T)^{-1}$ in eq (27). The profiler voltage V_f gives W_T. So eq (24) can be used with the known constant B to calculate W_B. This W_T is also needed to calculate the remaining analytic correction factors in eq (29).

4.3 Results

Figure 5 shows the corrected and uncorrected profiles on n-type bulk Si slice which had a resistivity of 0.05 ohm-cm as determined by the 4 point probe technique. The impurity concentration was known to be independent of depth. Capacitance-voltage data was measured on a diffused p-n junction from 0.3 volts in the forward direction to the breakdown voltage in the reverse direction. The junction depth was 4.4 microns, A was 8.0×10^{20} cm^{-3}, and $B^{-1} \sim 4.2 \times 10^{-4}$ cm. As can be seen in Fig. 5 if these corrections are not taken into consideration the C-V data would lead to an incrorrect magnitude and shape of the profile.

5. Large Capacitance Intervals

5.1 Introduction

The fundamental resolution limit in these techniques is the Debye length at a given concentration. The actual precision achieved in profile determination is normally limited by the measurement technique to be less than this. There are two experimental factors which can be adjusted to optimize the precision of the profile measurement for both Schottky barrier and p-n junctions. The first factor is the need to take small intervals in capacitance. These small capacitance intervals lead to large random errors due to the imprecision in measuring capacitance at low signal levels (<50 millivolts) and frequencies greater than 100 kHz. The second factor affecting the precision is that as the capacitance interval is enlarged, the finite approximation for dV/dC is no longer valid. For measured profiles the error in dV/dC is not only a function of the interval size but is also related to the shape of the concentration profile.

The diode capacitance technique for profile determination was first used by Hilibrand and Gold [2] and Thomas, et al., [3]. Other authors [18,19] have had reservation about the use of the differential method in regions where the doping profile was changing rapidly as a function of distance. An error analysis of the technique for flat profiles was published by Amron [20]. He calculated the random error in concentration as a function of interval size, diode area and error in capacitance. He was able to show that the absolute errors introduced by capacitance interval sizes up to 10% for a constant doping profile are small, but he did not consider the case of variable profiles. Jones [21], also considering flat profiles, developed a technique which permitted the use of large capacitance intervals in taking the C-V data yet yielded zero absolute error.

5.2 Theory

This section treats the case of a flexible interval size coupled with an arbitrary doping profile. The flexible interval size is needed to minimize the sum of the random error and shape distortion for different types of doping profiles. We assume the doping profile in the material can be approximated by

$$N(W) = KW^{S-2} \tag{30}$$

where K and S are constants. We allow K and S to vary from interval to interval but require that over a given interval in W, they are constant. By approximating the material's profile using eq (30), the capacitance interval can be enlarged while minimizing the increased error in the finite approximation for dC/dV.

This profile is inserted in eq (23), Poisson's equation, and the equation is solved using the assumptions: 1) There is no depletion layer widening into the heavily doped side of the junction. 2) There are no mobile carriers within the depletion region. 3) The depletion region edge abruptly terminates in the lightly doped side of the junction, and 4) There is no diffusion of charged carriers. A double integration of eq (23) yields an explicit expression for the voltage

$$V = \frac{q}{\varepsilon_s} \int_0^W \int_x^W N(x)dxdx = \frac{qKW^S}{\varepsilon_s S} \quad . \tag{31}$$

Differentiating the above equation with respect to W and solving for N(W), gives the standard step junction solution.

$$N(W) = \frac{\varepsilon_s}{q} \frac{1}{W} \frac{dV}{dW} \quad . \tag{32}$$

Rewriting eq (32) in incremental form yields

$$N(W_*) = \frac{\varepsilon_s}{q} \frac{1}{W_*} \left(\frac{V_2-V_1}{W_2-W_1}\right)_{W=W_*} \quad . \tag{33}$$

In the last equation, V_1 and V_2 represent the experimental voltages and W_1 and W_2 represent the depletion depths at either end of the interval, ΔW, corresponding to experimental capacitances C_1 and C_2. W_* is the point at which $dV/dW = \Delta V/\Delta W$. The essence of the new solution is to relate dV/dW to $\Delta V/\Delta W$, evaluated at W_*. The location of W_* in the interval ΔW is not a function of interval size. The relationship between W_1 and W_2 and W_* is only a function of how the concentration varies over the interval, ΔW. At this point we make the assumption that W_* is single valued over the interval, ΔW. Calculate the theoretically expected voltage interval, ΔV, to cause a change in depletion depth (W_2-W_1) for the profile assumed in eq (30), using eq (31).

$$\Delta V = \frac{qK}{\varepsilon_s S} \left(W_2^S - W_1^S\right) \quad . \tag{34}$$

Substituting eqs (30) and (34) into eq (33) and solving for W_*

$$W_* = \left[\frac{W_2^S - W_1^S}{S(W_2-W_1)}\right]^{\frac{1}{S-1}} \quad . \tag{35}$$

Equation (35) still contains the unknown S which can be evaluated from eq (31)

$$S = \frac{\ell n(V_2/V_1)}{\ell n(W_2/W_1)} \tag{36}$$

W_* has now been expressed as a function of the measured voltage and capacitance by eqs (35), (36) and (37) where the measured capacitance per unit area, C, is related to the depletion width, W, by

$$W = \frac{\varepsilon_s}{C} \quad . \tag{37}$$

Therefore, the impurity concentration $N(W_*)$ can now be calculated from C-V data using eqs (33) and (35).

5.3 Results

The large interval approach asuumes that the material's profile can be represented by an analytic function given in eq (30) The errors in calculating the doping profile from C-V data by the standard method:

$$N(W) = 2(q\varepsilon_s d(1/C^2)/dV)^{-1} \tag{38}$$

$$W \equiv \varepsilon_s/C \tag{39}$$

and the large interval method of eqs (33) and (35) depend on the shape of the profile being measured. To illustrate the improved accuracy of the large interval technique, consider material having doping profiles of the form $N(W) \propto W^z$ in which z is a constant and $N(W) \propto e^{DW}$. A set of capacitance-voltage data for diodes with these profiles is generated by taking a series of fixed depths W, and using eq (37) to determine the capacitances and calculating the double integral of Poisson's equation

$$V = q/\varepsilon_s \int_0^W \int_x^W N(x')dx'dx \tag{40}$$

to determine the voltages. Here $N(x')$ is the postulate form of $N(W)$. This capacitance-voltage data is analyzed by both C-V analysis techniques. The results are compared with the known initial profile to determine the error in the two analysis methods.

In each case, an error analysis has been calculated to determine the 3σ error in concentration due to random errors in capacitance measurement of 0.5% and 1.0% and diode diameter of 0.3% and 1.0%. These results show the increased error in profile determination due to imprecision in the C-V data when the capacitance interval is decreased.

Consider a real doping profile which varied with depth, W, as $N(W) \propto W^z$ for z = o, 3, 10, and 20. Figure 6 gives the error in concentration introduced when determining this real profile by the standard analysis of eqs (38) and (39). The figure shows that for fixed z, the error in concentration increases as the percentage change between successive capacitance measurements increases from 10% to 95%. The standard analysis technique introduces errors in determining the doping concentration for even the simple case of a constant profile, $N(W) \propto W^0$. The large interval solution gives zero absolute error in determining a real profile, $N(W) \propto W^z$, since this analytic form was used to derive eqs (33) and (35).

Figure 7 depicts the case where the real doping profile in the material is $N(W) = 10^{15} e^{DW}$ where $D = 1\ \mu m^{-1}$. This value of D gives representative profiles in devices after processing. The error in this profile as determined by the standard technique, eqs (38) and (39) and by the large interval technique, eqs (33) and (35) at two depletion depths 2.7 and 5.4 microns are shown. At both depths the classical solution introduces larger errors than the large interval technique described here.

6. Summary

Two variations of the basic C-V impurity profiling techniques have been presented. It was shown that the pulsed MIS C-V technique can be used to measure the point by point profile under an oxidized silicon surface. In addition, this technique has proved to be sensitive to small profile changes induced by the oxidation itself. The second harmonic profiling technique has proved valuable as a rapid, inexpensive, and simple characterization method for device material.

Corrections to C-V profile data obtained from a diffused p-n junction to correct for the nonabrupt nature of the junction have also been presented. Finally, for those limited to determining profiles by the standard C-V technique, a large capacitance interval technique has been developed to minimize the error in profile determination.

7. References

[1] Grove, A. S., Leistiko, O., Jr. and Sah, C. T., J. Appl. Phys. 35, 2695-2701 (1964).

[2] Hilibrand, J. and Gold, R. D., RCA Rev. 21, 245-52 (1960).

[3] Thomas, C. O., Kahng, D. and Manz, R. C., J. Electrochem. Soc. 109, 1055-61 (1962).

[4] Amron, I., Electrochem. Tech. 12, 327-9 (1964).

[5] Decker, D. R., J. Electrochem. Soc. 115, 1085-9 (1968).

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

[7] Goetzberger, A. and Nicollian, E. H., J. Appl. Phys. 38, 4582-8 (1967).

[8] Deal, B. E., Grove, A. S., Snow, E. H. and Sah, C. T., J. Electrochem. Soc. 112, 308-14 (1965).

[9] Copeland, J. A., IEEE Trans. on Electron Devices, p. 445, May, 1969

[10] Weisberg, L. R. and Schade, H., J. Appl. Phys. 39, 5149 (1968).

[11] Carballes, J. C. and LeBailley, J., Solid State Comm. 4, 33 (1966).

[12] Schade, H., J. Appl. Phys. 40, 2613 (1969).

[13] Carballes, J. C., Diguet, D. and LeBailley, J., 1968 Symposium on GaAs, Dallas.

[14] Williams, R., J. Appl. Phys. 37, 3411 (1962).

[15] Senechal, R. and Basinski, J., J. Appl. Phys. 39, 4581 (1968).

[16] Woodall, J. M. and Woods, J. F., Solid State Comm. 4, 33 (1966).

[17] Lawrence, H. and Warner, R. M., Jr., Bell Sys. Tech. J. 39, 389 (1960).

[18] van Opdorp, C., Solid State Electronics 11, 397 (1968).

[19] Kennedy, D. P., Murley, P. C. and Kleinfelder, W., IBM J. R and D, 12, 399 (1968).

[20] Amoron, I., Electrochem. Technology 5, 94 (1967).

[21] Jones, T. D., Masters Dissertation in Electrical Eng., Lehigh Univ., 1968.

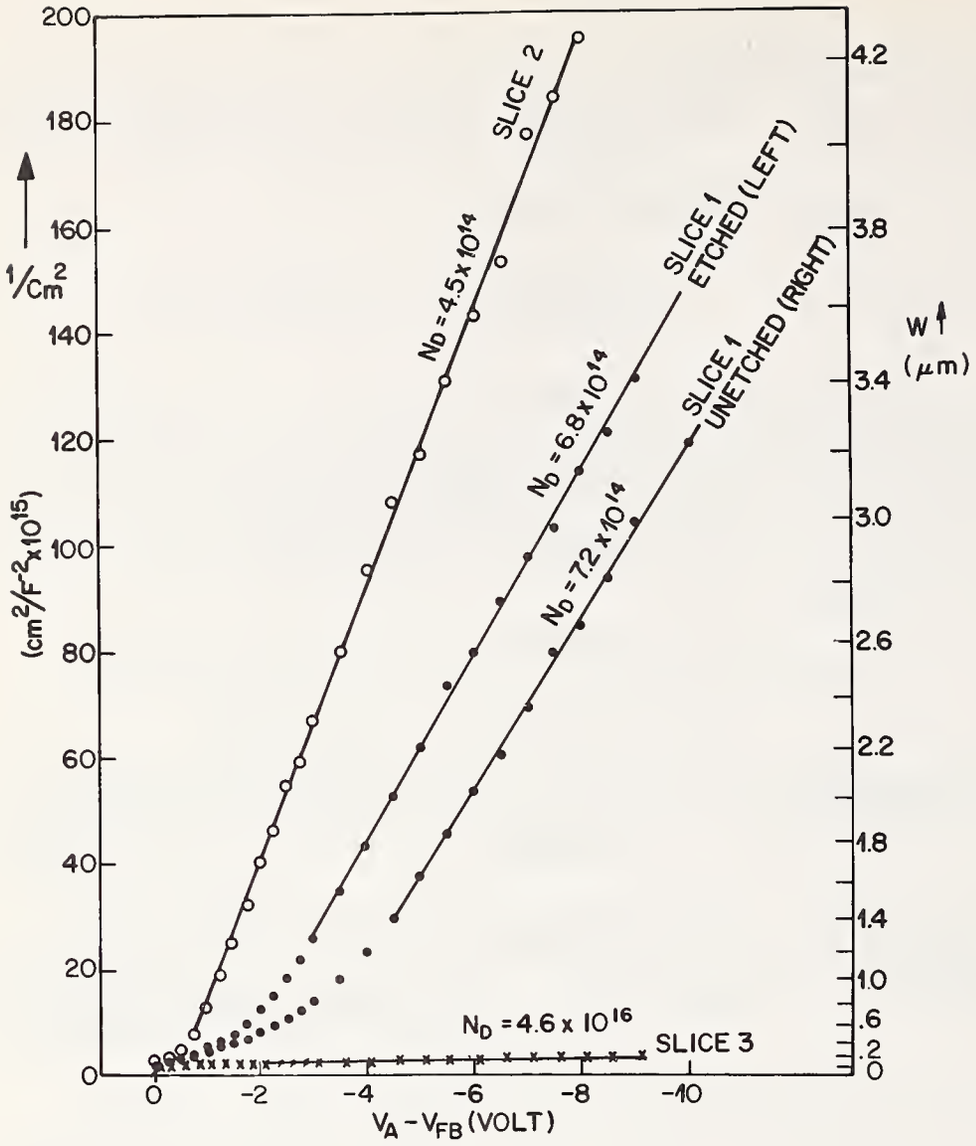


Fig. 1. Plots of $1/C_m^2$ versus the voltage $(V_A - V_{FB})$. The right hand ordinate indicates the depth (W) of the measurement below the Si surface.

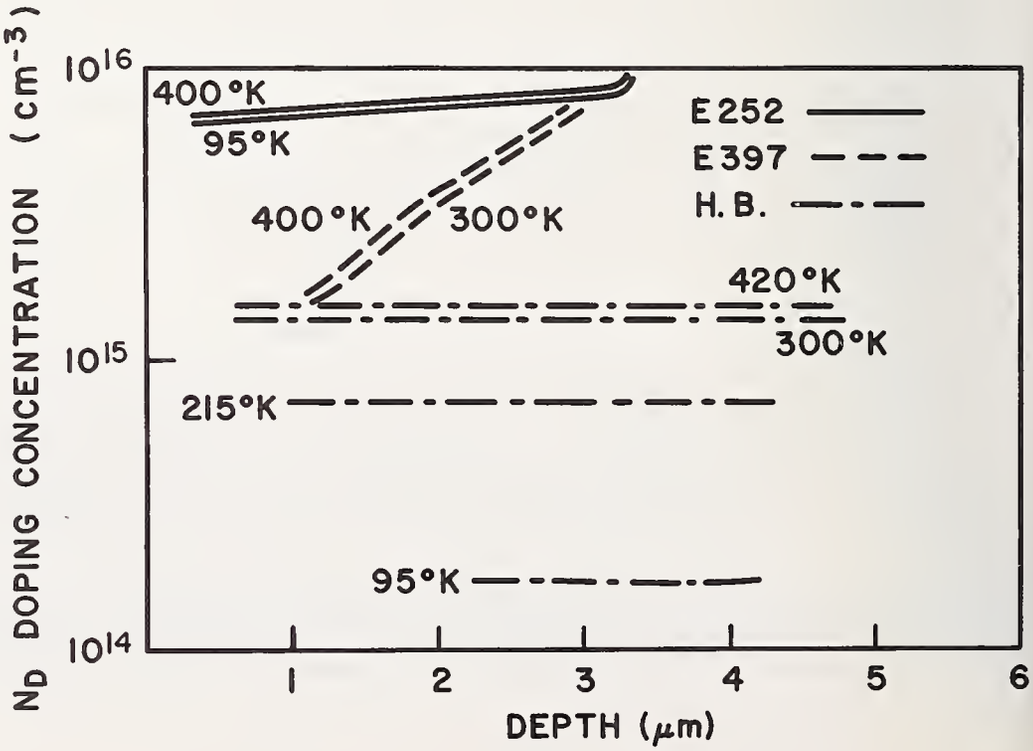


Fig. 2. Profiles of typical H.B. crystals and vapor phase epitaxial layers of n:GaAs at 95°K, 300°K, and 400°K. Also shown is an epitaxial layer unintentionally doped with oxygen.

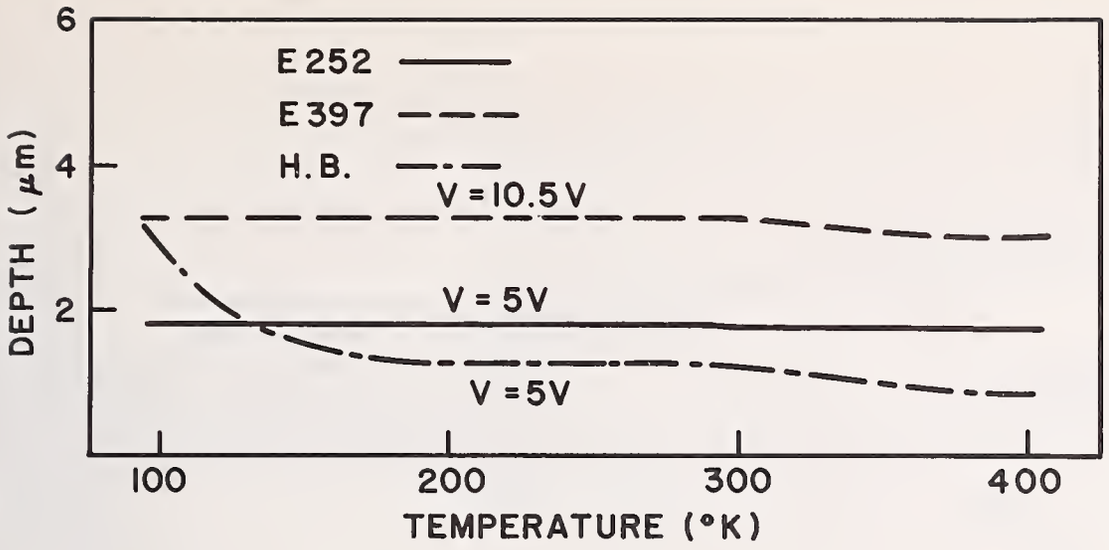


Fig. 3. Thermal variations in depletion depth or equivalently diode capacitance as a function of temperature.

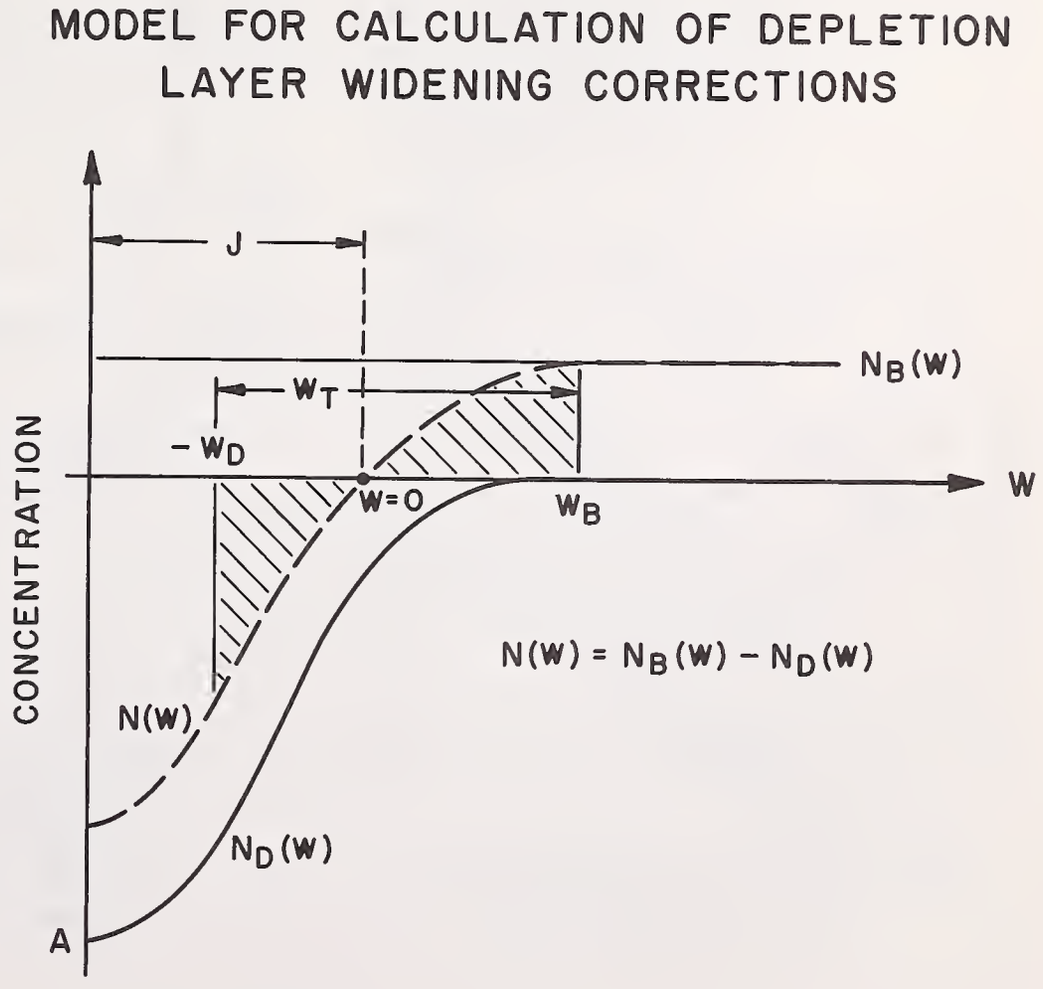


Fig. 4. Model for the calculation of depletion layer widening and compensation near the junction due to the impurity diffusion tail itself.

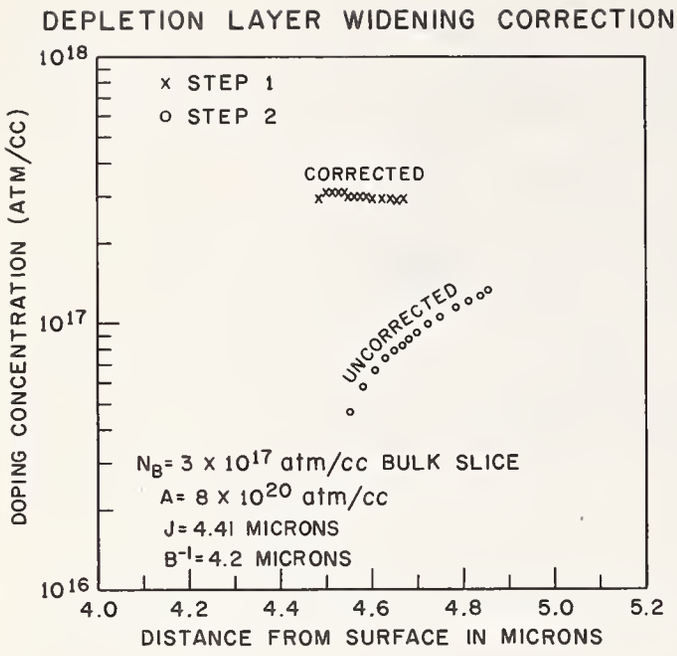


Fig. 6. Errors in the classical solution for a profile as a function of capacitance interval for an experimental profile of the form $N(W) \propto W^Z$. The random error is introduced by lack of precision in the experimental data. The finite approximation errors arise in calculating dV/dC over a finite interval.

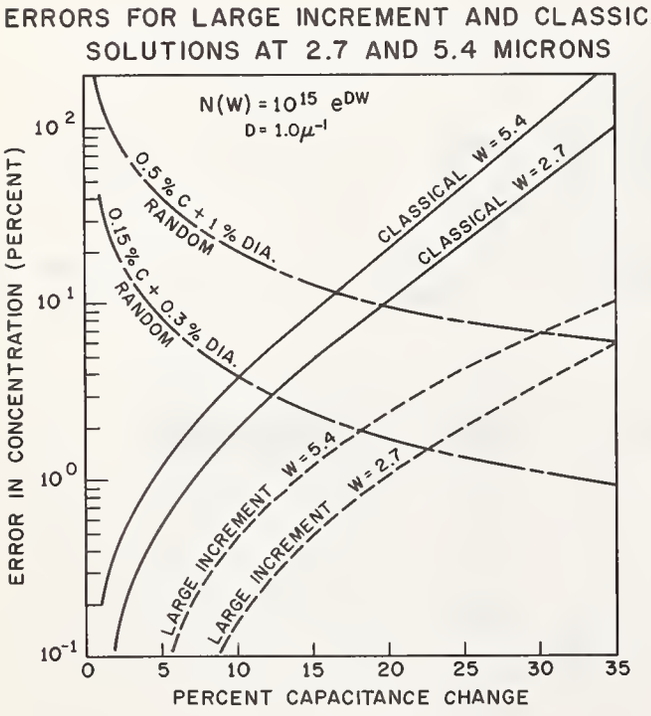


Fig. 5. Profiles calculated from C-V data on a diffused junction. The impurity concentration in the bulk of the semiconductor before diffusion differs significantly from the profile determined by the standard analysis of the raw data.

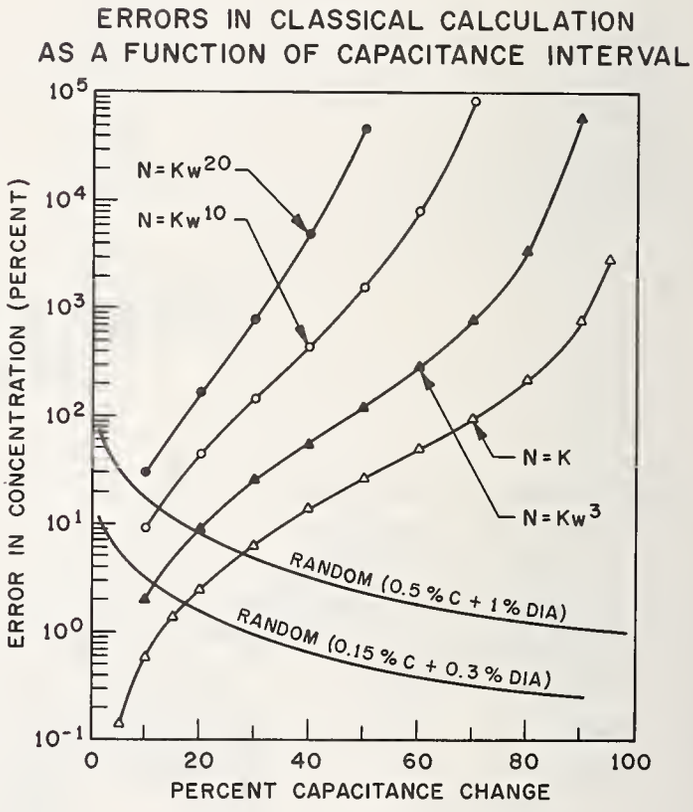


Fig. 7. The random error and finite interval errors at 2.7 microns and 5.4 microns from the diode junction for the classical and large interval solutions to an experimental profile $N(W) = 10^{15} e^{DW}$. Here the assumed random errors in C are 0.5% and 0.15% and in diameter to 0.5% and 1%.

A New Impurity Profile Plotter For Epitaxy And Device

B. J. Gordon and H. L. Stover

University of Southern California
Los Angeles, California

R. S. Harp

California Institute of Technology
Pasadena, California

Impurity profile diagnostics are an indispensable part of semiconductor material and device studies. This paper describes an on-line measuring instrument which is accurate and flexible to use. The instrument can measure doping profiles of a finished device (even when packaged) or of a material slice. The doping profile (N versus W) is plotted directly onto an X-Y recorder. The instrument can cover a variation of at least three decades of doping on a given sample without changing the settings.

The instrument operation is based on simple physics and complex electronics. The key portion of the instrument is a precision automatic capacitance measuring network operating at 1 MHz. A swept DC bias and a controlled AC "dither" signal are applied to the diode simultanously, resulting in DC and AC outputs proportional to C and dC/dV respectively from the capacitance measuring network. These signals are then processed by precision but inexpensive analog circuitry to produce outputs of Log $C^3/(dC/dV)$ and Log 1/C, or more simply Log N and Log W after junction area and permittivity are appropriately incorporated into the display coordinates

$$\left(\text{i.e., } N(W) = \frac{C^3}{e\,\epsilon A^2 \frac{dC}{dV}} \text{ and } C = \frac{\epsilon A}{W}\right)$$

Plots of C versus V are also available from the instrument. Profile data, C, and dC/dV data are presented for several device samples, including well punched-through material and Read diodes. An X-band Read diode profile illustrates the large range of doping measurable, as N varies from 2×10^{16} cm^{-3} and then up again past $10^{17} cm^{-3}$, all in five microns of material.

This new profile plotter was built as a precision instrument for device research. The biggest source of error generally lies in the determination of sample area. A profiler recently reported by the Royal Radar Establishment, England, is based on C-V modulation and analog processing as is the profiler reported here but the design details and performance differs substantially.

Key Words: Impurity profile, doping profile, profile plotter, epitaxial material measurements, C-V technique, automatic profile plotter.

1. Introduction

Impurity profile diagnostics are an indispensable part of semiconductor material and device studies. When the devices under study are microwave devices which utilize epitaxial material with thicknesses from a few tenths of microns to several microns, doping and thickness controls and diagnostics become even more important. There are several convenient and basically accurate techniques for measuring thickness of epitaxial layers. There are also several techniques for estimating some "average" value of doping of an epilayer; most of the techniques assume homogeneous material and are not very accurate.

The two most common means of measuring epilayer thickness are by use of an ellipsometer or by lap-and-stain techniques. The measurements are reasonably accurate for thicknesses on the order of a micron or greater. It is difficult to measure accurately epilayer thicknesses on the order of a few tenths of micron by those techniques, however; it is in fact frequently difficult to ascertain the presence of such a thin epilayer by those techniques.

The most common means of measuring the doping level of epitaxial material are by use of a four-point probe or by use of avalanche breakdown information (perhaps obtained from a three-point probe). These methods and results are generally of limited use and frequently inaccurate. In addition, for thin epitaxial material, the probes may penetrate the epilayer. The thickness of the epilayer should be known in order to properly interpret the results of any probe electrical data. Probe measurements are generally destructive to the material in the region where the measurements were made; etch pits can be found where the probe made contact.

A common electrical measurement which can yield an accurate impurity profile is the capacitance versus voltage data of a back-biased junction. The junction can be a Schottky barrier (i.e., a metal-semiconductor interface) or it can be a P-N junction where one side of the junction is heavily doped in comparison to the other side. Schottky barrier junctions, P^+N junctions and PN^+ junctions are frequently referred to as one-sided junctions since essentially all of the bias-voltage drop is across the lightly doped side of the junction. The capacitance of a back-biased junction is given by

$$C(V) = \frac{\epsilon A}{W(V)} \tag{1}$$

where ϵ and A are the semiconductor permittivity and junction area respectively, and where $W = W(V)$ is the space-charge region width (or depletion width); as the back-bias voltage V is increased, the space-charge width W increases thereby decreasing C. The doping at depth W from the junction is given by

$$N(W) = \frac{C^3}{e \epsilon A^2 \frac{dC}{dV}} \tag{2}$$

where e is the electronic charge and dC/dV is the slope of the C-V characteristic at that particular bias voltage. For a given back-bias voltage then, a corresponding C and dC/dV are measured; the depth W is obtained from eq (1) ($W = \epsilon A/C$) and the doping N is obtained from eq (2). The area of the junction must be accurately known or a substantial error can be introduced into the doping determination. Other sources of error will be discussed later in this article. One key advantage of C-V profiling is that the probing rf voltage (say on a bridge) can be on the order of kT (25 mV); profile errors due to rectification and diffusion capacitance at low bias are thereby minimized.

Another method for impurity profiling also utilizing a one-sided junction is the rf second harmonic method. The best known profiler using this technique is one built by Copeland. [1]* Superimposed on a DC back-bias voltage across the junction is an rf signal from a constant current source. The fundamental rf voltage across the junction and the second harmonic are each measured. The fundamental voltage may be related to the space-charge width W. The second harmonic voltage may be related to the reciprocal of doping, $1/N$. Thus as the DC back-bias voltage is swept, doping versus depth may be continuously plotted on an X-Y recorder. Some disadvantages of the Copeland profiler relate to the fact that the rf second harmonic is a second order effect; a rather large rf

* Figures in brackets indicate the literature references at the end of this paper.

fundamental voltage swing is required in order to generate the harmonic. The large voltage swing causes a loss in spatial doping resolution, especially at low bias voltages. A further disadvantage is the rather complicated filter system required to separate the DC bias, the large rf fundamental, and the small second harmonic; achieving this separation while retaining the proper circuit Q and other circuit parameters for measurement accuracy over many doping and junction area ranges is difficult. An advantage of the Copeland profiler is that if the doping is essentially constant, the second harmonic signal does not change inordinately (i.e., $v_2 \propto 1/N$). Preparation of the semiconductor samples is the same for the C-V profiling technique, and the errors introduced by uncertainty in junction area are the same for both techniques.

In the rest of this article we shall consider only the C-V profiling technique, its advantages, disadvantages, and sources of error. We shall then describe the practical development of a profiler which overcomes the disadvantages of the method and retains the advantages. Finally, impurity profiles of epitaxial material and devices which have been characterized by this new profiler are presented.

2. Details of C-V Profiling

The usefulness and accuracy of the junction capacitance measurement technique for determining doping profiles in a reverse-biased step junction has long been known.[2] To date, due to the difficulty in building an accurate analog profiler, it is still much more common to utilize discrete C-V data in conjunction with a computer program; profiles obtained in this manner represent time-consuming, tedious measurements and plotting; expensive computer time; and frequently slow turn-around of the processed data.

When discrete C-V data is acquired manually (by use of a capacitance bridge) and processed by a digital computer, the determination of dC/dV at a given point is difficult. The capacitance increments are generally small and when plotted would deviate somewhat from a smooth C-V curve. The voltage increments are also small. Thus when ΔC between two points is divided by ΔV between the same two points, an irregular scatter of dC/dV points results; the value of dC/dV so determined can be zero, can have the wrong sign, or can differ greatly from the true value. A plot of N versus W so determined (from eqs (1) and (2)) is very irregular due to the sensitivity of N on dC/dV.

Most computer programs which process C-V data utilize a <u>smoothing routine</u> to avoid those complications with dC/dV; dC/dV is averaged over many points. Another smoothing approach involves plotting C versus V and then making an analytic approximation to the C-V plot; this analytic form is then processed by the computer without dC/dV anomalies. Both of the above smoothing techniques result in a loss of doping resolution; small doping nonuniformities are not detected, and abrupt doping interfaces are frequently "rounded off." An instrument which measures C-V and dC/dV continuously avoids most of the problems with the granular C-V data; in particular, if a dynamic measurement of dC/dV is made, the resolution then depends on the amplitude of the ac voltages across the junction.

A far more serious disadvantage of C-V profiling is the <u>range</u> over which C and dC/dV must vary, even if N is uniform. Let us consider a uniform material simply as an example. As the back-bias voltage is increased, C commonly changes by at least a factor of 10 before avalanche breakdown occurs (i.e., $C(V) \propto (1 + V/\varphi)^{-\frac{1}{2}}$, where φ is the barrier potential). Inspection of eq (2) reveals that while C is changing that <u>one</u> order of magnitude, dC/dV must change <u>three</u> orders of magnitude; the ratio of C^3 over dC/dV, however, must remain constant if the material is uniform. If in addition to the above considerations N also varies several decades (for instance, three decades), then dC/dV may change a total of <u>six</u> orders of magnitude. The electronic development required to have circuits which can measure C and dC/dV accurately over the desired ranges, and to have logarithmic amplifiers and analog circuits which are accurate over at least three, four, or five decades, is far from easy. These developments are possible, however, and are in fact a major portion of the profiler we are reporting here.

3. The Basic Profiler Circuits and Principles

The profiler described here uses automatic, continuous, and accurate C-V measurements in conjunction with precision analog circuitry to process the electrical data; C-V data is converted

to N-W data on-line. In the profiler development we have maintained capacitance measuring simplicity and accuracy at the expense of analog processing complexity; the usage and calibrations of the profiler are simple and convenient, however. With the availability of precision integrated operational amplifiers and temperature compensated log modules, the reduction of principles to practical circuitry has been greatly facilitated.

Recently workers at the Royal Radar Establishment, England, reported the development of an analog C-V profiler;[3] we find several things in common with their profiler, but we also have some important differences.

The basic block diagram of our profiler is shown in Figure 1. The heart of the instrument is the commercially available capacitance meter, the Boonton Model 71-AR. The sample junction (diode) is connected across the Boonton terminals by means of a micro-manipulator probe station. An analog DC voltage proportional to the sample capacitance is available from an output jack at the rear of the Boonton. As the reverse DC bias across the junction is swept, analog outputs proportional to C and V are available. An output proportional to a dynamic measurement of dC/dV is made possible by superposition of a 1 KHz ac "dither" signal on the DC bias; the Boonton was modified to pass the 1 KHz signal. At the output of the Boonton then, a 1 KHz signal proportional to dC/dV is easily separated from the DC signal proportional to C . These analog outputs are the essential data which are processed to obtain N and W . The key block diagrams and their operation are discussed separately in the following paragraphs.

3.1 The Boonton Capacitance Meter and Its Modifications

The Boonton meter essentially measures the ac current passing through the sample diode at the 1 MHz crystal controlled frequency. The 1 MHz probing signal is set sufficiently low (≈15 MV) that neither rectification nor second harmonic generation is a problem. The capacitance measuring network has many C ranges, up to 1000 pf, and is quoted to have an accuracy of 1%-2% by the manufacturer depending on the Q of the device measured. The meter features phase-locked detection which allows accurate capacitance measurements down to a Q of 3; in terms of sample-junction leakage, this corresponds to 100 μa of current for a 10 pf capacitance at 5 V reverse bias. The Boonton meter also has the advantage of three-terminal capacitance measurements and the capability to null our stray capacitance, as well as package capacitance of a finished device.

Modifications of the Boonton are only to improve the instrument's response to the 1 KHz signal and do not affect the normal operation of the meter. Low pass filters in the Boonton were redesigned to accommodate a pass band at 1 KHz. The 1-KHz signal as it comes out of the capacitance meter is amplified and rectified by a phase-lock detector.

3.2 The 1 KHz Modulator for dC/dV

As discussed in an earlier section, dC/dV can change more than five orders of magnitude as the sample diode is back-biased. For a fixed input "dither" signal then, the output 1 KHz signal can drop to an exceptionally low value. Experimentally it was found on an earlier version of our profiler that the output 1 KHz signal would drop so low that it was at the same level as background noise and pickup; in fact, if the epitaxial material of the junction fully depletes before avalanche breakdown (i.e., if the epitaxial region "punches through"), dC/dV should drop to zero. Since dC/dV enters into the denominator of eq (2), the output value for N depends in a very sensitive way on noise and pickup on the dC/dV signal when the signal is near zero.

To avoid the problem outlined above, whereby the large range covered by dC/dV results in the 1 KHz signal dropping to a value so low that accuracy is impaired, a modulator was developed which increased the input dither signal while the output 1 KHz signal was decreasing. The input dither signal is proportional to ΔV ; the output 1 KHz signal is proportional to ΔC . The recovery of these two signals thus yields dC/dV . Two phase-sensitive detectors are employed to recover the ΔC and ΔV signals. The detectors derive their reference signal from an FET amplitude-controlled 1 KHz oscillator. Their circuitry is straight-forward, employing easily available integrated circuits and annular switching transistors. An integration constant of 0.25 second is used.

As the ΔC signal is amplified and rectified by the phase-lock detector, the rectified signal

is maintained at the same order of magnitude by the modulator which increases the input dither signal. The amplitude limits of the modulator are completely adjustable; a convenient range extends from 20 mV p-p (peak-to-peak) to 2 V p-p. At low bias levels, the dither voltage is thus on the order of kT; as the reverse bias becomes large (and dC/dV small) the dither voltage can increase to 2 V p-p, but is still only a small fraction of the DC bias voltage.

3.3 Logarithmic Amplifiers and Arithmetic Unit

The logarithmic amplifiers' performance is critical to the acquisition of accurate profile plots. In the finalized profiler, commercial temperature-compensated logarithmic modules were employed in conjunction with FET operational amplifiers with low input bias currents in order to achieve accurate logarithmic tracking over five decades.

The arithmetic unit consists of simple resistive voltage-divider networks made from precision resistors. The voltage inputs of $3 \ln C$, $\ln \Delta V$, $\ln \Delta C$, and $\ln C$ are combined appropriately (as per eqs (1) and (2)) to yield two outputs proportional to N and W.

3.4 Calibration Procedure

The signals at the output of the arithmetic unit are $3 \ln C + \ln \Delta V - \ln \Delta C$ and $- \ln C$, which are seen from eq (2) to differ from the actual $\ln N$ and $\ln W$ by constants involving ε, A and e. Since the outputs are plotted on an X-Y recorder with logarithmic graph paper, calibration involves adding the appropriate constants or simply shifting the X and Y position controls on the recorder. The calibration procedure requires the knowledge of the area and dielectric constant of the material to be measured beforehand.

Figure 2 displays the front panel of the profiler, including the Boonton capacitance meter, the power supply for the junction back-bias voltage, and the X-Y recorder. For our applications it was convenient to use 3-cycle by 2-cycle log-log paper. The recorder was arranged so that N could cover 3 cycles and W could cover 2 cycles. The gain controls of the recorder were adjusted so that a decade change from a logarithmic amplifier corresponded to a decade change on the graph paper; those controls were then locked.

The zero-bias capacitance of the sample junction is measured and the correct scale of the capacitance meter is selected; the capacitance scale also enters into the calibration. The function switch seen on the panel is then turned from the "OPERATE" position to the "CALIBRATE" position, and fixed calibrated C and dC/dV signals are automatically introduced at the input to the analog circuitry; the output to the X-Y recorder then corresponds to a calibrated doping, N_{cal}, and a calibrated depletion layer width W_{cal}, given by

$$N_{cal} = \frac{V_{C_{max}} C_{max}^2}{\varepsilon e A^2} \tag{3}$$

and

$$W_{cal} = \frac{\varepsilon A}{C_{max}} , \tag{4}$$

where $V_{C_{max}}$ is the DC analog voltage out of the capacitance meter corresponding to the full scale capacitance C_{max}. The position controls of the X-Y recorder are used to align the W_{cal} and N_{cal} point on the log-log graph paper. This completes the calibration. The function switch is returned to "OPERATE". The reverse bias voltage can now be varied manually or swept automatically to complete the profile.

4. Epitaxy Characterization

The purpose of developing our profiler was to characterize epitaxy for microwave device applications, and to measure doping profiles of finished devices. It was anticipated that multi-epitaxial material and devices would be studied, and that doping levels on a given sample could vary three or four decades. In addition, we wished to retain the capability of subtracting out package parasitic capacitance. It was anticipated that impurity concentrations between 10^{14} and 10^{18} per cm^3 would be the most prevalent.

We have characterized much epitaxial material with the profiler, much of it "garden variety" graded material. The profile characterizations we report here are those which have something unique about them. They therefore either demonstrate the versatility of the profiler, or they indicate the capability of a certain epitaxy growth technique or of a certain epitaxial reactor. The thickness and doping of all samples have been cross-checked by other measurement techniques whenever possible.

The sample preparation has been of two types:

1.) Metal dots evaporated through a mask or formed by photolithographic techniques act as Schottky barrier junctions. This method of preparation is not in itself destructive to the material; the metal can be removed. A Schottky barrier under reverse bias tends to be leaky due to edge effects; if the sample is leaky at too low a back-bias voltage, an accurate profile may not be obtained. The Boonton capacitance meter with its low-Q capability allows us to obtain useful profiles for junctions that are not exceptionally leaky.

2.) A P^+N or an N^+P step junction is formed by diffusion, ion implantation, or by growth of a heavily doped epilayer. Mesa junctions are then etched. A sample prepared in this manner can reach the theoretical avalanche breakdown without leakage. This method is destructive to the material sample unless mesa diodes are the desired end product, which is frequently the case. This preparation method allows accurate area determination and leads to the most accurate characterizations; for that reason, most of the samples reported here had P^+N junctions.

For both types of sample preparation, the current-voltage characteristics of the sample diodes were monitored carefully before and after profiling. This was done to insure that junction leakage was not a problem.

4.1 The 4-Layer Read Diode

A unique material sample which demonstrated the versatility of the profiler is the Read microwave oscillator diode. The Read diode has a total of four layers, two of them being the N^+ substrate and the P^+ layer, as seen in the inset to Figure 3. All of the voltage is dropped across the center two N and N^- layers; these are the two layers which can be characterized by C-V profiling. At low reverse bias, the high electric field is entirely in the N region; as the reverse bias increases, the N region depletes and the field extends into the N^- region. The capacitance changes rapidly with voltage when the electric field first reaches into the N^- layer, as seen in Figure 3.

Figure 4 presents the doping profile of the active regions of a multi-epitaxial Read diode which operated in X-band (≈10 GHz). The circled dot near the upper left-hand corner of the figure is the calibration point (N_{cal}, W_{cal}). The N layer is nominally a micron thick and doped at $2(10)^{16}$ cm^{-3}. The doping drops more than 2 orders of magnitude over the next micron, down to $8(10)^{13}$ cm^{-3}. The doping remains near 10^{14} cm^{-3} for the next 3 microns and then goes off scale to 10^{19} cm^{-3}, which of course represents the substrate.

The Read diode described above was fabricated in a vertical reactor using silicon tetrachloride. The N^{--}, N and P^+ layers were grown on the N^+ substrate. There was no other high temperature processing step, such as diffusion. The devices fabricated from this material showed excellent rf performance as microwave generators.

4.2 Silane Grown Epitaxy

Figure 5 shows a profile of some epitaxy grown in a vertical reactor using silane at 1000° C.

The growth rate was 0.25μ per minute. A P^+ region 0.1μ deep was formed by ion implantation of boron, followed by suitable annealing. The epilayer is reasonably uniform near $3(10)^{15}$ cm^{-3} over the first 4.5μ; near the N^-N^+ interface, however, there is a region approximately 0.3μ thick where the doping drops to $1.5(10)^{15}$ cm^{-3}. This lightly doped interface region actually exists on this sample; it might not have been detected if granular C-V data had been processed by a computer program with a smoothing routine.

That the lightly doped region does exist may also be seen from the discrete raw C-V data presented in Figure 6. These data were obtained on an entirely different Boonton bridge. If the material were uniform, the slope of the straight line would be $-\frac{1}{2}$. A slope less steep than $-\frac{1}{2}$ indicates a positive doping gradient (i.e., doping increasing with depth). A slope steeper than $-\frac{1}{2}$ indicates a negative doping gradient. The data points below the solid line near the "punch-through" point indicate a steeper slope than $-\frac{1}{2}$ and therefore a transition to a lighter doping.

Figure 7 displays a profile of another sample grown in the same reactor and at the same temperature and rate as the above sample. This sample is only a micron thick, however, and is doped near $4(10)^{16}$ cm^{-3}. Again near the N^-N^+ interface a 0.3μ region is observed with lighter doping than over the rest of the epilayer. Figure 8 shows raw C-V data from an independent C-V measurement on a capacitance bridge. Again the change in slope indicates the lighter doped region at the substrate interface.

In separate papers presented at the symposium, similar lightly doped regions at interfaces were observed by spreading resistance measurements [4] and by infrared reflectance techniques [5]. The cause for the formation of this lightly doped region during epilayer growth remains a matter for conjecture.

4.3 Submicron Epilayers

Figures 9 and 10 show profiles of two submicron epilayers grown in a horizontal reactor using silane. The two samples were grown at different times and for different applications but otherwise appear very similar. The epitaxy is nominally 0.7μ thick with a $4(10)^{16}$ cm^{-3} doping level. These samples had ion implanted junctions 0.1μ deep. Both samples underwent avalanche breakdown just after full depletion of the epilayer.

4.4 Hyperthin Epilayers

It should be recalled that most of the common techniques for measuring epilayer thickness and doping do not apply well for epilayers which are only a few tenths of a micron thick. The C-V profiling technique, however, is able to give considerable information about these hyperthin layers. Due to the depletion by the barrier potential, the profiles (which begin at zero volts bias) do not show all of the region close in to the junction.

Figure 11 shows the impurity profile of an epilayer which is at most 0.3μ thick. The material is heavily graded. An extrapolation of the curve to the material surface suggests a surface concentration near 10^{15} cm^{-3}. This material sample was grown in a horizontal reactor using silicon tetrachloride. The material was purchased commercially; both doping and thickness are very remote from the seller's specifications, probably due to outdiffusion from the substrate and to autodoping.

Figure 12 shows a hyperthin epilayer only 0.1μ thick. The material appears to have considerable detail, as may be seen by expanding the scale on the X-Y recorder. The material within the first 0.08μ of the junction is depleted by the barrier potential; that material must have a doping lighter than 10^{17} cm^{-3}. The profile shows considerable detail on the doping tail; that detail may be examined more closely by increasing the gain on the X-Y recorder to expand the W scale, thereby yielding the second curve shown in Figure 12.

The detail displayed on both curves in Figure 12 was repeatable and did not appear to be due to leakage or anything electrical. Whether the detail was put there during epilayer growth, or whether it was put there during later processing is uncertain. This sample was prepared by ion implantation of boron to a 0.1μ depth, followed by a 600° C annealing for 15 minutes.

5. Error Analysis

Two major classes of errors should be discussed in relation to the profiler's performance--errors arising from instrumentation and errors due to physical measurements used in calibration of the profiler. The analog circuitry can be conservatively rated for 5% accuracy.

The integration time constant of the phase sensitive detectors, 0.25 sec., can be a source of considerable error if the sweep rate on the DC bias voltage is too fast, but this is not a serious source of error.

The Boonton capacitance meter has a rated accuracy of (0.5% of reading + 0.5% + (3/Q)% full scale - 0.01 pf) which could account for up to 2% error under the most unfavorable low-Q conditions. For calibration purposes the DC analog output from the meter must be determined and this is done to an accuracy of 0.2%.

In actual use, however, the greatest source of error was found to be area measurement of the devices under test. In most cases the diameter of the junction was measured under a metallurgical microscope usually to within 2 or 3 microns resolution. Note that in eq (2) the A^2 factor involves the fourth power of the diameter. For a 100μ junction diameter as much as 12% error could be realized in doping and a 6% error in junction depth. An overall worst case estimate of accuracy would be 15%, dropping to less than 10% with more precise junction area determination.

6. Summary

An impurity profile plotter has been developed which is on-line, accurate, and flexible to use. The profiler operation is based on simple physics and complex, although straightforward, analog circuitry. A novel feature of this profiler is a modulator which allows dC/dV to be measured accurately over six decades. The modulator also allows the profiler to retain high resolution near zero bias, and allows measurement of doping over many decades from a single calibration.

A broad range of material samples have been characterized. Profiles are presented here of multi-epitaxial samples, silane grown samples, and submicron layer samples. On many of the silane vapor grown N on N^+ samples, a thin (≈0.3μ) high resistivity layer was observed near the NN^+ interface. The profile detail on some of these samples indicates the doping resolution of the profiler.

This profiler has been in regular use in the USC Microwave Device Laboratory for the study of epitaxy and devices in both silicon and III-V compound semiconductors. It is simple to use and has required very little alignment since the initial assembly. The initial assembly and alignment required considerable care, but from then on the instrument is virtually trouble-free. The preparation of samples for profiling also requires great care, but this is common to all profilers.

Acknowledgment: This research was sponsored by the Joint Services Electronics Research Program under Contract No. AF-OSR-69-1622-A. The authors are indebted to the Hughes Aircraft Research Labs for assistance in preparing some of the samples measured and reported here.

7. References

[1] Copland, J. A., A technique for directly plotting the inverse doping profile of semiconductor wavers, IEEE Trans. Elect. Dev. ED-16.

[2] Hildebrand, J., and Gold, R. D., Determination of the impurity distribution in junction diodes from capacity-voltage measurements, RCA Review, June 1960.

[3] Low cost British instrument plots impurity profiles of semiconductors, Electronics International, May 31, 1969.

[4] Severin, P. J., On the interpretation of some measurement methods for epitaxially grown layers, Proceedings of the 1970 Symposium on Silicon Device Processing.

[5] Schumann, P. A., Jr., Measurement of semiconductor epitaxial layer thickness by infrared reflectance, Proceedings of the 1970 Symposium on Silicon Device Processing.

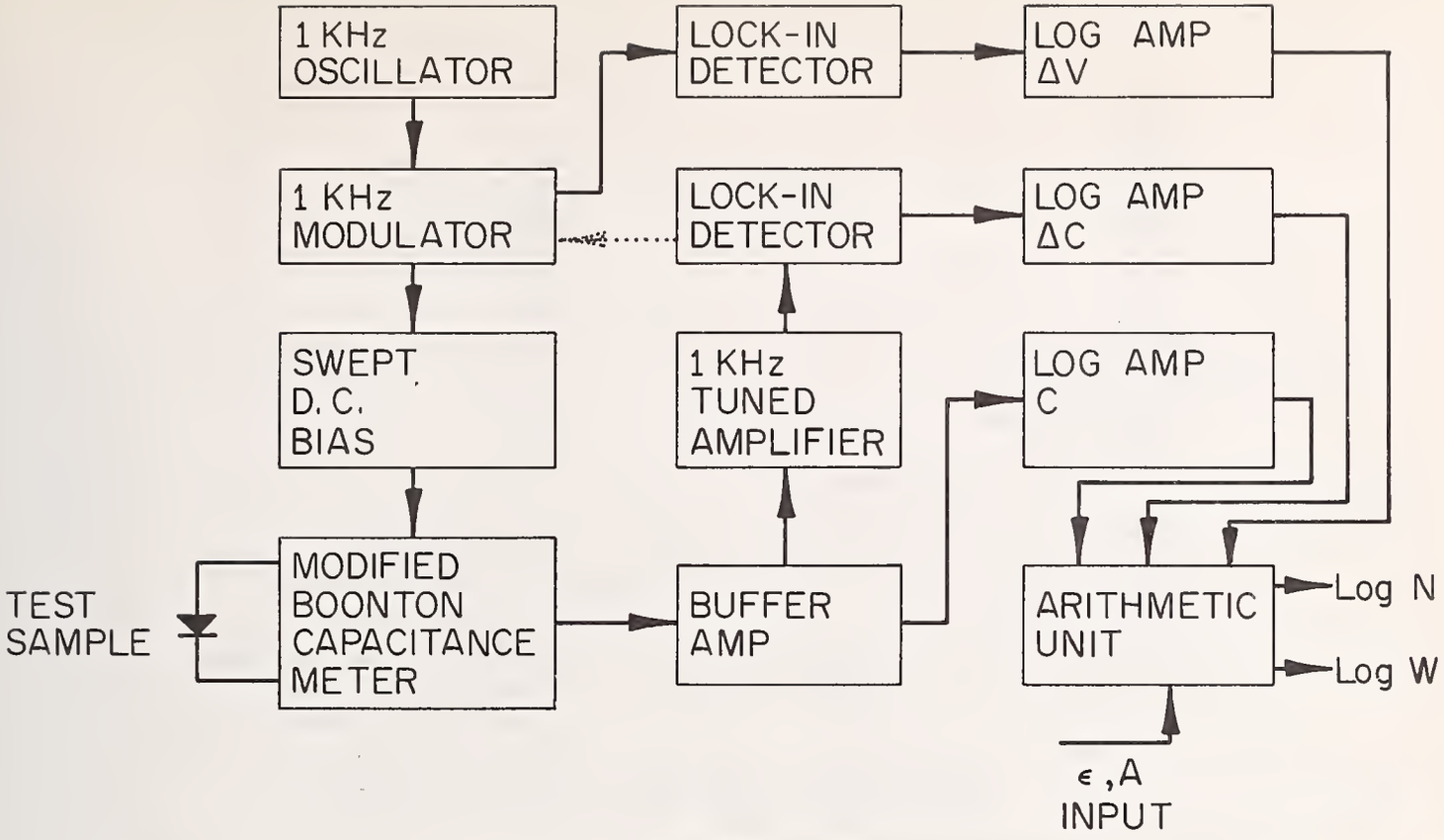


Fig. 1. Basic block diagram of profile plotter. Solid lines indicate main signal paths. The dotted line is the modulator feedback signal path.

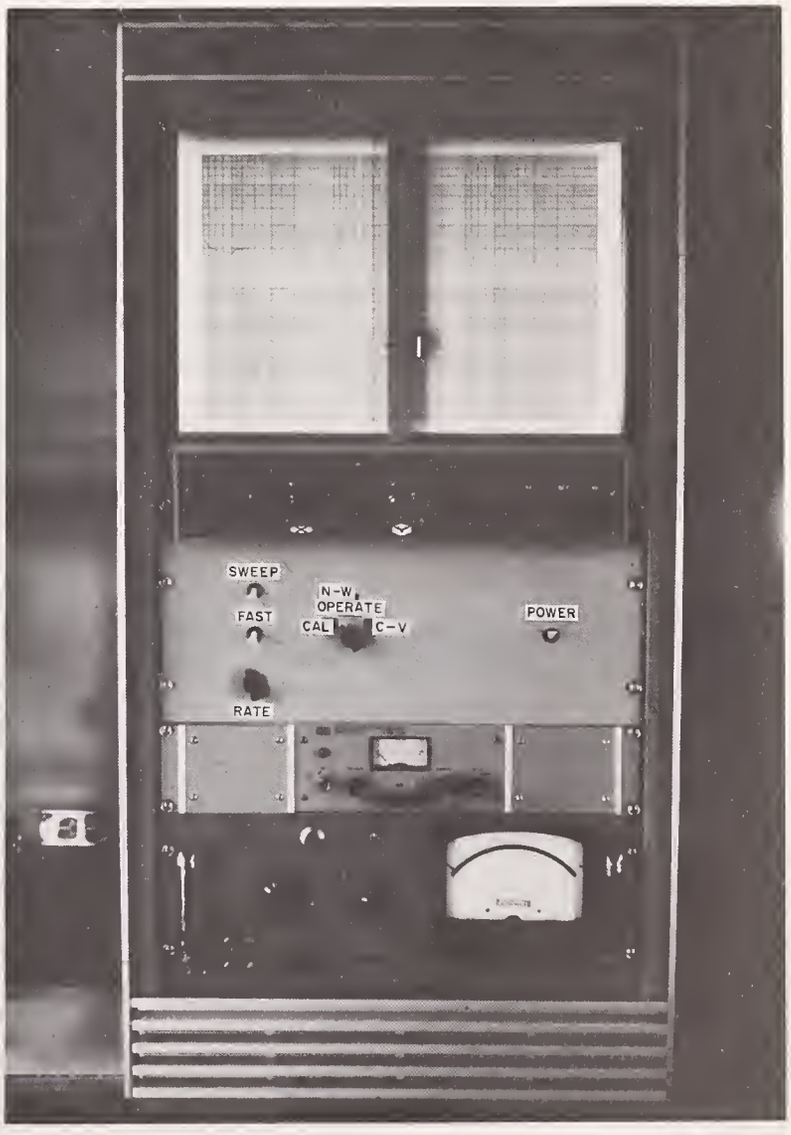


Fig. 2. Photograph of the front of the profile plotter showing, from bottom to top, the Boonton capacitance meter, a power supply, the main panel with sweep rate, function and power switches, and the X-Y plotter.

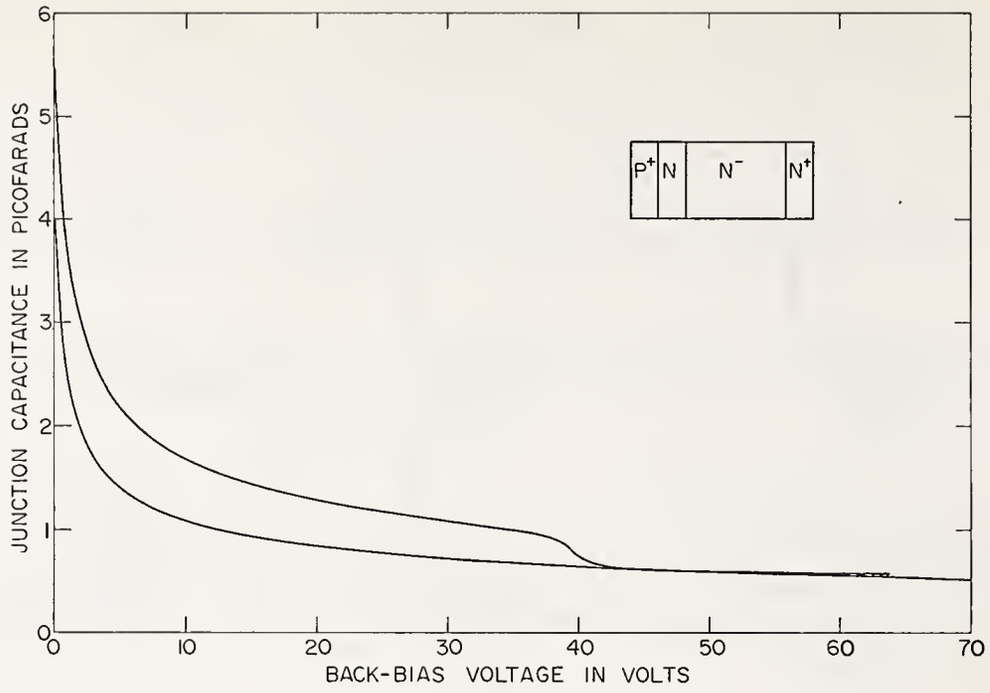


Fig. 3. C-V curves of a Read diode ($P^+NN^-N^+$, upper curve) and a regular IMPATT diode ($P^+N^-N^+$, lower curve). The inset illustrates the Read 4-layer structure.

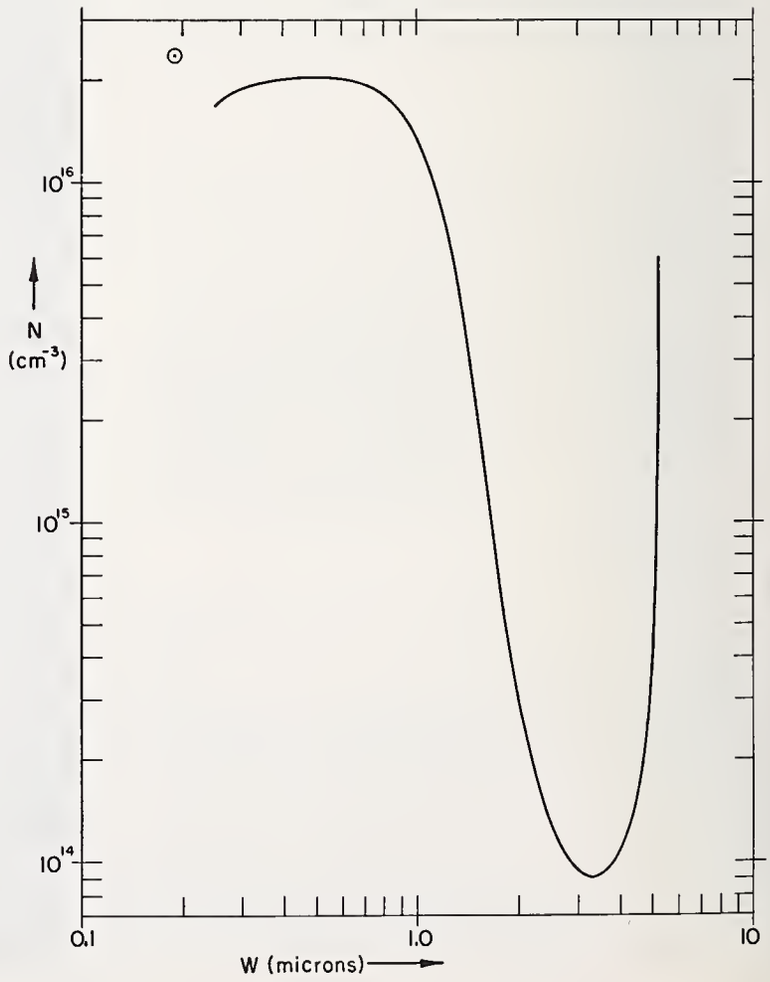


Fig. 4. Doping profile of X-band Read diode grown by multi-epitaxial techniques.

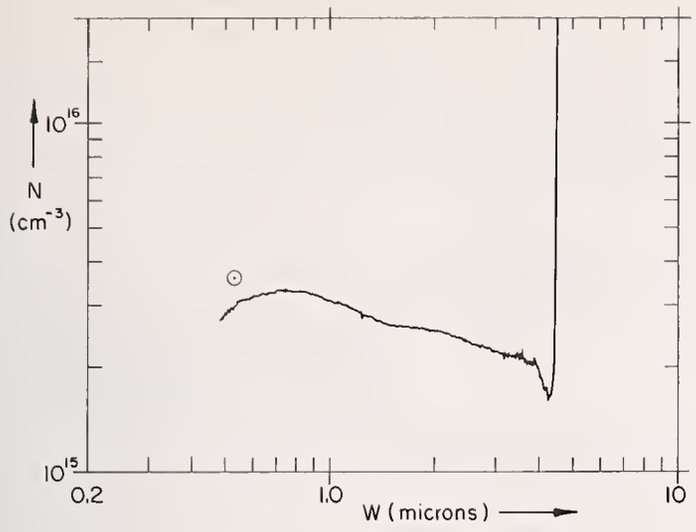


Fig. 5. Doping profile of silane grown epitaxy (1000 °C, 0.25μ/min.). Note lighter doped region at NN^+ interface.

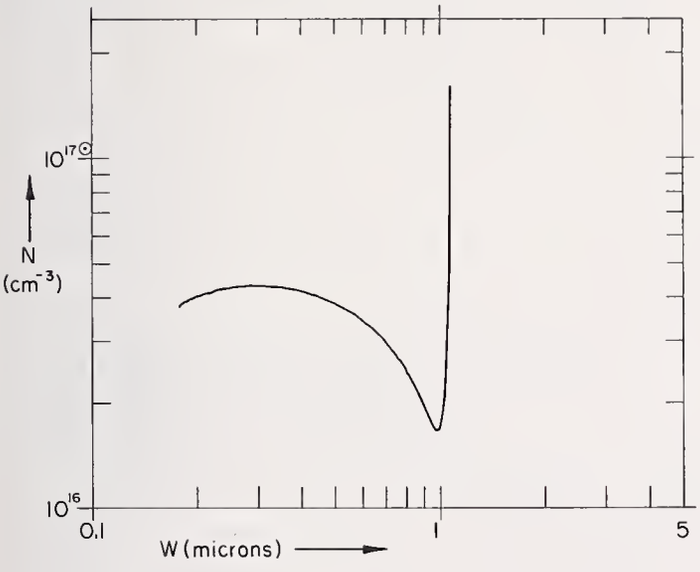


Fig. 7. Doping profile of a 1 micron layer of silane grown epitaxy (1000 °C, 0.25μ/min.). Note the distinct drop in doping near the NN^+ interface.

Fig. 6. Log C vs. log V plot of discrete C-V data for material whose doping profile is shown in Fig. 5.

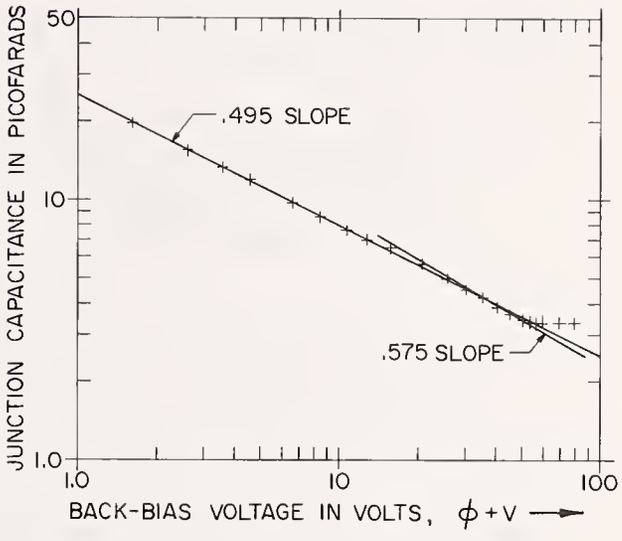
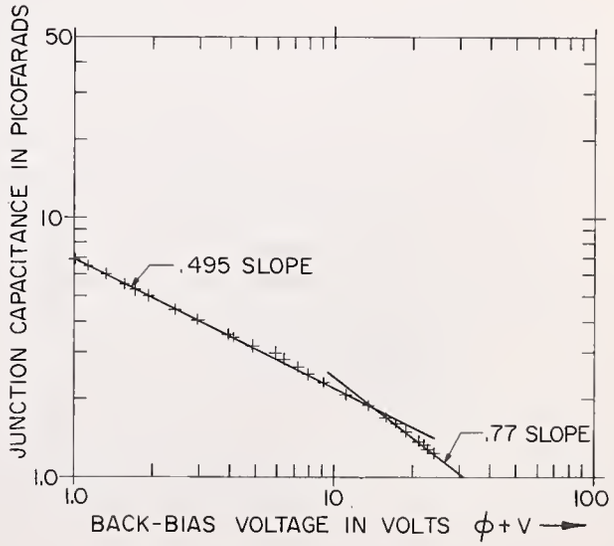


Fig. 8. Log C vs. log V plot of discrete C-V data for material whose doping profile is shown in Fig. 7.



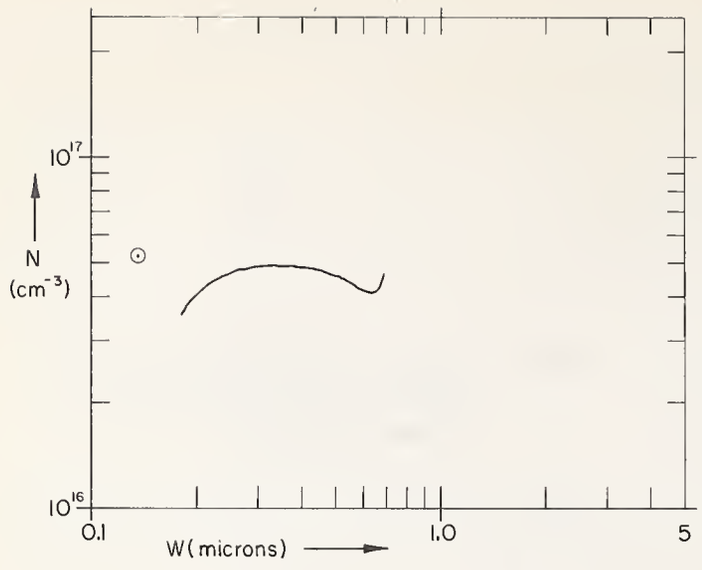


Fig. 9. Doping profile of sub-micron silane grown epitaxy used for millimeter-wave devices. Depletion and avalanche occur simultaneously.

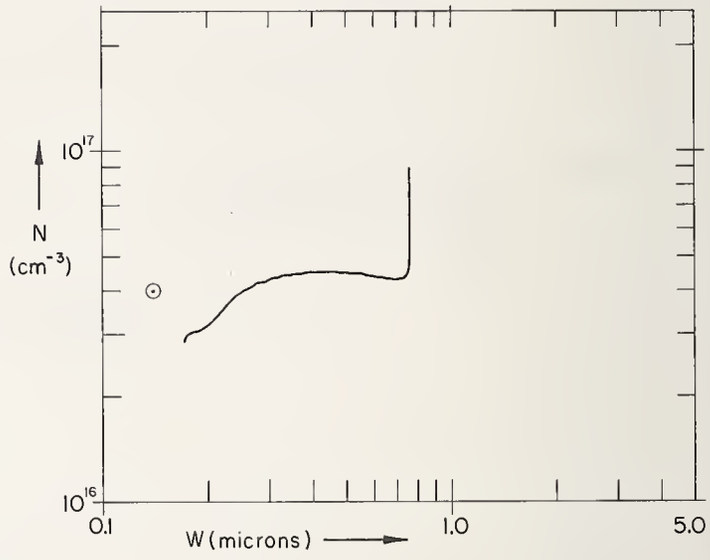


Fig. 10. Doping profile of punched-through silane grown epitaxy used for millimeter-wave devices.

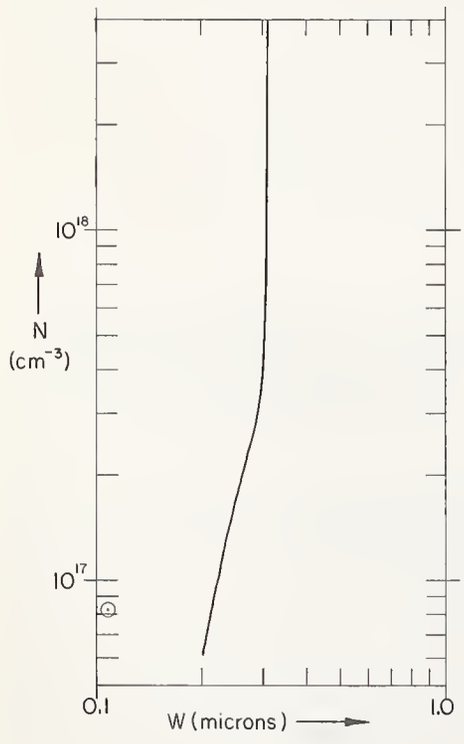
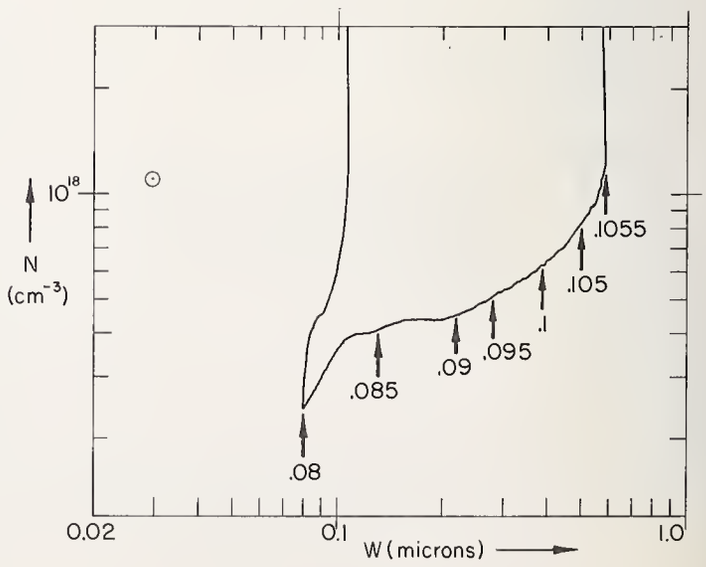


Fig. 11. Doping profile of hyperthin, graded epitaxy, $SiCl_4$ grown.

Fig. 12. Doping profile of hyperthin, graded epitaxy, $SiCl_4$ grown. Labeled curve is expanded to show detail.



STRUCTURAL FAULTS IN EPITAXIAL AND BURIED LAYERS IN SILICON DEVICE FABRICATION

P. Wang*, F. X. Pink and D. C. Gupta

The Waltham Research Center of the
General Telephone and Electronics Laboratories Incorporated
100 Sylvan Road, Woburn, Mass. 01801

Introduction

It has been well established that structural imperfections in silicon material affect its physical, chemical and metallurgical properties. Since device fabrication processes include physical, chemical and metallurgical treatments, the structural changes throughout the entire sequence of processing steps should be measured. At present, the ability to describe simply, precisely and completely the structural properties of silicon is still limited. A review of material problems for silicon power devices was written by John (1). Observations and properties of lattice defects in silicon were recently reviewed by Queisser (2). In small-area devices and integrated circuits (IC), similar problems exist, and processing difficulties such as photomask precision and alignment become more critical for the fine geometry and high component density used in IC and large scale integration (LSI). Bearing in mind the cause and effect relationship between process-induced defects and lattice imperfections and between impurities and defects, the complexity of IC and LSI designs can produce structural imperfections in a finished wafer so severe as to defy identification and correlation. In the present paper, the situation is simplified by excluding geometry problems and chemical impurities except when these affect silicon lattice perfection. The emphasis will be on structural defects observed in silicon material in device fabrication especially after epitaxial layer formation and after diffusion. The defects are correlated with the yield and performance of the product using a NAND/NOR IC as a demonstration vehicle.

Techniques for Detecting Silicon Structural Defects

Techniques commonly used for the detection of structural defects for single crystal silicon and similar material may be loosely classified into three groups:

1) Surface characterization methods, e.g., scanning electron microscopy (SEM) and replica electron microscopy (REM).

2) Bulk and surface characterization methods, e.g., x-ray transmission and reflection topography.

3) Thin film characterization methods, e.g., transmission electron microscopy (TEM).

---

*Present address: Texas Instruments, Inc., Box 6325, Dallas, Texas

X-ray transmission and reflection topography is most valuable for revealing defects and strains in silicon produced prior to and resulting from production processing. Transmission or Lang topography provides information in the bulk and surface of the material, whereas reflection or Berg-Barrett topography samples shallow surface layers in the single crystal material. A review of experimental procedures was given by Austerman and Newkirk (3) and a technical discussion by Lang (4). The Lang technique requires more complex instrumentation and for the same area requires longer exposure times than the Berg-Barrett technique. Figure 1 is a description of a Berg-Barrett scanning camera designed to obtain topographs of silicon wafers up to 6.3 cm in diameter (5). Employing Cu K_{α}, full-area topographs of two-inch-diameter silicon wafers may be obtained on 25μm thick nuclear emulsions. Direct display of an x-ray diffraction image by means of a conventional phosphor vidicon and closed circuit television (CCTV) has not really been successful, although the recently developed silicon diode array vidicon has shown promise for displaying x-ray diffraction images with improved contrast and resolution (6). Some of the advantages of an x-ray CCTV display system are the following: the study of dynamic events, rapid photographic recording, real time information for quick decision and enlarged view for simultaneous observation at different locations. This is an area where further work may be very rewarding.

Infrared microscopy has been a powerful tool to study material structural imperfections, as shown by the early work of W. C. Dash (7) and the recent work of Dale (8).

Infrared microscopy was used in the present work to detect structural defects such as impurity precipitates and segregates, second phases, scratches, etc. A He-Ne laser with emission at 1.15μm was evaluated as an alternate source of radiation in place of the tungsten lamp. However, considerable difficulty was encountered in the use of this nearly monochromatic, well-collimated light, mainly due to scattering in the light path and interference from the backside surface while examining the top surface. An infrared scanned laser microscope (9-11) using a 3.39μm emission from the He-Ne laser was also used to examine material homogeneity. A shadowgraph display of variation in the IR transmittance or reflectance of a specimen is obtained on the oscilloscope. In addition, the spreading resistance probe has been applied for the detection of inhomogeneity in silicon wafers after various processing steps correlated to microdefects (12). Finally, optical microscopic examination of etched specimens can also furnish very useful information.

Structural Faults in Silicon Epitaxial Materials

In the high-temperature silicon epitaxial deposition process commonly used in the industry, a variety of faults may be introduced in the deposited structure. Tables 1, 2 and 3 list commonly observed structural defects in polished silicon wafers, epitaxial silicon and wafers after junction diffusion. Point defects are not listed because their presence has not been directly observed. However, they may play important roles in impurity diffusion and affect the finished products significantly. The structural faults in the epitaxial silicon wafers used in IC fabrication, for instance, are influenced by surface preparation, buried layer and isolation diffusion and epitaxial deposition conditions. These are strongly interdependent, cumulative effects between the processing parameters and structural perfection.

Stacking Faults and Twinned Regions in Epitaxial Films

Stacking faults are the most common defects in silicon epitaxial films. In (111)-oriented silicon they are generally initiated at the layer-substrate interface, grow along inclined $\{111\}$ planes and intersect the surface along $\langle 110 \rangle$ directions, forming lines, V-bends and single and complex triangular configurations, when viewed at low magnifications. Their origin and growth mechanisms were studied by Booker (13) and by Mendelson (14). The presence of stacking faults has been attributed to various types of surface perturbations such as residues, films, embedded particles or particulate matter from the reactor ambient. Mechanical damage and nonuniform surface features may enhance stacking fault formation. Examples are seen in Figs. 2 and 3 where stacking faults may occur either in selective areas or in general over the entire epitaxial film. High concentration boron diffused regions used in isolation techniques may also be the cause of high densities of stacking faults.

The presence of twinned regions is another fault usually observed in epitaxial films grown on (111)-oriented substrates; appearing as pyramidal growth or microtwin lamella. Booker and Joyce (15) have shown their origin to be due to beta-silicon carbide contamination on the silicon surface in this case. Additionally, foreign material on substrate surfaces is known to induce these growth spikes.

The occurrence of stacking faults and growth pyramids has been greatly reduced with improved substrate surface preparation and deposition processes.

Growth Striae

Growth striae are commonly observed in highly doped, single crystal semiconductors. Impurity heterogeneity is one of the causes for their formation as discussed by Witt and Gatoes (16). The striations in the substrates may be detected by means of x-ray transmission topography and 3.39μm scanned laser microscopy (10,11) and the fluctuation in the resistivity may be determined by spreading resistance measurements (9).

When the materials having growth striae are used as substrates in epitaxial deposition, the striae propagate into the layer and introduce resistivity variation and structural inhomogeneity (17), thus affecting the electrical properties of the device. To determine the propagation of the growth striae, n/n+ and p/p+ epitaxial structures were selected, n+ and p+ substrates being heavily doped ($n \simeq 3 \times 10^{19}$ atoms/cm^3) cut in a $\langle 100 \rangle$-direction from $\langle 111 \rangle$-grown silicon single crystals. These wafers were then beveled and stained to reveal the layer-substrate interface. Fig. 4 shows the beveled surface of the n/n+ epitaxial wafer depicting the evidence of propagation of striae from substrate into the layer. Fig. 5 gives the top view of the epitaxial layer surface showing the regular arrays. Bright and dark field photomicrographs (Fig. 6) present the beveled surface of the p/p+ epitaxial wafer showing the propagation of dislocation arrays from the substrate into the layer as well as mismatched dislocations at the layer-substrate interface which shows more prominently in the dark field micrograph (Fig. 6b). Thus, it is believed that the propagation of striae into the epitaxial layer is associated with a defect mechanism.

Slip Dislocations

The generation of slip dislocations is another problem in silicon epitaxial material, due to rapid and nonuniform (18) cooling, mismatch of silicon and SiO_2 thermal expansion coefficients and lattice mismatch between n+ and p+ regions and the deposited layers (19). Fig. 7 is a Berg-Barrett x-ray topograph (20,21) showing slip formation in an epitaxial IC wafer after isolation diffusion. The transmission topograph of a plastically deformed silicon wafer with severe cross slip is shown in Fig. 7 . Device yields were found to be lower in the slipped regions as compared to other regions.

Pattern Washout in Silicon Epitaxial Wafers

Another problem encountered in IC fabrication is pattern washout after epitaxial layer deposition on substrates with selective buried diffusion for transistor collector regions. This phenomenon is covered in the literature (22-24). Briefly, during the isolation buried layer diffusion, small depressions in the substrate occur in the diffused areas from the etching action of the diffusing ambient. Upon subsequent epitaxial growth from a $SiCl_4$ source both the etch and growth rates occur slowest along $\{111\}$ facets on the concave or convex areas of the depressions. The result is a washout or loss of definition of the depression pattern in a defined crystallographic direction. The loss of this definition makes it difficult or impossible for subsequent accurate mask alignment. Washout may be reduced by orienting the surface 3° off (111) toward a nearest (110) on (111) slices and orienting the $\langle 001 \rangle$ direction to within 0.25° of the surface normal on (001) slices. The use of silane as a source for epitaxial deposition also reduces or eliminates washout. Another critical effect, which depends upon the etch-growth mechanism, is a shift and/or distortion of the

resultant configuration on the epitaxial layer surface from the configuration on the substrate surface. A large shift may cause shorting in tight tolerance devices. Fig. 8 is an example of a severe case of washout.

Structural Faults and Yield Correlation

Using NAND/NOR gates as test vehicles observed faults in silicon wafers were correlated with the yield and performance of circuits. Three groups of silicon wafers were cut from the same p-type, (111) single crystal of 10 ohm-cm resistivity. The back surfaces of the wafers were given various surface treatments; such as, lapping, etching and chem-mechanical polishing. Dislocation density was of the order of 1500 to 2000/cm^2. These wafers underwent the processing steps given in Fig. 9. X-ray and optical examinations of wafers were carried out at the conclusion of every batch processing step.

After the completion of the processing steps, the wafers were probe-tested on a Teradyne J259 automatic testing system on a "go"-"no go" basis. Upon rejection, the cause was registered. The yields on the basis of probe testing are given in Table 4.

Figs. 10-14 are the x-ray transmission topographs showing typical faults in five wafers after various fabrication stages. Fig. 10 shows a severely plastically deformed wafer with a chemically polished back surface. The topograph shows a high density of slip dislocations after cross-over diffusion. Fig. 11 shows the high density of slip dislocations after the base diffusion in a wafer with a mechanically-chemically polished back surface. Probe yield was found to be zero in these two wafers. Fig. 12 is the topograph of a bent wafer with a radius of curvature of about 500 cm. This wafer had a lapped back surface. Fig. 13 shows a wafer with a small slipped region. The wafer had a chemically polished back surface. Probe yield was found to be 31 percent. Fig. 14 shows a wafer with a mechanically-chemically polished back surface. No gross slip was detected and the probe yield was 56 percent.

IC units were packaged from the wafers in each group of different back surface treatments and tested for dc characteristics at -55, 0, 25, 75 and 125^{o}C, ac switching characteristics, transfer characteristics and other parameters such as output sink current, V_{cc} vs. I_{cc}, inverse beta, output and input leakage currents. The following conclusions may be drawn from these measurements:

1. Distribution of failures in parameters was higher in circuits fabricated from the wafers with chemically and mechanically-chemically polished back surfaces.
2. Circuits fabricated from the wafers with lapped back surfaces had lower output sink currents, soft junctions of the output transistors and higher output leakage currents.

Two hundred circuits made from acceptable wafers, judged by in-process structural defect measurements, were placed on a 1000 hour 125°C ring counter life test; there were no electrical failures.

Discussions

The quantitative analysis of structural faults in a silicon wafer can be very complex. It necessitates the estimation of geometric properties of micro-structures of a three dimensional specimen from observations made on a two dimensional section (such as in the case of optical microscopy, SEM, etc.) or on a projection of these structures upon some plane of observation (25) such as in case of x-ray topography, TEM, scanned laser IR microscopy). Frequently, a clear distinction between the various types of faults is impossible due to interactions and overlapping of these faults. When one particular type of fault exists in a specimen, a quantitative estimation is not usually difficult. However, one must state the technique used and bear in mind that the result depends on the measurement method. For example, dislocation density is usually determined by etch pit counting and reported in the form of etch pit counts per unit area. If x-ray topography or TEM is used, the sum of line lengths per unit volume projected onto one plane of observation is measured. The unit is then the length of a linear feature per unit volume. Estimations from optical or x-ray techniques are based on different approaches and the results are not generally similar.

An estimation of linear faults such as dislocation loops and complex dislocations are not easy to describe. Stacking faults generally seen by optical microscopy are two-dimensional images projected from three-dimensional geometric features. Twinned regions, inclusions, precipitation and other bulk defects are truly three-dimensional where the shape, orientation, size distribution, density and chemical and metallurgical nature of the faults should all be determined for full characterization. This is, for all practical purposes, impossible to accomplish manually by a human observer. It may be feasible to some extent by using a CCTV inspection system (26).

The counting procedure as it is used for the determination of dislocation etch pit density may be followed in the determination of process-introduced structural faults such as stacking faults or pyramidal growth and a few other types of imperfections in polished silicon wafers. In contrast, many device fabrication-process induced defects are not randomly distributed but highly localized. Typical examples may be damage due to orientation flat formation, dislocation loops adjacent to pn junctions with steep impurity gradient strains at the corners of diffusion windows, collector shift, probe chuck or tweezer marks. A wafer may have a slipped region only in a localized area. Therefore, it is preferred to judge the "in-process" wafers by means of an estimate of percentage of area affected. If the faulty area is over 50 percent, it may be wise to reject the wafer. If it is a wafer of simple

circuit design or a discrete device, a value of 30 percent may be acceptable. With high density IC's or LSI, the requirements are much more stringent.

Conclusions

Today's silicon device technology is still limited by structural imperfections introduced during device fabrication. Defects introduced by processing are cumulative, and imperfections may generate structural inhomogeneities on further processing (27). However, a one-to-one correspondence between the presence of these defects and the electrical characteristics of the devices has not been feasible. Even in seemingly defective silicon wafers, there are areas producing acceptable devices which display good life-test reliability. This is true at least for low power digital integrated circuits operating within specifications.

The interplay between the physical properties and structural perfection of the silicon and the electrical characteristics of the devices is still not fully understood.

Acknowledgment

The authors wish to acknowledge the assistance of J. Weltchek of Sylvania Semiconductor Division and of E. D. Jungbluth of the Bayside Research Center of the GT & E Laboratories. The production facilities of Sylvania Semiconductor Division were used to fabricate and test the IC units. A portion of this work was performed under the Air Force MATE Contract AF 33-615-67-C-1508 for Wright Patterson Air Force Base, Ohio.

REFERENCES

1. H. F. John, Proc. IEEE, Vol. 55, p. 1249 (1967)
2. H. J. Queisser, "Semiconductor Silicon", p. 585, The Electrochemical Soc., Inc., New York, N. Y. (1969)
3. S. Austerman and J. B. Newkirk, "Advances in X-ray Analysis", Vol. 10, p. 1, Plenum Press, New York (1967)
4. A. R. Lang, "An Investigation of Crystal Imperfections by X-ray Diffraction". Final Report, Contract AF 61(052)-449, April 20, 1966
5. F. X. Pink and P. Wang, Electrochemical Soc. Meeting, New York, N. Y., May 1969
6. A. N. Chester and F. B. Koch, "Advances in X-ray Analysis", Vol. 12, p. 165, Plenum Press, New York (1969)
7. W. C. Dash, Jour. Appl. Phys., Vol. 27, p. 1193 (1956)
8. J. R. Dale, "Semiconductor Silicon", The Electrochemical Soc., Inc., New York, N. Y., (1969) p. 622
9. J. F. Black, B. Sherman and V. Fowler, ibid. p. 693
10. E. D. Jungbluth and P. Wang, J. Electrochem. Soc., Vol. 116, p. 249C (1969)
11. B. Sherman and J. F. Black, Applied Optics, Vol. 9, p. 802 (1970)
12. D. C. Gupta, J. Y. Chan and P. Wang, J. Electrochem. Soc. 116, 227C (1969)
13. G. R. Booker, Discussions Faraday Soc. 38, 298 (1964)
14. S. Mendelson, Mat. Sci. Eng., 1, 42 (1966)
15. G. R. Booker and B. A. Joyce, Phil. Mag. 14, 301 (1966)
16. A. F. Witt and H. C. Gatos, "Semiconductor Silicon", p. 146, The Electrochemical Soc., Inc., New York, N. Y. (1969)
17. P. Wang, F. Pink and J. Sciola, J. Electrochem. Soc. 114, 879 (1967)
18. K. Morizane and P. S. Gleim, J. Appl. Phys. 40, 4104 (1969)
19. E. D. Jungbluth and P. Wang, "Physics of Failure in Electronics Symposium Proceedings", Vol. 4, p. 379 (1966)
20. W. Berg, Naturw., Vol. 19, p. 391 (1931)
21. C. S. Barrett, Trans. A.I.M.E., Vol. 161, p. 15 (1945)
22. D. W. Boss and V. Y. Doo, Jour. Electrochem. Soc. 15, 59C (1966)
23. C. M. Drum and C. A. Clark, J. Electrochem. Soc. 115, 664 (1968)
24. T. Enomoto, K. Yukawa, Y. Iwata and T. Ohkubo, Jap. J. Appl. Phys. 8, 130 (1969)
25. R. T. De Hoff, STP 430, p. 63, American Soc. for Testing and Materials, Phila., Pa. (1966)
26. R. R. Langhoff and A. R. Johnson, ibid. p. 96
27. E. D. Jungbluth and P. Wang, J. Appl. Phys., Vol. 36, p. 1967 (1965)
28. "Nomenclature and Recommended Practices for the Identification of Structures and Contaminants Seen on Specular Silicon Surfaces", Working Document of ASTM Committee F-1(6) Sec. 6.6.

Table 1

Commonly Observed Defects (28) in Silicon Substrate Slices After Polishing

Contaminants:	solvent or etch residues, particles of dirt, surface films, mottled surface, hazy surface (may be due to high density of polishing pits) embedded abrasive particles.
Mechanical Defects:	cracks, chips, crinkles or wavy surfaces, dimples, protrusions, scratches, tweezer marks, chuck or probe marks, orange peel marks (may be due to impurity distribution in silicon or poor practice in etching).
Structural and Crystallographic Defects:	dislocations, slip, lineage, pits (due to preferential etching), grown-in strain or microstrain, growth striae (due to impurity segregation).

Table 2

Structural Defects in Silicon Epitaxial Wafers for Device Fabrication

Defects(21)	Possible Causes
Dislocations	Local stress.
Slip grid work	Large area stress relief due possibly to improper temperature cycling or non-uniform heating.
Scratches	Careless handling.
Pits	Surface depression due to foreign matter on surface during growth or etching during N^+ drive.
Voids	Holes in epitaxial layer due to silicon particles on substrate surface.
Dimples	Shallow surface depression due to uneven support and sag of wafer during growth.
Spikes	Dendritic growth, due to dirty surface.
Stacking faults	Related to contaminated surface or embedded particles.
Orange peel	Insufficient polishing of substrate.
Taper	Nonuniform growth over wafer due to process.
Edge-ledge	Decreased growth at edge due to improper flow control.
Haze	High concentration of stacking faults or polycrystalline growth.
Pattern washout or collector shift	Substrate misorientation from 111 direction is incorrect, (in IC wafers only).
Flexure of wafer	Plastic deformation or elastic stress.
Growth striation	For striae in heavily doped substrates (in N/N^+ and P/P^+ structures).

Table 3

Structural Defects in Diffused Silicon Wafers

Defects (21)	Possible Causes
Flexure of wafer	Plastic deformation or elastic stress.
Strain	Local stress.
Slip dislocation network	Stress-relief due to thermal conditions.
Dislocation loops	Stress-relief generally associated with very high concentration diffusion and high temperature.
Precipitates	Formation of a new phase, e.g., Au_2P_3 or above solid solubility condition.
Pits in silicon	Etching in the deposition step or moisture.
Surface tarnish	Contaminants or poor photoresist process.
Pitted passivation film	Water vapor introduced in diffusion.
Loss of device geometry	Poor photoresist-etch definition.
Cracks, chips, scratches	Poor handling, fracture of wafer.
Nonuniform diffusion front	Emitter push-out in npn devices, poor flow control in deposition or residue oxides.
Diffusion pipes	Poor diffusion masking.
Dislocations (random)	From dislocations in substrate.
Dislocation arrays	Lattice mismatch.

Table 4

Yield of Circuits as a Function of Back Surface Treatment of the Wafer

Back Surface Treatment	Total Circuits Tested	No. of Acceptable Circuits	Yield %
Chemical Polish	1675	530	31.6
Mechanical-Chemical Polish	1474	648	44.0
Lapped	4108	1026	25.0

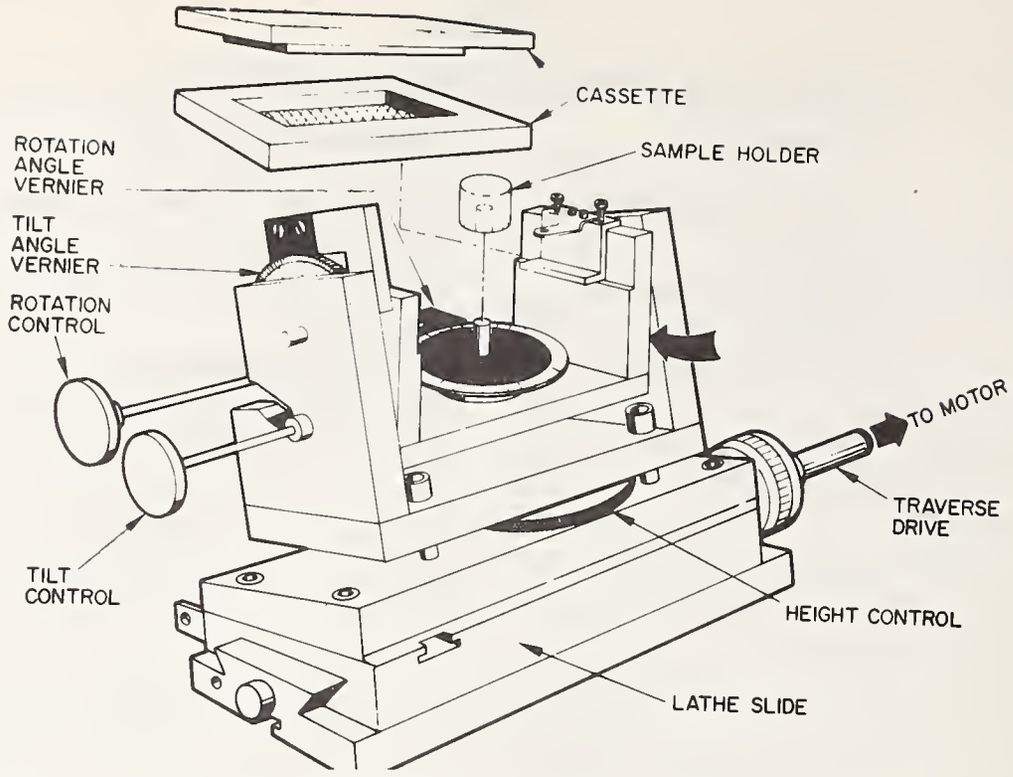


Fig. 1. Schematic diagram of Scanning Berg-Barrett camera.

Fig. 2. Stacking faults in an n/n+ epitaxial layer aligned with crystal lattice defects.

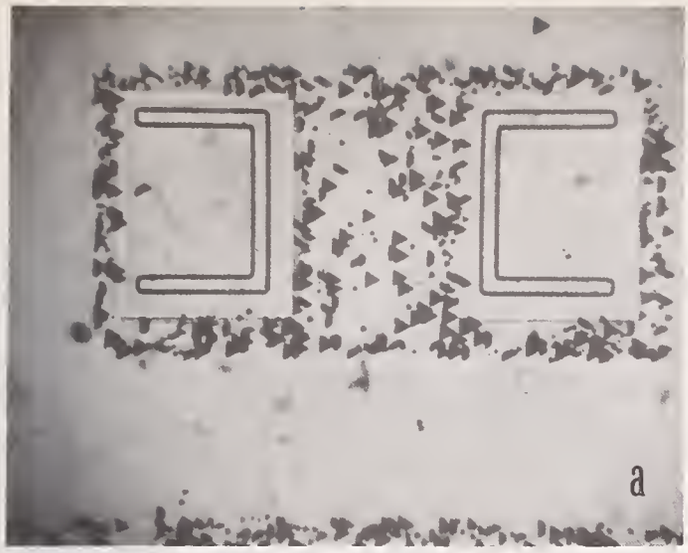
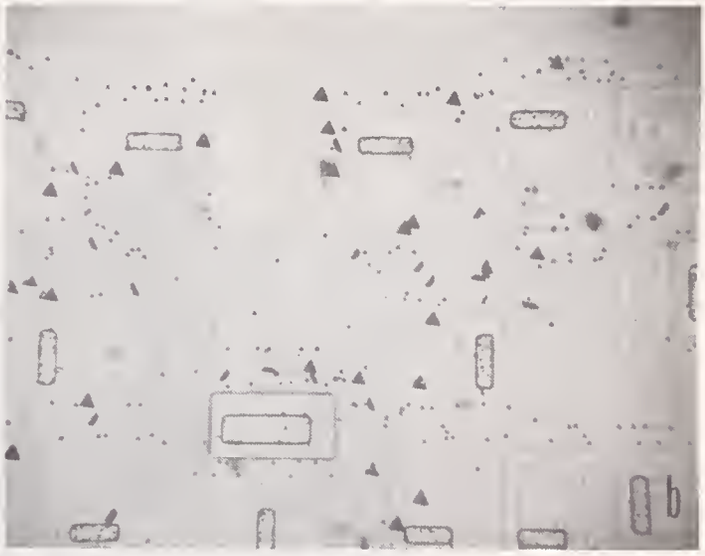


Fig. 3.(a) Stacking faults in isolation regions due to high concentrations of an impurity.
(b) Stacking faults due to poor cleaning of the wafer before epitaxial deposition.



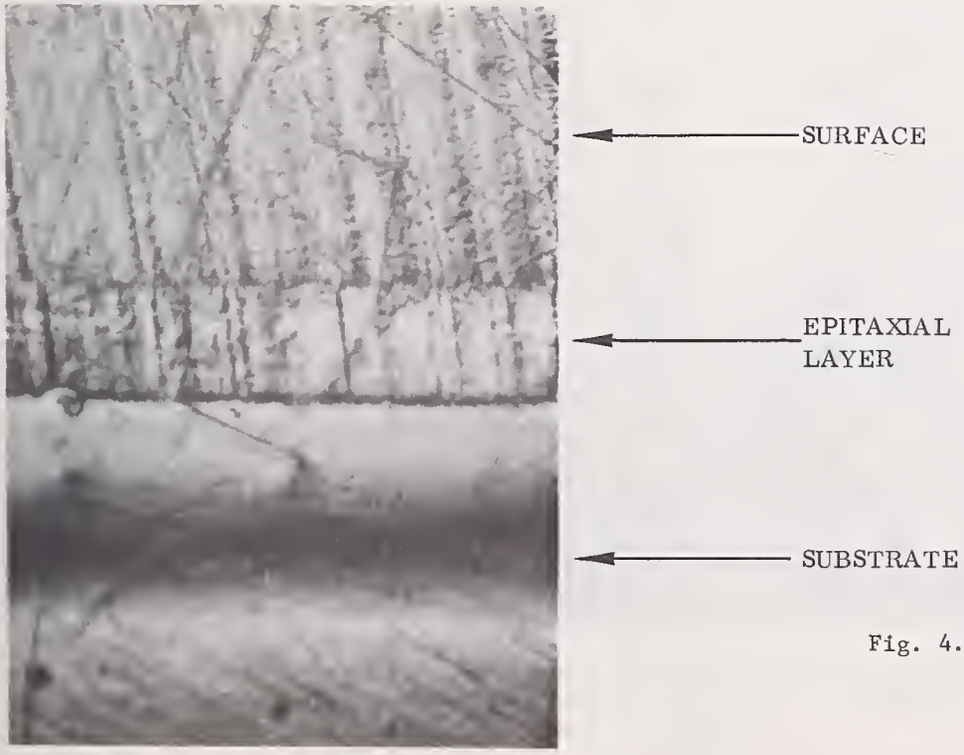


Fig. 4. Growth striation in n/n+ epitaxial layer. Layer thickness = 20μm.

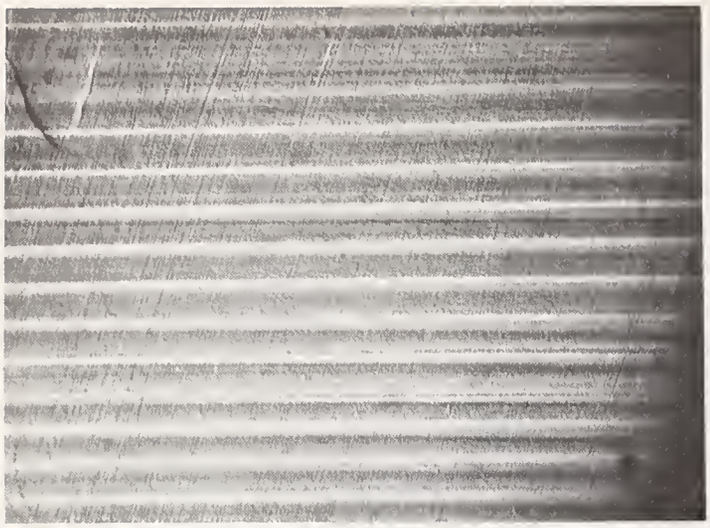


Fig. 5. Top view of the surface of the epitaxial layer shown in Fig. 4. The surface was etched in B-etch to reveal the arrays.

Fig. 6. Beveled surface of the p/p+ epitaxial wafer showing the propagation of dislocation arrays. Layer thickness = 20μm. (a) Bright-field (b) Dark-field.

(a) Bright-field

(b) Dark-field

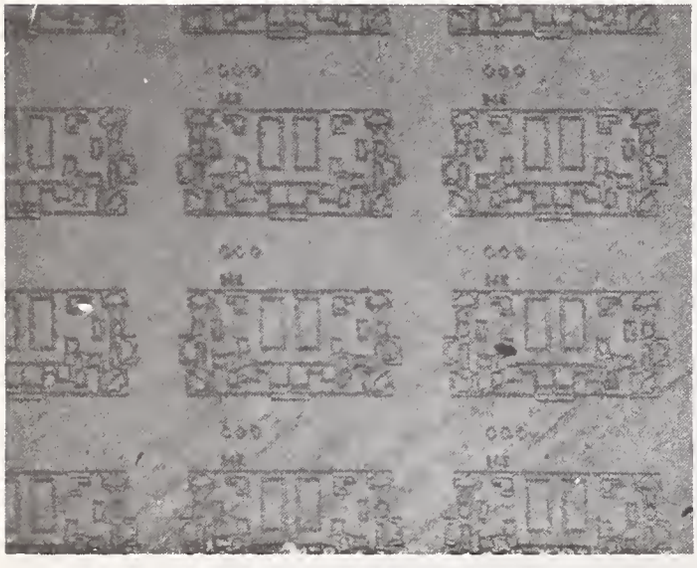


Fig. 7. X-ray reflection topograph showing the slip formation in an epitaxial IC wafer after isolation diffusion.

Fig. 8. Pattern washout in an epitaxial wafer. [111] direction oriented 1° from surface normal.

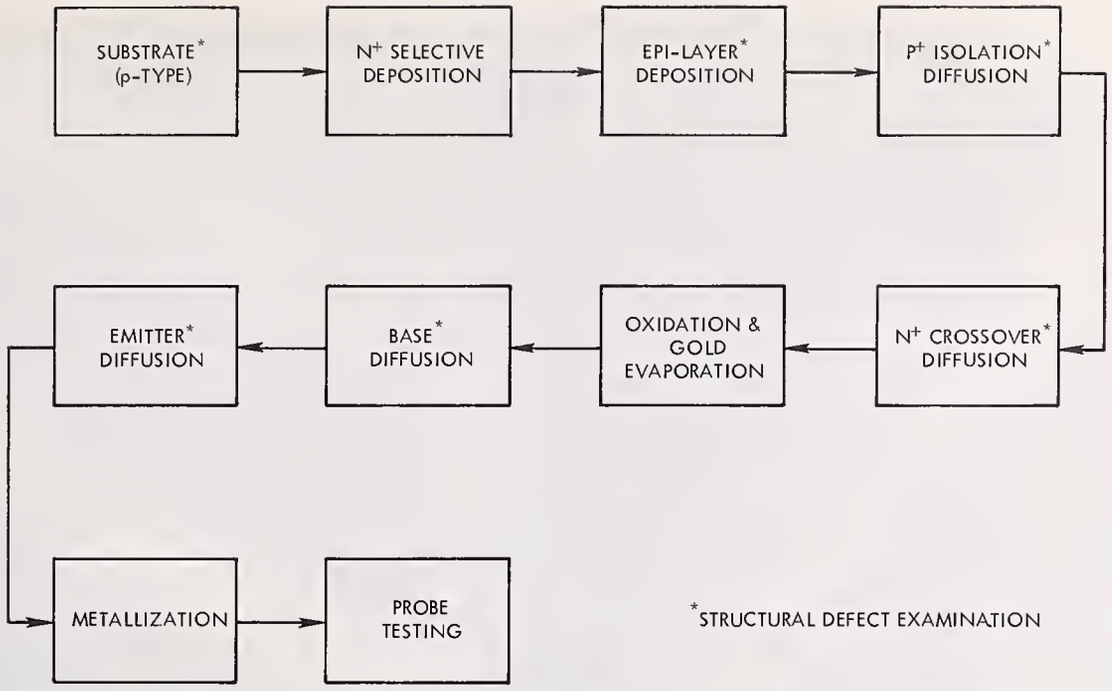


Fig. 9. Flow chart of wafer processing for fault-yield correlation.

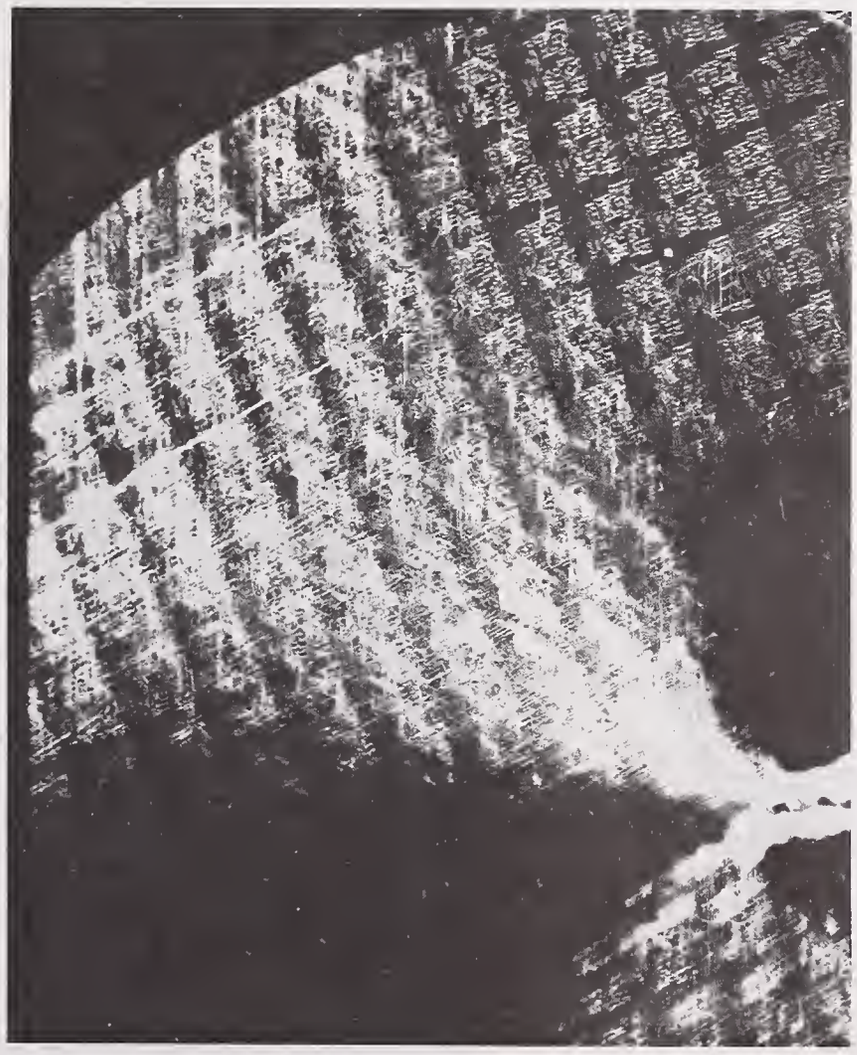


Fig. 10. A severely plastically deformed wafer showing high density of slip dislocations after cross-over diffusion. Back surface chemically polished.

Fig. 11. Topograph showing high density of slip dislocations after base diffusion. Back surface mechanically-chemically polished.

Fig. 12. A bent wafer with radium of curvature of about 500 cm. Back surface lapped.

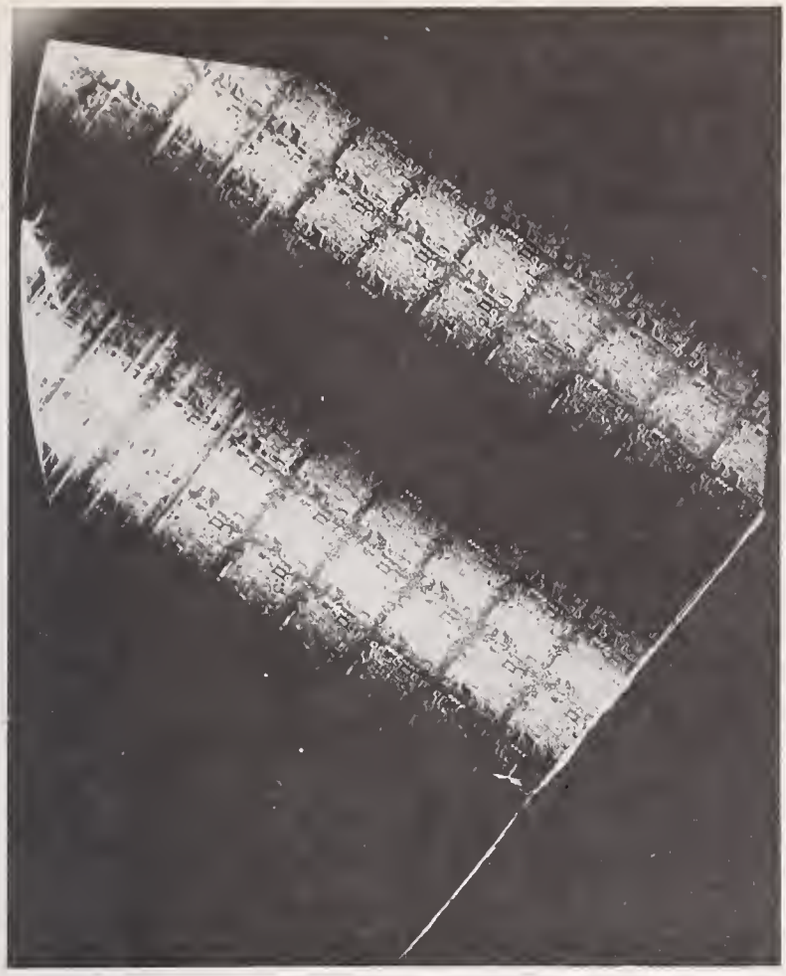


Fig. 13. A wafer showing small slipped region. Back surface chemically polished.

Fig. 14. Wafer showing no gross clip. Back surface mechanically-chemically polished. Device yield 297 out of 534.

An Instrument for Automatic Measurement of Epitaxial Layer Thickness

Allison Roddan[1] and Vitali Vizir[2]

This paper describes a commercially available instrument system designed to rapidly and accurately measure the epitaxial layer thickness of silicon wafers. Designed to operate in the production QA environment, the system utilizes the time proven infrared scan method while incorporating an electronic data processor to provide computation, storage and control functions.

The infrared spectrophotometer is especially optimized for epitaxial measurement and provides measurement rates unobtainable with other scanning instruments. Design parameters and performance requirements are discussed in detail along with several sample and control features.

To eliminate operator error from the measurement interpretation and data reduction, all peak picking, thickness calculating and phase shift correcting tasks are provided by a fully dedicated general purpose data processor. The processor also provides batch storage and hard copy data output printing. Since the processor hardware is general, the in depth discussion will be limited to the unique program features.

The system design targets a 2% accuracy specification and a 480 sample per hour measurement rate will be discussed in light of in situ performance. Even though the principle system described here is limited to the 3 to 50 thickness range, modifications to cover the .5 to 100 thickness range will be discussed. An appendix will be used to present parameter derivation and error analysis.

Key Words: Computer Automation, Epitaxial layer, Infrared measurement, Nondestructive testing, Spectrophotometer, Thickness measurement.

As the semiconductor industry continues to expand production of more advanced and more sophisticated devices and circuits, it requires an increasing number of accurate measurements to characterize devices from basic material preparation through to final products. Despite this expansion, most devices produced today use epitaxial layers at one stage or another, and the thickness of these layers is one of the devices' most important parameters. Methods of epitaxial layer thickness measurements must therefore be fast and accurate.

There are a number of ways to evaluate the epitaxial layer thickness:

1) Lap and stain or cleavage (1, 2)[3]
2) Stacking fault method (3)
3) Diode capacitance Method (4)
4) Electronic thickness gauge (5)
5) Infrared interference method (6, 7, 8, 9, 10)

Methods 1) through 3) are destructive and are therefore reluctantly used for production tests, but the remaining two are nondestructive and are recommended for production measurements. The electronic thickness gauge method is not sufficiently accurate due to backside deposition as well as wafer etching during epitaxial layer growth, the amount of which is not accurately known. This leaves the infrared method as the most likely candidate for production testing and is generally accepted as being fast and sufficiently accurate.

---

[1] Beckman Instruments, Inc., Fullerton, California
[2] Fairchild Semiconductor, Mountain View, California
[3] Figures in parentheses indicate the literature references at the end of this paper

The infrared method does, however, suffer from several limitations. These limitations are:

1) Refractive index of the epitaxial layer should be higher than that of the substrate.
2) The phase change at the epitaxial-substrate interface due to impurity concentrations of the substrate (12) introduces an error in the thickness measurement that must be corrected.
3) A resistivity gradient in the area of the epitaxial-substrate interface must also be corrected (13).

This paper describes an automatic system designed to optimize the epitaxial layer thickness measurement by minimizing the above limitations.

Up to now, general purpose spectrophotometers operating in the 11μ to 35μ wavelength range were employed to make IR thickness measurements. The instruments were slow (2-3 minutes to scan each wafer) and required a great deal of "hand" calculation of layer thicknesses. This method not only added greatly to the measurement time but created errors and ambiguity in the measurement. Due to all this the measurements were costly; about 8¢ per wafer. The system described here, operating at half its design capacity, has already reduced this cost to less than 4¢ per wafer.

This paper will describe the above system as to design considerations hardware and software components and in situ performance. Since expansion of this system leads to improved performance and extended ranges, several proposed systems will also be discussed. An appendix will show the derivation of several pertinent equations.

1. General Description of the Epitaxial Thickness Monitor System

The Epitaxial Thickness Monitor System (Beckman Instruments, Inc.) shown in Figure 1 is designed to rapidly and accurately measure the epitaxial layer thickness of n/n^+ or p/p^+ Silicon wafers in production environment. The infrared scanning technique was chosen because of its emphasis on speed, convenience and familiarity within the industry.

The spectrophotometer reflects monochromatic infrared energy off of a 1 x 5 mm sample area and detects the resultant spectral energy distribution. Connected to the spectrophotometer output is a data processor to provide all control, data reduction and storage functions, freeing the operator from all tasks except sample handling.

Batch reports are provided by the teletype and each consists of up to 30 wafers. The auxiliary recorder is provided to allow visual spectral data interpretation and act as a backup for the data processor in case of failure. One or two operators are able to monitor up to 480 wafers per hour, making 100% monitoring of a production batch economically attractive, with a full system of four spectrophotometers and one electronic data processor.

It should be noted that even though the system designed here is limited to the 3μ to 50μ thickness range, systems providing thickness limits of .5μ to 5μ and 12μ to 100μ may be built based on similar principles.

2.0 The Spectrophotometer

Until now, no commercial infrared spectrophotometer was optimized for epitaxial thickness measurement in the production situation. There are a number of trade-offs available and when made, the result is the spectrophotometer described here. Keep in mind that the purpose is to measure thin film silicon rapidly and with the utmost simplicity.

2.1 The Scan Range

Very little instrument design may be done until the wavelength (or frequency) scan range is established. Since the ray ABC shown in Figure 4 of the appendix must transverse two thicknesses of Si, the Si must be transparent, and thereby limiting the scan range to the infrared region. Furthermore, the lower wavelength limit is set at about 11μ(900 cm^{-1}) due to the general lack of detectable interference fringes caused by a diminishing difference between the refractive indices of the epitaxial layer and the substrate. On the other hand, the upper wavelength limit of the spectrophotometer is determined by the need for a continuous and broad spectral scan range. A prism monochromator can provide this continuity and range but is limited to transmission wavelengths below 35μ (300 cm^{-1}). It turns out that this 11μ to 35μ range is just enough to yield one fringe for a

3μ thick sample, a sufficient requirement for operation.

2.2 Scan Function and Direction

As shown in the appendix, the sample thickness is inversely proportioned to the fringe spacing and is in turn linear with frequency. By synchronizing the spectrophotometer and data processor with time, the abscissa information may be assumed, thereby simplifying the thickness measurement. To accomplish this, a synchronous motor scans the prism linearly with frequency in a direction contrary to general practice. The direction of scan is from low wavenumber to high wavenumber, allowing the data processor to monitor the output signal/noise ratio and terminate the measurement when sample fringes become too weak. The acceptance of the weak fringes would only add to the uncertainty of the measurement.

2.3 The Scan Rate

One of the objectives of this instrument is to increase the rate at which a wafer can be measured. This increase, dependent upon the response of the spectrophotometer, comes from the total energy gain of the instrument. This gain may be increased either by adjusting the electrical gain, resulting in more noise or by widening the monochromatic slit width, resulting in a second order increase in energy. It should be noted that the concomitant reduction in instrument resolution is of no consequence because epitaxial measurement requires poor resolution from a spectroscopy standpoint. This gain trade off results in an instrument response of 1/2 second for general conditions, and in turn allows a scan rate of two wafers per minute per spectrophotometer.

Since the response is affected by atmospheric water absorption, the optical portion of the spectrophotometer is sealed and must be purged with dry air or nitrogen. However, for thicker samples to 100 thick, resolution becomes a more important consideration. It is therefore necessary to narrow the slit program by about 3X, resulting in a 9X reduction in system energy and requiring an increase in electrical gain and a decrease in scan speed for compensation. Except for the scan motor, period filter and the slit program, the thick sample spectrophotometer is the same as the standard instrument.

2.4 The Sample Holding Fixture

At times, wafer profiles and buried layer thickness measurements are desired. For this reason, the infrared beam at the sample plane is focused to a 1 x 5 mm size. Since the sample platen (in the focus plane) is horizontal, the sample is simply placed face down over the target. However, when measurements are made on production wafers instead of test wafers, as in this case, a special vacuum system is used to hold the samples over the target but not in contact with the platen. The only contact with the sample is from the substrate side, preventing contamination from electrostatically charged dust.

2.5 Spectrophotometer Controls and Output

The measurement scan is initiated by depressing a single momentary Start Scan pushbutton and is automatically terminated at the scan end. Instrument parameters, such as gain and balance, normally do not require adjustment and are isolated from the operator while the other operator controls provided, such as Enter New Thickness and Report Request, are used only to command the data processor. The Enter New Thickness control changes a parameter in the computation algorithm while the Report Request control initiates a batch report. All operator controls are color coded, illuminated, and interlocked to minimize operator error.

2.6 Spectrophotometer Output

The spectrophotometer portion of the Epitaxial Thickness Monitor System produces analog output data proportional to the sample reflectance as a function of time (or wavenumber). This data output may be collected by the data processor or displayed on a linear strip chart recorder. A special Beckman Ten-Inch Recorder will display the entire spectrogram on 12-1/2 inches of chart paper, and when used with a transparent overlay, allows the thickness to be read directly from the recording. Figure 2 shows a sample recording with the overlay in position. A minimal amount of operator skill is required to measure uncorrected sample thickness using the overlay at rated system accuracy. Each recording is initiated synchronously by the spectrophotometer and is separated by a 2-1/2 inch dead zone.

3.0 The Data Processor (Hardware)

The Epitaxial Thickness Monitor uses a general-purpose data processor to reduce the data from the monitoring spectrophotometer and to generate reports that present this data in final form. In selecting the data processor included in the Epitaxial Monitor System, it was necessary to evaluate the following parameters:

1) Memory Capacity
2) Speed
3) Input/Output Capabilities
4) Cost

This section contains a brief discussion of the data processor and its associated peripheral equipment.

3.1 Input/Output Requirements

The data processor must have the capability of accepting four data signals per spectrophotometer:

1) Spectrophotometer Data (analog)
2) Expected Thickness Data (analog)
3) Enter New Thickness (command)
4) Generate Report (command)

By making use of the fact that the two analog signals occur at different times, and placing both on a single line, the number of analog signals to the analog to digital converter (ADC) is reduced to four. The actual ADC chosen was 11-bits (giving sufficient sensitivity) plus sign, where the sign bit is used to differentiate between data and expected thickness data. A positive voltage on an ADC channel is interpreted as meaning that the spectrophotometer is scanning and presenting valid data. A negative voltage signifies that the spectrophotometer is idle or between scans and the voltage present is proportional to the expected sample thickness set on that instrument.

The sampling rate of the ADC is very critical as it determines a major portion of the system error and is therefore derived from the same crystal-controlled oscillator in the data processor that generates the system interrupt signals. Because of this precision time base and the synchronous scanning motor in the spectrophotometer, the need for a precision wavenumber encoder is eliminated along with a substantial cost. The standard system has a sample rate of sixty-four times a second causing each input (four channels) to be interrupted sixteen times a second.

All output reports are generated on an ASR-33 Teletype which has no input responsibilities other than data processor program entry. The keys of the teletype are inoperative during program control to minimize error, allowing flexible report generation while freeing the system operator from having to monitor the teletype station.

3.2 Power Failure Monitoring

To eliminate the possible loss of batch sample data measured with a loss of power, the processor is equipped with a power fail monitor. In case of a power failure, or if the system is turned off for any reason, the processor is protected from losing any memory information. When power is restored, the program can be restarted with no information loss except the "in process" measurements.

3.3 Beckman 816 Data Processor

The Epitaxial Thickness Monitor System uses the Model 816 Data Processor manufactured by Electronic Instruments Division of Beckman Instruments, Inc. The Model 816 is a low cost, medium performance processor containing 2,048 words of 16-bit memory. The processor has a serial internal organization which accounts for its relatively slow processing time of approximately 40μ seconds. The Processor has a hardware bootstrap loader, making troublesome hand loading unnecessary. All controls necessary to start and load the data processor are lockable to prevent inadvertent operator errors. Those who wish a complete description of the processor are referred to the "Model 816 Reference Manual".

4.0 Data Processor, Software

The real power of the Epitaxial Monitor System lies in the software (or program), implemented by the data processor, that supplies system control, data processing and report generating functions. This section will address itself to describing the basic structure of the software using the Standard Film Thickness System as an example.

The software package consists of seven major routines. These are:

1) Executive
2) Boxcar
3) Peak Sense
4) Thickness
5) Report
6) Wait
7) Dbug

4.1 Executive Routine

The Executive routine is the master control for the entire system. It monitors all incoming data, requests for service and queues all jobs by use of a priority system. If at any time during the processor operation it becomes time to enter new data (highest priority job), the present job is halted, sufficient information to re-enter the job is saved, and control is transferred to the Enter Data portion of the Executive. If the data is spectrogram data (positive), it is stored in a buffer for use by the Peak Sense routine. The Executive will re-enter the interrupted job after it transfers the new data and processes any higher priority jobs that the entry of new data created. If the data is the expected thickness data (negative) it is stored for use by the Boxcar routine.

4.2 Peak Sense & Boxcar Routine

The Boxcar routine uses the above data to optimize the digital filter algorithm, in the Peak Sense routine, to locate the fringe extrema. The filter algorithm computes the area of two segments the spectrogram determined by the parameter calculated in the Boxcar routine and shown in Figure 3. The width of filter area is optimum when it is approximately one-half the period of the fringe to be filtered and is, of course, a function of the operator's estimate of sample thickness. Inaccuracy of this estimate simply causes a larger variance.

As the spectrogram is measured at each new data point, the oldest element in B is discarded and replaced with the oldest element from A. The new data then replaces the vacancy in A. When the filter algorithm equation .

$$\sum_{i=0}^{N} (A_i - B_i)$$

changes sign, indicating equal areas under the peak, a peak (minimum or maximum) has been detected. To insure against detecting noise around the desired peak, sign changes more frequently than every N/4 data points are counted only once.

The new peak magnitude is constantly compared with the previous peak magnitude. If their absolute difference is equal to or greater than 0.02 (2%), the old magnitude is replaced with the new magnitude and the peak sensing process is continued. When the difference becomes less than the threshold value, the process is terminated. This feature ensures sufficient sign-to-noise ratio for the filter to perform accurately, without which the calculated thickness values would have a greater variance.

Amplitude and peak data is not the only information needed for a thickness calculation. Each time a new data point is measured, the data processor increments a counter. When a fringe peak is reached, the number of counts corresponding to the first peak is stored in memory as Peak I and the number of counts corresponding to the peak prior to process termination is stored in memory as Peak II. Since the data points occur every 1.5 cm^{-1} (24 cm^{-1} second/divided by 16 data points/second),

$$V_j = 1.5 \text{ PEAK I} + 290\ cm^{-1}$$

and

$$V_i = 1.5 \text{ PEAK II} + 290\ cm^{-1}$$

where V_j and V_k are the wavenumber values of the first and last peaks, respectively, and 290 cm^{-1} is the starting point of the scan. These frequency values, along with the number of peaks encountered, are stored for future use in the Thickness routine.

4.3 Thickness Routine

The Thickness routine uses the above data to solve equation 8 of the appendix. Since the angle of incidence and index of refraction are assumed constant, they are stored in the data processor for use for calculations and not generally changed. Constants likely to vary from sample to sample and spectrophotometer-to-spectrophotometer (S_k, S_j, K_j and K_k) are stored with respect to specific spectrophotometers and are changed when the substrate resistivity is changed. It should be noted that each spectrophotometer is separate and independent, requiring its Thickness routine computation to use the proper constants assigned to the spectrophotometer and permitting samples differing in substrate and expected thickness to be measured on different machines at the same time. The output of this routine, the thickness values, is stored in memory for use by the Report routine.

4.4 Report Routine

When a report command is given by a specific spectrophotometer (operator initiated), the entire batch of thickness values is reported. The report page is headed, identifying the spectrophotometer origin, and includes sample number and thickness units. Each report may then be removed from the teletypewriter following its completion and attached to the sample batch for further processing.

A batch is any group of up to thirty samples after the last report from a given spectrophotometer. Since reporting is based on a first come, first served basis when handling multiple spectrophotometers, a given instrument may be required to wait two minutes (4 x 30 seconds per report) before its report is complete. To allow work to proceed without delay, each batch storage buffer has an alternate bank. When the number of samples of the batch exceeds thirty, a report is automatically generated and the thirty-first sample becomes sample number one of the next batch.

4.5 816 Dbug Routine

To allow for modification of program parameters and program initialization, the "816 DBUG" routine is resident in core memory with the Epitaxial Software package. This routine has many features and is described in detail in Appendix B-3 of the Model 816 Reference Manual. Its main feature, as far as the Epitaxial System is concerned, is the ability to modify the program from the teletype. This package is also used to set the substrate resistivity value for each spectrophotometer. (See phase shift correction discussion in the appendix).

5. System Performance

The above system has logged about 6000 hours of operation since its installation on a N and P type Si Wafer production line about 1 year ago. Some difficulty was experienced initially due to the extremely corrosive environment presented by the production process. The spectrophotometers and computer have been sealed and purged with clean air to protect them from corrosion.

To evaluate the performance of the Epitaxial Thickness Monitor, several test wafers were prepared and periodically measured. The resultant data quite adequately describes the in situ performance. To prepare the test wafers, three nominal thicknesses were chosen to represent the range of measurements encountered. Their physical epitaxial layer thicknesses were measured by the sirtl etch or copper etch method and found to be 5.0μ , 15.0μ and 25.0μ.

The test Wafers were measured and recorded on each spectrophotometer 30 times per month. The correlation between the metallurgical and IR measurement average is shown in Table 1. The test also showed a .15μ difference between the two spectrophotometers and a .1μ difference between the computer caluclated thickness value and values obtained using the analog recorder and overlay.

Table 1. Correlation of Metallurgical and ETMS thickness values

Test Wafer	Epi Layer Thickness (μ) Physical	ETMS	Diff (%)
1	5.0	5.6	.6 (12)
2	15.0	15.9	.9 (6)
3	25.0	26.3	1.3 (5)

Table 2 shows the month to month measurement averages along with their 2 sigma limits. In general, thickness measurements above 7μ showed a precision and repeatability of 2%. Thinner sample measurements showed larger variances due to poorly defined fringes and precision degradation due to uncertainty in the knowledge of substrate resistivity and boundry conditions. In all fairness to the equipment, it should be remembered that an error of 5% at 5μ thickness represents .25μ ; a respectable variation for production measurements.

Table 2 Monthly Instrument Standardation Results

Wafer # Month	Avg. Thick (μ)	2Σ%	Avg. Thick (μ)	2Σ%	Avg. Thick (μ)	2Σ%
Aug. 69	5.44	4.0	15.33	1.7	24.50	1.6
Sept. 69	5.55	5.6	15.25	1.7	24.55	1.9
Oct. 69	5.61	6.4	15.28	1.7	24.55	1.6
Nov. 69	5.63	5.0	15.35	1.2	24.56	2.0
Dec. 69	5.57	6.0	15.34	1.7	24.51	2.9
Jan. 70	5.63	4.4	15.39	1.6	24.63	2.1
Feb. 70	5.53	6.1	15.11	1.5	24.33	2.2
Mar. 70	5.54	3.9	15.09	1.6	24.30	2.0
Avg.	5.56	5.1	15.26	1.6	24.49	2.0

As was mentioned earlier, the above system designed to operate with two spectrophotometers in the 3μ to 50μ thickness range, may be modified to provide other characteristics. The most logical expansions would be to increase the number of spectrophotometers or extend the lower thickness limit to about 1μ. The expansion to four spectrophotometers has been implemented for other systems by simply modifying the Executive routine to index four input channels instead of two. Of course, the additional storage for constants and calculated thickness values had to be available.

A more involved expansion, however, is the extension of the lower thickness limit. In the area of 1μ thickness measurements, the contribution of phase shift and reflectivity envelope tangent contribute as much to the answer as does the location of the fringe peaks. This is the subject of new generation monitors and will be reported at a later symposium.

Appendix

The derivation of thickness equations for infrared measurement of epitaxial layer thickness is general and is found in many publications. It is given here to establish the units used in the hardware/software package of the Epitaxial Thickness Monitor.

Beam I, shown in Figure 4, incident on the film at the angle of ϕ, is reconstructed after reflection at plane P. The intensity of the reconstructed beam will be due in part to the phase difference between the first and second surface reflections. The phase difference may be described as

$$\delta = \frac{4\pi t}{\lambda}\left(N^2 - \sin^2\theta\right)^{1/2} + \phi_1 - \phi_2 \qquad (1)$$

In the analysis of this relationship, several assumptions will be made. These are:

1. Epitaxial layer resistivity is large enough to ensure a constant index of refraction in the range scanned.
2. The epitaxial layer and substrate boundary is sharp and definite
3. The incident angle is small enough to neglect polarization effects
4. The material is absorption free in the range scanned

When the phase difference, δ, is an integer multiple of 2π, constructive recombination (as opposed to destructive recombination or cancellation) takes place and the reflected beam has maximum intensity. Letting m be the order or multiple of recombination, Equation 1 becomes

$$2\pi M = \frac{4\pi t}{\lambda}\left(N^2 - \mathrm{SIN}^2\theta\right)^{1/2} + \phi_1 - \phi_2$$

or

$$M = \frac{2t}{\lambda}\left(N^2 - \mathrm{SIN}^2\theta\right)^{1/2} + \frac{\phi_1}{2\pi} - \frac{\phi_2}{2\pi} \qquad (2)$$

Assumptions 1 and 4 allow $\phi_1 = \pi$, therefore Equation 2 becomes

$$M - \frac{1}{2} = \frac{2t}{\lambda}\left(N^2 - \mathrm{SIN}^2\theta\right)^{1/2} - \frac{\phi_2}{2\pi}$$

By converting to units of frequency, V (cm^{-1}), the final form is then

$$M - \frac{1}{2} = \frac{2tV}{10^4}\left(N^2 - \mathrm{SIN}^2\theta\right)^{1/2} - \frac{\phi_2}{2\pi} \qquad (3)$$

where t is in units of microns. Figure 5 shows the theoretical relationship between frequency, intensity and order number for the reflected energy. The figure neglects the influence of the second surface phase shift, $\phi_{2/2\pi}$. A study of the figure and Equation 3 will permit the following conclusions:

1. For V = 0, m - 1/2 = 0 or m = 1/2
2. For t = constant, m is linear with V
3. As t gets larger, the locations of m approach V = 0

It can be seen that a knowledge of the value and location of m will yield the epitaxial layer thickness

Rewriting Equation 3 in terms of a variable subscript i, we have

$$M_i - \frac{1}{2} + \frac{\phi_{2i}}{2\pi} = \frac{2tV_i}{10^4}\left(N^2 - \mathrm{SIN}^2\theta\right)^{1/2}$$

where i designates the ith order. Letting i = j for the first peak and i = k for the last peak in the determinations, the difference is

$$M_k - M_j + \frac{\phi_{2k}}{2\pi} - \frac{\phi_{2j}}{2\pi} = \frac{2t}{10^4}\left(N^2 - \mathrm{SIN}^2\theta\right)^{1/2}\left(V_k - V_j\right) \qquad (4)$$

where m_j, m_k, V_j, V_k. A peak is either a maximum or a minimum. Rearrangement of Equation 4 yields the final form of the thickness computation equation:

$$t = 10^4 \cdot \frac{\left(M_k - M_j\right) + \left(\frac{\phi_{2k}}{2\pi} - \frac{\phi_{2j}}{2\pi}\right)}{2\left(N^2 - \mathrm{SIN}^2\theta\right)^{1/2}\left(V_k - V_j\right)} \qquad (5)$$

where

m = order of peak,
t = thickness of epitaxial layer in microns
n = index of refraction
ϕ = angle of incidence
V = frequency of cm^{-1}, and
$\phi/2\pi$ = unitless phase shift
$m_k - m_j = N$, the number of whole orders between the first and last reading.

The phase shift term ($\phi_{2k}/2\pi - \phi_{2j}/2\pi$) may be evaluated using data provided by Schumann (12) or ASTM procedures. Figure 6 shows a typical plot of $\phi_{2i/2\pi}$ vs. for a sample of substrate resistivity = 0.003Ωcm.

For simplicity, the curve may be approximated by a pair of least square linear segments. The equations of these segments may be used to solve for $\phi_{2/2\pi}$ when the frequency is known.

The general form of each linear segment is

$$\frac{\phi_{2i}}{2\pi} = Si\lambda i + Ki \qquad (6)$$

where S_i is the slope and k_i is the ordinate intercept. Since

$$\lambda(\mu) = 10^4 / V\,(cm^{-1})$$

Equation 6 may be written as

$$\frac{\phi_{2i}}{2\pi} = \frac{10^4 Si}{Vi} + Ki \qquad (7)$$

The value of S_i and k_i will depend upon whether the location of the desired peak is greater or less than V_b.

Again, by letting i = j for the first peak and i = k for the second peak (V_k and V_j) and combining Equation 5 and Equation 7, we have

$$t = \frac{10^4}{2(N^2 - \sin^2\theta)^{1/2}(V_k - V_j)}\left(M_k - M_j + \frac{10^4 S_k}{V_k} - \frac{10^4 S_j}{V_j} + K_k - K_j\right) \qquad (8)$$

which is the general form for thickness calculations.

It can be seen from Figure 6 that it is possible for $S_k = S_j$, $k_k = k_j$, if both V_k and V_j are either less or greater than V_b. This situation would only further simplify Equation 8.

References

(1) C. S. Fuller & T. A. Ditzenberger, F. App. Phys. 27, 544 (1956)
(2) ASTM Standards, Park 8, F110-697
(3) S. Mendelson, J. Appl. Phys. 35, 1570 (1964)
(4) C. O. Thomas, D. Kahng and R. C. Manz J ECS, 109, 1055 (1962)
(5) C. C. Allen & E. G. Bylander, AIME Metallurg. Sec. Confer. 15, 113 (1962)
(6) W. G. Spitzer and M. Tanenbaum, J. Appl. Phys., 32, 744 (1961)
(7) M. P. Albert and J. F. Combs, J. ECS, 109, 709, (1962)
(8) ASTM Standards, Part 8, F95-68T
(9) H. J. Sloane, Beckman Application Data Sheet IR-8077
(10) T. E. Reichard, Electronics, March 18, 1968, 101
(11) W. Spitzer & H. J. Fan, Phys. Review, 108, 268 (1957)
(12) P. A. Schumann, Jr. et al, J. ECS, 113, 368 (1966)
(13) K. Sato et al, Solid St. Electr. 9,771 (1966)

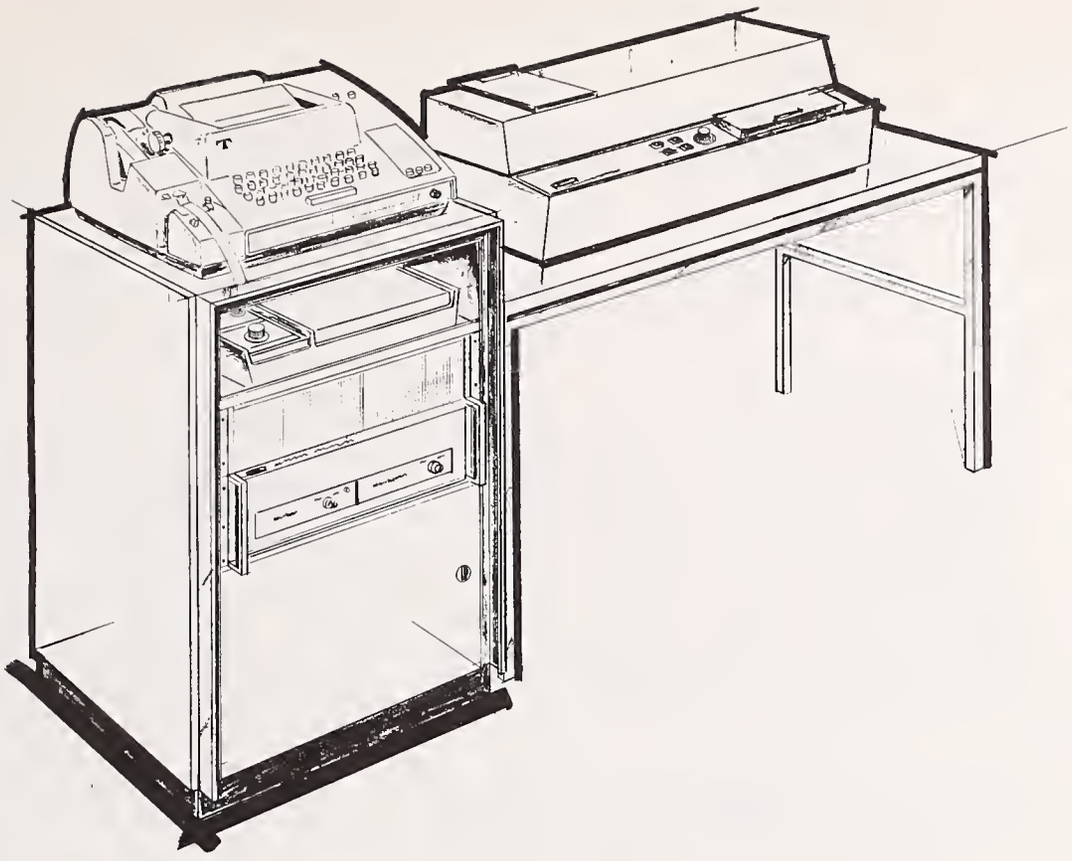


Fig. 1. Epitacial layer thickness monitor system.

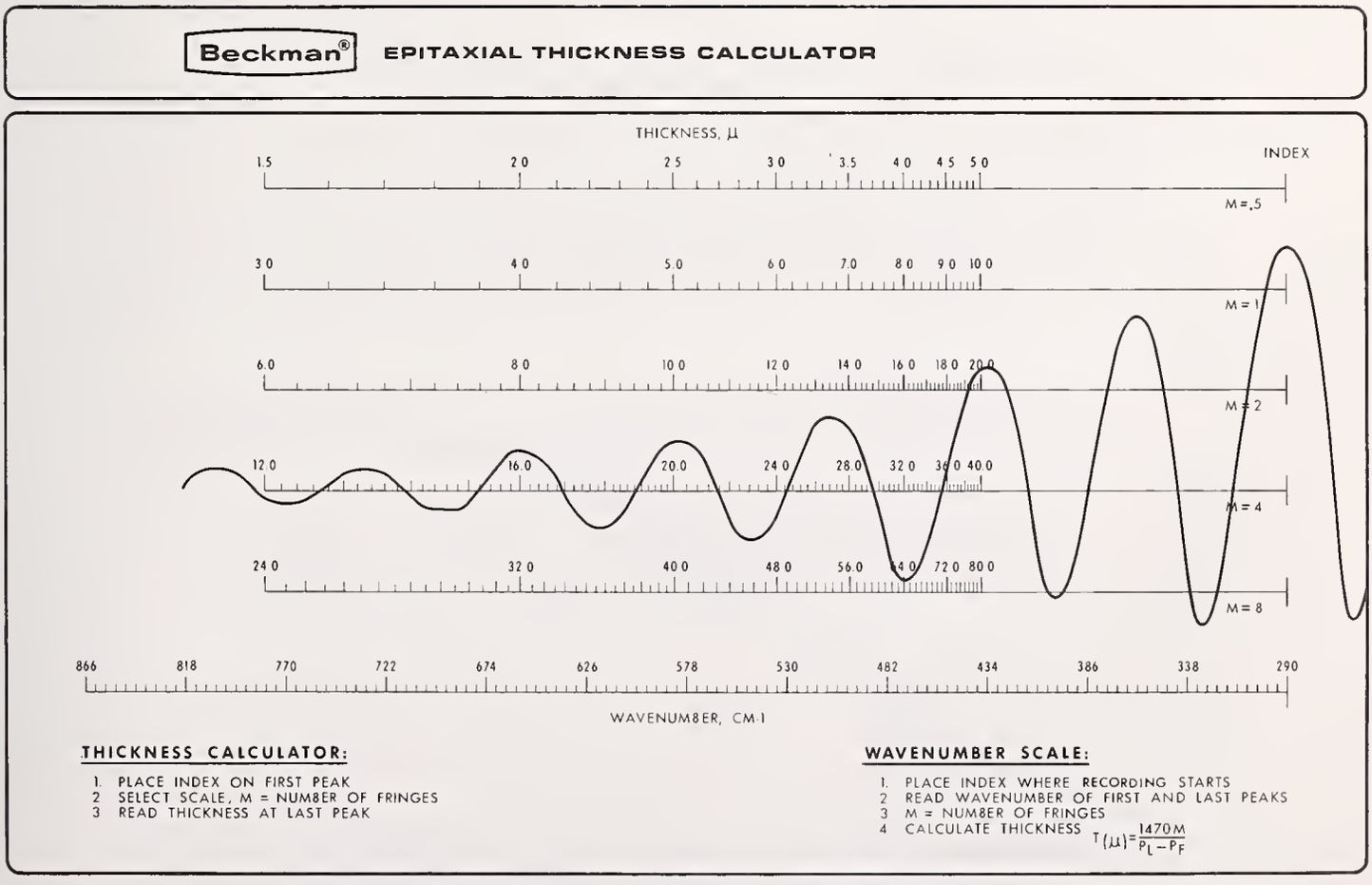


Fig. 2. Sample recording with overlay thickness calculator in place.

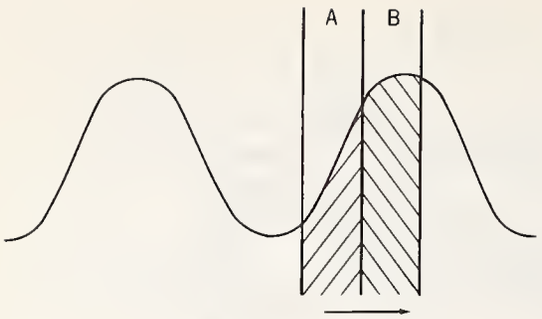


Fig. 3. Boxcar digital filter.

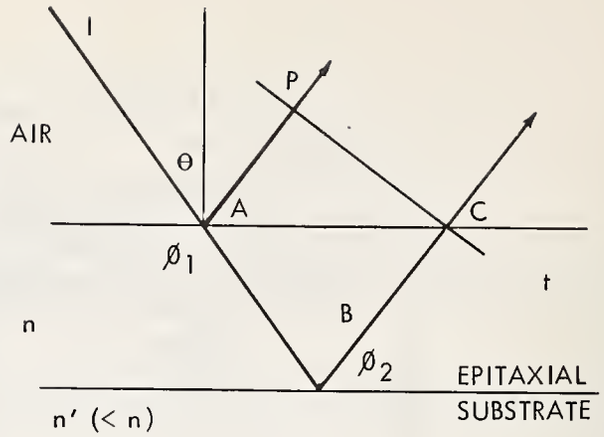


Fig. 4. Light interaction with a thin film.

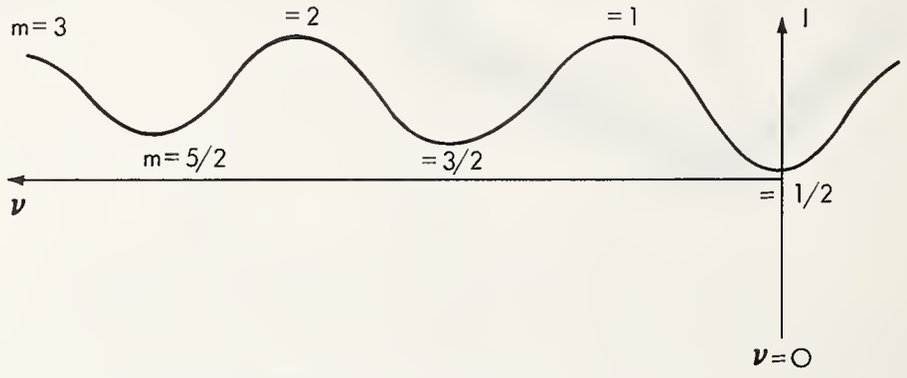


Fig. 5. Fringes, wavenumbers and orders.

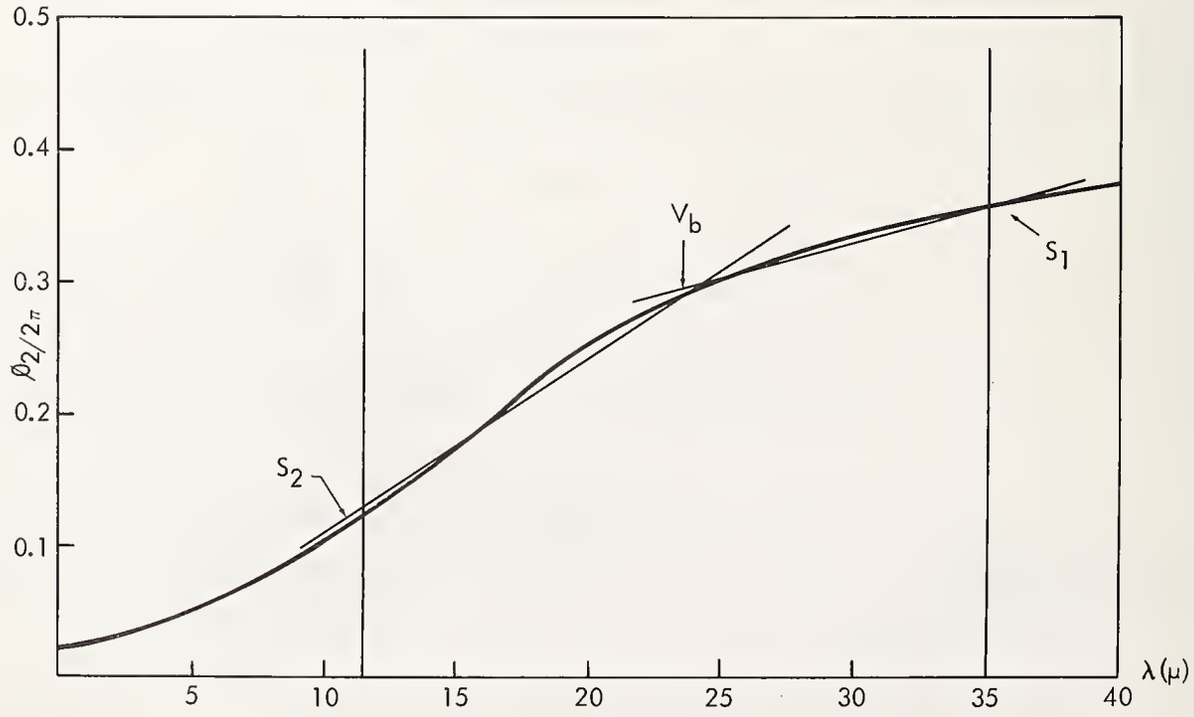


Fig. 6. Phase shift correction function for Si with p = .003 Ω cm.

Defects Induced in Silicon Through Device-Processing

M. L. Joshi

IBM Burlington, Components Division
Essex Junction, Vermont 05452

J. K. Howard

IBM E. Fishkill, Components Division
Hopewell Junction, New York 12533

Growth mechanisms of defects induced in silicon through various process steps involved in fabrication of integrated circuits need to be critically understood since among the active parts of integrated circuits, silicon is of primary importance. Structural defects (e.g., dislocations and stacking faults), may cause diffusion spikes which, in turn, result in microplasmas and lower breakdown voltage in junctions. They usually come into existence during the various process steps starting from the very first step of crystal growth. The most serious defects usually cause yield losses during processing rather than field failures. A brief review of the defects and their influence on the electrical quality of devices is presented.

The role of residual strains in creating defects is brought into focus in this article. Operations on silicon crystal, such as slicing and lapping, introduce damage in crystal structures adjoining the surfaces of the silicon slices. The subsequent polishing operation is expected to remove the damage; the consequences of the failure of this operation to remove the damage is creation of residual strains. The defects created by these strains are examined.

Strains of both the plastic and the elastic types can be introduced in silicon through diffusion and precipitation of impurities. Diffusion is used for fabrication of structures; such as base, emitter, isolation, subcollector, resistor. Diffusion of both the odd-size atom impurities (like phosphorus and boron) and the same-size atom impurity, arsenic, introduces plastic deformation in silicon. A new synthesis with regards to origin and morphology of diffusion-induced deformation is described. Effects of the plastic deformation on impurity distributions are also analyzed. Cooperative diffusion effects are discussed in terms of the size of an impurity.

Finally, the effects of the process steps involving heat-treatment of silicon (e.g., oxidation and metallization) are examined from the point of view of defect generation in substrate silicon and in various diffused structures.

Key Words: Diffusion, electron microscopy, epitaxy, imperfections, junction quality, oxidation, silicon, wafer shaping, x-ray topography.

1. Introduction

Quality of a silicon p-n junction is intimately related to the structural perfection of the junction region. [1-21][1] Crystal-growth conditions and device-processing steps both introduce structural imperfections in silicon. [22-73] This article is essentially a brief survey and a synthesis of the vast data on defects in silicon introduced due to processing steps commonly involved in any device fabrication. It has to be emphasized that this article does not intend to present an exhaustive review of the defects induced by all the common processes, since excellent critical reviews on defects induced by some of the important ones, namely, crystal-growth [74-76] and epitaxial-growth [35] are available. It is also not our intention to give a comprehensive review of the different tools and methods of observations of the defects. We will mostly present our experimental work on process-induced defects observed through x-ray reflection topography (XRT) and transmission electron microscopy (TEM).

Besides the objective of a brief survey, the next but more important objective of this paper is to present our experimental work that points out to the major role in defect generation played by the strains in silicon. Saw damage at times may leave a remnant strain-damage in finished wafers. These finished wafers, if used for processing, may introduce uncontrolled slip. The growth faults of various kinds in epitaxial films are well understood. We show a profound but not well-understood effect of mechanical damage on the epitaxial-growth faults. Silicon oxide and metal layers are used on silicon for the purposes of masking and electrical conduction. They invariably introduce elastic strains in silicon. We will show the uses of x-ray transmission topography (XTT) in characterizing the strain effects due to saw damage in silicon substrates, due to mechanical damage on epitaxial layers, and due to oxide and metal layers on silicon substrate. We will further point out that it is considerably advantageous to adopt the XRT method instead of the often used XTT method for device defect analysis. [20,67] We will prove this point by illustrating the application of XRT to device-processing.

Introduction of strain in silicon also occurs during diffusion of odd-sized impurities such as P or B, resulting in generation of misfit-dislocation nets. We feel that the cooperative diffusion effects such as push-out and retardation of base-collector junctions are strongly influenced by the nature of the size of the impurity atoms. Oxide precipitation in silicon is also associated with strain effects. In this article, we will present a coherent status of the most recent understanding of the effects of strain and the strain-induced defects on diffusion and oxidation in silicon. We will also be able to throw a new light on defect mechanisms responsible for diffusion in silicon.

Before we describe our experimental work, a brief review of the work in the field of relationship between structural imperfections and device-junction quality done by various researchers is presented. This brief review is essential for understanding the significance of the part played by the structural defects in affecting device yield in manufacturing.

This article, hereafter, is divided into the following sections:

A. Structural Imperfections and Junction Quality

B. X-Ray Topography – A Defect-Analysis Tool

C. Process Characterization of Wafer-Shaping Operations

D. Residual Damage and Epitaxial Film Perfection

E. Adhesion of Oxide and Metal Films

F. Diffusion

G. Oxidation

A. Structural Imperfections and Junction Quality

The breakdown phenomenon in silicon p-n junctions is usually characterized by microplasma effects. A microplasma pattern is a very useful tool for correlating

[1] Figures in bracket indicate the literature references at the end of this paper.

physical defects with electrical effects. These effects were first reported in 1953 by McKay and McAfee. [1] Since then a considerable amount of work has been done in the field resulting in deeper understanding of the phenomenon. [1-21] Visible light is emitted by the junctions both in avalanche breakdown and breakdown by internal field emission. Recombination of the junction-field-accelerated electrons in the space-charge region of a Si p-n junction emits light in the visible range. This light is seen if a shallow p-n junction is made in order to avoid light absorption. In the case of avalanche breakdown, the carriers are injected from the end regions of the junction. On the other hand, in the case of internal field emission, the carriers are created by the direct action of the field. In avalanche breakdown, the light comes from a large number of localized spots (microplasmas) distributed over the junction region. The microplasmas usually occur in the prebreakdown region. Light emission comes from a large number of closely spaced spots with internal field emission (soft reverse characteristics) giving the impression of a more-or-less uniform glow over the whole junction region. It was speculated quite early that both of these light emission effects involve structural defects like dislocations. In a review on p-n junction problems in 1960, Shockley, [8] after analysis of large data, concluded that the microplasma effects, including the "lock-on" or negative-resistance phenomena and noise in p-n junctions, do require structural imperfections, such as dislocations, which provide traps capable of storing and immobilizing high density charges.

The value and sharpness of the breakdown voltage is the most widely used criteria for evaluating the junction quality. The reverse avalanche breakdown voltage of a good diode is determined by the base dopant magnitude and the concentration gradient at the junction. Defects will lower this value. This is a well-accepted result. Defects affect the electrical behavior [22] of a diode in basically four ways: (1) decreased reverse voltage, (2) increased reverse current, (3) forward or reverse current tunneling, and (4) decreased minority carrier life time. The type of adverse electrical behavior is determined by the type, the size, and the distribution of the defects. Small-insulating particles cause microplasmas characterizing avalanche breakdown. Large-insulating particles cause local destructive breakdown. Small, but dispersed, metallic particles lead only to soft junctions (exhibiting high-reverse leakage current and lower breakdown voltage) characterized by current tunneling corresponding to Zener-effect.

The basic models by Shockley [8] for microplasmas and soft junction behavior are to this date still valid. For microplasma behavior, he proposed a symmetrical-system model (Fig. A.1a) and an asymmetrical-system model (Fig. A.1b). A system exists in the space-charge region and contains a structure-imperfection producing a local high field and possessing traps for immobilizing high density electric charges. A dielectric particle like SiO_2 in the space-charge region, because of its lower dielectric constant, produces a local high electric field. The SiO_2/Si interfaces act as traps and create stationary charges as the carriers generated in the local high-electric-field region sweep by. This action increases the strength of the electric field at the location of the carrier generation and produces microplasmas with "on-off" characteristics through "lock-on" mechanism. In the unsymmetrical system, we have a space-charge region crossed by a dislocation decorated with small SiO_2 particles of various sizes and shapes. It is assumed that electrons cause the small spherical precipitates to become positively charged. This system is also expected to cause microplasmas. The SiO_2/particle explanation for microplasmas accounts very well the prevalence of microplasma formation on the silicon surface where the junction meets since this is the silicon-SiO_2 interface. Metal precipitates in the space-charge region are expected to produce soft junction behavior according to Shockley since they are conducting. Goetzberger and Shockley [15] have shown that softness of diodes does correlate with metallic precipitates in the active regions of the diodes. Along with the excessive leakage currents at sub-avalanche voltages, metallic precipitates also have been thought to be responsible for anomalous forward characteristics by Queisser. [21]

The actual reverse or leakage current is usually controlled to a significant amount at the surface boundary of the junction. No matter how we form a junction, it must end at some surface. The surface is a discontinuity in the crystal structure; so are dislocations and surface-absorbed contaminants, such as water vapor or metal precipitates. The major problem is that surfaces introduce a short life-time by introducing traps. It is well understood that the common-emitter current gain β_o is inversely proportional to the surface current when a high rate of surface recombination prevails. Duffy and his associates [83] have done certain experimental work to prove this point. Through a plot of transistor gain vs V_{be} taken at I_e = 0.5 mA for transistors made with high and low phosphorus emitters, the transistors show lower gains for high-P emitters. High-phosphorus emitters usually contain high density of dislocations in the surface. Some of these dislocations originating in the emitter

should extend well into the base region. Our observations on Duffy's high-P emitters, through TEM, show such dislocations as can be seen from Fig. A.2. Duffy's low-P emitters did not show such dislocations.

One of the major-yield reducers in transistor manufacturing is the "pipe" defect. A pipe can be defined as any defect that causes the emitter junction to be locally shorted to base. Electrically, the pipe can be considered present when the collector-emitter breakdown with open-circuit base, BV_{ceo} is less than 4 volts at 10 μA current for most of the present day small-area transistors. This value of BV_{ceo} could be as low as 1 volt depending upon device specifications. The intensity of pipe activity and therefore the junction quality, depends on the proportion of the emitter-area being shorted to the collector junction. Figure A.3 qualitatively illustrates this point showing the relation between pipe activity and I_{ce} vs V_{ceo} curves. Such an electrical characterization, however, gives us no insight into the nature of pipes. Pipes are known to result from either masking-failure or metal-decorated dislocation crossing the emitter-base and the base-collector junctions. Emitter-junction raggedness also is known to give the piped behavior of I_{ce} vs V_{ce} curves. Extensive slip generated in the transistor regions at the wafer-edges during various processing steps [20,73] could be held responsible for pipe-effects. Slip-process involves local creation of surface steps by relative sliding of the {111} planes inclined to the surface of the wafers and can cause junction raggedness. In Figs. A.4a, b, and c, various device-configurations that can be affected by slip are shown. No models have yet been developed to assess the electrical effects on the device performance. One can imagine the consequences on the performance of a diode with a slipped region as shown in Fig. A.4d. The slip increases the area of the junction and introduces an impurity gradient in the direction $\bar{g}$. The latter consequence is not expected to affect the electrical characteristics of the diode in any serious manner. The increase in area, however, if excessive would affect the current-carrying capacity and the junction capacitance. An excessive slip in the transistor region could affect far more seriously the transistor parameters. Excessive slip in the transistor regions could cause the following effects:

(1) Local constrictions in the base region, resulting in local narrow base widths.

(2) Enhanced emitter diffusion along the slipped planes, giving rise to pipe-like structures.

(3) Ragged junctions, resulting in increased junction areas.

The first effect would tend to increase the base-resistance, current gain, and the frequency response; and decrease the punch through voltage and V_{be}. The second effect would change in the manner, pointed out before, I_{ce} vs V_{ceo} curves. The third effect would alter the junction capacitances. It must be emphasized, however, that these changes in the electrical parameter of the transistor would be serious only if the slip is excessive.

From the above discussion, it can be concluded that precipitates — metallic or dielectric — dislocations and slip are the most important crystal imperfections with possible adverse effects on device properties. However, a one-to-one correlation either between dislocations and microplasma or between SiO_2 precipitates and microplasmas has never been established. On the basis of the statistical correlations only, the demand for dislocation and precipitate-free grown crystals has grown strong in recent years. Fortunately, silicon crystals can be grown extremely pure and highly dislocation-free. The properties of silicon — such as low dislocation mobility, large thermal conductivity, and relatively small thermal expansion — aid the crystal growers to produce dislocation-free crystals with relative ease. In modern silicon device manufacturing we indeed use dislocation-free crystals with well-defined purity limits.

Some researchers feel that we buy dislocation-free crystals at an enormous cost with dubitable advantages in junction quality and yield. Device-processing steps introduce, in large densities, lattice defects; a comprehensive survey has been compiled by H. F. John. [22] Processes such as imperfect chemical polishing after lapping, [23,24] wafer handling, epitaxy, [25-38] diffusion, [39-53] and oxidation/heat-treatment [54-66] have been carefully studied by various researchers with regards to their effects on structural perfection of silicon. X-ray topography has been used in quite a revealing manner by Jungbluth and Wang, [67] and Schwuttke and his associates [20,68] in studying in-process generation of structural defects in Si. Electron, x-ray, and optical microscopy studies by the various researchers [23-66] have shown:

(1) Slight deviations from ideal epitaxial processing introduce a considerable amount of dislocations, stacking faults, twins, growth pips, anomalous pits and tripyramid hillocks. If one observes with TEM in the substrate-epitaxy interface, one finds the stacking faults in association with faulted dislocation loops. A typical example is given in Fig. A.5.

(2) High concentration diffusion of P and B unavoidably introduces strains which are relieved by creation of misfit dislocations nets, interstitial loop dislocations, and precipitates. In Figs. A.6a, b, and c, we have typical examples of these for a (100) orientation silicon. On (111) silicon similar results, but with hexagonal symmetry, are observed. Later we will discuss the phenomenon of diffusion-induced defects in detail.

(3) Oxidation introduces through surfaces stacking faults and oxide platelets into silicon, particularly in p-crystals, which often contain supersaturated oxygen. A typical TEM example of such faults is given in Fig. A.7 where the Frank partials F-penetrating silicon are shown decorated by the oxide particles.

(4) Certain heat-treatments of silicon create Si_xO_y precipitates, which punch out dislocation loops. A TEM illustration is given in Fig. A.8.

(5) Trace-mechanical damages in wafers are one of the major sources of dislocations and slip in silicon wafers.

(6) Remnant-lapping damage, which is discussed later in detail, creates dislocations when silicon gets heated. The remnant damage, in the form of scratches after insufficient chemical polishing, often results into rows of dislocation loops as illustrated in the TEM view in Fig. A.9. Such scratches may also be produced by wafer handling later during device-processing. Chipped wafer edges are well-known precursors for slip in silicon wafers.

This short description of the process-induced damage in silicon wafers leads us, naturally, to the question: Why should we obtain expensive dislocation-free grown silicon crystals for device-processing. We can only answer this question after our examination of the present literature related with the effect of the process-induced defects on the junction quality.

The x-ray topographic work [20,73] indicates quite a good correlation between low breakdowns and excessive reverse currents for devices in the slip regions of the edges of wafers. Queisser and Goetzberger [69] have shown that stair-rod dislocations are electrically conductive and are the sites of microplasmas. Our TEM observations, depicted in Fig. A.6, on the epitaxial stacking faults responsible for microplasmas show that the precipitates accumulate at the faults. In Fig. A.10, we observe two stacking faults. After heating the sample in the TEM, these faults were annealed out, leaving the precipitate rows visible. Microplasmas have also been observed along the periphery of a collector junction after reoxidation by Barson et al. [70] and the defects responsible for these are thought to be oxidation-induced defects of Joshi-description. [54] These observations reinforce the question raised regarding the usefulness of expensive dislocation-free grown silicon crystals. There is no doubt that we must have silicon crystals with uniform doping. Goetzberger's results, [17] showing striated-light emission patterns from diodes built in silicon crystals of striated boron-doping, indicate that we need crystals with uniform doping. We, however, have no demonstrated proof that we must have dislocation-free crystals to start off. Lawrence [71,72] has given convincing arguments and experimental evidence for choosing crystals of moderate ($\approx 10^7/cm/cm^3$) dislocation densities. His argument is as follows: The interactions between contaminants of immense variety and amounts introduced during device-processing and these structural defects play a profound role in junction quality and yield problem. Direct electrical effects of uncontaminated dislocations in devices are very weak. Device failures are primarily associated with impurity precipitation in depletion regions from supersaturated concentrations of dopants. These failures are aggravated by dislocation-free silicon substrate. The impurities will precipitate at the crystal surface and process-induced (lattice) disorders near the depletion regions of a device if the bulk lattice is free of stable bulk dislocations that could aid in dispersing the junction-degrading impurity. This conclusion seems to be further reinforced by two apparent discrepancies existing in the reported data which correlate device failures to dislocations: (1) device failures and x-ray topographic slip/dislocation structure is less meaningfully correlated away from the wafer edges [20], and (2) Goetzberger et al. [17] have shown that neither lattice imperfections nor oxygen in the crystal are

necessary for microplasmas at premature breakdown sites. There is also other experimental evidence in favor of the above argument. Low-dislocation-density Si is less advantageous for application in solar cells which need dislocations for absorbing radiation-induced point defects. [73]

The discussion so far has emphasized the intimate connection between the structural defects and the junction quality of devices. Now, we proceed to our experimental work as outlined in the introduction.

B. X-Ray Topography — A Defect-Analysis Tool

B.1. Introduction

The analysis of process-induced defects in silicon monocrystals requires a nondestructive method to "track" the wafer through the process and then define the critical step for defect generation. The method of x-ray diffraction topography is generally applicable to most process studies of this type, since it is nondestructive and sensitive to crystal disorder; such as dislocations, precipitation, and mechanical damage.

In the following discussions of process-induced defects, we will first introduce the reader to a simplified discussion of the x-ray methods which are utilized, and of more importance, why we use a specific technique to investigate a particular problem. For a more detailed description of the x-ray techniques and the interpretation of contrast effects, we will include references to appropriate articles. The format for the discussion will also be somewhat different from most review papers of this type. Rather than try to do justice to every investigator that has contributed to the understanding of process-induced defects, and risk confusing the reader with a voluminous amount of detail, we will include specific examples of process studies. It is anticipated that these individual studies will promote a coherent understanding of how to characterize a process, and what the information means in terms of the field of process analysis.

B.2. X-ray Topography Methods

X-ray topography is a nondestructive method to image crystal flaws in the entire crystal interior or at the surface layers. The defect is revealed on a high resolution photographic plate because the strain field of the defect diffracts more intensity than the perfect crystal matrix. Figure B.1 depicts a schematic of the transmission (Lang geometry) method. The crystal is aligned to diffract planes nearly normal to the wafer surface, and the diffracted beam is passed through a slit and registered onto a photographic plate. [77] The entire crystal is examined in this manner by translating the crystal and film unit parallel to the film. The resultant image on the photographic plate yields the superposition of the majority of defects in the crystal. However, the Lang method is somewhat difficult to employ in the evaluation of semiconductor-grade silicon during processing, as the wafer becomes distorted and curved due to diffusion, oxidation, and metallization stresses. These stresses cause the wafer to bow and when the crystal is translated, the reflection angle is not preserved across the wafer surface. In other words, the process-induced stresses do not permit the entire wafer to be examined with the conventional Lang method.

Schwuttke [78] has modified the transmission method to record large area topographs of highly-strained wafers. To accomplish this, the wafer is oscillated (Fig. B.2) about the Bragg angle during translating. The oscillation is set to "pick-up" the distorted regions and, thus, record a uniform density topograph. This method, Scanning Oscillator Technique (SOT), has been particularly successful for process characterization of defects. [79]

The third method, which is useful for process analysis, is the reflection method [67,68,80,81] as seen in Fig. B.3a. In this case, the incident beam is reflected from planes nearly parallel to the wafer surface and the diffracted beam is recorded on a photographic plate placed near the sample. An alternate method [82] of imaging large area wafers in reflection topographs is shown schematically in Fig. B.3b. This method yields the defect structure at the surface, to depths of 5-20 μM for silicon; the experimental method is simple but problems such as fluorescence scattering and multiple images have been problematical. However, the development of this method is particularly attractive because the surface layer is of primary importance in relating the substrate perfection to the epitaxial film [82] and correlating crystal defects in

the surface layer to the active junction regions in the epitaxial film. Examples of these methods are presented in the text, and the advantages and disadvantages of each are discussed in relationship to specific problems. These dislocations were introduced during crystal growth or some oxidation cycle prior to epitaxial growth and some fraction of the slip penetrated the epitaxial film surface.

B.3. Application of X-ray Topography to Process Analysis and Yield Improvement

As previously discussed, the major interest in x-ray topography stems from its nondestructive nature and defect sensitivity. These two features are of importance in critically evaluating a faulty process in regard to locating the steps responsible for defect activation. To assess the magnitude of the problem, the electrical results can be directly correlated to the defect map in the x-ray topograph. In this manner, several studies have clearly shown the deleterious effects of process-induced dislocations on diode reverse characteristics [67,68] and transistor gain. [83] To demonstrate the attribution of transmission SOT topography to process studies, we will first discuss some results of Raheja and Troutman [84] of our laboratory. Next, we will attempt to show the correlation between transmission and reflection topography and the relationship between crystal defects and emitter-collector shorts (pipes). [85]

B.4. Process Analysis with Transmission SOT Topography

Figure B.4 illustrates the concept of process analysis with x-ray topography. In this example, a single wafer is tracked from the polished state to after epitaxial growth. Note, the as-polished wafer is of moderate perfection ($\approx 10^3$ dislocation per cm^2) and is typical of most in-coming material.

However, the crystal perfection is drastically altered after the first few processing steps; the heavy dislocation pattern (in slip array) responds to thermal stresses produced by withdrawal from the oxidation furnace. [86] The wafer shows some alteration in the defect density after epitaxial growth, but the defect density at the periphery is still several orders of magnitude greater than the grown-in level. This point illustrates the fact that the controlling ingredient in the final defect density is the process rather than the quality of the as-grown crystal. This is confirmed by experiments with dislocation-free wafers which were simultaneously subjected to the same thermal stress. Upon completion of the transistor structure, the yield results from devices in the center and periphery could be related to the defect map in the transmission topograph to assess the magnitude of the yield loss. Based on these and similar experiments, the process can be optimized to reduce the amount of dislocations induced by the process step.

However, there are other problems to be encountered in processing, even if the dislocation generation is under control. Figure B.5 depicts the creation of yet another defect through continued thermal cycling. The as-polished wafer is dislocation-free, the spiral contrast regions are due to handling damage. Note, after the first oxidation, the wafer is essentially dislocation-free. The dislocations, which are observed at the ground flat, are probably due to embedded damage, not removed by etching, or due to position of the wafer in the oxidation boat. [7] Another contrast effect is beginning to appear in the crystal interior and the effect is amplified after epitaxial growth. The dark structureless region is thought to indicate precipitation of some dissolved impurity (probably oxygen). Note, the crystal strain is so intense that the crystal does not diffract near the central region. This type of microdefect was recently discussed by John [22] in terms of dissolved oxygen or carbon and the potential effect of these imperfections on device reliability. X-ray topography can not establish the exact nature of the defect, which illustrates the point that every method has limitations. It should be realized that these problems can only be understood by integrating other methods such as the TEM. However, this example clearly shows that dislocation-free crystals can still be highly imperfect. Hart [87] has employed x-ray topography and precision lattice parameter measurements to demonstrate that significant variations in lattice parameter or $\Delta d/d$ (strain) exist in dislocation-free silicon wafers. The variations were in the range of $\sim 10^{-4}$ in $\Delta d/d$ for a single sample and indicate that nonuniform distribution of impurities can give rise to crystal degradation even if the crystal is dislocation-free. These measurements suggest that dislocation density alone is insufficient to characterize the crystal perfection. The effect of the nonuniform impurity concentration (resistivity variations) on diode performance has been described by Goetzberger and co-workers. [88]

B.5. Correlation of Crystal Defects to Device Failure

The application of transmission x-ray topography to process-activated defects was described previously in this section. In these studies, the transmission technique was very effective in characterizing the volume defect state of the wafer. As was previously discussed, Schwuttke [79] has correlated device performance to yield loss through a yield map construction, i.e., superimpose the electrical data on the x-ray topograph to correlate defect density to device failure. However, since the transmission method records the superposition of the defects in the substrate and the epitaxial film, the resultant image (Fig. B.4b) is so dense that an exact correlation of defects in the active device region (epitaxial film) is obscured by defects in the substrate. A problem of correlating unresolved defect structures at the edge of the wafer to the yield map is that numerous effects are involved near the wafer edge (mask defects, alignment problems, etching, etc.) due to the wafer curvature. Smith [89] of our lab has also observed that numerous "good" devices are found at the highly defect wafer edge in transmission SOT topographs. The good devices were in many cases adjacent to failed devices. There are many possible explanations to these effects; however, rather than speculate on these unproven ideas, we will present data to relate surface defects to emitter-collector shorts. [85]

We will first demonstrate that when transmission and reflection topography are applied to the same wafer, then the defect analysis is considerably simplified. Figure B.6 depicts an SOT topograph of an integrated circuit wafer after electrical testing for emitter-collector shorts based on collector-base breakdown voltage measurements and collector-emitter leakage. The topograph reveals the defect extremes which occur in most defect-device correlations. The dislocation density in the center is low but increases dramatically toward the wafer periphery. Figure B.7 is an enlargement of a chip in the low dislocation region (a in Fig. B.6). Several dislocation sources, due to a steep concentration gradient at the diffused-indiffused interface, can be identified (1 and 2 in Fig. B.7); however, the inclination of the loops relative to the surface is difficult to determine in a single transmission topograph. Note that the individual transistors within the chip are observed by the intense diffraction contrast [78-90] at the transistor edge, which obscures the active emitter region. The corresponding reflection topograph is displayed in Fig. B.8. The observed dislocation density is reduced since only the surface layer ($\approx$ 10 μM) is contributing to the image. The location of the loop images in the reflection topograph (1 and 2 in Fig. B.8) indicate that the loops are activated at the chip edge and are punched downward on {111} planes to relieve the stress buildup at the apex of the stress riser (sharp corner at the square edge of the chip).

Individual dislocations can frequently be mapped from transmission to reflection topographs. The dislocation line (3) in Fig. B.7 is seen to penetrate the surface layer (Fig. B.8). The attitude of the defect can be inferred from the contrast broadening (degree of diffuseness) with depth in the crystal (Fig. B.8). A one-to-one correspondence between defects in reflection and transmission topographs is complicated by the fact that the image contrast is dependent on the specific set of reflecting planes used to obtain the topograph. [77,78] However, of more practical importance is the fact that the transistor structures are easily resolved and dislocations are seen to either penetrate the transistor region or not. In the case of these chips, the transistor that was tested yielded a leakage current of 250 μA ($\sim$ 4000 shunt resistance). The high resistance value indicates that excess emitter-collector leakage is not occuring, which is in agreement with the low dislocation density at the surface. Figure B.9 represents an enlargement of several chips near the wafer periphery (b in Fig. B.6). The dark bands are dislocation bundles that propogate in ⟨110⟩ directions; the defects are activated by slip to relieve thermal stresses during an oxidation or diffusion cycle prior to epitaxial growth. The spatial resolution is poor because of the high defect density and the superposition of defects in the substrate and film. Therefore, it is difficult to estimate the defect density at the surface or in the active device region from the transmission topograph. The defects, which penetrate the surface, are easily resolved in a reflection topograph of these same chips (Fig. B.10). These dislocations are in slip array and are inclined at 70.5° to the surface. Leakage current measurements from transistors in this region are in the range of 2000-3000 μA ($\sim$ 500 Ω). The excess leakage can be explained by emitter-collector pipes, i.e., phosphorus diffusion-down dislocations to produce a low resistance path. The dislocations in slip array are inclined to the surface at a large angle and have been identified by Barson [91] and Plantinga [92] as sources of pipes.

Another wafer with a higher pipe density was also investigated with x-ray topography to facilitate the correlation between dislocation density and leakage current. An enlarged area of a reflection topograph which depicts several chips and two Kerf

transistors is shown in Fig. B.11. Note that the observed dislocations are not in slip array but represent grown-in defects in the epitaxial film (a transmission x-ray topograph verified that the wafer contained minimal slip). The leakage current of the upper Kerf transistor was 240 μA (4170 Ω); the large resistance is in agreement with the low density of dislocations in the reflection topograph. Note, however, observation showed that the lower chip contained more dislocations and the leakage current increased to 1300 μA (770 Ω) in agreement with the increase in diffusion paths. Figure B.12 depicts a reflection topograph from the same wafer, but closer to the edge. The dislocated density is considerably higher than in the previous example; the leakage current is also increased to 4000 μA (250 Ω) and is consistent with the magnitude of surface defects.

The location of surface defects, as seen in reflection topographs, is shown to agree with electrical pipe data which implies that dislocations provide regions of enhanced diffusion of phosphorus (emitter diffusant) down the dislocation line to provide a low resistance path from emitter to collector.

The results indicate that device-defect interactions can be readily established with x-ray topography. The reflection mode is particularly attractive if the dislocation density obscures the surface region. However, the information of volume and surface defect distribution can be reconstructed by employing both transmission and reflection methods. The methods are compatible and provide information for process analysis and yield studies.

C. Process Characterization of Wafer-Shaping Operations

C.1. Introduction

In the previous sections we were concerned with illustrating process characterization from the as-polished wafer up to, but not including, the dicing operation. However, the concern about process-induced damage exists from the time the wafer is initially shaped by sawing, lapping, and polishing operations.

In this discussion, we will first review the concept of depth of damage, existing methods to measure the worked layer, and finally the structural model of surface damage. Following this review, we present an approach to characterize the sawing process from the point of view of the entire crystal, rather than the depth of damage on an individual wafer basis. [93]

The intent of most damage measurements is to associate a characteristic depth of damage with a particular wafer-shaping operation, i.e., sawing, lapping, polishing, etc. If this value is known, the damaged layer can be chemically removed with a minimum of material wasted. Damage problems accompanying the shaping processes in semiconductor manufacturing can be formulated as follows: (a) determine the depth of damage after various wafer-shaping operations, (b) define the true nature of the crystal defects, and (c) characterize the crystal-shaping process.

The investigations reported in the following section of this paper are directed toward the problems listed under (a) and (c). To understand the sawing process, we must examine the depth of damage on either side of the blade at various positions along the crystal length. In this manner, various anisotropies in damage depth can be exposed, and probable explanations formulated. At present, the characterization of the sawing process is incomplete and many questions remain unanswered; however, it is our purpose to expose a method for systematically studying these effects. It will also become obvious that the same method can be used to characterize the lapping and polishing operations, i.e., define the significant parameters that affect damage-depth anisotropies.

After investigating the sawing process, we will examine some effects of polishing damage on the quality of epitaxial films.[94] The pertinent references on defects in epitaxial films will be included. No attempt has been made to classify and review each defect type and its origin, since several excellent articles on this subject exist as pointed out earlier.

C.2. Review of Mechanically Induced Damage in Silicon

The concept of "depth of damage" has not been interpreted consistently in the literature. Some investigators assign this term to the maximum extent of the worked layer that can be measured directly. In comparison, others refer to "depth of damage" as the depth detectable by electrical phenomena, and here the depth of damage is deduced indirectly from the electrical measurements.

Principal direct methods used to evaluate damage are: the taper section metallographic method,[95] transmission electron microscopy, [57] and the etch rate method. [96,97] The principal indirect methods are the photomagneto electric (PME) effect [98] and the photoconductive decay (PCD). [99,100] The electrical techniques were used in germanium, but little was accomplished through them in silicon.

C.2.1. Direct Methods

Taper Section Method: The taper section method involves sectioning the abraded surface at a nominal angle of ~ 5°, which produces a geometrical magnification of ∿ 10:1 perpendicular to the section line. Optical magnification of the selectively etched surface [95] reveals the damage depth directly. The depth of damage is taken from the prominent crest of the surface irregularities to the extremities of the perturbed layer.

Transmission Electron Microscopy: The silicon sample is abraded on one side; the undamaged side is chemically thinned to ~ 2 μM. A tilting stage is employed to determine the exact nature of the damage, i.e., dislocations or cracks. The depth-of-damage values are obtained by progressively removing the abraded surface with a CP4 ($5HNO_3$:$3CH_3OOH$:3HF) etch to reveal the structure. The resultant surface is examined with an optical microscope, and the damage depth is determined by changing the focus (±1 μM estimated accuracy). [57]

Etch Rate Method: The dependence of etch rate on crystal perfection is used to determine an average damage depth. This method depends on the fact that damaged material etches more rapidly than undamaged material. The abraded samples are weighed and then etched to remove a thin surface layer; the etch rate is computed from the weight loss per unit etch time. This procedure is continued until the etch rate becomes constant. The depth at which this constancy occurs is termed the depth of damage. [97,98]

C.2.2. Indirect Methods

Electrical Measurement: The photomagneto electric (PME) effect and the photoconductive decay (PCD) methods are used to detect variations in the recombination rate of excess minority carriers below the surface. The surface recombination rate is enhanced at the damaged surface because of the high density of recombination centers associated with the worked layer. As the surface layer is removed through etching, the carrier recombination rate decreases and approaches a steady-state condition. The weight loss method is used to determine the depth at which the rate of carrier generation and recombination reaches equilibrium.

C.2.3. Surface Damage Models

Present models of surface damage respond either to the type of structural defects or to the effects of these crystal flaws on the electrical properties of the material.

The structural model relates the abrasion process to gouging particulate matter from the crystal surface by a cleavage and/or fracture mechanism. Wolff demonstrated the existance of cleavage in abraded material through optical goniometry. [99] The orientation dependence of damage, as shown by Faust, [96] also substantiates the conclusion above.

According to the structural model, three irregular damage zones are created at different depths in the crystal when the surface is abraded. The surface layer consists of microcracks, dislocation networks, and elastically strained regions. The interior zone contains dislocation networks and/or single dislocations. The density of dislocations decreases with increasing depth into the crystal until finally only

unperturbed material exists. This structure was first suggested by Faust [100] and later confirmed by Stickler and Booker using TEM. [57]

The electrical model of surface damage, as formulated from Buck's experiments [98] with PME and PCD, requires a thin layer of very low-lifetime material at the surface. The recombination centers in the distorted layers are assumed to be dislocations and vacancies. The acceptor action of these defects in Si is well-known.

After this brief review of the damage problem encountered in shaping silicon wafers for semiconductor manufacturing, we are ready to describe the experimental method which was employed by us to measure the depth and distribution of saw damage.

C.3. Experimental

Sample Preparation: Dislocation-free silicon crystals were chosen for this investigation. The (111) crystals were grown boron-doped to a resistivity of 0.07 to 0.09 ohm-cm. These crystals were prepared for sawing by fixing the crystals onto a wax-covered ceramic block. The lower ends of the crystals were uniformly covered with wax after mounting. A commercial HAMCO-ID[2] saw was used, with a 200-grit blade, 10 mils thick. The saw blade rotation was maintained at 3600 rpm with an observed cutting rate of 1-1/4 in./min. The saw was operated in a production environment, and no special cutting procedures were employed in the experiments. This suggests that process control (prior to emitter diffusion) of crystal perfection should consider other grounds for wafer rejection.

Figure C.1 depicts the schematic of a wafer identification system we used. The location of the wafer face relative to the saw blade is identified by arbitrarily labeling the wafer surface which is attached to the crystal bulk during cutting as A. The wafer surface on the opposite side of the blade is designated B. Samples are selected from the front, middle, and end of the crystal, and their surfaces are identified accordingly as A or B samples.

Experimental Approach: SOT topography records topographs of entire crystal wafers several inches in diameter even in the presence of elastic strains. Saccocio and McKeown [101] have shown that a single SOT topograph of a step-etched wafer can reveal the extent and also the distribution of damage in a saw-cut silicon slice. We have applied this technique to characterize crystal-shaping processes in terms of saw damage. Our approach is shown schematically in Fig. C.2. Accordingly, the sawed wafer contains a damage zone on both the A and B surfaces (not necessarily equal in depth). For the analysis of contiguous A and B surfaces, two adjacent wafers are used. The A surface of one wafer and the B surface of the other slice are step-etched in a solution of $3HC_2H_3O_2:2HNO_3:1HF$. Thus, a series of steps or mesas is made through the damaged layer. Each step exposes a different level of damage. These steps are shown schematically in Fig. C.2 and are designated from zero to 4. The step height is measured by interferometry, as described by Saccocio and McKeown. [101] It should be noted that Talysurf[3] measurements of the step heights are within 10% of the interferometry values. This method is preferred in the case of gross damage, whereas the interferometer can be well used for measuring steps in a mechanically polished wafer (planar surface). Samples can be examined to a greater depth in the same way, simply by masking one-half of the wafer on a line normal to the first set of steps and subsequently step-etching the material from the exposed half. In this manner, samples were evaluated in increments of 2-5 μM down to a depth of 40 μM or more. A typical topograph of two sawed wafers after step-etching is shown in Fig. C.3. It is clearly seen that the damage contrast delineates the boundary between damaged and undamaged material. Note the uniformity of the damage distribution in the wafer surface at the top of Fig. C.3 as compared to the wafer at the bottom. This damage anisotropy occurs in wafer surfaces on opposite sides of the saw blade and will be discussed in the following sections.

The measurements will be described in terms of the location along the crystal and the chosen adjacent surfaces. In this manner, the wafer damage can be monitored during the cutting process.

Trademark
[2]Hamco Machine & Elect. Corp.
[3]Taylor-Hobson Co.

Front Cuts: Figure C.4 is an SOT topograph of saw damage in the B surface of a wafer cut from the front of the crystal. The intensity of the x-ray diffraction contrast indicates the degree of deformation in the crystal. The dark contrast layer in Fig. C.4a at position 0 represents the crystal damage immediately after cutting. The reduction of contrast after the first step (1) is appreciable and extends uniformly across the wafer. It is interesting to note that after the second step (2) and after each following step, the damage contrast is nonuniformly distributed across the wafer. The areas in the topograph that show uniform or nonuniform contrast are designated I and II, respectively. The individual saw damage per wafer is displayed in a more quantitative way through damage maps constructed as shown in Fig. C.4b. Such maps are obtained by measuring the step height of the mesas and relating this information to a schematic of the damage contrast as obtained from the SOT topograph. The damage contrast at known increments below the surface represents a direct measure of the extent and distribution of saw damage. The damage map in Fig. C.4b indicates the direction of cutting as well as the two damage modes. For the purpose of discussion, we assume that these two modes have different origins; i.e., (I) represents a nonuniform mode of damage, which is related to the direction of the saw cut, and (II) is a uniform damage source, which reflects blade-wear, grit-size, rpm, etc. In this case (II) represents the nominal damage depth for this cut (~ 7 μM), whereas (I) is a deviation from the process optimum (~ 27 μM).

An SOT topograph of the corresponding A surface is shown in Fig. C.5a. (Note that the surfaces in Figs. C.4a and C.5a were separated by the same cut.) As shown previously, the maximum contrast occurs in the unetched mesa. In Fig. C.5b, the damage contrast is reduced uniformly after each step until its final elimination at 24 μM. It is interesting to note that the A surface contains only the uniform mode of damage (I), but the magnitude ∿ 24 μM is nearly identical to the extent of the nonuniform damage (II) in the opposite B surface (~ 27 μM).

Middle Cuts: An SOT topograph of the A surface of a sample selected from the middle of the crystal is seen in Fig. C.6a. Note that this time the damage is more extensive and nonhomogeneous across the A surface than across the B surface. (Recall that the A surface in the front cut did not contain nonuniform damage.) The damage map[4] (Fig. C.6b) depicts that the uniform damage ends 20 μM below the surface, whereas the nonuniform component extends deeper than 30 μM. The approximate depth of the nonuniform damage was evaluated from an adjacent wafer and found to be 39 μM.

The corresponding B surface shows homogeneous damage plus strong contrast striations (Fig. C.7a). In addition, the twice-etched half of the wafer displays peripheral damage. The depth of damage, as displayed in the map (Fig. C.7b), is 20 μM for the striations and 35 μM for the edge damage. Note that the underlying uniform damage disappears at about the same depth as the striations and that the individual striae end at nearly the same depth.

End Cuts: The x-ray topographic images of the A and B surfaces of wafers cut from the end of the crystal are essentially identical to Figs. C.6a and C.7a. The nonuniform damage in the A surface has a polished appearance similar to the one shown in Fig. C.6a; the nonuniform damage in the B surface consists of striations and edge damage of the type shown in Fig. C.7a. The damage map of the A surface (Fig. C.8) depicts the extent of the damage; the uniform damage is removed after 19 μM, and the nonuniform damage is ~ 38 μM deep. The B surface map (Fig. C.9) indicates ~ 19 μM of uniform damage and 27 μM of nonuniform edge damage.

Damage Profiles: An analysis of adjacent surfaces (separated by the saw blade during the same cut) permits a complete evaluation of the shaping process in terms of saw damage. Such an evaluation is made through damage profiles. The profile of the sawing process is obtained through recording and evaluating SOT topographs and damage maps of A and B slices taken at various locations along the crystal throughout the complete shaping procedure. Such damage profiles are shown in Figs. C.10 and C.11. Figure C.10 depicts the variation of nonuniform damage depth on A and B surfaces during the slicing procedure. In the initial phase of slicing, the depth of nonuniform damage in the A surface is considerably less than in the corresponding

[4]The damage map is constructed such that the mesa opposite the original surface is at nearly the same depth as the final step of the first etching (-19.5 μM). This is accomplished by measuring the total step heights of the first set of steps (-19.5 μM) and then thinning chemically one-half of the wafer to the same depth. This depth can be measured relative to the original surface (adjacent mesa) and the remaining step heights are measured relative to this value.

B surface. The damage profile shows also that during this phase, the nonuniform damage in the A surface rises sharply and becomes larger than the damage on the B surface after approximately 1-3/4 in. An almost reciprocal relationship is observed for the change of uniform damage depth shown in Fig. C.11. In the initial phase, the maximum uniform damage occurs in the surface of the A slice, but not of the B slice. Figure C.11 shows also that the depth of uniform damage for A and B surfaces approaches a value of 20 μM and stays constant throughout the rest of the slicing operation.

C.4. Discussion

The damage profiles and the topographs of the A and B surfaces are coupled to illustrate how the saw process can be evaluated. The information derived from this experiment is not sufficient to answer all of the questions that occur regarding the process; however, they are helpful in defining new experiments that are directed to a particular facet of the problem. The profiles of the cutting process are interpreted in a preliminary manner; then additional experiments are discussed in regard to specific questions here in this article.

The damage profiles shown in Figs. C.10 and C.11 reveal at a glance two main characteristics of the sawing process. The first is a damage anisotropy as evidenced through the different distributions of saw damage on A and B surfaces. A possible explanation is that the free surface (B) is more susceptible to blade deviation and inhomogeneous damage than the fixed (A) surface. The existence of inhomogeneous damage — flutter damage in the form of striations (Fig. C.7a) and peripheral damage at the point of initial blade contact (Fig. C.7a) — on the B surfaces is in agreement with this explanation. Note that the magnitude of the nonuniform damage in the A surface eventually exceeds that of the B surface after the transition region (1-3/4 in. after the initial cut). This can be related to the reduction of bulk crystal through slicing; consequently, the effect of vibration sinking attenuates during slicing. The decrease in nonuniform damage in the B surface (Fig. C.10) also reflects this effect.

The second significant characteristic of the sawing process is the cross-jumping of the damage modes. This is indicated in the cross-over of the A-B profiles in Fig. C.10 and the occurrence of flutter damage in the B surface, as indicated by the SOT topograph in Fig. C.7a. The exact source of the flutter-induced damage and the general amplification of damage in the crystal is not known. It is most likely that a change in blade tension during cutting is responsbile for the striation contrast on the B surface. The curvature of the striations on the B surface relative to the direction of sawing [102] suggests that this mode of damage occurs during blade removal. The polished appearance of the corresponding A surface would then reflect abrasive contact with the convex side of the deflected blade. The edge damage on the B surface (Fig. C.7a), at the point of initial blade contact, is also an indicator of either irregular blade removal and/or excessive feed-rate.

To determine the effect of feed-rate on nonuniform damage, another crystal of similar doping concentration, orientation, and size, as the crystal previously investigated was sliced in the same ID saw, but with a different blade. The blade had been used to cut 50 slices but was dressed prior to cutting the test crystal.

The test crystal was sliced in three zones, each two inches apart, where each zone contained approximately twenty wafers. The first zone was sliced at a feed-rate of 0.5 in./min, the second at 1.5 in./min, and third zone was processed at 2.5 in./min. The wafers were step-etched and SOT topographs were recorded to determine the damage depth. The damage contrast was eliminated after the first step of ~ 12 μM in each case. In Fig. C.12, some evidence of striations is still noticeable, which apparently indicates that feed-rate has a minimal effect on nonuniform damage as samples from all three zones revealed the same magnitude of striations. However, the fact that the striation contrast is evident when a new blade is employed suggests that the irregular blade-removal, coupled with a lack of blade tension, is the major contributor to the problems.

The effect of cutting with a 400-grit (~ 37 μM particle size) blade and the variability in damage depth at a specific location is illustrated in the damage profile in Fig. C.13. The range of depths for three A wafers and three B wafers is plotted at different locations along the crystal.

It is interesting to note that the maximum damage depth is $\leq$ 13 μM compared to $\leq$ 40 μM for the 200-grit blade. However, there is reason to believe that the 200-grit blade was faulty, which prevents any exact comparison of the depth of damage for different grit size blades. The existence of an effect of grit size on damage depth is, however, to be expected from the work of Faust [96]. Note, the maximum variability in damage depths also occur near the front of the crystal.

C.5. Summary

A systematic approach for characterizing sawing procedures of crystals in semiconductor manufacturing is demonstrated. This practice relies on damage profiles obtained from the cut crystal. Such profiles describe the entire shaping procedure of the crystal and are obtained through evaluation of SOT topographs of adjacent wafer surfaces. Damage profiles of crystals cut by standard shaping procedures reveal that the depth of uniform damage on either side of the crystal can vary up to 75%, that the variation in damage depth depends on the blade size, spatial location of the wafer, and probably blade tension.

D. Residual Damage and Epitaxial Film Perfection

D.1. Introduction

In the sawing operation, we saw that the damage depth could exceed 35 μM and that the mode of damage was dependent on processing parameters such as grit-size and blade tension. We recall that the purpose of the damage depth measurements was to determine the minimum amount of material-removal to insure a damage-free surface. However, the damage was seen to be highly variable across the wafer surface, from one side of the wafer to the other, and along the crystal length. The lack of a unique depth of damage for the sawing operation increases the possiblity of accepting damaged wafers after lapping and polishing. We will briefly examine some effects of residual damage, in the form of a mechanically polished surface, on the structural perfection of the epitaxial film. [27] Residual damage is, of course, only one form of process-induced flaws; others include: residual oxide particles, cleaning stains, handling damage, etc.

The analysis of structural defects in silicon epitaxial films has followed the development of the epitaxial process. Mendelson [31] discusses the nucleation of stacking faults at region of crystal distortion or surface discontinuity, such as scratches, dislocation slip, impurity segregation regions, oxide particles, and cleaning stains. It is also generally accepted that the majority of stacking faults originate at the substrate interface due to some type of mismatch or faulted layer. [27, 31] The mismatch is relieved by the formation of stacking faults along the {111} planes to form a tetrahedron whose intersection with the (111) surface exposes a triangle with ⟨110⟩ edges (Fig. D.1). Whereas, a considerable amount of experimental work has been directed to understanding the fault structure, the nature of stacking fault nucleation on damaged surfaces is not well understood.

D.2. Experimental

A qualitative description of the effect of residual damage is displayed in the SOT topographs in Fig. D.2.

The starting wafer was dislocation-free and damage-free; the wafer was mechanically polished with 1/4 μM Al_2O_3. The wafer was then step-etched to reveal the damage distribution in the same manner as the sawed wafer. The SOT topograph in Fig. D.2a indicates that the damage is removed after the first step of ~ 1.1 μM (as determined by interferometer and Talysurf measurements). Note that the polishing striations are evident in the topograph; the dark line at the damage-undamaged interface indicates that the deformed layer is under tension. [90,103] When the epitaxial film (~ 10 μM thick) was grown onto the partially damaged substrate, several distinct changes are noticed in the SOT topograph in Fig. D.2b. The most noticeable feature is that the diffraction contrast is increased in the previously polished section, which indicates the increase of crystal deformation and strain. Note, also, that the contrast at the damaged-undamaged interface has reversed and a slim white line can now be detected. The contrast reversal (tension-compression) of the faulted layer is in agreement with the relaxation of stress through plastic flow. To investigate the crystal strain throughout the film, the sample was step-etched at right angles to the initial steps. Figure D.2c displays the distribution in deformation in a

single-topograph. This type of sample preparation permits a direct comparison of faults across the damaged-undamaged interface as a function of depth in the film. The details of this study will be published in a separate article, [94] but significant observations are given here on the nature of the faults.

Figure D.3a is an optical photograph of the damaged-undamaged interface following a 15-sec Sirtl etch. [104] Several features are of interest.

1. The triangular faults are of the same size in the undamaged material (left half of photograph).

2. Some of the linear faults in the damaged region are nearly the same size as the faults in the undamaged side, which indicates that the film thickness is nearly uniform across the interface.

3. The triangular and linear faults in the damaged side are of various sizes.

The relative difference in dislocation density is displayed in Fig. D.3b, following a 1-min etch.

The stacking fault structure at specific locations in the film (grown on the polished surface) was also analyzed by isolating several faults of interest and coating the remainder of the sample with black Apiezon[5] wax (etch resistant). The fault was chemically sectioned with a solution of $3HC_2H_3O_2$:$2HNO_3$:1HF followed by a Sirtl etch to reveal the fault structure. An example of this is seen in Fig. D.4, which depicts a triangular fault. Note the dislocation "trails" at two of the apices. Figure D.5 illustrates two etch pits that terminate a linear fault. These are not altogether typical examples for all cases as numerous triangular faults did not reveal any etch pits at the apices. [27, 105].

D.3. Discussion

In summary, the existence of a thin-damaged layer can seriously degrade the quality of the epitaxial film. The degradation is in the form of dislocations and stacking faults (Fig. D.3b). The existence of variable-size triangular and linear faults in the film grown on the polished substrate clearly shows that faults can be generated in the epitaxial film. In the case of the triangular fault, bounded by stair-rod partials, this is probably the consequence of the tensile stress in the epitaxial film.

The effect of dislocations produced in the epitaxial process on emitter-collector shorts has been described by us earlier in this paper. We recall Queisser and Goetzberger [69] demonstrated that stair-rod partials could act as sites for microplasma breakdown when activated by impurities. Barson and co-workers also presented evidence that phosphorus diffusion-down the stair-rod dislocations can induce emitter-collector shorts (pipes). It is evident that the cooperation of anisotropic damage, experienced in wafer-shaping and defect generation during thermal cycles in a process, can be a factor in determining the component yield and possibly the reliability.

E. Adhesion of Oxide and Metal Films

E.1. Introduction

In this section of process-induced damage, we will discuss the characterization of thin-film adhesion, one of the most important but least understood areas of device processing.

Evaporated Al films are employed for interconnections in semiconductor-integrated circuits. The degree of local adhesion is one of the most important parameters to characterize in thin film technology. The durability of the film during device-processing operations is dependent upon the adhesion between substrate and film. In essence, adhesion is a measure of the ease of film removal.

Trademark

[5]James G. Biddle Co.

Numerous methods have been developed to investigate adhesion in thin films. The most prominent of these methods is the "Scotch tape"[6] test. This technique consists of pressing an adhesive tape onto the film and then removing it. The film is considered to have good adhesion if it is not lifted with the tape. [106] Heavens [107] developed a more quantitative method to measure film adhesion. He employed a chrome-steel point that was drawn across the film surface. A vertical load was gradually increased until the film ruptured. [108]

A new nondestructive method is reported to evaluate film adhesion before and after processing operations. [90,109] X-ray diffraction topography provides a simple adhesion map of the wafer. The adhesion map is actually an elastic imprint of the stress sites in the film mirrored by the silicon substrate. Figure E.1 depicts the origin of x-ray contrast at stress risers in thin films. [90] In this case, stress pile-up occurs at a window in a compressive film on a single-crystal substrate, i.e., SiO_2/Si. Since the thermal coefficient of expansion is greater than its oxide, the substrate will contract more than the film upon cooling (Figs. E.1a, b, and c). The resulting slice shape is nearly parabolic and convex to the film surface. The film prevents the substrate from returning to its original shape; therefore, the film is under compression. A simplified stress diagram is shown in Fig. E.1b. Since the substrate is constrained by the film, an opposing tensile stress is generated in the silicon. The silicon lattice constant is locally expanded at the substrate-film interface (Fig. E.1b). When a window is etched into the oxide layer (Fig. E.1c), the tensile stress is partially relieved. If the oxide is completely removed, the silicon returns to its original shape. The reflecting planes are bent to compensate for interfacial stress at the window-edge or at the interface between adhesion and no adhesion.

E.2. Experimental

A conceptual understanding of the contrast observed in x-ray topographs is depicted in Fig. E.2. The incident beam I_o strikes the crystal at the proper reflection angle. Two standing wavefields are excited by Bragg reflection, however, only one wavefield survives. The resulting energy flow ($I_o/2$) is parallel to the Bragg planes. [110] If the curvature at the exit surface is small, then the wavefield bends and energy is conserved. Energy or x-ray intensity is channeled in the direction of the curvature at the exit surface of the crystal. The net effect is that the film discontinuity locally alters the integrated intensity. The intensity is either increased above normal, or decreased below normal, depending on the sense of curvature. The effect of intensity channeling at a window in a SiO_2 film on silicon is shown in Fig. E.2. Note that the relative contrast is reversed in the direct beam and diffracted beam topographs. In a similar manner, variations in adhesion will alter the x-ray intensity relationships in the x-ray topograph and black-white images will be observed. An SOT transmission topograph of a metal film on an oxidized silicon substrate is shown in Fig. E.3. The contrast (black-white) relationship at a triangular window in the metal film is identical to the black-white contrast in the regions of poor adhesion that outcrop near the window. [90] Another example of poor adhesion is depicted in Fig. E.4; note the numerous contrast lobes that decorate the interface between good-poor adhesion. It should be realized that the x-ray contrast delineates the extent of the elastic strain field produced by poor adhesion rather than the exact size of the local film rupture.

The effect of surface contaminants on adhesion can be seen in x-ray topographs before and after evaporation. If the silicon substrate is subject to a light etch in buffered HF before evaporation, then the film adhesion is good. The evaporated films were deposited at 200 °C onto oxidized (111) Si wafers. An x-ray topograph after evaporation yields a uniform contrast film imprint (Fig. E.5). In a similar deposition cycle, other wafers were not pre-etched; the resulting effect is shown in Fig. E.6. The stress pockets are delineated by black-white contrast images along the film periphery. The wafers were inspected optically after deposition and no differences could be detected.

The next step in processing is to form patterns in the film with photo-etch methods. The resist was exposed through masks to produce the desired pattern. After developing, the resist was baked at 180 °C for 40 minutes. The pattern was then etched into the film and the resist layer was removed. The residue was cleaned by ultrasonic cleaning in alcohol. The effect of the photo-etch steps on adhesion of aluminum film deposited without the predeposition-etch step is shown in Fig. E.7.

Trademark

[6]3M Minnesota Mining & Manufacturing Co.

Figure E.7 shows regions of black-white contrast in the stripe images which still remain after processing. This result is expected because of poor adherence of the evaporated film. Note that the periphery of the Al film was completely removed in normal processing stresses, whereas the central region is intact but locally ruptured.

To illustrate the process analysis capabilities of stress topography, a single wafer is tracked through the evaporation-photoresist cycle to determine the extent and location of process-induced damage. Figure E.8 represents an SOT transmission x-ray after aluminum deposition; the sample was cleaned prior to evaporation to simulate the process step. However, there are indications of poor adhesion near the edge of the film as evidenced by the contrast centers. Figure E.9 depicts the same sample after the resist was developed and cured prior to etching. Note the stress imprint of the resist pattern that extends over the edge of the aluminum film. The topograph also yields information about the adhesion of the photoresist to the aluminum, photoresist to SiO_2, and the adhesion of aluminum to SiO_2. The absence of black-white contrast sites in the image indicates that the film adhesion has not been modified significantly at this stage in processing. The wafer after etching, resist-strip, and ultrasonic cleaning is shown in Fig. E.10. The stress contrast in the stripe is obvious. Note that the lower region of the film is actually removed, which is in agreement with the location of adhesion contrast centers in Fig. E.8 (after deposition). Of more importance, however, is the fact that processing stresses can affect the adhesion of previously adherent regions in the center of the wafer.

Additional experiments were conducted to determine the effect of other processing parameters on adhesion degradation. The essential philosophy of the experiment involved "treating" one-half of an oxidized wafer in a manner designed to produce poor adhesion, then fabricate patterns in each one for direct comparison.

Dislocation-free silicon wafers were chosen for the adhesion experiments to prevent the masking of adhesion contrast by intrinsic imperfections. The oxidized wafers were then treated in a manner to simulate conditions that could arise in processing and lead to adhesion degradation; i.e., improper cleaning, contamination, and handling. The samples were divided into two lots: one-half of the oxidized wafer was cleaned, as usual, in buffered HF; the remaining half was either (a) not cleaned in HF, or (b) contaminated with tap water and fingerprints. The treated wafers and a control were then employed as substrates for aluminum deposition. The deposited film was 20,000 Å thick and was grown in the sequence: 12,000 Å at a substrate temperature of 300 °C-8000 Å at 150 °C substrate temperature. The SOT topographs were obtained after opening contact holes in the oxide and after blanket metallization.

Figure E.11 is an SOT topograph of sample EF-9; the region marked (A) was cleaned in the conventional manner and (B) was treated with dried tap water and fingerprints prior to aluminum deposition. The topograph clearly shows adhesion degradation in the contamination region of the wafer (B), whereas the adhesion contrast effect is not observed in the A (cleaned) region. The wafers were then examined after formation of the device patterns through sub-etching. Figure E.12 displays an SOT topograph of EF-4. The adhesion degradation is difficult to discern because of the small, complicated device pattern. However, this difficulty was anticipated and resolved by permitting the contaminated region to include one large alignment block (A) and bypass the other (B). The adhesion contrast is clearly evident in Fig. E.13a that depicts block (A) and is absent in Fig. E.13b, which similarly depicts block (B). A similar effect was observed on the other wafers.

E.3. Results

A summary of the experimental results are summarized in Table E.1 (partial list). The table lists the sample number, the surface treatment, and the results of the adhesion map (SOT topograph). A "yes" response in the adhesion degradation column indicates the existence of contrast features typical of poor adhesion.

Table E.1. Adhesion data.

Sample Number	Surface Treatment (half-wafer)*	Adhesion Degradation (after deposition)
EF-3	Tap water - Fingerprint	Yes
EF-8	Tap water - Fingerprint	Yes
EF-9	Tap water - Fingerprint	Yes
EF-11	Tap water - Fingerprint	Yes
EF-2	No buffer etch	Yes
EF-6	No buffer etch	Yes
EF-7	No buffer etch	No
EF-4	No buffer etch	Yes
EF-1	No buffer etch	Yes

*One-half of the wafer was treated as listed and the other half was cleaned in buffered HF.

E.4. Discussion

X-ray topographic results indicate that the quality of deposited aluminum film can be significantly altered through faulty processing conditions. The observations are in agreement with previous SOT observations in which the effect of buffer HF cleaning on the adhesion of aluminum film was noted. The effect of process-induced adhesion faults on device failure have not been determined, but the correlation is now possible. However, the ability of the SOT technique to detect and diagnose potential failures before testing is partially confirmed; and further work along this line may provide a method to screen films, which are susceptible to specific failure modes.

F. Diffusion

F.1. Review

There is an extraordinarily large amount of published literature concerned with impurity-diffusion and self-diffusion in silicon. The reader is referred to the many good reviews on the subjects for a total view and detailed references. [111-114] Considerable confusion, however, regarding the diffusion mechanisms in silicon is still prevalent.

The most commonly used impurities in transistor technology are: P and As for n-type diffusion and B for p-type diffusion. We will concern ourselves mostly with P and B in this paper because of the availability of the trustworthy data. The four most important points which have not yet been reasonably explained about the diffusion of these elements in silicon are: (1) phenomenal concentration dependence of diffusivity, (2) total impurity profiles vs electrical impurity profiles, (3) anomalous junction depth behavior in open-tube phosphorus diffusions, and (4) cooperative diffusion anomalies.

F.1.1.A Phenomenal Concentration Dependence of Diffusivity

The diffusion coefficient of P is a constant up to the concentration of approximately 10^{20} cm^{-3}, but is a strong function of concentration above that value. With regard to this point, Tannenbaum's results [115] are given in Fig. F.1 for one temperature, T = 1050 °C. Maekawa's results [116] for boron diffusion at the same temperature, following similar trends, are shown in the same figure. The concentrations mentioned in this paper are always total, and not electrical, unless mentioned otherwise.

F.1.1.B Unexpected Impurity Distribution

As a consequence of the concentration dependence of diffusivity, the high surface concentration impurity distributions deviate considerably from complementary error functions expected for the constant source condition. For P and B, these deviations start for the values of $C_S \approx 10^{20}$ and $\approx 10^{19}$ cm^{-3}, respectively. Both the open-tube and the evacuated quartz-capsule diffusion systems give similar results with respect to points (1) and (2).

F.1.2. Total Impurity Profiles vs Electrical Impurity Profiles

At surface concentrations higher than $C_S \approx 10^{20}$ cm^{-3}, the total impurity profiles measured through the neutron activation or the radio-tracer techniques tend to remain higher on the concentration scale than the electrical impurity profiles measured usually by the four-point probe technique. The difference in the total impurity and the electrical impurity profiles is thought by many to be the electrically inactive or precipitated P. Similar results are true for boron.

F.1.3. Anomalous Junction Depth Behavior in Open-Tube Phosphorus Diffusions

Duffy et al. [83] have found a remarkable result that, in general, characterizes open-tube phosphorus diffusions. As shown in Figs. F.2a and F.2b, the junction depth (x_j) increases with phosphorus source concentration (in the ambient gas) up to a point and later on, remains constant. This happens in the deposition stage of diffusion. Following both deposition and drive-in cycle (Fig. F.2b), the values of x_j initially increase with source concentration, then decrease, and later on remain constant. The photo inserts in Fig. F.2b show that the junctions are ragged in the decreasing x_j region. An understanding of these effects is extremely important for controlling transistor fabrication processes.

F.1.4. Cooperative Diffusion Anomalies

Localized push-out and retardation of the base-collector junction underneath the emitter of the planar n-p-n transistors are well-known cooperative diffusion anomalies. In spite of a large amount of work, the anomalies still are unexplained. [117-126] There exists a number of unconnected theories; Willoughby has made an excellent review on this subject. [127] Later in this section, we will show a consistent explanation for all these four points. The basis of our explanation is the strains induced through diffusion of odd-sized impurity atoms.

The most important point yet unresolved with regards to self-diffusion in silicon is the nature of atomistic defects responsible for diffusion. We have a very reliable data [112] on the frequency factor D_o, activation energy ΔH and diffusivity D in the range of T = 1100 and 1400 °C. Out of the seven usually invoked mechanisms (namely, monovacancy, divacancy, interstitial, extended-interstitialcy, extended-vacancy, Frenkel disorder, and ring exchange), the most discussed ones are monovacancy, divacancy, and extended-interstitialcy for diffusion in silicon at high temperatures. Seeger and his associate [111] are in favor of extended-interstitialcy; while Kendall and his associates [112] favor divacancy. The monovacancy mechanism has stayed challenged consistently over the last few years. The resolution of the question of the mechanism of self-diffusion depends upon the resolution of the mechanism of impurity-diffusion. Later we will introduce the TEM evidence on the nature of diffusion-induced dislocation loops in silicon which, we think, gives support to Seeger's mechanism.

F.2. Synthesis

Let us now discuss the four points that we have considered crucial for understanding of the diffusion of P and B in silicon. Since results on boron diffusion are known to be similar to those on phosphorus diffusion, we will concentrate on phosphorus diffusion only.

F.2.1. Point 1

With regards to phosphorus impurity profiles, our own work [52] has shown that the concentration dependence of D is ficticious. In Fig. F.3, we have electrically measured profiles of phosphorus. From this figure, it will be seen that we do obtain complementary error function fit even for high C_S phosphorus profiles, such as A and B, if the junctions are shallow. As the value of x_j increases beyond 3750 Å, the reduction in C_S begins, although we expect its value to remain constant. In Fig. F.4, we have the plot of D vs 1/T (°K) based on the old data by Mackintosh. [129] The constant value of D obtained from our high-concentration curves A and B (Fig. F.3) and Tannenbaum's low-concentration ($C_S < 10^{20}$ cm^{-3}) value of D are about the same as can be seen in Fig. F.4. The agreement of these values of D, with those of Mackintosh, indicates that the well-known dependency of D on concentration is ficticious.

The so-called concentration dependence of D is apparent also from our curves C and D; it, however, ensues only after $x_j \approx 4000$ Å. The explanation for this is that the dislocation generation begins at this value of x_j. Figure F.5 illustrates this point. No dislocations could be observed with TEM in diffused samples corresponding to A- and B-impurity profiles given in Fig. F.3. Dislocation loops and patches of dislocation lines appeared in samples corresponding to impurity profile C given in Fig. F.3. The usual dense misfit-dislocation nets appeared in samples corresponding to impurity D given in Fig. F.3. We conclude, therefore, that dependency of D begins with generation of dislocations. The dislocation generation itself is not dependent on concentration C_s alone, it is Q dependent. These observations suggest that D is dependent on the total impurity atoms Q cm^{-3} penetrating silicon. The observations on normal phosphorus distributions with $C_s < 10^{20}$ cm^{-3} and x_j far deeper than 4000 Å (i.e., micron-range) are well documented in literature. Consequently, dependency of D on Q remains the only valid conclusion.

F.2.2. Point 2

We present in Fig. F.6, the electrical and the total profiles corresponding to the electrical profiles A and D given in Fig. F.3. We observe that there is only a little difference between An and A; however, the difference grows larger with deeper diffusion penetration. The critical widening of this difference begins with the generation of dislocation indicating inhomogeneous precipitation as the mechanism for the electrical inactivation of phosphorus. Precipitation of an impurity should naturally occur with cooling. This, however, is not very significant for shallow-junction high-concentration diffusions. Dislocation generation in the lattice is far more efficient in deactivating P as can be inferred from Figs. F.5 and F.6. It seems that dislocations behave like sponges in silicon.

F.2.3. Point 3

Results by Duffy et al. given in Fig. F.2b for the deposition and drive-in case are consistent with the sponge behavior of dislocations. In Fig. F.7, we have a TEM view of dislocations showing profuse precipitation of phosphorus. In further support of the theory of dislocations acting like sponges, Yoshida and Kanamori [130] very recently provided a remarkable evidence. They conclude that the junction retardations observed in Fig. F.2b are due to misfit-dislocation nets. We have in Fig. F.7 a view of an n-p-n junction bevel after application of Sirtl etch. The retardation of the emitter-base (E-B) junction is clearly visible here under the dislocation grooves. Raggedness of junctions observed by Duffy et al. in the source range — 1500 to 3000 ppm — implies nonuniform, or patchy, spread of dislocation nets. The uniformity of the junctions in the 3000, and higher, ppm-source range implies uniform spread of dislocation nets.

The explanation for the constant junction depth behavior after 1000 ppm source during deposition stage (Fig. F.2b) is again simple. These results are for essentially deep junctions ($x_j > 5000$ Å). Below 1000 ppm, the value of C_s is low and determined by dislocation-free junctions. Above 1000 ppm of C_s is high; however, a considerable amount gets absorbed in dislocation. These competing processes essentially maintain a constant C_s value irrespective of the source concentration in the glass on the top of silicon.

F.2.4. Point 4

In Table F.1, we have given a summary of cooperative diffusion effects observed by us. [128] The detailed results of our investigation will appear in a later publication. We used Sn and Ge, which are electrically neutral in silicon, to separate the possible field effect and to assess the effect of atom-size of diffusing atoms alone. High amounts of Sn and Ge, when diffused in silicon, caused misfit dislocations just as P or B. The change of sign of the cooperative diffusion phenomenon, when we switch Sn or Ge for P in the emitter and also when we change the diffusion sequence, is a reliable proof of the dominant role of the atom-size of the emitter impurity in controlling the collector-base junction depth. These effects are illustrated in Figs. F.8, F.9, and F.10.

Table F.1. Push-out effects (+ΔX) or retardation effects (-ΔX) obtained from different sequences of diffusions.

Normal Sequence

2nd Diffusion → / 1st Diffusion ↓	Emitter	
	P	Sn or Ge
B (Base)	+ΔX	-ΔX

Abnormal Sequence

1st Diffusion → / 2nd Diffusion ↓	Emitter	
	P	Sn or Ge
B (Base)	-ΔX	+ΔX

The three basic causes that one can skim out of the present various theories [9-17] for cooperative diffusion are: (1) lattice disorder, (2) electrostatic field, and (3) flooding of base-collector region by vacancies. We feel in the light of our evidence the first cause, namely lattice disorder (due to atom-size differences between the impurity atoms in the emitter and the silicon atoms) is the most significant cause for the cooperative diffusion effects. Since neutral impurities, like Sn and Ge, produce cooperative diffusion effects, the electrostatic field effect is not valid. Reiss et al. [131] have argued theoretically against the possibility of this effect. They say that at the early stages of a high-concentration diffusion, a space charge should be developed. This can happen only if a sufficient Fermi-level change occurs in a distance less than the Debye length at the diffusion temperature. They calculate this length to be too small (≃ 5 Å) at 1100 °C to be of significance. The third cause — the flooding of the base-collector region by vacancies — is felt to be too speculative to be valuable in further understanding.

Up to this point, we have shown a consistent picture relating impurity profiles and dislocations. We have also shown that the dislocations appear as loops in their first phase. We have good evidence [52] to show that these loops are sessile; and in the case of both phosphorus or boron diffusions, these interact to form misfit-dislocation nets. Generation of sessile loops due to condensation of point defects is well understood since the beginning of understanding the Kirkendall diffusion-effect.

In the case of phosphorus diffusion, we find that most of the sessile dislocation loops have their Burgers vectors parallel to the plane of diffusion. The loops induced via arsenic diffusion do possess Burgers vectors oriented in all possible directions. These considerations allow phosphorus and boron diffusions to produce misfit-dislocation nets parallel to the wafer surface. The considerable atomic misfit of P and B in Si makes this necessary for relief of misfit strains. No such dislocation arrangement is necessary in the case of As diffusion.

The most valuable information one can obtain from the diffusion-induced sessile loops is: (1) the activation energy for self-diffusion in silicon, if one obtains through annealing studies the growth-rates of these loops as a function of inverse absolute temperature, and (2) the defects responsible for diffusion if one finds the contrast of the loops in a systematic manner in TEM. We have done both of these things; the detailed results are in the publication given in reference [52]. The conclusions are that: (1) the dislocation loops are interstitial in nature and (2) the activation energy for silicon self-diffusion is found to increase from 2.12 eV near the surface to the mean value of about 5 eV in the vicinity of the diffusion junction. These values are valid for both P- and As-doped silicon. The value of 5 eV in

comparatively undoped silicon agrees well with the published results obtained through radio-tracer diffusion analysis. The method of proving the interstitial nature of the As-induced loops is illustrated in Figs. F.11a, b, and c. This result is equally valid for P-induced loops. The immediate conclusion one can draw from this result is that the Seeger mechanism – the extended interstitialcy mechanism – seems to be nearer to truth. We, however, cannot prove the "extendedness" of the interstitial defects. The interstitial-type contrast of these loops necessarily implies the interstitial nature of the defects responsible for diffusion during both the self-diffusion or impurity-diffusion.

G. Oxidation

G.1. Introduction

Studies in the formation of stacking faults bound by Frank partials in both the damage-free and damaged silicon after surface-oxidation at high temperatures have led to considerable understanding of the phenomenon of precipitation of oxygen in silicon. [54-59] Precipitation of oxygen in silicon in neutral ambients is quite a complex subject; the investigation was first started in 1957 by Kaiser and his associates. [60,62,132-134] It has now been realized that oxygen precipitation through surface-oxidation, or heat-treatments in neutral ambients, need not be considered unconnected. Oxygen trapped in silicon during crystal growth in a quartz crucible is known to precipitate at temperature as low as 450 °C. Kaiser and his associates gave some insight into the complicated relationship between the polymerization (i.e., formation of Si_xO_y complexes) of silicon and the associated electrical activity. Infrared-absorption band of silicon is strong, 9.1 μ, for silicon containing oxygen. The polymerization reactions, the nature of their products, and the amount of oxygen have been interpreted to a large extent by the infrared-absorption (IRA) techniques. [134-136]

Oxygen-precipitation in silicon through heat-treatment in neutral atmospheres or in vacuum has been the major concern of many researchers since the pioneering work of Kaiser and his group. Recently, Patrick and Dash [66] have done remarkable work through both TEM and IRA, which has brought considerable insight into the process of Si_xO_y formation. Patrick and Dash ruled out any precipitation in oxygen-bearing crystals at 450 °C. They found that oxygen atoms only leave their interstitital positions and migrate merely a few interatomic distances to occupy other lattice positions at that temperature. They also found that prolonged heat-treatment at 1025 °C gives oxygen-rich silicon clusters similar to three-dimensional Guinier-Preston zones. Some of these clusters grow into particles that produce pure-edge-type dislocations in otherwise dislocation-free crystals as can be seen in the TEM view given in Fig. G.1.

G.2. Synthesis

All the above discussion points out the important consideration that the growth of SiO_2, or similar particles, create precipitation stresses in silicon bearing supersaturated oxygen. The specific volume of SiO_2 is about ten per cent higher than that of silicon. The creation of pure-edge dislocations, or the creation of stacking faults bounded by the 1/3 ⟨111⟩ Frank partial-dislocations, involves accommadation of precipitation strains.

Stacking faults are created in the surfaces of silicon containing remnant surface damage following the treatment of oxidation. The experimental work by Thomas, [55] Queisser and Van Loom, [56] Stickler and Booker, [57] and Lawrence [65] regarding these stacking faults conclusively shows that the faulted dislocation loops pinned at $Si-SiO_2$ interface provide an extremely fast heterogeneous mode of precipitation for oxygen, which is illustrated in Fig. A.7. Silicon is a low-stacking fault energy matrix. In the oxygen-precipitation process, nucleation is controlled to a large extent by strain-energy consideration. The interstitial sites of oxygen occupy ⟨111⟩ direction sites. As the Burgers vector of the Frank partial of the faulted dislocation loop is also of the 1/3 ⟨111⟩ type, accomodation of the misfit strains of the Si-O-Si molecular units in the matrix, therefore, is easily accomplished. The growth of SiO_2 platelets on the {111} planes inclined to the (111) silicon surface in Joshi's observations [54] is consistent with these considerations. The growth of precipitation through the generation of stacking faults was first formulated by Silcox and Tunstall. [137] Although there was no remnant damage in Joshi's samples, the previous annealing heat-treatment at 1000 °C had already initiated nucleation of Si_xO_y complexes. In Fig. G.2, we observe through TEM small dislocation loops lying in the field of these complexes. Oxidation of the surface only accelerated the growth of the already nucleated Si_xO_y complexes through growth of the loops into stacking faults.

Patrick and Dash discovered that the precipitation of oxygen is observable for long-time heat-treatment in the temperature range 890 to 1100 °C, and that the maximum precipitate rate is at 1000 °C, which is the same heat-treatment at which Joshi saw profuse precipitation of oxygen. Through the TEM method of vanishing contrast for stacking faults, the profuse precipitate-growth in the stacking faults can be seen in Fig. G.3.

All the researchers agree that the steam-oxidation far more accelerates the generation of stacking faults and, therefore, the oxygen-precipitation than the dry-oxidation. This is most possibly related to the rate of consumption of silicon, which is faster during steam-oxidation than during dry-oxidation. Since the faults are extrinsic, they will either grow by vacancy emission or interstitial absorption at the Frank partials. It was pointed out in the diffusion section that the self-diffusion in silicon is via some kind of interstitial silicon defect. The authors feel that these types of interstitial defects would be absorbed at the partials, causing them to grow and penetrate into silicon.

H. References

[1] K. G. McKay and K. B. McAfee, Phys. Rev. 91, 1079 (1953).

[2] K. G. McKay, Phys. Rev. 94, 877 (1954).

[3] A. G. Chynoweth and K. G. McKay, Phys. Rev. 102, 369 (1956).

[4] A. G. Chynoweth and K. G. McKay, Phys. Rev. 106, 418 (1957).

[5] A. G. Chynoweth and G. L. Pearson, J. Appl. Phys. 29, 1103 (1958).

[6] K. S. Champlin, J. Appl. Phys. 30, 1039 (1959).

[7] R. L. Batdorf, A. G. Chynoweth, G. C. Dacey and P. W. Foy, J. Appl. Phys. 31, 1153 (1960).

[8] William Shockley, Solid State Electronics 2, 35 (1961).

[9] M. Kikuchi and K. Tachihawa, J. Phys. Soc. Japan, 14, 1830 (1959).

[10] M. Kikuchi, J. Phys. Soc. Japan, 15, 1822 (1960).

[11] B. McDonald, A. Goetzberger and C. Stephens, Bull. Am. Phys. Soc. 4, 455 (1959).

[12] A. Goetzberger and W. Shockley, J. Appl. Phys. 31, 1891 (1960).

[13] H. J. Queisser, K. Hubner, and W. Shockley, Phys. Rev. 123, 1245 (1960).

[14] H. J. Queisser, J. Electrochem. Soc. 110, 52 (1963).

[15] A. Goetzberger and W. Shockley, J. Appl. Phys. 31, 1821 (1960).

[16] A. Goetzberger, Bull. Am. Phys. Soc. 5, 160 (1960).

[17] A. Goetzberger, B. MacDonald, R. Haito and R. M. Scarlett, J. Appl. Phys. 34, 1591 (1963).

[18] G. H. Schwuttke, J. Appl. Phys. 36, 2717 (1965).

[19] J. M. Fairfield and G. H. Schwuttke, J. Electrochem. Soc. 113, 1229 (1966).

[20] G. H. Schwuttke, Proc. of the Internat'l Sym. Test Methods and Measure., Budapest, Hungary (April 1967).

[21] H. J. Queisser, Solid. State Electronics 5, 1 (1962), L. Physik 176, 313 (1963).

[22] H. F. John, Proc. IEEE 55, 1249 (1967).

[23] R. Stickler and G. R. Booker, J. Electrochem. Soc. 109, 743 (1962); Phil. Mag. 8, 859 (1963); J. Electrochem. Soc. 111, 485 (1969).

[24] J. W. Faust, Jr. and R. Stickler, Electrochem. Tech. 4, 399 (1966).

[25] T. B. Light, "Metallurgy of Semiconductor Materials," Met. Soc. Conf., Vol. 15 (Interscience, New York, 1962) pp. 137-158.

[26] J. M. Charig, B. A. Joyce, D. J. Stirland, and R. W. Buckness, Phi. Mag. 7, 1847 (1962).

[27] G. R. Booker and R. Stickler, J. Appl. Phys. 33, 3281 (1962).

[28] G. R. Booker and R. Stickler, Appl. Phys. Letters 3, 158 (1963).

[29] G. R. Booker and B. A. Unvala, Phil. Mag. 9, 691 (1964).

[30] S. Mendelson, M. H. Francombe and H. Sato (Eds.), Conf. Single Crystal Films, Blue-Bell, Pa. (Pergamon Press, New York, 1964) p. 251.

[31] S. Mendelson, J. Appl. Phys., 35, 1570 (1964).

[32] R. H. Finch, H. J. Queisser, G. Thomas and J. Washburn, J. Appl. Phys. 34, 406 (1963).

[33] G. H. Schwuttke and V. Sils, J. Appl. Phys. 34, 3127 (1963).

[34] H. J. Queisser, R. H. Finch and J. Washburn, J. Appl. Phys. 33, 1536 (1962).

[35] S. Mendelson, Mat. Sci. Eng. 1, 42 (1966).

[36] N. Holonyak, Jr., D. C. Jillson and S. F. Bevacaua, "Metallurgy of Elemental and Compound Semiconductors," AIME Met. Soc. Conf. (Interscience, New York, 1902) Vol. 12, p. 81.

[37] J. Amick, M. H. Francombe and H. Sato (Ed.), Conf. Single Crystal Films, Blue Bell, Pa. (Pergamon Press, 1964) p. 241.

[38] R. S. Wagner, W. C. Ellis, K. A. Jackson and S. Arnold, J. Appl. Phys. 35, 2993 (1964).

[39] H. J. Queisser, J. Appl. Phys. 32, 1776 (1961).

[40] S. Prussin, J. Appl. Phys. 32, 1876 (1961).

[41] G. H. Schwuttke and H. J. Queisser, J. Appl. Phys. 33, 1540 (1962).

[42] J. Washburn, G. Thomas, and H. J. Queisser, J. Appl. Phys. 35, 1909 (1964).

[43] E. Levine, J. Washburn, and G. Thomas, J. Appl. Phys. 38, 81 (1967).

[44] M. L. Joshi and F. Wilheim, J. Electrochem. Soc. 112, 185 (1965).

[45] R. J. Jaccodine, Appl. Phys. Letters 4, 114 (1964).

[46] W. Czaja, J. Appl. Phys. 37, 3441 (1966).

[47] M. Yoshida, H. Arata, and Y. Terunuma, Japan J. Appl. Phys. 7, 209 (1968).

[48] M. Yoshida, Japan J. Appl. Phys. 9, 338 (1970).

[49] M. C. Duffy, F. Barson, J. M. Fairfield and G. H. Schwuttke, J. Electrochem. Soc. 115, 84 (1968).

[50] M. L. Joshi, B. J. Masters and S. Dash, Appl. Phys. Letters 7, 306 (1965).

[51] M. L. Joshi and S. Dash, IBM J. Res. Develop. 11, 271 (1967).

[52] S. Dash and M. L. Joshi, Diffusion-Induced Defects and Diffusion Kinetics in Silicon, Am. Soc. Testing Mater. Proc., to be published.

[53] J. E. Lawrence, J. Electrochem. Soc. 113, 819 (1966).

[54] M. L. Joshi, Acta Met. 14, 1157 (1966).

[55] D. J. D. Thomas, Phys. Sta. Sol. 3, 2261 (1963).

[56] H. J. Queisser and P. G. G. Van Loom, J. Appl. Phys, 35, 3066 (1964).

[57] R. Stickler and G. R. Booker, Phil. Mag. 11, 1303 (1965).

[58] V. A. Phillips and W. C. Dash, J. Appl. Phys. 33, 568 (1962).

[59] F. Wilhelm and M. L. Joshi, J. Appl. Phys. 36, 2592 (1965).

[60] W. Kaiser, Phys. Rev. 105, 1751 (1957).

[61] G. R. Booker and W. J. Tunstall, Phil. Mag. 14, 71 (1966).

[62] C. S. Fuller and R. A. Logan, J. Appl. Phys. 28, 1427 (1957).

[63] H. Reiss and W. Kaiser, Properties of Elemental and Compound Semiconductors (Interscience, New York, 1960), pp. 103-119.

[64] W. Czaja and J. R. Patel, J. Appl. Phys. 36, 1476 (1965).

[65] J. E. Lawrence, J. Appl. Phys. 40, 360 (1969).

[66] W. Patrick and S. Dash, "Oxygen Precipitation in Silicon," to be published.

[67] E. D. Jungbluth and P. Wang, J. Appl. Phys. 36, 1967 (1965).

[68] Semiconductor Junction Properties as Influenced by Crystallographic Imperfections: Scientific Reports, No. 1 and No. 2, Contract No. AF 19(628)-5059/AFCRL-67-0564, Prepared for Air Force Cambridge Research Laboratories, Bedford, Mass. 01730

[69] H. J. Queisser and A. Goetzberger, Phil. Mag. 8, 1063 (1963).

[70] F. Barson, W. J. Armstrong, and W. E. Mutter, J. Electrochem. Soc. 111, 1263 (1964).

[71] J. E. Lawrence, J. Electrochem. Soc. 115, 860 (1968).

[72] J. E. Lawrence, Semiconductor Silicon (The Electrochemical Society,, New York, 1969), pp. 596-607.

[73] H. J. Queisser, Semiconductor Silicon (The Electrochemical Society, New York, 1969) pp. 585-595.

[74] R. G. Rhodes, Imperfections and Active Centers in Semiconductors (Macmillan, New York, 1964).

[75] W. R. Runyan, Silicon Semiconductor Technology (McGraw-Hill, New York, 1965).

[76] Transistor Technology II (Van Nostand, New York, 1958).

[77] A. R. Lang, J. Appl. Phys. 29, 597 (1958); A. R. Lang, J. Appl. Phys. 30, 1748 (1959).

[78] G. H. Schwuttke, J. Appl. Phys. 36, 2712 (1965).

[79] G. H. Schwuttke, Sym. Manufacturing In-Process Control and Measuring Technique for Semiconductors, Phoenix, Arizona (1966).

[80] C. S. Barrett, Structure of Metals (McGraw-Hill, New York, 1952); J. B. Newkirk, Trans. AIME 215, 483 (1959).

[81] V. Bonse, M. Hart, and J. B. Newkirk, Advances in X-Ray Analysis (Plenum Press, New York, 1966) Vol. 10, p. 1.

[82] J. K. Howard and R. D. Dobrott, Appl. Phys. Letters, 1, 101 (1965); J. K. Howard and R. Cox, "Advance in X-Ray Analysis," (Plenum Press, New York, 1965) Vol. 9, p. 35.

[83] M. C. Duffy, F. Barson, J. Fairfield, and G. Schwuttke, J. Electrochem Soc. 115, 84 (1968).

[84] R. Rahia and H. Troutman, unpublished results.

[85] J. K. Howard, submitted to external journal.

[86] S. M. Hu, J. Appl. Phys. 40, 4413 (1969).

[87] M. Hart, Proc. Royal Soc. A309, 281 (1969).

[88] A. Goetzberger, B. McDonald, R. Haitz, R. M. Scarlett, J. Appl. Phys. 34, 1591 (1963).

[89] P. S. Smith, unpublished results.

[90] G. H. Schwuttke and J. K. Howard, J. Appl. Phys. 39, 1581 (1968).

[91] F. Barson, M. S. Hess, and M. Roy, J. Electrochem. Soc. 116, 304 (1969).

[92] G. H. Plantinga, IEEE Trans. 16, 394 (1969).

[93] J. K. Howard, J. Regh, and G. H. Schwuttke, J. Electrochem. Soc. 114, A227 (1967).

[94] J. Regh and J. K. Howard, unpublished results.

[95] E. N. Pugh and L. E. Samuels, J. Electrochem. Soc. 108, 1043 (1964).

[96] J. W. Faust, Electrochem. Tech. 2, 339 (1964).

[97] R. Stickler and J. Faust, Electrochem. Tech. 4, 399 (1966).

[98] T. M. Buck, The Surface Chemistry of Metals and Semiconductors (John Wiley & Sons, Inc. New York, 1960), p. 107.

[99] G. A. Wolff, J. M. Wilber, and J. Clark, Z. Electrochem. 61, 101 (1957).

[100] J. W. Faust, "Symposium on Cleaning of Electronic Device Materials," (ASTM special Tech. Pub. 246, Philadelphia, Pa., 1959), p. 66.

[101] E. J. Saccocio and W. McKeown, J. Appl. Phys., 37, 2702 (1967).

[102] W. R. Runyan, Silicon Semiconductor Technology (McGraw-Hill, New York, 1965), p. 231.

[103] E. S. Meieran and I. A. Blech, J. Appl. Phys. 36, 3162 (1965).

[104] E. Sirtl and A. Adler, Z. Metallk 52, 529 (1961).

[105] T. L. Chu and I. R. Gavaler, J. Electrochem. Soc. 110, 338 (1963).

[106] J. Strong, Rev. Sci. Inst. 6, 97 (1958).

[107] O. S. Heavens, J. Phys. Radium 11, 355 (1950).

[108] P. Benjamin and C. Weaver, Proc. Roy. Soc. (London) A 254, 163 (1960); A 274, 267 (1963).

[109] J. K. Howard and R. F. Ross, J. Electrochem Soc. 114, 142c (1967).

[110] M. Hart and A. R. Lang, Acta. Cryst. 16, A102 (1963).

[111] A. Seeger and K. P. Chik, Phys. Stat. Sol. 29, 455 (1968).

[112] D. L. Kendall and D. B. DeVries, Semiconductor Silicon (The Electrochemical Society, Inc., New York, 1969), p. 358.

[113] H. Reiss and C. S. Fuller, Semiconductors (Reinhold Publishing Corporation, New York, 1959), p.244.

[114] B. I. Boltaks, Diffusion in Semiconductors (Academic Press, New York, 1963).

[115] E. Tannenbaum, Solid State Electronics 2, 123 (1961).

[116] S. Maekawa, J. Phys. Soc. Japan 17, 1592 (1962).

[117] L. E. Miller, Properties of Elemental and Compound Semiconductors (Interscience Pub. Ltd., New York, 1960), p. 303.

[118] T. H. Yeh, Bull. Amer. Phys. Soc. 7, 332 (1962).

[119] Y. Sato and H. Arata, Japan J. Appl. Phys. 3, 511 (1964).

[120] P. Baruch, C. Constantin, J. C. Pfeister, R. Saintersprit, Discussions Faraday Soc. 31, 76 (1961).

[121] R. Gereth, P.G.G. Van Loom and V. Williams, J. Electrochem. Soc. 112, 323 (1965).

[122] K. H. Nicholas, Solid State Electronics 9, 35 (1966).

[123] R. Gereth and G. H. Schwuttke, Appl. Phys. Letters 8, 55 (1966).

[124] J. E. Lawrence, J. Electrochem. Soc. 113, 819 (1966).

[125] J. E. Lawrence, J. Appl. Phys. 37, 4106 (1966).

[126] J. E. Lawrence, J. Electrochem. Soc. 115, 860 (1968).

[127] A. F. W. Willoughby, J. Mater. Sci. 3, 89 (1968).

[128] M. L. Joshi and T. H. Yeh, to be published.

[129] I. M. Mackintosh, J. Electrochem. Soc. 109, 392 (1962).

[130] M. Yoshida and S. Kanamori, Japan. J. Appl. Phys. 9, 338 (1970).

[131] H. Reiss, C. S. Fuller and F. J. Morin, Bell System Tech. J. 35, 535 (1966).

[132] W. Kaiser, H. L. Frisch, and H. Reiss, Phys. Rev. 112, 1546 (1958).

[133] H. J. Hrostowski, R. H. Kaiser, J. Phys. Chem. Solids 9, 214 (1959).

[134] W. Kaiser, and P. H. Keck, JAP 28, 882 (1957).

[135] A. Smakula, and J. Kalnajs, J. Phys. Chem. Solids, 6, 46 (1958).

[136] T. Arai, J. Phys. Soc. Japan 17, 246 (1962).

[137] J. M. Silcox and W. J. Tunstall, Phil. Mag. 10, 361 (1964).

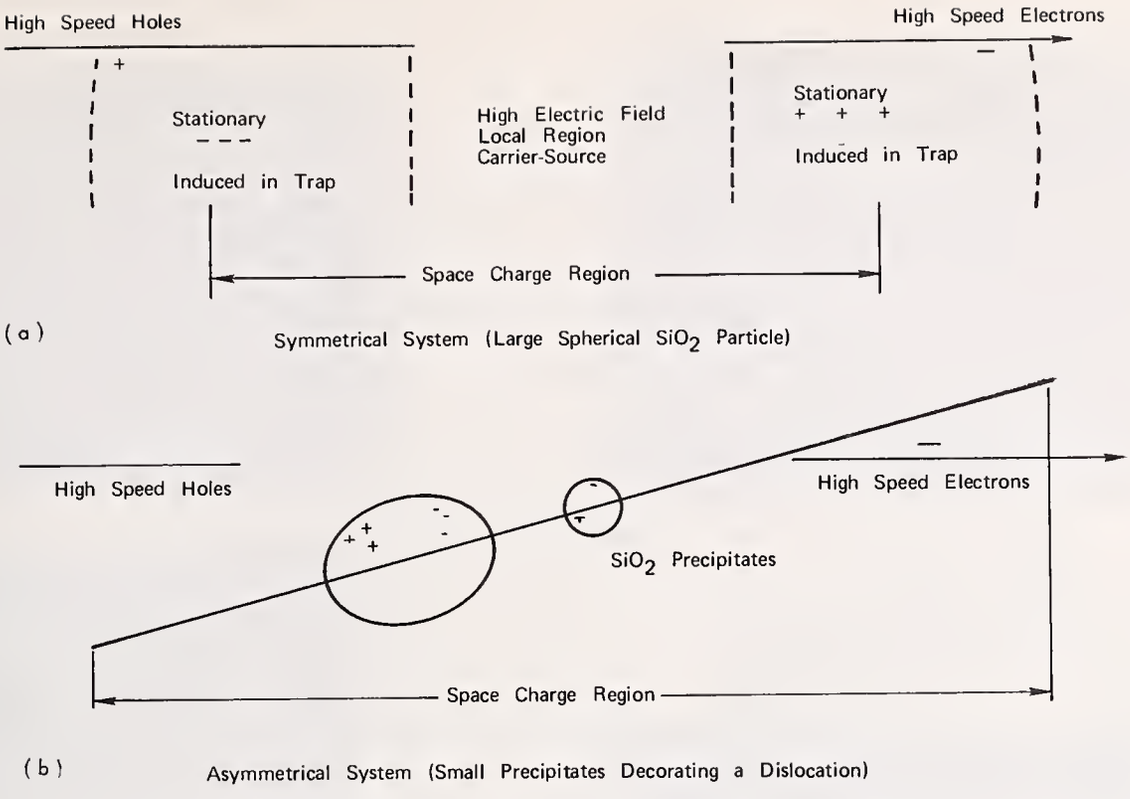


Fig. A.1. Shockley's microplasma models for p-n junctions.

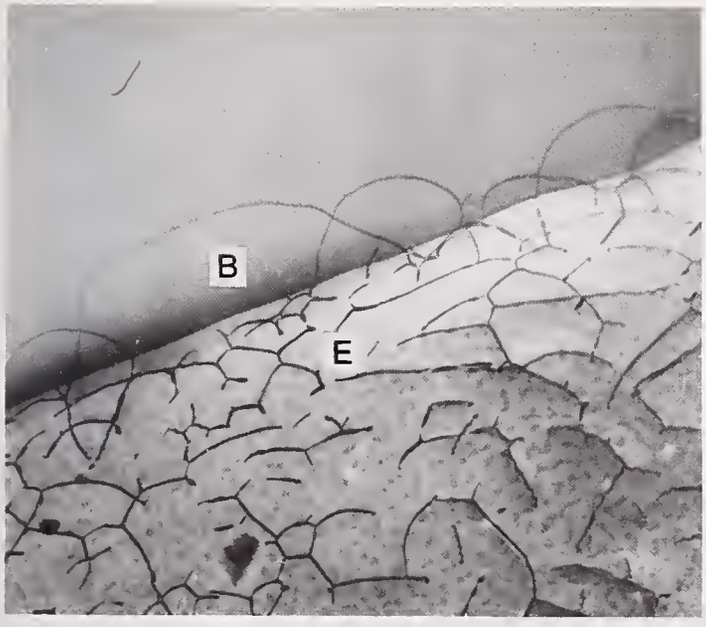


Fig. A.2. Spreading of dislocations from emitter E into base B.

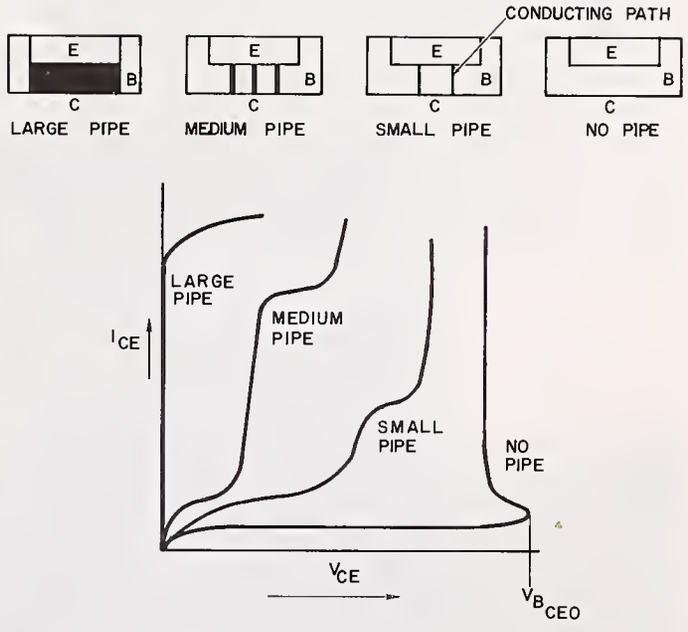


Fig. A.3. Effect of varying degree of pipe formation on the transistor function: I_{ce} vs V_{ce} with base floating. Note the reduction in $V_{b\ ceo}$ with the increased intensity of pipe activity.

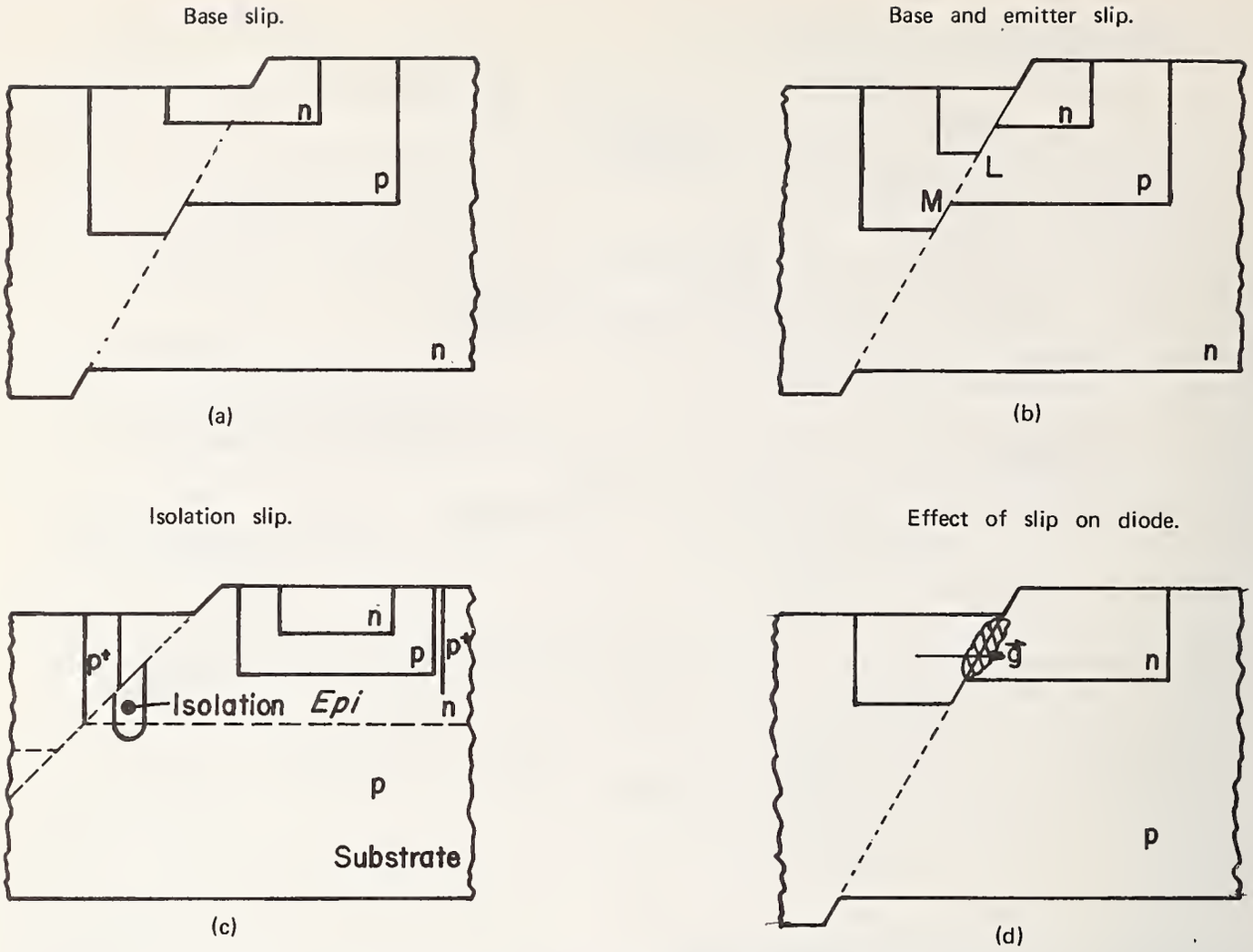


Fig. A.4. Effect of slip on vertical structures of transistors.

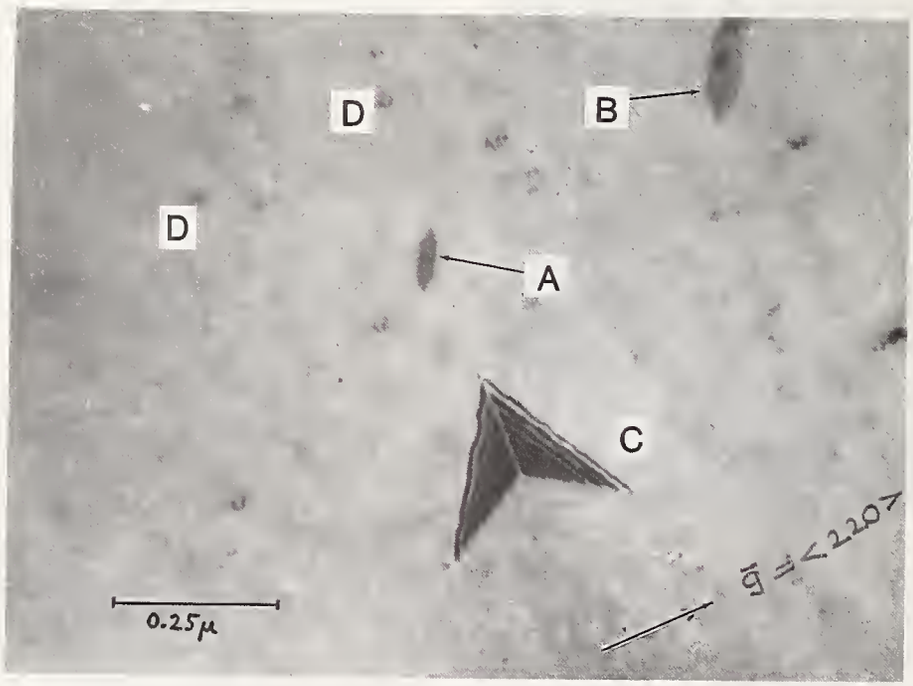
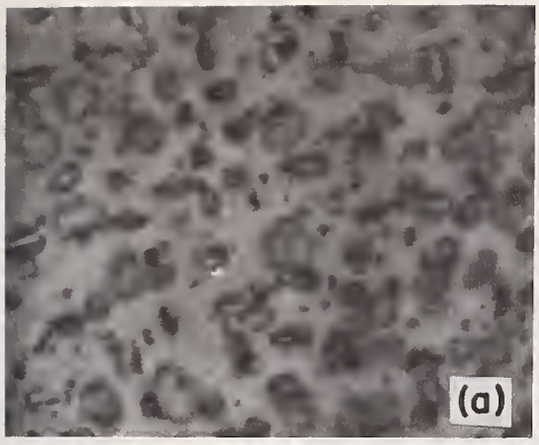
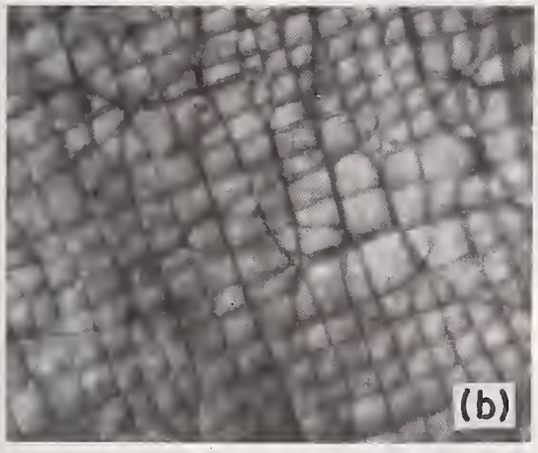


Fig. A.5. Note association of faulted dislocation loops A and B of hexagonal nature and triangular loops D with the epitaxial stacking-fault C. Origin of C is thought to be such dislocation loops by some researchers.





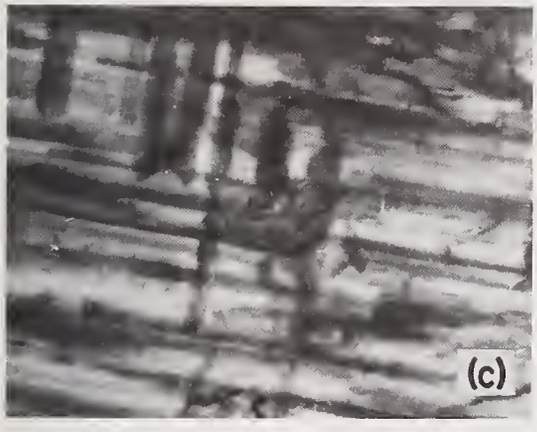


Fig. A.6. Various kinds of structural defects observed in (100) silicon diffused with high amounts of phosphorus.
(a) Dislocation loops are formed in initial diffusion stage.
(b) Dislocation nets are formed after interaction within loop dislocations in final diffusion state.
(c) Thin precipitate platelets of phosphorus-silicon complexes are often observed in diffused layers.

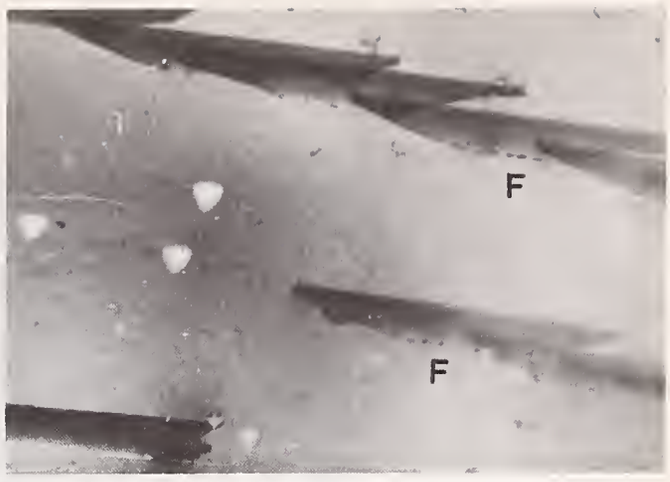


Fig. A.7. Formation of stacking faults in silicon, heat-treated at 1000 °C in a neutral ambient and then steam-oxidized. Observe the Frank-partials F decorated with oxide particles and penetrating into silicon.

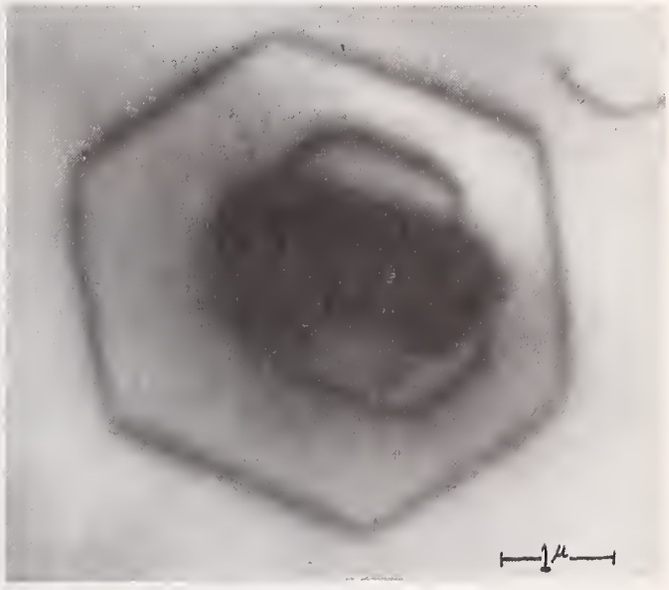


Fig. A.8. Formation of dislocation loops in heat-treated silicon after development of a large SiO_2 particle.

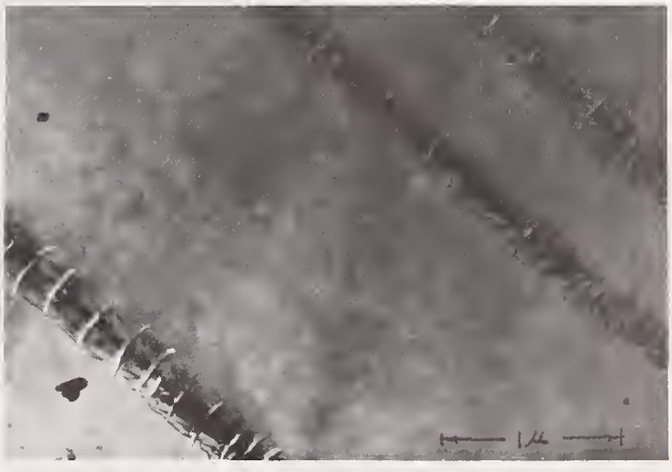


Fig. A.9. Scratches on silicon surfaces have developed into rows of dislocation loops after heat-treatment at 1000 °C.

Fig. A.10. Observation of stacking fault roots in a hot-stage TEM.
(a) Two stacking faults of tetrahedron-type. Observe that one side of these faults vanishes for parallel setting of diffraction vector $\bar{g}$.
(b) Same two stacking faults after heating in the TEM. Note that stacking faults are annealed out of silicon and only impurities delineating the faults are revealed.

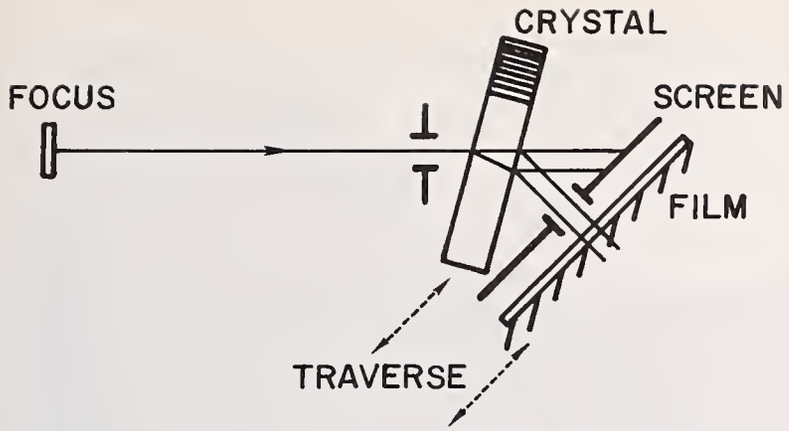


Fig. B.1. Scanning method developed by Lang.

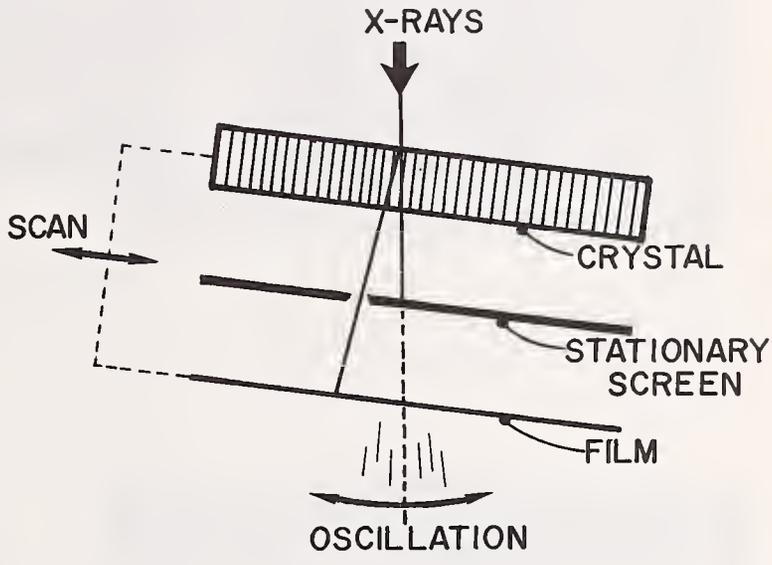


Fig. B.2. Transmission SOT method of Schwuttke. Crystal and film scan and oscillate to record large area topographs in preserve of wafer strain.

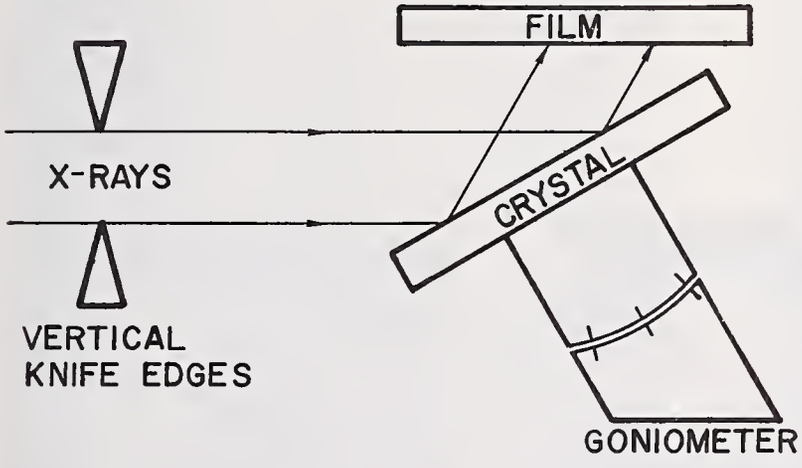


Fig. B 3a. Experimental arrangement of Berg-Barrett method for investigating surface defects.

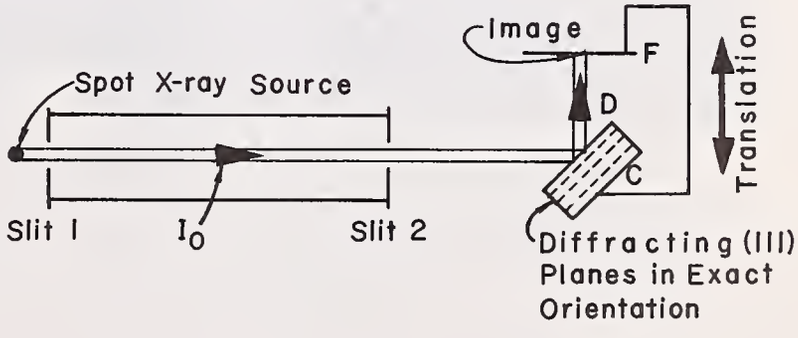


Fig. b 3b. Experimental arrangement for scanning-reflection x-ray topography.

I

X-RAY TRANSMISSION MICROGRAPHS

Sample: Chemically Polished P^- (Boron Doped) ⟨111⟩ Axially Oriented Wafer

Radiation: $Mo_{K_{\alpha 1}}$

Reflection: [110]

Magnification: 2.5X

As Polished Wafer | First Oxide | Epitaxial Growth

Subcollector Application,
Expose, Develop, Etch, Remove

Arsenic Subcollector Diffusion

Oxide Drive In

Oxide Removal

Fig. B.4. Transmission SOT topographs of wafer perfection at various processing steps.

II

X-RAY TRANSMISSION MICROGRAPHS

Sample: Chem-mech Polished P^- (Boron Doped) ⟨111⟩ Axially Oriented Wafer

Radiation: $Mo_{K_{\alpha 1}}$

Reflection: [110]

Magnification: 2.5X

As Polished Wafer

First Oxide

Subcollector Application,
Expose, Develop, Etch, Remove

Arsenic Subcollector Diffusion

Oxide Drive In

Oxide Removal

Epitaxial Growth

Fig. B.5. Transmission topograph of process-induced flaws in silicon wafer. (See text for discussion.)

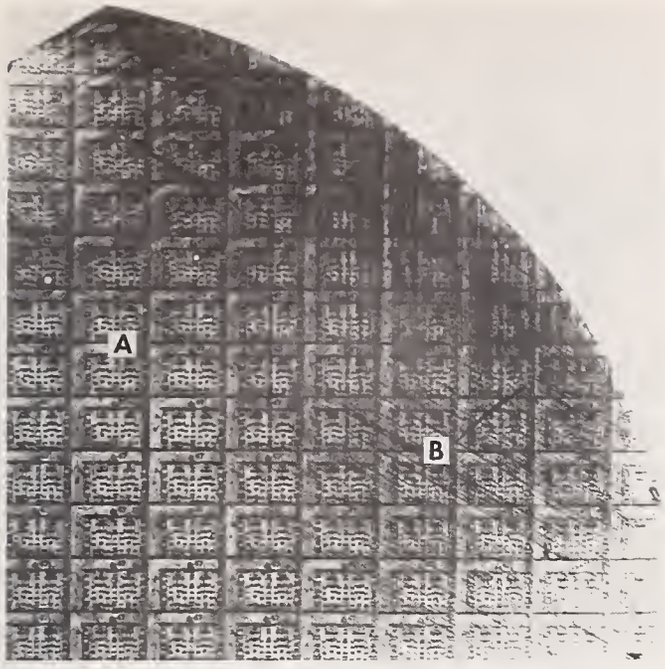


Fig. B.6. Transmission (SOT) x-ray topograph of a device-wafer section. Chip A is located in low dislocation interior of crystal, whereas B is near heavily slipped wafer edge.

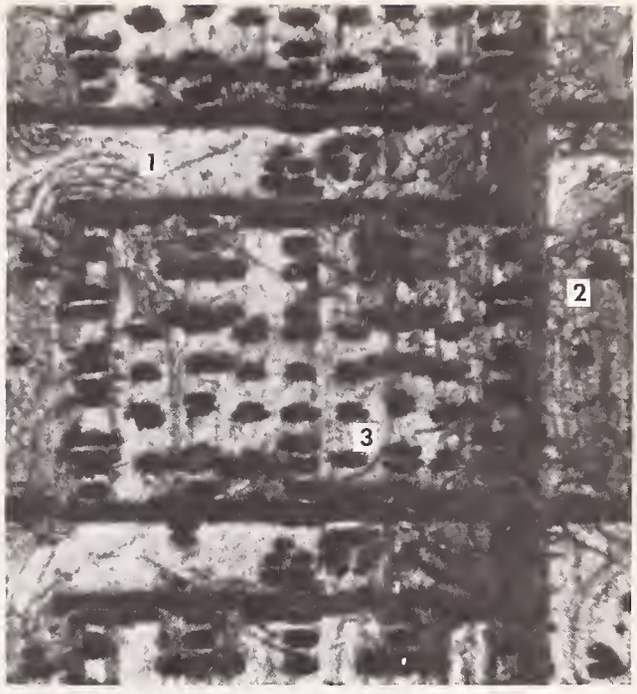


Fig. B.7. Enlargement of A in Fig. B.6. Note dislocation loops (1) which are located near edge of the isolation region and (2) which occur between adjacent chips.

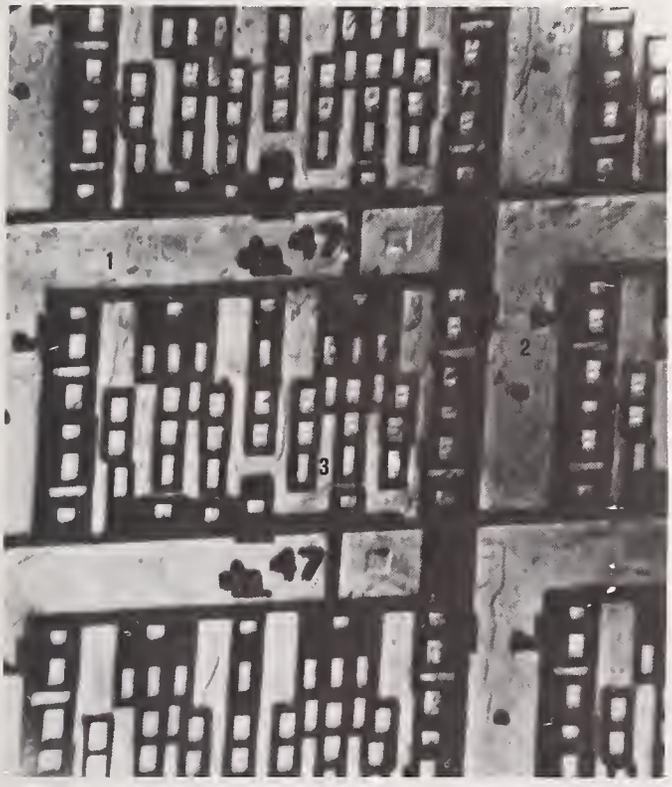


Fig. B.8. Reflection (surface) x-ray topograph of chips in Fig. B.7. Location of dislocation outcrops (1) and (2) indicates that loops are punched downward from observed location in surface topograph. Dislocation line (3) is identified from transmission image in Fig. B.7.

Fig. B.9. Enlargement of chip B in Fig. B.6. Heavy dislocation slip (superposition of substrate and epitaxial film defects) prevents any defect-device correlation.

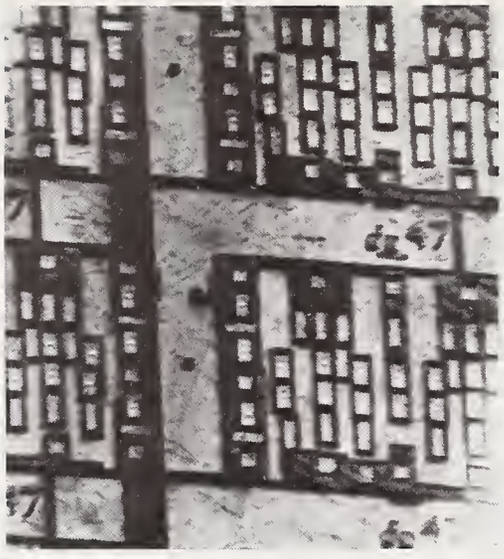


Fig. B.10. Surface x-ray topograph of region in Fig. B.9. Short line segments represent steep-inclined dislocation segments (slip array), which intercept the surface. Note disparity in volume and surface defect density transmission (volume) topograph can exaggerate surface defect density.

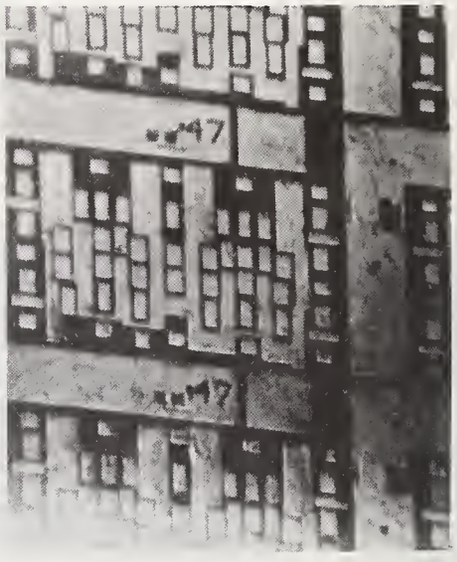


Fig. B.11. Surface topograph of another device-wafer section. Kerf transistor at the top yielded a leakage current of 240 μA as compared to 1300 μA for the lower chip, being consistent with observed dislocation density in each chip.

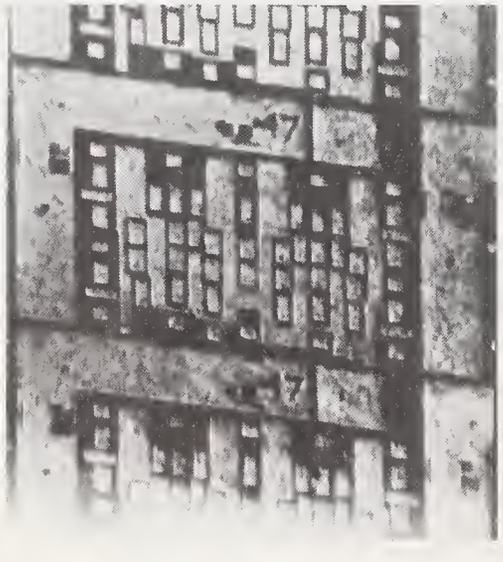


Fig. B.12. Surface x-ray topograph of a chip near wafer edge. Top Kerf transistor yielded a leakage current of 2000 μA as compared to 4000 μA for lower Kerf.

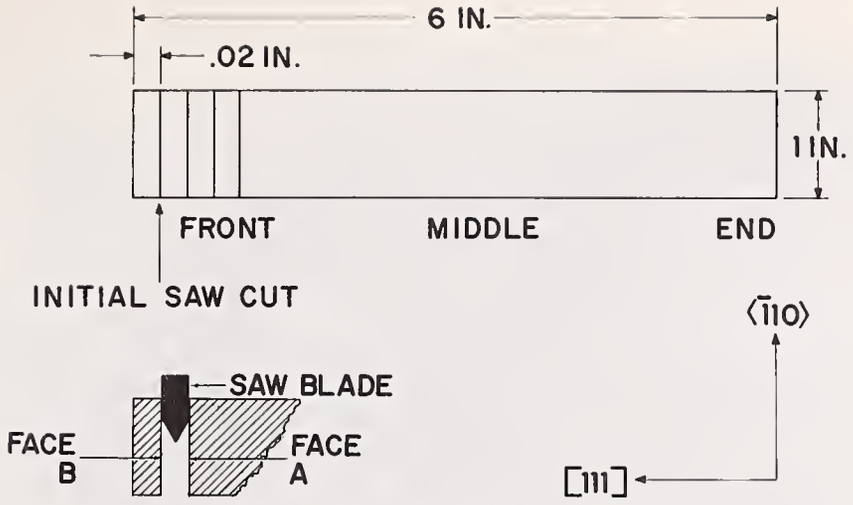


Fig. C.1. Method of crystal slice identification.

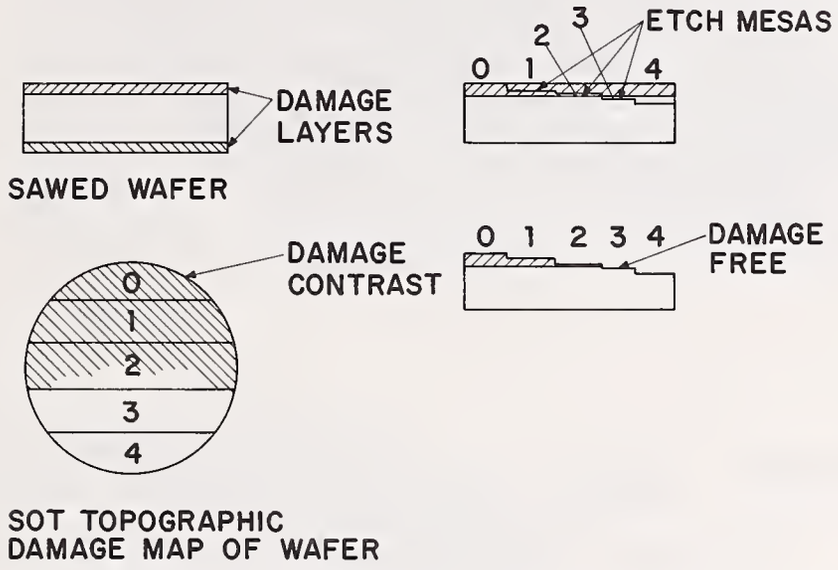


Fig. C.2. Depth of distribution of damage through SOT topography.

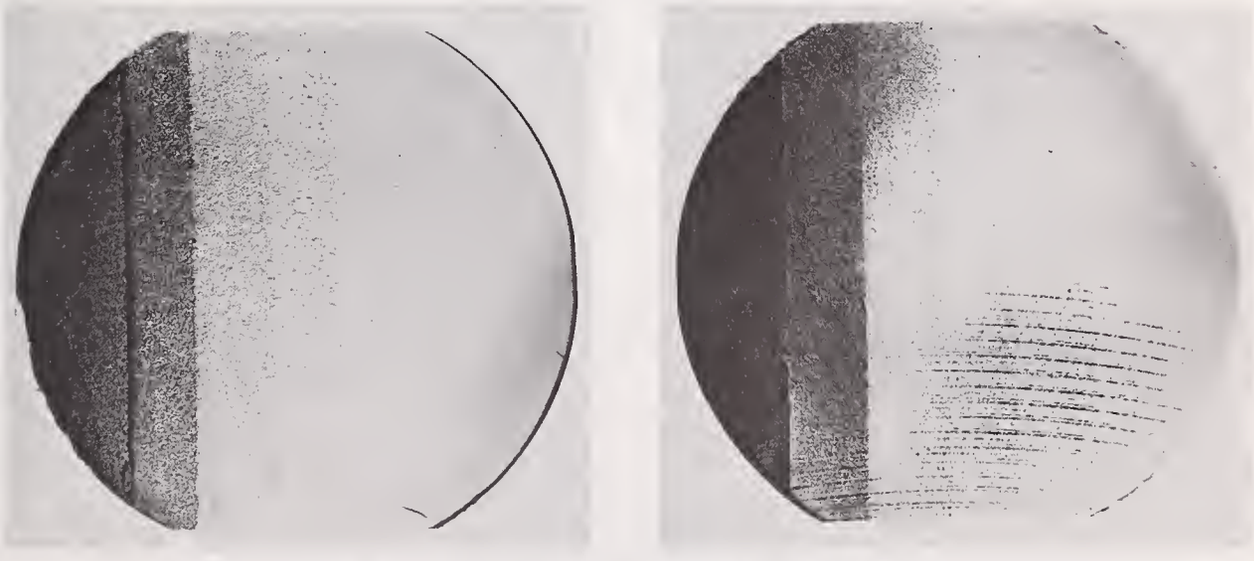


Fig. C.3. SOT transmission topograph showing damage contrast.

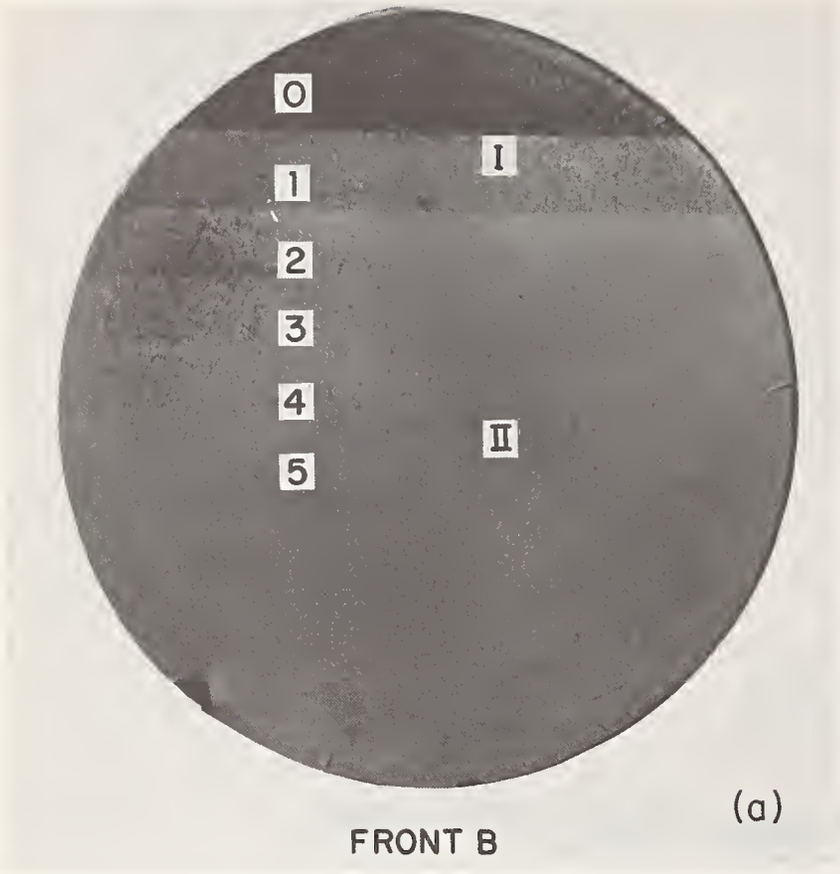


Fig. C.4a. SOT topograph of saw damage in the B surface of wafer cut from front of the crystal. Note contrast reduction after initial step. Damage zones (I) and (II) represent two different modes of saw damage. (See text.)

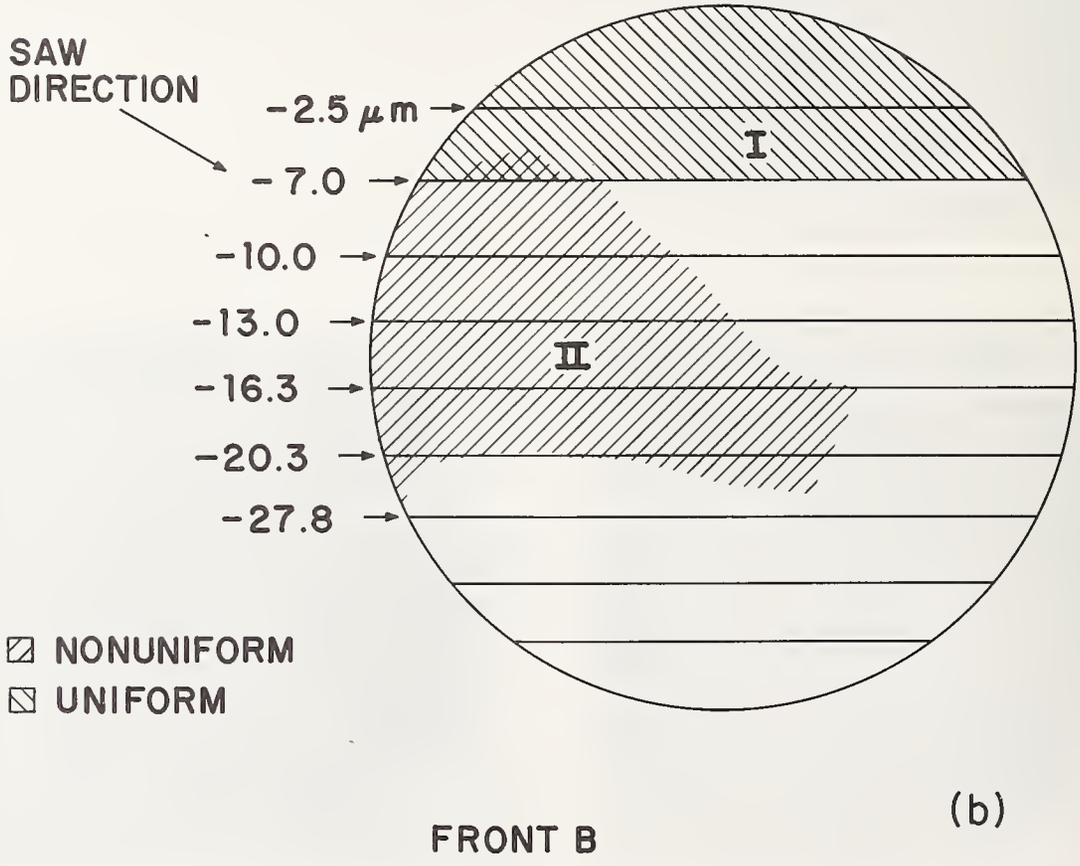


Fig. C.4b. Damage map, front B. Depth of damage indicated after each step is average of ten measurements at different locations along step.

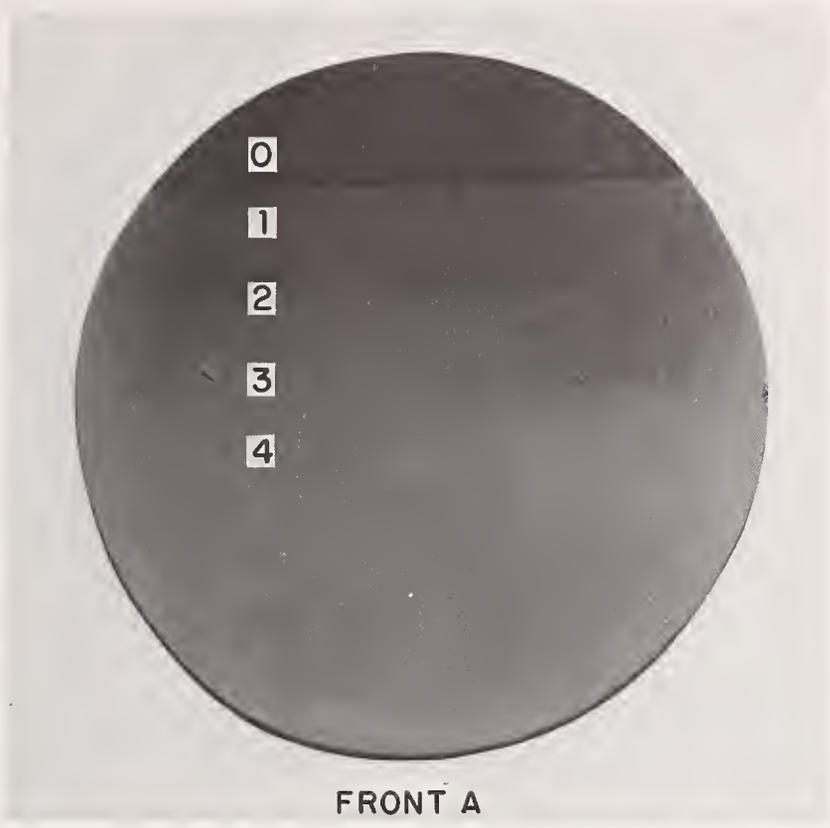


Fig. C.5a. SOT topograph of A surface damage (front), reduction in contrast after each step represents a direct measure of distribution of damage below surface.

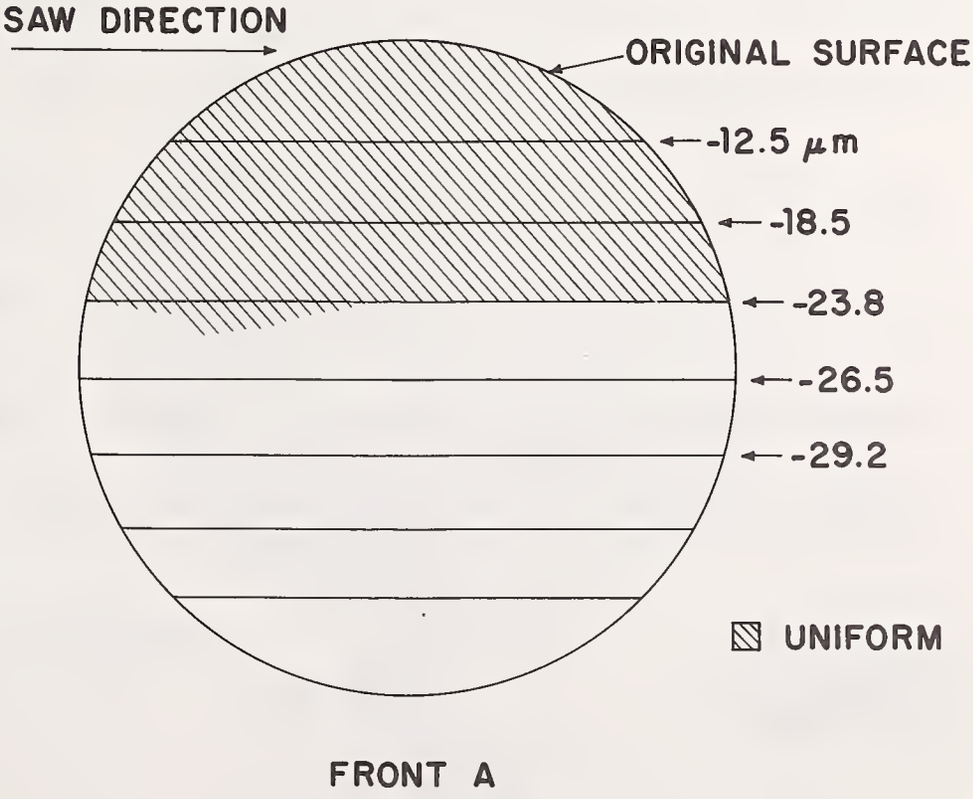


Fig. C.5b. Damage map, front A.

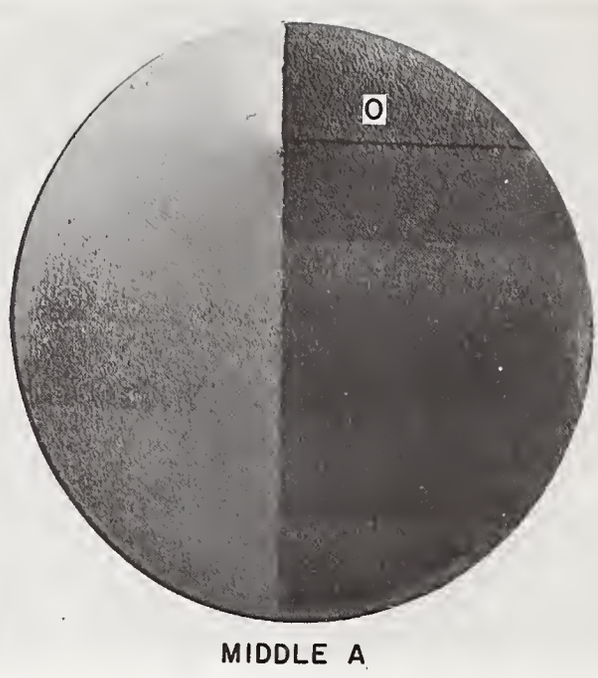


Fig. C.6a. SOT topograph of saw damage in A surface (middle).

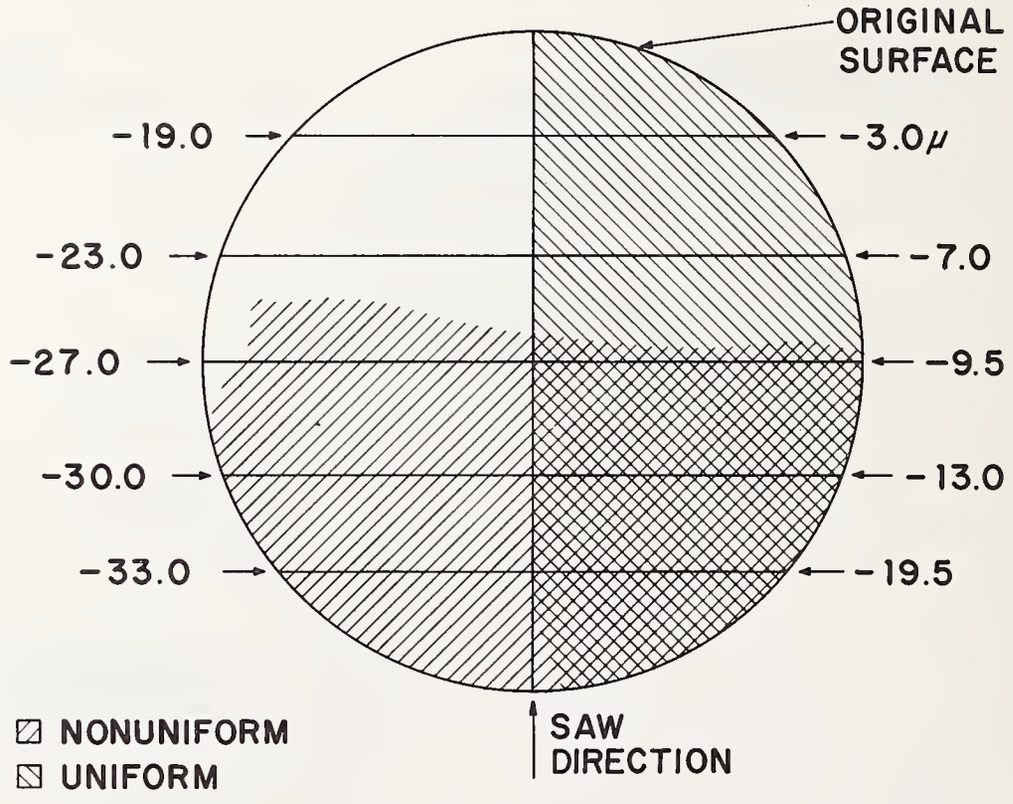


Fig. C.6b. Damage map, middle A.

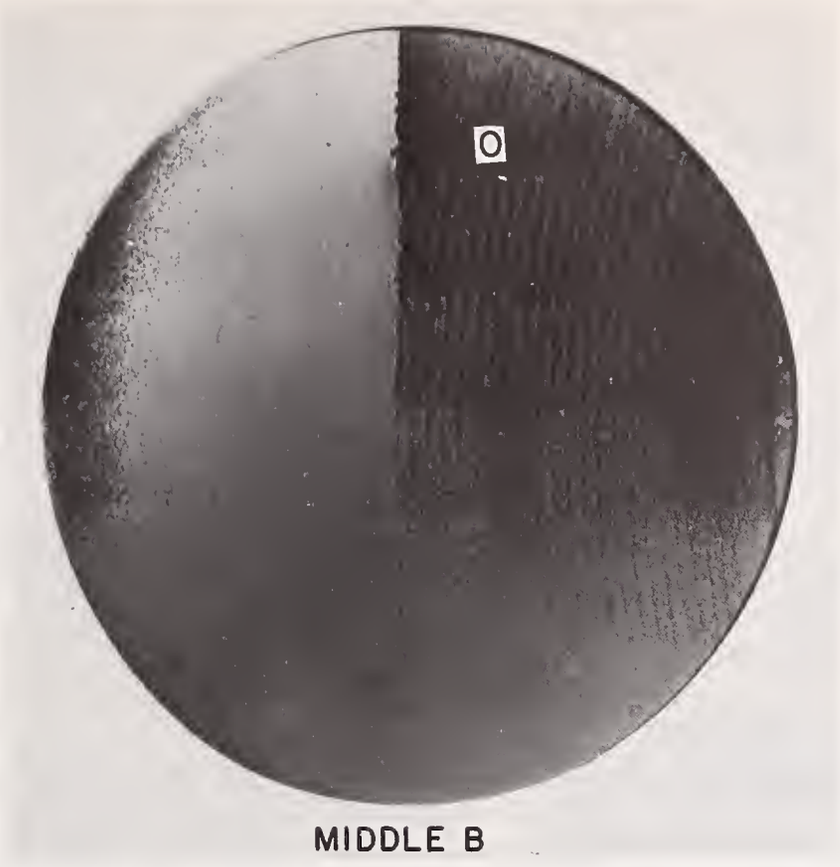


Fig. C.7a. SOT topograph of saw damage in B surface (middle).

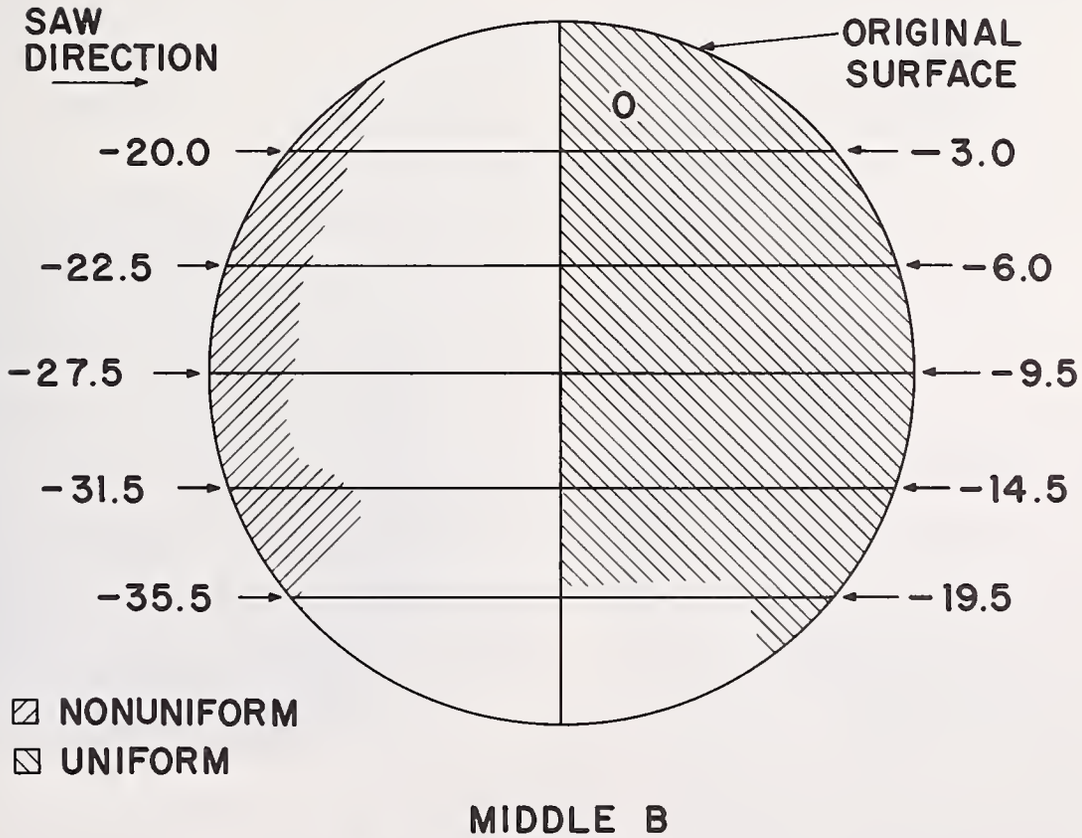


Fig. C.7b. Damage map, middle B.

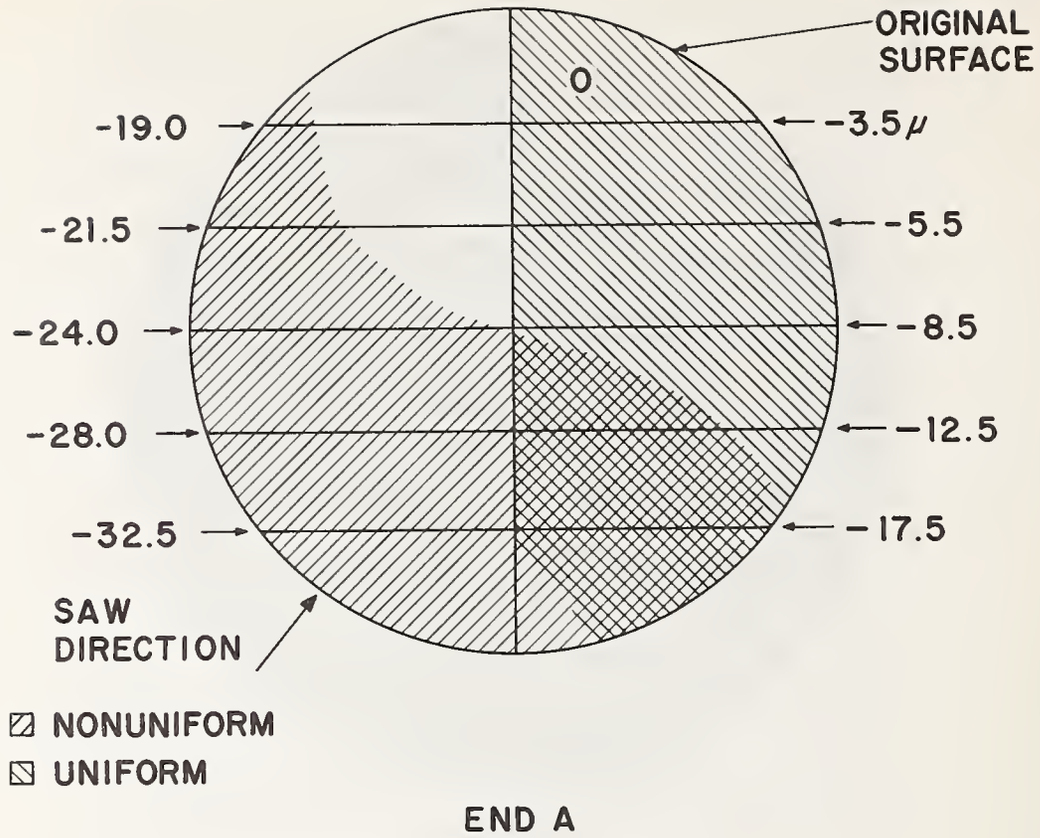


Fig. C.8. Damage map, end A.

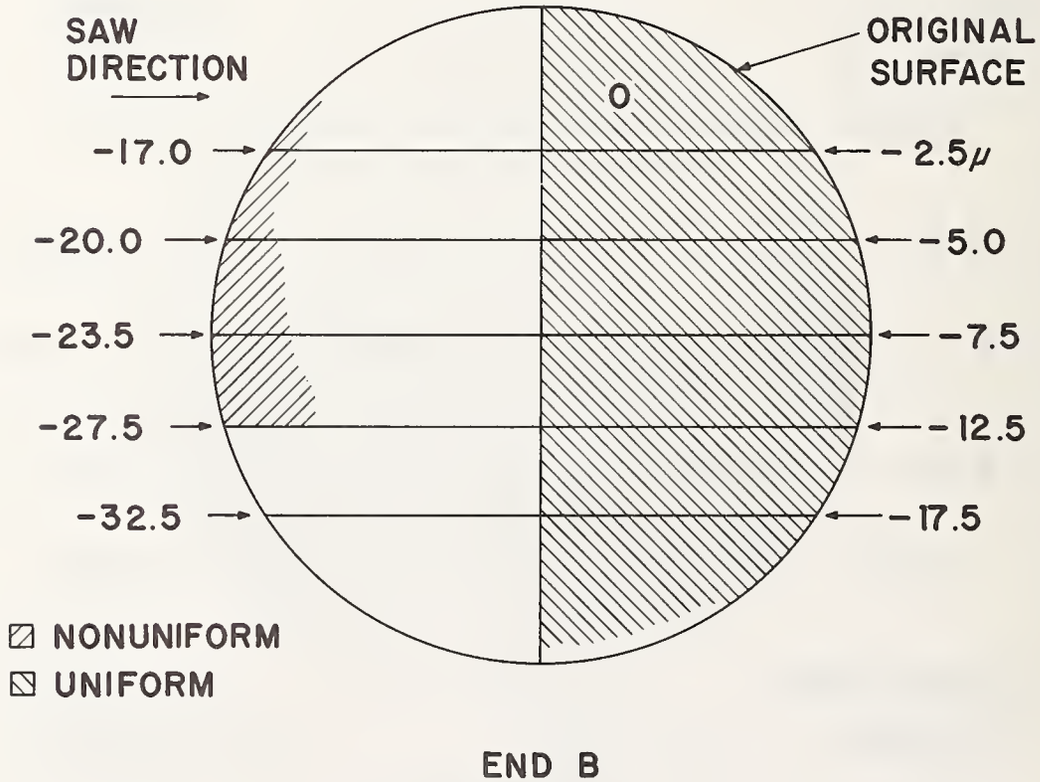


Fig. C.9. Damage map, end B.

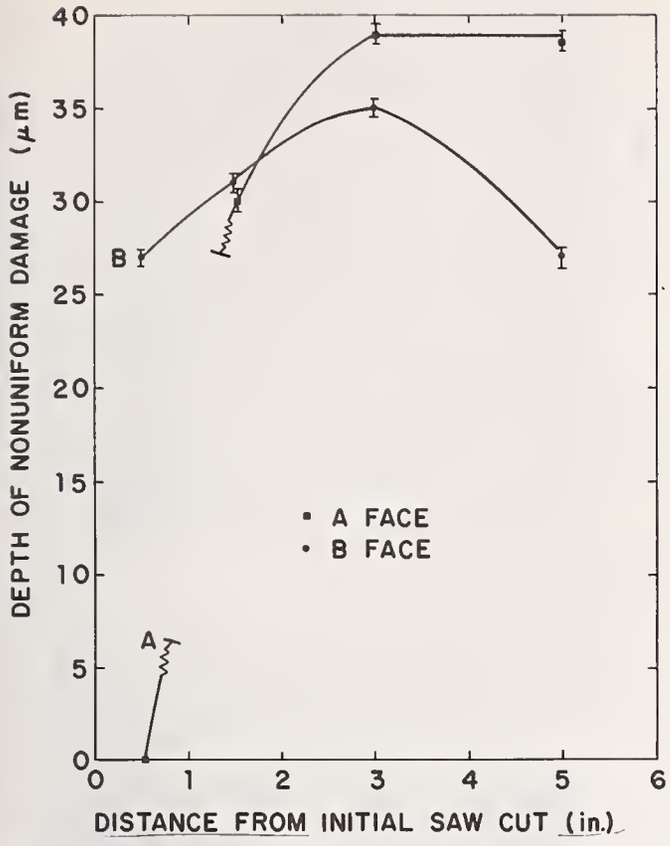


Fig. C.10. Damage profiles showing nonuniform damage depth as a function of wafer face and location.

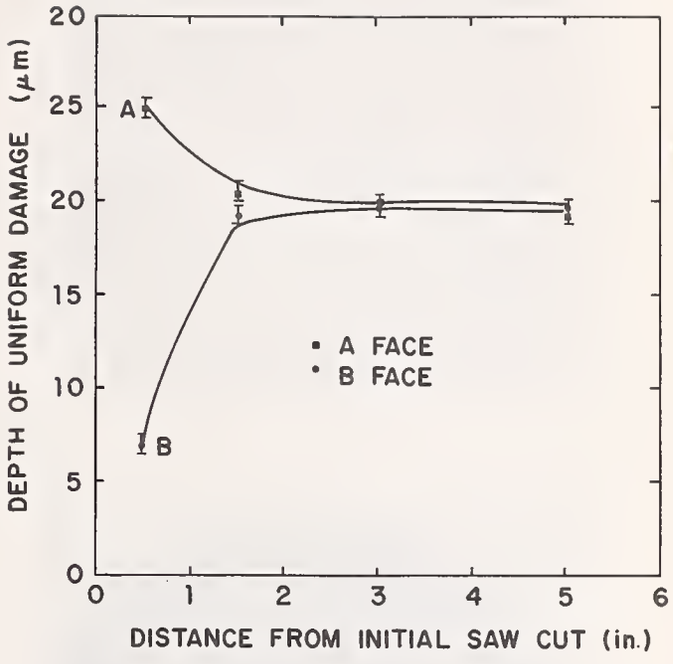


Fig. C.11. Damage profiles showing uniform damage depth as a function of wafer face and location.

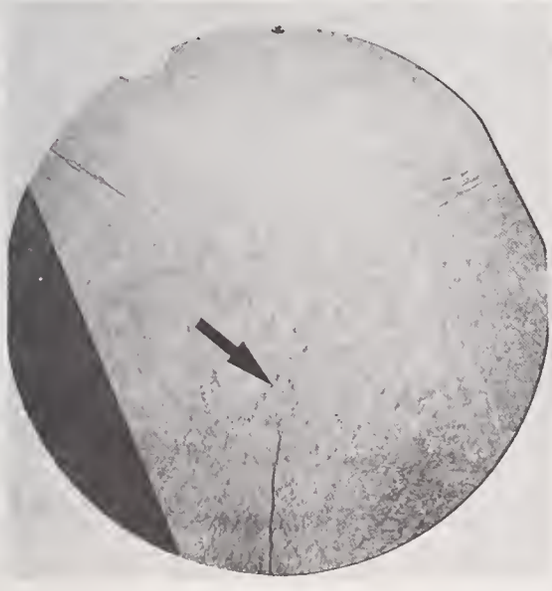


Fig. C.12. Transmission SOT topograph of wafer cut at feed-rate of 1.5 in/min. Note striations (arrow) and fact that damage contrast is absent after first step.

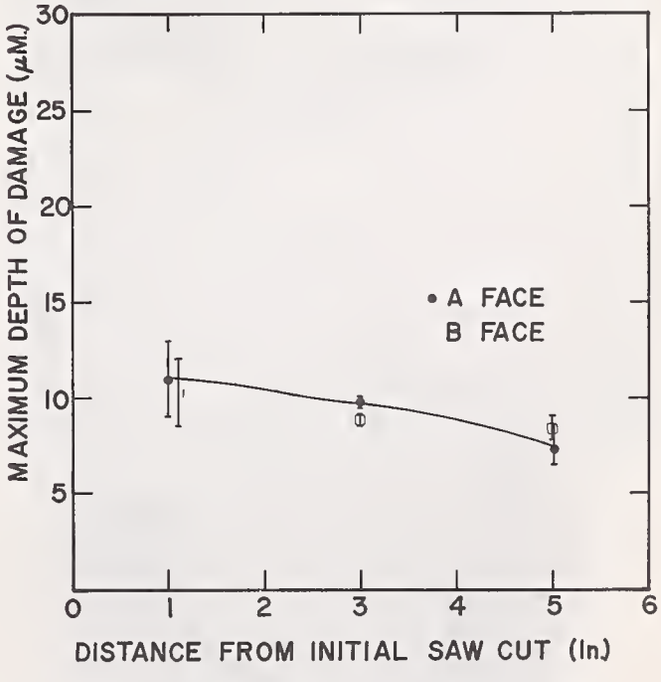


Fig. C.13. Damage profile of crystal cut with 400 grit blade.

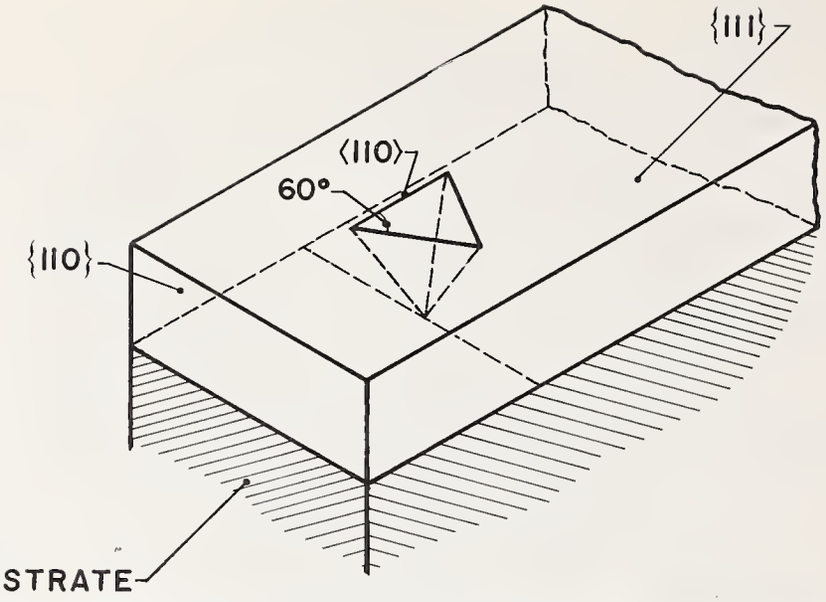


Fig. D.1. Geometry of stacking faults relative to the substrate.

Fig. D.2a. SOT topograph of damage contrast due to mechanical polishing

Fig. D.2b. Epitaxial film over polished region is highly defective as evidenced by strong diffraction contrast.

Fig. D.2c. SOT topograph after chemically sectioning through epitaxial film; degree of crystal deformation is indicated by reduction in contrast.

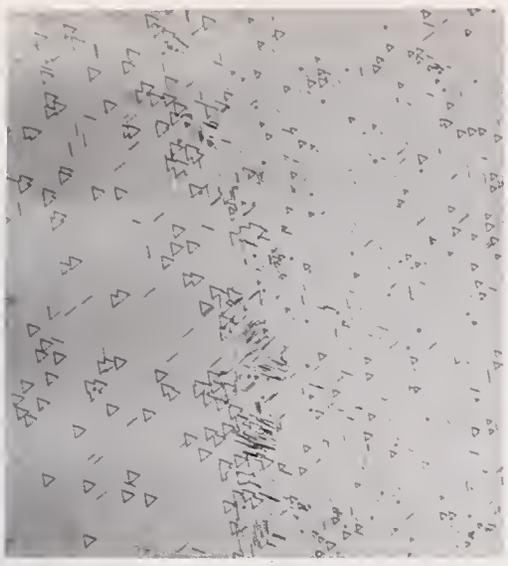


Fig. D.3a. Optical photograph of fault structure across damaged (upper)-undamaged interface.

Fig. D.3b. Relative difference in dislocation density is revealed after a 1-min Sirtl etch.

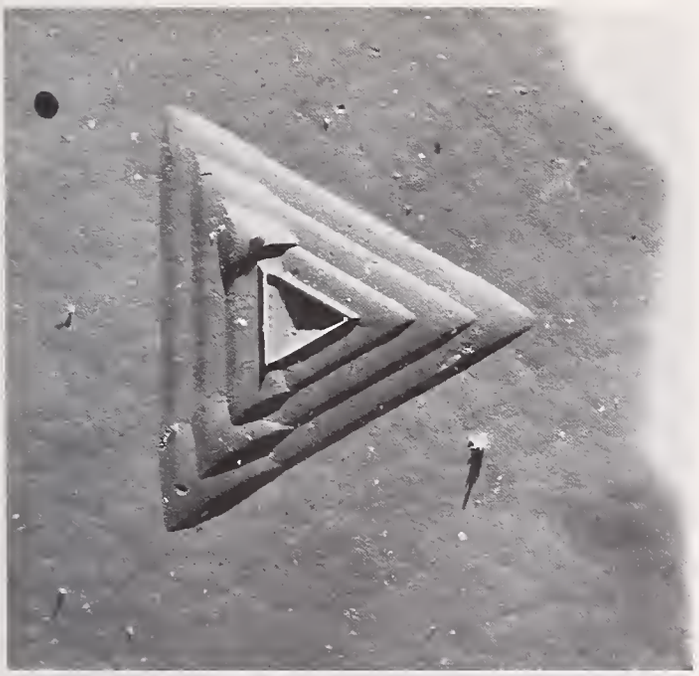


Fig. D.4. Replication electron micrograph of a stacking fault.

Fig. D.5. Replication micrograph showing a linear stacking fault terminated by two dislocation-etch pits.

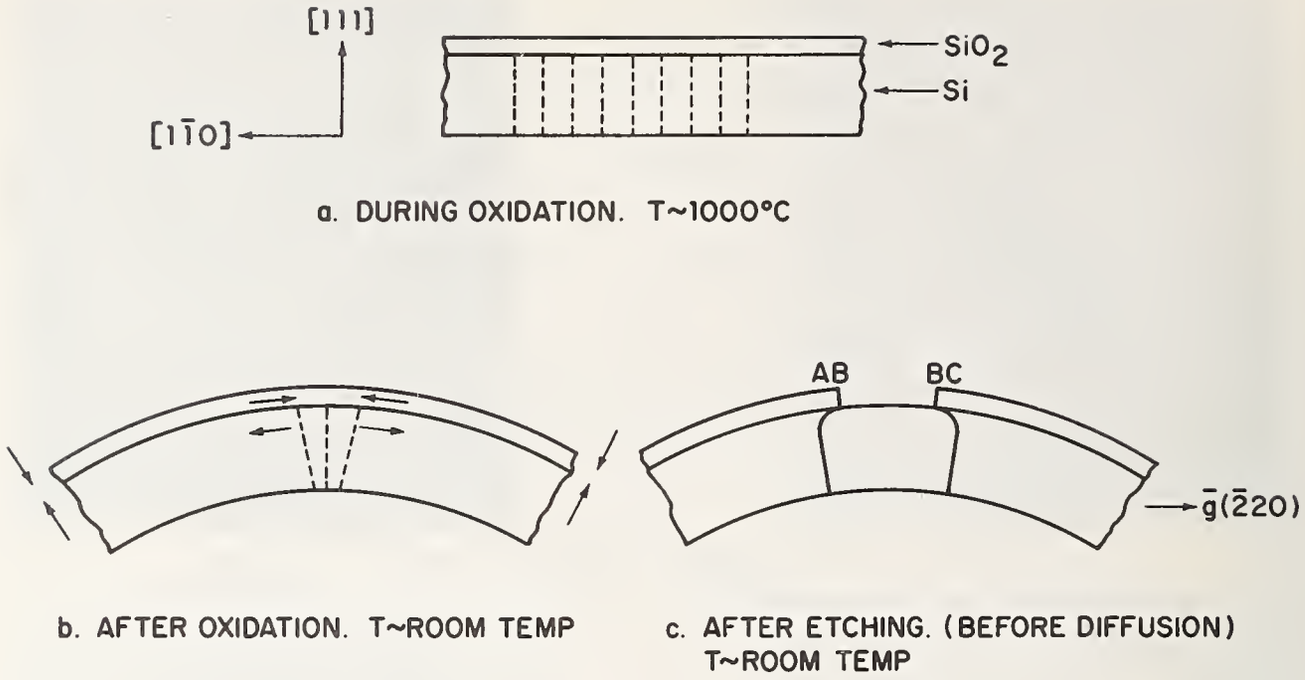


Fig. E.1. Exaggerated view of crystal deformation near window edge.

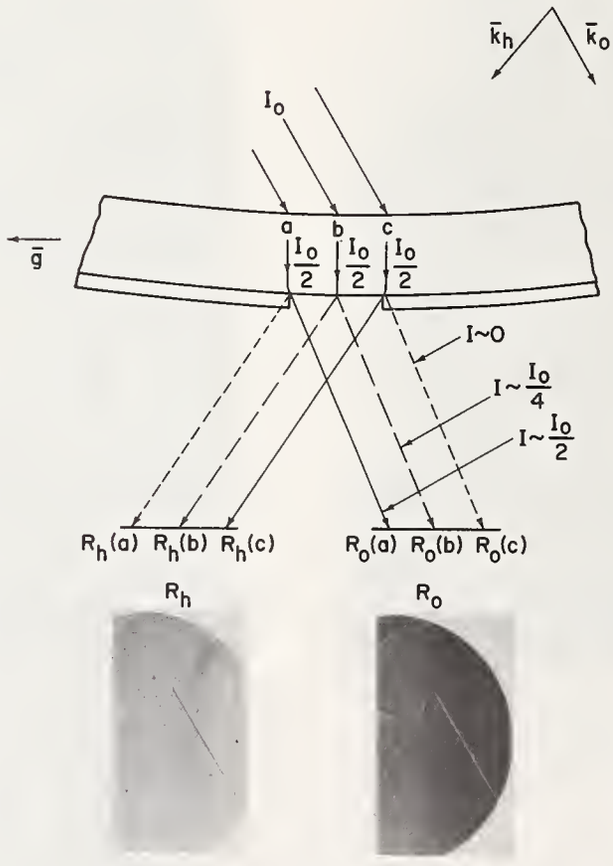


Fig. E.2. Schematic of x-ray intensity alteration at triangular window cut into an oxide film (SiO_2/Si). Corresponding x-ray topographs indicate that simple curvature model[2] is satisfying.

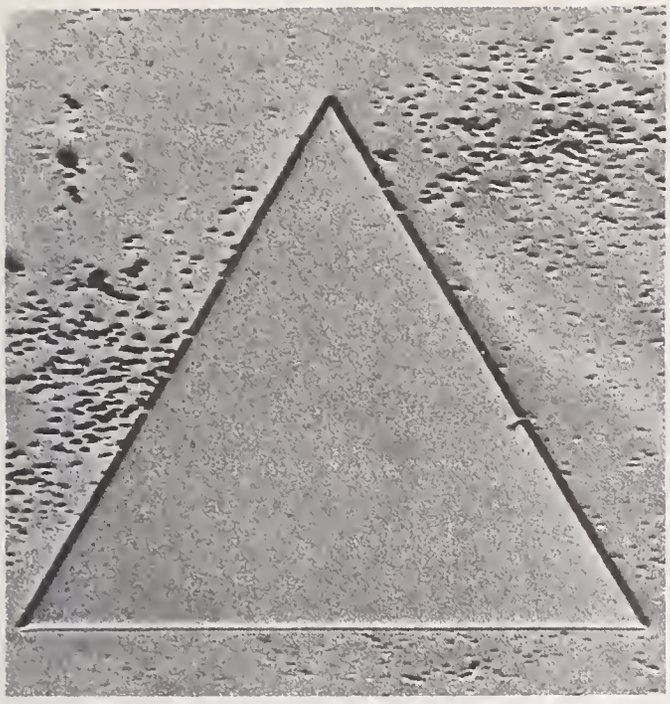


Fig. E.3. SOT topograph of triangular window in metal film. Note the black-white contrast at local regions of adhesion loss.

Fig. E.4. Example of contrast lobes due to poor adhesion between metal and SiO_2.

Fig. E.5. SOT topograph of oxidized Si sample after Al evaporation (circular region 0); the lack of contrast lobes in the film region indicates good adhesion.

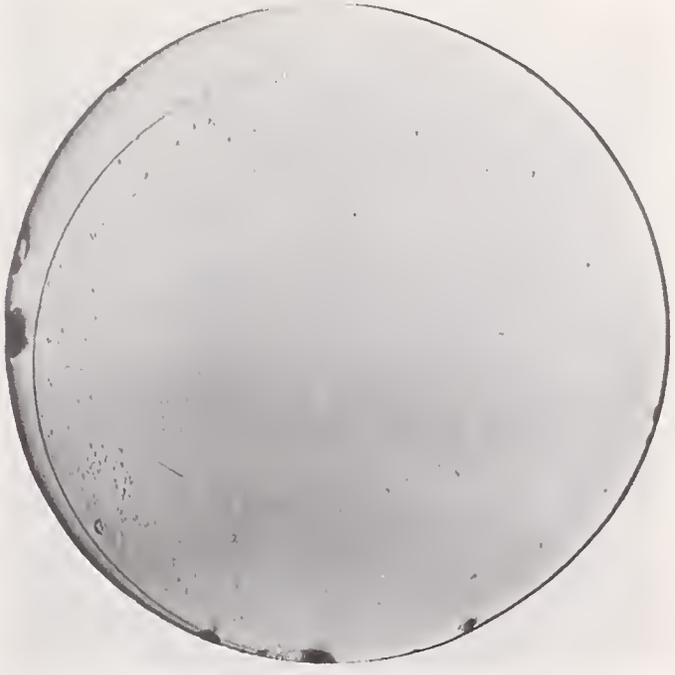


Fig. E.6. Topograph of similar wafer; the contrast images at the film periphery indicate local adhesion loss.

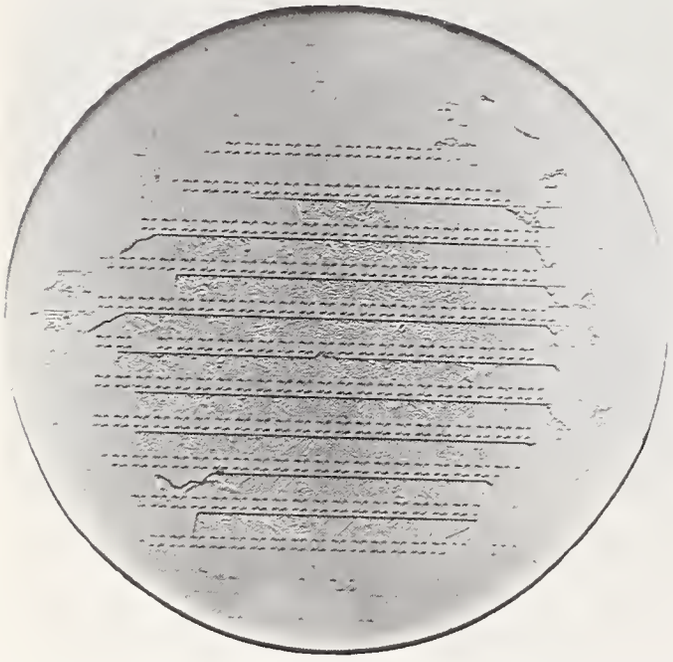


Fig. E.7. The poor adhesion at the wafer periphery results in local film loss and degradation during photoresist operation.

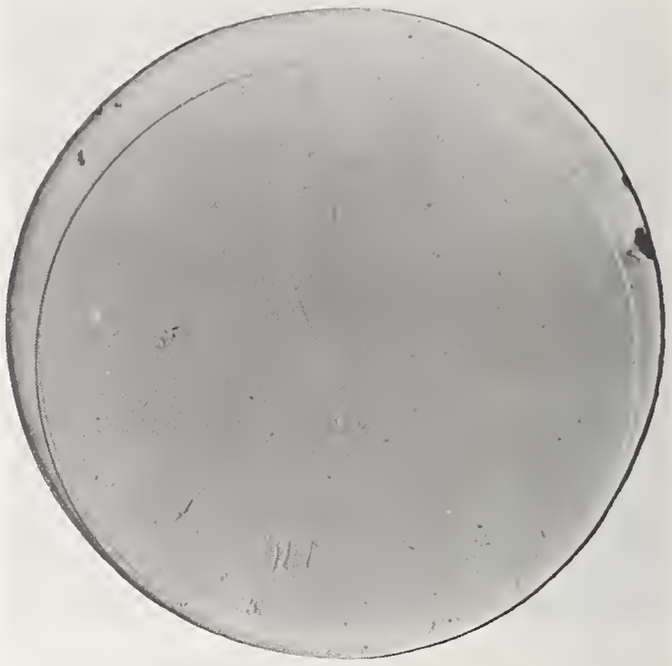


Fig. E.8. Topograph after Al deposition. Note several contrast lobes near edge of film.

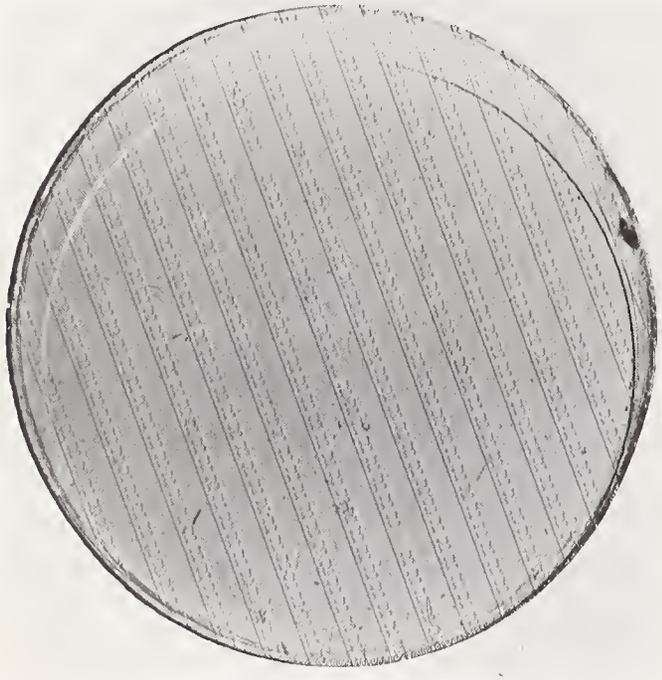


Fig. E.9. The same sample is examined after the photoresist is exposed and developed.

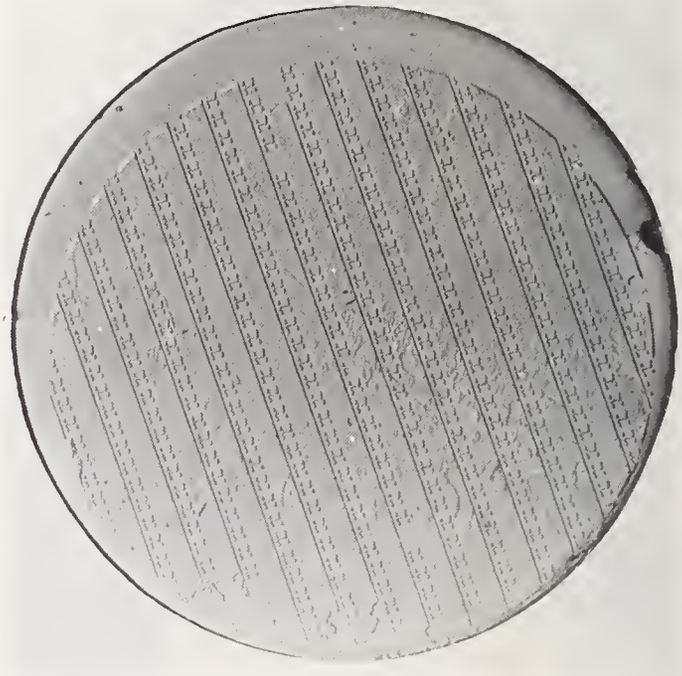


Fig. E.10. After etching and cleaning, film is removed in same region of contrast lobes (Fig. E.8.) and degraded in film center.

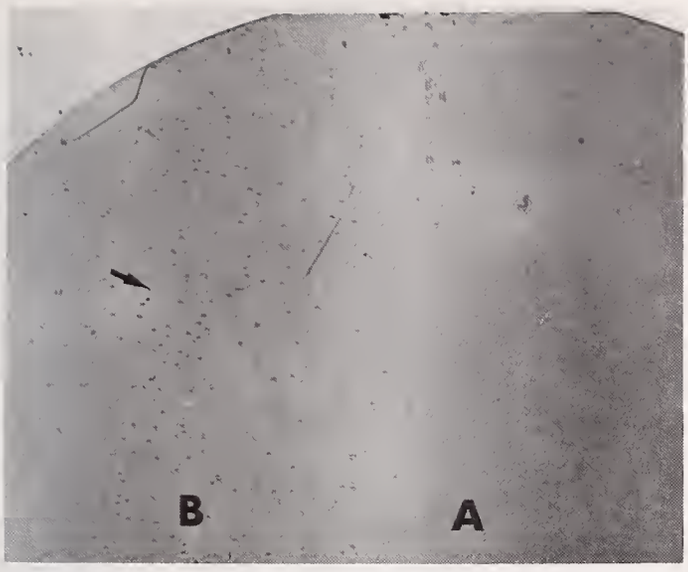


Fig. E.11. SOT of EF-9 after blanket metallization reveals characteristic adhesion contrast (arrow) on half-wafer that was contaminated prior to deposition. Note that half-wafer, which was properly cleaned (A), is free of this type of adhesion fault.

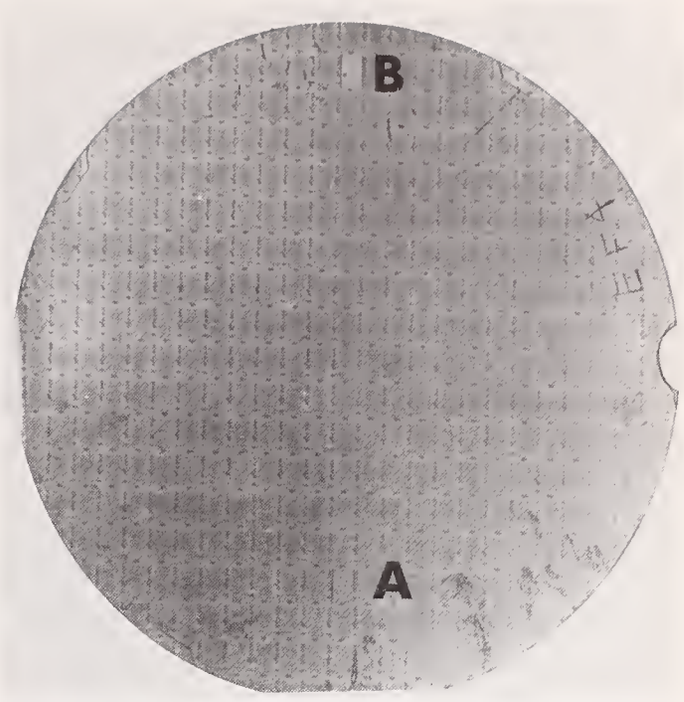


Fig. E.12. SOT of EF-4 reveals an array of test sites after sub-etch. Contaminated region of wafer alignment block (A) and bypassed (B).

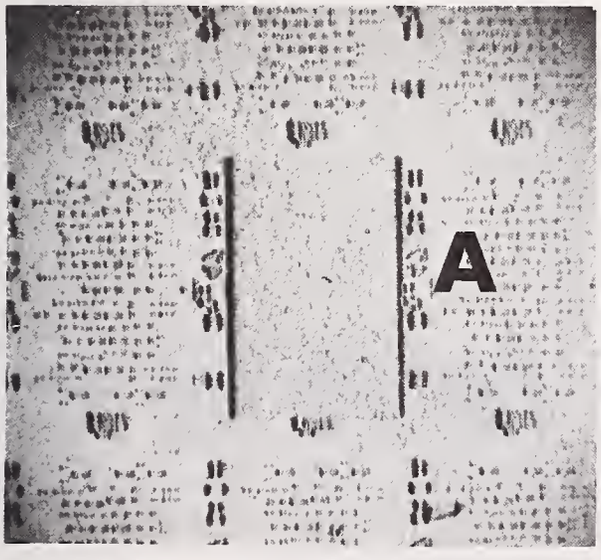


Fig. E.13a. Enlargement of block (A) which displays the adhesion contrast after deposition.

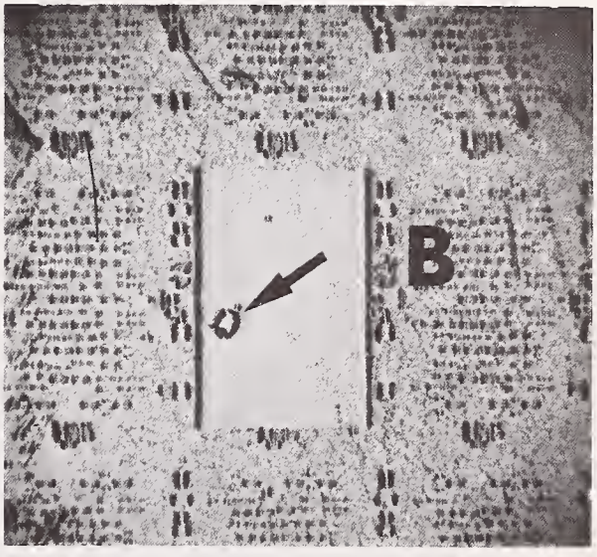


Fig. E.13b. Alignment block (B) is free of adhesion contrast; dark region (arrow) in block is a pin-hole (optically visible).

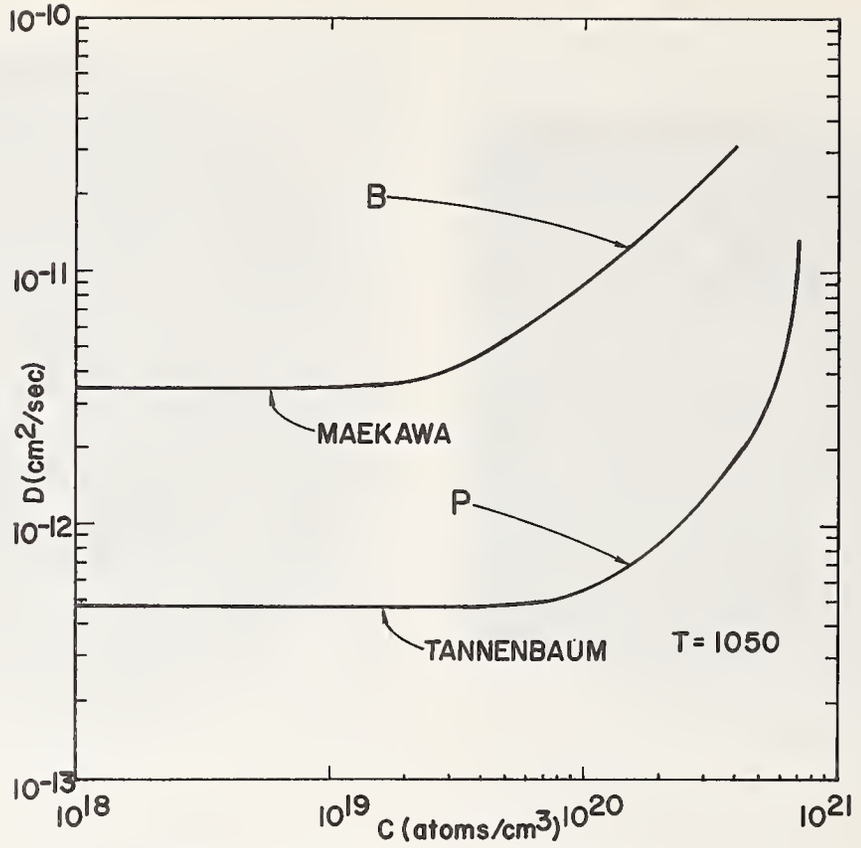


Fig. F.1. Concentration dependence of B and P diffusivity as observed by Maekawa and Tannenbaum, respectively.

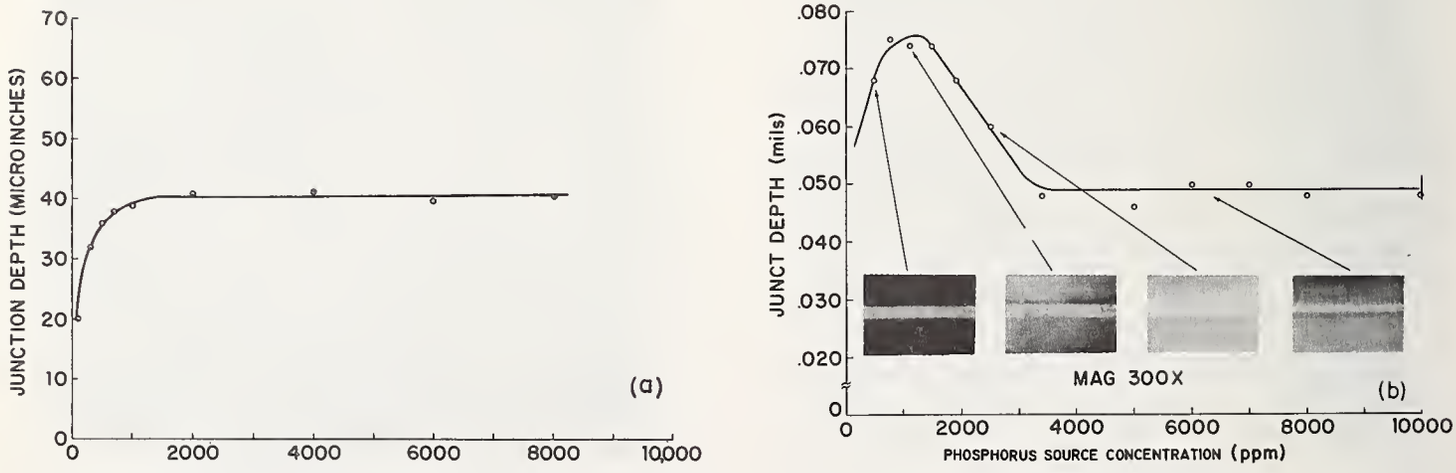


Fig. F.2. Dependence of junction depths in open-tube P-diffusion systems. [83]
a. Junction depth vs phosphorus concentration after the deposition cycle only.
b. Junction depth vs phosphorus concentration after both the deposition and drive-in-cycle. Photo inserts illustrate the junction depth from beveling and staining at representative points.

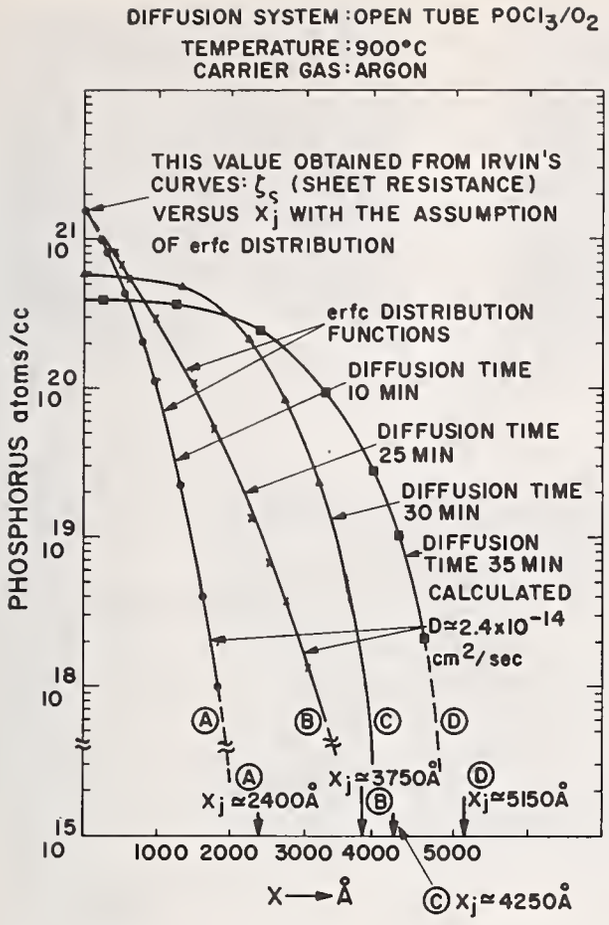


Fig. F.3. Phosphorus distribution obtained through open-tube diffusion processes. Observe the reduction of C_s after exceeding $x_j \simeq 3750$ Å.

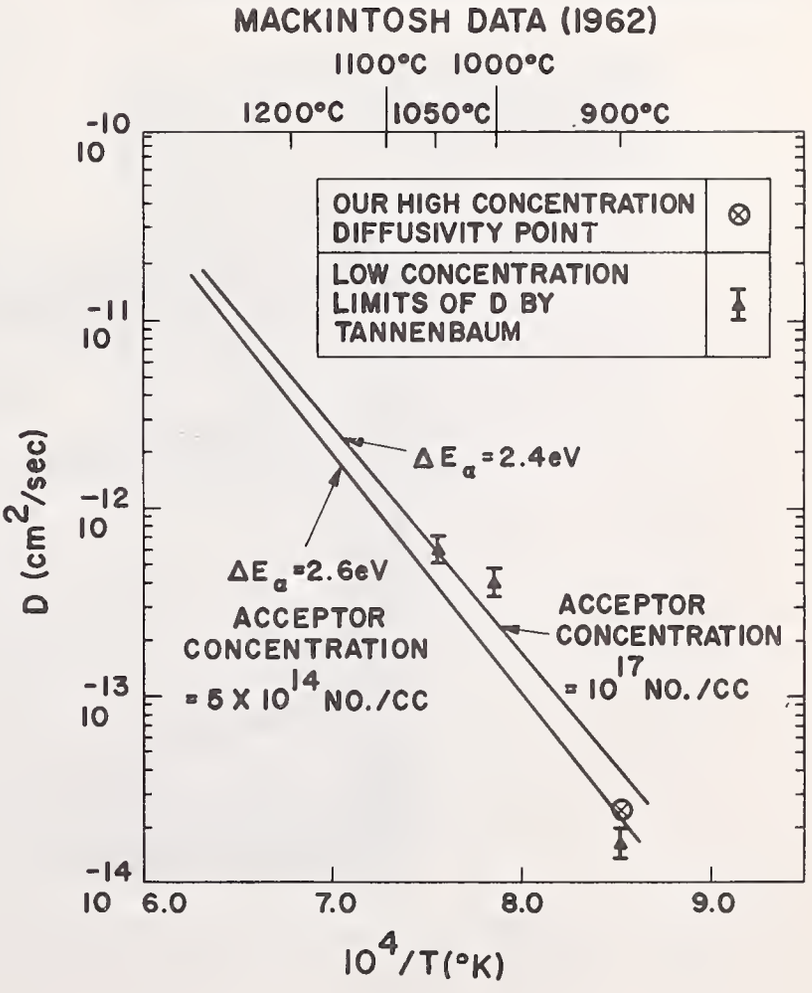


Fig. F.4. Phosphorus diffusivity as a function of reciprocal absolute temperature.

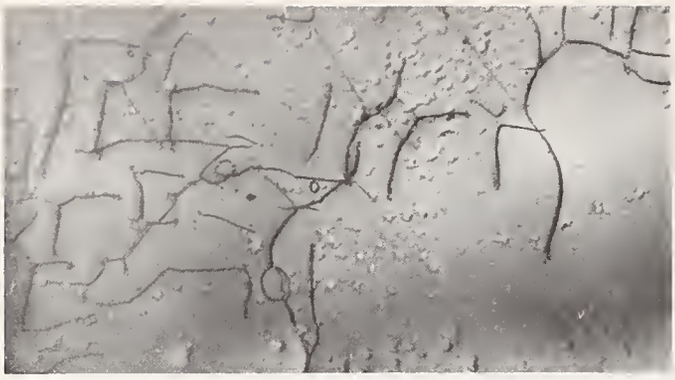


Fig. F.5a. Initial stage of dislocation-loop interactions resulting in formation of misfit-dislocation nets. Samples correspond with C-curve diffusion given in Fig. F.3.

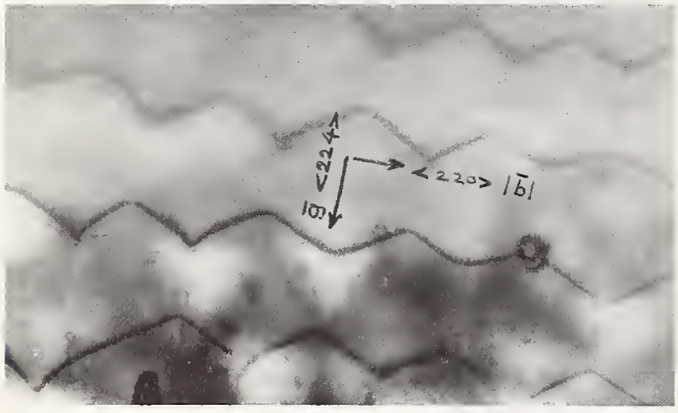


Fig. F.5b. Misfit-dislocation net with one set of vanishing dislocations for $\bar{g}$ = [220]. The Burgers vectors of these dislocations are obviously a/2 [110] type and are lying in plane of net.

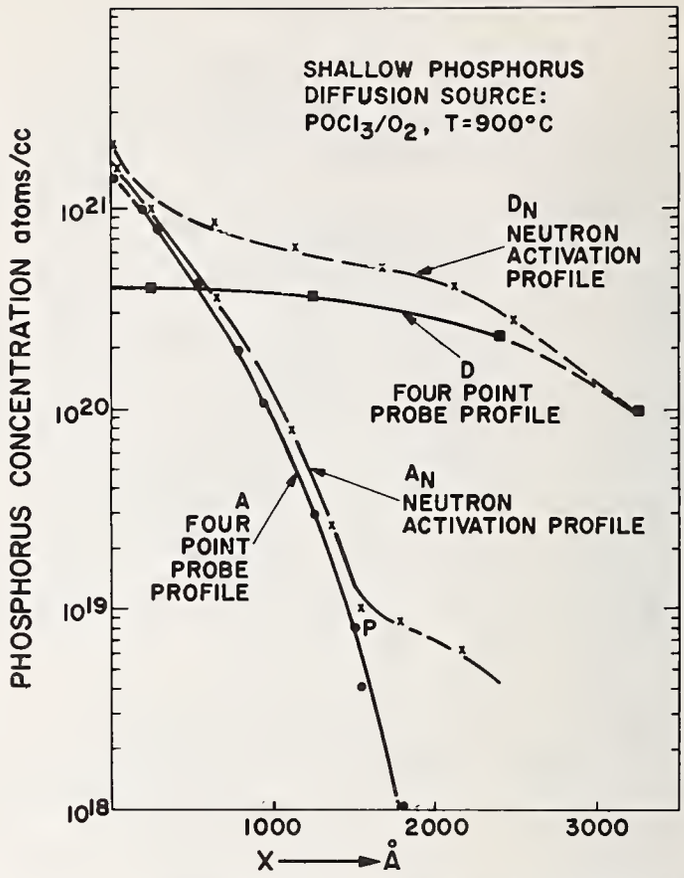


Fig. F.6. Comparison of phosphorus distributions obtained by the techniques of neutron activation and four-point probe. Observe the increased difference in the distributions obtained by the two techniques for the deeper diffusion.

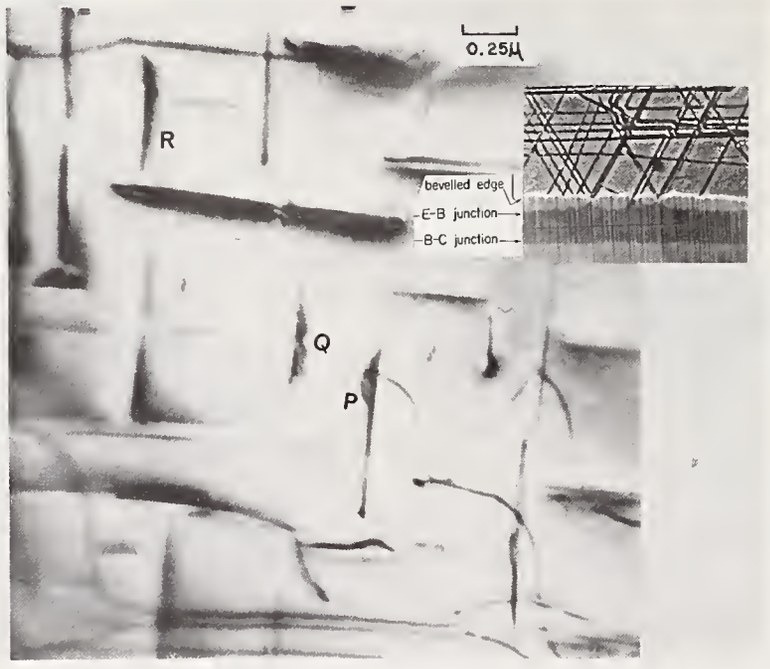


Fig. F.7. Observe diffusion-induced dislocations decorated with phosphorus at P, Q, and R. Right-corner insert in figure shows emitter-base junction (E-B) is seen retarded at every point under the dislocations. Insert, courtesy of Yoshida and Kanamori.

Fig. F.8. Cooperative diffusion effects. Retardation by diffusing, first P; then B.

Fig. F.9a. Cooperative diffusion effects. Push-out by diffusing, first Sn; then B.

Fig. F.9b. Cooperative diffusion effects. Push-out by diffusing, first Ge; then B.

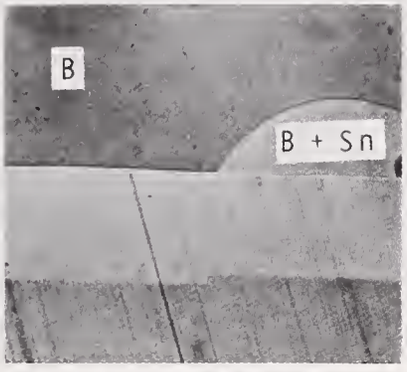


Fig. 10a. Cooperative diffusion effects. Retardation by diffusing, first B; then Sn.

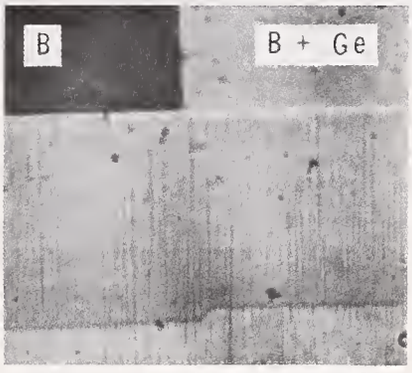


Fig. 10b. Cooperative diffusion effects. Retardation by diffusing, first B; then Ge.

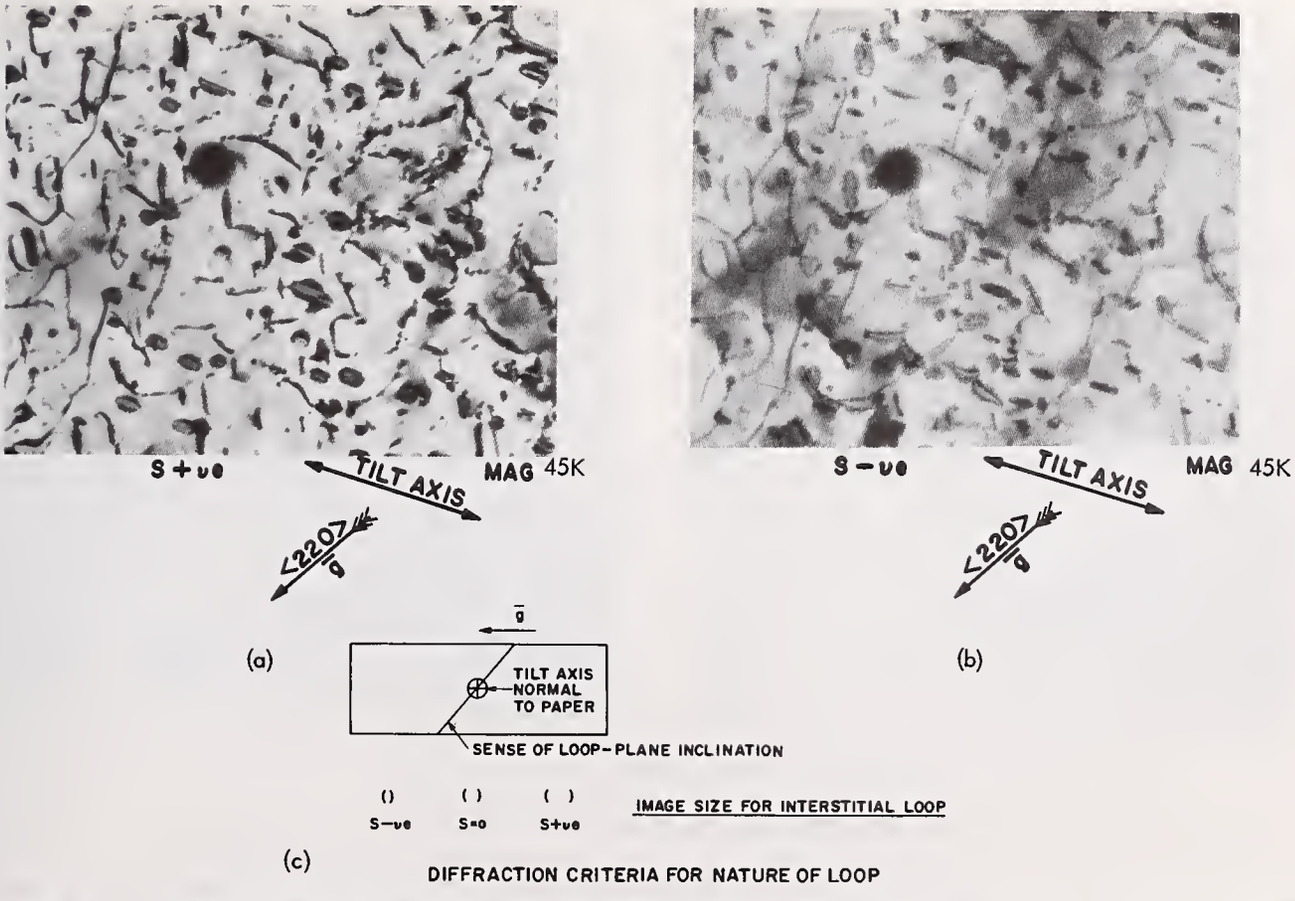


Fig. 11. Diffraction contrast experiments for demonstrating the interstitial nature of dislocation loops.

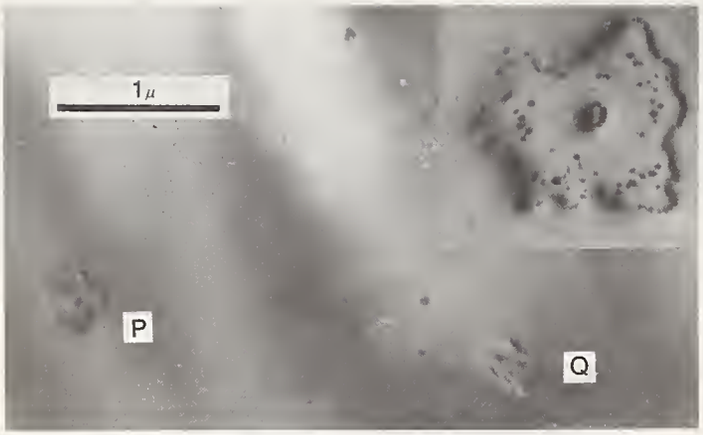


Fig. G.1. Observe Si_xO_y precipitate clusters P and Q in oxygen-bearing silicon after 1000 °C/ neutral ambient heat-treatment. See that cluster fields contain small sessile dislocation loops. Right-corner insert is a magnified view of cluster P.

Fig. G.2. Observe dislocation generation around Si_xO_y complexes. Dislocations were found purely edge-type. TEM observation, courtesy of Patrick and Dash.

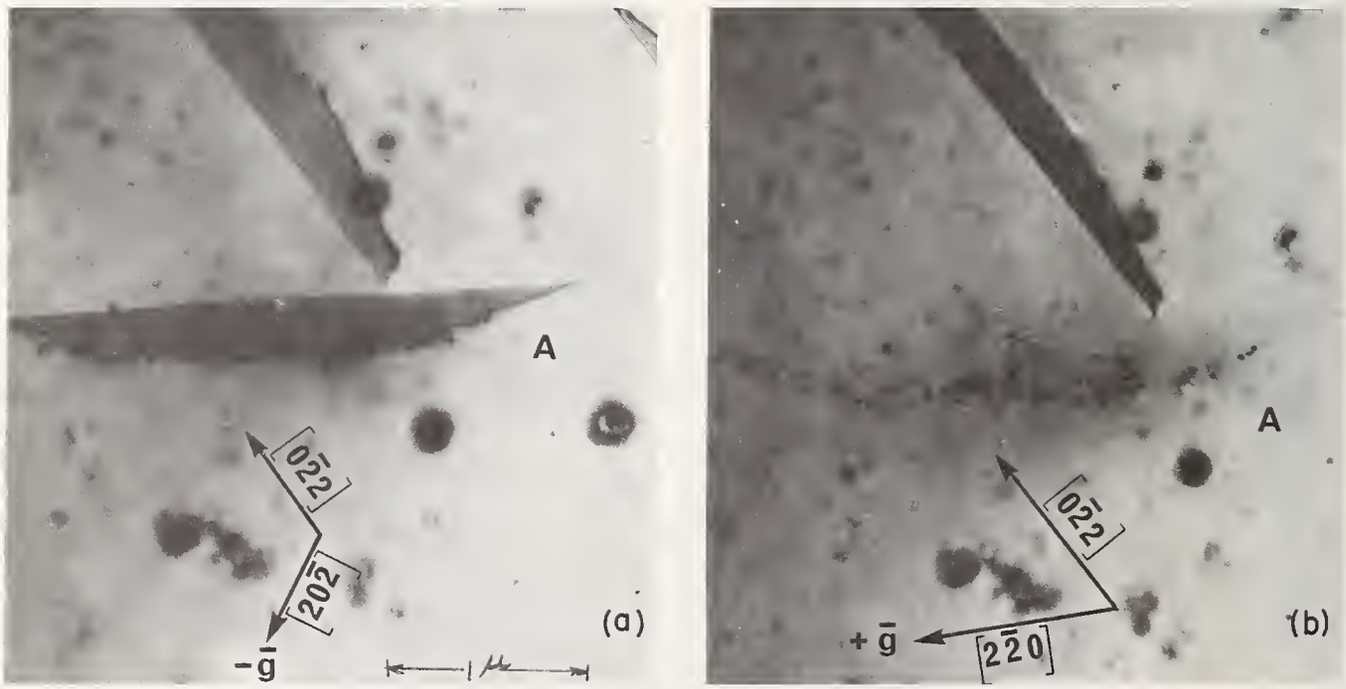


Fig. G.3. Stacking faults in steam-oxidized silicon.
a) Observe precipitates decorating fault A.
b) Observe only precipitates in region of vanishing fault.

A Study Relating MOS Processes to a Model of the Al-SiO_2-Si System

Marvin H. White, Franklyn C. Blaha, and David S. Herman[1]

Westinghouse Electric Corporation
Baltimore, Maryland 21203

A model of the Al-SiO_2-Si system is presented to aid in the control and understanding of integrated circuit fabrication. The model assumes ionic contamination is distributed through a thin phosphosilicate glass under the Al electrode. The ionic contamination is composed of alkali and hydronium ions which result from the unit processes. The level and type of ionic contamination is inferred from capacitance-voltage (C-V) measurements and the interface state density is monitored by capacitance-time (C-t) measurements. The intrinsic surface-state charge Q_{ss} is less than $10^{10}/cm^2$ in the process sequence described in this paper. The surface-state charge densities greater than $10^{10}/cm^2$ have been attributed to the retention of the fluoride ion on the silicon surface.

Key Words: aluminum, capacitance-time, capacitance-voltage, fluoride ion, ionic contamination, model, phosphorous glass, polarization, silicon, silicon dioxide, surface states, unit processes.

1. Introduction

The Al-SiO_2-Si system is the basic process sequence for present-day integrated circuit technology, although other metalization [1] and dielectric processes [2] have been employed. A thorough understanding and control of this system provides the foundation for more sophisticated devices and circuits on a silicon surface. In this paper we present a model of the Al-SiO_2-Si system based upon ionic contamination distributed through a thin phosphosilicate glass sandwiched between the Al and SiO_2. In addition to the model, we discuss the sources of ionic contamination and the interdependence of such unit processes as chemical cleaning, oxidation, diffusion, metalization and photoengraving. The level and type (e.g., alkali and hydronium) of ionic contamination are determined by capacitance-voltage (C-V) measurements [3] and temperature-bias stress (TBS) techniques [4]. The occurrence of interface states [5] is detected and monitored by C-V and capacitance-time (C-t) measurements [6]; the latter allows a determination of surface recombination velocity and bulk generation time of minority carriers.

2. Classification of Surface States in the Al-SiO_2-Si System

The general terminology "surface states" includes oxide charge and interface states. The original concept of interface states consisted of fast and slow states [8]. Fast states describe the states associated with the Si-SiO_2 interface through which charge exchange with the Si surface is possible (i.e., electrical communication with the semiconductor). Slow states, in the historical sense, applied to semiconductor surfaces with very thin oxides (e.g., less than 50 Å). These states were then and are today associated with the movement of surface ions (e.g., ions on the SiO_2 surface or at the Al-SiO_2 interface). Today, the term interface states is restricted usually to fast states since the integrated circuit devices have relative thick oxides.

Figure 1 illustrates the classification of surface states in the Al-SiO_2-Si system. The oxide charge is separated into several sub-categories: fixed charge, alkali and mobile ions, and ionized traps. The fixed oxide charge has been reported to lie within 200 Å of the Si-SiO_2 interface [9]. This fixed charge is dependent upon crystal orientation [e.g., largest for (111) and least for (100)] and independent of resistivity and type of silicon [3]. We have not observed this so-called intrinsic fixed charge in our studies on (100) silicon, although a residual fixed charge is observed when the fluoride ion is retained by chemical adsorption on the silicon surface [9-10]. This is a result of the tenacity and approximately 70% ionic character of the Si-F bond [11].

1 Fellow Engineer, Associate Engineer, and Fellow Engineer, respectively.
2 Figures in brackets indicate the literature references at the end of this paper.

The alkali ion contamination has been attributed to the presence of sodium ions [12] while the mobile ion contamination has been traced to the proton species, such as the hydronium ion [13]. The activation energy of the sodium ion is approximately 1 e.V., thus, a temperature-bias stress (TBS) of the SiO_2 is necessary to alter the sodium distribution within the SiO_2. The mobile hydronium ions have an activation energy on the order of 0.3 eV which enables this charge distribution to be altered by the application of an electric field at room temperature. Finally, the ionized traps may be introduced in the SiO_2 or at the Si-SiO_2 interface by irradiation with X-rays [14]. In general, these traps can be annealed out by short term heat treatments in the range from 300 - 500 °C. X-ray damage may be introduced in the electron beam deposition of the Al [15].

3. Redistribution of Space-Charge in the SiO_2

The charge distribution in the SiO_2 may be altered with an electric field and/or temperature for a specified time interval. Figure 2 illustrates the distribution of space-charge density $\rho(x)$ in the SiO_2. Consideration of minimum energy for a sheet charge located at a distance "x" from the Al-SiO_2 interface leads to the expression [12]

$$V_{FB} = \phi_{MS} - \frac{1}{C_{ins} x_o} \int_0^{x_o} dx \; x \; \rho(x) \qquad (1)$$

for the flat-band voltage V_{FB}. ϕ_{MS} is the metal-semiconductor work function difference, C_{ins} the oxide (insulator) capacitance per unit area and x_o the oxide thickness. Figure 3 indicates the redistribution of charge within the SiO_2 under a TBS sequence. Part (a) shows the system immediately after processing which indicates a distribution in favor of the Al-SiO_2 interface. This is due to the observation that the ionic charge occurs after the SiO_2 is formed over the Si surface. The first step in the TBS sequence is to move all of the mobile charge at room temperature to the Si-SiO_2 interface with the application of a field strength of approximately 2×10^6 V/cm as shown in Part (b). The field is reversed and Curve (c) is obtained. The difference between Curves (b) and (c) is the untrapped mobile hydronium ion concentration (i.e., ions/cm^2) in the SiO_2. Part (d) represents the charge distribution after the application of an electric field strength of 2×10^6 V/cm at an elevated temperature such as 200 °C for 3 minutes. All of the ionic charge is moved over to the Si-SiO_2 interface under this TBS operation. (The system is cooled to room temperature under bias.) Next, the field is reversed at room temperature to determine the total mobile charge in the SiO_2 as shown by Curve (e). The difference between Curves (d) and (e) is the total mobile charge which consists of any trapped species released by the TBS operation. A comparison of this voltage shift with the voltage shift obtained from Curves (b) and (c) will indicate the level of trapped mobile charge in the SiO_2. This trapped charge generally exists at the Al-SiO_2 interface since the initial Curve (a) is often shifted at flat-band by the amount $V_{FB} = \phi_{MS}$. This also indicates the alkali contamination is at the Al-SiO_2 interface. (i.e., if $\rho(x) = Q_{ss}\,\delta(x)$ is a sheet charge located at the Al-SiO_2 interface, then equation (1) indicates $V_{FB} = \phi_{MS}$.)

The final step in the TBS sequence is the application of the electric field at 200 °C for 3 minutes to move all of the ionic charge to the Al-SiO_2 interface. The difference between Curve (e) and (f) represents the alkali ion contamination, whereas, the difference between Curve (f) and the ideal or theoretical curve (i.e., theoretical curve is formed for $\phi_{MS} = Q_{ss} = 0$) yields the level of fixed intrinsic charge and the ϕ_{MS} contribution. As mentioned previously, the level of fixed charge at the Si-SiO_2 interface will vary dependent upon the removal of the fluoride ion. If the fluoride ion is properly removed from the Si surface, then there is an absence of measurable fixed charge. Figure 4 illustrates a TBS sequence applied to a Al-SiO_2-Si sample with N_{ss} (mobile) $> 10^{12}$ ions/cm^2, N_{ss} (alkali) $\simeq 10^{12}$ ions/cm^2 and N_{ss} (fixed) $\sim 5 \times 10^{10}$ ions/cm^2 to illustrate the contamination present in "unclean" unit processes. Figure 5 illustrates the TBS sequence applied to a sample with N_{ss} (mobile) $< 10^{10}$ ions/cm^2, N_{ss} (alkali) $< 10^{10}$ ions/cm^2 and N_{ss} (fixed) $\simeq 0$ (too small to be measured -- less than 10^{10} ions/cm^2, if present). The sample shown in Figure 5 represents the contamination levels observed in "clean" unit processes.

4. Mixed Dielectric Systems to Inhibit Ion Migration

The migration of ionic charge may be inhibited with several dielectric systems:

(i) $P_2O_5 \cdot SiO_2$ / SiO_2

(ii) Si_3N_4 / SiO_2

(iii) Al_2O_3 / SiO_2

Phosphosilicate glass ($P_2O_5 \cdot SiO_2$) was the first mixed dielectric system discussed [16] with regards to surface stabilization in silicon integrated circuit technology. Such a glass inhibits the migration of ionic charge and effectively getters [17] any sodium ions in the underlying SiO_2. Next, the Si_3N_4 / SiO_2 system was introduced [18] with the Si_3N_4 as a protective insulator for the SiO_2. Other features of this

mixed dielectric system have been discovered, such as the ability to store charge at the interface between the two dielectrics [19]. The most recent mixed dielectric system has been the Al_2O_3 / SiO_2 combination [20]; however, this system is still under development. In this paper we will discuss the $P_2O_5 \cdot SiO_2$ passivation glass.

The phosphosilicate glass is characterized by an internal polarization [21] and a susceptibility to mobile ion contamination as it is hygroscopic. In practice, a thin layer of phosphosilicate glass is used (i.e., $x_g \ll x_o$) over the SiO_2. This glass traps the alkali and mobile charge close to the Al interface as shown in Figure 6. The model shown in Figure 6 depicts the ionic contamination Q_{ion} uniformly distributed thoughout the phosphosilicate glass. For a constant volume charge density we may write,

$$\rho(x) = \frac{Q_{ion}}{x_g} \qquad 0 < x < x_g$$

$$= 0 \qquad x_g < x < x_g + x_o \tag{2}$$

Substitution of Equation (2) into (1) yields,

$$V_{FB} \doteq \phi_{MS} - \frac{Q_{ion}\, x_g}{2K_o\, \epsilon_o} \tag{3}$$

where K_o is the relative dielectric constant of the SiO_2. The so-called fixed charge is neglected in Equation (3) since this was not observed in our experiments. Equation (3) indicates the importance of monitoring Q_{ion}, which is composed of alkali and mobile ions, and the glass thickness, x_g.

A method has been developed to monitor the glass thickness x_g by a determination of the ratio of x_g/x_o through low temperature polarization measurements [21] and the total thickness $x_g + x_o$ through standard capacitance measurements. Figure 7 illustrates the method of TBS applied to a phosphosilicate glass sample. The saturation polarization voltage ΔV_{sat} is relatively independent of temperature and time and is linearly proportional to the polarizing voltage V_p. From electrostatics [21],

$$\frac{-\Delta V_{sat}}{V_p} = \frac{K_o\, \chi_p \left(\frac{x_g}{x_o}\right)}{K_g \left(K_o \frac{x_g}{x_o} + K_g + \chi_p \right)} \tag{4}$$

where χ_p is the volume polarizability and K_g the dielectric constant of the phosphosilicate glass. From Equation (4) the ratio x_g/x_o is experimentally obtained with $K_o \doteq K_g \doteq 3.9$ and $\chi_p \doteq 0.75$, although both quantities depend upon the mol % of phosphorous in the SiO_2. [22]

5. Experimental Techniques to Characterize the Al-SiO_2-Si System

There are two distinct measurement techniques to investigate the quality of the Al-SiO_2-Si system:

(i) Equilibrium C-V Characteristics [3]
(ii) Non-Equilibrium C-t Characteristics [6]

The equilibrium C-V curves are measured by the application of a fixed d-c voltage across the MOS capacitor and the superposition of a small a-c voltage to determine differential capacitance versus d-c bias voltage. In this method, the majority carrier fermi level at the semiconductor surface is constant and $n_o p_o = n_i^2$. The most easily measured C-V curve is the so-called high frequency curve in which the frequency of the test signal is considerably higher than the reciprocal of the bulk generation time τ_g of minority carriers in the depletion region. In these measurements the minority carrier charge distribution in the surface inversion layer is unaffected by the test signal.* As a result of these measurements the threshold voltage V_T and stability with temperature $\frac{\partial V_T}{\partial T}$ can be determined for a MOS Transistor.

---

* test signal in this paper obtained from 1 MHz Boonton L-C Meter

A convenient method of measurement is to refer all surface states to the $Si-SiO_2$ interface with the assumed distribution,

$$\rho(x) = Q_{ss}\,\delta(x - x_o) \tag{5}$$

Substitution of eq (5) into (1) yields,

$$V_{FB} = \phi_{MS} - \frac{Q_{SS}}{C_{ins}} \tag{6}$$

where Q_{ss} is an effective surface state density placed at the $Si-SiO_2$ interface. The threshold voltage V_T may be written in terms of the effective surface state density as, [23]

$$V_T = 2\phi_F + \phi_{MS} - \frac{(Q_{ss} + Q_s)}{C_{ins}} \tag{7}$$

where ϕ_F is the Fermi level measured from the intrinsic Fermi level and Q_s is the bulk semiconductor charge density. Figure 8 illustrates a method of obtaining V_T experimentally once the high frequency C-V minimum is known.

A study of the MOS structure in a non-equilibrium situation will yield information regarding the dynamics of minority carrier movement at the $Si-SiO_2$ interface and in the bulk depletion region. The basic idea of the C-t measurement is shown in Figure 9. For $t < 0$, a positive step voltage has placed the N-type semiconductor surface in accumulation. At $t = 0$, a large negative step voltage drives the device into deep depletion and a net thermal generation of electron-hole pairs occurs. Holes drift to the interface to form the inversion layer while electrons drift towards the neutral bulk to help neutralize the ionized donor atoms in the collapsing depletion layer. A computerized data reduction technique [24] has been employed to analyze the data to obtain surface recombination velocity and bulk generation lifetime; however, if the surface effects can be neglected, then an approximate expression for the bulk generation lifetime becomes,

$$\tau_g = 2\,\frac{n_i}{N_D}\left(\frac{C_F + C_o}{4\,C_{ins}}\right)^2 \frac{C_{ins}}{C_F}\; t_F \tag{8}$$

where t_F is the length of duration associated with the decay of the depletion layer to its equilibrium value. In general, lifetimes less than 20 μsec may be estimated with equation (8) and this corresponds to a transition frequency of 8 KHz. Thus, as we near 8 KHz with our test signal frequency there will be an increasing response of the minoirty carriers in the inversion layer of the C-V measurements to the applied signal. This will result in the low frequency C-V curve. Surface recombination velocities in the deep depletion region will generally be less than 10 cm/sec for the fabrication process discuss in section (6). Figure 10 illustrates a typical C-t measurement.

6. Experimental Preparation of the $Al-SiO_2-Si$ System

Figure 11 illustrates the gate oxidation procedure for low effective surface state MOS Transistors. [25] The silicon wafers are (100) orientation, N-type, 4-8 ohm-cm. obtained from Monsanto Co., St. Louis, Mo.. The surfaces are prepared with a chemical-mechanical polish to provide low dislocation counts (i.e. less than $500/cm^2$). The gate oxidation procedure is monitored with an undoped SiO_2 and a phosphosilicate glass/SiO_2 sample. The undoped SiO_2 provides information about the level of alkali and hydronium ion contamination while the glass sample yields the values of V_{FB}, V_T, x_g, x_o and device stability $\partial V_T/\partial T$. The glass sample is also used for the determination of surface recombination velocity and bulk generation lifetime. Table I illustrates the values of surface parameters obtained with the process shown in Figure 11.

Table I (Surface parameters)

$V_{FB} = -0.9 \pm 0.1v$	$x_g \doteq 100 \pm 50A^o$
$V_{TP} = -2.4 \pm 0.2v$	$x_o \doteq 1100 \pm 50A^o$
$\phi_{MS} = -0.18 \pm 0.01v$	$\partial V_T/\partial T = \pm 0.2v$

To monitor the gate oxidation the surface must be prepared in a manner identical to device fabrication. Thus, a monitor wafer has an initial oxide (thermal) which is removed prior to the gate oxidation procedure. This is called an "oxide etch" since the growth of this initial oxide removes a few thousand angstroms of Si. Pregate cleaning consists of hot H_2SO_4 and HNO_3 baths followed by an HF dip (10% for 15 sec.) to remove the chemical oxide from the surface. A final rinse in boiling double distilled, H_2O for 20 min. removes the fluoride ion from the Si surface. [9-10] The phosphosilicate glass is derived from a gaseous phosphine source (PH_3) and the furnace is pre-doped to insure uniform thickness of glass from run to run. The interface states are annihilated [5] by the hydrogen present in the decomposition products of the phosphine reaction.

The photoengraving is performed to open the contact windows in the device fabrication. This process introduces significant amounts of alkali and hydronium ion contamination (i.e. greater than $10^{11}/cm^2$) on the wafer surface. A chelating agent EDTA is employed to remove heavy metal contamination prior to gate metalization. The aluminum metalization process is very important as it is a source of alkali and hydronium ion contamination. [26] Table II illustrates the results obtained with a combination of evaporation systems and various crucibles. It can be seen that the electron beam evaporation (direct) provides alkali and hydronium ion levels below $10^{10}/cm^2$ while other systems are an order of magnitude higher.

Table II (Ionic charge in Evaporation Systems)

	Filament		Electron Beam	
	Standard	Low Alkali	Crucible A TiB_2-BN	Crucible B TiB_2-BN
N_{ss}(Alk.)	$> 10^{12}/cm^2$	$< 10^{12}/cm^2$	$> 10^{11}/cm^2$	$< 10^{10}/cm^2$
N_{ss}(mob.)	$> 10^{11}/cm^2$	$> 10^{11}/cm^2$	$> 10^{11}/cm^2$	$< 10^{10}/cm^2$

7. Conclusions

A model of the Al-SiO_2-Si system has been presented to aid in the understanding and control of the silicon surface. The results of our experiments show the absence of measureable fixed charge near the Si-SiO_2 interface and a distribution of ionic charge in the phosphosilicate passivation glass. This distribution is responsible for the shift in V_{FB} from the value ϕ_{MS}. We discuss the use of C-V and C-t measurements to examine the quality of the silicon surface.

8. References

[1] Vadasz, L. L., Grove, A. S., Rowe, T. A., and Moore, G. E., IEEE Spectrum, 6, 28 (1969).

[2] Frohman-Bentchkowsky, D. and Lenzlinger, M., J. Applied Phys., 40, 3307 (1969).

[3] Grove, A. S., Deal, B. E., Snow, E. H. and Sah, C. T., Solid-State Electronics, 8, 145 (1965).

[4] NASA Microelectronics Subcommittee, NASA-STD-XX-3, Method 6041 A, 182 (1968).

[5] Brown, D. M. and Gray, P. V., J. Electrochem Soc., 115, 760 (1968).

[6] Zerbst, M., Z. Angew. Phys., 22, 30 (1966).

[7] Many, A., Goldstein, Y. and Grover, N. B., Semiconductor Surfaces, p. 11, North-Holland Pub. Co., (Amsterdam, 1965).

[8] Deal, B. E., Grove, A. S. and Snow, E. H., J. Electrochem Soc., 114, (1967).

[9] White, M. H. and Cricchi, J. R., Solid-State Electronics, 9, 991 (1966).

[10] Larrabee, G. B., Heinen, K. G., and Harrell, S. A., J. Electrochem Soc., 114, 867 (1967).

[11] Pauling, L., The Nature of the Chemical Bond, Cornell University Press, Ithica, N. Y. (1960).

[12] Snow, E. H., Grove, A. S., Deal, B. E. and Sah, C. T., J. Appl. Phys., 36, 1664 (1965).

[13] Hofstein, S. R., IEEE Trans. Electron Dev., ED-14, 749 (1967).

[14] Mattauch, R. J. and Lade, R. W., Proc. IEEE, 53, 1738 (1965).

[15] Collins, D. R. and Sah, C. T., Applied Phys. Letters, 8, 124 (1966).

[16] Kerr, D. R., Logan, J. S., Burkhardt, P. J. Pliskin, W. A., IBM J. Res, and Dev., 8 376 (1964).

[17] Yon, E., Ko, W. H. and Kuper, A. G., IEEE Trans. Electron Dev., ED-13, 276 (1966).

[18] Dalton, J. V., J. Electrochem. Soc., 113 1650 (1966).

[19] Chu, T. L., Szedon, J. R. and Lee, C. H., Solid-State Electronics, 10, 897 (1967).

[20] Cheney, G. T., Jacobs, R. M., Korb, H. W., Nigh, H. E. and Stack, J., IEEE Int'l Electron Dev. Meeting, Paper 2.2, Washington, D. C. (1967).

[21] Snow, E. H. and Deal, B. E., J. Electrochem Soc., 113, 2631 (1966).

[22] Kerr, D. R., Reliability Physics Symp., Session 1.2, Las Vegas, Nevada, (1970).

[23] Sah, C. T. and Pao, H. C., IEEE Trans. Electron Dev., 13, 393 (1966).

[24] Mack, I. A. G., Blaha, F. C., White, M.H., GOMAC 1970, Ft. Monmouth, N. Y. (1970).

[25] Development of the 1000°C process for CMOS Transistor fabrication is due to J. R. Cricchi and R. M. McLouski of our laboratory.

[26] Hoenig, S. A. and Pope, R. A., Solid-State Technology, 11, 53 (1968).

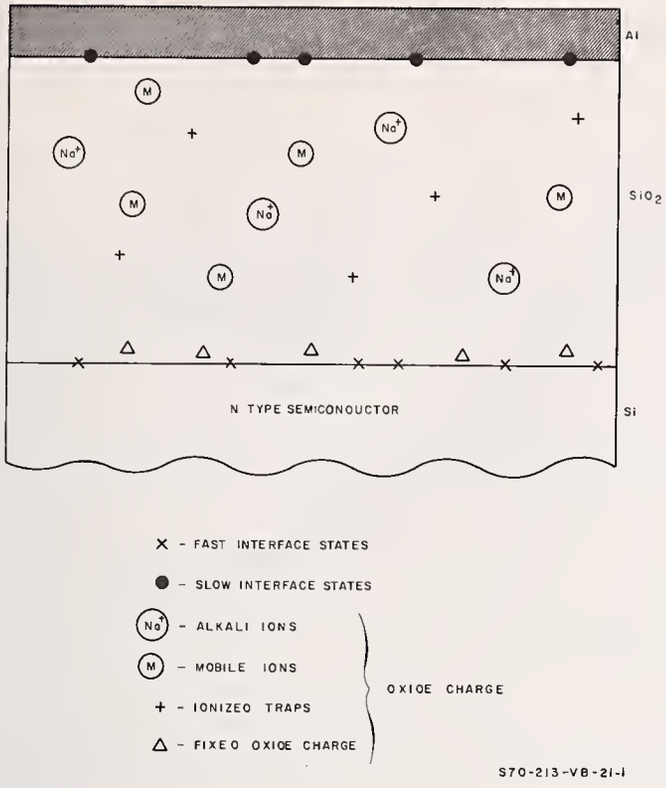


Fig. 1. Classification of States and Charges for the Al-SiO_2-Si System.

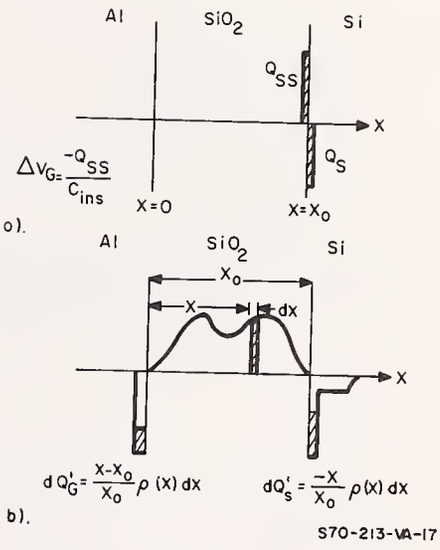


Fig. 2. Induced Charge in the Al-SiO_2-Si System.

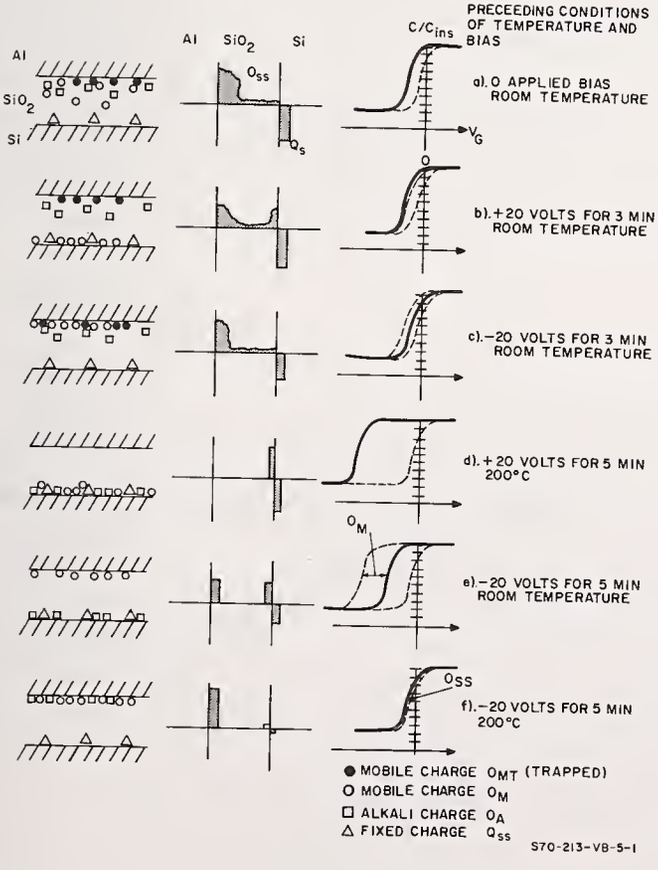


Fig. 3. Charge Distribution in a Non-Doped SiO_2 MOS Structure.

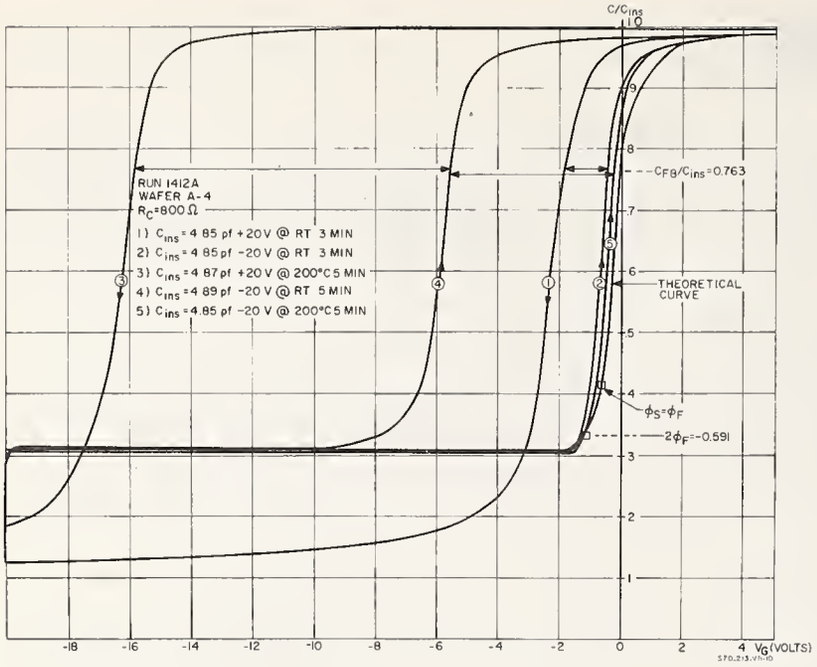
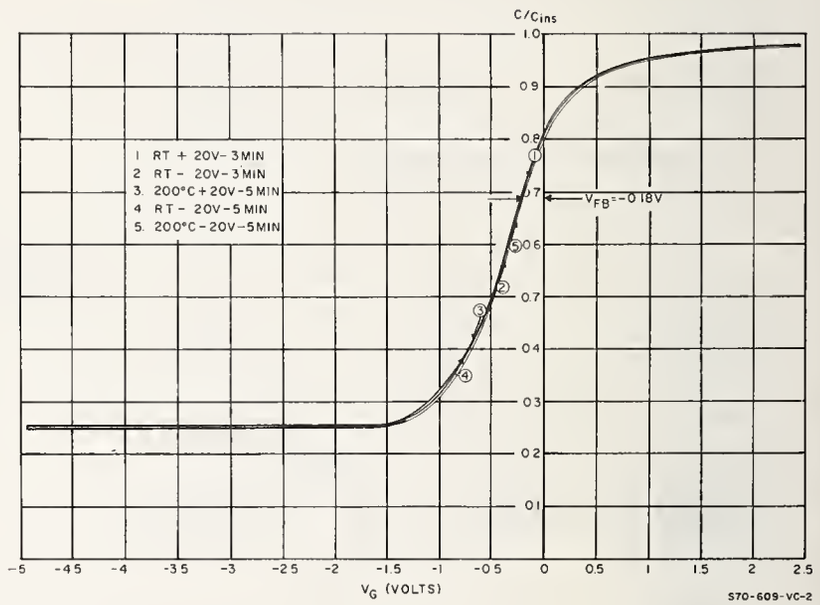


Fig. 4. Sample TBS Curves on a Non-Doped, Unclean, SiO_2 MOS Structure.

Fig. 5. Sample TBS Curves on a Non-Doped, Clean, SiO_2 MOS Structure.



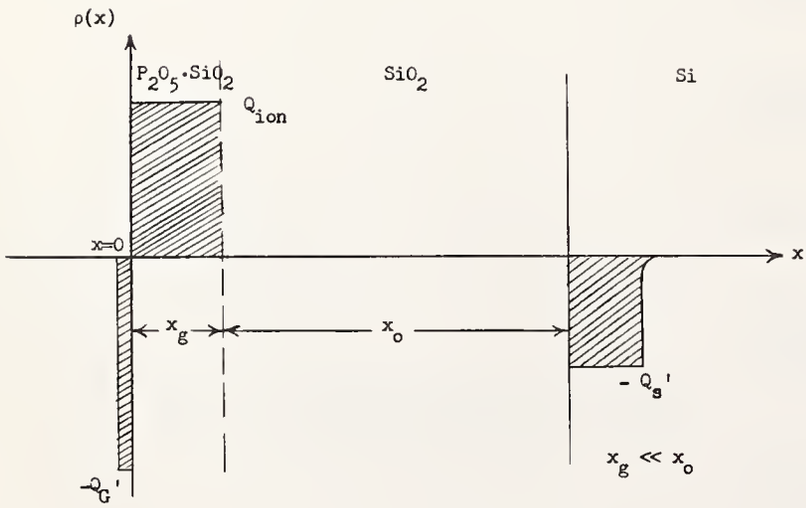


Fig. 6. A Model of the Al-SiO_2-Si System.

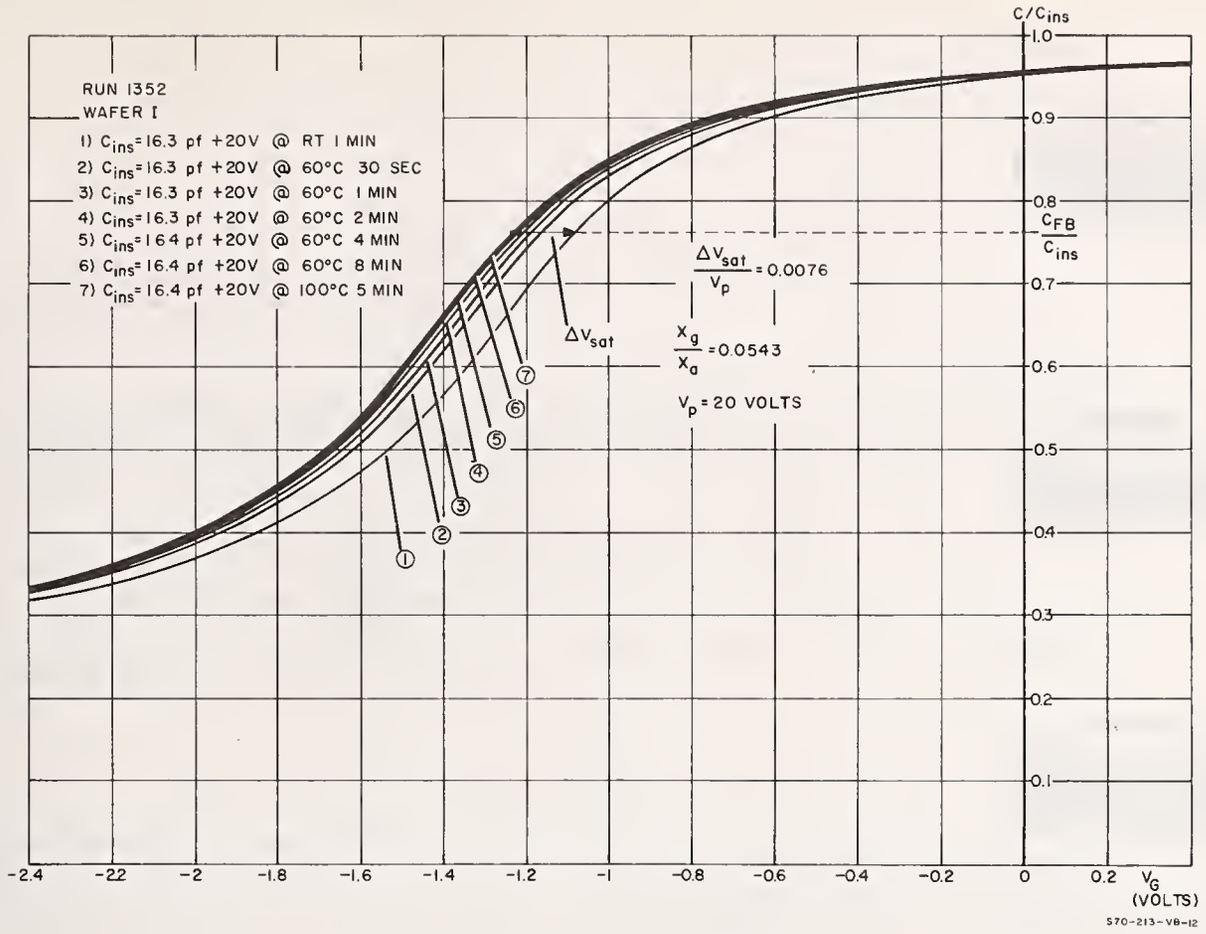


Fig. 7. Polarization Test Sequence.

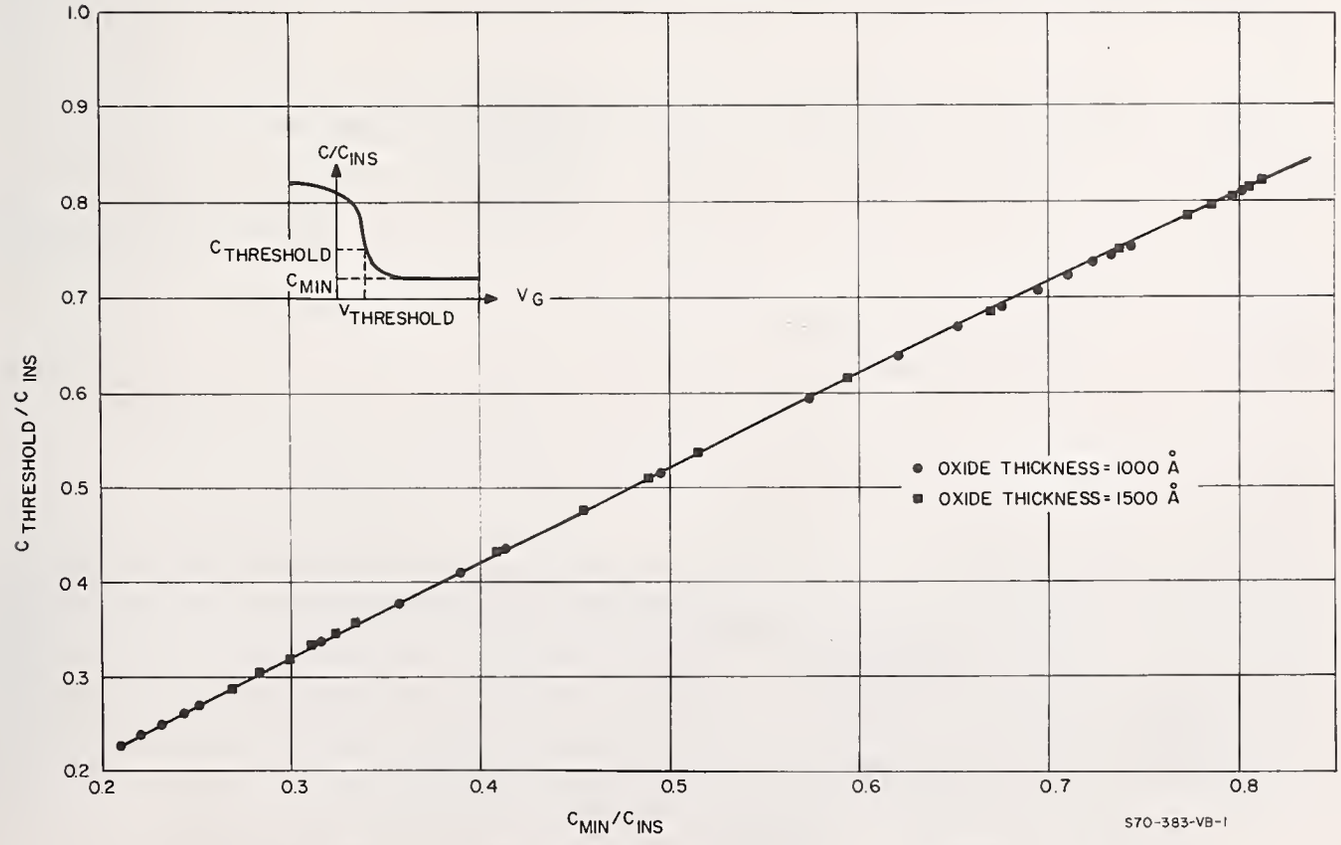


Fig. 8. Threshold Voltage Capacitance C_T versus High Frequency Minimum Capacitance C_{min}.

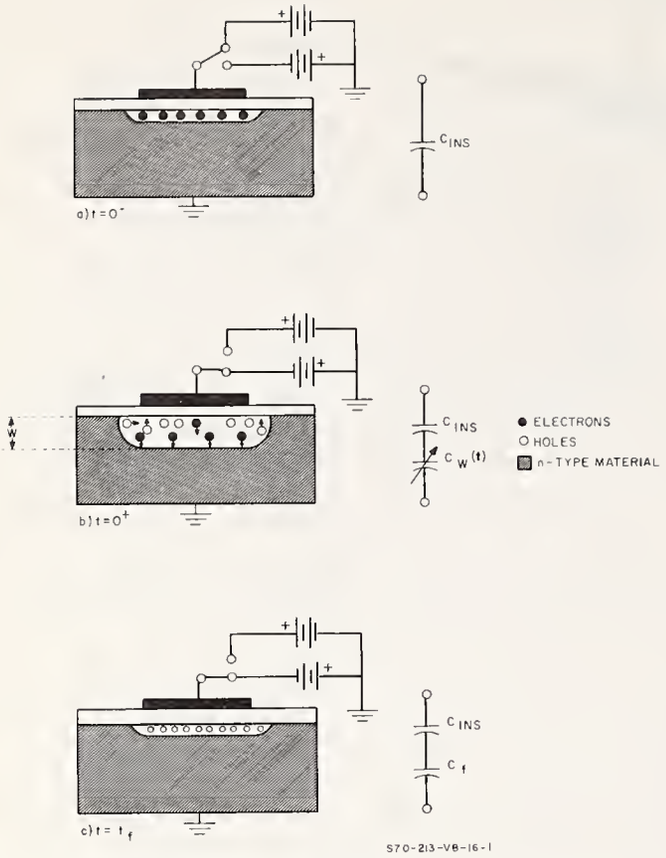


Fig. 9. Non-Equilibrium Capacitance-time Operation of a MOS Structure.

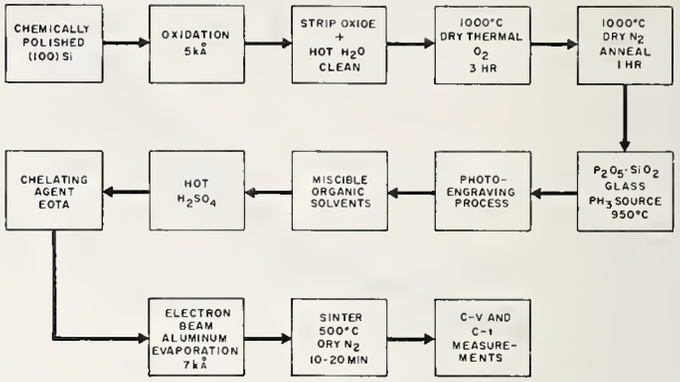


Fig. 11. Al-SiO_2Si System Fabrication

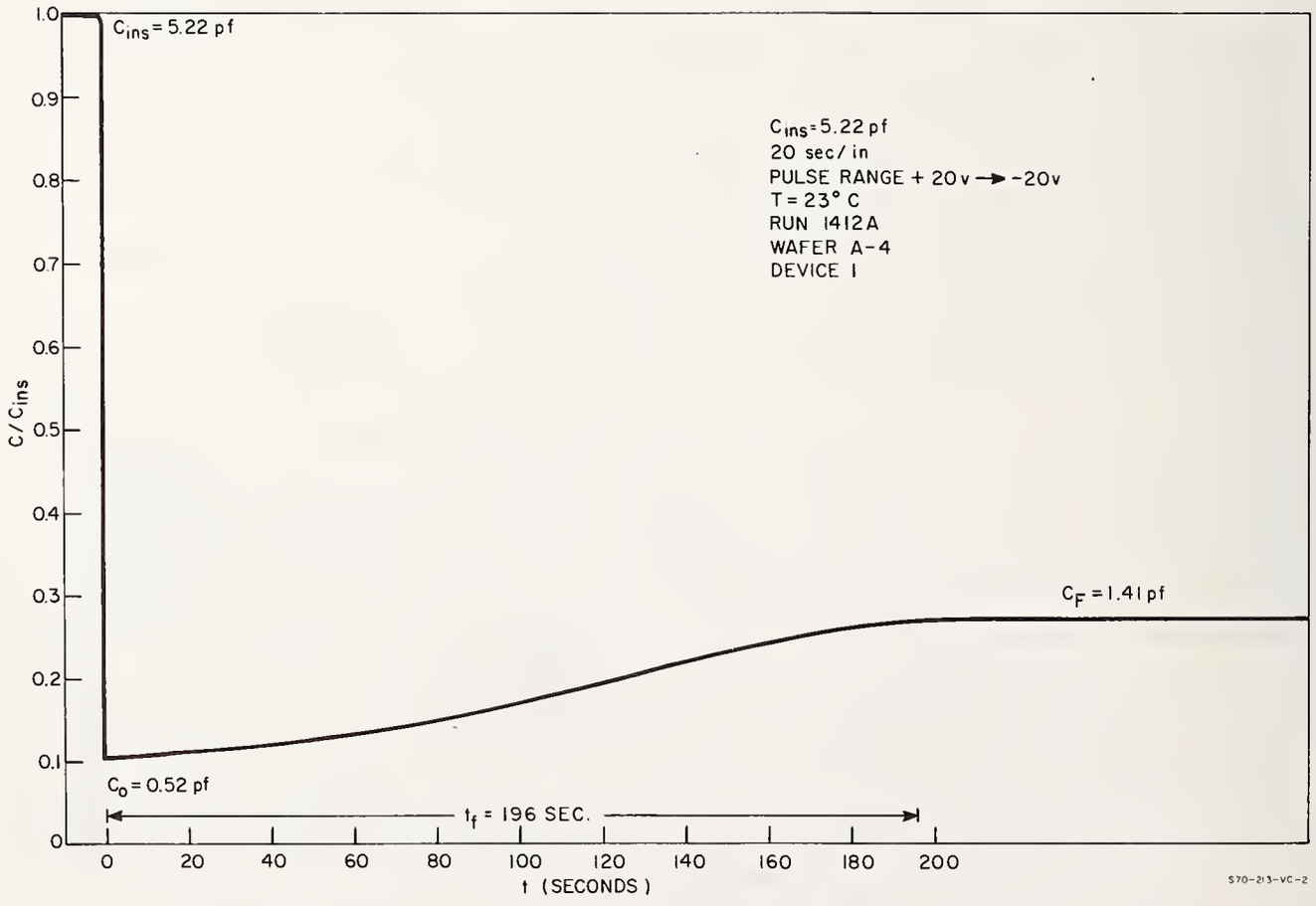


Fig. 10. Pulsed C-t Measurement of a MOS Structure

Activation Analysis In Silicon Device Processing

Graydon B. Larrabee and H. Gary Carlson

Texas Instruments Incorporated
Dallas, Texas

The application of neutron activation analysis to silicon device processing is discussed. A review of the technique is given describing the effects of irradiation conditions, sample handling, and radioassay on the ultimate sensitivity.

Examples of the introduction of ppm to sub-ppb levels of both dopant and unwanted impurities in device processing steps are given.

Key Words: Activation analysis, device processing, dopants, impurities, silicon.

1. Introduction

Silicon device processing is basically the controlled introduction and at the same time extrication of part per billion to part per million levels of impurities. It is necessary to introduce low ppm levels of dopants, such as phosphorus and boron, while at the same time making every effort to exclude unwanted sub-part per billion impurities, such as gold, copper, sodium, etc. Knowledgeable control of the small impurity contents is particularly difficult if it is necessary to rely solely upon the electrical characteristics of the final device. The role of chemical imperfections on final device parameters is not well understood. As a result, it is necessary to utilize a direct analytical technique with high sensitivity. The reasons for this high sensitivity requirement involve not only the low levels of impurities but also the very small sample sizes typically measured in semiconductor process characterization.

Neutron activation analysis supplies most of these requirements and, when properly applied, can give valuable insight into the chemical imperfection problem in silicon device processing.

This paper will treat the general aspects of neutron activation as applied to silicon and then will cover examples of how this characterization technique has been applied to problems involved with silicon device processing.

2. Neutron Activation Analysis

Neutron activation analysis is essentially a four step process involving:

(1) sample preparation for irradiation
(2) irradiation
(3) sample preparation for radioassay
(4) radioassay

The ultimate sensitivity of the technique is controlled by the attainable specific activity, that is, the number of disintegrations per second per unit mass of impurity. This is shown in eq (1), where it can be seen that for any given number of impurity atoms (N), the amount of radioactivity produced (N_{dps}) is a function of irradiation conditions. This controls the analytical sensitivity. The flux and irradiation time are the major variables. However, the flux is composed of slow and fast neutrons in differing amounts which are a function of individual reactors and of any shielding in those reactors. For a given impurity the capture cross section and isotopic half-life are constants.

$$N_{dps} = \phi\sigma N\left(1-e^{\left(\frac{-0.693t}{t_{\frac{1}{2}}}\right)}\right) \quad (1)$$

where ϕ = flux, n/sec/cm^2
σ = capture cross section, cm^2
N = number of impurity atoms
t = irradiation time
$t_{\frac{1}{2}}$ = half-life of isotope being formed

The effect of irradiation time (at constant flux) is shown in Figure I for several impurities. In the upper right of this figure the exponential growth of any radioactive species is shown as a function of half-life of the isotope being formed. As can be seen, after one half-life 50% saturation has been achieved and continued irradiation to an infinite number of half-lives will do no more than double the sensitivity. Since each reactor facility has a different operating cycle, the value of moving from one facility with a 14 hour cycle to one with a 4 day cycle must be evaluated on the basis of what information is required. In Figure I it is obvious that for short lived isotopes such as sodium-24 little is gained by a longer irradiation. But for longer lived isotopes such as chromium-51 and iron-59 there is a significant increase in sensitivity.

Frequently it is possible to couple a longer irradiation cycle with a higher flux (with concomitant irradiation costs) and achieve a significant increase in sensitivity. This is shown in Table I for several selected impurities.

Table I. Typical Sensitivities for Bulk Silicon
Showing the Effect of Varying Flux and Irradiation Time

Isotope	Half-life	Texas A&M Reactor 14 Hours, 1.5×10^{13} n/sec/cm^2	Union Carbide N.Y. Reactor 96 Hours, 7.5×10^{13} n/sec/cm^2
Cu-64	12.6 hrs.	4.2×10^{12}	4.4×10^{11}
Au-198	67.0 hrs.	1.2×10^{10}	5.2×10^{8}
Fe-59	45 days	1.2×10^{16}	3.4×10^{14}
Cr-51	27 days	7.7×10^{13}	2.2×10^{12}
Na-24	14 hrs.	1.6×10^{13}	1.6×10^{12}
Ag-110^m-110	260 days	6.0×10^{13}	1.7×10^{12}
As-76	26 hrs.	9.1×10^{11}	6.0×10^{10}
W-187	24 hrs.	8.6×10^{11}	5.9×10^{10}

At the same time that all of the impurities are being activated, the silicon matrix is also undergoing activation (^{30}Si (n,γ)^{31}Si). Fortunately silicon lends itself to neutron activation because of the comparatively short 2.6 hour half-life activity of the silicon-31 matrix. The stable silicon-30 has only 3.12% isotopic abundance and a relatively small, 0.1 barn, neutron capture cross section. Following irradiation the 2.6 hour silicon-31 is allowed to decay before performing the radioassay for the other activities of interest.

This laboratory has made extensive use of gamma ray spectroscopy [1][1] utilizing 3 x 3" NaI (Tl) detectors with computer reduction of the gamma spectra. The computer program used to analyze the data is a modified form of that developed by Helmer et. al. [2]. A multiprogramming scheme is utilized on a CDC 6400 computer as shown in Figure II. In this scheme the data, starting from gamma spectra, is processed directly to impurity concentrations in a single pass through the computer. Typically 8 to 10 samples can be processed in 20 to 30 seconds of central processor computer time.

[1]Figures in brackets indicate the literature references at the end of this paper.

The absolute sensitivity for activation analysis, as we use it at Texas Instruments Incorporated, has been calculated by Keenan [3]. The values, shown in Table II, are based on a 0.01 intensity ratio from the gamma spectral analysis program, assuming a 100 minute radioassay, and correcting for the efficiency of a 3 x 3" NaI detector. The sensitivity data is conservative, since under favorable conditions better sensitivity is consistantly obtained; but it is sample dependent.

Table II. Detection Limits for Impurities in Silicon. Counted on a 3 x 3 NaI spectrometer. (14.00 H irradiation at 1.50 E + 13 neutrons per cm^2 per sec. 36.000 H decay).

Element	Atoms	Element	Atoms
Antimony 122	2.297E+11	Neodymium 147	7.863E+12
Antimony 124	5.732E+12	Osmium 191	1.697E+12
Arsenic 76	2.106E+11	Palladium 109	2.061E+12
Barium 131	3.510E+14	Phosphorous 32 Beta	*
Barium 133	2.185E+16	Platinum 197 0.19 MeV	1.881E+13
Bromine 82	4.403E+11	Platinum 197 77 KeV	5.248E+12
Cadmium 115	6.088E+13	Potassium 42	1.208E+14
Cadmium 115M	9.735E+14	Praseodymium 142	3.138E+12
Calcium 45 Beta	1.880E+14	Rhenium 188	4.817E+10
Cerium 141	6.833E+12	Rubidium 86	7.643E+13
Cerium 143	4.665E+12	Ruthenium 97	2.478E+13
Cesium 134	3.543E+12	Ruthenium 103	1.109E+13
Chromium 51	3.251E+13	Samarium 153	1.279E+10
Cobalt 60	2.368E+13	Scandium 46	7.224E+11
Copper 64	4.971E+11	Selenium 75	5.185E+13
Dysprosium 165	2.809E+14	Silver 110M	2.549E+13
Erbium 171	1.311E+12	Sodium 24	2.279E+12
Gadolinium 159	2.495E+12	Strontium 85	1.617E+15
Gallium 72	4.254E+11	Sulfur 35 Beta	1.272E+14
Germanium 77	1.531E+14	Tantalum 182 0.1 MeV	2.155E+12
Gold 198	4.075E+09	Tantalum 182 1.0 MeV	3.702E+12
Hafnium 181	1.135E+12	Tellurium 127	1.107E+16
Holmium 166	5.361E+10	Terbium 160	3.815E+11
Indium 114	7.533E+12	Thallium 204 Beta	4.449E+12
Indium 116	4.533E+20	Thulium 170	1.671E+12
Iridium 192	5.762E+10	Tin 113	1.305E+15
Iridium 194	3.527E+10	Tungsten 187 0.1 MeV	1.856E+11
Iron 59	4.982E+15	Tungsten 187 0.7 MeV	1.637E+11
Lanthanum 140	1.398E+11	Ytterbium 169	3.685E+11
Lutecium 177	1.527E+11	Yttrium 90M	8.149E+16
Manganese 56	8.858E+13	Zinc 65	5.281E+14
Mercury 197	1.017E+12	Zinc 69M	2.706E+13
Mercury 203	3.790E+12	Zirconium 95	1.387E+15
Molybdenum 99	4.692E+13	Zirconium 97	4.968E+14

* Limited by secondary reaction on silicon to 4.98E+13.

This sample dependence is shown in Table III where it can be seen that the detection limit (expressed in concentration units) can vary as much as a factor of 400 depending on the silicon process being evaluated. A summary of the type of sample and sample size that can be expected in silicon processing is shown in Figure III. Notice that the three steps, oxidation, photoresist-etch and diffusion are repeated a number of times, depending on the process. Extraneous impurities can be introduced in each step.

Table III. Detection Limit as Determined by Sample Analyzed*

	Single Crystal Silicon	Epitaxial Film	Oxide Film	Diffused Area
Range of Sample Size	0.1 to 10 grams	2.5 to 25μ	1000 to 10,000 Å	0.5 to 5μ
Typical Sample 2 inch slice	1 gram	10μ	5,000 Å	2.5μ
Gold Detection Limits				
(μg)	1.3×10^{-6}	1.3×10^{-6}	1.3×10^{-6}	1.3×10^{-6}
(atoms)	4.0×10^{9}	4.0×10^{9}	4.0×10^{9}	4.0×10^{9}
(ppma)	1.9×10^{-7}	4.0×10^{-6}	8.0×10^{-5}	1.6×10^{-5}
(atoms/cm^3)	9.3×10^{9}	2.0×10^{11}	4.0×10^{12}	8.0×10^{11}

* 14 hour irradiation at 1.5×10^{13} n/sec/cm^2, 36 hour decay before gamma ray spectroscopic analysis

3. Analysis of Silicon

In the analysis of impurity levels introduced during device processing it is frequently sufficient to analyze silicon slices. A set of undoped, high resistivity, silicon slices, with known impurity content, can be introduced at the beginning of the process and slices removed as the batch is processed. The slices are then analyzed as described earlier. A typical sequence is shown in Table IV for two processes that were experiencing difficulty. As can be seen, each processing step adds more impurities. It is then possible to change the processing to eliminate these impurities and thus improve the process.

4. Analysis of Dielectric Films

Frequently, in device processing it is necessary to know the spatial distribution as well as the total concentration of impurities. This is particularly true for dielectric materials such as silicon dioxide and silicon nitride which are a fundamental part of the fabrication, passivation, and in some cases actual operation of silicon devices (FET). In the case of such films, more than anywhere else in device processing, the sample size is very small (< 1 mg), yet ppm levels of impurities are important.

The role of sodium, a highly mobile species, has been extensively studied in silicon dioxide [4, 5, 6]. Only neutron activation analysis can provide the sensitivity required for this type of analysis. A typical sodium distribution in a "clean oxide" is shown in Figure IV. Contrast this with the distribution in Figure V for silicon dioxide formed in a wire wound furnace on silicon surfaces prepared with normal semiconductor grade chemicals. The bulk levels differ by a factor of ten, and most of the difference occurs on the oxide surface and at its interface with silicon. Activation energy for sodium diffusion in the more contaminated sample is 1.0 ev while it is 1.4 ev in the cleaner sample.

Sample handling of these samples prior to irradiation is a critical problem for the analyst because we are dealing with so few total impurity atoms. For example, ultrapure vapor etched silicon surfaces with $<10^{11}$ atoms Na/cm^2 accrues to about 10^{14} atoms/cm^2 when dipped in 15 meg-ohm deionized water. Similarly silicon samples resting in a non-clean room environment, such as found in a normal laboratory, accrue 10^{15} atoms Na/cm^2.

Table IV. Neutron Activation as Used for Evaluation of Process Induced Chemical Imperfections

A. Discrete Device Process

Operation	Copper	Gold
1. After Slice Cleanup	1.2×10^{13}	1.2×10^{12}
2. After Oxidation	3.1×10^{13}	0.6×10^{12}
3. After N Deposition	8.4×10^{13}	1.5×10^{12}
4. After KMER	*	2.0×10^{14}
5. After Pre-P Cleanup	1.8×10^{14}	2.3×10^{12}
6. After P Diffusion	2.5×10^{14}	3.4×10^{12}

B. Integrated Circuit Process

Operation	Copper	Gold	Sodium
1. After Slice Cleanup	8.0×10^{13}	7.6×10^{10}	1.6×10^{14}
2. After Oxidation	3.1×10^{15}	9.5×10^{12}	8.5×10^{13}
3. After Epi Layer	2.4×10^{15}	7.5×10^{12}	4.7×10^{15}
4. After Epi Layer with Duf	5.0×10^{15}	3.6×10^{13}	5.0×10^{16}

* High gold content precluded copper analysis.

A highly effective preirradiation packaging technique was developed that has been found to be applicable to all silicon neutron activation analyses. Special quartz irradiation containers were fabricated, (see Figure VI) and are approximately the same diameter as the silicon slice and about 1 cm in height. There is a quartz cover that fits inside the container, resting on the slices. Using either ultrapure silicon slices or silicon slices identical to those for analysis as spacers, the slices which have been processed through clean oxidation furnace are stacked into the quartz container. The quartz container is then wrapped in aluminum foil and submitted with the flux monitor for irradiation.

The advantages of this preirradiation packaging system are many. The process engineers can package their own slices thus ensuring the engineer that his samples were not inadvertently contaminated during packaging. Packaging can also be performed in the same ultraclean rooms in which the slices were processed.

Sodium-24 activity is produced not only from the primary thermal neutron reaction, [$^{23}Na(n,\gamma)^{24}Na$], but also from the fast neutron reactions on sodium [$^{23}Na(n,\gamma)^{24}Na$], aluminum [$^{27}Al(n,\alpha)^{24}Na$] and magnesium [$^{24}Mg(n,p)^{24}Na$]. The magnitude of the isotopic masking in any set of samples can be determined by the classical cadmium shielding experiment to exclude thermal neutrons. In our work [4] we had separately measured by spark source mass spectroscopy the impurity levels to be 22ppm for magnesium and 13 ppm for aluminum for a particular set of samples. The efficiency of conversion is sufficiently poor by fast neutrons to make no significant contribution to the sodium-24 counted. This was confirmed by the cadmium shielding experiments [4].

The spatial and concentration distribution of several impurities can be measured simultaneously. Figure VII shows the distribution of both phosphorus and sodium in silicon dioxide where the phosphorus was diffused into the silicon dioxide during a procedure typical of an emitter diffusion for an integrated circuit. In this case copper and gold and arsenic were present and were gettered into the phosphosilicate phase. It was necessary to chemically separate these, before counting as sulfide precipitates. Incremental volumes etched off the film are represented by the points on the curves. Each increment was counted for gamma and beta activity. The extremely small sample volume can be inferred from the number of data points. In some cases volumes were combined prior to phosphorus counting.

Figure VIII shows an analysis of an arsenic doped deposited oxide applied over a 4000 Å, undoped, silicon oxide film thermally grown on silicon. Prior to thermal diffusion the concentration was about 1.5×10^{21} atoms/cm^3 and uniform throughout the deposited layer. After diffusion the curve shows outdiffusion with arsenic escaping from the sample as well as diffusion into the underlying undoped silicon oxide. Detail at the oxide-silicon interface is not shown. However, the arsenic surface concentration on the silicon was about 2×10^{17} atoms/cm^3 (metal-oxide silicon capacitor measurement of the silicon surface doping density) and this agreed well with that measured by activation analysis. For studies of this type it is fortunate that extremely pure boron doped silicon is available, wherein neither the boron nor the silicon form long lived isotopes.

Numerous other examples can be cited for study of chance and intentional impurity introduction. Those given have shown the sensitivity of the method and have treated significant process variable not resolvable by a more conventional method.

5. Analysis of Diffusion Processes

The literature is replete with papers dealing with the application of radiotracers to the study of the diffusion of impurities in silicon. It is then only a natural extention to apply neutron activation analysis to the diffusion processes in device fabrication. The experimental techniques for performing this analysis are described in detail by Kane and Larrabee [7]. There is one problem with the use of neutron activation analysis in this area and that is the non-specificity of the radioactivity produced. For example, a study of gold diffusion into an arsenic doped slice with say a phosphorus emitter in a boron base or collector would be extremely difficult. Further, autoradiography could not be used. However, by careful control of experimental conditions and using sophisticated radioassay techniques, it is possible to obtain very good data on impurity distribution.

6. Future Trends

The use of activation analysis is silicon device process studies will continue and when coupled with the high resolution available with the new lithium drifted germanium (Ge(Li)) detectors will overcome the problem of one impurity masking another. While Ge(Li) detectors do not have the efficiency of 3 x 3" NaI(Tl) detectors, the full impact of these high resolution systems in this type of work is only starting to be developed.

In addition other forms of activation using heavy particle bombardment will find application to study of impurities which do not yield a sufficient radioactivity after neutron bombardment. Boron, aluminum, oxygen and carbon are species of high interest to workers in semiconductor device processes. However, an immense number of device process await elucidation by neutron activation analysis as herein described.

7. References

[1] Heinen, K. G. and Larrabee, G. B., Anal. Chem., 38, 1853 (1966).

[2] Helmer, R. G., Heath, R. L., Metcalf, D. D., and Cazier, G. A., IDO 17015 (1964).

[3] Keenan, J. A., Private Communication (1970).

[4] Osborne, J. F., Larrabee, G. B., and Harrap, V., Anal. Chem., 39, 1144 (1967).

[5] Yon, E., Ko, W. H., and Kuper, A. B., IEEE Trans. Electron Devices, ED-13, 276 (1966).

[6] Carlson, H. G., Fuller, C. R., Meyer, D. E., Osborne, J. F., Harrap, V., and Brown, G. A., Phys. Failure Electronics 5, 265 (1967).

[7] Kane, P. F., and Larrabee, G. B., "Characterization of Semiconductor Materials" p. 265 McGraw-Hill, N. Y. (1970).

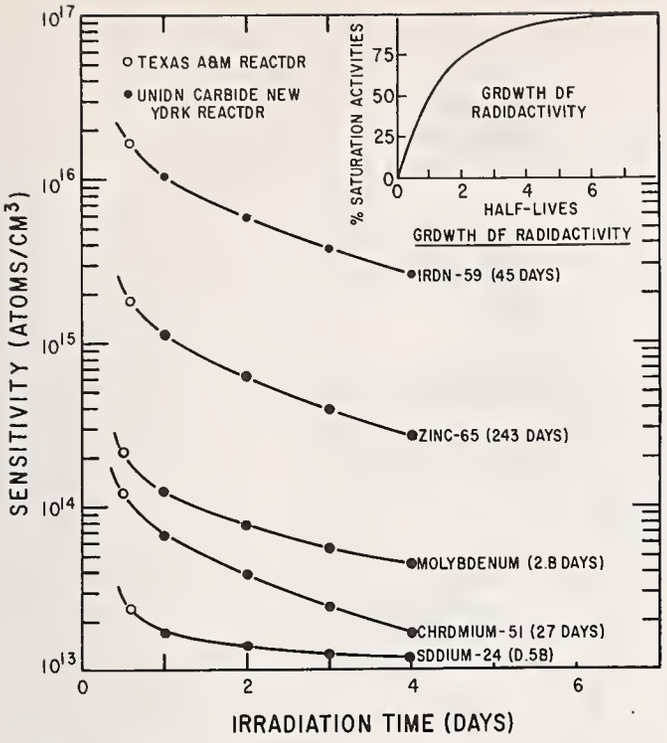


Fig. I. Effect of irradiation time on the sensitivity for bulk silicon analysis. (Based on a one gram sample and a flux of 1×10^{13} n/sec/cm^2)

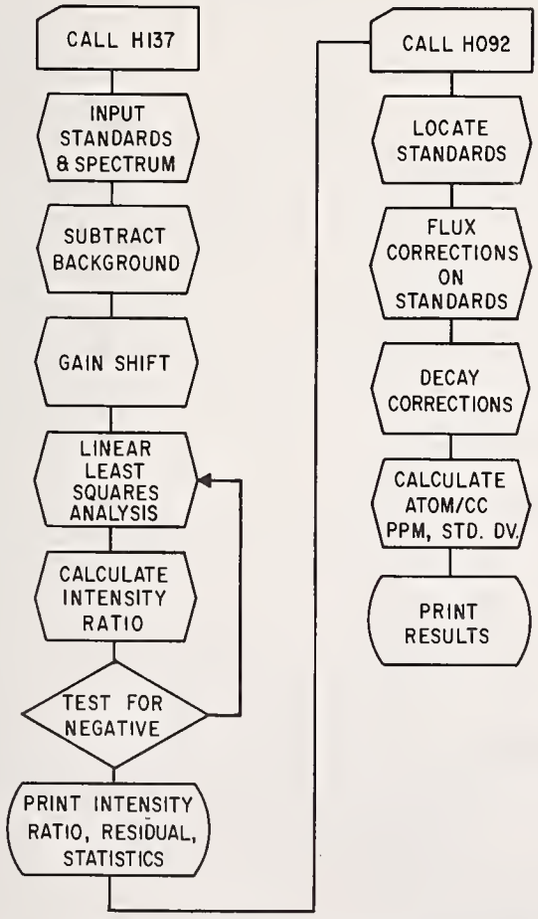


Fig. II. Multiprogramming used to analyze gamma spectral data and calculate impurity concentrations.

PROCESSING STEP	LOCATION OF IMPURITY	SAMPLE SIZE (mg)
POLYCRYSTALLINE	BULK	1000
SINGLE CRYSTAL	SURFACE (1-4μ) BULK	0.5 - 2.0 1000
EPITAXIAL	SURFACE (1-4μ) FILM (10μ)	0.5 - 2.0 5.0
OXIDATION	FILM (5000Å)	0.25
PHOTORESIST & ETCH	SURFACE (1-4μ)	0.5 - 2.0
DIFFUSION	FILM (2.5μ) BULK	1 1000
METALLIZATION		

Fig. III. Flow diagram illustrating silicon processing and showing location of impurities and probable sample size.

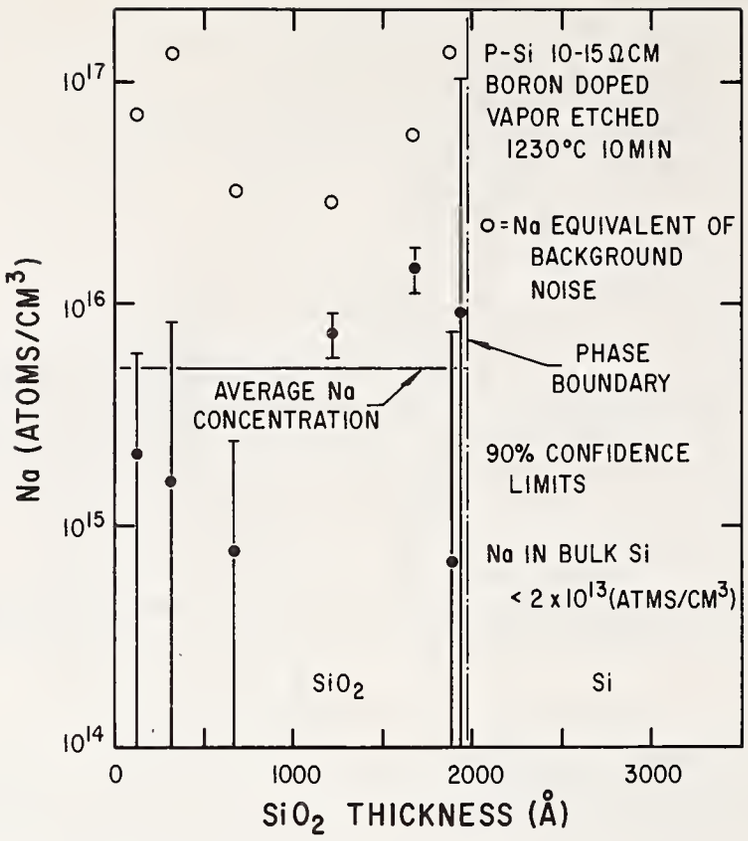


Fig. IV. Sodium distribution in oxide film formed in a cold tube oxidation furnace.

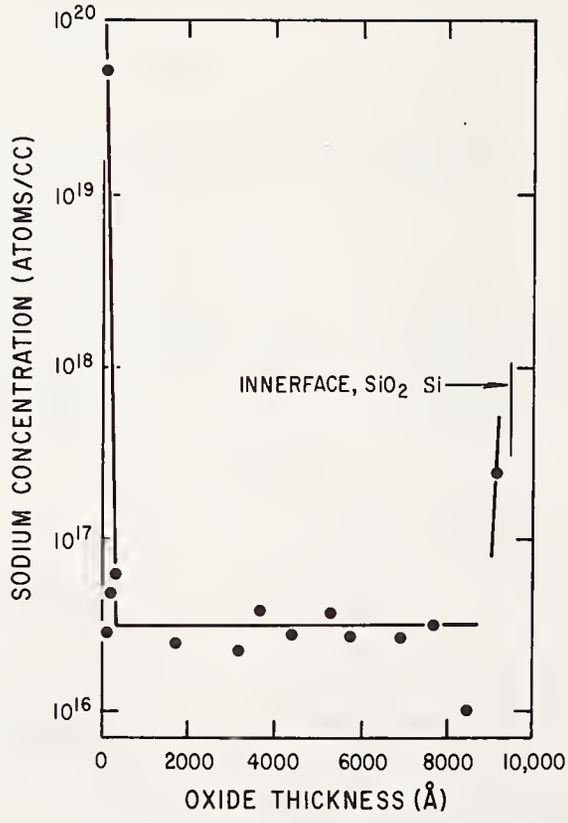


Fig. V. Sodium distribution in dry steam oxide formed in a conventional wire wound furnace.

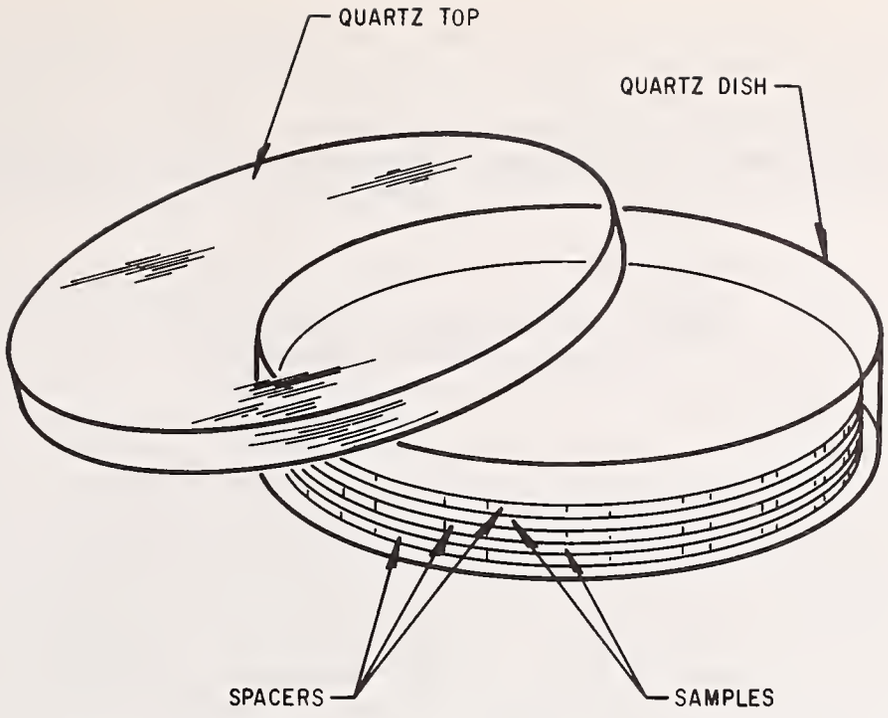


Fig. VI. Quartz irradiation container and high purity silicon spacers used for neutron activation analysis.

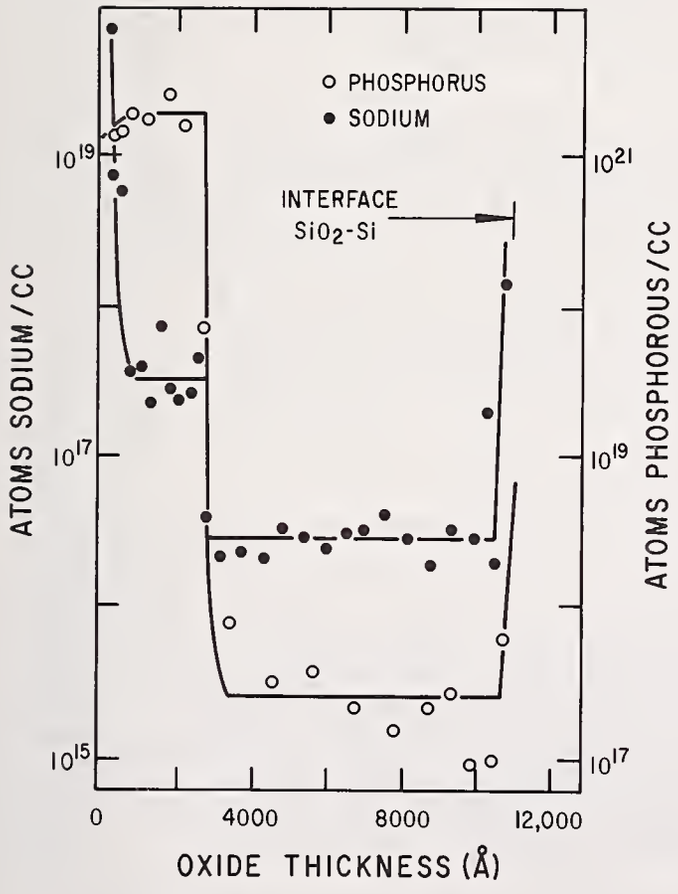


Fig. VII. Sodium and phosphorus distributions in dry steam oxides.

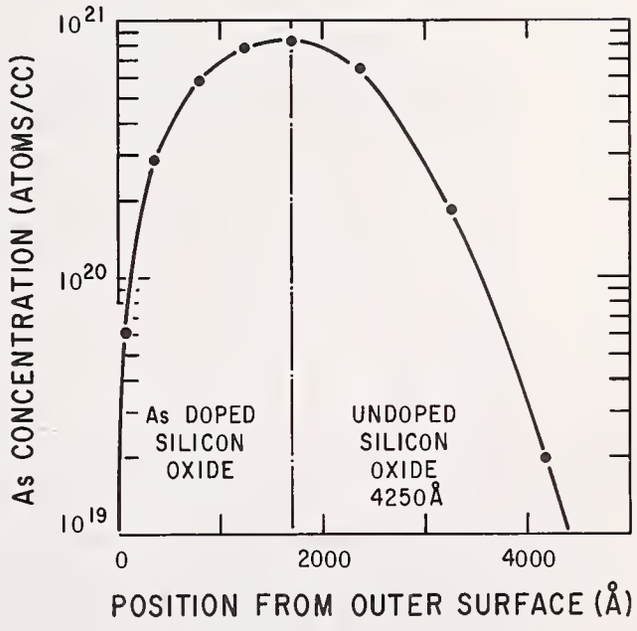


Fig. VIII. Arsenic doped deposited silicon oxide used as a diffusion source. C_s on silicon was 2 x 10^{17} atoms/cc after diffusion.

The Use of the Scanning Electron Microscope As a Semiconductor Production Line Quality Control Tool

John W. Adolphsen and Robert J. Anstead

NASA/Goddard Space Flight Center
Greenbelt, Maryland 20771

The use of the scanning electron microscope (SEM) as an in-line, quality control tool to determine the integrity of metallization on semiconductor devices is recommended in future procurements by the Goddard Space Flight Center of NASA. Experience with failed devices at the GSFC has demonstrated that opens in the metallization at oxide steps is not confined to a single vendor, a single point in time, nor to a particular device type, but rather to small geometry contact window devices in both discrete transistors and integrated circuits. The extent of the problem and its seriousness is illdefined, although one user claims that ".. contact window defects probably constitute the most significant single group of defects (excluding workmanship defects) causing failures of integrated circuit devices in electronic systems today".[1] The problem is compounded by the lack of any screening procedure for packaged devices which is highly effective in detecting those units which can and do cause latent in-system failures.

Heretofore, the procedure has been to sample inspect devices with a SEM following incoming inspection. Should these samples show evidence of opens or partial opens in the metallization at the oxide steps, the user faces two unattractive alternatives. He may either gamble and fabricate the pieces into his system, or he may reorder and incur a delay of six to eight months (for devices built to high reliability specifications), with no assurance that that order - or its successor - will be free of this failure mode. The scheme herein discussed should allow high confidence in the metallization quality of an acceptable lot with little additional delay in the manufacturing cycle. The detection of a bad metallization lot reduces lost time to four to six weeks instead of thirty weeks. While these are benefits which accrue to the user, the paramount requirement of the scheme is that pertubation to the vendor's normal processing be minimal.

A nonstatistical sampling plan specifies the quantity and location sites of dice to be viewed by the SEM. These factors are determined by the order size, device (die) size, geometrical configuration of the slice holder in the metallizing chamber, number or type of metallizing source elements, and static or dynamic position of the slice holder relative to the source(s). The premise of the scheme is that opens or potential latent opens in the metallization at the oxide steps are batch process problems. This scheme,

---

[1] E.E. Maiden and R.D. Enquist, "Integrated Circuit Interconnect Metallization Failure Analysis." 12th Annual ASQC/CAL POLY Quality Control Conference, January 25, 1969, San Bernardino Section, California State Polytechnic College, Pomona, California.

then, is intended to detect poor metallization; it does not attempt to identify the failure mechanism, which might be due to improper metallization thickness, over-alloying, masking and/or etching characteristics or errors, effects of dopants, simple shadowing effects, etc. (Hopefully, description of the metallization by the SEM will eventually lead to vendors identifying the failure mechanism(s).)

The sampling plan is as follows. Following metallization, three to five slices from specific sites on the slice holder are identified. Processing proceeds normally through alloying and scribe-and-break operation. At this last stage, and before relative dice locations are lost, three to five dice from specific site locations on each of the previously identified slices are removed for viewing by the SEM.

Examination of the sample dice is made using a SEM at a minimum of 6000X magnification and at prescribed die orientation and angles of viewing. Judgment of metallization quality is made to written and pictorial standards, with examples of acceptable, unacceptable, and marginal metallization. Acceptance or rejection is made of the entire metallization lot, except for very small geometry dice and a high slice yield, in which case acceptance is on a slice basis.

A procurement specification has been written which identifies the purpose of the scheme, the advantages to user and vendor, the critical requirements, and acceptance-rejection criteria denoted above. It also specifies qualification of SEM operator techniques, SEM facilities (in-house or rented), and assigns responsibility to the vendor for judgment of metallization quality.

Vendor interest and active cooperation has been gratifying. Use of the procurement specification is expected in early 1970 by NASA-Goddard and its contractors, and is being recommended to other NASA Centers.

Key Words: Metallization, scanning electron microscope, screening, semiconductor devices.

1. Background

During recent years electrical failures have occurred in integrated circuits and transistors in spacecraft systems at the Goddard Space Flight Center, its contractors, or other laboratories, which are due to discontinuities in the interconnection metallization at the silicon-dioxide step surrounding a contact window. This problem has, in fact, been a recurring one throughout the semiconductor industry, notably appearing in both discreet transistors and integrated circuits with small geometry contact windows. The extent of the problem and its seriousness is ill-defined, although one user claims that "... contact window defects probably constitute the most significant single group of defects (excluding workmanship defects) causing failures of integrated circuit devices in electronic systems today."(1)[2] Optical microscopes are incapable of describing the problem characteristics and only the recent advent of the scanning

[2] Figures in parentheses indicate the literature references at the end of this paper.

electron microscope (SEM) has made it possible to identify, verify, and study the metallization anomaly. Most important, it should now be possible to correct the problem from a basic material and processing standpoint.

A number of these failures in bipolar integrated circuits (2) in the spring of 1968 precipitated considerable interest and efforts on the part of both users and vendors in defining and remedying the problem. While not all production lots contained devices which could and did fail, the problem was compounded because there was (and is) no known truly effective screening test for this anomaly in encapsulated devices. That is, the failure rate for tested devices appeared to be constant. When latent failures occurred, these devices normally had previously passed more than one, and many times several, electrical and environmental testings. They have been detected as early as in users' incoming inspections, and as late as in flight-qualified, packaged systems. (3)

Since the spring of 1968 a common procedure to guard against this failure mode in a lot has been to open a sample number of units and inspect them with a SEM. Should these samples shown evidence of opens or partial opens in the metallization at the oxide step, the user faces two unattractive alternatives. He may either gamble and fabricate the devices into his system, or he may reorder, thereby incurring a delay of six to eight months (for devices built to high reliability specifications), with no assurance that that order - or its successor - will be free of this failure mode.

2. An Inspection Test

The nonscreenable nature of this anomaly in encapsulated devices and the long fabrication cycle make imperative the need for early detection of bad devices. To this end a scheme has been devised which will test for metallization anomalies and which will be employed near the beginning of the manufacturing process. The premise of the scheme is that opens or potentially latent opens in the metallization at the oxide steps are batch process problems. This scheme is intended to <u>detect</u> poor metallization; it does not attempt to <u>identify</u> the failure mechanism, which might be due to improper metallization thickness, over-alloying, masking and/or etching characteristics or errors, effects of dopants, simple shadowing effects, etc. Hopefully, description of the metallization by the SEM will eventually lead to vendors identifying the failure mechanism(s) and implementing corrective procedures. The test inspection plan utilizes unique, non-statistical sampling plans, specifies instrumentation procedures, lists specific acceptance/rejection criteria, and defines the procedure for qualifying a vendor's facilities and personnel.

Although the most serious metallization processing mode of failure has been either a permanent or intermittent break as it passes over the relatively sharp edge of the oxide step, other metallization processing anomalies have been experienced by users, and some of these can also be screened for by this plan. The first of these additional anomalies is a dotting of the metallization pattern with a series of small irregular holes or thinned-out areas, best described as "swiss cheese" in appearance. If or when the holes are so situated that they meet tangent to one another across an interconnection, the result is an electrical discontinuity in the metallization pattern. Another anomaly has been the peeling of metallization along interconnect stripes. When severe this situation can result in open circuits. Even when less severe it can result in loosely attached metallic particles which under vibration or shock can separate and cause shorting. The last anomaly guarded against is that of abnormally thin metallization, (4) sometimes due to improper masking and/or etching.

If the intent of this scheme is fulfilled, it will:

a. Give reasonable assurance of the quality of the metallization in a particular procurement lot;

b. Eliminate most of the time heretofore lost by users when a delivered shipment proves to be defective (lost time is reduced from 30 weeks to 4 to 6 weeks);

c. Accomplish (a) and (b) with minimal perturbation to any vendor's normal process flow.

The importance of minimal disturbance to vendors' normal processing cannot be overemphasized. It is recognized that users probably cannot realize both points (a) and (b) unless this is so. The attraction for vendors is the relative ease of the scheme together with elimination of manufacturing costs incurred from working on bad lots from the beginning of the assembly operations, through electrical testing and screening to shipping.

3. Sampling Plan

A nonstatistical sampling plan specifies the quantity and location sites of dice to be viewed by the SEM. These factors are determined by the order size, device (die) size, geometrical configuration of the holder of the slices in the metallizing chamber, number or type of metallizing source elements, and static or dynamic position of the holder relative to the source(s).

The procedure to be followed in selecting dice to be viewed varies for discrete transistors and integrated circuits. Because of the very large number of transistor dice - 15,000 to 20,000 - on a 2 inch diameter slice, it is anticipated that many NASA procurement orders can be filled by 1 to 3 slices, assuming reasonably good yields. Therefore, qualification of metallization for transistor procurements is on a slice basis. For integrated circuit slices (or large geometry RF or power devices) die size is much larger and a single slice may contain only 20 to several hundred dice; yield is normally lower also. Typical size procurements of integrated circuits will thus require many slices, and therefore, qualification of metallization for integrated circuits is on a metallization run basis.

For discrete transistors, dice to be viewed under the SEM are selected in the following way. Metallized slices are selectively etched and sintered in accordance with the vendor's normal processing. Immediately following the scribe and break operation (see Figure 1, point 2), and before relative die location on the slice is lost by additional processing, five dice from specific locations on the slice are selected. Four dice are selected from near the periphery of the slice and 90° apart; i.e., one from near the edge in each quadrant. The fifth die is selected from the middle of the slice (see Figure 2). All five dice must be judged to have acceptable metallization. Unacceptable metallization on any one of the five dice is cause for rejection of that whole slice as a source of material to fill an order. Should a slice be rejected, a second slice from the same metallization run may be examined by selecting dice from it in the same manner. Should this second slice also be judged to be unacceptable, that entire metallization run is thereby designated unacceptable. A slice from another metallization run then is examined for metallization quality. This procedure is followed until sufficient slices to supply dice for the order are found which have acceptable metallization. The procurement order will be filled with dice from this slice or slices. Should the number of slices required to fill an order exceed five, an entire metallization run is to be qualified in the manner prescribed for integrated circuits.

For integrated circuits the procedure to be followed differs somewhat. Slice identification (traceability) at the metallization step is necessary (see Figure 1, point 1). Specifically, the relative positions of the individual slices in the metallizing chamber must be known. The normal procedure at this point is to record the identification of three to five slices in selected positions. Preliminary discussions with a number of semiconductor vendors disclosed a wide variety of metallizing systems which varied in size, metallizing source and type, geometry of the holder of the slices, and whether they were stationary or moving. For this reason the sample size and sample slice sites will necessitate individual negotiations with each vendor. For example, in the case shown in Figure 3, with stationary slice holder and six (6) point sources, the four slices shown at the edge, center and intermediate sites on the holder would be identified. Similarly, with a rotating slice holder and three (3) point sources as shown in Figure 4, a slice would be selected (identified) in each of the (three)

concentric circular rows. In Figure 5, four slices are identified which should be at "worst case" positions for a metallization system using a cylindrical slice holder and a line filament.

After slice identification processing procedes normally again until the scribe and break operation. Following this operation and again prior to further processing which would prevent determination of die location on the slice, five dice are selected from each of the previously identified slices. Dice selection is identical to that for selection of dice from a slice with discrete units. That is, four dice are selected from near the periphery of the slice and one die from the center (see Figure 2). The 25 dice thus obtained from the five slices are then viewed under a SEM. All 25 dice must be judged to have acceptable quality metallization. If not, that entire metallization run is designated unacceptable and no dice from this metallization run may be processed further to fill a procurement order. Other metallization run(s) are examined, following the same procedure for slice and dice selection, until an acceptable metallization run is found. The procurement order is filled from dice from this acceptable metallization run. If the order is for different device types from the same family (i.e., gates and flip-flops from LPDTL or gates and flip-flops from LPTTL, etc.), the vendor may mix slices proportionately as he determines necessary to achieve the required yield. Different device families (LPDTL, DTL, LPTTL, TTL, etc.) may not be mixed in a metallization run.

4. Critical Factors

A number of critical, interdependent factors affect judgments on quality of metallization. In order to establish standards and minimum requirements to be complied with by vendors throughout the industry, the inspection specification prescribes procedures and limits for the following factors: (a) SEM viewing parameters, (b) contact window identification, (c) magnification, (d) vendor qualification, and (e) acceptance/ rejection criteria.

4.1 SEM Viewing Parameters

The angle of viewing, or the tilt, is the angle between the primary electron beam and a line normal to the surface of the die (see Figure 6). This angle is extremely important in examining for metallization discontinuities because of the sharp angle (45° which is common in the contact windows at the oxide-steps). If the tilt angle is too low, an insufficient proportion of the edge of the oxide-step is visable for examination. For this reason the tilt angle should be no less than 60°, and angles of 70° are often advisable.

The second viewing parameter of importance is the direction of sample rotation. Discontinuities at contact windows occasionally are related to particular sides of the windows. That is, in many of the cases in which electrical opens have occurred at an edge (step) of the window, this edge was adjacent to the largest mass of the aluminum metallization stripe. Therefore, the viewing rotation should be such that the electron beam of the SEM is directed into the step of the window where the aluminum metallization leads up to the interconnection strip or wire bond pad. In Figure 7 the viewing direction shown by arrow "1" is incorrect. Direction "2" is normally the correct one, and direction "3" frequently is used to determine the quality of questionable metallization (cracks or tunnels) at the point indicated.

4.2 Contact Window Identification

Results of failure analyses indicate that electrical discontinuities resulting from metallization separation at the oxide step have normally occurred at base contacts in transistors and/or resistor contacts in integrated circuits. This is believed to result from the fact that, except in the ground contact to the substrate in integrated circuits, the steepest step in the oxide cuts for contact windows occurs at base or resistor contacts, rather than at emitter or collector windows. Thus, the site of examination for bipolar devices is the base contact window both for discrete transistors

and for transistors in integrated circuits. For the same reason SEM examination of MOS field-effect transistors or MOS integrated circuits is made at source or drain contact windows.

4.3 Magnification

Different metallization anomalies require different viewing magnifications. Inspections for "swiss cheese" effects or peeling should be made at a power of 1500X. Examination for opens at the oxide step or for thin metallization should be made at a magnification of 6000X or greater.

4.4 Vendor Qualification

Perturbation to vendors' normal processing would not be minimal if the procuring activity were to perform the actual SEM inspections and judge metallization quality. Delays would probably be frequent and lengthy, not under control of the vendor, and thus this situation would be intolerable. It is therefore necessary to individually qualify each vendor. This includes the SEM facility itself, whether in-house or rented, the evaluation of the experiences and techniques of the SEM operators who perform the examination, and designation of the appropriate person at each vendor facility with sign-off, or approval, authority. While the SEM is capable of high magnifications, great depth of field, and resolution much in excess of that available in conventional optical microscopy, these capabilities can be nullified by inexperienced operating personnel, laxity in properly maintaining the SEM, or inexperience or haste in interpreting the results (pictures). Obviously this infers expertise in SEM operation, maintenance and interpretation of results on the part of the first activity to procure devices submitted to this inspection, inasmuch as that activity would make the initial qualification. Expertise in picture interpretation and a thorough understanding of the viewing parameters is also necessary for all succeeding procurements from this vendor by other activities, in order to verify that standards are being maintained. The duplicate photographs supplied to the buyer as a part of each procurement enable this evaluation of continued qualification.

4.5 Acceptance/Rejection Criteria

Judgment of metallization quality requires the definition of what is acceptable and what is not. Some devices are clearly good, some are clearly very bad, and some marginal. Many marginal units probably would continue to perform satisfactorily, but confidence in their reliability over extended time would be low. The large percentage of devices examined at Goddard Space Flight Center are not in the category of being clearly acceptable. This requires that criteria be specified so that devices with minimal metallization problems are classified as acceptable.

In acceptable devices there should be no question that metallization (a) is continuous at the oxide step and along the interconnect stripe, (b) is of sufficient thickness, and (c) exhibits no peeling or flaking. Unacceptable devices unquestionably exhibit one or more of these anomalies. In marginal devices metallization voids at the oxide step, for example, may be very small and scattered, and justify categorization as acceptable. If the voiding characteristic is a crack of extensive or indeterminate length along the edge, judgment as unacceptable would be logical. Figures 8 through 19 demonstrate typical gradations of metallization quality which have been encountered. The detailed comments should explain the rationale used in their categorization.

Figures 8 and 9. The metallization in these examples is clearly acceptable. The aluminum is smooth and continuous over the oxide step with no sign of discontinuities. There is no evidence of pitting, peeling, or poor etching processes.

Figure 10. Although the metallization at this contact window has some irregularity along the oxide step, it is judged to be acceptable. Comparison should be made here with a similar device shown in figure 9.

Figure 11. This metallization has more serious irregularities along the oxide step of the contact window and is therefore judged to be unacceptable. Again, comparison should be made with figure 9.

Figure 12. Although there is some evidence of "opens" along the oxide step of the contact window in this example, this unit would be judged acceptable because there appears to be sufficient metallization to assure electrical continuity.

Figures 13, 14, and 15. At these windows very little, if any, metal is bridging the edge of the oxide step. These units would be judged unacceptable.

Figure 16. This is another variation of metallization which is unacceptable. The aluminum, either by design or misalignment, covered only three of the window edges. As often happens in that case, there was gross lateral etching which opened the metallization along the sides leading away from the uncovered edge. This, coupled with intermittent opens along the remaining side, formed the basis for the unacceptable rating.

Figure 17. The "undercut" metallization shown here resulted from an improper combination of etchant and photomask. This condition is well contrasted by comparing figures 17a and 17b. The device represented in figure 10b would be judged unacceptable.

Figure 18. Shown in this picture is an unacceptable die with the "swiss cheese" effect. The presence of a single void or multiple voids whereby a metallization stripe is reduced to less than two-thirds its design width is cause for rejection. In this picture voiding occurs in all metallization in the field of view, and there are many paths which may be chosen across the stripe width which fail to satisfy the two-thirds requirement. In other instances, where voiding incidence is lower or where clusters of voids occur, decisions to accept or reject will be more difficult, and vendor discretion in judgment, understanding of the problem, and agreement with this approach to screening for it are critical factors.

Figure 19. This clearly shows unacceptable metallization due to peeling. No peeling whatsoever is allowed.

4. Buyer Approval

When the vendor has judged the metallization of a slice or run to be acceptable, he is required to send 1:1 duplicates of the pictures upon which he based his decision (approval) to the procuring activity. (Pictures of dice from rejected slices and runs are <u>not</u> sent.) The pictures are to be accompanied by appropriate data, such as slice and run information, magnification, date of viewing, and formal approval, together with technical evaluation comments. The pictures and comments must clearly show the rationale for acceptance. This data package must be sent to the buyer no later than seven working days after being viewed and approved by the vendor, <u>not</u> several months later with the delivered devices themselves. In order not to unduly delay the vendor and his product line, the buyer, who has final approval, is required to notify the vendor within ten working days after receipt of the pictures if he disagrees with the vendor's acceptance and decides to reject the lot. Otherwise the vendor may assume approval by the buyer, and further processing may be resumed.

5. Implementation

Preliminary contacts were made with a number of semiconductor vendors to test their response to the scheme. The extent of interest and active cooperation on their part was surprising and most gratifying. A preliminary procurement specification was then written which embodies the comments and requirements denoted in this paper. This document has been sent to most of the semiconductor vendors and many users for comment before final revision and issue as a procurement specification. It should be available for use by June 1970. Expectation is that it will be recommended for use in procurements made by NASA/Goddard Space Flight Center and its contractors and to other NASA Centers.

6. Summary

Recent history of faulty metallizing processes and the nonscreenability of devices with poor metallization at oxide steps has prompted the use of the scanning electron microscope as a tool to detect poor metallization in the manufacturing cycle of semiconductor devices. Perturbation in vendors' normal processing is minimal and assembly operations on bad lots is obviated. At the same time the user avoids accepting bad lots and/or drastically reduces lost time (to refabricate a procurement), and has high confidence in the integrity of the metallization of approved lots. A procurement specification has been negotiated with semiconductor vendors and other users which completely describes the inspection plan, including its purpose, advantages to user and vendor, description of nonstatistical sampling plans, critical requirements, qualification of SEM facilities, operating personnel and tecniques, written and pictorial standards to be used for acceptance/rejection criteria, and assigns responsibility to the vendor for judgment of metallization quality. The document should be issued by June 1970 and will be recommended for use by NASA/Goddard and its contractors and to other NASA Centers.

7. References

(1) Maiden, E. E. and Enquist, R. D., "Integrated Circuit Interconnect Metallization Failure Analysis," 12th Annual ASQC/CAL POLY Ouality Control Conference, Jan. 25, 1969; San Bernardino Section, California State Polytechnic College, Pomona, Calif.

(2) Goddard Space Flight Center Failure Analysis Report, FAR 10-003, June 28, 1968.

(3) Anstead, R. J. and Floyd, S. R., "Thermal Effects on the Integrity of Aluminum to Silicon Contacts in Silicon Integrated Circuits," IEEE Trans. Electron Devices, April 1969, Vol. ED-16, No. 4.

(4) Goldberg, E. J. and Adolphsen, J. W., "A Failure Mechanism of Semiconductor Devices and Its Analysis," 6th Annual Reliability Physics Symposium, Nov. 6-8, 1967, Los Angeles, California.

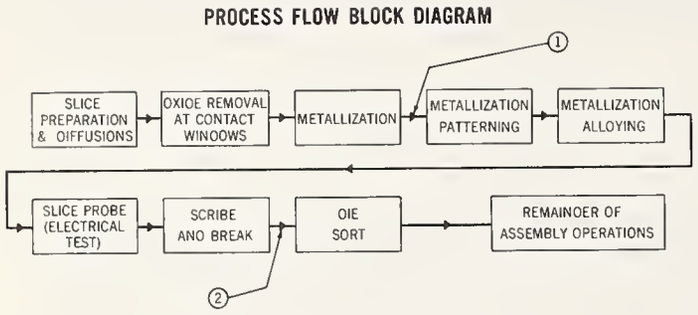


Figure 1

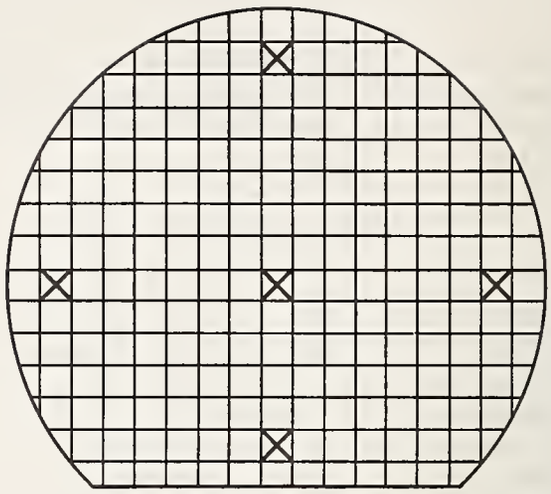


Figure 2

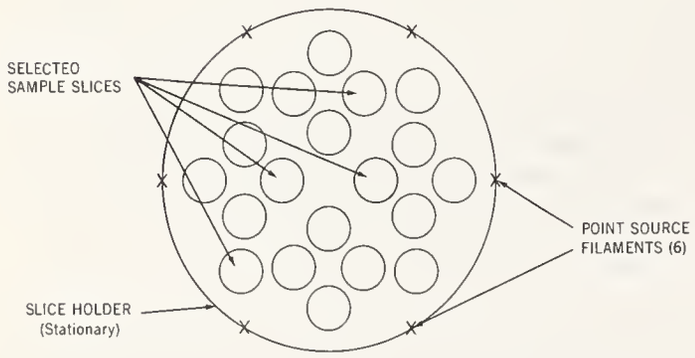


Figure 3

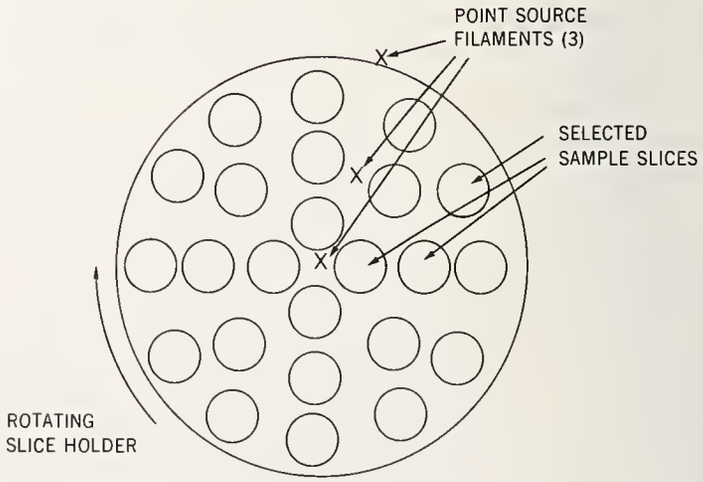


Figure 4

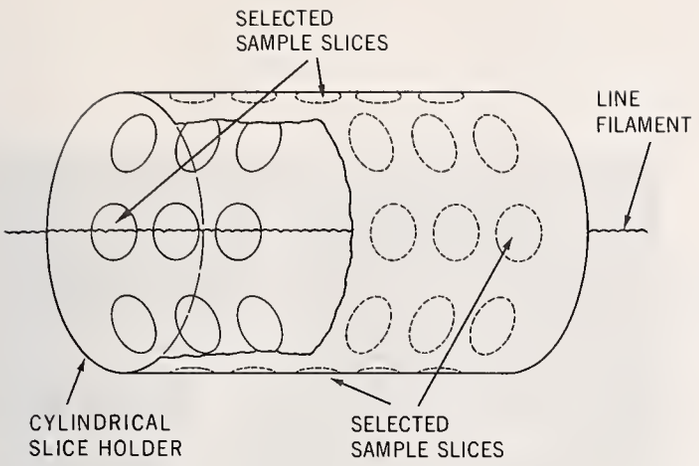


Figure 5

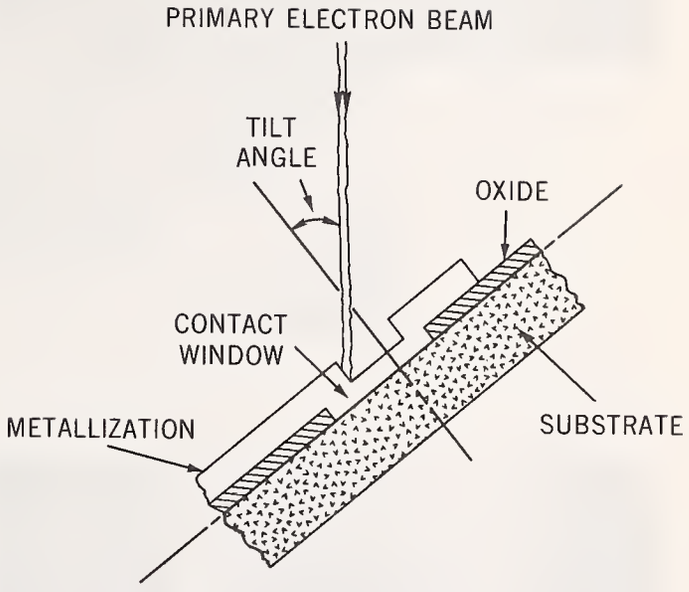


Figure 6

METALLIZATION AT A CONTACT WINDOW

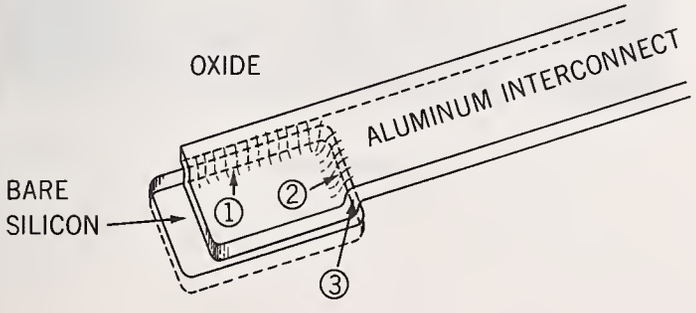


Figure 7

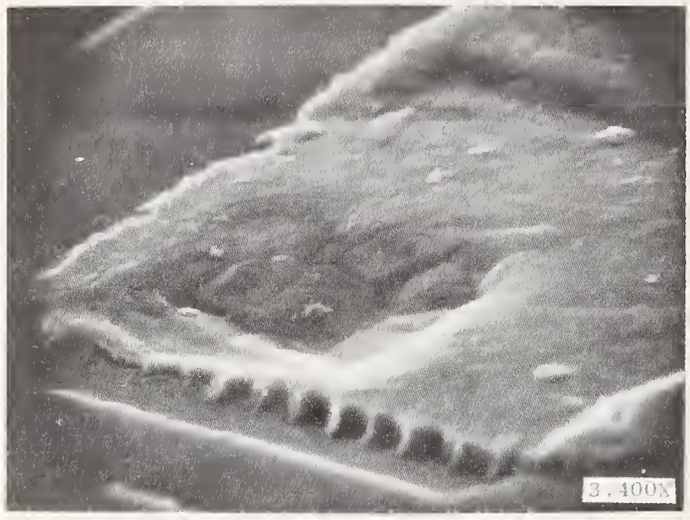


Figure 8

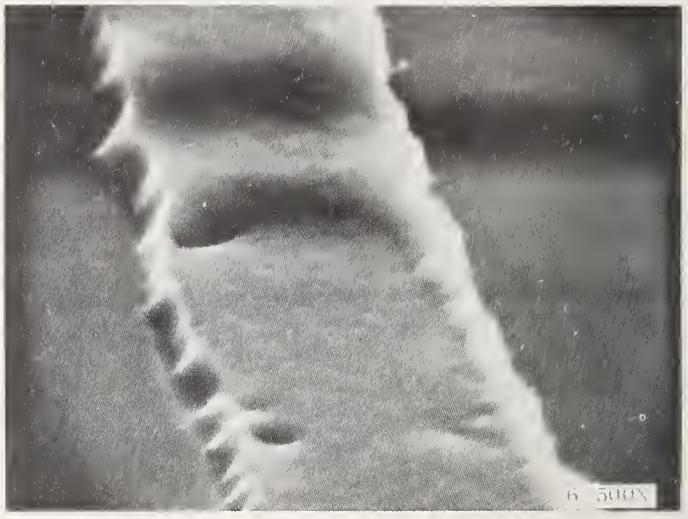


Figure 9

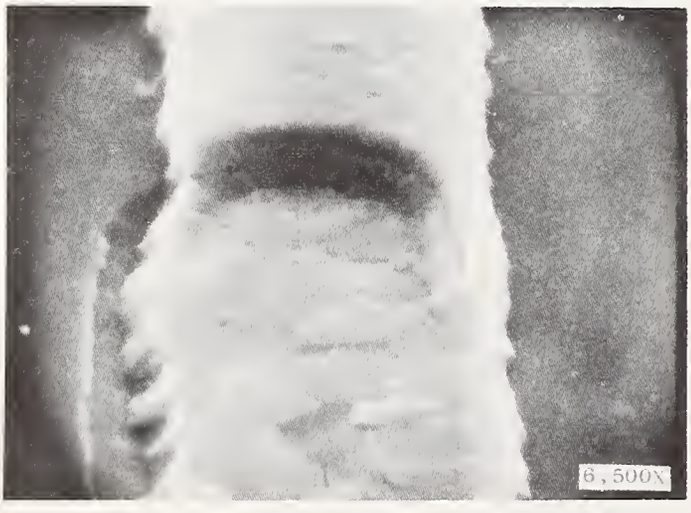


Figure 10

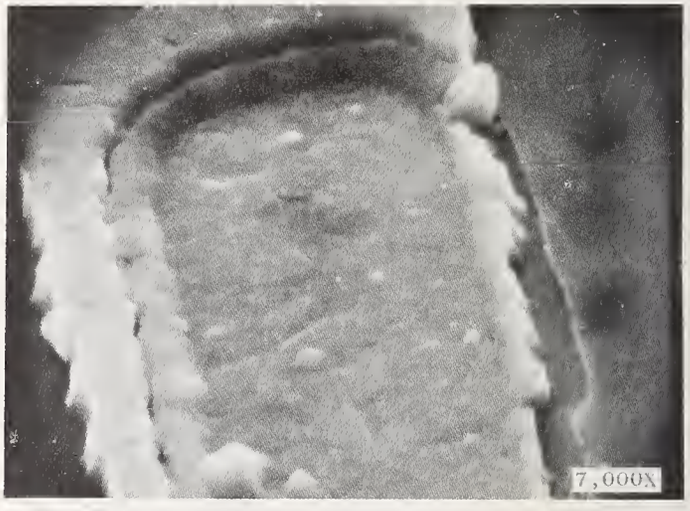


Figure 11

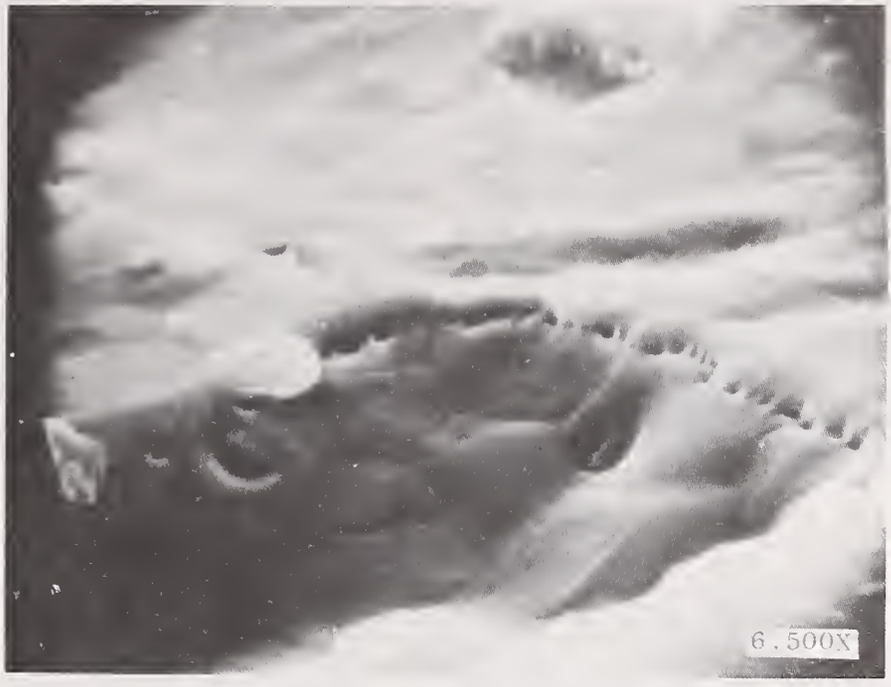


Figure 12

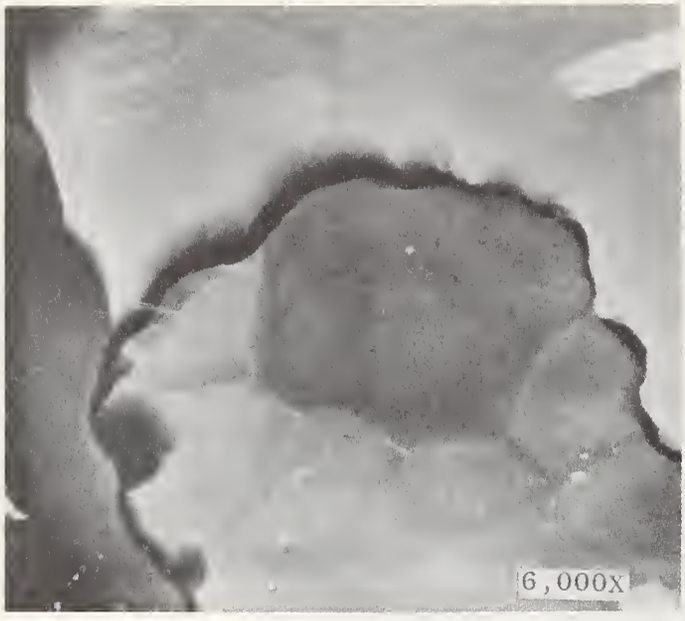


Figure 13

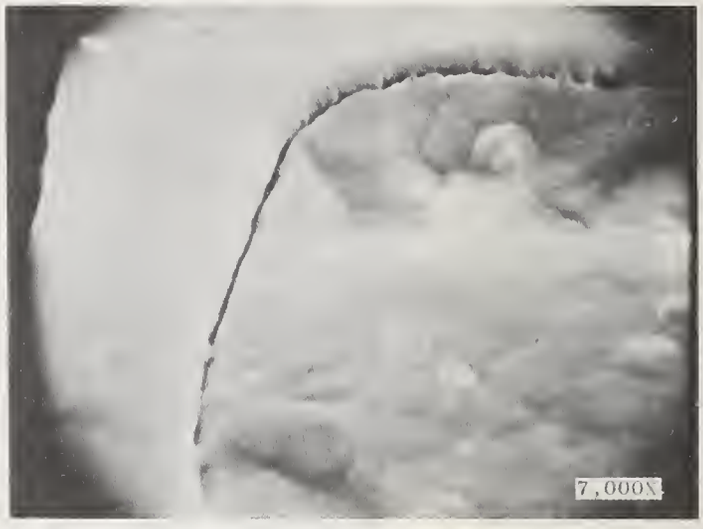


Figure 14

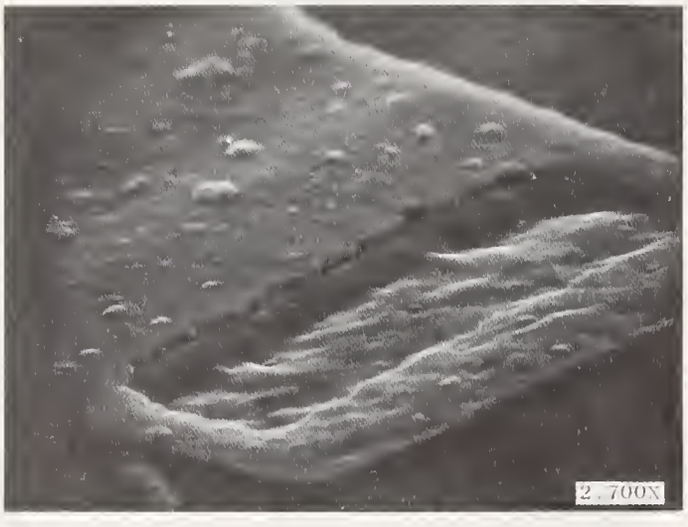


Figure 15

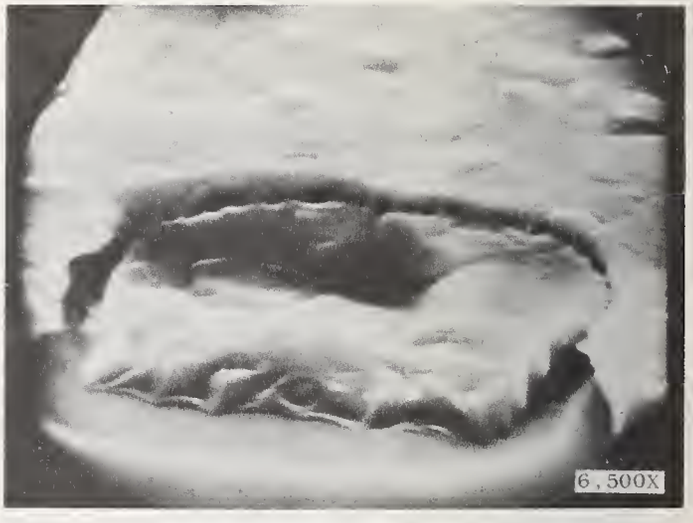


Figure 16

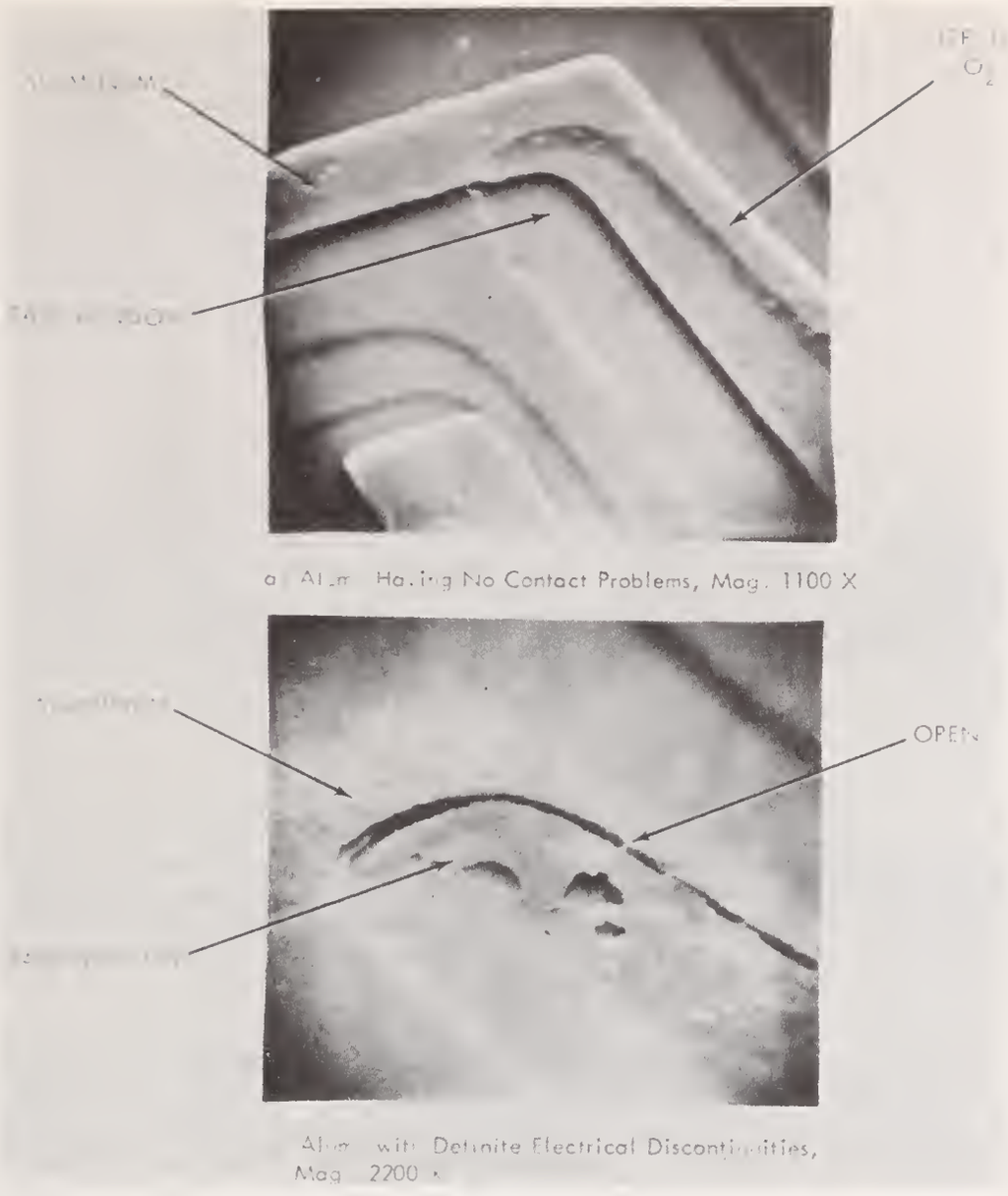


Figure 17

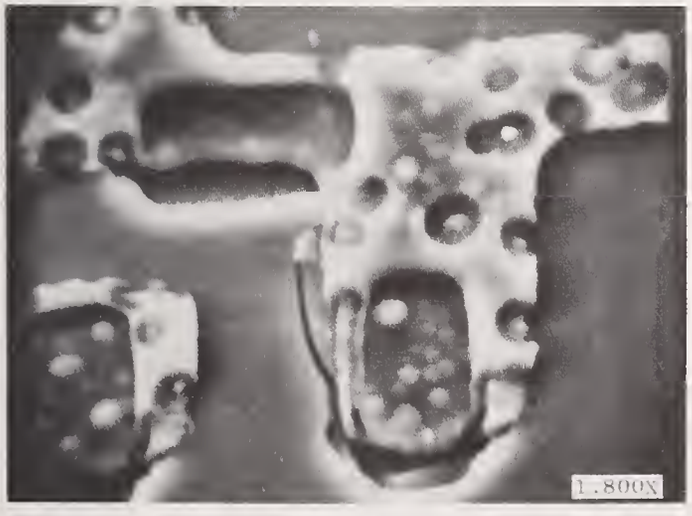


Figure 18

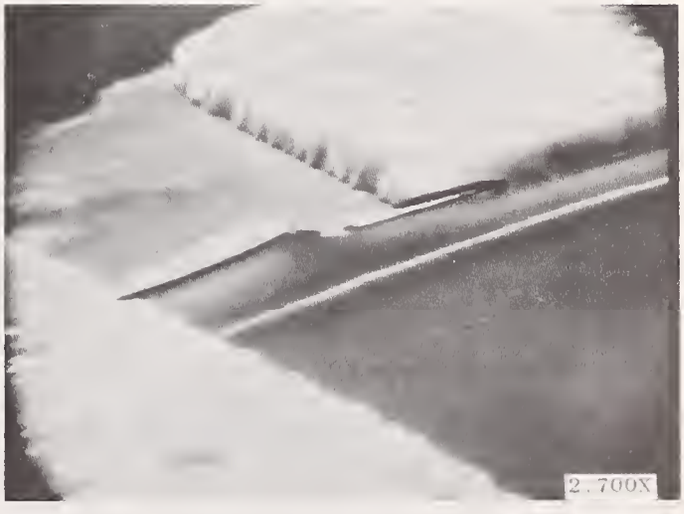


Figure 19

Metallization Deposition Parameters and Their Effect On Device Performance

J. R. Black
Motorola Inc.

1. Introduction

Aluminum films are extensively used as conductors for semiconductor devices and integrated circuits. These are deposited by condensation of the vapor in a vacuum ambient and etched into patterns using lithengraving techniques. The processing parameters employed for the formation of these films can greatly effect their quality which in turn effects subsequent processing steps and the reliability of devices. This paper briefly describes aluminum films deposition parameters and how they may effect film properties. These film properties as well as conductor design parameters are then related to various factors which can degrade the device properties or lead to device failure.

2. Vapor Deposition of Al Films

Generally an aluminum source is heated in a vacuum environment to raise its vapor pressure to the point where it quickly evaporates and condenses on cold surfaces to form aluminum films. A vacuum is required to prevent rapid oxidation of the aluminum either at the source, as a vapor or as a condensed film. Typical vacuum deposition equipment employs vacuum levels ranging from the low 10^{-4} Torr to the high 10^{-7} Torr level.

A good "rule of thumb" for vacuum work is that the approximate mean free path length of gas molecules is:

$$\lambda = \frac{5 \times 10^{-3}}{P} \text{ cm} \qquad \text{(Eq. 1)}$$

where λ is the mean free path length between collisions of gas particles in centimeters and P is the gas pressure in Torr.[1] For the common dimensions of evaporation equipment and at pressures stated above collision of a particle with the walls of the system is more probable than collision with other molecules. Thus, the aluminum particles can be considered to travel in straight lines from the source, to the walls of the vessel and to the work substrates. Under these conditions the possibility of chemical reaction between the aluminum vapor and the gas ambient is very low.

A second approximate but useful relationship to consider is the rate of arrival of residual gas within a partial vacuum to a surface as a function of the ambient pressure. This can be expressed as:

$$N_M = 4 \times 10^5 \, P \, \frac{\text{monolayers}}{\text{second}} \qquad \text{(Eq. 2)}$$

Where N_M = rate of gas arrival in monolayers/second
P = gas pressure in Torr

At a pressure of 2.5×10^{-6} Torr the residual gas arrives at a surface at a rate to deposit one monolayer of that residual gas per second. Also at a pressure of 2.5×10^{-4} Torr 100 monolayers of residual gas arrive at a surface each second.

Figures in brackets indicate the literature references at the end of this paper.

Typical aluminum film depositions are performed at a rate of about 1 micron per minute (or 166 angstrom units per second.) Since a monolayer of aluminum in the close packed plane occupies 2.5 angstrom units the aluminum is deposited at a rate of 67 monolayers per second. It is seen then, that at the elevated pressure ranges the rate of gas arrival to the work substrate can be comparable to the rate of arrival of aluminum atoms. Although not all of the gas molecules react with or are incorporated in the film it is clear that the film properties are dependent upon the deposition rate and vacuum level. High deposition rates at good vacuum result in the purest films. The reaction of fresh surfaces of aluminum with the ambient gas is observable by the dramatic drop in pressure during evaporation. Active gas molecules strike the surface and react chemically with the aluminum to form low vapor pressure compounds. Partial pressures of oxygen have a great effect on the properties of depositing films, however it has been reported that similar partial pressures of water vapor have only a slight effect.(2)

When aluminum vapor condenses on a surface, energy is given up in distinct steps. The vapor arrives at the surface with a temperature very close to that of the source aluminum. When it sorbs on the surface it releases energy equivalent to its heat of vaporization and forms what may be considered a two dimensional liquid. The atoms stick to the surface, but can travel over the surface some distance before they cool to the freezing point of aluminum, fall into a potential well in a growing crystallite, and release energy equivalent to the heat of fusion for aluminum. The higher the substrate temperature, the longer is the path over which the atoms can travel as a two dimensional liquid. Finally as a solid the atom cools from its melting point to the temperature of the substrate.

As aluminum atoms condense on a foreign surface they nucleate at many sites initiating crystal growth. The longer the migration path of the atom as a two dimensional liquid the lower is the density of these nucleating sites. Thus relatively large crystallites will grow on substrates held at elevated temperatures while small crystallites form on the cold substrates. At substrate temperatures in excess of 400°C crystallites as large as 10-12 microns form in films 5,000 angstroms thick. Grain growth is inhibited by impurities, thus the rate of arrival of aluminum vapor must be high compared to the rate of arrival of the residual gas ambient if large crystallites are desired.

3. Step Coverage

A metallization failure mode which recently has received attention concerns an electrical open at the step in oxide films where the metal drops to make contact with silicon. This failure mechanism has been shown to be due to a discontinuity which is formed during the deposition process and not during subsequent processing.(3,4) The discontinuity may initially form an open circuit, or later during operation an open would appear because of electromigration in the thin metal region or because of mechanical stress when thermally cycled.

Figure 1 presents a micrograph of a transistor which possesses metal defects at the two base contacts. The device was electrically stressed, fusing the thin metal at the defect. An enlarged view of one of these steps as shown in Figure 2 indicates that the long narrow defect section that was not fused originally was formed as an electrical open.

The metal defects at steps in the substrate are attributed to shadowing of the metal vapor by the film as it grows. This is shown in the diagram of Figure 3. Under the vacuum ambient employed during deposition of the film the metal atoms travel from the source to the substrate in straight lines. Vapor from source "A" deposits along the top of the step and along a portion of the bottom of the step; however, shadowing by the step prevents vapor from that source from reaching the step rise and that portion of the bottom of the step which is adjacent to the rise. Vapor from source "B" can reach the entire surface of the step if only source "B" were operating. The step riser is coated thinner than the step top or bottom due to the angle of incidence of the vapor. With both sources operating, that portion of the bottom of the step which received vapor from both sources grows more rapidly than the shadowed region. This step in film thickness shadows vapor from source "B" forming a region which receives no vapor. This process results in the formation of a long narrow void in the metal as shown in Figure 3.

To accomplish step coverage it is required that the angle of the step be reduced, thick films of aluminum are employed and that careful attention be paid to the geometry of the furniture within the vacuum station. The use of multiple sources and rotating

substrate holders to provide a diffused source aide step coverage but are not sufficient to assure adequate coverage. The use of elevated substrate temperatures which increase the mean free path of the two dimensional liquid phase of the condensing aluminum reduces the shadowing effect but to 300°C does not eliminate the defect. Figure 4 shows step coverage accomplished by a diffused source, a reduced step angle and a high substrate temperature. The large crystallites of the film are seen to extend over the step indicating excellent coverage.

Care must be taken when etching films with slight defects at the step. The etchant can enter the defect causing severe erosion of the aluminum along the defect. A drastic example of the results of this process is shown in Figure 5.

4. Electrical Resistance of Aluminum Films

Figure 6 presents a graph relating experimentally determined aluminum film volume resistivity as a function of thickness.[5] The films were deposited from tungsten filaments in the mid 10^{-8} Torr region on substrates held at 425°C. The thicker films possessed grain sizes of about 8 microns, however it was difficult to maintain the grain size as the film thickness fell below 4,000 angstrom units. Under the given deposition conditions to maintain near bulk resistivities the films should be at least 4,000 angstroms thick. d'Hurle et al have reported that as oxygen is added to the system a rapid increase in resistivity results.[2] This indicates that to obtain near bulk resistivity either a high vacuum must be employed or the film purity can be maintained through the use of rapid deposition rates. Films deposited at slow rates at high ambient pressures can exhibit resistivities several times that of the bulk.

5. Adhesion To Silica

The free energy change in the reaction

$$Al + SiO_2 \quad Al_2O_3 + Si \qquad \text{(Eq. 3)}$$

is -150 kilocalories per mole at 500°C indicating that the reaction will proceed. When aluminum is allowed to come into intimate contact with silica it competes with silicon for the oxygen in silica forming a strong chemical bond. The aluminum is chemically sorbed onto the surface resulting in a very high adhesive force. A monolayer of organic material will be sufficient to maintain the aluminum and silica apart and will result in low adhesion of the aluminum. Also several monolayers of water sorbed on the silica will react with the aluminum to form aluminum oxide. The interfaces of aluminum-aluminum oxide and aluminum oxide-silica in series in this case serve to attach the aluminum to the substrate. Highly adherent films are obtained when clean substrates are used and heat is applied to the substrates to drive off residual organic or water vapor comtaminating films. It should be noted that even under a vacuum, temperatures in excess of 150°C are required to drive off the water sorbed to silica.

6. Chemical Reaction With Silica

The chemical reaction between aluminum and silica (Eq. 3) is used extensively by the semiconductor industry to make ohmic contacts to silicon through the 20 to 30 angstrom thick layers of native oxide which form on freshly etched silicon. The process commonly is termed "alloying", however, this is a misnomer. The rate of reaction of the aluminum through the silica as a function of temperature is of interest to process engineers and is important to devices utilizing aluminum on thin oxide films.

Figure 7 shows a diagram of test structures that were employed to determine an equation for the rate of penetration of aluminum through silica.[5] Here aluminum was deposited onto a known thickness of silica on silicon. The samples were then heat treated at a given temperature for a prescribed time and the depth of penetration was determined by angle lapping the structure and by microscopic inspection. Figure 8, 9 and 10 present micrographs of beveled test samples that have been heated at 523°C for various time periods. The glass thickness was 9,000 angstrom units. The interface between the glass and the silicon is clearly seen and the wedge of glass as viewed with white light shows the blue fringe interference patterns. These figures show the increase of penetration into the silica as time proceeds. After 7½ hours the

reaction penetrated the 9,000 angstroms of glass.

A plot of the penetration depth versus time with temperature as a parameter appears in Figure 11, where the temperature was varied from 543°C to 421°C. The slope of the lines drawn through the experimental points gives the rate of penetration for the various temperatures. The log of the rate of penetration is plotted as a function of inverse temperature (Kelvin) to form the Arrhenius Plot presented in Figure 12. The equation derived from this plot is shown in the figure. The high activation energy of 2.56 electron volts indicates that the penetration rate is very sensitive to temperature.

It is interesting to note that aluminum on silica can melt at 580°C, a temperature well below its melting point of 660°C. At this temperature the aluminum rapidly reduces silica to form silicon and alumina. The silicon diffuses by solid state processes into the aluminum to form the eutectic composition of 12% silicon which melts under equilibrium conditions at 577°C.

7. Solid State Reaction With Silicon

Aluminum is a solvent for silicon with the solubility falling from a maximum of 1.6% silicon at 577°C to near zero below 200°C. When an aluminum film is deposited onto silicon and the couple is heated, the amount of silicon which dissolves into the aluminum by solid state processes is determined by the solid solubility and the amount of aluminum available.

A cross section of a typical ohmic contact to a silicon device is shown in the diagram of Figure 13a. When the couple is heated, etch pits will from in the silicon as the silicon dissolves into aluminum (Figure 13b). Small etch pits filled with aluminum appear under most of the contact region. Their size is limited because of the small volume of aluminum above them. At the edge of the contact region however, where the aluminum film rises and extends over the silicon dioxide film larger etch pits can form. The relatively large volume of aluminum of the stripe can consume a greater quantity of silicon and the source of that silicon is adjacent to the overlay stripe. The micrograph of Figure 14 shows a contact that has been heated and the aluminum subsequently etched exposing the pitted silicon. The larger pits are seen next to the overlay stripe (which has not been totally removed). Also of interest is the recrystallized silicon which formed preferentially at aluminum grain boundaries as the aluminum-silicon couple was cooled. When the aluminum was etched the small silicon crystallites which decorated the grain boundaries remained.

8. Conclusion

The quality of aluminum films greatly depend upon deposition parameters and post deposition treatments. Semiconductor devices manufactured without adequate engineering of aluminum processing contain unnecessary uncontrolled variables which may effect the device performance and reliability.

9. References

(1) Saul Duchman, Scientific Foundations of Vacuum Technique, Wiley and Sons, 2nd ed., (1962)

(2) F. d'Heurle, L. Berenbaum, R. Rosenberg, On the Structure of Aluminum Films, Trans. of the Metallurgical Society of AIME, Vol. 242, (March 1968)

(3) T. Yonagawa, I. Takikoshi, Failure Analysis of Evaporated Metal Interconnection at Contact Windows, Eighth Reliability Physics Symposium, (April 1970), Las Vegas

(4) A. Blech, J. Campbell, Jr., W. Shepherd, Microcracks in Aluminum Microcircuit Metallization, Ibid.

(5) J. R. Black, Metallization Failures In Integrated Circuits, RADC Tech. Report, TR-68-243, (October 1968)

10. Acknowledgments

The scanning electron micrographs of Figure 1, 2, 4, and 15 were made by Mrs. L. Millican of our laboratory while Figure 5 was prepared by Mr. J. Devaney of the Jet Propulsion Laboratory. Also, it is a pleasure to acknowledge the assistance of Mr. R. Mattox and Mrs. V. Rysso in the experiment on the reaction of aluminum with silica.

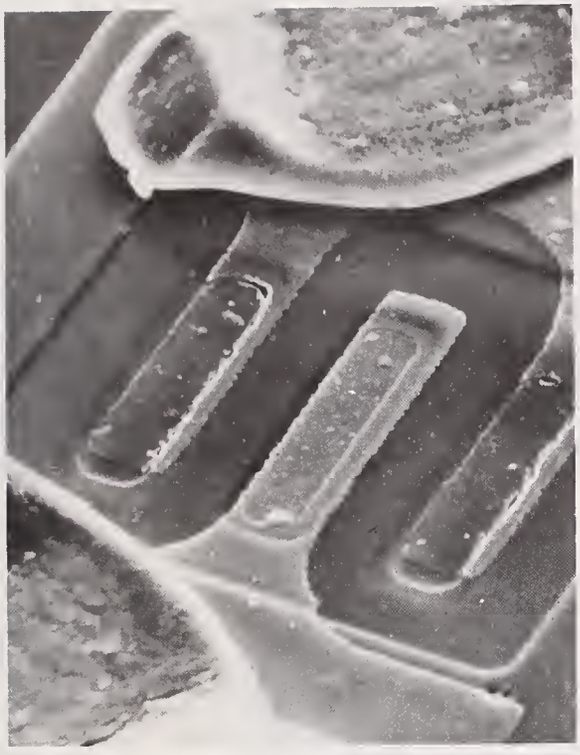


Fig. 1. Transistor with Base Metal Step Defects.

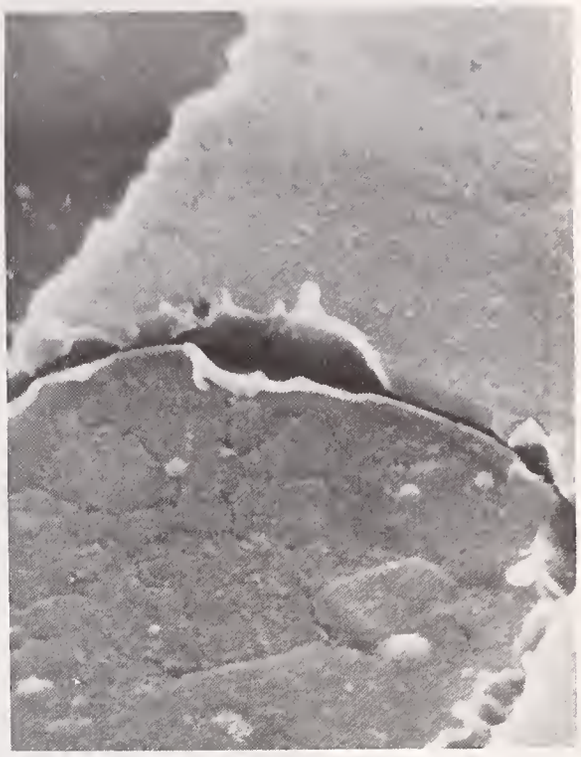


Fig. 2. Enlarged view of Metal Step Defect.

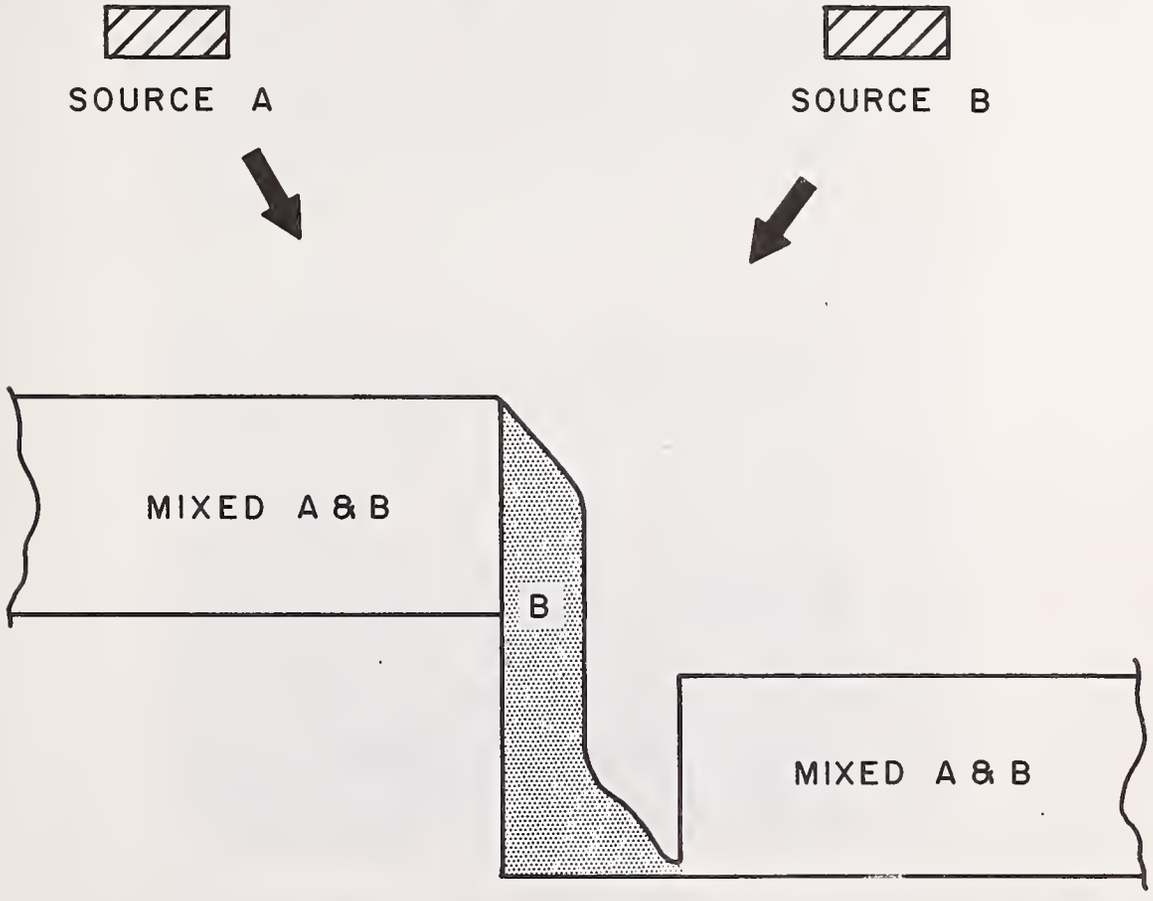


Fig. 3. Step Coverage From 2 Sources.

Fig. 4. Good Metal Step Coverage.

Fig. 5. Damage Due to Etchant Entering Step Defect.

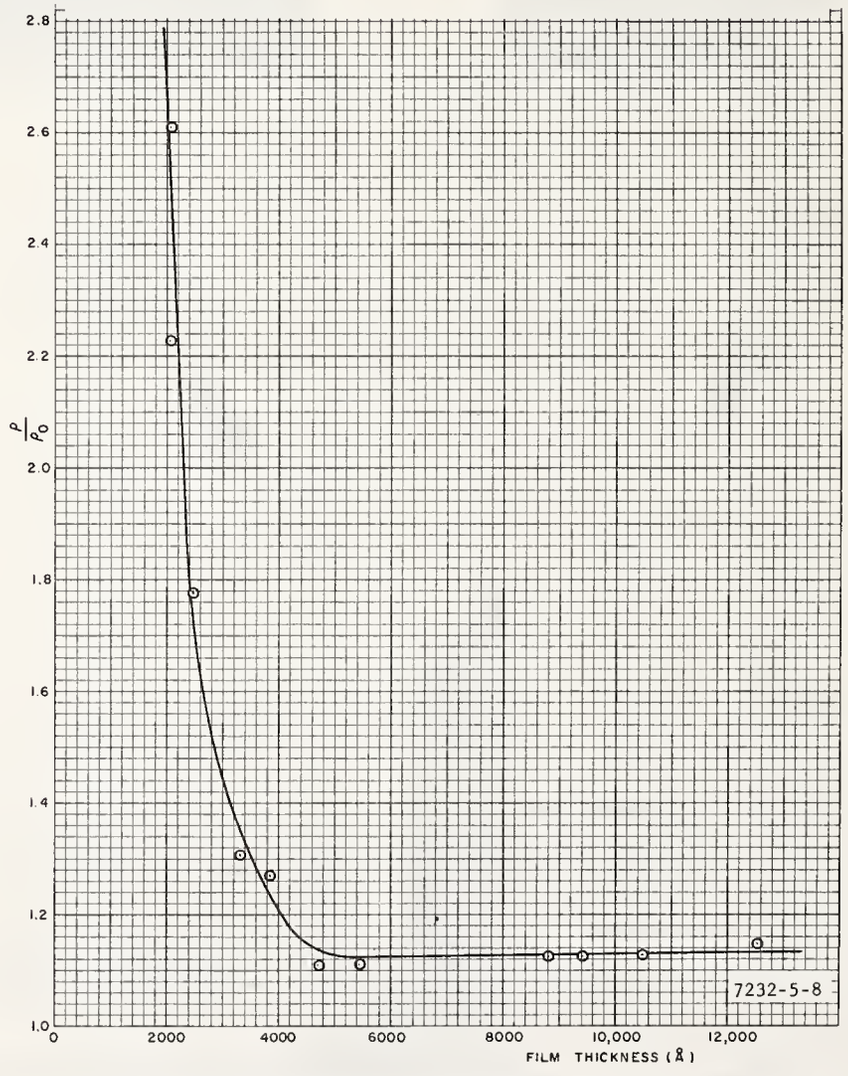


Fig. 6. Aluminum Film Resistivity Normalized to Bulk Resistivity as a Function of Film Thickness.

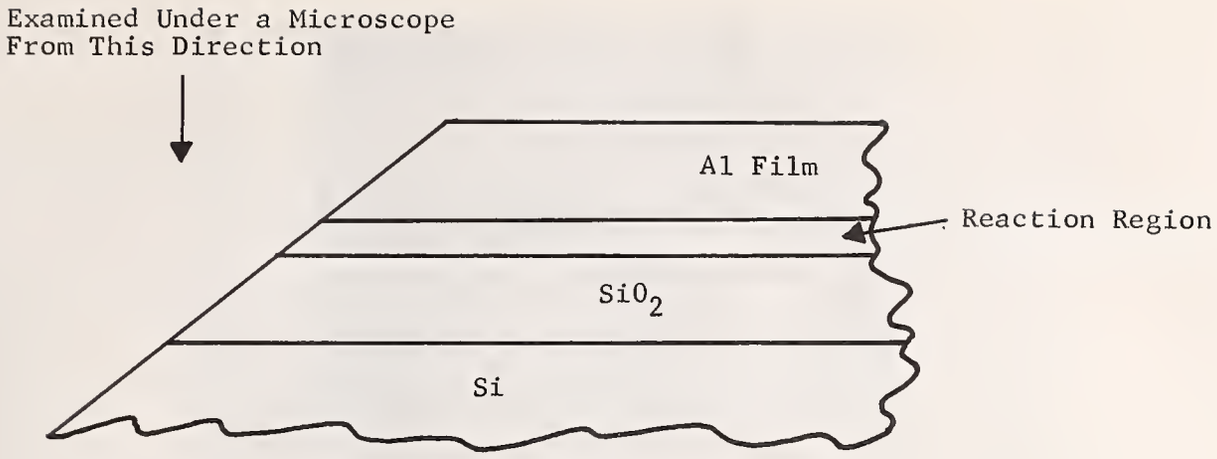
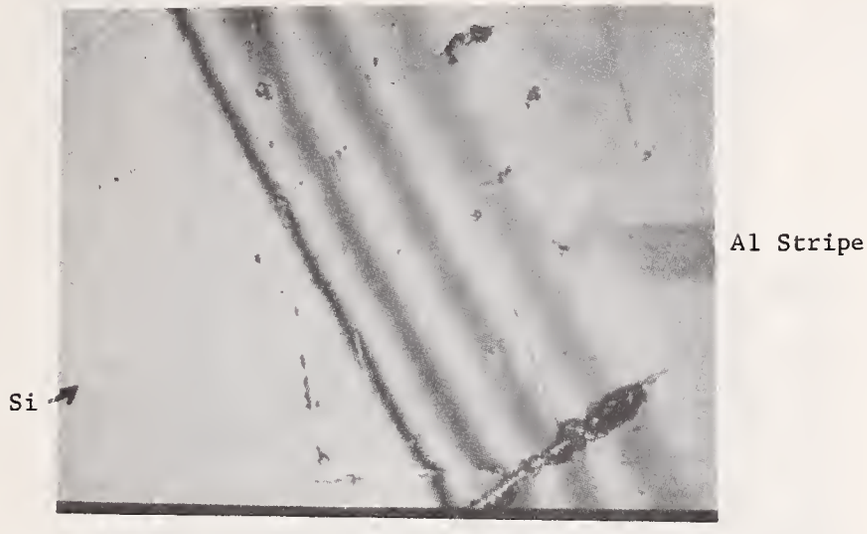
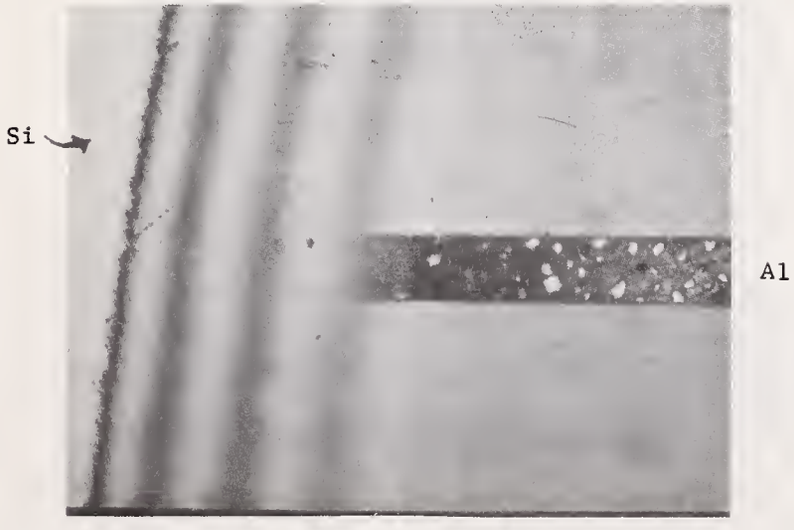


Fig. 7. Cross-Sectional Diagram of Beveled Al on SiO_2 on Si Sample

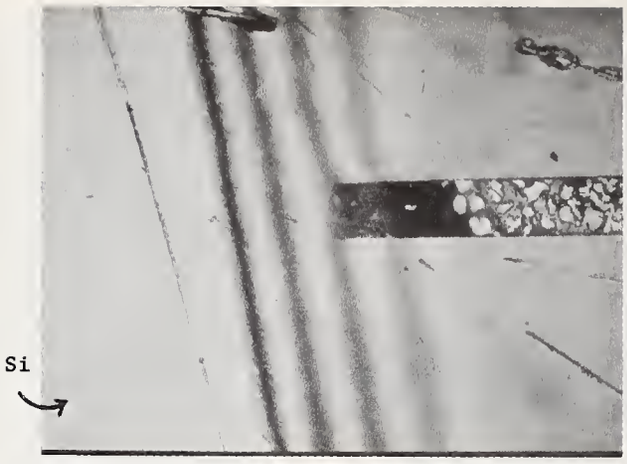


(270X) (a) 1-3/4 hours



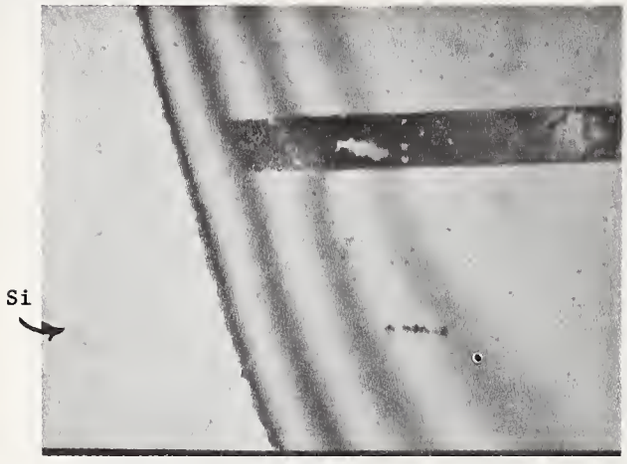
(270X) (b) 3 hours

Fig. 8. Beveled Al on SiO_2 on Si Samples Aged at 523 °C.



(270X)

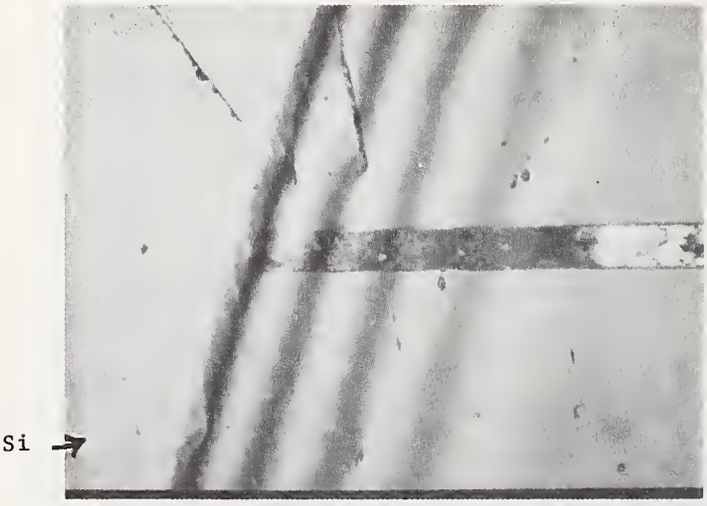
(a) 4-1/2 hours



(270X)

(b) 6 hours

Fig. 9. Beveled Al on SiO_2 on Si Samples Aged at 523 °C.



(270X)

7-1/2 hours

Fig. 10. Beveled Al on SiO_2 on Si Samples Aged at 523 °C.

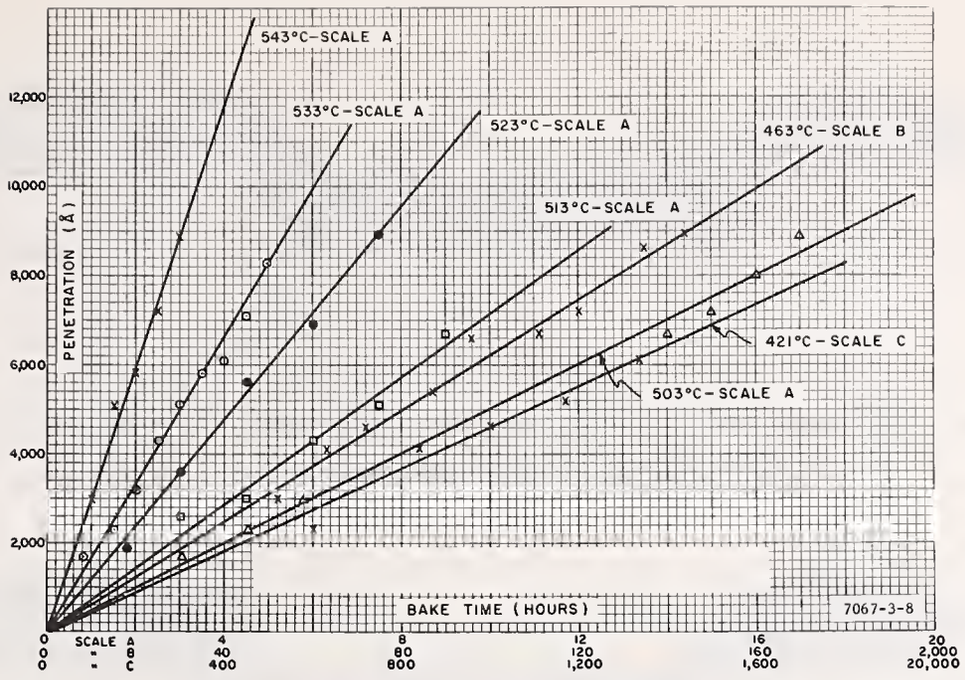


Fig. 11. Penetration of Al into Thermally Grown SiO_2 as a Function of Time and Temperature.

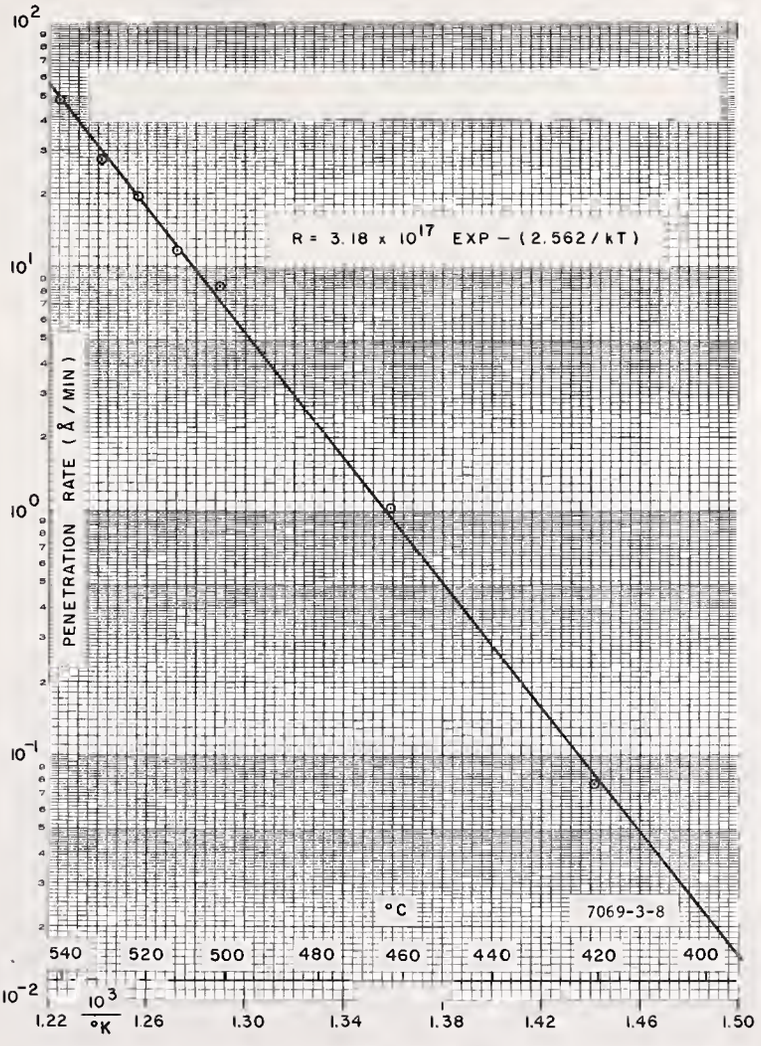
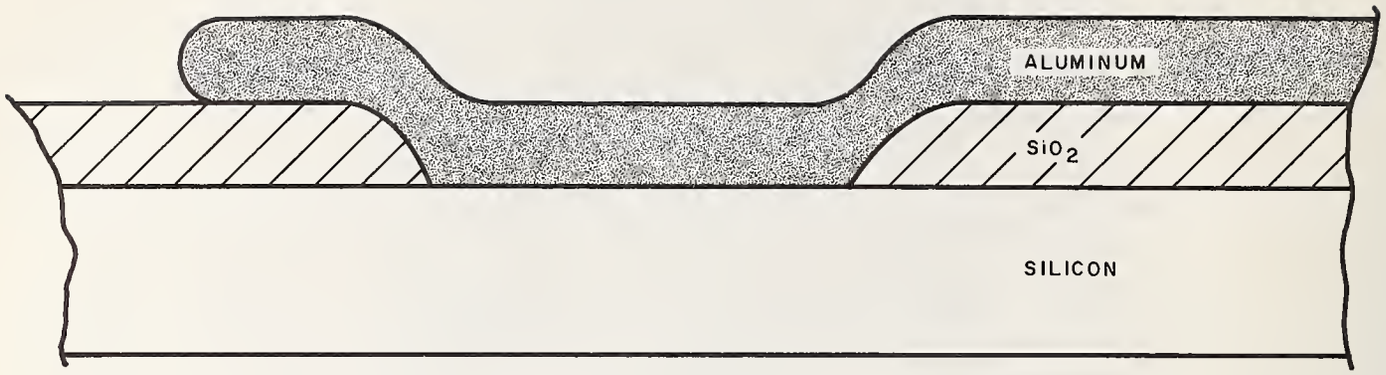
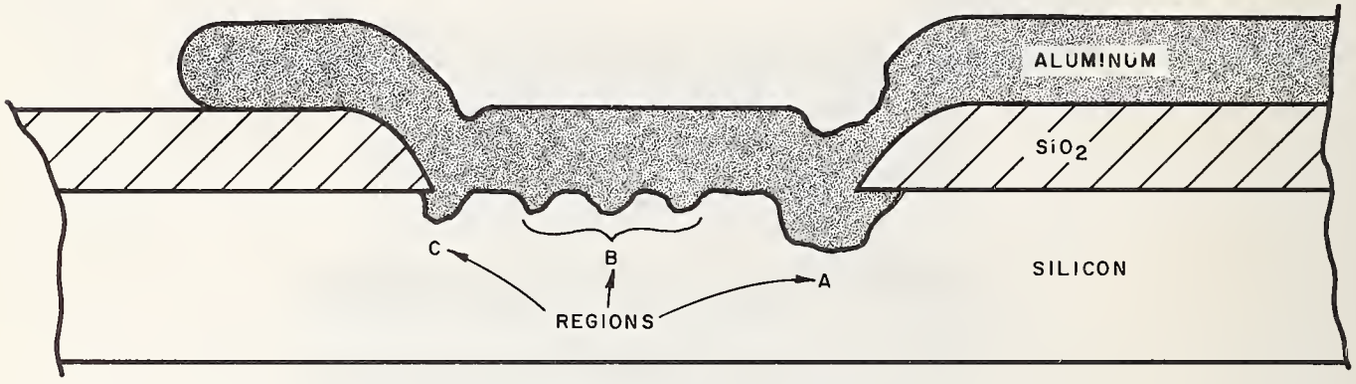


Fig. 12. Penetration Rate of Aluminum Reaction into Thermally Grown SiO_2 as a Function of Temperature.



(a) As Deposited



(b) After Sintering

Fig. 13. Silicon Etch Pit Formation by Dissolution into Al.

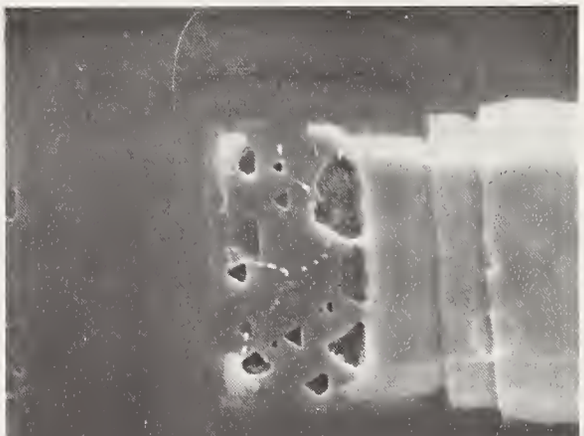


Fig. 14. Si Etch Pits Caused by Dissolution of Si into Al.

Methods for Determination of the Characteristics of Hyper-Pure Semiconductor Silicon and Their Information Content for the Device Production

Fritz G. Vieweg-Gutberlet

Wacker-Chemitronic GMBH
Burghausen,West Germany

Key Words: Silicon, Measurements, Resistivity, Lifetime Mobility, Crystal Perfection.

1. Introduction

The characterization of bulk silicon as a starting material for semiconductor device production is mainly based on the information about conductivity type, resistivity, minority-carrier-lifetime and crystal-perfection.

The methods for detection of the resistivity differ from one another in their basic structure and the information content. In the case of the lifetime measurement, even more than with the resistivity test, the methods and the information content therefrom, differ between measurements on the starting material and the results obtained by measurements later in the processing of the material, for example, after the first diffusion step. More than in the field of the electrical tests, the crystal perfection varies from the starting material due to defects introduced during the different device production steps.

In this paper we shall try to review the different methods for the description of the material and the information obtainable from these.

2. Conductivity Type

In the past, two methods became standard detection methods for type testing and are mostly applied to starting material; the rectification test and the hot-probe test, F42[1].

The application of one of these two methods depends mostly on the resistivity of the sample under test. The hot-probe test is more applicable in the range of lower resistivities, i.e., the range from substrate material up to 100 Ω-cm. The upper limit occurs because the number of charge carriers decreases with the increase of resisitivity, and the signal level decreases with the decrease of charge carriers. For higher resistivities, from 10 to 100 Ω-cm up to 10,000 Ωcm., the rectification test is more applicable, because the rectification effect is more effective on higher resistivity material.

In the special case of very high resistivity material(resistivity higher than 5000 Ω-cm)one can frequently obtain more reliable bulk conductivity type readings with the hot-probe test or a modification of this, the cold probe test, despite the reduced signal level. This occurs because on high resistivity silicon surfaces sometimes one can find a very thin layer of opposite conductivity type. Because the rectification test depends on the rectification characteristics of the surface of the sample under test, it will lead to wrong information in such a case. However, the value of the reading obtained in this hot-probe test depends on the number of free carriers which is very low in the high resistivity material. This requires the use of a very sensitive amplifier.

To obtain reliable readings, the temperature difference between hot and cold probe must be in the range of 50 to 100°C. With the normal cold probe at room temperature, the hot probe must be heated to values near 100°C. However, at 100°C one is reaching a temperature near the intrinsic range of silicon. To eliminate the effect of intrinsic conduction in some cases the cold-probe modification of hot probe test is more applicable. In this case the "hot probe" is at room temperature and the "cold probe" is cooled down to the liquid nitrogen temperature. The testing procedure is the same, but the temperature

[1]Published in American Society for Testing Materials Book of Standards, Part 8.

difference between hot and cold probe of the test equipment is greater than the temperature difference possible in the normal hot-probe test.

3. Resistivity

Two methods of measurement of resistivity of a given silicon sample are usually employed: the two-probe method F43[1],[1,2] and the four-probe method F84[1],[3,4]

The two-probe method is usually applied to detect the axial resistivity profile of a silicon rod. The four-probe method is mostly applied to silicon slices, but in some cases it is also used to detect the axial profile of a silicon rod mostly to the surface of the seed and bottom ends and to the cylindrical skin of the rod. Because the two-probe and four-probe methods differ significantly in the volume which is tested, the readings obtained differ depending on the radial profile of the ingot. In addition differences in readings obtained from a silicon sample by the four-probe test occur if the volume under test varies because of differences in sample size or point spacing of the probe.

It does not seem to be easy to correlate readings obtained from the two-probe method on the ingot and the four-probe reading obtained from slices with different thicknesses or diameters, or from four-probe measurements made with different point spacings. Correction factors are given in several papers (e.g., Smits,[5] Valdes,[3] and Swartzendruber[6]) to transfer the readings obtained from different methods and sample sizes to a "standard reading", but these are generally applicable only to homogeneous specimens.

Differences between the different test methods are caused by the nonuniformity of dopant distribution in a given silicon sample. The main variation which occurs is in the radial resistivity profile which is caused by the conditions during the crystal growth. It is easy to see that with different point spacings of the probe, different volumes are covered under test; a different way of integration leads to different mean values.[7] In addition to the large scale variation of resistivity along the radial and axial profiles, short range variations are observed in the form of "striations".[9,10] These inhomogeneous dopant concentrations are qualitatively detectable by etching techniques, or quantitatively by application of the spreading resistance method. Experiments directed towards reliable detection of the real variation factors are still going on. With the decrease of device size and the advent of large scale integration, or the increase of slice diameter (for high power devices) resistivity variations over the diameter of a starting silicon slice become more and more important. The extent to which "micro-inhomogeneities" such as striations may directly effect the device efficiency is not yet well-known. Therefore, future experiments and developments in measuring techniques leading to refinement of the measurement methods for highly localized areas will become necessary.

4. Lifetime

The most frequently used method for determining minority carrier lifetime of the parent material is the photoconductive decay method F26[1] applied to the "as-grown" ingot. Following many refinements of this method,[11] we obtained more reproducible and more reliable data regarding the original lifetime in silicon material. Consequently, and taking into account that the minority carrier lifetime depends very roughly on impurities (e.g., dopant concentration) and crystal perfection, it now seems to be very difficult to arrive at an absolute correlation between the minority carrier lifetime values measured in the parent material and the values measured after device fabrication steps (such as diffusion or any heat treatment with temperatures higher than 600°C). In the early days of silicon technology when the crystals contained very high dislocation densities, lifetime could be correlated with the dislocation density. With dislocation densities higher than 40,000 per square centimeter one could observe a very definite decrease of lifetime with an increase in dislocation density. Lifetime therefore, was a good indication of crystal perfection in this early stage of silicon technology. As technology develops in the direction of dislocation-free materials, lifetime measurements will become increasingly important as an indicator of the purity of the material. With the increase in the purity of the starting material, minority carrier lifetime after heat treatment and diffusion, is tending to become a problem associated with device technology (including gettering processes and/or lifetime killing processes such as gold diffusion) rather than a problem of the starting material.

5. Crystal Perfection

In the past, crystal perfection was one of the chief problem areas in device fabrication. An excellent method for the detection of crystal defects such as dislocations, or their arrangement in the form of slippage or lineage, is the etching technique. The most frequently applied method F47[1], is

chemical polishing of the test sample, and then etching it with a typical dislocation etch like Sirtl's solution[12] (solutions of the form $n CrO_3 + m H_2O + \ell HF$). This means of detection yields good information about dislocation density and crystal perfection.

However, with the transition from starting material with high dislocation density to dislocation-free materials, this problem area has lost much of its importance, and process-induced crystal defects may assume the status of the main problem. Process-induced defects may be viewed as defects beginning at the surface due to mechanical treatment during the slicing procedure, as well as dislocations and slippage introduced during heat treatment and diffusion. On the other hand, zero-dimensional defects ("point defects") like vacancies and clusters of them, may also become a serious problem in dislocation-free material. The extent to which they may interact with impurities, e.g., oxygen, is not sufficiently understood at this time. Moreover, it now seems that dislocation-free material may react more strongly with temperature differences than material having a significant dislocation density.

In addition to the aforementioned etching techniques, X-ray methods, e.g., the SOT-Technique, are very sensitive tools for the detection of crystal defects in both starting and processed material.[13,14]

6. Conclusion

At the present status of silicon technology, many fine measurement methods for resistivity, lifetime and crystal perfection are well known. But the sufficiency of these methods for characterizing the starting material for increasingly sophisticated devices is a question which scientists and engineers charged with the responsibility for evaluating the starting material are concerned. Many relationships between the characteristics of starting material and device efficiency may not be recognized with sufficient clarity. The relationship between resistivity and avalanche break-through voltage is well known, but the extent to which micro-inhomogeneities may affect the break-through voltage is not well understood. These considerations have to lead to the development of test equipment for the detection of micro-variations of dopant concentration. The same case obtains with minority carrier lifetime. Micro-variations of this characteristic, and the relationship between it and lattice imperfections and impurities must be explored.

Last but not least, process-induced lattice defects and impurities may change the characteristics of the material. For the development of more and more sophisticated devices, the relationships between starting material characteristics and process-induced defects will become more important than in the past. Relative to this, development of detection tools, such as spreading resistance and X-ray techniques, are a challenge to scientists and engineers chiefly concerned with materials and process specification.

7. References

[1] Zerbst, M. u. Heywant, W., ATM V3515-1 (Oct.1961).

[2] Vieweg-Gutberlet, F.,ATM V941-1 (April 1964).

[3] Valdes L. B.,Proc. of IRE, (Feb. 1954).

[4] Valdes, L. B., Transistor Tech.,Chapter 12.

[5] Valdes, L. B., D. van Nostrand Co. (1958).

[6] Swartzendruber, L.J., NBS Tech. Note 199, (April 15, 1964).

[7] Vieweg-Gutberlet, F. u. Schonhofer, F. X., ATM V 941-4 and V 941-5 (Oct.-Nov. 1966).

[8] Muhlbauer, A., Kappelmayer, R. u., Keiner, F., Z. Naturf. 20a 1089 (1965).

[9] Mazur, R. G., J. Electrochem. Soc. 114 (3)225 (1967).

[10] Vieweg-Gutberlet, F., Solid State El. 12 (1969).

[11] Vieweg-Gutberlet, F., ATM, V941-3(1966).

[12] Sirtl, E., u., Adler, A., Zeitschr. f. phys. Chemie 22 (1959).

[13] Schwuttke, G. H., et. al. Air Force Cambridge Res. Lab., Scientific Rep. No. 2.

[14] Meieran, E. S., Siemens Report, Eg 4/207e (Nov. 1969) and references therein.

Mechanical Damage - Its Role in Silicon Surface Preparation

R. B. Soper

Semiconductor Processing Co., Inc.
Hingham, Mass. 02043

Silicon is mechanically damaged during the centerless grinding, slicing, lapping and mechanical polishing operations used to shape the ingot into wafers. Surface damage, removed by chemical processes and peripheral defects such as conchoidal fractures, indents, and microcracks are discussed. A comparison of various polishing methods and how they relate to mechanical damage is given.

Key Words: Centerless grinding, etching, lapping, mechanical damage, peripheral damage, polishing, silicon, slicing, surface damage, surface preparation.

1. Introduction

The silicon surface quality sufficient for the manufacture of different devices varies widely; for instance, a simple diode may require only a surface obtained from the free etching of a sawn slice, whereas the slice used for the vidicon image tube should be of precise diameter, polished on both sides, free of any peripheral or surface damage, thin (four to six mils), flat and parallel.

To prepare the high quality silicon surface required for today's sophisticated devices, consideration must be given to the damage created by the necessary abrasive operations such as centerless grinding, orientation flat generation, slicing and lapping. Operations such as polarity probing, resistivity measurements, thickness measurement and handling with tweezers can also cause thermal or impact damage to the silicon surface. Abrasive, thermal and impact damage is best removed by chemical processes. The final polishing operation should leave the surface as damage-free as possible.

In the last decade mechanical damage induced in semiconductor materials has been investigated by many workers. (1-10)[1] There has been a difference in the depths of damage reported by different workers because of the variables involved in the abrasive operations and in the techniques used to determine the depth of the damaged layer. These points will be further discussed by others at this meeting.

The primary concern of this paper will be the damage created at the periphery of the wafer and how it can be controlled.

2. Peripheral Damage

Below is a description of the different types of mechanical damage which is located or is generated at the periphery of the wafer.

Conchoidal Fracture - A conchoidal fracture is a spalled flake of silicon. It may be shallow or deep, and multiple fractures may be on both sides of the wafer. In some cases partial spalling results due to incomplete fracturing.

Indent - An indent fracture is any irregularity from the normal profile of the wafer. It may be bounded by crystalographic planes or be random in shape.

Microcrack - A microcrack is a minor break which does not involve any appreciable separation of silicon. The depth usually penetrates the entire thickness of the wafer. The break normally follows a crystalographic plane, but the initiating force can be directed so as to yield a multi-directional break.

[1] Figures in brackets indicate the literature references at the end of this paper.

Scratch - A scratch is a very narrow surface groove caused by an oversized abrasive particle or other sharp object. Scratches can be classified into three categories:

a) An invisible scratch beneath a specular surface
b) A microscratch with shallow damage and
c) A macroscratch with deep damage which often causes wafer breakage.

Saw Mark - A saw mark is a definite damaged line which follows the curvature of the saw blade.

These damage defects often appear in combination with each other such as a microcrack extending from the edge of a conchoidal fracture or from the apex of an indent.

The peripheral damage is frequently de-emphasized because it appears outside of the complete device pattern; however, this damage is a prime source of high density dislocations and of ultimate breakage of costly wafers which have been processed to the device stage. A sawn slice with peripheral damage can harbor abrasives and other foreign materials which may be released during polishing and which in turn may scratch the surface of the wafer.

3. Shaping Operations

3.1 Centerless Grinding

The peripheral damage begins with the centerless grinding of the silicon ingot which is required when a uniform slice diameter is desired. Severe damage can be created if great care is not taken.

The technology of grinding has advanced a great deal in recent years, but it is still a complicated art which is not well understood. Grinding implies that the abrasive is fixed or bonded to the grinding tool; therefore, the silicon is abraded away by a cleaving action.

Centerless grinding machines use either abrasive belts or abrasive wheels. The latter is normally associated with more massive and stable machines. Some of the other variables in centerless grinding are type of abrasive, size of abrasive particles, surface speed of abrasive, type of coolant, thickness of silicon removed on each pass and, of course, the skill of the operator. These variables influence the degree of damage because large abrasive particles cleave out chips from the silicon surface; and high surface speed, improper coolant and gross thicknesses of silicon removed with each pass cause excessive frictional heat which drives the damage deeper into the crystal.

In properly ground ingots the damaged layer does not exceed 5 mils and can be removed by chemical etching. A slow etch rate is preferred since the heat generated by fast etches may propagate the damage still further into the surface. (11)

Orientation flats or notches are also generated by abrasive methods and must receive the same considerations as the ground cylindrical surface of the ingot.

3.2 Ingot Mounting

The next operation to be considered is the mounting of the silicon ingot on a fixture for sawing. This will vary with the machine and method used for sawing but usually the ingot is held in place with a wax or plastic. In using wax, it becomes necessary to heat the ingot above the melting point of the wax for good adhesion; but care should be employed to heat the ingot slowly and evenly to reduce the thermal gradient within the ingot, otherwise thermal damage may be caused - especially in any area in which mechanical damage exists. If a plastic such as an epoxy is used, the ingot need not be heated because the plastic will cure at room temperature even though it takes longer than at elevated temperatures.

Often times the etched ingot is abraded in the region where it is bonded to the mounting fixture so as to increase the adhesion. Of course, this abrasion introduces mechanical damage, and this procedure should be eliminated if possible. If the etched ingot is clean and free from soils deposited onto the surface during handling or storage and the proper mounting medium is chosen, the abrading step will not be necessary.

3.3 Slicing

The effect of peripheral damage is first observed at slicing. Obviously, peripheral damage can be initially caused by slicing, but often it is the result of the preceding operations. An improperly mounted and damaged ingot will create many slicing problems and defects.

Semiconductor Processing Co. receives from many customers as-sawn silicon wafers to be polished. The quality of the wafers vary widely. Some wafers are flat, parallel, free of peripheral defects and have no saw marks, where as other wafers have gross damage which extends throughout the thickness of the wafer. Bowed and non-parallel wafers must receive special consideration so as to insure the complete removal of all surface damage.

Both peripheral and surface damage is influenced by the sawing technique used. The conventional ID diamond wheel is capable of producing less damage than the OD diamond wheel. The Norton multi-blade abrasive saw and the various wire saws produce a minimum of damage because of the more gentle lapping action involved, but one must be aware of the possibility of abrasive being imbedded into the silicon surface. Electrochemical and chemical slicing may be the ideal approach in respect to damage produced, but as yet these techniques have not gone beyond the stage of laboratory investigation.

If the silicon ingot has been properly prepared and the slicing procedure has been optimized, the sawn wafer will have a minimum damaged layer. A signal that the damaged layer in much deeper is the observance of numerous edge defects.

Also there is evidence that the extent of damage differs between the two surfaces A & B (Figure 1) of the sawn wafer. The A surface, Figure 1, is more susceptible to damage as it is flexing while in intimate contact with the diamond saw. This is especially true as the slicing cycle approaches completion. Severe damage will cause the wafer to break away prematurely. It is in this area that the mounting material plays an important role by giving the sawn wafer structural support. As previously mentioned, when the ingot is abraded to improve the adhesion of the mounting material, the damaged region is subject to greater damage during slicing.

It is a common notion that between the two orientations (100) and (111), the (100) silicon is more easily chipped and broken than (111). The author has found that any difference in physical strength between the two orientations is of little significance in producing good polished wafers. What is more significant is how the silicon wafers were processed. Grossly damaged regions will contribute to breakage regardless of the orientation.

3.4 Etching

The etching of sawn silicon wafers before any further processing is highly desirable. The advantages are:

1. It removes surface damage.
2. It facilitates the inspection for peripheral defects.
3. It removes small silicon particles and residual sawdust which may lead to scratches during polishing.
4. It reduces the propagation of saw damage deeper into the wafer when abrasive lapping is required.
5. It generates a clean, smooth surface.
6. It decreases the polishing time required to remove the damaged layer.
7. It facilitates the cleaning of the final polished wafer.

3.5 Lapping

On occasion silicon wafers are lapped to improve flatness, parallelism, surface finish or to remove saw marks or to reduce thickness. Some of the variables in the lapping operation are type of machine (planetary, coplanar), abrasive size (2-25 micron), type of abrasive (diamond, silicon carbide, aluminum oxide, garnet), type of vehicle (oil, water), type of lapping plate (steel, glass), pressure, and previous history of the silicon wafers. These variables obviously affect the nature of damage generated.

In planetary lapping the wafer moves freely within the boundary of the cut-out in the thin metal carrier. Unless the cut-out is just slightly larger than the diameter of the wafer, the periphery of the wafer strikes against the metal carrier and create numerous chips.

In co-planar lapping the wafers are normally waxed to a plano, circular disk giving greater protection to the wafer. However, the wafers can easily be damaged if the disk is handled too roughly while positioning the disk onto the lapping plate.

There is always the possibility of imbedding abrasive particles into the silicon; consequently, the abrasive lapping process should be eliminated if at all possible.

3.6 Polishing

The polishing of silicon is accomplished by various methods:

A) abrasives

B) abrasives and chemicals reacting simultaneously and

C) chemicals.

In all three cases, the polishing media is administered to the silicon mechanically. The fourth method, D, is a chemical reaction not involving mechanical action such as slices immersed in a liquid etch. The term "Mechanical-Chemical polishing" is misleading because the word "mechanical" may refer to the machine or the abrasive.

Sub-micron diamond or aluminum oxide is an example of Method A because the silicon is abraded from the surface. Gross damage results by this method as evidenced by the scratch density observed after the surface has been lightly etched.

Zirconium oxide is commercially available suspended in an aqueous chemical solution. Since silicon is removed by the abrasive action of the zirconium oxide as well as the chemical action of the solution, this is an example of Method B. The degree of damage resulting from this method depends largely upon the actual processing technique but is considerably less than Method A.

The role played by colloidal silicon dioxide in an aqueous system is not fully understood, but it can be considered an example of Method C because the silicon dioxide particles appear to have little or no abrasive action on the silicon. The polishing rates are low but good surfaces are obtained with a minimum damaged layer. Another example of Method C involves the cupric ion. (12-13) The cupric ion, in a fluoride solution, displaces the silicon atom; and the thin copper layer is subsequently removed by the mechanical action of an appropriate pad on the polishing wheel. Two advantages of the cupric ion process are the fast removal rate and damage-free surface.

Conventional acid etching and vapor etching are examples of Method D.

All of the previous abrasive and chemical processes are in preparation of the final polishing step which should leave the silicon surface damage free. The ideal polishing method is one using only chemicals which will reveal damage in the form of surface defects and blemishes. One such method is the cupric-ion process which produces a polished surface only after all of the damaged layer has been removed. Microcracks associated with peripheral indents or conchoidal fractures are easily revealed with this polishing method because the chemical activity is greater in these damaged areas. In contrast when abrasive or abrasive-chemical polishing methods are used, the microcracks may not be optically visible. The peripheral defects also act as sites to harbor abrasive particles which may be released during the polishing cycle and cause surface scratches.

4. Wafer Handling

During the various shaping operations, the wafers are subjected to many cleaning and etching steps with subsequent turbulent water rising. Proper baskets or other fixtures should be chosen to prevent the wafer edges from coming into sharp contact with hard surfaces such as glass or metal. The contact may be severe enough to cause peripheral damage.

5. Summary

In summary the presence of peripheral damage in a polished silicon wafer can be related to wafer breakage during device processing. Surface damage can be related to an irregular surface obtained in an epitaxial growth process which in turn affects the quality of photo masking. Surface damage may also lead to irregular diffusion and alloying depths causing electrical shorts and high leakage currents.

Abrasive shaping operations should be controlled so as to minimize both surface and peripheral damage and this damage should be chemically removed before any further abrasive operations are performed.

Figure 2 shows the many processing routes that can be taken to prepare silicon surfaces for devices.

6. References

(1) Pugh, E.N. and Samuels, L.E., A Metallographic Investigation of the Damaged Layer in Abraded Germanium Surfaces, J. Electrochem. Soc., 108, 1043 (1961).

(2) Sticker, R. and Booker, G.R., Nature of the Damaged Layer on Abraded Silicon Specimens, J. Electrochem. Soc., 109, 743 (1962).

(3) Pugh, E.N. and Samuels, L.E., Etching of Abraded Germanium Surfaces with CP-4 Reagent, J. Electrochem. Soc., 109, 409 (1962).

(4) Stickler, R. and Booker, G.R., Transmission Electron Microscope Investigation of Removal of Mechanical Polishing Damage on Si and Ge by Chemical Polishing, J. Electrochem. Soc., 111, 485 (1964).

(5) Maruyama, S. and Okada, O., Crow Track Formed by Mechanical Force on Silicon Crystal Wafer, Japan. J. Appl., Phys., 3, 300 (1964).

(6) Faust, J.W. Jr., Factors That Influence the Damaged Layer Caused by Abrasion on Si and Ge., Electrochem. Tech., 2, 339 (1964).

(7) Pugh, E.N. and Samuels, L.E., Damaged Layers in Abraded Silicon Surfaces, J. Electrochem. Soc., 111, 1429 (1964).

(8) Stickler, R. and Faust, J.W. Jr., Comparison of Two Different Techniques to Determine the Depth of Damage, Electrochem. Tech., 4, 399 (1966).

(9) Meek, R.L. and Huffstutler, M.C. Jr., ID-Diamond-Sawing Damage to Germanium and Silicon, J. Electrochem. Soc., 116, 893 (1969).

(10) Whitten, W.N. Jr., Heitz, A.J. and McNamara, J.E., Depth of Work Damage Resulting From Shaping Operations on Silicon, E.C.S. Meeting, Washington, D.C., Oct. 1964, Abst. 162.

(11) Fairchild Semiconductor, 5th Quarterly Report, Section III, Task 2, Exhibit I, U.S. Army Electronics Material Agency, Contract No. DA-36-039-SC 86726 (1963).

(12) Regh, J. and Silvey, G.A., Silicon Planar Chemical Polishing, Electrochem, Tech., 6, 155 (1968).

(13) Mendel, E. and Yang, K., Polishing of Silicon by the Cupric Ion Process, Proc. IEEE, 57, 1476 (1969).

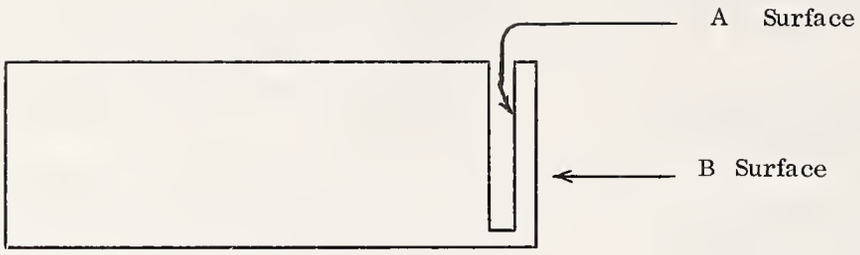


Fig. 1. Identification of the two sawn surfaces.

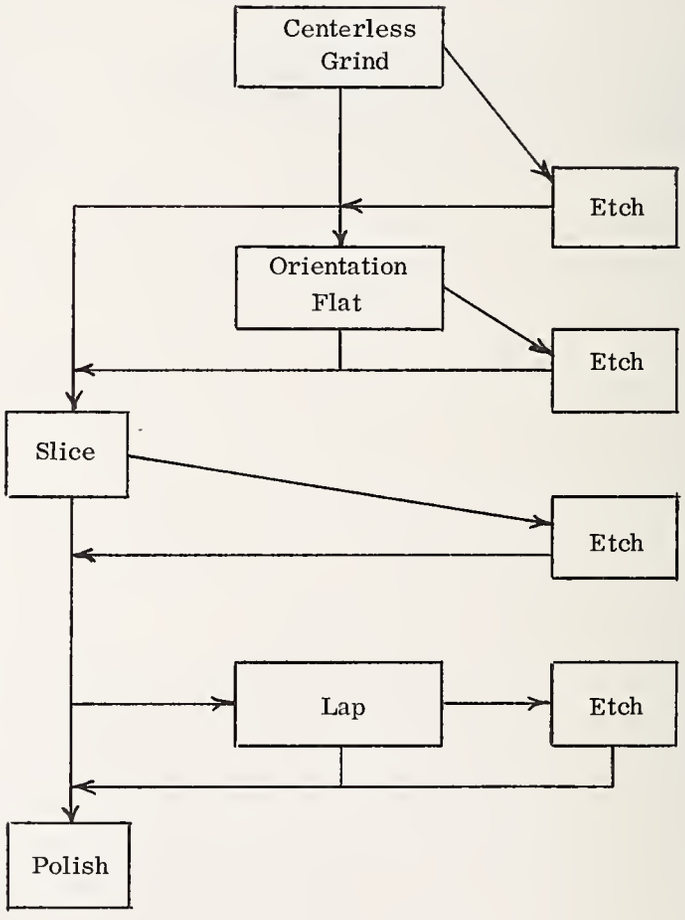


Fig. 2. Processing routes for the preparation of silicon substrates.

Crystallographic Damage to Silicon by Typical Slicing, Lapping, and Polishing Operations

T. M. Buck and R. L. Meek

Bell Telephone Laboratories
Murray Hill, New Jersey

This paper surveys some of the literature on surface damage on silicon resulting from shaping operations such as sawing, lapping, and polishing. Topics covered include depths of damage by various processes, techniques for detecting damage, effect of residual damage in high temperature operations, the present picture of the microscopic nature of abrasion damage, proposed mechanisms of fine polishing operations, a procedure for achieving a smooth, relatively damage-free surface (by silica-sol polishing), and recent results of ion scattering experiments on polished surfaces.

Key Words: Abrasion damage on silicon, cutting silicon, depth of damage, detection of damage, lapping silicon, polishing silicon, silica-sol polishing, silicon surfaces, surface damage.

1. Introduction

The manufacture of most silicon devices begins with shaping operations such as sawing, lapping, and polishing. These mechanical treatments produce a damaged surface layer which must be removed before further processing since its continued presence may seriously compromise performance requirements of finished devices, or prevent attainment of an economic yield.

A detailed explanation of the nature of the damage which would quantitatively explain all of the changes in physical and chemical properties which have been observed at mechanically damaged silicon surfaces has not yet been possible. Macroscopic efforts such as chipping, cracking, embedded abrasive particles, and elastically strained regions are apparent. However, these do not seem to explain, for example, the observed large increase in effective surface recombination velocity (i.e., low minority carrier lifetime) at a damaged surface (1). It has, therefore, been proposed that microscopic defects such as dislocations, vacancies, and "dangling" bonds at the new surface are also produced and some experimental verification of their existence, which will be further discussed below, has appeared.

Damage introduced by mechanical shaping may be expected to interact with impurities and defects existing in the bulk crystal if the wafer is later subjected to high temperature treatments. Koehler (2) has discussed the alteration in the distribution of defects and impurities produced by a stress field and has shown that two effects are of importance: (1) equilibrium defect density is changed, and (2) the height of the activation barrier for defect migration may be altered. The latter could result in defect motion at lower temperatures than might otherwise be thought possible. Such effects become especially important when diffusion or oxidation is performed on wafers retaining some mechanical damage. Dislocation formation takes place when damaged surfaces are annealed, and other defects, which are thought to be stacking faults, are observed when lightly damaged (Lustrox or diamond polished) wafers are oxidized (3,4). Clearly an increase in dislocation or stacking fault density or an impurity segregation may have deleterious effects on device properties.

John (5) has reviewed some of the device malfunctions which may be traced to grown-in or process-induced defects and impurities. Dislocations, precipitates and stacking faults, which may arise from residual abrasion damage, appear to be especially important (5-8) in particular when associated with oxygen. The associated device defects include: (1) decreased reverse breakdown voltage, (2) increased reverse leakage current, (3) excess forward drop, (4) microplasma breakdown at precipitates and inclusions, and (5) excess noise. Poor reverse junction characteristics have been correlated with dislocations introduced by mechanical shaping (9), and enhanced diffusion along such dislocations will reduce the yield of narrow base transistors (10). Dislocations may also be expected to affect field effect devices through the surface recombination velocity and the mobility (11). The economic impact is that

[1] Figures in parentheses indicate the literature references at the end of this paper.

the greater the density of defects which cause a device to fall below specifications, the greater is the probability of a given device on a wafer having such a fatal defect, and, consequently the yield is lowered.

2. Surface Damage Studies

We have not attempted a complete tabulation of results that have been published on depth of abrasion damage by various treatments, as determined by different techniques. Rather, Table 1 gives results which are fairly representative for four different abrasion treatments: fine polishing, 5 micron lapping, 25 micron lapping, and saw damage. Other compilations can be found in references 1 and 12 through 16.

Several observations can be made about the data of Table 1. 1.) Depth of damage increases with nominal particle size of abrasive, although there is considerable variation among values for one particle size, 25μ. 2.) In only one case (item 4) was the depth of damage greater than the nominal particle size and in that exception polishing was done on the dry diamond powder. However, the results tabulated refer to reasonably gentle application of the abrasive; excessive pressure or heating during abrasion can undoubtedly cause deeper damage. 3.) Metallographic techniques which detect individual flaws, usually give larger values than the techniques, e.g., PME and etch rate, which determine an average property over a large area, although there are some exceptions. 4.) Method of abrasion is influential, e.g., items 11 and 12, 13 and 14. Also, the crystal orientation seems to have an effect, items 15, 16, 17. 5.) The depth measurement may make a small difference. For example, in the taper-sectioning method, items 13 and 14, depth is measured from the outermost crests of surface irregularities to the extremities of the deepest damage features, whereas the weight-loss measurement, used in the PME and etch-rate techniques, assumes bulk density in the damaged layer.

Progressive etching and measurement of an average property sensitive to damage is illustrated in Fig. 1, which shows change of photomagnetoelectric voltage as the damaged layer is removed by etching from the illuminated surface of a silicon slice (18,19). Chemical effects on recombination velocity are important in this case. They cause scatter beyond the knee in the curve. Special treatments can be used to assure low S in the absence of damage (24,25). Fig. 2 illustrates a metallographic technique for monitoring residual damage as material is removed (14).

Detection of damage through an electrical property, e.g., the effective surface recombination velocity or minority carrier lifetime near the surface, as in the PME and photoconductivity decay methods (1) gives direct information about a property which is obviously important in a silicon device and also says something about the electrical nature of the damage. Detection of individual residual cracks and/or dislocations, as in the TEM, taper-sectioning and other metallographic techniques is also important since these defects may result in deeper damage during heating steps of subsequent processing. Lifetime techniques appear to be more sensitive than the metallographic methods in detecting damage due to fine polishes.

While many measurements have been made, and tabulations such as this one can give useful estimates for common abrasion treatments there will continue to be a need for measurements on particular processes of practical interest. Metallographic techniques are convenient and sensitive. Other electrical measurements based on lifetime can probably be used. For example, efficiency of a photodiode illuminated from the side opposite the junction is very sensitive to recombination velocity at the illuminated surface (25), and the wave-length dependence of efficiency gives information about diffusion damage and other "dead layer" effects at the illuminated surface (26).

X-ray diffraction topography is a sensitive technique (27,56). Optical reflectance spectra have been used to detect implantation damage in silicon (28) and GaAs (29) and also abrasion damage in GaAs (29). This measurement is rapid and does not require an optically smooth surface. The energy spectra of back-scattered ions, He^+ or H^+, give information about surface damage and also the impurities present in the surface layer. This technique will be discussed further in section 4.5.

3. Microscopic Nature of Abrasion Damage

The detailed microscopic nature of abrasion damage on silicon, and germanium, has been a matter of interest for some time, even though many of the practical considerations which arise in semiconductor processing can be dealt with successfully without a complete understanding of this matter. One question to be answered is: what are the defects, the recombination centers, which account for the very high recombination rates for minority carriers at a mechanically damaged semiconductor surface? Actually, in the case of silicon, the surface recombination velocity (S) is usually very high even on an etched undamaged surface, but various chemical treatments and atmospheres can reduce S to low values by shifting the surface potential (24,25), whereas this is not possible for damaged surfaces. A damaged surface has a considerably higher density of recombination states. In the late 1950's dislocations and also point defects such as vacancies and interstitials in abraded surfaces had been suggested by various workers. A difficulty in considering dislocations was pointed out by Faust (33), i.e., that they are ordinarily produced only by deformation at elevated temperature, 600°C for Si (34). Some of

the evidence and speculation regarding such defects was reviewed by Buck (1), but no very direct evidence for dislocations or any other defect in a mechanically damaged germanium or silicon surface had been reported at that time.

In the past decade considerable experimental work, particularly on silicon, has been brought to bear on the problem and some of this is summarized below.

Prussin (35) analyzed stresses in lapped silicon surfaces, measuring the radius of curvature of thin wafers, which were lapped on one side, by an interference fringe technique. He could find no evidence of dislocations by etching or X-ray techniques and presented a picture of a lapped surface consisting of microcracks filled with debris.

Pugh and Samuels (13) studied damaged layers on silicon by metallographic techniques. Using a chromic-hydrofluoric acid etch (36) they differentiated between cracks and dislocation arrays on the basis of the width of etch features, cracks producing wider fissures than did dislocations. The dislocation arrays were described as being similar to "dislocation cracks" which they had found in germanium (30). Dislocation cracks, first proposed by Allen (37), are supposed to result from cracks which open and reseal quickly, leaving a wall of dislocations.

Stickler and Booker (17,38) used transmission electron microscopy (TEM) to study the dislocation question in abraded surfaces of silicon, as illustrated in Figs. 3-5. The sample had been lapped unidirectionally with 400 mesh SiC paper and thinned to a few thousand A by etching from the undamaged side. Fig. 3 shows interference fringes which are attributed to slightly misaligned regions of material, one above the other, which result from cracking. Fig. 4 shows features which are attributed to a strained dislocation network running along the length of a groove produced by lapping. Fig. 5 shows the effect of annealing at 850°C. An approximately hexagonal network has resulted from rearrangement of a strained dislocation network (Fig. 4) while additional single dislocations have propagated deep into the surrounding material. Stickler and Booker proposed that these latter dislocations were of the type observed by others (31,32,35) after annealing abraded Ge samples. Stickler and Booker (17) emphasized the effect on the TEM image of tilting the specimen as a means of distinguishing between dislocations and cracks. The dark line representing a dislocation changes in contrast but not in position relative to the specimen surface, while the interference fringes due to cracked material move across the surface. Stickler and Booker also used a metallographic technique, progressive etching with CP4 and optical microscopic examination at each stage, to study the damage. This method did not distinguish between dislocation networks and cracks; however, it provided a convenient way of determining the depth of damage comprised of both cracks and dislocations.

In addition to dislocation studies there has been considerable effort on another type of defect, a paramagnetic center associated with mechanical damage on silicon surfaces. Fletcher, et al. (40), reported a spin resonance line with electronic g-value (gyromagnetic ratio) = 2.006 which was associated with the silicon surface. Feher (41) studied it on sandblasted silicon surfaces and found that removal of about 1μ of silicon by etching was required to eliminate the resonance line. Various chemical treatments which did not remove silicon (HCl, HNO_3, or HF alone) failed to make the resonance line disappear. (It might be noted here that the recombination centers associated with mechanical damage behave in a similar fashion.) Walters and Estle (42) studied the line, for which they determined a g-value of 2.0055, on lapped silicon surfaces and crushed silicon powders. Again they found the resonance behavior unaffected by atmospheres, even when crushing was done in vacuum and the powder subsequently exposed to atmospheres. From this and the etching results of Feher (41) they concluded that the paramagnetic centers were not localized at the silicon surface but rather were distributed in a depth of $\sim$ 1 micron.

Pursuing this matter further, Chung and Haneman (43) worked with silicon powders prepared by crushing in ultra-high vacuum, and found that the g = 2.0055 resonance did change when the powder was exposed to either hydrogen or oxygen at 10^{-2} Torr. The signal increased by 40-80% with hydrogen and 75-148% with oxygen, at room temperature. That is, the effective number of spins was increased by these amounts. Air and water vapor both reduced the signal. Silicon surfaces cleaved in high vacuum showed a similar resonance and sensitivity to atmospheres and Haneman (44) proposed a surface structure with alternate rows of raised and lowered atoms. The raised atoms with s dangling bonds overlapping 80%, contribute to the resonance. The depressed rows have p dangling bonds which overlap fully and do not contribute to the resonance. Mechanically polished surfaces and air-crushed powders were similar to each other but differed significantly from vacuum crushed powder (45).

The resonance behavior of vacuum-cleaved and vacuum-crushed silicon toward gas exposure was interpreted as showing that the paramagnetic centers were at the very surface, not in the interior, and were associated with dangling bonds formed in creating new surfaces. The increase in signal on adsorption of oxygen or hydrogen was attributed to a reduction in mutual overlap of the electron wave functions of the clean surface (45). The possibility was considered (44) that surface potential changes, i.e., band bending induced by adsorbed molecules, might influence the resonance of centers distributed in the bulk by changing their occupation probability, but this possibility was discounted because oxygen exposures sufficient (10^{-5} Torr min) to cause substantial band-bending did not affect the resonance, whereas heavy exposures (10^{-1} Torr min) which cause little additional change in barrier

height, did cause large changes in the resonance. Results of the earlier workers (42) in which the signal from crushed silicon was unaffected by atmosphere exposures, were explained by suggesting that the vacuum was not good enough to prevent rapid contamination by oxygen which desensitized the centers to further exposure. The EPR behavior of crushed powders and mechanically polished surfaces was thus related to creation of extra surface area with unpaired but partially overlapping electrons whose resonance behavior could be enhanced or reduced, but not destroyed, by interaction with atmospheres.

Taloni and Rogers (46) measured EPR signals from silicon surfaces lapped with diamond abrasives ranging in size from 0.5 to 60μ and found, rather unexpectedly (Fig. 7), that the signal height was not a monotonic function of abrasive particle size although the surface area of the substrate should be. The maximum in the curve at about 8μ was attributed to a competition between surface area of cracks and fissures (greater for large particles) and surface area contributed by silicon debris stuffed into the cracks. There would be more of this debris, with greater surface to volume ratio, for smaller abrasive particles.

Thus, there is evidence for dislocations beneath the cracked layer in abraded surfaces, and for a paramagnetic center with a very reproducible g-value, which seems to be at the outer surface. As yet there has been no direct correlation of either of these defects with electrical properties of a damaged surface.

4. Description of Processes (47,48)

4.1. Slicing

The several different methods which might be used for silicon wafering may be roughly classified as abrasive and nonabrasive. Nonabrasive slicing, for example chemical (49) or electrochemical (50,51) cutting or spark machining (52), is, in principle, preferred because of its potentiality for production of slightly damaged although perhaps very uneven, surfaces. Abrasive methods, on the other hand produce highly damaged, but quite flat, surfaces. Except for a few isolated experimental applications, all silicon wafers are currently produced by abrasive cutting methods. These methods may be further classified as to whether they involve loose or fixed abrasive; that is, whether 3- body or 2- body abrasion is involved (53). Wire sawing (54) is an example of the former, and sawing with bonded abrasive blades exemplifies the latter.

Loose abrasive wire and band sawing has been studied by Kobayashi (23), who has measured the cutting rate vs. wire speed and load and in addition has determined the resultant parallelism, flatness, kerf loss, and damage depth. Apparently, no analysis has been made of the damage nature but it is expected to be the same as on a lapped surface which was discussed in the previous section. Ikeda, et al. (55) have found, by use of electron microprobe X-ray analysis that abrasive grains are embedded in the surfaces of wire sawed silicon slices so that care must be taken to eliminate these in the next process step.

Wire saws are only now becoming commercially practical; consequently most silicon has been, and is at present, wafered by use of diamond cutoff wheels (47). The inside diameter (ID) saw has received widespread acceptance because of the material savings resulting from reduced kerf loss since thinner ID than OD blades may be used for given flatness and parallelism requirements.

A study was recently made, utilizing metallographic examination and mechanical characterization of the damaged surface of ID-diamond-sawed silicon (and germanium) (22). The damage, which was interpreted as being chiefly microcracks, was found to extend 50±10 microns below the as sawed surface. The important damage sources were identified as out-of-plane blade vibration and periodic abrasive contact. The damage depth has been confirmed by recent work of Blish (56) using Berg-Barret X-ray diffraction topography (27) and of Kobayashi (23) using the etch rate method. The depth of damage on both fixed abrasive (ID diamond wheel) and loose abrasive (bands and wires) are summarized in Table 1. For reciprocating wires, the damage depth is about equal to the abrasive size as expected, but for reciprocating bands it is considerably greater which probably eliminates it as a competitive method.

4.2. Removal of Abrasion Damage

Some sequence of lapping, etching, and polishing steps is used to remove saw damage and produce flat, nearly perfect, wafers for further processing. Lapping with progressively smaller particle sizes is sometimes used in an attempt to minimize the layer which must be polished off and optimize the time involved (57). Considerable care must be taken, however, to prevent the saw damage from being propagated deeper into the slice (68,59). For example the authors of the latter reference found that when an ID-sawed wafer was lapped with 14μ abrasive (75μ removed) then with 5μ abrasive (25μ removed) and finally polished (Lustrox, 25μ) that the dislocation density after annealing the wafer was $\sim 10^3$ whereas the background level was ~ 10. However, if the sawed wafer was chemically etched (HF, HNO_3; 60μ removed) and then polished as before the background density of dislocation could be reclaimed. Adjusting the lapping parameters, load, etc., does not materially affect the conclusion which is the necessity of relieving the saw or lap damage by chemical etching. Whitten, et al. (58) found that etching as sliced wafers prior to lapping with 15μ alumina reduced the damage depth, found by annealing the damage into

dislocations, from 60-75 to 15-20μ. Similarly if wafers were polished with .5μ diamond after slicing or lapping without benefit of an intermediate etch the damage depth was found to be 45-100μ whereas it was less than 5μ with etching.

Furthermore, it is known that abrasive particles are often embedded in the lapped surface (60) and it appears that wafers are contaminated with many foreign substances, especially heavy metals, during lapping (61,62). Nigh (63) found that a lapped back surface introduced a fast diffusing donor contaminant during oxidation at 1050°C in wet oxygen, unless care was taken to remove lapping residues with a detergent, before they dried. The levels introduced were studied by MOS measurements on the oxidized surface opposite the lapped surface, and were found to be photosensitive and to anneal out at relatively low temperature, 400°C. Based on the foregoing observations, it would appear the best procedure for removing saw damage is to chemically etch and then polish, eliminating lapping. This requires that sawed wafers be quite flat and parallel but this can be done.

Some care may be needed even in the choice and application of an etchant. Using Berg-Barret X-ray diffraction topography, Blish (56) examined ID-sawed wafers which had been etched with "purple" etch (1HF: 5 HNO_3: 6 HAc; saturated with iodine) and/or silica-sol (Syton) polished. Polishing to a depth of about 50 microns produced a damage free surface whereas etching a sawed surface to a depth of 200 micron left a damaged surface. Application of the etch to a damage free Syton polished surface did not produce damage and it was concluded that the etch caused the saw damage to propagate into the slice.

4.3. The Nature of Mechanical Polishing

There are still two viable theories on the nature of the mechanical polishing process. The first, and older, of the two views holds that polishing is just very fine scale abrasion differing only in degree from lapping. The other holds that polishing is quite different from abrasion and involves removal of layers no more than a few monolayers thick.

Polishing of both metals and glasses has been widely studied. In the case of metals the principal objective is usually to attain a polished surface in the least possible time without much regard to resultant crystallographic perfection (64). When the polished surface is characterized attention is paid mostly to the areas under surface scratches (65,66). It seems however, that an ideal polishing process should eliminate scratches; that is, that they are a defect in the polishing to be avoided. While Samuels (64,66) concludes that metal polishing differs only in degree from abrasion, Rabinowicz (67) proposes that the polishing particles adhere to the surface and that removal is on a molecular scale if a critical load is not exceeded.

Of more direct relevance to the problem of silicon polishing is the extensive work that has been done on polishing of glass. Of special interest is the theory proposed by Kaller (68) which is that by complex chemical processes involving the chemical nature of the surface, the polishing agent, the polishing media, and the polishing pad, a silica gel layer is formed on the surface. This is molecularly removed by bond formation with the polishing agent which as it is comminuted makes available reactive sites. This is of course quite similar to the idea of Rabinowicz (67) except in his model the particle adheres to the original material. The experimental evidence, which is too extensive to discuss here but which has been reviewed adequately elsewhere (69,70) indicates that these mechanisms are important, but that other processes take place to some extent depending on the conditions. In particular, abrasion is more or less important depending on the hardness and shape of the particles involved.

These considerations suggest that an attractive process for polishing a silicon (or any other) surface to near perfection might involve production of a thin reaction layer which would be removed by particles or pads not capable of abrading or deforming the underlying crystal - chemical/mechanical polishing. That is, the polishing process should be designed to minimize undesirable effects, especially abrasion.

4.4. Polishing Silicon

A review of methods and materials for polishing silicon has been given by Mendel (15). Comparison of results led to the conclusion that the most satisfactory surfaces for device fabrication, that is those most nearly perfect, were obtained with the silica-sol (syton) polishing method (71). Electrochemical (72,73) and copper displacement (74-76,83) polishing also produce quality surfaces but they fall outside the scope of the present discussion.

4.4.1. Diamond and Alumina Polishing

In the past diamond and alumina have been used extensively for polishing silicon; diamond produces the more severly damaged surface for the same particle size (15,77). Although these surfaces may appear free of defects when examined by ordinary illumination, Nomarski optics or electron microscopy reveal the presence of many fine scratches and cracks. Adams (78) has observed, by use of ellipsometry, that diamond polished surfaces have a large refractive index and extinction coefficient due to a thin layer of damaged silicon. Dislocations have been observed when as polished specimens were examined by electron

microscopy (4) after annealing, and oxidation or epitaxial growth on such polished surfaces results in stacking faults (3,4,77)(Fig. 6).

4.4.2. Silica-Sol Polishing

The silica-sol polishing medium is a colloidal suspension of silica gel in sodium hydroxide solution having a pH of 9.9. The suspension is about 30 percent silica and the particle size range is 100-400 A. With a 12 inch wheel speed of 60 rpm and a 5 inch block speed of 45 rpm, the n-type silicon polishing rate is about 10 μ/hr. for a load of 200 gm/cm^2 of polished area when a napless polyurethane pad is used. Loads of 50 gm/cm^2 and 1 kg/cm^2 give polishing rates of 2 μ/hr. and 30 μ/hr. respectively. Doubling the block rotational speed has no measurable effect.

Heavily doped p-type (Boron) silicon, however, is removed about four times slower, and although the polishing rate is always less than for n-silicon, it does increase with increasing resistivity. It is interesting to note that it has been reported (4) that p-type silicon is diamond polished more rapidly than n-type and that the higher the boron concentration the greater the polishing rate. This is in both respects opposite to the silica-sol behavior, but the explanation in either case is unknown. In order to further illustrate the chemical nature of the silica-sol process we note that when it is used to polish other semiconductors, additions must be made to the solution (H_2O_2 for Ge and Br_2 for GaAs) to prevent formation of an orange peel surface.

When silica-sol polished surfaces are examined by phase contrast, dark field or electron microscopy or by X-ray diffraction topography no evidence of surface damage is found. When the slices are etched either before or after annealing no dislocations produced by the polishing are observed and it appears that stacking fault formation during oxidation or epitaxial growth is reduced to levels comparable to etched surfaces.

Adams (78) has studied silica-sol polished surfaces ellipsometrically. He finds the refractive index to be a function of the amount removed but that it stabilizes after approximately 2 mils have been removed from a sawed/etched wafer. The final value is, however, slightly higher than for an etched surface. Also, the refractive index is found to be a function of the removal rate, being smaller for the lower rate. The polished wafer has a surface film about 20 Å thick. Methanol, but not trichloroethylene or acetone, or strong oxidizers reduces the film thickness 6 Å. An HF rinse then reduces the film 10 Å. The conclusion is that the film is an oxide layer contaminated by an organic which is presumably a wax residue since the wax used is soluble in methanol but not acetone or trichloroethylene.

4.5. Ion-Scattering from Polished Surfaces

Back-scattering of energetic light ions, He^+ and H^+, is an interesting technique for studying damaged surfaces which has been developed recently. The method involves ion-channeling and Rutherford scattering. It has been used to study ion implantation damage in solids (79) and has also been employed specifically on surfaces to detect surface impurities and disordered substrate atoms in oxide layers (80,81). A system (82) utilizing 100-200 keV He^+ ions has been used in the present work to compare silica-sol (Syton) polishing and Linde B polishing with heavily etched surfaces.

In this method the silicon crystal is oriented so that a major crystal axis is parallel to the collimated ion beam. This causes channeling of ions through the spaces between rows of atoms and reduces back-scattering from atoms deep in the crystal, i.e., atoms farther in along the rows are shadowed by surface atoms. The reduction in back-scattering yield is quite marked for a well-aligned crystal with a well-ordered surface, as compared with the non-aligned or random case (see Fig. 8). Disorder at the surface increases the scattering yield in the aligned case since more atoms are exposed to the ion beam. Impurity atoms at the surface produce scattering peaks at different energies depending on the atomic masses, peaks for heavy elements occurring at higher energy.

Fig. 8 shows energy spectra, measured with an electrostatic analyzer, of 100 keV He^+ ions back-scattered from silicon surfaces through an angle of 120°. The Si peaks at 62-63 keV in the aligned spectra are due to Rutherford scattering from surface atoms of the lattice and from disordered atoms in the usual 20-50 Å oxide films on the surface. These ions bounce back with 62-63% of the incident energy. The scattering yield at lower energy is due to ions which have penetrated beyond the surface and have lost energy to electrons going in and back out, in addition to the loss in a back-scattering collision with a silicon atom inside. This yield is low for the aligned cases, only a few percent of the random yield. In Fig. 9 a Syton polished region is compared with a heavily etched region on the same (111) silicon surface. In each case 75μ of silicon had been removed from a saw-cut surface. There is no significant difference in the silicon spectra either in the surface peak or at lower energy, which says that the Syton polished surface has no more damage detectable by this technique as presently applied than does the etched surface. There is, however, a higher yield beyond 65 keV, from the polished surface; this indicates impurities heavier than silicon at densities of $< 10^{14}$/cm^2, including iron at 77 keV. Both surfaces had been carefully cleaned in organic solvents, acids (HCl, HF, HNO_3 separately), and deionized water before mounting in the scattering chamber.

Fig. 8 shows the spectrum from a surface which had been polished (after heavy etching) with Linde B slurry, on the same type of wheel and pad employed for Syton polishing. This is compared with the spectrum for a heavily etched region on the same (111) surface. In this case the silicon spectrum from the polished surface is definitely higher than that from the etched surface, indicating more damage. A much greater change would have been produced by a coarser abrasive; 5μ alumina produces a spectrum similar to the random spectrum (82). The polished surface also shows a substantial impurity peak at 77 keV; it is in the right position for iron and is traceable to an iron ring used to hold the polishing block on the pad. This iron contamination ($\sim 5 \times 10^{14}/cm^2$) was retained on the polished surface in spite of subsequent cleaning with acetone, d.i. water, boiling nitric, d.i. water, HF, d.2. water. Etching, on the other region of the sample, did remove it. Both regions show a small peak at 91-92 keV believed to be gold ($\lesssim 10^{12}$ atoms/cm^2). (Details of the impurity detection sensitivity will be published elsewhere (82)).

5. Acknowledgements

The authors wish to thank G. H. Wheatley for his collaboration in the ion scattering experiments, R. O. Miller and H. J. Robinson for discussions and data on silica-sol polishing, W. F. Flood, Jr. for preparation of samples and discussion on cutting techniques, and several other colleagues for permission to mention unpublished results.

6. References

(1) Buck, T.M., The Surface Chemistry of Metals and Semiconductors, H.C. Gatos, Editor, p. 107, John Wiley and Sons, New York (1960).

(2) Koehler, J.S., Phys. Rev. 181, 1015 (1969).

(3) Fisher, A.W. and Amick, J.A., J. Electrochem. Soc., 113, 1054 (1966).

(4) Lawrence, J.E., JAP 40, 360 (1969).

(5) John, H.F., Proc. IEEE 55, 1249 (1967).

(6) Stickler, R. and Booker, G.R., J. Electrochem. Soc., 111, 485 (1964).

(7) Goetzberger, A. and Finch, R.H., S.S. Electron., 7, 543 (1964).

(8) Queisser, H.J., Semiconductor Silicon, P.R. Haberecht and E. L. Kern, Editors, p. 585, Electrochemical Society, New York (1969).

(9) Jungbluth, E.D. and Wang, P., JAP 36, 1967 (1965).

(10) Queisser, H.J. and Shockley, W., Phys. Rev., 123, 1245 (1961).

(11) Lagowski, J. et al., Surf. Sci., 19, 205 (1970).

(12) Faust, J.W., Electrochem. Tech.,2, 339 (1964).

(13) Pugh, E.N. and Samuels, L.E., J. Electrochem. Soc.,111, 1429 (1964).

(14) Stickler, R. and Faust, J.W., Electrochem. Tech., 4, 399 (1966).

(15) Mendel, E., S.S. Tech.,10, 27 (1967).

(16) Faust, J.W., ASTM Special Tech. Pub.,246, 66 (1959).

(17) Stickler, R. and Booker, G.R., Phil. Mag., 8, 859 (1963).

(18) Buck, T.M. and McKim, F.S., data tabulated in reference 1.

(19) Buck, T.M. and McKim, F.S., unpublished data.

(20) Andrus, J. and Bond, W.L., data reported in reference 1.

(21) Andrus, J. and Bond, W.L., Am. Mineralogist, 37, 622 (1952).

(22) Meek, R.L. and Huffstutler, M.C., J. Electrochem. Soc.,116, 893 (1969).

(23) Kobayashi, A., paper presented at the 8th National Conference of the American Society for Abrasive Methods, Hartford, Conn., (March 2, 1969).

(24) Buck, T.M. and McKim, F.S., J. Electrochem. Soc.,105, 709 (1958).

(25) Harten, H.U., Phillips Res. Rep.,14, 346 (1959).

(26) Buck, T. M., Casey, H. C. Jr., Dalton, J.V. and Yamin, M., BSTJ, 47, 1827 (1968).

(27) Turner, A.P.L., Vreeland, T. Jr., and Pope, D.P., Acta Cryst., A24, 452 (1968).

(28) Kurtin, S., Shifrin, G.A., and McGill, T.C., App. Phys. Lett., 14, 223 (1969).

(29) Sell, D.V. and MacRae, A.U., to be published.

(30) Pugh, E.N. and Samuels, L.E., J. Electrochem. Soc., 109, 1197 (1962).

(31) Pugh, E.N. and Samuels, L.E., J. Electrochem. Soc., 108, 1043 (1961).

(32) Pugh, E.N. and Samuels, L.E., J. Electrochem. Soc., 109, 409 (1962).

(33) Faust, J.W., paper presented at the Electrochemical Soc. Mtg., Buffalo (1957).

(34) Pearson, G.L., Read, W.T. Jr., and Feldman, W.L., Acta Met., 5, 181 (1957).

(35) Prussin, S.A., ASTM Symp. on Met. of Semicond. Mat., Los Angeles, 219 (1961).

(36) Sirtl, E. and Adler, A., Z. Metall., 52, 529 (1961).

(37) Allen, J.W., Phil. Mag. 4, 1046 (1959).

(38) Stickler, R. and Booker, G.R., J. Electrochem. Soc., 109, 743 (1962).

(39) Faust, J.W., ASTM Symp. on Cleaning of Elect. Device Comp. and Mat., Special Tech. Publ. No. 246.

(40) Fletcher, R.C. et al., Phys. Rev., 94, 1392 (1954).

(41) Feher, G., Phys. Rev., 114, 1219 (1959).

(42) Walters, G.K. and Estle, T.L., JAP, 32, 1854 (1961).

(43) Chung, M.F. and Haneman, D., JAP, 37, 1879 (1966).

(44) Haneman, D., Phys. Rev., 170, 705 (1968).

(45) Haneman, D., Chung, M.F. and Taloni, A., Phys. Rev., 170, 719 (1968).

(46) Taloni, A. and Rogers, W.J., Surf. Sci., 19, 371 (1970).

(47) Runyan, W.R., Silicon Semiconductor Tech., McGraw-Hill Book Co., New York (1965), p.231.

(48) Frankl, D.R., Electrical Properties of Semiconductor Surfaces, Pergamon Press, New York (1967).

(49) Maddin, R. and Asher, W., Rev. Sci. Instr., 21, 881 (1950).

(50) Uhlir, A., BSTJ, 35, 333 (1956).

(51) Sheff, S., Electrochem. Tech., 5, 47 (1967).

(52) Livshits, A.L., Electro-Erosion Machining of Metals, Butterworths, London (1960).

(53) Rabinowicz, E., Friction and Wear of Materials, John Wiley and Sons, New York (1965).

(54) Jensen, E.W. and Jacobsen, H.R., S.S. Tech., 11, 53 (1968).

(55) Ikeda, M., Inomata, K., and Kawada, M., Bull. Electrotech. Lab. (Japan), 31, 72 (1967).

(56) Blish, R.C., private communication.

(57) Mayberg, S., J. Electrochem. Soc., 116, 509 (1969).

(58) Whitten, W.M., Heitz, A.J. and McNamara, J.R., paper presented at the Electrochem. Soc. Fall Mtg., Washington, D.C. (1964).

(59) Benson, K.E. and Stewart, W.W., Chem. Engr. Prog., 64, 93 (1968).

(60) Faust, J.W., Surf. Sci., 13, 60 (1969).

(61) Martin, J.A., Semiconductor Silicon, R.R. Haberecht and E.L. Kern, Editors, p. 547, Electrochemical Society, New York (1969).

(62) Lesher, D.F. and Doll, R.F., private communication.

(63) Nigh, H.E., private communication.

(64) Samuels, L.E., Metallographic Polishing by Mechanical Methods, Sir Isaac Pitman and Sons Ltd., Melbourne (1967).

(65) Samuels, L.E., Res. Appl. Ind., 13, 344 (1960).

(66) Samuels, L.E., The Surface Chemistry of Metals and Semiconductors, H.C. Gatos, Editor, p.82, John Wiley and Sons, New York (1960).

(67) Rabinowicz, E., Scientific American, 218, 91 (1968).

(68) Kaller, A., Naturwiss., 43, 156 (1956). Silikattech., 1, 38 (1956). Glastech. Ber., 30, 35 (1957). Jenaer Jahrbuch, 145 (1957). Jenaer Jahrbuch, 181 (1959).

(69) Holland, L., The Properties of Glass Surfaces, Chapman and Hall, London (1964).

(70) Cornish, D.C., The Mechanism of Glass Polishing, B.S. I.R.A. Research Report R267 (1961).

(71) Walsh, R.J. and Herzog, A.H., U.S. Patent No. 3,170,273 (Jan. 10, 1963).

(72) Sullivan, M.V. et al., J. Electrochem. Soc., 110, 412 (1963).

(73) Baker, D. and Tillman, J.R., S.S. Electron., 6, 589 (1963).

(74) Regh, J. and Silvey, G.A., Electrochem. Tech., 6, 155 (1968).

(75) Mendel, E. and Yang, K., Proc. IEEE, 57, 1476 (1968).

(76) Blake, L.H. and Mendel, E., S.S. Tech., 13, 42 (1970).

(77) Thomas, D.J.D., Phys. Stat. Sol., 3, 2261 (1963).

(78) Adams, A.C., paper presented at the Los Angeles Mtg. of the Electrochem. Soc. (May 13, 1970).

(79) "Atomic Collision and Penetration Studies," Proc. Int. Conf. on Atomic Coll. and Penetration Studies, Chalk River, Canada, Can. J. Phys., 46, 449-782 (1968).

(80) Davies, J.A., Denhartog, J., Eriksson, L. and Mayer, J.W., Can. J. Phys., 45, 4053 (1967).

(81) Bøgh, E., Can. J. Phys., 46, 653 (1968).

(82) Buck, T.M. and Wheatley, G. H., Am. Vac. Soc. Mtg., 10/27/69, Seattle, Washington, to be published.

(83) Soper, R.B., p. This volume.

TABLE 1

Summary of Depth of Damage Results

Abrasive Treatment		Details	Measurement of Damage	Depth of Damage	Reference
Fine Polish	(1)	0.25μ diamond slurry on etched surface	TEM (Transmission Electron Microscope)	0.2 Micron	17
	(2)	Linde A Al_2O_3 (1.0μ)	TEM	< 0.2	17
	(3)	Linde B Al_2O_3 (0.1μ)	TEM	< 0.2	17
	(4)	0.25μ diamond, dry, cloth lap on wheel	PME (Photo Magneto Electric measurement; progressive etching)	0.8	19
5 Micron Lap	(5)	No. 305 Al_2O_3; water slurry on glass plate	PME	3-4 Microns	18
	(6)	Ditto	X-ray Rocking Curve	3	20, 21
25μ Lap	(7)	Al_2O_3; (100) surface, water slurry	Etch Rate	3-4 Microns	12
	(8)	SiC; (100) surface; water slurry	Etch Rate	7-8	12
	(9)	Diamond; (100) surface; oil slurry	Etch Rate	3-4	12
	(10)	SiC; water slurry on glass plate	PME	8-10	18
	(11)	SiC; water slurry on cloth	PS (Progressive Section method. Microscopic examination.)	3	17
	(12)	SiC; abrasive paper	PS	9-12	17
	(13)	SiC; water slurry on glass plate; unidirectional	TS (Taper-Sectioning)	15	13
	(14)	SiC; abrasive paper; unidirectional	TS	10	13
	(15)	Al_2O_3; (111) surface; water slurry on glass plate	PP (Planar Polishing - successive fine polishes and light etches on lapped surface, microscopic examination)	14-18	14
	(16)	Al_2O_3; (110) surface; water slurry on glass plate	PP	18-23	14
	(17)	Al_2O_3; (100) surface	PP	17-22	14
Saw Damage	(18)	220 mesh diamond, O.D. saw, 1 cm/min.	Etch Rate	10-15 Microns	12
	(19)	I.D. saw	Etch Rate	7-11	12
	(20)	50μ diamond; I.D. blade; 1-4 cm/min.	Preferential etch - microscopic examination	50	22
	(21)	40μ diamond; I.D. blade	Etch Rate	40-45	23
	(22)	15μ diamond; recip. bands	Etch Rate	65-70	23
	(23)	37μ diamond; recip. wires	Etch Rate	50-55	23
	(24)	20μ diamond; recip. wires	Etch Rate	20-30	23
	(25)	15μ diamond; recip. wires	Etch Rate	24-26	23
	(26)	10μ diamond; recip. wires	Etch Rate	12-15	23

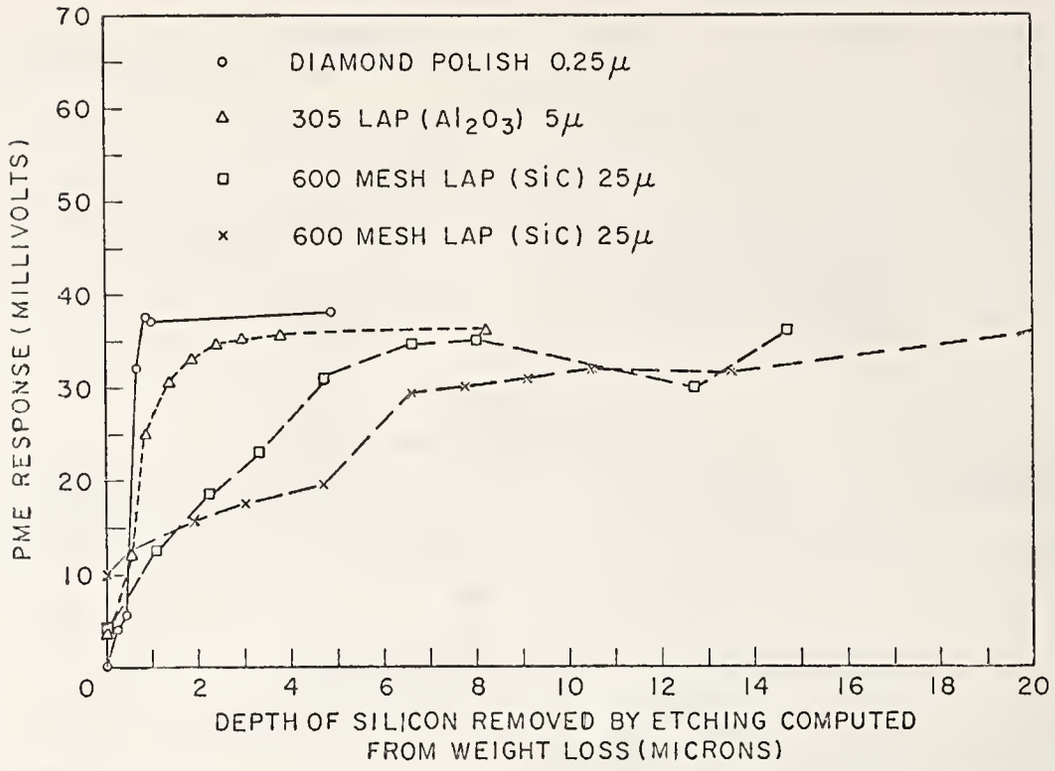
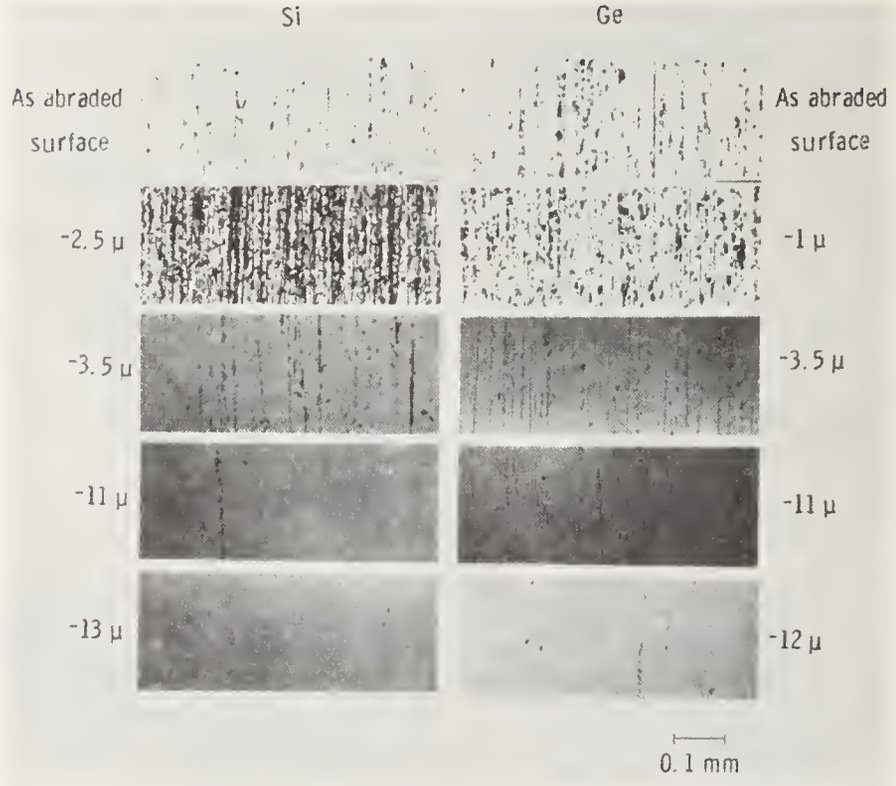


Fig. 1. Open Circuit PME Voltage vs. Depth of Silicon Removed from Damaged Surface for Polished and Lapped Silicon Surfaces (Ref. 19). Effective Surface Recombination Velocity is $\gtrsim 10^4$ cm/sec at Low End of Curves and $\lesssim 10^2$ cm/sec at High End.



Si and Ge wafer, abraded on 600 Al_2O_3, water as lubricant; areas of surfaces after "planar polishing" with ¼μ diamond, 1 sec etched in CP4; depths below max elevation on as abraded surface as indicated.

Fig. 2. Illustration of Use of Metallographic Examination for Determination of Damage Depth (Ref. 14).

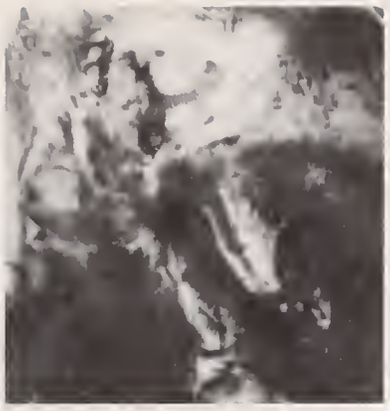


Fig. 3. Transmission Electron Micrograph of Silicon Abraded with No. 400 SiC Paper Showing Interference Fringes Arising from Cracked Material (Ref. 38).

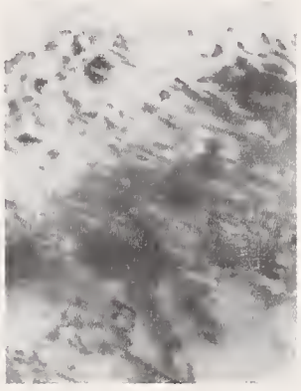


Fig. 4. Transmission Electron Micrograph of Silicon Abraded with No. 400 SiC Paper Showing Severely Strained Dislocation Network (Ref. 38).

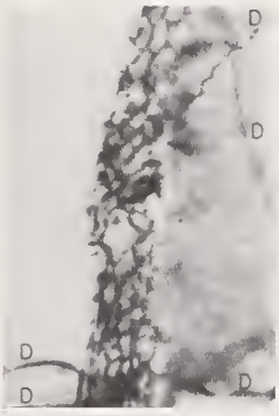


Fig. 5. Transmission Electron Micrograph of Silicon Abraded with No. 400 SiC Paper after Annealing 1 hr. at 850 °C Showing that Dislocations have Rearranged Themselves into Approximately Hexagonal Form and that Single Dislocations (D) have Propagated into the Surrounding Regions (Ref. 38).

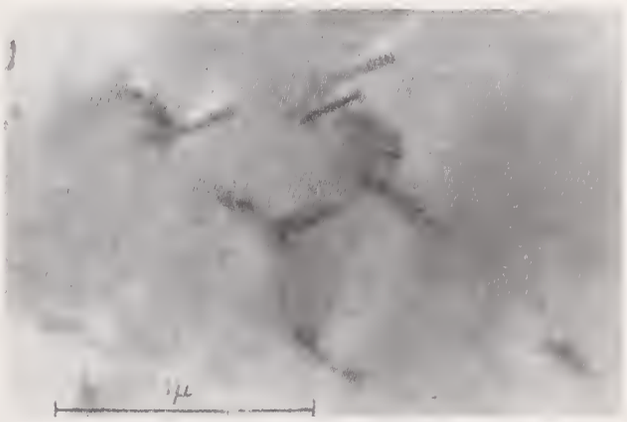


Fig. 6. Transmission Electron Micrograph of Diamond-Polished Silicon Surface after Oxidation at 1100 °C Showing Stacking Fault Formation (Ref. 4).

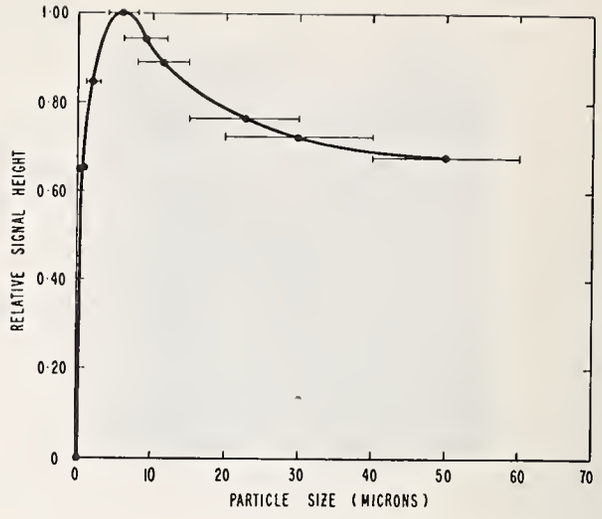


Fig. 7. The Normalized Relative E.S.R. Signal Height from Abraded Silicon Surfaces as a Function of Abrasive Particle Size (Ref. 46).

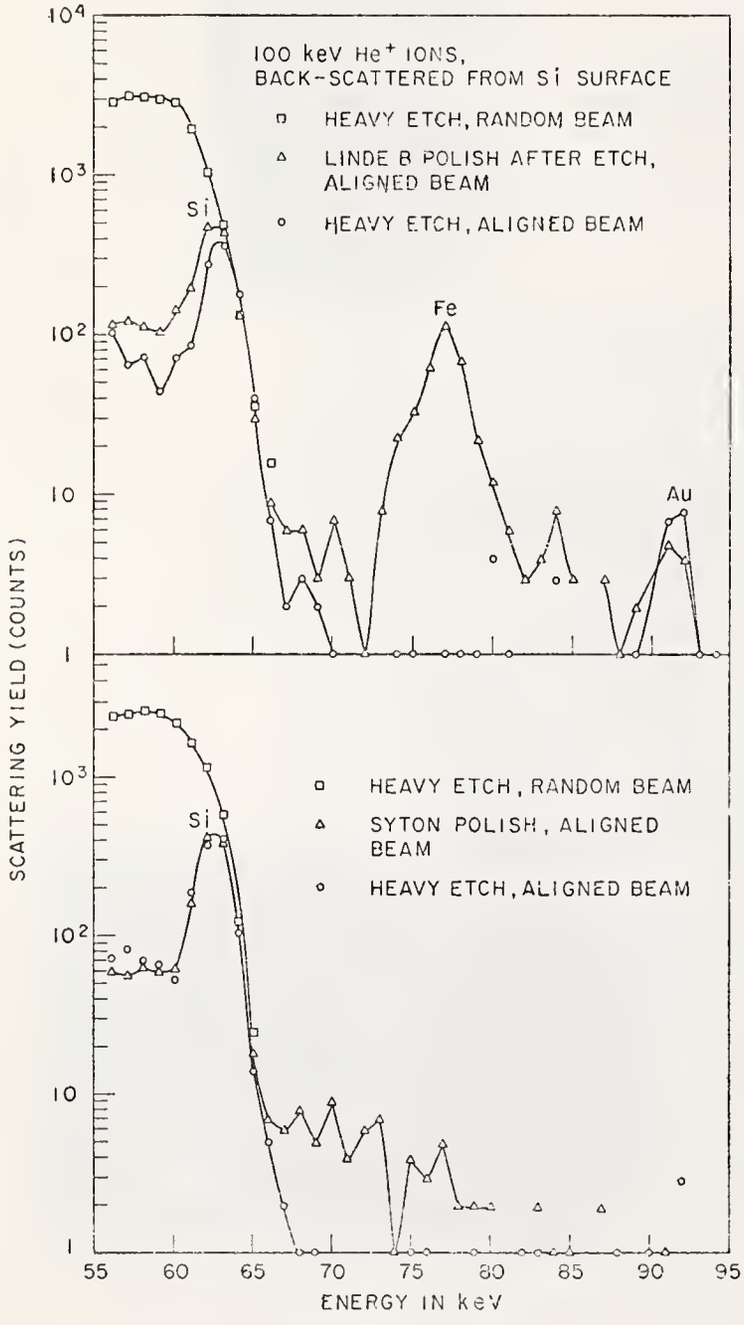


Fig. 8. Scattering Yield vs. Energy of He^+ Ions Back Scattered from Etched and Linde B Polished Silicon Surfaces.

Fig. 9. Scattering Yield vs. Energy of He^+ Ions Back Scattered from Etched and Syton Polished Silicon Surfaces.

THE PREPARATION OF PRACTICAL, STABILIZED SURFACES FOR SILICON DEVICE FABRICATION

by

A. Mayer and D. A. Puotinen
RCA Solid State Division
Somerville, N.J.

Polished silicon surfaces always contain a thin oxide layer. Methods for chemically treating real surfaces to contain a minimum of oxide impurities are described, together with suitable use tests. The influence of subsurface damage is also examined and methods of detection are described. The effect of crystal defects is briefly considered.

Advances in the art of preparing silicon wafer surfaces are coupled closely to the increased use of various x-ray, electron-beam scanning, and diffraction techniques, as well as visible light and infrared microscopy. The recognition of the origin and propagation of crystal defects during cutting, lapping, and polishing led to the development of chemical polishing methods that allow the silicon manufacturer to offer essentially blemish-free surfaces for regular sale. The purchaser, however, has to assure himself that the polished wafers not only conform to the basic materials specifications, but also can be adequately cleaned by relatively simple methods just prior to use in order to leave no adsorbed ionic impurities on the surface, and as few total dust particles as possible. Low-energy electron diffraction, electron-beam probing, and Auger spectroscopy are some of the excellent diagnostic tools which are sufficiently simple to be used regularly by skilled personnel for investigational purposes.

This paper suggests a reasonable method for providing a reproducible, stable, thin oxide-covered surface, and describes the procedures employed in characterizing and monitoring the process routinely.

Although largely focused on the preparation of silicon wafers destined for oxidation, the arguments presented are just as valid in consideration of other process steps. Also, because it is a well known fact that scratches, microcracks, and other crystal defects and impurities can rapidly propagate through the single crystal slice during heating, the remarks presented are confined to material in which the reverse side has been chemically etched to remove such blemishes. Lastly, it is assumed that incoming wafers have been cleaned by the vendor to remove gross contamination such as mounting wax or stain films, and were screened to exclude wafers with gross physical defects.

Preparation and Stabilization of Desirable Surface Films

Surface Oxide. When silicon is cleaned or etched and then exposed to air it is instantly covered with an oxide film, often in the hydrated form. This oxide film is usually thin enough for electrons to tunnel through, but for the purpose of making good ohmic contact with metal, for certain diffusions, for making Schottky diodes, or other uses, it is important to either remove or penetrate the film. The rates of formation of this film at room temperature under various conditions[1, 2] are compared in Fig. 1.

The oxide inevitably contains elements left from the preparative steps such as fluoride, copper, sodium, and iron. In addition to this oxide film, the surface of any wafer will carry dust, fiber particles, and often an organic film derived from degreasing, packaging, or handling.

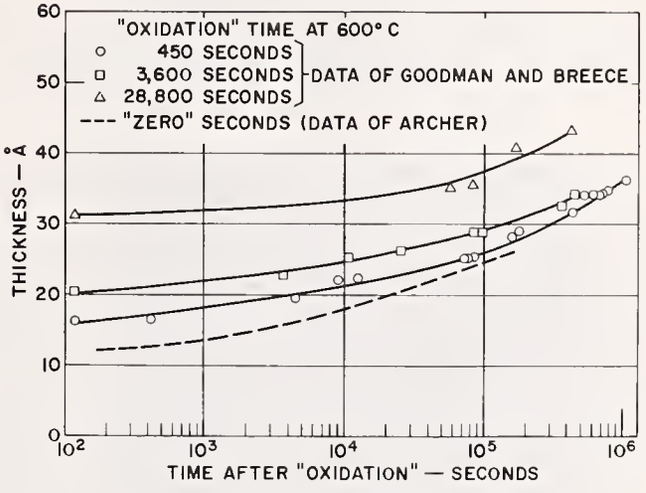


Fig. 1 - Oxide growth on bare silicon at room temperature with and without dry pre-oxidation at 600° C.

Most wafers are, or become, hydrophobic within a short time after etching. Because it is desirable to exercise control over the thin oxide film, it is best to remove the original film, and then to reform a new film under controlled conditions. Because aqueous reagents are to be used, the logical sequence of events is first to render the oxide face hydrophilic, then to etch off the oxide, and finally to desorb or dissolve impurities such as fluoride and metallic contaminants. Cleaning reagents are chosen to be completely volatile at a temperature low enough that they do not react with silicon.

Cleaning Steps. The most commonly used cleaning reagents are hot nitric acid, and sulfuric or hydrochloric acid-hydrogen peroxide whose function is mainly to destroy organic films and to remove metal ions by dissolution and dilution. The cleaning process consists of a primary treatment with a dilute aqueous solution of ammonia-hydrogen peroxide designed to take advantage of the solvent action of ammonia while organics are being oxidized, followed by a treatment with hydrochloric acid-hydrogen peroxide solution to dissolve and retain ions in solution by the chelating action of the chloride. Fig. 2 compares the redox potential of various wet oxidizing agents[3] and shows that the cleaning reagents are quite closely matched in redox properties. The dilute NH_4OH–H_2O_2 and HCl–H_2O_2 offer several theoretical as well as practical advantages. Some advantages[3, 4, 5] are that the dilute solutions are strong chelating agents, and are safer to handle than either the hot, concentrated nitric acid or the hot sulfuric acid-hydrogen peroxide mixture. The dilute cleaning solutions can also be used safely in fluorocarbon or polypropylene ware at temperatures up to

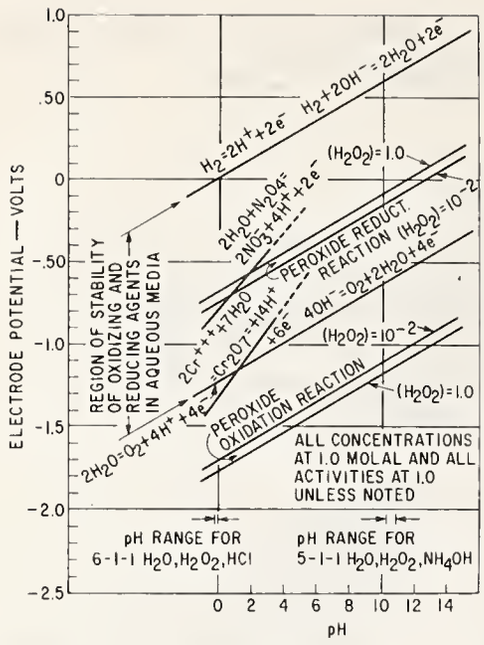


Fig. 2 - Potential vs pH diagram for hydrogen peroxide solution.

90°C, by contrast the strong nitric or sulphuric acid must be handled in the more fragile glass or quartz apparatus. Another advantage of the HCl–H_2O_2 solution is that any metal chloride residues formed are more easily volatilized than corresponding nitrates or sulfates formed from the action of the hot, concentrated acids. In all cases, it is good practice to add silicon chips at the acid storage vessels[6] to remove trace impurity atoms by displacement plating. Fluorine, equivalent to about a monolayer[7], is inevitably present in the thin oxide films after etching in hydrofluoric acid as well as after polishing by virtually all useful chemical methods. It may well be that it is the adsorbed fluoride ion which renders these "bare" silicon wafers hydrophobic. NH_4OH–H_2O_2 solutions invariably render the silicon or oxide surface hydrophilic, which is a strong indication that fluorine is reduced to a very low level in this treatment. Fluorine normally cannot be removed in a reasonable time even by volatilization at 1200°C from an SiO_2 surface[8].

A method[9] has been designed to stabilize the silicon surface for up to 60 minutes[10] by treating the freshly etched wafer with iodine dissolved in methanol. This method appears to link iodine to the surface of the wafer in such a way that the remaining oxide, which is thin at this point, rapidly volatilizes in hydrogen when heated to over 600°C, leaving a silicon surface ready for epitaxial deposition.

Contomination After Desorptive Cleoning. Triple-distilled or highest-quality deionized rinse water must be used to assure that no metallic ions or bacteria are present. The major decontamination problem, however, is the removal of dust and the prevention of particles settling on the clean surface during handling from the cleaning station to the next processing station. Four varieties of dust particles are commonly present:

a. Simple organic dust, such as lint, paper fibers, and oil mist. These dust particles are taken care of by the wet oxidation step.

b. Inorganic dust such as small particles of common metals and their compounds. These particles are generally dissolved and complexed by the acid treatment.

c. High molecular-weight polymer particles such as silicones, fluorocarbons, polyimides, and human skin flakes. These particles are difficult to remove and often disrupt the growing silicon oxide film when the particles do not burn off completely during high-temperature oxidation.

d. Refractory materials such as clay, sand, quartz, and silicon scribing dust. These particles are too slowly soluble for chemical removal. It is common practice to use nylon brushes wetted with detergent to remove them. Other implements have been used, but there is evidence that they can introduce crystal damage and leave residues.

Many of these contamination problems can be minimized by clean handling methods, use of white rooms, laminar flow boxes, filtered gas lines, and other methods, but no acceptable or convenient methods have been found for removing dust particles reliably and cleanly from silicon wafers.

Evoluation of Effectiveness of Cleoning Methods. The "use test" is still the best routine procedure for the evaluation of cleaning methods. Silicon, epitaxially deposited on a dirty surface, will show bumps and polycrystalline areas, visible in dark-field or with phase-contrast microscopy with or without differential etching. Also, resistivity variation can often be detected by four-point probe measurements or by capacitance-voltage plots on Schottky diodes. Deposition of a thin film of silicon dioxide by the oxidation of silane at 300 to 400°C is a sensitive, qualitative way of showing up poorly cleaned surface, and is useful as a spot check for surface films and particles, as shown in Fig. 3.

Fig. 3 - SiO_2 Deposit on inadequately cleaned silicon. Oblique illumination.

The most useful check[11, 12] that provides semiquantitative results is based on the growth of a thin SiO_2 layer in a clean furnace, formation of metal-oxide-semiconductor (MOS) capacitor by metallization, and measurement of the capacitance-voltage (C-V) relationship at room temperature before and after bias heat treatment (BT) at 300°C for one minute with a field of $\pm 10^6$ volt/cm. Normally, when 1000Å-thick SiO_2 is grown, a 0.1-volt change is detectable, which is equivalent to 2×10^{10} ions/cm^2. This method is sensitive to less than 10^{-3} monolayer equivalents of uniformly adsorbed monovalent contaminant. This SiO_2 procedure also samples the original surface to a depth of about 500Å and interrogates the unoxidized silicon to a depth of a few Debye lengths. Subsurface damage will distort the C-V plot but, with few exceptions, will not affect the ionic drift under BT treatment. The validity of the test was determined by the deliberate use of contaminated reagents, rinses, and appropriate countermeasures.

The minimum-to-maximum capacitance ratio of the C-V plot can be used to check the uniformity of carrier density of the Si-SiO_2 interface. Also, the dielectric breakdown voltage of

the capacitor can be used for the detection of defects. However, this dielectric method cannot distinguish between defects caused by poor surface preparation and foreign particles introduced later. Defect detection will be discussed in detail in a later section.

Results. On the basis of two years of study of oxide properties, the following ovservations can be made:

1. The procedures are capable of routinely producing oxidized wafers with uniform and reproducible properties.
2. The impurity level is sufficiently low that further refinement of the total procedure is unwarranted at this time.
3. The evaluation procedures are sufficiently reliable that deviations can be traced to operator error, equipment failure or contamination of reagents, water or gases. Also, the cleaning procedure using NH_4OH–H_2O_2 and HCl–H_2O_2 is effective in producing silicon surfaces with very low ionic contamination levels, as shown in Fig. 4.

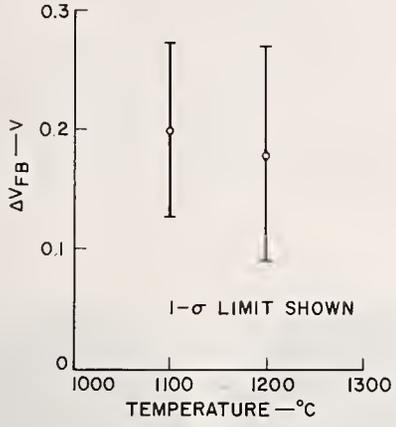


Fig. 4 - Stotisticol summory of CVBT test doto for o nine month period. Test conditions: 300°C ot 10^6 V/cm bios stress.

Subsurface Defects

Some crystal damage inevitably occurs in surface preparation, but good chemical polishing technique is capable of removing essentially all the mechanical damage. If insufficient stock is removed, however, an apparently well-polished surface may contain microcracks, strain fields, dislocation bundles, and embedded particles of polishing agents or debris. Besides the relatively sophisticated x-ray and electron-beam techniques, defects may be revealed by destructive tests such as differential etching before and after oxidation. One method of inspection that has been used is to sample the silicon at the rate of 3 to 5 wafers per hundred of incoming units. In general it is found that very few wafers are entirely free from defects. Typical distributions of residual damage revealed by differential etching for an insufficiently polished wafer is compared with a well-polished wafer in Fig. 5.

The combination of oxidation at temperatures greater than 1100°C and Sirtl[13] etching is a particularly sensitive method[14, 15] for revealing subsurface defects. Oxidation followed by removal of the oxide with hydrofluoric acid and the taking of a surface profile reveals a generally clean surface with some projecting spikes. After three minutes Sirtl* etching, the vast majority of microscopically significant figures project up to 2 μm out of the surface, as shown in Fig. 6 and 7.

*The particular Sirtl etch used was (50gCrO_3 100ml (49%) HF, 100ml H_2O)

Fig. 5 - Comparison of Sirtl etched silicon wofers with ond without gross surfoce domoge. 360 x differentiol interference micrographs.

Fig. 6(a) - Differentiol interference micrograph of silicon surfoce after stripping 1.0μm of 1100°C steom grown oxide ond Sirtl etching. 360 x

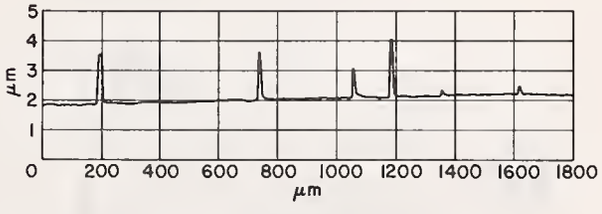


Fig. 6(b) - Tolysurf profile of o sample ofter etching.

The effect of oxidative heating, in contrast to neutral heating, propagates damage sites presumably because the 2:1 volume expansion during oxide formation acts as a stress razor. It is this propagation effect that makes oxidation and etching a valuable tool for the detection of subsurface defects.

Another method[16, 17] of examining oxide layers grown on polished silicon surfaces is to electrolytically decorate oxide

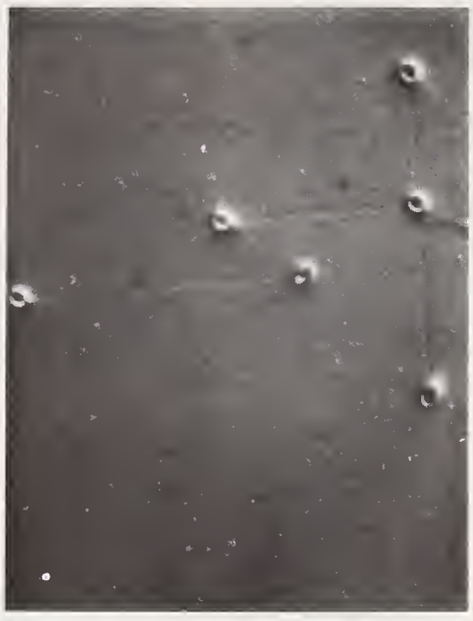


Fig. 7(a)- Differentiol interference micrograph of mechonicolly polished silicon surfoce ofter stripping 0.8μm of 1100°C dry oxygen grown oxide ond Sirtl etching. 360x

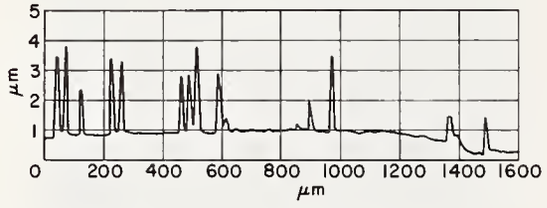


Fig. 7(b)- Tolysurf profile of sample after etching.

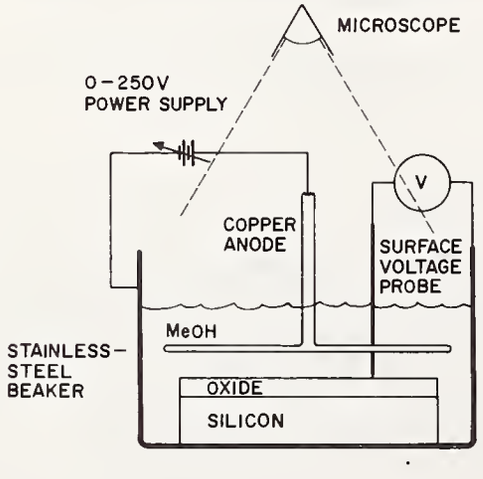


Fig. 8(a)- Copper decorotion opporotus.

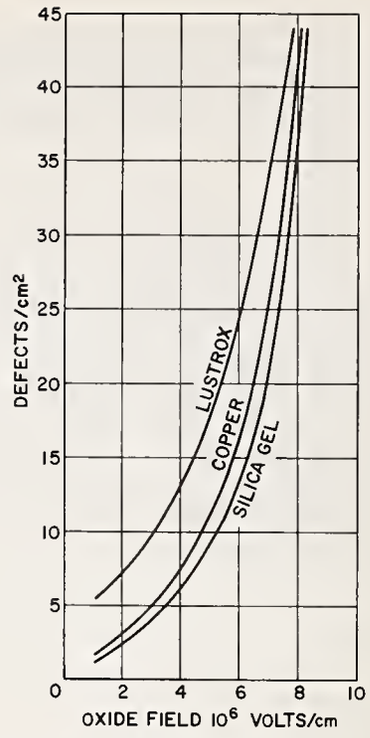


Fig. 8(b)- Field dependence of defect density.

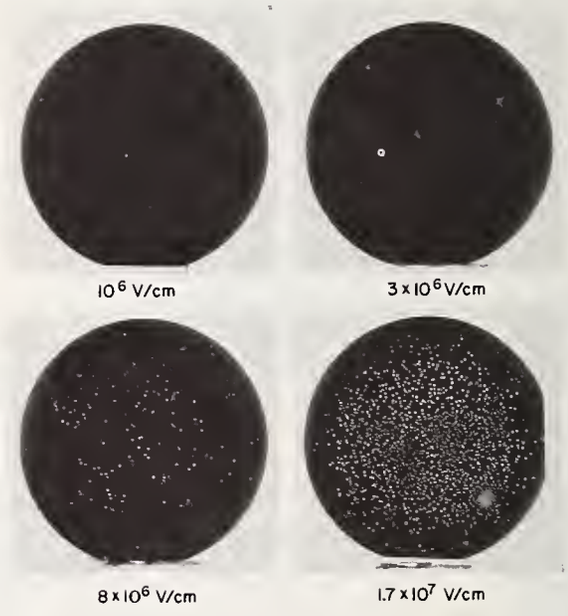


Fig. 9- Silicon wofer with 1000 Å of thermally grown SiO_2 decoroted with copper ot successively higher field strengths.

defects at a known surface potential, as shown in Fig. 8, 9, and 10. The sites revealed by this decoration method are areas of reduced dielectric strength caused by inclusions, dust particles forming glass regions and cracks, thin spots in the oxide caused by differences in the oxidation rates of the inclusions or other silicon defects with their environment and the silicon surface. It was felt that the copper decoration technique is simpler than the capacitor breakdown tests[18, 19] in monitoring self-healing discharges on MOS capacitors. Connecting pinholes in the oxide are revealed by chlorine etching[20] at 600°C, or by etching in boiling aqueous ammonia.

As a final comment, it can be said that the charge trapping usually associated with defects near the Si-SiO_2 interface results in distortion of the ideal shape of the C-V curve of an MOS capacitor.

The Influence of Defects, Inclusions, and Impurities in Bulk Silicon on Surface Properties

Three types of surface problems can derive from bulk silicon defects: structural defects such as low-angle grain boundaries, slip, and stacking faults; precipitated inclusions such as silicon carbide, oxide, or nitride; and, finally, intermetallic compounds precipitated from a solid solution. Severe structural defects in the crystal cannot be removed during normal surface preparation and therefore can lie close enough to the device region to cause charge trapping, to change the surface recombination rate, and also to act as centers for the precipitation of contaminant ions with diffusion coefficients large enough to move to these sites during device processing.

Carbides, oxides[21], and nitrides can occur as precipitates or clusters, although some heat treatments can cause substantial amounts to be taken into solid solution. Precipitates and clusters can occur close enough to the surface not only to cause defective oxide to grow during oxidation because of the differences in oxidation rate, but also to allow the precipitates and clusters to act as charge-trapping and recombination centers. Also, high concentrations of dissolved metals such as gold, copper, nickel, or sodium tend to move rapidly towards discontinuities such as crystal defects, precipitates, and the $Si\text{-}SiO_2$ interface. Again, charge trapping, ion movements, lifetime change, and electrical noise are the most serious effects on device parameters.

The art of crystal growing and selection has advanced far enough that crystal with bulk defects is generally not fabricated into silicon wafers for polishing. It is normal practice for laboratory work to order polished silicon slices by the ingot, and to request that the thick end slices be supplied for reference so that the crystal can be characterized in terms of resistivity, dislocation, oxygen and carbon content, and orientation of the reference flat.

References

1. Archer, R.J., J.Electrochem. Soc., 104, 619, (1957).
2. Goodman, A.M. and Breece, J.J., to be published.
3. Kern, W. and Puotinen, D., to be published in the RCA REVIEW, June 1970.
4. Schramm, J.A., U.S. Patent 3281915 (1966).
5. Integrated Silicon Device Technology, Volume 7, p.192 (1965), Research Triangle Institute, Durham, N. Carolina.
6. Faust, J.W. Jr., "Cleanliness Factors in Mechanical Processes and Etching of Semiconductors", Symposium on Cleaning of Electronic Device Components and Materials, ASTM STP No. 246, Am. Soc. Testing Mats., p.66, (1959).
7. Larrabee, G.B., Heinen, K.G., and Karrell, S.A., J. Electrochem. Soc., 114, 867, (1967).
8. Kern, W., RCA Review, 31, No. 2, (1970).
9. Lieberman, R. and Klein, D.L., J.Electrochem. Soc., 113, 956, (1966).
10. Adams, A.C., and Kaiser, R.H., Abstract 134, Spring Meeting Electrochemical Society, Los Angeles, May, 1970.
11. Snow, E.H., Grove, A.S., Deal, B.E., and Sah, C.T., J. of Appl. Physics, 36, 1664, (1965).
12. Revesz, A.G. and Zaininger, K.H., RCA Review, 29, 22, (1968).
13. Sirtl, E. and Adler, A., Zeitschrift f. Metallkunde, 52, 529, (1961).
14. Fisher, A.W. and Amick, J.A., J.Electrochem. Soc., 113, 1054, (1966).
15. Mayer, A., to be published in the RCA REVIEW, June 1970.
16. Final Report Contract DA36-090 AMC-03617 (E) Elkridge, Md. (U). AD614604. Improvement of Production Techniques to Increase the Reliability of Silicon Planor Epitaxial Techniques, Westinghouse Corp. (1965).
17. Shannon, W.J., to be published in the RCA REVIEW, June 1970.
18. Worthing, F.L., J. Electrochem. Soc., 115, 88, (1968).
19. Klein, N., J. Electrochem. Soc., 116, 963, (1969).
20. Atalla, M.M., Tannenbaum, E., and Scheibner, E.J., BSTJ, 38 749, (1959).
21. Patrick, W.J., ASTM Symposium on Device Processing, June 1970, Gaithersburg Md.

Surface Contamination

J. W. Faust, Jr.[1]

College of Engineering
University of South Carolina
Columbia, S.C. 29208

The difficulty in studying the oxide layer, and the belief that an oxide that looked clean was clean, were the contributing causes to the little activity in the studies on the oxide layers in the early days of silicon. Furthermore, the industry was plagued with material and device fabrication problems. Improvement in these problems led to more sophisticated devices. Oxide layers were purposely grown to act as masks and for surface stabilization. Many device processing steps take place through an oxide layer. Contamination in or on this layer can enter the silicon at elevated temperatures causing device degradation.

New techniques and modifications of existing techniques have been used to gain information about impurities and particulate matter in the oxide layer formed by different techniques. Some of the techniques mentioned are: transmission electron microscopy, electron microprobe, ellipsometry, selected area diffraction, neutron activation analysis, radio tracer studies, etc. These techniques show a variety of impurities. Sodium, potassium, and heavy metals have been identified. In some cases, particulate matter such as abrasive particles are present. Some suggestions for reducing or eliminating these contaminants are given.

Key Words: Silicon, oxides, impurities, thermal oxidation, surface measurement, silicon oxide, contamination, surfaces.

1. Introduction

The first work on silicon and silicon devices was aided by the techniques already worked out for germanium. The transition of techniques was not in reality simple because of the higher temperatures and greater reactivity of silicon. The industry was troubled at first with materials problems, such as purity and defects, and with device fabrication steps. Initially, the materials problems were dominant. Then, as the materials improved, the devices were limited by fabrication steps. This cycle continued over the years as the technology became more and more sophisticated. At first, little attention was given to the surface. It was known, of course, that an oxide layer was present and felt that it was in reality a help in stabilizing the surface. It was also suspected from earlier work on germanium[1][2][2] that certain ions; for example, sodium, were adsorbed or absorbed in the oxide layer. It was generally accepted that if after etching the surface looked clean, that the oxide layer was clean. To further complicate the issue, suitable techniques had not been available for the study of surface layers.

As the materials and device processing techniques improved the devices became more sophisticated. The planar technology brought with it thermally grown oxide layers to act as masks for diffusion. It became apparent that oxide layers formed on silicon by different techniques, in some cases even by the same technique, yielded quite different results. Attention then was directed toward the oxide layer. Aided by new techniques or modification of existing techniques much has been learned about the oxide layer on silicon.

The cleanliness and uniformity of the oxide layer is very critical in that many of the more widely used processes in the fabrication of silicon devices take place through an oxide layer. Even after a treatment to remove the oxide layer is performed, an oxide quickly forms on exposure to environments containing oxygen. This oxide layer was considered to be very thin - the order of a few atomic layers. Many steps in the processing of devices utilize high temperatures; for example, alloying, diffusion, and epitaxial growth. If the oxide layer contains impurities, these impurities can diffuse into the semi-

---

[1]Professor of Engineering.

[2]Figures in brackets indicate the literature references at the end of this paper.

conductor during these processes. Furthermore, impurities and structural faults in the oxide layer may have an influence on the long-term stability of the devices.

In surveying the subject of contamination in semiconductor device and electron tube manufacture, Biondi[3] listed four types of contamination: (a) physical dirt such as dust; (b) water-soluble impurities; (c) water-insoluble impurities, and gases, both adsorbed and absorbed. These areas are broad but as will be seen later should be extended.

It is the purpose of this paper to discuss briefly the various techniques for studying oxide films on silicon and to review the information gained from these techniques.

2. Techniques

Studies of contamination generally involve chemical species rather than structural defects. Thus, the techniques of interest will be those that reveal inclusions, adsorbed impurities, segregation, absorbed species, and the like. For a material such as the SiO_2 layer on silicon, one is not generally interested in structural defects. There is, however, one type of structural defect of importance to devices - namely pores or discontinuities in the oxide layer. Also, one would like to know the thickness and uniformity in thickness of the oxide layers. The techniques described below are not meant to be exhaustive, but merely illustrative.

The thickness of the oxide layer can be determined by ellipsometry and also by the electron microprobe. The uniformity of the thickness of the oxide layer can also be determined by either technique although the microprobe shows variations on a much smaller area (1 to 50 μ as opposed to 1.5 mm.). In the ellipsometric technique, plane polarized light is directed on a clean absorbing substrate at an arbitrary angle. The reflected light will be elliptically polarized. When a thin non-absorbing film covers the surface, the ellipticity parameters Δ and ψ of the reflected light are considerably different. These changes are utilized in measuring the thickness with great precision (i.e., typically $\pm 3A^o$).

The electron microprobe is an instrument for elemental chemical analysis on a scale of a few cubic micrometers; however, it has been shown in the case of silicon that it can be used to study the thickness of the oxide layers[4]. In this instrument a beam of high energy electrons is focused to a 1 μ diameter spot on the surface of the sample to be analyzed. Characteristic x-rays result and are monochromatized by curved crystal x-ray optics and detected by gas filled proportional detectors. X-ray count rates measured for the characteristic x-ray wavelengths are used to determine amounts of each element present.

Radio tracer techniques are used to study the adsorption of various materials. In this technique a radioactive isotope of the element to be studied is put in a suitable solution. The sample is placed in the solution and after a given time is removed, washed, and the activity it has picked up is counted. It is also possible to make autoradiographs to investigate any possible unevenness in the adsorption. One can also obtain information on the distribution of the elements in the layer by successive removal of the oxide and counting the etch solution (either dilute HF or buffered HF) and rinse waters.

Neutron activation analysis gives information about the various impurities in the oxide layer and also about the distribution within the layer. In this technique the sample is sealed in a quartz tube and irradiated with thermal neutrons at a low temperature. The sample, after removal, is allowed to sit until the Si radiation is down. Successive layers of oxide are removed as above and the etch and rinse waters counted for whatever elements are desired.

Studies of ion scattering have been used to study mainly impurities in the bulk of a material. Recently, it has been adapted to studying impurities on the oxide layer and also the thickness of the oxide layer[5]. The technique consists of bombarding the surface with 100 to 200 keV He^+ ions in a vacuum of 10^{-7} to 10^{-8} torr. The back scattered ions are measured over a mass spectrum. Peaks at the various mass numbers identify the element and amount. This technique is still being perfected but appears to be very powerful as will be seen later.

Other techniques which have been used to study impurities in the oxide layer are x-ray fluorescence analysis, spark source mass spectroscopy, and laser beam mass spectroscopy.

The above techniques give information about the chemical nature of the contaminants but do not give any indication of particulate matter. This must be done by other means. The optical microscope, replica electron microscopy and scanning electron microscopy can be used to observe surface irregularities. Particulate matter can best be studied by transmission electron microscopy. If the volume fraction of the particulate matter is large enough, the contamination can be identified using selected area electron diffraction. To study particulate matter in the oxide layer by transmission electron microscopy, it is necessary to remove the oxide from the material. Although this is not generally possible, it is in the case of silicon. The silicon can be etched away in hot Cl_2 gas without hurting the oxide layer. The oxide is then floated off on water and carefully picked up on electron microscope grids[6]. This technique can be used on oxide layers from 30 to 3000 A^o.

3. Survey of Results

Silicon is always covered with an oxide layer except in ultrahigh vacuum after special surface treatment. Even when the silicon undergoes a treatment to remove the oxide, such as HF, it quickly forms on exposure to air and according to Archer[7] builds up slowly for a day. The oxide layer after etching or mechanical polishing is transparent and was thus originally thought to be very thin, of the order of a few atomic distances. Thicker layers do show colors. Thermally grown oxide layers used in planar technology are opaque. Using special treatments a brown layer can form on the silicon surface. This layer has the average chemical composition, SiO, and was reported to be SiO. Several people, using different techniques, have shown that this layer is actually Si and SiO_2. One can remove the layer with HF and obtain a clean looking surface. Examination with an optical microscope using dark field illumination, however, shows the surface to be covered with a large number of particles. When such a surface is examined with transmission electron microscopy, it is seen that these particles are shaped like burrs. Selected area diffraction patterns of such particles shows only a silicon pattern[8]. The electron microprobe shows this more definitely[9]. Furthermore, from such data, one can estimate the relative amounts of the two components. The same instrument was used to show that thermally grown oxides were indeed SiO_2 as compared with peaks from Quartz[4].

The thickness of the oxide layer, its uniformity, and surface topgraphy such as the uniformity of etched trenches and pores in the oxide are not contamination, but for completeness, these topics will be briefly discussed here because of their importance to devices. Perhaps the most precise method of measuring the thickness is by ellipsometry. Other investigators use a series of standard colors calibrated by ellipsometry to estimate the thickness of layers that are thick enough to have color. Such a technique was used to measure the thickness of oxide layers by Yon, et al[10] in their study of the distribution of Na in thermally grown SiO_2 by successive etching. Knausenberger et al[4] reported that the electron microprobe could also be used to measure thickness by measuring the counts per second given off by the oxygen Kα x-rays. This had to be calibrated against the ellipsometric technique. When the counts per second were plotted against the thickness measured by ellipsometry, a linear (3) relation was found over the range from 50 A° to 1600 A°.

The electron microprobe is very powerful for investigating the uniformity of thickness of a layer. Although beam diameters of approximately 50 μ were used in the thickness studies mentioned above, a beam of 1 μ diameter can be obtained. When such a beam was used to make spot checks over the layer, the count rate did not change, indicating a very uniform layer. Transmission electron microscopy cannot be used to determine oxide thickness. It is important, however, in studying "tears" and large pores in the oxide layer[11]. This technique also showed pin holes along the edges of trenches etched in thermally grown oxides. Such pin holes may play a role in the lateral diffusion found under the trenches[11]. It has been reported that defects (i.e., large pores) can be detected by use of an electron microprobe[12). Small pores in the oxide layer can be detected by the etch pits formed on etching at elevated temperatures in gaseous chlorine[13]. Such treatment does not effect the oxide layer but does attack the silicon. Studies of this type have shown that many oxide layers have small pores in them. Pin holes were also reported in low temperature pyrolytic SiO_2 on silicon. The pin holes were detected by an electrophoretic decoration technique[14]. How deleterious these small pores are is open to question. Certainly more studies should be made.

Optical microscopy, both replica and transmission electron microscopy, and scanning electron microscopy, were used to study the uniformity of etched trenches in thermally grown oxide layers[11]. Optical microscopy generally showed that the trenches or windows looked clean with uniform edges except occasionally one would find areas where the trench had not etched due presumably to areas where the mask material had not been properly removed. When the same trenches were examined at high magnification using replica electron microscopy, the oxide still looked clean and uniform but the edges were seen to be serrated. Such serrations are probably of little importance to wide trenches and large windows, but may indeed have an adverse effect on very narrow trenches. The oxide layer was removed from the same samples by etching in chlorine gas and examined under transmission electron microscopy. The serration on the edges were still evident; however, instead of the clean looking oxide, "dirt" and holes were found in the oxide in the trenches in addition to the pin holes along the edges as mentioned above. The term, "dirt", was used for lack of a better name because it was not possible to identify it using selected area electron diffraction because of the small volume ratio. The electron diffraction patterns showed broad diffuse rings characteristic of amorphous material. Heating samples with such a thermal layer to temperatures used in diffusion (but with no diffusants present) in vacuum from 2 to 90 hours, resulted in a thermal etching of the surface and "dirt" as shown by transmission electron microscopy. Selected area diffraction of these oxide layers still showed the ring pattern; however, the rings were not as broad, suggesting a slight ordering of the oxide.

Stickler and Faust[6] studied the cleanliness of oxide layers on silicon formed in abrasion and mechanical polishing, several acid etchants, and hydroxide etchants. The oxide layers were removed from the silicon as mentioned previously. Examination of the oxide formed on abrasion showed a large amount of "dirt" and abrasive particles in the layer. The abrasive particles were proven to be so by selected area electron diffraction. Although the thickness could not be determined, it was possible to compare the thicknesses of the various oxide layers. The layers, after abrasion, were found to be

thicker than expected. Not as thick as the oxide layers after etching, but approximately as thick as those after alkaline etching. Cocca and Carroll[12] found high concentrations of aluminum at localized spots on abraded surfaces and ascribed this to abrasive particles imbedded in the surface.

Examination of layers formed by several different acid etches showed similar results, i.e., "dirt" and abrasive particles. The "dirt", however, was not as much as found in the layers after abrasion. Even after five minutes of etching in CP_4, the abrasive particles were still found. Such a long etching time removes quite a lot of material. Perhaps the alumina remains close to the surface by electrostatic attraction during etching. If an acid etched surface is etched on a hydroxide etchant, such as KOH, the oxide layers appear very clean and free of particulate matter.

The transmission electron microscope technique is a very powerful one for particulate matter. Special treatments are sometimes given to the oxide layer to help stabilize the surface. One such treatment is to fuse lead oxide with the silicon oxide layer. Examination of such a surface by optical microscopy and replica electron microscopy indicated a good layer, presumably homogeneous. Examination of the same layer by transmission electron microscopy showed it to be very inhomogeneous, possibly a two phase layer[11].

Transmission electron microscopy is very useful in detecting particulate matter but gives no information about chemical species adsorbed or absorbed in the layer. A number of techniques have been used to identify individual impurities. There are two general modes of attack on this problem. One is to put a suspected impurity into a solution, in a gas, or on the surface and then use some technique to investigate if the impurity was taken up by the oxide layer. The other technique is to identify the impurities already in the oxide layer after various device fabrication steps. Usually in the former, radio tracer techniques are used. Radioactive sodium was used in solution to show that sodium was adsorbed on the layer surface and may enter the lattice. Sodium in the oxide layer is detrimental to MOS devices. Several investigators, [10], [15], [16], [17], reported that sodium at the air-oxide interface may be removed by an acid or water rinse.

Radio tracer techniques were also used to study water contamination in the oxide layer. In one study[18], samples of silicon with an oxide layer were boiled in tritiated water and the profile of the water contamination studied. Holmberg et al[19] studied the profiles of water contamination in thermally grown oxide layers under conditions approximating device operation as a function of temperature (50° to 250° C), time (1/4 to 8 hr.), and pressure (25 to 400 Torr water vapor). The hydration was carried out using tritiated water. The effect of hydration on the diffusion of sodium (using radioactive sodium) and of sodium in the diffusion of water in the oxide layer were also studied. In all cases they found a high concentration of hydrogen at the air-oxide interface. Within the layer the concentration is lower and fairly uniform. In some samples the concentration increased slightly at the oxide-silicon interface. The diffusion coefficient was much lower than expected from studies on bulk silica. This was explained by the layer having an interlacing pore sturcture. The activation energy was found to be low, and it is suggested that this means that the hydration rates could be appreciable at room temperature. Water was found to enhance the sodium transport and positive charge production at moderate temperatures. On the other hand, sodium enhances the water transport at device temperatures. That water has a detrimental effect on the electrical properties has been shown by several investigators [20], [18], [21].

The distribution of sodium in the oxide layer was studied by neutron activation analysis[10], [15]. This analytical technique demonstrated that the sodium had a U-shaped distribution with the highest concentration at the air-oxide interface.

Cocca and Carroll used the electron microprobe to study defects and impurities in thermally grown oxide layers[12]. They report that impurities uniformly distributed throughout the layer could not be detected unless they were in concentrations greater than $10^{19}/cm^3$. In addition to the aluminum already mentioned, they found defects (i.e., pin holes) in the thermally grown oxide layer. At the sites of these defects they always found a high concentration of sodium and/or potassium. The concentration was of the order of $10^{21}/cm^3$. Occasionally, magnesium and calcium were found in concentrations an order of magnitude lower. No gold or chlorine were found in the defects. No explanation was given for the source of sodium, but it was reported to be removed by hot HCl. The relative values of various impurities observed in the defects agrees with the ion beam mass spectrometer studies of Gorton[22]. Gorton's values, however, were approximately two orders of magnitude lower due presumably to the much broader beam used by him.

Even the purest of chemicals used in device manufacture contain minute amounts of impurities. Heavy metal ions that lie below silicon in the electromotive force series can replace silicon and be an impurity in the oxide layer or at the Si-oxide interface. Such metals are found, for example, in reagents used for etching[23]. Bemski and Struthers[24] using neutron activation analysis reported 10^{12} to $10^{13}/cm^2$ of gold on the surface of silicon. This was enough to cause substantial degradation in the lifetime of carriers in silicon upon subsequent heat treatment. On the other hand, Carlson[25], using a different treatment but the same analytical technique, found only $10^{11}/cm^2$ of gold on silicon surfaces. He also suggested the presence of other fast diffusers such as iron, copper, manganese, and zinc. Buck reported that the ion beam scattering technique can show impurities on the surface,

especially the heavy elements[5]. He was able to show the presence of 1/1000 of a monolayer of gold purposely put on the surface. This was confirmed by neutron activation analysis. Thicker layers of gold (4 A°) were confirmed[12] by x-ray fluorescence analysis. He also identified iodine and its effect on the oxidation of silicon. By measuring the oxygen peak, he was able to estimate the thickness of the oxide layer after etching (23 A°). Ellipsometric measurements were in good agreement with this value.

Nitride layers are grown on silicon in some cases rather than oxide layers. It is beyond the scope of this paper to discuss contamination in these layers. It is, however, necessary to mention nitrogen as a contaminant. If nitrogen is present as an impurity in a vacuum system or gas stream at elevated temperatures, silicon nitride can form on the silicon surface. This was shown by transmission electron microscopy[26]. It was shown that silicon nitride could grow over SiO_2 but that SiO_2 could not grow over silicon nitride. The presence of silicon nitride was proven by selected area electron diffraction.

4. Preventative Measures.

It is not possible, in reality, to give general preventative methods for eliminating the contamination given above because conditions vary from plant to plant and from location to location. From the number and types of contaminants, it appears obvious that one must work in as clean an environment as possible and with the purest of chemicals. Water soluble impurities adsorbed on the surface can be removed by highly purified water. It was mentioned earlier that hot HCl would remove sodium[12]. Some places use other chemicals such as versene. It must be pointed out, however, that caution must be exercised in using other chemicals; they may have other types of impurities. Organics may be removed by rinsing in highly purified methanol, for example. The purity of methanol was discussed by John[27]. It seems best not to let heavy metal ions get in contact with the surface since they are extremely difficult to remove. Techniques have been given for reducing or removing these trace elements in etching reagents[23].

One method of removing abrasive particles has been given. It appears that the abrasive particles are held so tenaciously that even ultrasonic treatment will not remove them. Another treatment that has been reported to remove the abrasive particles is the use of a copper etch. It is reported that HNO_3 will remove the copper completely. When this was tried by the author, it was found by spark source spectrometry that approximately four monolayers of copper were left. The best solution to the contamination problem is not to let the oxide become contaminated in the first place.

5. References

[1] Keyes, R. J. and Maple, T. G., Variation of the recombination velocity of germanium surfaces with ambient gases, Phys. Rev. 94, 1416(A) (1954).

[2] Wolsky, S. P., Rodriguez, P. M. and Waring,W., Adsorption of sodium ions by germanium surfaces, J. Electrochem. Soc. 103, 606 (1956).

[3] Biondi, F. J., Cleaning electron device parts, Bell Laboratories Record, August 1958, p. 288.

[4] Knausenberger, W. H., Vedam, K., White, E.W., and Zeigler, W., Thin film characterization by electron microprobe and ellipsometry: SiO_2 films on silicon, Appl. Phys. Letters 14, 43 (1969).

[5] Buck, T. M., and Wheatley, G. H., Studies of silicon surfaces with 100-200 keV He^+ and H^+ ion beams, presented at AVS Surface Science Symposium, Seattle, Washington, October, 1969.

[6] Stickler, R. and Faust, Jr., J. W., The perfection and cleanness of oxide layers on silicon, Electrochem. Techn. 2, 298 (1964).

[7] Archer, R. J., Optical measurement of film growth on silicon and germanium surfaces in room air, J. Electrochem. Soc. 104, 619 (1957).

[8] Faust, Jr., J. W. and Stickler, R., unpublished results.

[9] White, E. W. and Roy, R., Silicon valence in SiO_2 films studied by x-ray emission, Solid State Commun. 2, 151 (1964).

[10] Yon, E., Ko, H., and Kuper, A. B., Sodium distribution in thermal oxide on silicon by radiochemical and MOS analysis, IEEE Trans. Electron Devices ED13, 276 (1966).

[11] Stickler, R. and Faust, Jr., J. W. The effect of device processes on the perfection of silicon oxide layers, Electrochem. Techn. 4, 277 (1966).

[12] Cocca, F. J. and Carroll, K. G., Electron microprobe analysis of impurity heterogeneities in thermally grown silicon oxide, IEEE Trans. Electron Devices ED15, 962 (1968).

[13] Atalla, M. M., Semiconductor surfaces and films; the silicon-silicon dioxide system, in Properties of Elemental and Compound Semiconductors, p. 163 (Interscience, 1960).

[14] Doo, V. Y. and Sun, V. M. L., Pin holes in pyrolytic oxide deposited on silicon and metals, Met. Trans. 1, 741 (1970).

[15] Fewer, P. R. and Carlson, H. G., Report on surface studies, Rome Air Development Center, Griffiss AFB, Rome, New York, Techn. Rept. RADC-TR-66-776 (March, 1967).

[16] Buck, T. M., Allen, F. G., Dalton, J. V., and Struthers, J. D., Studies of sodium in SiO_2 films by neutron activation and radio tracer techniques, J. Electrochem. Soc. 111, 862 (1967).

[17] Kooi, E., The surface charge in oxidized silicon, Philips Res. Repts., 21, 477 (1966).

[18] Hofstein, S. R., Proton and sodium transport in SiO_2 films, IEEE Trans. Electron Devices, ED14, 749 (1967).

[19] Holmberg, G. L., Kuper, A. B., and Miraldi, F. D., Water contamination in thermal oxide on silicon, J. Electrochem. Soc. 117, 677 (1970).

[20] Kuper, A. B. and Nicollian, E. H., Effect of oxide hydration on surface potential of oxidized p-type silicon, J. Electrochem. Soc. 112, 528 (1965).

[21] Hofstein, S. R., An investigation of instability and charge motion in metal-silicon oxide-silicon structures, IEEE Trans. Electron Devices ED13, 222 (1966).

[22] Gorton, H. C. et al, Battelle Memorial Institute, Columbus, Ohio, 2nd ECRC Res. Rept., Feb. 15, 1967.

[23] Faust, Jr., J. W.,Cleanness factors in mechanical processes and etching of semiconductors, ASTM Special Techn. Pub. 246, p. 66 (1958).

[24] Bemski, G. and Struthers, J. E., Gold in silicon, J. Electrochem. Soc. 105, 588 (1958).

[25] Carlson, R. O., Gold on silicon surfaces, J. Appl. Phys. 29, 1001 (1958).

[26] Stickler, R., Private communication.

[27] John, H. F., The preparation of high-purity methyl alcohol for use with semiconductor devices, ASTM Special Techn. Pub. 246, p. 195 (1958).

The Precipitation of Oxygen in Silicon, and its Effect on Surface Perfection

William J. Patrick

IBM Components Division
East Fishkill Facility
Hopewell Junction, N. Y. 12533

The behavior of oxygen in silicon grown by the Czochralski technique has been investigated quite extensively. When a sample is annealed at a temperature such that an appreciable oxygen supersaturation is maintained, precipitation processes occur if there is sufficient thermal energy available to allow appreciable diffusion. The precipitation of the oxygen is related to the formation of crystal defects which can intersect the surface and thereby affect surface perfection. Since, in the fabrication of complex integrated circuits, substrates are often subjected to temperature-time cycles which cause oxygen precipitation, the associated defects can affect the crystalline perfection of epitaxial layers grown after their generation.

In this paper we review the behavior of oxygen in silicon and report on some experimental observations on the nature of the precipitated phase and the kinetics of its growth. The defects appear to be related to the surface defects reported by a number of investigators to occur during steam oxidation of silicon. The nature of the defects as revealed by metallographic and transmission electron microscopy techniques is discussed. Finally, experimental evidence of their effect upon the perfection of epitaxial layers is presented.

Key Words: Crystal defects, oxygen precipitation, phase transformations, surface perfection.

1. Introduction

Oxygen is an impurity normally found in silicon single crystals grown from a melt contained in a silica crucible[1][1]. A reaction between the melt and crucible yields SiO, which dissolves in the melt and is picked up in the growing crystal. Oxygen, dissolved in the silicon lattice, intertupts a Si-Si valence bond, resulting in a nonlinear Si-O-Si configuration [2]. Maximum solubility of the oxygen apparently occurs at the melting point and is about 2×10^{18} oxygen atoms/cc [3]. The solubility decreases with decreasing temperature, reaching a value of approximately 2×10^{17} atoms/cc at 1000°C. The oxygen concentration in solution can be larger than the equilibrium value, giving rise to a supersaturation which tends to drive the oxygen out of solution. An actual decrease in dissolved oxygen depends not only on the driving force for precipitation, but also on the available thermal energy, since a number of energy barriers must be overcome before a second phase can form. There barriers can, for example, be due to differences in structure, composition and molar volume between the precipitate particle and the surrounding silicon matrix. A certain amount of energy is required to form the surface of the particle; and finally, the diffusion process whereby the oxygen atoms are brought out of the surrounding matrix into the vicinity of the growing particle requires a certain activation energy.

In this paper, results of a number of studies are discussed, the objectives of which were to provide some insights into the details of oxygen precipitation and determine if this could have undesirable effects in integrated circuit processing. The precipitation of excess oxygen occurs in two steps: nucleation, and then growth of the pre-

[1] Figures in brackets indicate the literature references at the end of this paper

cipitate particles. The results given here are concerned with the growth process. Some of the details of the IR spectrum of oxygen in silicon will be reviewed since IR absorption measurements were the primary investigation tool in the precipitation kinetics studies. Since differences exist in the physical properties of the precipitate particle and the surrounding silicon matrix, accommodation of the particles can give rise to the formation of dislocations. Metallographic studies which revealed these defects are reported here.

It has been pointed out that p-n junction quality may be degraded by the presence of oxygen precipitate particles in the depletion region [4]. However, in some integrated circuit processing schemes, the active areas of the devices are confined to an epitaxial layer grown on the oxygen-contaminated substrate. In these instances, oxygen precipitation in the vicinity of p-n junctions is probably rare. Lattice damage in the substrate occurring prior to epitaxy during high temperature processes can, however, affect the defect density in the epitaxial layer. This can, in turn, affect device properties without the direct introduction of oxygen precipitate particles into the epitaxial layer. We are thus interested not only in the production of these defects during oxygen precipitation, but also in their intersection with the surface just prior to epitaxy. Experimental results of investigation on the effects of the defects on epitaxial layer perfection will be presented in the last portion of this paper.

2. IR Spectra

Infrared absorption spectra were obtained using a double-beam grating spectrophotometer. Oxygen-free wafers, matched in thickness to the sample wafers to within 1%, were used in the reference beam to cancel the silicon lattice absorptions. Sample thicknesses were typically 2mm. Figure 1 shows the absorption band at 1106 cm^{-1} due to oxygen dissolved in the silicon. The absorption coefficient is obtained from the experimentally observed quantities by the relationship

$$\alpha = \frac{1}{t} \log_e \frac{T_o}{T} ,$$

where T is the percent transmission observed, T_o the percent transmission when absorption band is not present (estimated by the baseline method [5]), and t the sample thickness. This relationship neglects multiple reflections within the sample, and the observed α's are thus from 6 to 10% higher than the true values depending upon the band intensity and the sample thickness. This accuracy was considered adequate for these studies. The true absorption coefficient has been demonstrated to be a linear function of the oxygen concentration. In the instances where the oxygen concentration is given here, it was computed from the observed absorption coefficient by the relationship

$$[O] = 2.73\alpha \times 10^{17} \text{ atoms/cc} . \quad (1)$$

Slit widths used were one-tenth the half band width in the room temperature spectra.

The oxygen band changes shape and position when observed at low temperature (see Fig. 1). The low temperature spectra were taken in a twin-tail dewar with liquid nitrogen as the coolant. This arrangement allowed cancellation of the lattice bands, as before, by using oxygen-free reference samples.

When the sample is annealed for lengthy periods such that an oxygen supersaturation exists, the intensity of the 1106 cm^{-1} band decreases and its shape changes [6]. Investigations into the nature of the precipitated phase have shown that two distinct spectra appear upon prolonged heat treatment [7]. The first is characterized by the appearance of a new band at about 1225 cm^{-1} as shown in Fig. 2. When examined at 80°K. the band at 1106 cm^{-1} is seen to change in the same manner as shown for dissolved oxygen in Fig. 1. We conclude, therefore, that in this instance the intensity of the band at 1106 cm^{-1} represents the concentration of dissolved oxygen alone and the 1225 cm^{-1} band is due to the precipitate alone. This IR spectrum does not correspond to any of the published spectra for SiO_2 polymorphs [8]. Figure 3 shows, in addition to dislocation networks, small rectangular or square strain contrasts which can be identified with small coherent particles. These particles have been seen only when the 1225 cm^{-1} band is also present.

The other type of spectrum characteristic of precipitated oxygen is shown in Fig. 4. The 1107 cm^{-1} band is considerably broadened and reference to the spectrum at 80°K shows the precipitates give rise to a band which lies directly over the dissolved oxygen band at 290°K. This spectrum can be formed by up-quenching (to 1150°C or

above) a sample containing the first type precipitate, then re-annealing at approximately 1000°C. When the oxygen concentration is above about 12 x 10^{17} atoms/cc, or the anneal temperature above 1100°C (still maintaining an oxygen supersaturation, however), this spectrum is formed during the initial heat treatment. TEM studies have not given any definite indication of the morphology of this type precipitate. The IR absorption spectrum near 1100 cm^{-1} is quite similar to that of a thermally grown surface oxide layer[8].

Most samples investigated have shown mixtures of these two types of spectra; that is the 1106 cm^{-1} band broadens slightly while the 1225 cm^{-1} band increases in intensity. Sometimes the 1225 cm^{-1} band is not well separated from the 1106 cm^{-1} band, appearing almost as a shoulder. Occasionally, the 1225 cm^{-1} band appears only briefly before the second type spectrum appears. Once this surface oxide type spectrum is formed, no anneal cycles have been found which result in the formation of the first type.

3. Precipitation Kinetics

The change in intensity of the IR absorption bands in the vicinity of 1100 cm^{-1} were measured as a function of annealing time in a number of samples. These wafers were heat-treated for two-to five-hour cycles, with the IR spectrum taken between annealings. Wafers from two crystals were chosen as being typical of moderate oxygen concentration samples. Both crystals were dislocation-free and doped with 3 to 4 x 10^{14}B atoms/cc. In addition, both contained 7 x 10^{17}O atoms/cc and TEM examination showed they were almost completely free of precipitate particles before annealing. These samples received the pre-treatments summarized in Table I.

Table 1. Pre-treatment of silicon wafers

Sample	04A	04B	56A	56B
Pre-treatments	700°C/2 hrs plus 1300°C/15 mins	700°C/2 hrs	1300°C/15 mins	None

The decrease in intensity at 1106 cm^{-1} is shown in Fig. 5. Plotted here is the fraction excess oxygen remaining after a time t, where:

$$\frac{C(t) - C_f}{C_o - C_f} = \frac{\alpha(t) - \alpha_f}{\alpha_o - \alpha_f} .$$

Here, C_o is the initial oxygen concentration, C_f the concentration when precipitation was complete, and the α's the corresponding absorption coefficients. Measurements were made approximately every three hours up to 45 hours. The last two measurements were made at 115 and 185 hours to determine α_f.

Samples 56A and 56B behaved almost identically except for a slight difference in slope of the straight-line portion of the curve after approximately 20 hours. This latter portion of the curves is an exponential decay and can be fitted to a formula of the type $Ae^{-t/\tau}$. For comparison, the values of τ obtained are indicated. The initial transient occurred sometime during the first 5-hour anneal cycle and its exact shape is not known. The main precipitation began after about 10 hours for both samples.

Samples 04A and 04B showed considerably different behavior. Sample 04A, which received a 15-minute 1300°C treatment, started precipitation at t=0 and continued to completion without any incubation period. Sample 04B had a 12-hour incubation period and no initial transient.

These data closely fit Ham's theoretical formula for the diffusion-limited growth of spheroidal particles [9]. The solid curves in Fig. 4 for samples 04A and B were plotted using this theoretical relationship and were fitted to the data points in the vicinity of 45 hours. Since Ham's theory applies to particles growing from t=0, the time scale of 04B had to be shifted 12 hours to account for the incubation period. Similarly, if for samples 56A and 56B, C0 is taken as the value on the plateau after the initial transient decay, and the time scale is shifted 10 hours, then the experimental data also fits this relationship.

Two other samples were chosen as representative of high oxygen concentrations. Figure 6 shows the decrease in intensity at 1106 cm^{-1} observed for these samples. Sample 69 showed a very large initial transient followed by an exponential decay. Instead of approaching a constant value α_f, however, the absorption coefficient increased

slightly after 10 hours, giving the tail of points. This apparently was due to the formation of the absorption bands directly on top of the 1106 cm^{-1} dissolved oxygen band. Also, during these anneal cycles, the 1225 cm^{-1} band became quite prominent from about 10 to 16 hours, but disappeared rapidly after 18 hours. Sample 95 behaved more like the moderate concentration samples but with a τ of 9 hours.

Two samples from crystals 04 and 56 were heat-treated at 450°C for a total of 750 hours and the changes in the 1106 cm^{-1} band were followed. The band intensity decreased and the band broadened. These data also fitted Ham's formula for diffusion-limited precipitation but the values of τ were very long — 270 hours for 04, and 370 hours for 56. As pointed out earlier, these crystals were almost completely free of precipitates due to other causes. After these prolonged heat treatments, no evidence of precipitation was visible by TEM examination. The process giving rise to the spectral changes must have involved the rearrangement of the oxygen atoms as a local scale with no second phase formation. The thermal energy was insufficient to overcome the energy barriers to nucleation. An interesting observation made during these low-temperature anneals was the gradual disappearance of the band at 607 cm^{-1} due to carbon [10]. Whether or not this behavior is attributable to oxygen remains to be shown. However, annealing at 1000°C does not seem to affect the 607 cm^{-1} carbon band.

It has been demonstrated [6] that the introduction of dislocations by plastic deformation can greatly accelerate the decrease in intensity at 1106 cm^{-1}. It has also been reported [11] that dislocations formed in the crystal growth process tend to retard the oxygen precipitation. Some observations were made on the effects of dislocation density on precipitation rate in this study. In four different crystals with etch pit densities greater than 5000 per cm^2, no oxygen precipitation was noted upon heat-treating for up to 100 hours. The oxygen concentrations ranged from 7×10^{17} to 13×10^{17} atoms/cc in these samples. Precipitation could be initiated only by a brief (15-minute) anneal at approximately 1300°C. Heat treatments at 1000°C after this resulted in the formation of precipitates. Thus, in these cases the grown-in dislocations inhibited precipitation.

4. Defects Introduced

The precipitation process is accompanied by the formation of crystal defects. These consist mainly of dislocation networks as already seen in Fig. 3, dislocation loops, and, in some instances, stacking faults. These defects were routinely observed by means of a 2-minute Dash etch [12]. The samples were first etched in a rotating cup with $3HNO_3 : 2CH_3COOH : 1$ HF to remove any surface defects. Typically, 100 to 200 μm were removed before applying the Dash etch. Figure 7 is a photomicrograph of the etch figures typically seen. In the center of the photo is a loop approximately 30 μm in diameter. The plane of the surface is (111) and the line figures lie along $\langle 110 \rangle$. Examination of sections made of these wafers show the lines are the intersection of loops, identical to the one in the center, with (111). These loops lie on $\{ 11\bar{1} \}$. What appears to be a dislocation network can be seen within the loop. The line figures are in many instances identical to those observed by Joshi [13] on wafer surfaces after steam oxidation. These were identified as stacking faults which were generated during the steam oxidation process within a few microns of the surface. The defects in Fig. 7 were generated in the bulk of the silicon.

The precipitation kinetics data can be fitted to Ham's relationship describing diffusion-limited precipitation. More information is available from this theory since it also gives a relationship between the exponential decay portion of the curve to the number of precipitate particles. In Ref. 9 it is shown that for spheroidal particles we can write

$$\frac{1}{\tau} = \left[\left(3 \frac{\rho_o}{\rho_c} \right) \left(4\pi N_p \right)^2 A \right]^{1/3} D.$$

τ is the time constant of the exponential decay and was given in the figures for the precipitation kinetics, ρ_o/ρ_c is the ratio of the densities of the excess oxygen concentration to the oxygen concentration in the precipitated phase, N_p the number of precipitate particles per cm^3, D the oxygen diffusion coefficient, and A a geometric factor which depends upon the shape of the particles. Assuming the particles have SiO_2 stoichiometry results in a ρ_o/ρ_c ratio of about 5×10^{-5}. At 1025°C D is 2.5×10^{-11} cm^2/sec [2]. The value of A is not known. It is, however, 1 for spherical particles, and not very much different from 1 for spheroidal particles with a reasonable eccentricity [9]. Using the experimentally determined value of 9 hours for T in sample 95, we obtain $N_p^{2/3} \simeq 5 \times 10^7$ cm^{-2}. Therefore, roughly 10^7 particles per square centimeter should be visible on the surface after etching. Examination of the etch figures on this sample indicate only 1×10^5 clusters/cm^2 are present. Therefore these etch figure studies, as well as the TEM examinations mentioned above, indicate there is no one-to-one corresondence between observed etch figures and the number of precipitate particles.

5. Effects of Precipitates on Epitaxial Growth

To determine detrimental effects of these precipitates, epitaxial depositions were made on ten substrates heat-treated in various ways to precipitate the oxygen. Eight wafers contained oxygen and two were oxygen-free (Lopex). All were P^-, B doped with resistivities between 15 and 25 Ω/cm, 0.2-cm thick, and mechanically polished on both sides. The 10 wafers received the heat treatments indicated in Table II. After the anneal cycles were completed, a 100-μ layer was chemically etched from each surface to remove the mechanically damaged layer. Samples 1 through 5 were then placed on a small horizontal epitaxial reactor and were heated in H_2 for 5 minutes at 1168°C. This was followed by a 5-minute deposition of undoped Si, also at 1168°C (thickness about 5μm). After the microscopic examination (described later), wafers 3 and 4 were annealed at 1300°C for 1 hour to ensure dissolution of the oxygen precipitates. Then these two wafers, along with numbers 6 through 10, received a 5-μ undoped Si deposit under the same conditions as for numbers 1 through 5.

After deposition of this layer, each wafer was etched 30 secs. in Sirtl etch [14]. Stacking faults and etch pits were counted at 40 positions on each wafer, and the results tabulated are the average of all 40 readings. Ten readings, spaced 1mm apart, were taken from the center of each wafer along 2 lines intersecting at right angles at the wafer center.

Table 2. Substrates used in epitaxial deposition experiments

Wafer Number	Initial Oxygen Concentration (atoms/cc)	Anneal Cycles (Temp °C/time)	Final Oxygen State
1	11.2×10^{17}	none	dissolved
2	11.6×10^{17}	1300°C/15 min	dissolved
3	11.5×10^{17}	1300°C/15 min; 1000°C/64 hr	Ppt (Fig. 2)
4	11.7×10^{17}	1300°C/15 min; 1000°C/64 hr; 1200°C/24 hr; 1000°C/70 hr	Ppt (Fig. 4)
5	$< 10^{16}$	1000°C/70 hr	oxygen-free
6	10.6×10^{17}	none	dissolved
7	11.4×10^{17}	1300°C/15 min	dissolved
8	11.0×10^{17}	1300°C/15 min; 1000°C/65 hr	Ppt (Fig. 2)
9	11.7×10^{17}	1300°C/15 min; 1000°C/65 hr; 1200°C/24 hr; 1000°C/70 hr	Ppt (Fig. 4)
10	$< 10^{16}$	1200°C/24 hr; 1000°C/70 hr	oxygen-free

The observed stacking fault and etch pit densities are given in Table III. Apparently, the damage occurring during precipitation, rather than the precipitate particles themselves, affects the defect densities. Samples 3' and 4' were annealed to dissolve the oxygen precipitates. The observed etch pits did not have a tendency to cluster as they did on the substrates. Their density is about the same as the cluster density. There is thus the possibility that a number of defects, originally present at the surface, were annealed out during the initial heating in hydrogen. In any event, the defect density is much smaller than the number of precipitate particles.

Table 3. Stacking fault and etch pit densities after epitaxial deposition

	Wafer	State of Oxygen before Deposition	Stacking Faults (cm^{-2})	Etch Pits (cm^{-2})
Deposition #1	1	dissolved	140	350
	2	dissolved	100	950
	3	PPt #1	9,100	250,000
	4	PPt #2	1,200	240,000
	5	oxygen-free	260	820
Deposition #2	6	dissolved	400	970
	7	dissolved	270	4,000
	8	PPt #1	4,700	1,200,000
	9	PPt #2	9,300	740,000
	10	oxygen-free	930	690
	3'	dissolved PPt #1	2,700	1,400,000
	4'	dissolved PPt #2	1,300	100,000

6. Conclusions

The primary question asked in these studies was whether oxygen precipitation can affect the processing of integrated circuits. One possible effect that was found was that the precipitation is accompanied by the formation of dislocations. Epitaxial layers grown on substrates containing these defects will tend to have a higher defect density than layers grown on precipitate-free substrates. The number of defects introduced into the epitaxial layer is, however, much smaller than the number of defects observed in the substrates.

Analysis of the precipitation kinetics data was made in the hope of being able to make reasonably accurate forecasts as to whether the oxygen will or will not precipitate given a specific set of process temperature-time sequences. The rate at which the oxygen leaves solution is accurately given in the relationship for $1/\tau$. However, uncertainties exist in Np, the number of precipitate particles. This number is directly related to the nucleation process which must precede the growth phase. Also related to the nucleation process is the existence of the incubation periods in the moderate oxygen concentration samples, and the complete lack of precipitation in the heavily dislocated samples. Further complicating the picture is the fact that the IR spectra indicate that a number of processes can occur during anneal cycles when the oxygen leaves the solution. More information is needed about this very difficult question of nucleation of the second phase before the possibility of predicting the behavior of the oxygen in solution can be realized.

7. Acknowledgements

The author is indebted to V. Lyons, E. Mendel, and M. Joshi for numerous helpful discussions. Particular thanks are due to S. Dash, who performed the TEM examination, to A. Spiro, who made the epitaxial depositions, and finally, to C. Hauerwas and T. Irish, who assisted in the experimental work.

8. References

[1] W. Kaiser and P. H. Keck, J. Appl. Phys., 28, 882 (1957).

[2] H. J. Hrostowski and R. H. Kaiser, Phys. Rev. 107, 966 (1957). Also J. W. Corbett, R. S. Mc Donald and G. D. Watkins, J. Phys. Chem. Solids 25, 873 (1964).

[3] H. J. Hrostowski and R.H. Kaiser, J. Phys. Chem. Solids 9, 214 (1959).

[4] J. W. Faust, H. F. John, R. Stickler in "Physics of Failure in Electronics" Vol. 4, (Rome Air Development Center) AD637529 (1966).

[5] See, for example, R. T. Conley "Infrared Spectroscopy", Chapter 6, Allyn and Bacon (1966).

[6] W. Kaiser, Phys. Rev. 105, 1751 (1957). Also S. Lederhandler and J. R. Patel, Phys. Rev. 108, 239 (1957).

[7] W. Patrick and S. Dash - to be published.

[8] E. R. Lippincott, A. Van Valkenburg, C. E. Weir, and E. N. Bunting, J. of Research of Nat'l Bureau of Stds. 61, (1958).

[9] F. S. Ham, J. Phys. Chem. Solids 6, 335 (1958), and J. Appl. Phys. 30, 1518 (1959).

[10] R. C. Newman and J. B. Willis, J. Phys. Chem. Solids 26, 373 (1965).

[11] I. L. Shul'pina, L. V. Lainer, M. G. Mil'vidskii and E. P. Rashevskaya, Soviet Physics - Solid State 9, 1291 (1967), and also M. I. Iglitsyn, G. P. Kekelidze, L. V. Lainer, and M. C. Mil'vidskii, Bulletin of the Academy of Sciences of the Georgian SSR 45, 77 (1967).

[12] W. C. Dash, J. Appl. Phys. 27, 1193 (1965).

[13] M. L. Joshi, Acta Met. 14, 1157 (1966).

[14] E. Sirtl and A. Adler, Zeitschrift Fur Metallkund 52, p. 529 (1961).

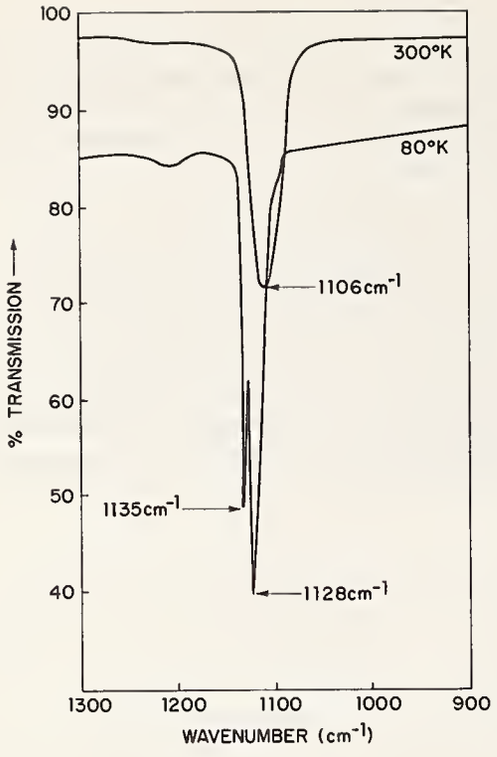


Fig. 1. The 1106 cm^{-1} band due to dissolved oxygen at room temperature and at approximately 80 °K. Upon cooling, a band at 1135 cm^{-1} becomes visible while the 1106 cm^{-1} band shifts to 1128 cm^{-1}.

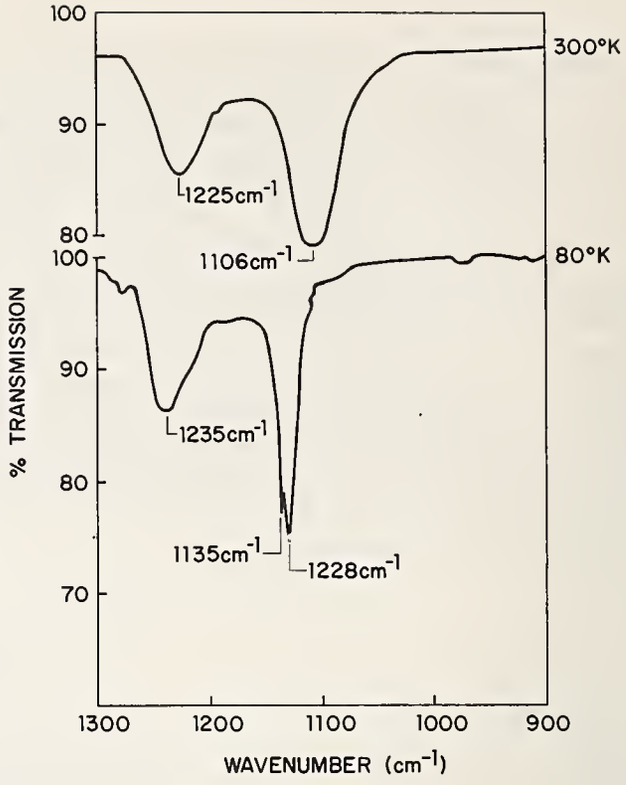


Fig. 2. IR absorption bands due to oxygen precipitates. The band at 1106 cm^{-1} is seen to be due entirely to dissolved oxygen while the new band at 1225 cm^{-1} is due to the precipitates.

Fig. 3. TEM photo of typical defects appearing upon precipitation of dissolved oxygen. Plane of the sample is (100). The small square or rectangular contrasts arise from particles too small to be resolved.

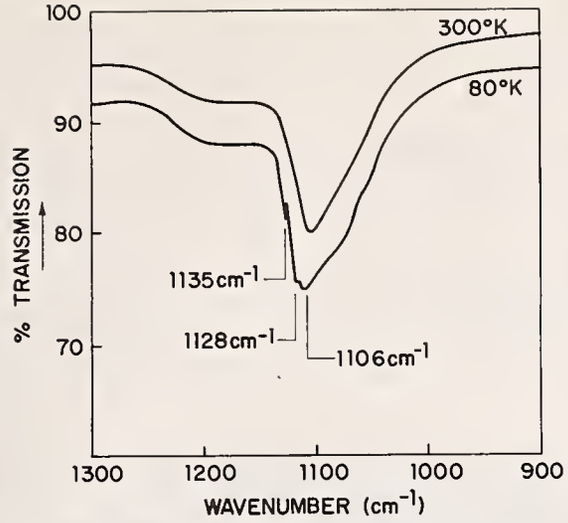


Fig. 4. Second type of IR absorption bands arising from oxygen precipitation. This spectrum resembles that of a thermally grown surface oxide. The low temperature spectrum shows the 1135 cm^{-1} and 1128^{-1} dissolved oxygen bands as the minor peaks on the high frequency side of the major absorption which still occurs at 1106 cm^{-1}.

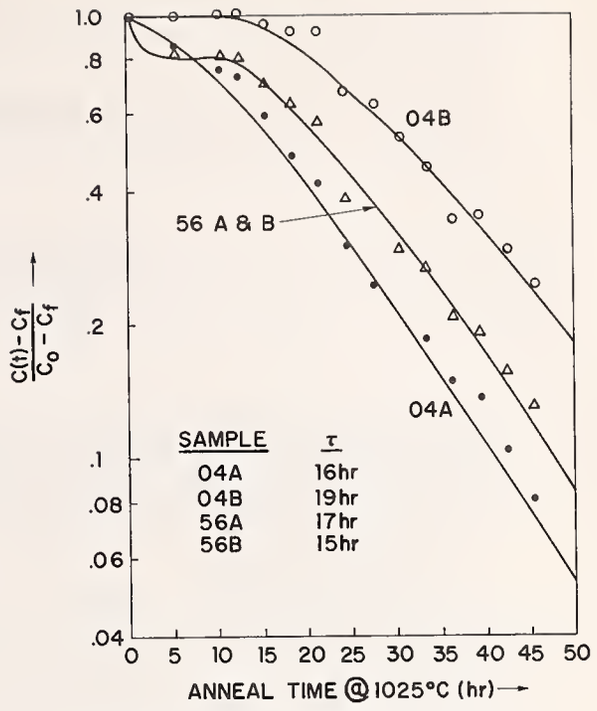


Fig. 5. Precipitation kinetics of moderate oxygen concentration samples. The fraction of excess oxygen remaining in solution is plotted against annealing time. The curves' drawn through the data are from Ham's relationship as discussed in the text. If the decrease at times in excess of 20 hours is plotted as a straight line $Ae^{-t/\tau}$, the values of τ so obtained are indicated. The data points for samples 56A & B lie so close together that only sample 56A is shown.

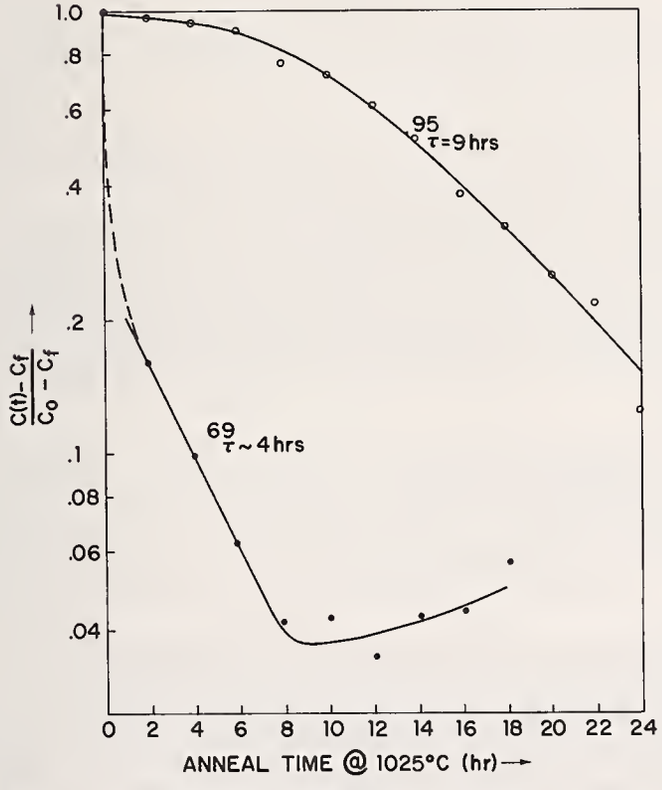


Fig. 6. Precipitation kinetics of high oxygen concentration samples. Sample 69 had an initial concentration near the solubility limit - 18 x 10^{17} atoms/cc. Sample 95 had an initial concentration of 12 x 10^{17} atoms/cc. The shape of the initial transient in 69 is not known since it occurred during the first 2-hour anneal.

Fig. 7. Etch figures produced by a 2-minute Dash etch on a sample in which the oxygen was precipitated by annealing at 1100 °C for 50 hours. Resistivity was 11.8Ωcm; boron doped; initial oxygen concentration of 12.9 x 10^{17} atoms/cc.

Auger Spectroscopy and Silicon Surfaces

J. H. Affleck

Semiconductor Products Department
General Electric Company
Syracuse, New York 13201

This paper reviews the basic principles of Auger spectroscopy, the equipment required and how it may be applied to silicon surfaces. During the course of this study a number of problems have been examined such as the effectiveness of surface cleaning, residues of photoresist material and various solvents, the effect of the processing environment on surface properties and the identity of contaminating films.

Several examples of the application of Auger analysis to semiconductor problems are cited.

Key Words: Auger analysis, silicon.

In 1925, the French physicist, Pierre Auger reported (1) a phenomenon while observing x-ray fluorescence in which electrons of characteristic energies are emitted from an irradiated material. This discovery has recently enjoyed a new interest and provides the basis for a new analytical technique for examining surfaces and determining their elemental composition.

When a material is bombarded with energetic electrons some of the inner shells of the atom may be ionized or loose an electron. Electrons from outer shells falling to the ionized level give up their energy as x-rays. The emitted x-rays as the result of the electronic transition will have energy that is unique for the excited atom. This phenomenon is well known and the basis of x-ray emission spectroscopy. In Figure 1A a simple diagram shows the transition between the K and L shells that give rise to the characteristic K_1 and K_2 x-rays for a particular atom. However, the electron making the transition from the L to K shell may release its energy to an electron in the L shell which is emitted with a unique energy, Figure 1B. This radiationless transfer of energy is known as the Auger effect and the emitted electron as the Auger electron. The energy of the Auger electron in this example is the difference in the energies between the K and L_{II} or K and L_{III} less the energy to ionize the particular level from which it originates.

Although, the Auger effect has been known for many years, it has been only within the past two years that its potential as an analytical tool has been realized (2, 3, 4).

In 1953 Lander (5) suggested that Auger electrons might be useful in material analysis. Later, in 1968, Harris (2) using a system with improved sensitivity demonstrated the usefulness of Auger spectroscopy.

Harris used a deflection type analyzer in which the sample is bombarded with an electron beam. The secondary and Auger electrons being directed through the analyzer as the voltage on the analyzer plates is varied slowly. Consequently, the output of the detector, an electron multiplier, gives an energy distribution curve of the electrons emitted from the sample. An illustration of the energy distribution curve is shown in Figure 2. By applying a small pertubation voltage to the deflection voltage and electronically differentiating the signal and the derivative of the energy distribution is obtained as depicted in Figure 3.

---

[1] Figures in parentheses ( ) indicate the literature references at the end of this paper.

The small Gaussian type peak at the top of the figure may be thought of an Auger transition as it appears on the energy distributuion curve. It will appear in the d N(E)/dE curve as shown where the background has been depressed and the derivative is clearly evident.

Weber and Peria (6) and later Palmberg and Rhodin (7) demonstrated that similar data could be obtained by minor modifications of a low energy electron diffraction apparatus. Using a three or four grid optics system as a retarding field analyzer the energy distribution curve may be obtained by electrical differentiation of the retarding field plot.

In this work a commercially available system has been used. It consists of a high vacuum chamber capable of operating in the 10^{-10} Torr range, an electron gun and retarding field analyzer and the electronics necessary to detect and display the Auger peaks. A spectra of clean silicon is shown in Figure 4. One major peak is seen at 87 ev resulting from $L_{I,II}$ VV transition and two minor peaks at 70 and 53 ev.

The significant feature of this technique is that the spectra originated from atoms on or very near the surface of the specimen. Auger electrons produced more than a few monolayers below the surface are scattered and experience energy losses so that they are undetectable. This feature makes the technique extremely useful in studying surface phenomenon. Since surface properties are so important in silicon device technology an obvious application of Auger analysis is to the numerous processing problems associated with silicon surfaces.

Recent studies have been directed toward the examination of silicon surfaces at various stages in the normal silicon device processing procedure. It has been convenient to examine whole or fractional wafers. Modifications of the sample holder has permitted mounting from 2 to 8 samples. The samples are placed in the high vacuum chamber as soon after a particular treatment or processing step as possible in order to avoid possible contamination. After exhausting the chamber to a pressure less than $5x10^{-9}$ Torr the spectra may be taken. Depending upon the sample it may be desirable to bake the system or heat the sample, however, one must be careful not to further contaminate or alter the surface by such treatments.

The spectra shown in Figure 4 is that of very clean silicon that was prepared by solvent cleaning, hydrofluoric acid etch and immersed in a methonal-iodine solution as described by Lieberman and Klein (7). This treatment minimizes the formation of an oxide film and has permitted the recording of very clean silicon spectra. There is no evidence of the iodine in the spectra since the primary beam is apparently responsible for removing the iodine from the surface. This type of surface provides a good reference and is useful in providing a basis for comparing various processing steps.

For example, in the normal photomasking processes the photoresist is removed by a stripping agent that leaves a residue of carbon (275 ev) on the silicon surface, as shown in Figure 5. Depending upon the type of stripping agent that was used the carbon peak will vary in amplitude frequently surpassing the silicon peak. When initially observed this might be the cause of concern, however, in a normal processing sequence such a wafer usually experiences a cleaning to remove organic and inorganic residues. Wafers monitored after such treatment are free of the carbon residue produced by the stripping agent.

Another technique that is often used to remove photoresist is by plasma ashing. Although this technique is effective in removing organic material it leaves inorganic material on the silicon surface. Using the Auger method of analysis tin is consistently found on wafers that have had the photoresist removed by the plasma ashing technique. Carbon is removed very effectively although a small amount of oxide is formed. Again the inorganic residue need not be a potential problem if it is removed properly before the next step in the process.

An interesting effect has been observed as the oxide film forms on silicon. In Figure 4 where the amount of oxygen is very small, the silicon peaks at 87 ev and 70 ev are distinct and well separated. As the oxygen peak increases as seen in Figure 6 a peak will appear on the low energy side of the 87 ev peak at approximately 80 ev and will continue to increase as the oxygen increases. In Figure 7 it has increased to the point where the Si (87 ev) peak is almost masked completely. This is apparently a chemical shift in the spectra resulting from the Auger electrons originating from a silicon atom in SiO_2 rather Si alone. Oxygen may be present in the spectra without the 80 ev peak. This may be useful in determining the composition of a particular surface and how it should be processed.

Unlike the x-ray emission spectra, the Auger transitions can not be predicted precisely. The spectra can best be obtained empirically. It is, therefore, necessary to run spectra of the elements normally encountered in order to identify the transitions and line shape. For example in Figure 8 the spectra of Chromium and oxygen is shown. The main peaks in both spectra occur near 520 ev but there is considerable difference in the shape and number of the secondary peaks. This fact was used in detecting the presence of Chromium and sulfur on the surface of a device wafer after a chromic acid etch.

This residue severely affected the resist removal in following steps. Improved cleaning procedures were instituted and successfully removed the residue, however, it was necessary to identify chromium in the presence of oxygen by the shape of the spectral lines.

At present a number of processing procedures are under investigation such as determining the effectiveness various types of solvents in removing wax residues, comparing different cleaning techniques after wafer polishing and determining the optimum length of time a chemical solution may be used before loosing its effectiveness. These problems and many others are important to silicon device technology and with the aid of Auger electron emission analysis considerable improvement in the processing technology may be expected.

The significant features of Auger analysis may be summarized as follows:

1. It detects those electrons originating on or near the surface. The mean free path, for example, of the Auger electron of silver at 72 ev is only 4Å (8).

2. It can detect fractions of a monolayer. As little as 1-2% of a monolayer of potassium has been detected (9).

3. It is very sensitive to elements of low atomic number but elements of high atomic number (Z=90) have been detected.

In conclusion, the author wishes to acknowledge the guidance and encouragement of Dr. L. A. Harris of the General Electric Company's Research and Development Center.

References

(1) Auger, P., J. Phys. Rad. 6, 205, (1925). Comp. Rend. 180, 65, (1925).

(2) Harris, L. A., J. Appl. Phys. 39, 1419, (1968). J. Appl. Phys. 39, 1428, (1968).

(3) Taylor, N. J., J. of Vacuum Sci. and Tech., 6, No.1, (1968).

(4) Palmberg, P. W. and Rhodin, T. N., J. Applied Physics, 39, 2425-32, (1968).

(5) Lander, J. J., Phys. Rev., 91, 1382-87, (1953).

(6) Weber, R. E., and Peria, T. W., J. Appl. Phys., 38, 4355-58, (1967).

(7) Lieberman, R. and Klein, D. L., J. of E.C.S., 113, 936, (1966).

(8) Weber, R. E. and Johnson, A. L., J. of Appl. Phys., 40, 314, (1969).

(9) Palmberg, P. W., Appl. Phys. Letters, 13, 183, (1968).

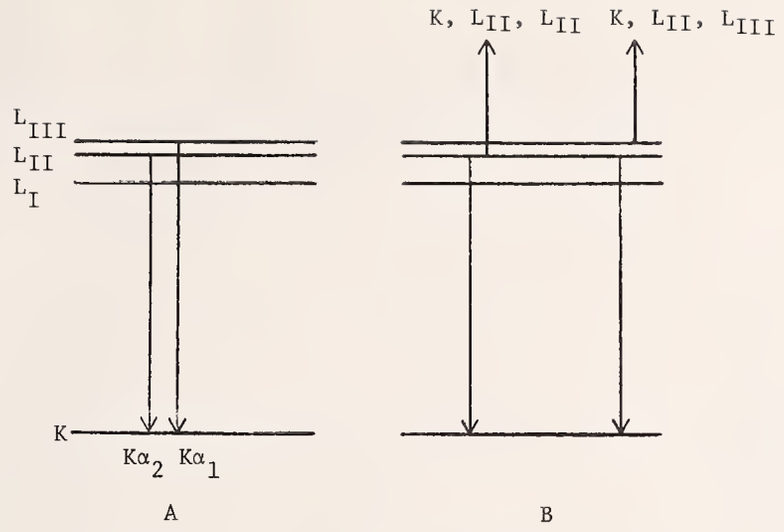


Figure 1 X-Ray and Auger Transitions

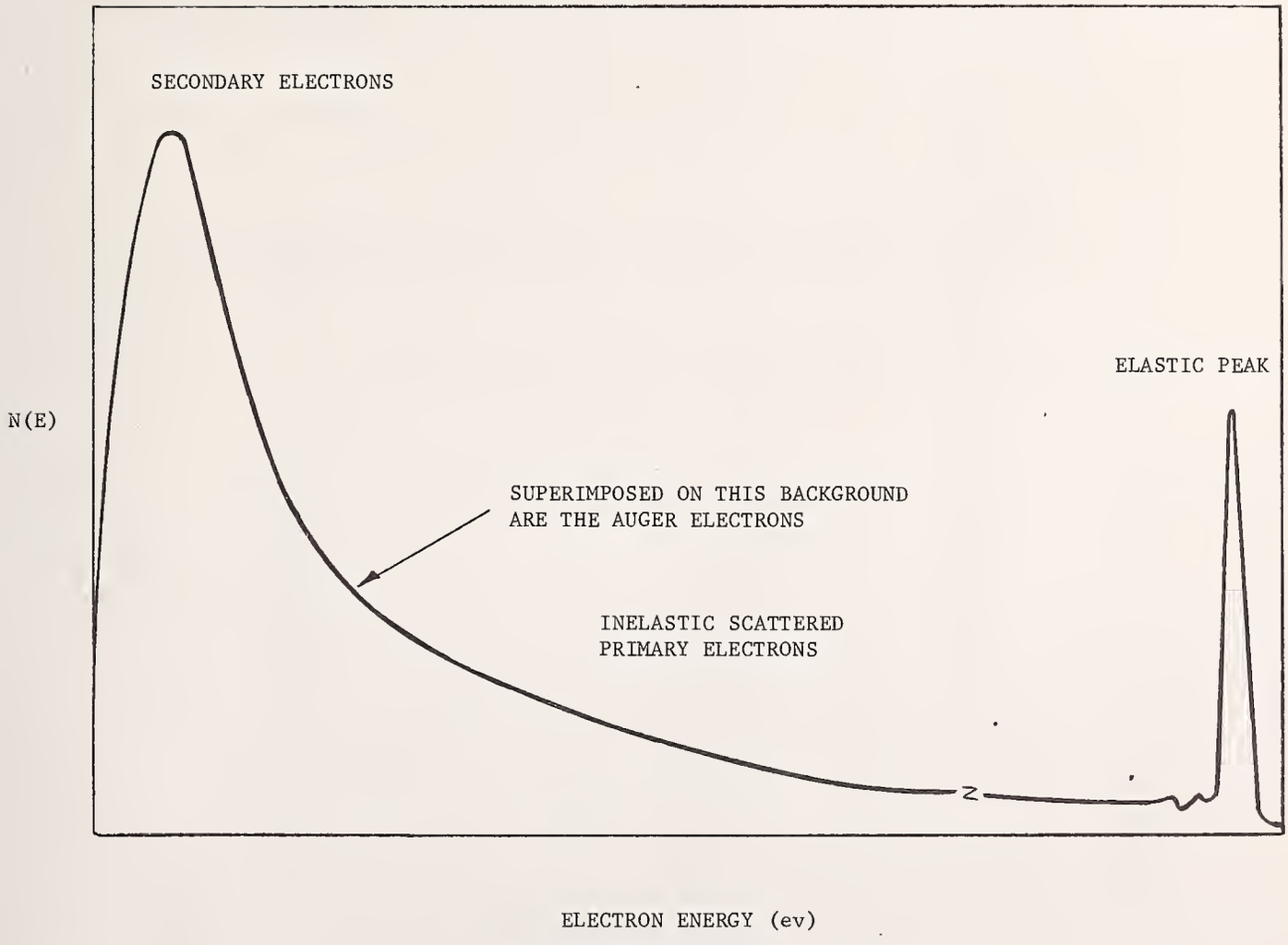


Figure 2 Electron Energy Distribution Curve

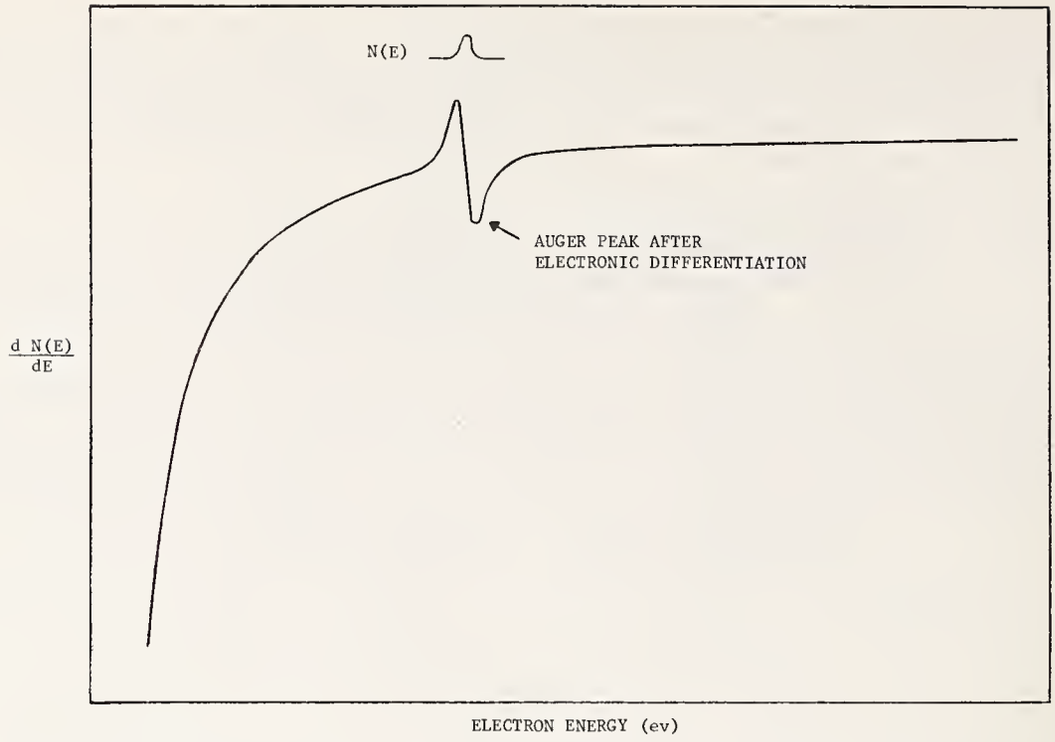


Figure 3 Illustration of the Derivative of the Electron Energy Distribution Curve

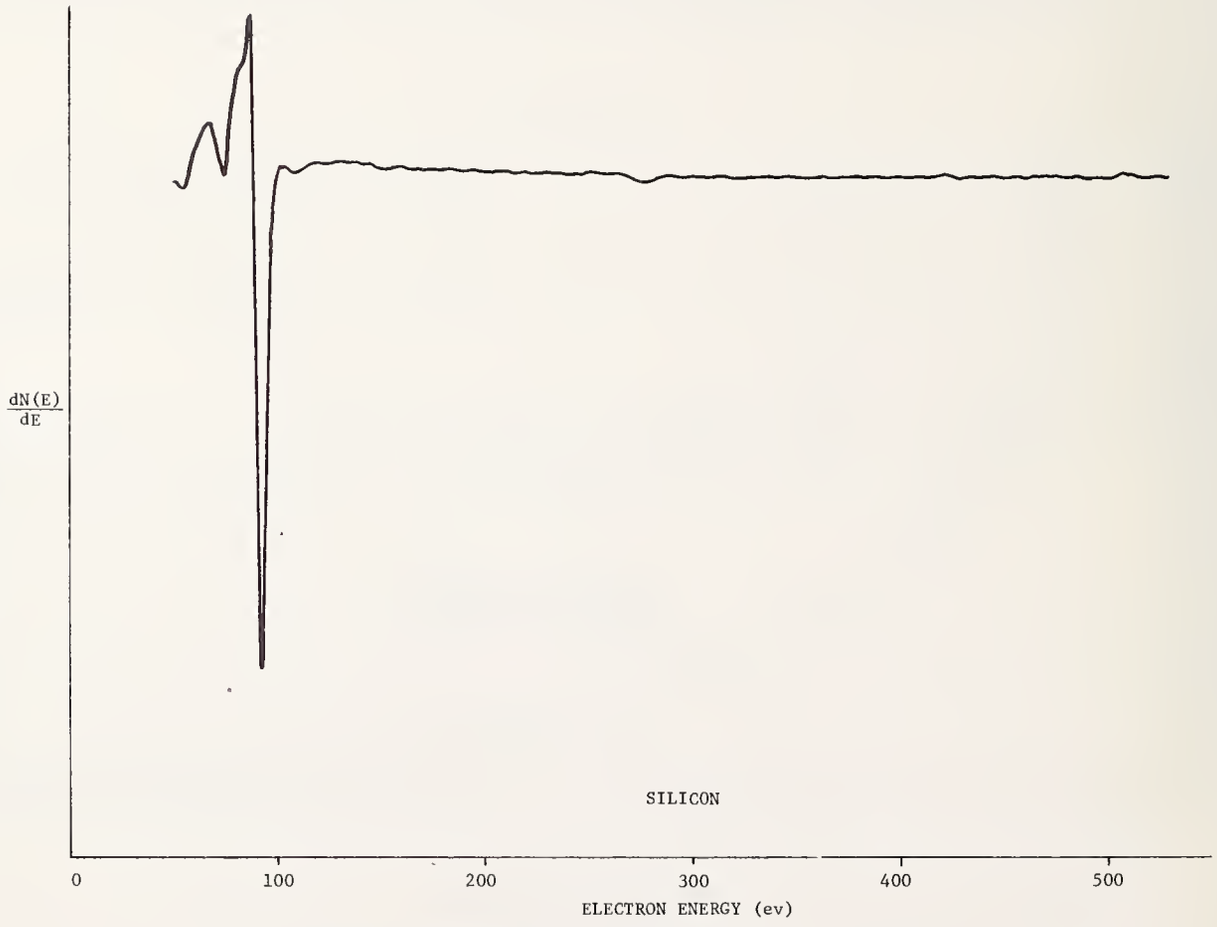


Figure 4 Auger Spectra of Clean Silicon

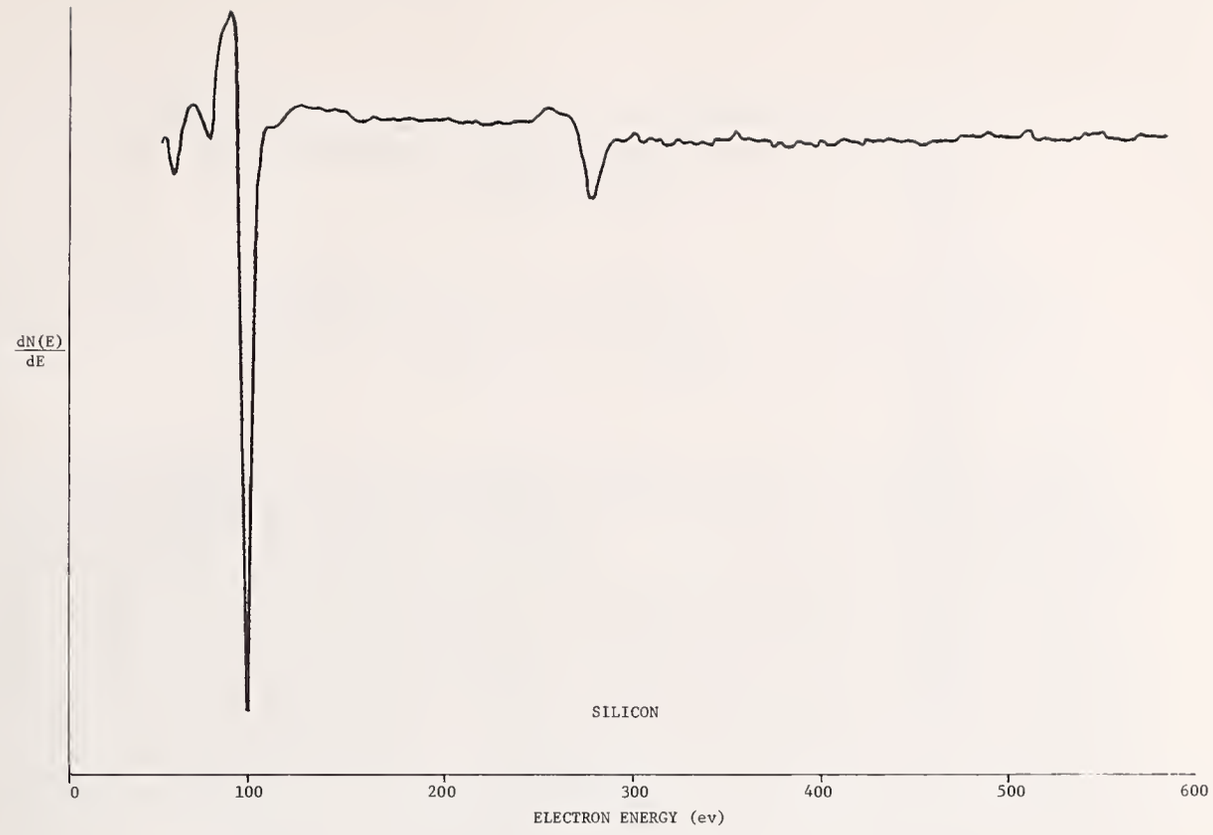


Figure 5 Auger Spectra of Silicon with Carbon Film

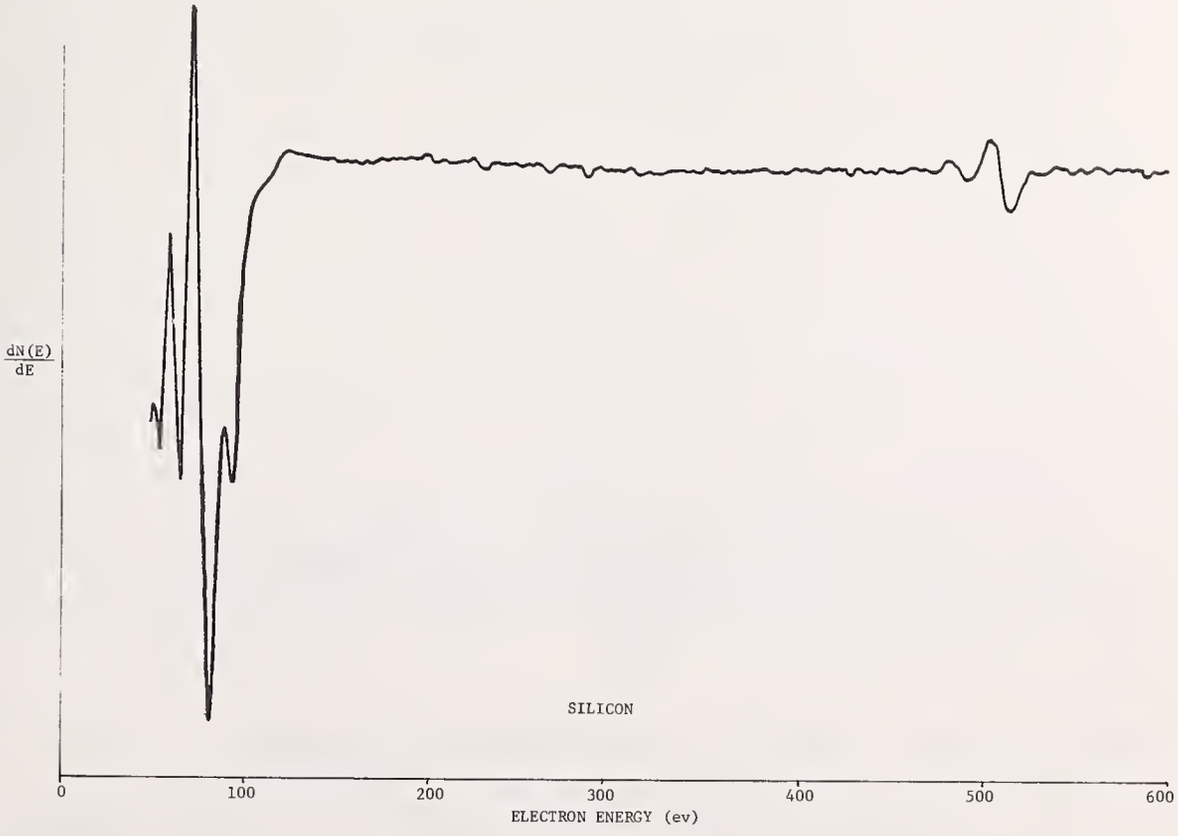


Figure 6 Auger Spectra of Silicon with Oxide Film

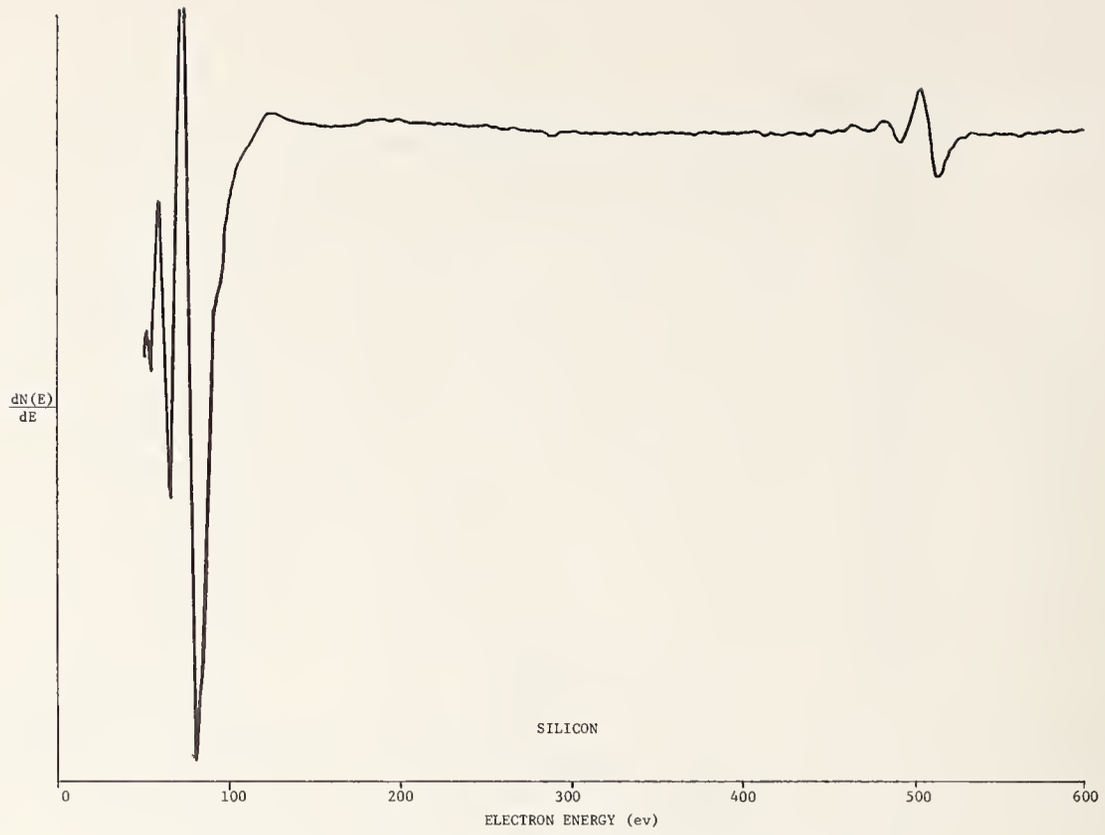


Figure 7 Auger Spectra of Silicon Dioxide Film

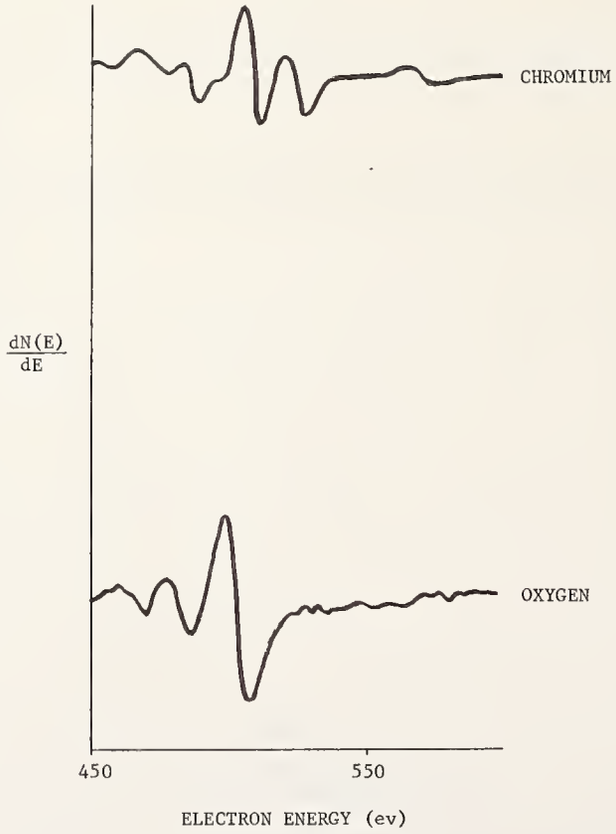


Figure 8 Auger Spectra of Chromium and Oxygen

Characterization of Semiconductor Surfaces and Interfaces by Ellipsometry

N. M. Bashara

Electrical Materials Laboratory* - Department of Electrical Engineering
University of Nebraska

Ellipsometry can be used to characterize contamination and defects of semiconductor surfaces. Two recent symposia, the first held at the National Bureau of Standards in 1963[1] and the second at the University of Nebraska in 1968[2] report typical applications and recent developments. In the latter symposium useful introductory background is provided in the papers by Muller[3] on preferred definitions and by Hall[4] on the early history. -- Increased attention is being given to investigation of various types of defects by ellipsometry. Experimentally changes in optical constants from a few percent[5,6] to a few hundreths of a percent are involved.[7-12] For example, the change in refractive index in transparent vitreous silica, induced by implantation of 150 KeV argon ions to depths of a few thousand angstroms, is easily detectable by conventional ellipsometry[5] as are the defects formed during evaporation in 100 angstrom gold films.[6] With more elaborate instrumentation perturbation methods detect changes of the order of 10^{-5} in optical constants. Although these techniques should be of value in defect studies, primary use to date has been in the investigation of band-structure.[13] Perturbation of the specimen by application of an electric field is especially useful where field-induced changes due to field penetration is of primary interest. Determinations have been made which range from penetration depths of the order of an angstrom in gold and silver[7-10] to a few thousand angstroms in germanium.[14] An ellipsometric determination has also been made of the field-induced change in the depth of the depletion layer in tin oxide.[11] ---- [1] - Ellipsometry in the Measurement of Surfaces and Thin Films, E. Passaglia, R. R. Stromberg and J. Krueger, Editors (Natl. Bur. Stds. Misc. Publ. 256, U. S. Govt. Printing Office, Washington, D. C., 1964). [2] - Surface Science, 16 (1969); also Symposium on Recent Developments in Ellipsometry, N. M. Bashara, A. B. Buckman, and A. C. Hall, Editors (North-Holland, 1969). [3] - Definitions and Conventions in Ellipsometry, R. H. Muller, reference 2, p. 14. [4] - A Century of Ellipsometry, A. C. Hall, reference 2, p. 1. [5] - Ellipsometric Analysis of Radiation Damage in Dielectrics, J. B. Schroeder and H. D. Dieselman, J. Appl. Physics 40, 2559 (1969). [6] - Study of Thin Film Annealing by Ellipsometry, J. R. Adams and K. K. Rao, reference 2, p. 382. [7] - Ellipsometry for Modulated Reflection Studies of Surfaces, A. B. Buckman and N. M. Bashara, J. Opt. Soc. Am. 58, 700 (1968). [8] - Modulated Ellipsometry for Energy Band Studies of Surfaces, A. B. Buckman, Ph.D. Thesis, Department of Electrical Engineering, University of Nebraska, May 1968. [9] - Electroreflectance Changes in Dielectric Constants of Au and Ag by Modulated Ellipsometry, A. B. Buckman and N. M. Bashara, Phys, Rev. 174, 719 (1968). [10] - Modulated Ellipsometry for Band Structure Study of Solids and Films, A. B. Buckman, reference 2, p. 193. [11] - Electroreflectance Studies of Semiconducting Films by Ellipsometry, D. G. Schueler, Ph.D. Thesis, Department of Electrical Engineering, University of Nebraska, June 1969. [12] - The Detection of Strain-Induced Optical Perturbations by Ellipsometry, L. G. Holcomb, Ph.D. Thesis, Department of Electrical Engineering, University of Nebraska, June, 1970. [13] - Modulation Spectroscopy, M. Cardona, Solid State Physics, Supplement 11. (Academic Press, 1969). [14] - A. B. Buckman, personal communication.

Key words: Ellipsometry, optical constants semiconductors, optical determination defects.

*Supported inpart by the U. S. Office of Naval Research and the National Science Foundation.

Latest developments in the subject area of this publication, as well as in other areas where the National Bureau of Standards is active, are reported in the NBS Technical News Bulletin. See following page.

HOW TO KEEP ABREAST OF NBS ACTIVITIES

Your purchase of this publication indicates an interest in the research, development, technology, or service activities of the National Bureau of Standards.

The best source of current awareness in your specific area, as well as in other NBS programs of possible interest, is the TECHNICAL NEWS BULLETIN, a monthly magazine designed for engineers, chemists, physicists, research and product development managers, librarians, and company executives.

If you do not now receive the TECHNICAL NEWS BULLETIN and would like to subscribe, and/or to review some recent issues, please fill out and return the form below.

(cut here)

Mail to: Office of Technical Information and Publications
National Bureau of Standards
Washington, D. C. 20234

Name _____

Affiliation _____

Address _____

City _____ State _____ Zip _____

☐ Please send complimentary past issues of the Technical News Bulletin.

☐ Please enter my 1–yr subscription. Enclosed is my check or money order for $3.00 (additional $1.00 for foreign mailing).
Check is made payable to: SUPERINTENDENT OF DOCUMENTS.

SP-337

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH reports National Bureau of Standards research and development in physics, mathematics, chemistry, and engineering. Comprehensive scientific papers give complete details of the work, including laboratory data, experimental procedures, and theoretical and mathematical analyses. Illustrated with photographs, drawings, and charts.

Published in three sections, available separately:

• Physics and Chemistry

Papers of interest primarily to scientists working in these fields. This section covers a broad range of physical and chemical research, with major emphasis on standards of physical measurement, fundamental constants, and properties of matter. Issued six times a year. Annual subscription: Domestic, $9.50; foreign, $11.75*.

• Mathematical Sciences

Studies and compilations designed mainly for the mathematician and theoretical physicist. Topics in mathematical statistics, theory of experiment design, numerical analysis, theoretical physics and chemistry, logical design and programming of computers and computer systems. Short numerical tables. Issued quarterly. Annual subscription: Domestic, $5.00; foreign, $6.25*.

• Engineering and Instrumentation

Reporting results of interest chiefly to the engineer and the applied scientist. This section includes many of the new developments in instrumentation resulting from the Bureau's work in physical measurement, data processing, and development of test methods. It will also cover some of the work in acoustics, applied mechanics, building research, and cryogenic engineering. Issued quarterly. Annual subscription: Domestic, $5.00; foreign, $6.25*.

TECHNICAL NEWS BULLETIN

The best single source of information concerning the Bureau's research, developmental, cooperative and publication activities, this monthly publication is designed for the industry-oriented individual whose daily work involves intimate contact with science and technology—*for engineers, chemists, physicists, research managers, product-development managers, and company executives.* Annual subscription: Domestic, $3.00; foreign, $4.00*.

*** Difference in price is due to extra cost of foreign mailing.**

NONPERIODICALS

Applied Mathematics Series. Mathematical tables, manuals, and studies.

Building Science Series. Research results, test methods, and performance criteria of building materials, components, systems, and structures.

Handbooks. Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications. Proceedings of NBS conferences, bibliographies, annual reports, wall charts, pamphlets, etc.

Monographs. Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

National Standard Reference Data Series. NSRDS provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated.

Product Standards. Provide requirements for sizes, types, quality and methods for testing various industrial products. These standards are developed cooperatively with interested Government and industry groups and provide the basis for common understanding of product characteristics for both buyers and sellers. Their use is voluntary.

Technical Notes. This series consists of communications and reports (covering both other agency and NBS-sponsored work) of limited or transitory interest.

Federal Information Processing Standards Publications. This series is the official publication within the Federal Government for information on standards adopted and promulgated under the Public Law 89–306, and Bureau of the Budget Circular A–86 entitled, Standardization of Data Elements and Codes in Data Systems.

Order NBS publications from:

Superintendent of Documents
Government Printing Office
Washington, D.C. 20402

☆ U. S. GOVERNMENT PRINTING OFFICE : 1970 O - 408-476