RADC SERIES IN RELIABILITY

PHYSICS OF FAILURE IN ELECTRONICS, VOLUME 2 (March 1964)
Edited by M. E. Goldberg and J. Vaccaro

PHYSICS OF FAILURE IN ELECTRONICS, VOLUME 3 (April 1965)
Edited by M. E. Goldberg and J. Vaccaro

PHYSICS OF FAILURE IN ELECTRONICS, VOLUME 4 (June 1966)
Edited by M. E. Goldberg and J. Vaccaro

AVAILABILITY NOTICE

Qualified requesters may obtain copies from the Defense Documentation Center (TISIR), Cameron Station, Alexandria, Va., 22314. Other requesters may order copies from the Clearinghouse for Federal Scientific and Technical Information (CFSTI), Department of Commerce, Sills Bldg., 5285 Port Royal Road, Springfield, Va. 22151.

Responsibility for the contents of papers presented at the Symposium and published herein rests solely upon the authors and not upon RADC or IIT.

Physics of Failure in Electronics

Volume 4

Reliability Series

Edited by

M.E. Goldberg
IIT Research Institute

and

Joseph Vaccaro
Rome Air Development Center
USAF

RADC SERIES IN RELIABILITY

AIR FORCE, Cato Show Prtg. Co. - 7/22/66 - 12

These proceedings are approved for publication as a part of the RADC Reliability Series.

A. T. CULBERTSON
Brig. Gen. USAF
Commander

PREFACE

This volume contains the proceedings of the 1965 Symposium on the Physics of Failure in Electronics, held at the Illinois Institute of Technology in Chicago on 16-18 November 1965. This is the fourth of an annual series of symposia jointly sponsored by the Rome Air Development Center and the IIT Research Institute.

The purpose of the meeting was to exchange information on fundamental physical and chemical processes which contribute to degradation, aging, and failure of electronic parts and materials. Emphasis was placed on the application of this information to the problems of reliability control, measurement, prediction, and improvement.

Introductory remarks were made at the opening session by G. T. Jacobi, on behalf of the IIT Research Institute, followed by an opening address given by Brig. Gen. A. T. Culbertson, Commander, Rome Air Development Center, USAF. The symposium luncheon address was delivered by Dr. Thomas P. Cheatham, Jr., Deputy Director of Defense Research and Engineering.

Papers were presented at six half-day sessions as follows: Session I - Interconnections; Session II - Test, Analysis, and Correlation; Session III - Device Physics; Session IV - Surface Effects; Session V - Bulk Effects; Session VI - Minuteman II CQAP Program. Session VI was co-sponsored by the Ballistic Systems Division, USAF. Included in this volume are also a number of standby papers which were not presented at the symposium.

The symposium co-chairmen were:

Morton E. Goldberg
IIT Research Institute

Joseph M. Schramp
Rome Air Development Center, USAF

Session moderators were:

N. M. Parikh
IIT Research Institute

J. Vaccaro
Rome Air Development Center, USAF

W. C. Dunlap
NASA, ERC, Cambridge, Massachusetts

P. Handler
University of Illinois, Urbana, Illinois

A. Tamburrino
Rome Air Development Center, USAF

D. F. Barber
Rome Air Development Center, USAF

Dr. J. S. Burgess, RADC, and Mr. J. D. Meindl, USAEC, schedule session moderators, did not attend.

The editors would like to acknowledge the assistance of Mrs. Margaret L. Warner in reading and correcting proofs.

M. E. GOLDBERG

J. VACCARO

TABLE OF CONTENTS

Contents

TABLE OF CONTENTS (CONT)

TABLE OF CONTENTS (CONT)

TABLE OF CONTENTS (CONT)

TABLE OF CONTENTS (CONT)

OPENING ADDRESS

Brigadier General A. T. Culbertson, USAF

Commander, Rome Air Development Center

Griffiss Air Force Base, New York

I am happy to have again the opportunity to join personally our co-sponsor, the IIT Research Institute, in welcoming you to this Fourth Annual Symposium on the Physics of Failure in Electronics.

The field of reliability physics is well on its way to becoming an acceptable and useful part of any major reliability program. This is not because we have found a simple short-cut to reliability; we all know we have not. Rather, it is because a fundamental physical approach represents the only rational basis for a systematic understanding of part failures. Such a fundamental approach offers a valid basis for transferring experience to new device situations. Last year, in my introductory address at the Third Annual Physics of Failure Symposium here in Chicago, I made several observations on some of the problem areas deserving additional attention. In a long range program such as reliability physics, we obviously did not expect to solve them all in a year. Nevertheless, I think it worthwhile to examine the progress that has been made in some of these areas.

First, let us consider those areas where progress is still inadequate. The statisticians remain ahead of us with their well developed mathematical tools. We, on the other hand, still need a better physical understanding of how these devices work. Furthermore, translation of this information to the reliability domain has not yet occurred. Perhaps this is because those people who have a good understanding would rather build a newer device than sharpen the tools of the reliability technology.

Detailed models of device degradation rapidly tend to become complicated and ineffective as engineering tools. There is an urgent need for the simpler models which describe gross device behavior, and which will provide the "quick and dirty approximations" needed by the reliability engineers.

Information transfer from life test data is still rather
I believe this is in part due to a lack of centralized attack
problem. We at RADC have most of the elements necessary for s
centralized attack--the reliability researchers, the reliabili
neers, computer software specialists, and the systems analysts
the next few years, these elements will be supplemented by ful
operation of the RADC Reliability Central--a central clearing
the collection, analysis, and dissemination of device reliabil
results. This Central is a closed loop system of analysts and
working together to provide reliability analyses and test resu
timely manner. The Central will not only aid in reliability i
transfer to the device designers, but it will also serve as th
back loop" to the reliability researchers and engineers so tha
concentrate their efforts in those areas needing the most work
of caution at this point--the Central is still in the early st
development. Do not expect us to solve all of your problems t

Now, let us discuss those areas where progress has been g
haps most encouraging has been a better integration of reliabi
composed of scientists, engineers, and parts specialists. Phy
and chemists seem to have become increasingly aware of the pra
side of the problem, and the reliability engineer now better a
the capabilities, as well as the limitations, of theoretical a
mental methods of failure mechanism analysis.

Our experience in the past few years has amply demonstrat
value of such team efforts in major systems programs. This va
rived in part from the impact made on the current system and,
the guidance it affords to future systems, directly and indire
through identification of necessary future research efforts.
hear the results of at least two such systems programs at this
in the next few days. The Ballistic Systems Division of USAF
sponsoring the last session, which will cover the physics of f
studies connected with the Minuteman II Component Quality Assu
gram (CQAP). You will also hear, during one of the earlier se
paper on the techniques developed under the Project Advent par
ability program. Investigations such as these, conducted with
context of a system development program and addressed to its p
problems, indicate how much can be done--not only to improve p
to characterize defects and relate them to device performance
ability. It is this type of information which affords a sound
for reliability prediction and testing.

There is one other area I mentioned last year that bears
the need for acceptable methods of evaluating the reliability
grated circuits. We all realize that the art of device fabric
variably leads systematic understanding. This is painfully ev
the rapid advances of solid state technology. We in the resea
development environment also realize the great potential that
circuits hold for our present and proposed military systems.
tems project officer, as you might expect, however, is always
to buy a "pig in a poke." There are several studies character
evaluating integrated circuit performance and reliability. Th
through RADC and the Air Force Materials Laboratory, is sponso
tractual programs in this area. We at RADC have also initiate
sive in-house project which in effect is a frontal assault on

of integrated circuit reliability. Areas of emphasis will include design of reliability tests, performance of tests on a representative circuit type, analysis of failed circuits, interpretation of test data, and application of results to the control, prediction, and assessment of the reliability of integrated circuits.

As always, problems in reliability get tougher as we progress. But, we remain convinced that the route we are taking is the most effective one. Despite our conviction, we are always willing and eager to hear how we can do a better job. So, I suggest we give the floor to those best qualified to tell us that. Again, I welcome all of you to this symposium.

SECTION I

INTERCONNECTIONS

RELIABILITY PHENOMENA IN ALUMINUM METALIZATIONS ON SILICON DIOXIDE*

W. M. Berger
R. S. Keen
G. L. Schnable

Philco Corporation
Lansdale Division
Lansdale, Pennsylvania

INTRODUCTION

This paper is a preliminary report on studies concerning the reliability limitations of aluminum metalization on the SiO_2 surface of integrated microcircuits.

The Al-SiO_2 interface is the most common metal-insulating oxide interface encountered in silicon planar devices and integrated circuits. Less widely used, but of increasing importance in equipments, are MOS devices with Al metalized gates, and SiO_2-dielectric thin-film capacitors with Al counterelectrodes.

The use of a high resolution thermal plotter has shown operating life failures to be due to localized heating resulting from IR drops at constrictions in the aluminum metalization. Storage problems result from a significant decrease in conductivity of the metalized layer. Also, reaction of aluminum with silicon dioxide, which is thermodynamically feasible, occurs to a significant extent during high temperature storage. In any application of Al metalization, even a localized Al-SiO_2 interaction is very undesirable, because it can adversely affect the electrical conductivity of the Al film, or because it can result in penetration or in decrease in thickness of the SiO_2 layer. These effects can change the electrical characteristics of the device on which they occur.

A number of phenomena have been observed on devices subjected to operating and storage stress which relate to

* The reported work was in part supported by the United States Air Force under Contract No. AF 30(602)-3610.

reliability limitations of aluminum metalizations on SiO_2 surfaces of silicon devices and integrated circuits. The empirical observations include:

1. Open metalization as a result of storage or operating life testing,
2. Ragged edges on aluminum metalization as a result of storage life testing,
3. Total disappearance of aluminum metalization from certain regions during operating and storage life testing,
4. Short circuits in capacitors with a SiO_2 dielectric and an aluminum counterelectrode as a result of storage at elevated temperatures,
5. Changes in aluminum sheet resistivity as a result of storage at high temperatures,
6. Reactions of aluminum in contact with other metals which result in formation of intermetallic compounds and/or loss of adhesion between Al and the underlying SiO_2.

The fundamental reactions which could produce the phenomena described above include the following:

1. Reaction of aluminum with SiO_2 to form silicon plus aluminum oxide,
2. Formation of aluminum-silicon eutectic phase,
3. Chemical reactions to form oxides, nitrides or other aluminum compounds,
4. Agglomeration (lateral migration) of aluminum on the SiO_2 surface of the device,
5. Diffusion of aluminum through SiO_2 layers,
6. The influence of gold bonded to aluminum on the properties of aluminum layers in the vicinity of the bond,
7. Electrolysis occurring as a result of an electric field across a dielectric and anodizing the aluminum metalization at positively biased aluminum electrodes.

The above-described reactions, while not a severe problem with relatively wide, 1-μ thick metalizations, would be expected to be a definite reliability limitation with 0.1-μ thick Al regions, particularly with narrow linewidths or, o operating life, with regions of circuits in which high currents flow in metalizations over areas which attain high temperatures.

This report discusses, in Part I, phenomena observed during operating life, and results of analysis of operating circuits using a high resolution thermal plotter as a tool to determine localized temperatures. In Part II, the effects of storage life are reviewed, with a discussion of the phenomena involved. Finally, the limitations of aluminum metalization on silicon integrated circuits and devices are summarized and conclusions are drawn relative to the mechanisms involved. Possible techniques for improving device or circuit reliability are outlined.

PART I - OPERATING LIFE RESULTS AND THERMAL PLOTTER STUDIES

Effect of Operating Life on Failure Modes

Monolithic silicon microcircuits demonstrate very low failure rates at the maximum rated storage temperature, for example, at 3V, 125°C ring circuit operating conditions. Therefore, high stress operation and high stress temperature storage testing is considered a necessary supplement to extended testing at rated conditions, and high stress testing is conducted as a normal part of the reliability evaluation of these devices. High stress testing can sufficiently accelerate some failure mechanisms that detection is possible within a relatively short period of time, compared to the duration required for the mechanism to exhibit itself at rated conditions. However, it is possible that the stress level applied to the device can activate failure mechanisms that would not normally occur at rated conditions. Any given family of monolithic silicon microcircuits consists of devices fabricated by utilizing standard family components (i.e., the same load or input resistors, the same output transistor, etc.) diffused into and interconnected on the same silicon chip according to a set of design rules unique to the family. Stress testing, therefore, in addition to whatever other information is derived from its application, will definitely establish the reliability limitations of any given microcircuit family.

The failure mechanisms indicated previously have been observed not only on stress storage life devices, but also on stress operating life devices. The time to failure for devices subjected to both operating and temperature stress is dependent upon the level of the stress.

Because increased temperature accelerates the failure mode, and because the physical appearance of the metalization is changed, it is concluded that these failure

mechanisms are chemical in nature. It is significant that certain areas of devices subjected to power stress testing are more susceptible to failure than others. Because increased temperature accelerates the failure mechanism, it is likely that the areas prone to failure during stress operation are regions of localized high operating temperature. The areas especially susceptible to failure during stress operation include: (1) metalizations crossing elements of the device which dissipate considerable power, (2) the metalization near contact cuts, and (3) the metalization near oxide steps. Because of the procedures used for depositing metalization, failures in the latter two areas may result because of constriction in the cross sectional area which causes increased resistance to current flow, and hence localized heating.

Thermal Plotter Studies

To determine if localized heating does occur in small areas on microcircuit devices, the temperature must be measured by a non-contacting detection system, since any of the contacting methods of temperature detection (such as thermocouples and temperature sensitive film) are of sufficient mass compared to the area being measured to seriously disturb thermal equilibrium. The Thermal Plotter which measures temperature by detecting the infrared radiation emitted from a given surface area is the type of instrument required[1,2,3,4].

The spatial resolution of the Thermal Plotter can be varied and is dependent upon which interchangeable Cassegrainian lens system is used. The resolution is an important consideration where a hot spot on the narrow metalization of a microcircuit is concerned because the Thermal Plotter integrates and averages the radiation emitted from the area within the circle of resolution. The effect of resolution on the Thermal Plotter output is illustrated in Figure 1.

To demonstrate localized heating, the Thermal Plotter was used to determine the operating temperatures of all operating devices reported in this paper. Consideration was given to the resolution of the instrument in all cases.

Temperatures were determined along the center line of the same 1-mil wide base metalization stripe on three integrated devices. The temperature resulted from the application of a current of 80 ma. The determinations reflect two effects: (1) that of the substrate heating due to device dissipation, and (2) localized heating of the metalization

due to constrictions. The temperature determinations, as shown in Figure 2, were made, utilizing a 0.8 mil mil resolution lens, immediately after the application of power, and again after the device reached thermal equilibrium. The difference in the temperature levels of devices 1 and 3, compared with device 11, are due to the difference in the power dissipated. Devices 1 and 3 dissipated 324 mw and 348 mw, respectively, as the result of forward biasing two transistors on the substrate, whereas device 11 dissipated only 148 mw because only one transistor on the substrate was utilized.

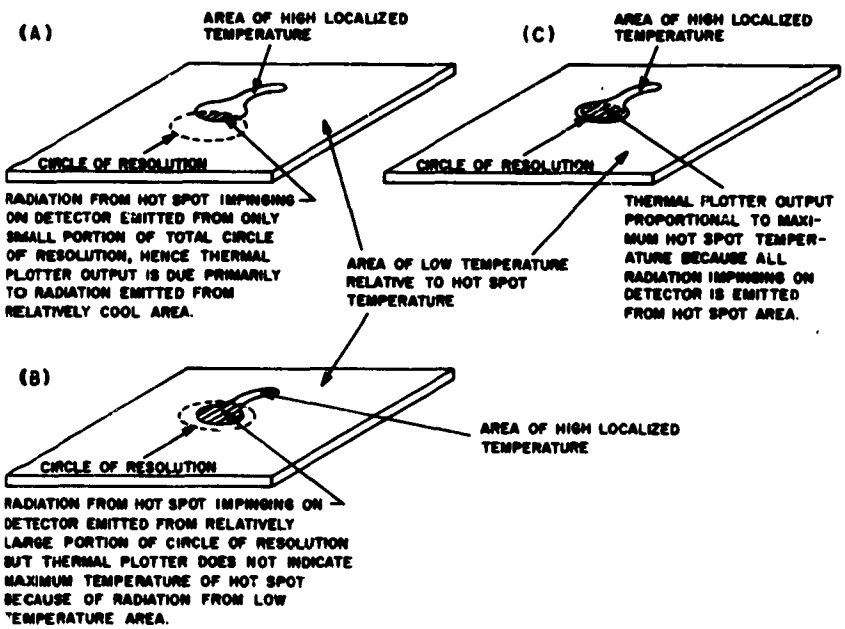


Figure 1

Illustration showing relationship of hot spot size and resolution of system to Thermal Plotter output.

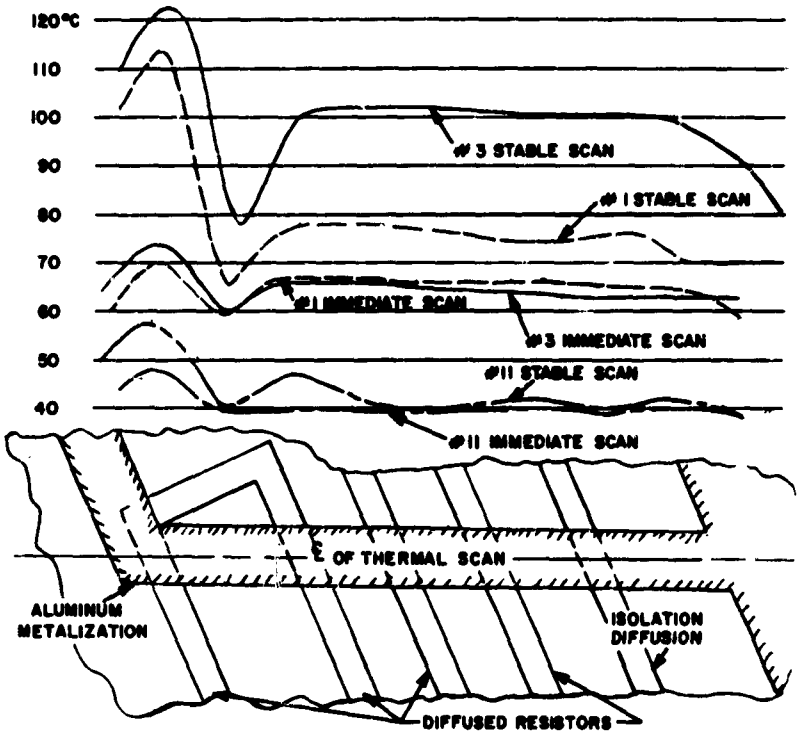


Figure 2

Temperature along center line of base metalization of operating devices.

Effects of Geometry on Localized Heating

The point of highest temperature on all metalizati regardless of the power dissipated in the device, occure an angle-point in the metalization at an oxide step caus by a resistor diffusion. Consideration of the geometry would lead one to anticipate a hot spot at this location because:

a. Current tends to crowd toward the inside of a corner of the metalization in searching for the shortest path between connections. Hence the increased current density would cause increased heating.

b. An oxide step can cause a constriction which also increases the current density.

The points of minimum temperature along the thermal scan also occur in similar locations for each device; at the bonding pad, and at points in the metalization where the distance between successive oxide steps is maximum. The locations of minimum temperatures are credible because they occur at points on the metalization that are a maximum distance from possible heat sources, such as oxide steps. The thermal scans demonstrate that localized heating does occur at oxide cuts and the effect is enhanced by sharp corners in current conductors. It is also observed that localized heating near an oxide cut at a corner in the metalization occurs more rapidly than heating due to oxide cuts alone. This is demonstrated in the scan shown for device 11 in Figure 2. The temperature determination made immediately after the application of power to the device does not show any appreciable temperature spike at oxide cuts not influenced by a corner, whereas at thermal equilibrium a spike is apparent at an oxide cut not at the corner of the metalization. Figure 3 shows the time rate of increase in temperature of an aluminum conductor after the application of power. The point selected for this determination was located on the metalization, shown in Figure 2, where it crosses the first resistor oxide step adjacent to the angle point.

The temperature determination along the centerline of an emitter metalization where one of the four contact cuts crossed by the metalization is forward biased is shown in Figure 4. Three integrated devices were used in this determination. The measurement was conducted with the 0.8 mil resolution lens after thermal equilibrium was achieved.

A temperature spike occurs at a sharp corner in the conductive metalization, and in general other temperature spikes occur near contact cuts where there is a step in the oxide surface. The maximum temperature in this case occurs at the contact cut to the isolation region, but this spike is further enhanced by the current flow around a sharp corner to the bonding pad for the external lead. Since the current passed by the metalization at this point is 160 ma, it is reasonable that here the temperature is the highest. The differences in temperature levels can again be attributed to differences in the power levels applied to the device.

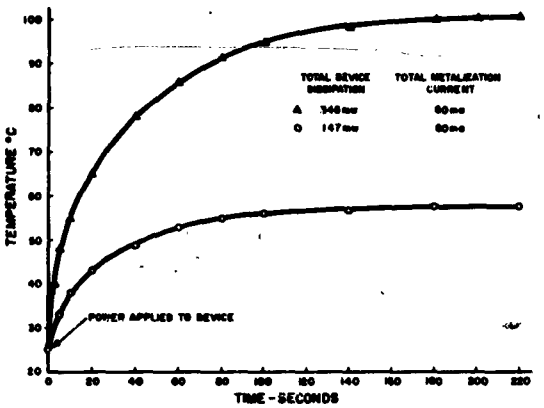


Figure 3

Time rate of metalization temperature increase after application of power.

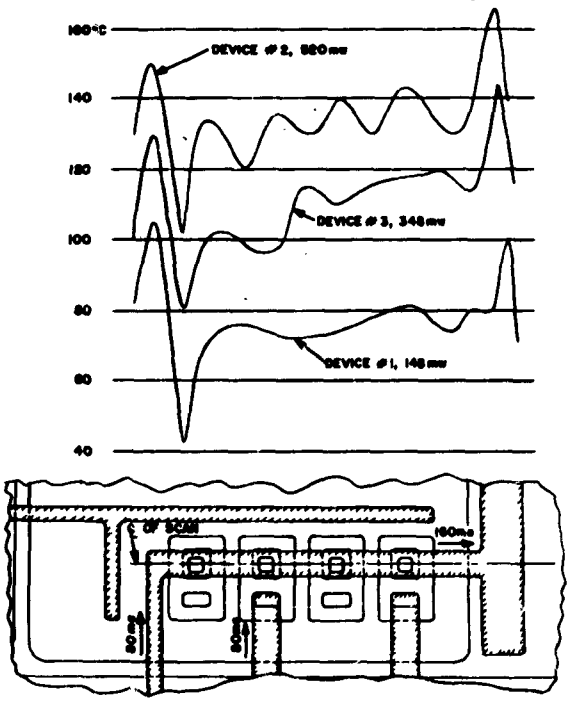


Figure 4

Temperature determination along center line of emitter metalization with scan crossing forward-biased emitter-base diode.

Disappearance of Metalization

The disappearing metalization failure mode has been observed on devices subjected to a 7000-hour, 900-mw room temperature operating test. It was significant that all the failures were almost identical. The emitter metalizations on all devices which failed because of this mechanism had opened at or near the outside edge of the emitter contact cut. This particular contact cut is a current node in the physical device where the current conducted by the metalization increases from 60 to 90 ma. The operating life circuit used during this testing is shown in Figure 5. Figure 5 also shows the equivalent point of failure in the physical device.

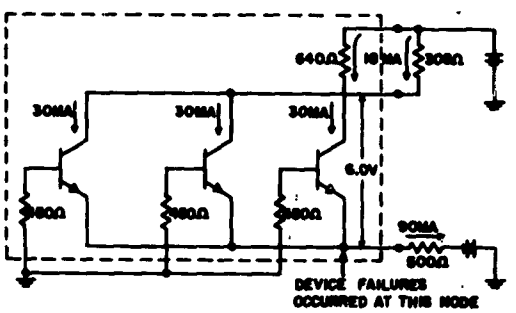


Figure 5

Stress circuit, 900 mw d.c.

Figure 6 is a photomicrograph of one of the failed devices showing the point of failure. This photograph is typical of all failures.

The physical appearance of these failures indicated heating in excess of temperatures that would be anticipated based on the thermal resistance of the device. Since the failure occurred at the edge of a contact cut where a constriction in the metalization could easily occur, and because the point of failure was at a current node, localized heating was hypothesized as the cause of failure. To verify this hypothesis, the two devices surviving after 7000 hours of 900 mw stress operation were opened, the stress conditions were imposed on the devices, and the Thermal Plotter was used to determine the temperature along the

centerline of the emitter metalization. The resulting thermal scan indicated that localized heating was occurri in the region where the metalization had opened on the failed devices. The temperatures determined along the ce terline of the emitter metalization are shown in the uppe part of Figure 7. These temperatures were determined usi the Thermal Plotter fitted with a 0.8 mil resolution lens

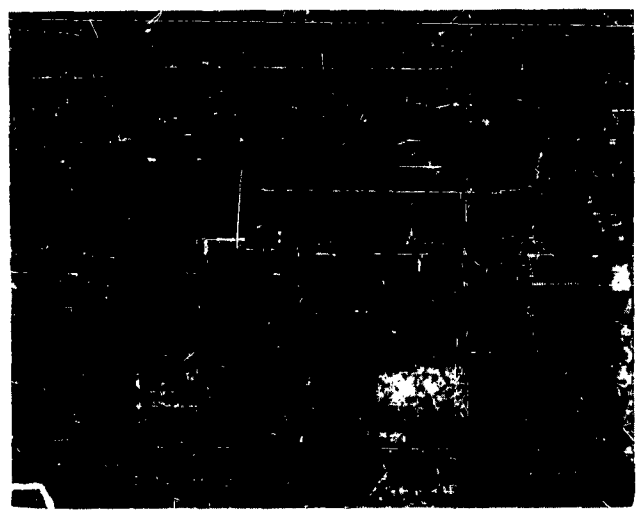


Figure 6

D: sappearing emitter metalization after '000 hours of test at 900 mw.

To establish whether all devices of this type exhibited a localized hot spot in the same region as tha responsible for the failure of the devices subjected to 900 mw stress test, a determination was made of the temp tures along the emitter metalization of a device from t same production lot. This device had been subjected to hours of 150°C storage, but not to any operating test. a small temperature spike appeared at the metalized area question, and was apparent only after the spatial resolu of the Thermal Plotter was increased to 0.3 mil. The te atures determined along the centerline of the emitter metalization for this device are shown in the lower part Figure 7.

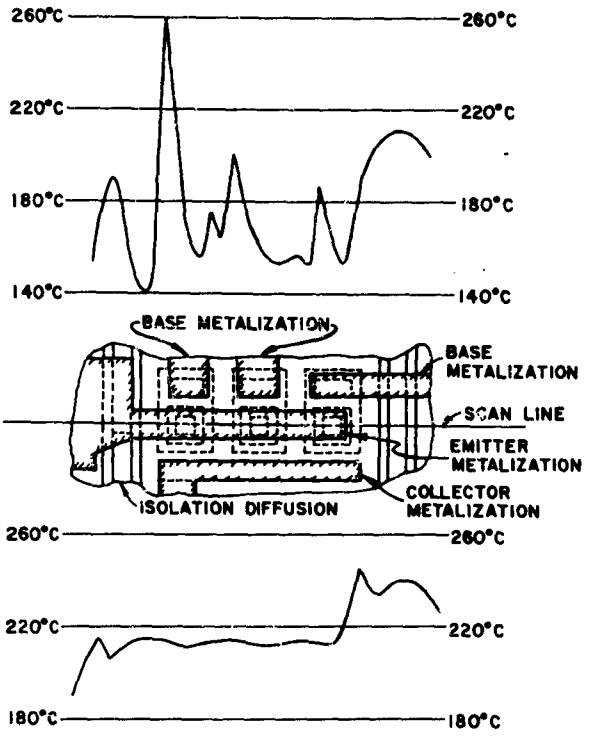


Figure 7

Temperature determination along center line of emitter metalization, with 0.8 mil resolution lens (upper) and with 0.3 mil resolution lens (lower).

These results indicate that the devices as initially subjected to operating life stress testing exhibited a small area of slight localized heating which increased in both size and temperature until the subsequent failure of the device due to chemical effects caused by excessive localized temperature.

A second example of a metalization failure is shown in Figure 8. This device failed after 1000 hours of 900 mw operating life at 125°C. It can be seen from the photomicrograph that the absence of metalization has an appearance similar to metalization failure shown in Figure 6. The d-c stress circuit in which this device was operated was arranged to dissipate approximately 750 mw in the load resistor immediately under the metalization that subsequently failed. A total of 900 mw was dissipated in the entire device, and the test was conducted in an ambient temperature

of 125°C. The cause of failure in this case is attributed to excessive localized heating of the metalization due to high localized dissipation in the 125Ω resistor immediately under the region of failure.

Figure 8

Metalization failure due to localized heating.

Considerable study has been made of ion migration effects in MOS capacitors and transistors at 300°C with a high electric field applied across the SiO_2 layer. There appears to be no tendency for migration of aluminum ions in amorphous SiO_2. It should also be pointed out that the aluminum oxide-silicon oxide equilibrium diagram, while indicating the existence of a compound mullite ($2Al_2O_3 \cdot SiO_2$), shows no phase with a melting point below 1700°C.[5] Evidence of electrolytic phenomena as a result of ion migration on dielectric surfaces such as thermally-grown SiO_2 or other insulating substrates used in microelectronics was recently reported by several investigators[6,7,8].

Al-Si Eutectic

Another metalization failure mode is a growth which begins in contact cuts and progresses outward along the metalization. This mode is also attributed to localized heating. Under microscopic examination utilizing vertical illumination, the growth appears as a black granular substance and is accompanied by a substantial increase in volume over the original aluminum metalization. The growth appears black under vertical illumination, but when viewed with dark field or side illumination, the growth is silver-white. Failures of this type normally occur during stress

operating life at dissipations far in excess of the device rating, and failure is due to short circuiting of the device. Figure 9 shows photomicrographs of this mode taken under vertical illumination at early and late stages of development. Figure 10 is a photomicrograph of the growth mode as it appears when using side illumination.

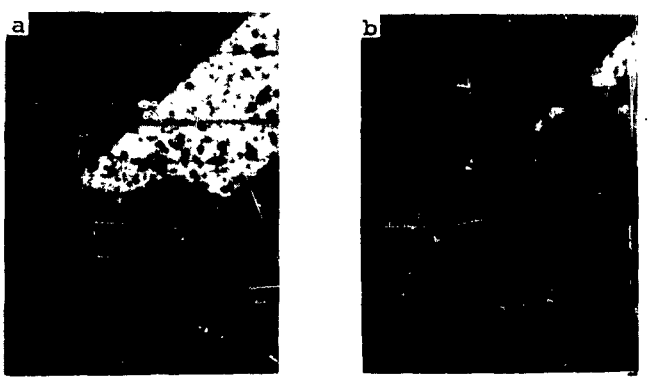


Figure 9

Metalization growth at (a) early stage of development, and (b) late stage. Both photomicrographs taken under vertical illumination.

The failure mode is believed to be the reaction of aluminum with silicon to form aluminum-silicon eutectic. It has been observed to occur in contact cuts of integrated circuits subjected to high dissipations. In general, this type of reaction starts at contact cuts which are near reverse biased junctions or other structures which produce considerable local heating.

It was possible to induce the phenomenon by forcing current through the reverse biased emitter diodes of two transistors on an integrated device. The electrical configuration of the device and the circuit used to induce the growth are shown in Figure 11. Figure 12 is a photomicrograph of the device showing the extent of the reaction at one base contact after 15 minutes of exposure to the operating conditions. Figure 13 is a photomicrograph of the entire device showing the extent of the reaction after 240 minutes of exposure to test.

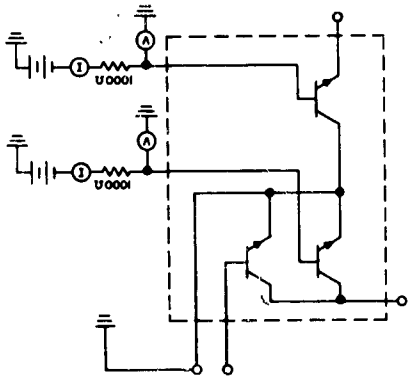


Figure 11

Electrical schematic of device and its test circuit.

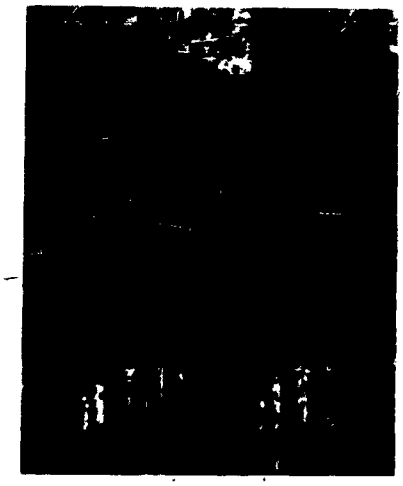


Figure 12

Extent of growth after 15 minutes.

Figure 13

Extent of growth after 240 minutes.

The rate at which the phenomenon occurred was quite rapid, but a considerable amount of power was required to cause it to proceed at this rate. A total of 1600 mw was dissipated in the entire device, 800 mw in each transistor. The progression of the front of the growth was accompanied by a dull red area which moved from spot to spot along the front as the phenomenon proceeded. The existence of the dull red spot indicates temperatures between 575 and 600°C at the leading edge of the growth. Since the Al-Si eutectic has a melting point of 577°C, the Al-Si eutectic is no doubt formed in this reaction.

The aluminum metalization not affected with the growth was removed when the device was immersed in an alkaline solution. However, the metalization in the area in which the growth had occurred was not disturbed. Subsequent etching of the device with a silicon etch did remove the growth. This indicates that the metalization had been converted from pure aluminum to some other metal. Figure 14 is a photomicrograph of the device after the metal was removed.

During the time the growth was occurring, a temperature scan was made with the Thermal Plotter and the scan

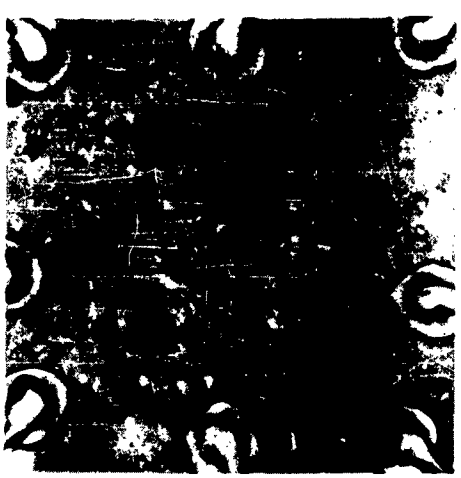


Figure 14.

Photomicrograph of device after metal removal.

indicated a hot spot at the leading edge of the growth. Temperature determinations were made along the centerline of the same metalization at lower power levels before the growth had started to occur. Figure 15 shows the temperature scan made with the device dissipating 1000 mw, approximately 500 mw in each transistor. A temperature of about 220°C is indicated at the oxide cut for the base metalization, the point from which the growth subsequently emanated. It is concluded that this phenomenon is the result of the formation of the Al-Si eutectic, and is induced by regions of high localized temperature near silicon sources.

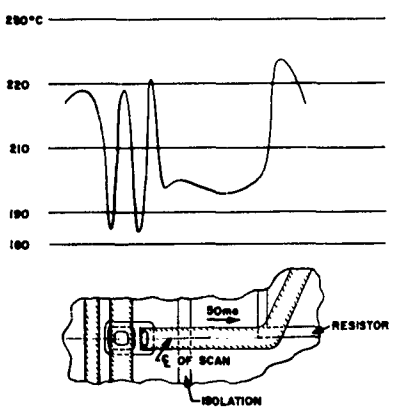


Figure 15

Temperature along centerline of metalization prior to formation of growth.

Figure 16 shows a photomicrograph of a device that failed after approximately 50 hours of 500 mw d-c stress at 125°C. The metalization growth is readily visible in the photomicrograph and is typical of the growth seen on the other devices.

To determine if localized heating was responsible for the growth, a device that had been subjected to this same test was thermally scanned across the growth-susceptible area while operating at room temperature under the voltage stress conditions. The temperature determination scans were made with the 1.4 mil resolution lens system. The results are shown in Figure 17.

It is observed from the scans in Figure 17 that temperatures in the range of 180°C to 200°C were detected

when the device was operated at room temperature. This would indicate that temperatures as high as 325°C could initially occur during operation in a 125°C ambient. An investigation of the operating circuit by probing has indicated that preferential electrical opening of certain internal nodes can cause a current as high as 120 ma to flow through the V_{CC} connection to the device. It has also been determined that leakage paths not indicated in the device schematic can be thermally activated, causing current flow as high as 100 ma in any of the blackened areas shown in Figure 16. Tests conducted at 25°C have shown that a current flow of 200 ma through the V_{CC} connection with no other stresses applied is sufficient to cause a black growth to occur in less than one minute. A current flow of 180 ma through all 8 diodes (in parallel as in the test circuit) is sufficient to induce the black growth at room temperature. It is concluded that the failure mode is one or more of the high temperature-activated mechanisms previously discussed, i.e., disappearing metalization which causes internal opens and hence increases the current flow through certain regions, which then raises the device temperature sufficiently to cause the formation of an Al-Si eutectic.

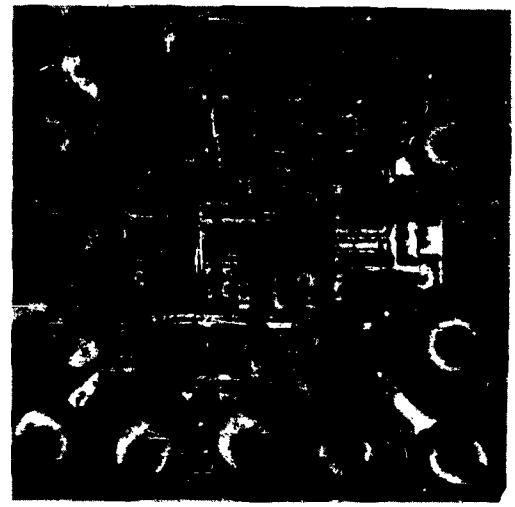


Figure 16

Metalization growth after 50 hours at 500 mw in 125°C ambient.

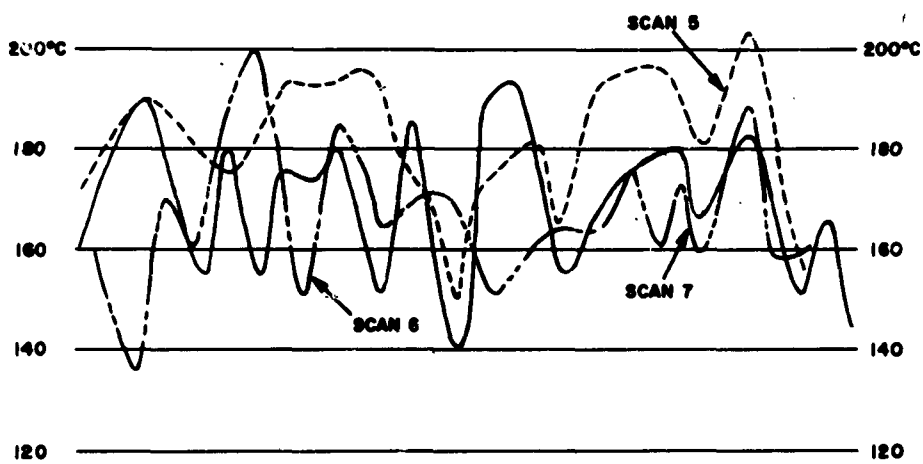


Figure 17

Temperature scans across device operating under stress voltage conditions.

PART II - STUDIES OF ALUMINUM REACTION AND MIGRATION DURING STORAGE LIFE

Al-SiO_2 Interactions

Studies of the Al-SiO_2 system have been concentrated on determining the extent to which the thermodynamically possible reaction

$$4Al + 3SiO_2 \longrightarrow 2Al_2O_3 + 3Si,$$

would occur at temperatures within the operating range of aluminum metalization measurements[9]. This reaction would be an important consideration in device reliability: (1) if silicon doped aluminum is a significantly poorer electrical conductor than pure aluminum, (2) if the reaction results in shorts through an SiO_2 layer in microcircuits or in SiO_2-dielectric capacitors, or (3) if the reaction results in an

oxygen deficiency in the silicon, which in turn induces an excess electron concentration in the underlying silicon.

It is well established that the Al-SiO_2 interaction occurs at temperatures in the range of 500 to 550°C[10]. Evidence of the interaction of the aluminum with SiO_2 has been obtained from photomicrographs of metalized patterns and from photomicrographs of angle-lapped structures consisting of aluminum on SiO_2. Figure 18 is a 1° angle lap of a 1μ aluminum pattern on a 1μ thick SiO_2 wafer after being subjected to 15 hours of storage at 550°C in a dry nitrogen atmosphere. It is evident that the reaction proceeded entirely through the SiO_2 layer, forming an ohmic path between the aluminum and the underlying silicon.

Figure 18

Angle lap showing Al-SiO_2 interaction.

This reaction has also resulted in formation of shorts in capacitors consisting of anodized n^+-type silicon with aluminum counterelectrodes. Shorts occurred in all capacitor structures tested in 20 minutes at 550°C with SiO_2 layers 0.03μ thick. When the same capacitor structure was temperature aged at 475°C for 20 minutes, very few shorts were observed.

Appreciable interaction of Al on SiO_2 at approximately 500°C has been experimentally shown to occur by Lesk[11] and Selikson[12]. More recently, there appears to be evidence of appreciable interaction between Al and SiO_2 as a result of lengthy, lower temperature storage[13].

A sensitive and effective technique for studying loss of conductivity of aluminum metalizations is to measure the sheet resistivity of thin aluminum layers which have been evaporated onto thermally oxidized silicon wafers. Table 1 indicates the change in the sheet resistivity of a 0.1μ thick aluminum layer on SiO_2 resulting from storage for various periods of time at temperatures ranging from 500 to 560°C. Empirical data indicates that the resistivity of 0.1μ Al films on SiO_2 surfaces increases by several orders of magnitude when stored for short durations in dry nitrogen at temperatures in excess of 500°C. The activation energy determined from this data is on the order of 20 kcal/mole.

Table 1

Change in Sheet Resistivity of Al on SiO_2
Measured with Four In-Line Probes

Conditions
Metal System: Al on SiO_2
Layer Thickness: 0.1μ Al
Contact Material: Al
Resistor Material: Al
Ambient: Dry N_2
Initial Resistance: 0.85 ohms/□
Measured with: 4 in-line probes

Time in Hours	Total Resistance Change (Ohms/□) for Each Time and Temperature		
	500°C	530°C	560°C
1.5	0.7	0.85	2.65
4.5	0.9	1.6	4.6
10.0	1.4	8.0	-
20.0	2.8	-	-
30.0	14.5	-	-

Whereas the reaction of Al with SiO_2 has been shown to occur, a reaction which produces Si only in solid solution in Al could not proceed, except on a localized basis, far enough to produce the observed order-of-magnitude increases in electrical resistivity of the Al film.

The solid solubility of Al in Si is 1.3 wt % at 550°C, and 0.8 wt % at 500°C. The specific resistivity of Al containing 1% silicon, however, would be less than 50% higher than that of pure Al, and thus Al in solid solution in Si would not, in itself, account for the observed resistivity increases.

With a knowledge of the Al-SiO_2 interactions occurring, techniques to confirm the reaction were introduced. The techniques were appropriately applied to both thin-film test structures and long-term storage devices.

Specimens were prepared for electron diffraction studies by evaporating a 400 Å layer of aluminum on silicon wafers which had a thermal oxide thickness of 1μ. The specimens were then baked for 20 minutes at 550°C in an atmosphere of dry nitrogen, and etched in concentrated HCl for varying lengths of time. Diffraction patterns obtained from samples etched for 30 minutes and for 16 hours exhibited a reasonable fit to the published γ-Al_2O_3 d values. These results indicate the presence of an Al_2O_3 phase in samples subjected to a 20-minute, 550°C bake in an inert atmosphere. It cannot be stated unequivocally that the Al_2O_3 is formed at the Al-SiO_2 interface because electron micrographs indicate some material from the original aluminum surface persists after etching.

A high temperature bake serves to increase the thickness of the surface oxide on the aluminum. This conclusion is drawn from the fact that the diffraction patterns from unbaked specimens show aluminum, while the patterns from baked specimens do not. Hence the aluminum has reacted with other elements, and therefore a volumetric increase is necessary. This reaction may be a consequence of an impurity in the atmosphere during baking. A high temperature bake also affects the aluminum film because electron diffraction patterns of unbaked films had a (111) fiber texture, whereas the baked ones did not. Sufficient evidence is not available to determine whether this is similarly due to atmospheric contamination during heat treatment, or whether this is due to a reaction at the Al-SiO_2 interface.

Figures 19(a) and 19(b) show a circuit in which the aluminum metalization overlying the SiO_2 has been removed by chemical means. This circuit had been stored at 250°C for 2600 hours. The photograph in Figure 19(a) was made using white light illumination, and shows no evidence of any product of an aluminum-SiO_2 interface reaction. Figure 19(b) shows the same circuit photographed under monochromatic light (Na_D = 5895 Å). In monochromatic light, the regions where the Al metalization had previously existed can be observed. It is believed that the pattern visible after removal of the aluminum is a result of interface reaction in the Al-SiO_2 system. The index of refraction of this oxidation phase has been determined to be approximately 1.51. Amorphous SiO_2 has an index of refraction of 1.456.

Thus this phase is concluded to consist of a mixed amorphous oxide of aluminum and silicon.

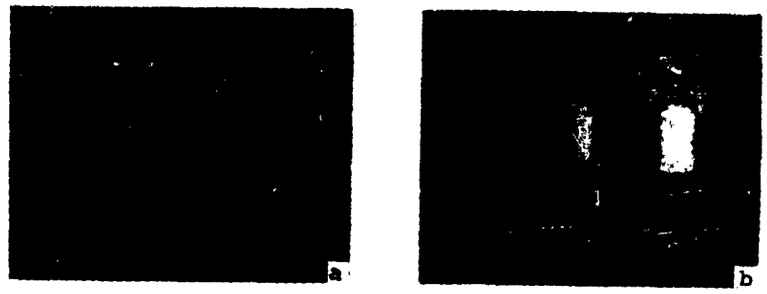


Figure 19

Device after 2600 hours at 250°C, aluminum metalization removed, photographed using (a) white light illumination and (b) monochromatic illumination.

A photomicrograph of an integrated circuit from which the metalization was partially removed by cleaning the surface with a cotton tipped swab is shown in Figure 20. The device had been subjected to 11,000 hours of storage at 200°C. A difference in oxide thickness under the removed metalization is apparent when one compares this oxide color with that of the oxide which had not been covered by aluminum.

Aluminum Migration

The disappearing metalization effect first detected on operating life has also been observed after extended storage life testing at temperatures at or in excess of the device rating.

Microscopic examination of integrated circuits stored at 150 to 350°C for times ranging from 2500 to 11,000 hours indicated disappearance of aluminum from areas of the metalization, particularly at the edge of the metalization, or at steps in the oxide. Loss of adhesion of Al to SiO_2 has also been observed. The effects at 150°C and at 250°C are shown

in Figure 21 (a) and (b). These phenomena occurred on area of Al metalization patterns which were not connected to gol bonds, as well as on areas directly connected to gold bonds.

Figure 20

Difference in oxide thickness under metalization after 11,000 hours at 200°C. Taken with white light illumination.

In the vicinity of Au bonds, a tendency for Al to disappear, forming holes in the Al pattern, was noted. Thi effect is shown in Figure 22 (a), (b), and (c). The resis ance increases given in Tables 2 and 3 are believed to be due, at least in part, to this type of phenomenon.

A similar phenomenon of disappearing Al has been observed with 0.1μ films of Al on SiO_2 in the vicinity of A bonds. This effect is a result of the storage of devices a elevated temperatures, and is not present immediately after the thermocompression bonding operation.

A test vehicle consisting of a resistor-type pattern 1-mil-wide stripes, 250 squares long, with contact pads at each end were delineated on thermally oxidized silicon. Th delineated Al resistor patterns were mounted in multiple-le TO-5 packages, with gold thermocompression bonds on the con tact pads. The structures were vacuum baked at 200°C for o hour and then sealed in dry N_2.

Recent evaluations of these 0.1μ thick aluminum patter evaporated on SiO_2 and on $P_2O_5 \cdot SiO_2$ surfaces have indicated substantial increases in electrical resistivity after being subjected to high temperature storage[13]. These data are given in Tables 2 and 3.

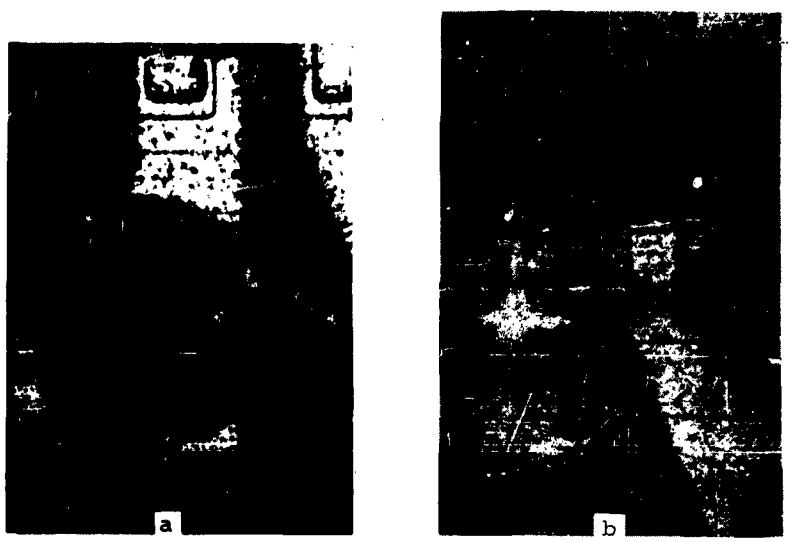


Figure 21

Photomicrographs of disappearing metalization (a) after 9500 hours at 150°C, and (b) after 4400 hours at 250°C. The smallest metalization is 0.5 mil wide.

Due to the observed increase in resistance, devices were opened for microscopic inspection. The structures revealed areas of voids in the metalization in the vicinity of the Au bonds. Voids were found in the metalization structures at temperatures of 150, 250, and 350°C after 1300 hours of aging.

Figure 22(a), (b), and (c) shows the effect of 1300 hours of storage at 140, 250, and 350°C, respectively. It is concluded that at these temperatures, lateral migration of aluminum toward the gold bond occurs, thereby causing a reduction of the cross sectional area of the metalization, and an associated increase in resistance.

It should be pointed out that no voids were observed in the areas of metalization that were connected to but not near Au bonds, or in areas where there was no gold at all.

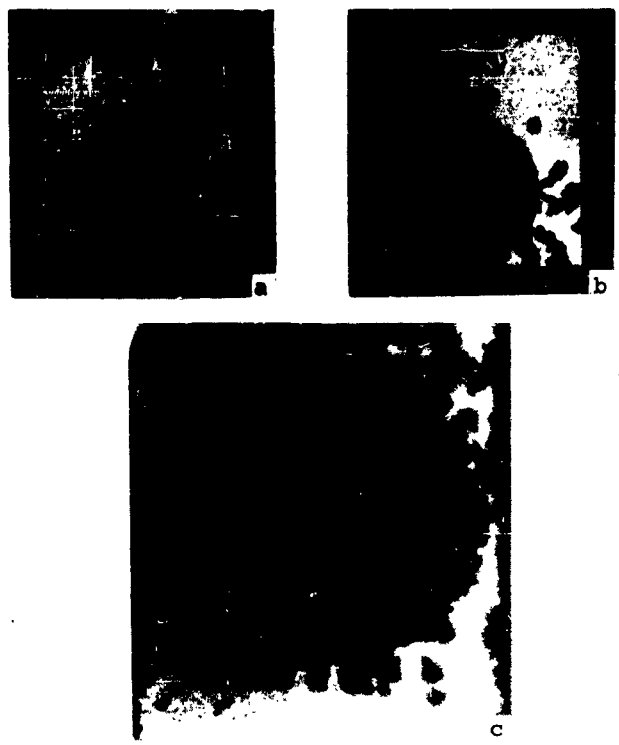


Figure 22

Metalization pattern after 1300 hours at (a) 150°C, (b) 250°C, and (c) 350°C.

Table 2

Change in Resistance of Al on SiO_2
Measured on Delineated Resistors

Conditions

Metal System: Al on SiO_2
Layer Thickness: 0.1μ Al
Contact Material: Al
Resistor Material: Al
Ambient: Dry N_2
Initial Resistance: 73 ohms
Measured With: Resistance Bridge

Time in Hours	Total Resistance Change for Each Time and Temperature					
	350°C		250°C		150°C	
	Ohms	%Δ	Ohms	%Δ	Ohms	%Δ
15	52	71	-	-	-	-
30	110	150	-	-	-	-
137	-	-	0	0	0	0
300	124	170	-	-	-	-
437	-	-	2	2.7	0	0

Table 3

Change in Resistance of Al on $P_2O_5 \cdot SiO_2$
Measured on Delineated Resistors

Conditions

Metal System: Al on $P_2O_5 \cdot SiO_2$
Layer Thickness: 0.1μ Al
Contact Material: Al
Resistor Material: Al
Ambient: Dry N_2
Initial Resistance: 131 ohms
Measured With: Resistance Bridge

Time in Hours	Total Resistance Change for Each Time and Temperature					
	350°C		250°C		150°C	
	Ohms	%Δ	Ohms	%Δ	Ohms	%Δ
30	17	13	-2	-2	-1	-0.8
94	11	8	-	-	-	-
300	32	24	1	0.8	0	0
600	35	26	11	8.4	0	0

Colteryan and Kersey[14], in a study of Au bonds to Au metalizations, showed that gold diffuses out of the bond and along the metalization stripe. They pointed out that aluminum would also seem to have diffused into the bond region, as the total thickness of the intermetallic layer was greater in bonds having external aluminum than with bonds with Al under the bond only.

SUMMARY AND CONCLUSIONS

Both the lateral diffusion of aluminum metalization and the reaction of Al-SiO_2 have been observed on devices subjected to both very long duration storage testing conducted at moderate temperatures, and short duration storage testing conducted at high stress temperatures. The end result in both cases is that the conductive metalization is reduced in cross sectional area due to the formation of voids in the metalization. The severity of these phenomena is apparently a function of time, temperature, metalization thickness, and proximity to gold nailhead bonds.

The same type of phenomenon has been observed on devices subjected to high stress operating tests. The phenomenon, as it occurs on devices subjected to operating tests, is somewhat preferential in that it normally appears at (1) metalization constrictions which have been demonstrated through temperature measurements made with the high resolution Thermal Plotter to be areas of high localized operating temperature, or (2) over areas on an operating device where the localized temperature is high because of high power dissipation in that area.

On operating devices, the phenomenon is self-accelerating in that the lateral diffusion of the metalization from an area reduces the cross section and thereby increases the electrical resistance and decreases the thermal conductivity. The increased resistance further elevates the localized temperature and increases the rate of the phenonmenon.

The Al-SiO_2 reaction, known to occur quite rapidly at temperatures in excess of 500°C, apparently also proceeds at a much slower rate at temperatures in the vicinity of 200°C. On test vehicles stored at 550°C for 15 hours, the reaction has occurred sufficiently to effect an ohmic path through a 1μ SiO_2 layer between the metalization and the substrate. The shorting has also occurred through a 0.03μ SiO_2 layer on a test vehicle after 20 minutes of storage at 550°C. There has not been any evidence, however, of this reaction

proceeding far enough to cause shorting on any devices fabricated as products by Philco, even after extended storage at 350°C.

The reaction of Al with Si to form the Si-Al eutectic has been induced by forcing very high current through back-biased junctions on silicon planar microcircuits. The reaction results in either shorting the junction or opening the metalization; the conditions at which this phenomenon has been observed are extreme and would not occur on devices operated within their ratings.

Present knowledge of reliability limitations of Al-SiO_2 metalization systems, and of degradation mechanisms, while not complete, permits consideration of possible means for improving device and circuit reliability. Several specific approaches which continue to utilize the advantages of aluminum include:

1. Use of a barrier layer between the Al and the SiO_2,
2. Use of multiple coil evaporators to insure good coverage of steps resulting from photoengraving,
3. Lateral migration of Al may be reduced by a suitable Al oxidation step,
4. Deposition of an amorphous oxide layer over the wafer surface to reduce both lateral migration and the tendency of Al to be scratched during wafer processing,
5. Elimination of sharp corners in conductive metalizations to avoid localized hot spots due to high current densities at these points.

Although not all of the reliability limitations of Al-SiO_2 have been evaluated, Philco Corporation is processing and evaluating devices and structures fabricated using a number of these suggested approaches for greater reliability.

REFERENCES

1. Schlegel, E.S., "Apparatus for Determining Temperature Profiles in Microstructures," Review of Scientific Instruments, Vol. 34, No. 4, April 1963.
2. Walker, M., J. Roschen, E. Schlegel, "An Infrared Scanning Technique for the Determination of Temperature Profiles in Microcircuits," IRE Transactions

on Electron Devices, Vol. ED-10, No. 4, pp. 263-267, July 1963.

3. Marks, B.G., G. Revesz, M. Walker, "The Thermal Plotter and Its Uses in Microcircuit Analysis and Testing," paper presented at the Third Annual Meeting of the Infrared Techniques for Electronics Committee at Huntsville, Ala., February 1964.
4. Berger, W., "Application of Isothermal Mapping as a Reliability Tool," paper presented at the Fourth Annual Meeting of the Infrared Techniques for Electronics Committee, at Los Angeles, Cal., February 1965.
5. Levin, E.M., C.R. Robbins, H.F. McMurdie, "Phase Diagrams for Ceramists," edited by M.K. Reser, compiled at the National Bureau of Standards and published by the American Ceramic Society, Columbus, Ohio, 1964.
6. Orr, W.H., and A.J. Masessa, "The Effect of Substrate Ion Migration on Thin Film Tantalum Nitride Resistors," paper presented at the Fall Meeting of the Electrochemical Society, Buffalo, N.Y., October 13, 1965; Extended Abstracts of Electrodeposition Division, Vol. 8, pp. 52-55 (1965).
7. Walker, M., G. Schnable, M. Sharp, A. McKelvey, "Accelerated Aging and Failure Mechanism Analysis of Thin Tantalum Film R-C Networks," paper prepared for presentation at the Fourth Annual Symposium on the Physics of Failure in Electronics, November 16, 1965.
8. Choi, S.S., "Electrochemical Study on Lateral Ion Migration in Insulating Substrates for Thin Film Microcircuits," paper prepared for presentation at the Fourth Annual Symposium on Physics of Failure in Electronics, Chicago, Ill., November 16, 1965.
9. Wicks, C.E., and F.E. Block, "Thermodynamic Properties of 65 Elements - Their Oxides, Halides, Carbides and Nitrides," Bureau of Mines, Bulletin No. 605, U. S. Govt. Printing Office, Washington, D.C., 1963.
10. "Study of Contact Failure in Semiconductor Devices," RADC-TR-65-205, dated August 1965, prepared by Philco Corporation, Lansdale Division, for Rome Air Development Center under Contract No. AF 30(602)-3610.
11. Lesk, I.A., "Metal-Interface Problems in Integrated Circuitry," Transactions of the Metallurgical Society of AIME, Vol. 233, pp. 578-587 (1965).
12. Selikson, B., "The Role of Compound Formation on Semiconductor Device Reliability," pp. 365-377, in Physics of Failure in Electronics, Vol. 3, edited by M.F. Goldberg and J. Vaccaro, Rome Air Development Center, USAF, (1965).
13. "Study of Contact Failure in Semiconductor Devices," Technical Documentary Report No. 2, submitted September 1965 by Philco Corporation, Lansdale Division, to Rome

Air Development Center, under Contract No. AF 30(602)-3610.

14. Colteryahn, L.E., and J.L.Kersey, "Failure Mechanisms and Kinetics of Intermetallic Formation," paper presented at Session 16-B, "A Failure Mechanism of Gold/-Aluminum Integrated Circuit Bonds," presented at Western Electronic Show and Convention, San Francisco, Cal., August 1965; WESCON/65 preprint No. 16 B.3, pp. 1-7 (1965).

THE ROLE OF METALLOGRAPHY IN THE ANALYSIS OF FAILURES OF ELECTRONIC COMPONENTS

William C. Coons

Lockheed Missiles & Space Company,
3251 Hanover Street
Palo Alto, California

INTRODUCTION

Historically, optical microscopy has played a major role in determining the causes of failures of materials of construction and of fabrication, for instance, automotive parts, aircraft components, and tools and dies. The study of failures in this area has been relatively simple in that specimens are generally of fair size; only one basic materia. is involved, e.g., steel, aluminum, copper; and the specimen preparation techniques are not particularly difficult since they are well established.

Today, however, in the electronic industries, failures occur in extremely small components. These components consist of several combined materials having widely different characteristics pertaining to mounting, grinding, polishing, and etching. The difficulties encountered in preparing specimens frequently have resulted in improperly prepared samples from which no information or misleading information has been obtained. When this happens, discontinuance of microscopic analysis usually follows. This must not happen because, in addition to the determination of the actual causes of failure, the basic microstructures of electronic component materials must be revealed and understood in order to predict other potential sources of failure.

Kehl in his standard reference book on metallographic laboratory practices has stated, "It has been only through diligent study of metals microscopically that many perplexin problems of physical metallurgy have been solved and it may be safely predicted that the contributions that will be made to the field of physical metallurgy in the future will depend in part, or solely, upon structural evidence revealed by the microscope."

Thus, a dilemma exists. It becomes apparent that within the framework of the physics of failure concept in electronics, microscopic study must be pursued. The question remains: how to do it properly so that a yield of maximum information is produced.

PRIME CONSIDERATIONS

Since the electronic engineer is beginning to rely more heavily on the microscopist, a closer relationship must exist between the two in order that their respective problems can be better resolved. To illustrate, the following situation is cited. A metallographic examination was requested on several integrated circuits. The electronic engineer, not understanding the ultimate problems of the metallographer, cemented the encased circuits to a glass plate so that the case surface could be removed by grinding. The cement used was Eastman 910. The glass plate, containing four exposed circuits, was presented to the metallographer for study. With considerable difficulty the circuits were pried loose from the glass plate with a razor blade preparatory to individual mounting in epoxy. When polished specimens revealed extensive cracking of the silicon wafers, the metallographer's fears were founded; the circuits were severely damaged during removal from the glass plate, and no definite or reliable interpretations could be made regarding the integrity of the observed joints or substrates.

While all microscopists are confident of their techniques and their ability to perform them, it becomes more apparent with each investigation that improvements are needed in this area. Because of the smallness and fragility of many of the current electronic components, pitfalls are encountered at every stage of handling and preparation. Pressureless mounting, careful fine grinding, elimination of embedded grit, maintenance of flat interfaces, and freedom from coarse scratches are mandatory before one can safely interpret what is revealed.

All of these techniques of preparation require the greatest need of the microscopist. This need is time. Whereas a sample of steel can be prepared in an hour, it may take two or three days for proper preparation of the cross section of a small integrated circuit, lead wire joint.

SUGGESTED PREPARATION TECHNIQUES

In the metallographic laboratory of the Lockheed Missiles & Space Company's Research Division, the following steps are taken to prepare samples:

Step 1. All samples are mounted in Maraset epoxy resin, a clear, moderately low-temperature setting material. Mounting is performed in two ways. If the components and associated

films are of known sufficient thickness, 90-degree cross sections through pertinent areas are exposed at or near the surface of the mount. When the components and associated films are very small and extremely thin, taper sections are prepared with angles of exposure on the order of 5 to 6 degrees, which yields a "mechanical" magnification of 10 to 12 in a direction normal to the cross-sectional trace.

Step 2. Grinding is performed by hand only on 600-grit SiC wet or dry papers. When lead wire joint cross sections are to be studied, the sample is examined under a binocular microscope after every few strokes of grinding to determine when the joint area has just been reached. At this point, polishing steps are started.

Step 3. Rough polishing is performed automatically on the back side of a slipper-satin cloth impregnated with 1-micron diamond paste. The sample is polished with a 240-gram weigh until the desired cross section is almost obtained. Frequent examinations are required.

Step 4. Final polishing is done automatically on Microcloth using Cer-Cre Metallographic Polishing Abrasive. Generally a light weight - about 140 grams - is employed and polishing should not be prolonged; otherwise, excessive relief will result. Grit, embedded in the soft phases, and fine scratch are almost impossible to eliminate entirely without causing excessive relief at the interfaces between the materials varying widely in hardness or abrasive resistance.

Step 5. In all the investigations performed in this laboratory, examinations were made using bright field and polarize light illumination on unetched samples. When the polishing was performed properly, no etching was necessary.

ILLUSTRATIVE AREAS AND RESULTS

To illustrate the role of metallography in the analysis of failures of electronic components, three examples will be presented. These examples will show the results of a nondestructive type of examination, a 90-degree cross-section examination, and a taper section study.

(1) Nondestructive Examination of a PbS Infrared Detector

An instance was encountered where PbS infrared detectors were unpredictable in regard to sensitivity and noise levels when subjected to temperature variations. One of these detectors and its pertinent regions is illustrated by Figure 1.*

By inverting one of the detectors onto the stage of a Bausch and Lomb Research Metallograph equipped with a

*The magnification shown for all figures is that preceding an 0.75 reduction by the printer.

carbon-arc light source, examination of the lead wire joints could be made. While examining the joints, movement was noted within one of the joint areas. From experience it was known that incompletely cured plastic and some thermoplastic materials could be affected by the small amount of heat passing through the optics of the microscope from the carbon-arc light source. In this case, the lead wires had been cemented into a V-notch in a quartz substrate by a conductive, metallic-filled plastic paste. This conductive paste then provided contact between the lead wire and gold conductive film. Apparently the plastic in the paste was not completely cured or it was a thermoplastic type which became plastic above some elevated temperature.

An experiment was conducted wherein a joint was chilled to 10°C. While at this temperature, the joint was photographed and its appearance is shown in Figure 2. The material appeared solid, it was severly crazed, and no evidence of movement was observed. After the joint had been allowed to become stabilized at the temperature produced by the carbon-arc light source, it was again photographed and its appearance is illustrated by Figure 3. The temperature in this instance was determined to be 41°C. The craze pattern was still evident and some minute evidence of plastic flow was noted.

When the temperature of the joint was raised to 50°C, the characteristics of the joint changed completely. The craze pattern disappeared, minute air entrapments coalesced, and the plastic became more transparent. These changes are shown in Figure 4.

Obviously any movement of the conductive media around the joint lead wire might be expected to produce detector noise. If the metal in the paste settled, by gravitation, away from the lead wire when the plastic was fluid, complete failure of the device could occur.

(2) Examination of 90-Degree Cross Sections Through Gold to Aluminum Thermocompression Bonds on Silicon Transistors

Owing to apparent random failures of TC gold to aluminum bonds, studies were undertaken to determine possible modes of failure. An understanding of the modes of failure was needed to suggest methods of improving the reliability of these parts.

In this instance, 90-degree cross sections were prepared because the size of the joints was large enough to permit this type of examination. One of the transistors is shown in Figure 5.

Others[1, 2] have reported failures of this type of joint due to fracture along interfaces of brittle intermetallic phases, through these brittle phases as a result of stresses

Figure 1

Photograph of PbS IR Detector Showing Lead Connections. 50×

Figure 2

Photomicrograph of PbS IR Detector Lead Wire Joint Area at 10°C. 250×

Figure 3

Photomicrograph of PbS IR Detector Lead Wire Joint Area at 41° C. 250×

Figure 4

Photomicrograph of PbS IR Detector Lead Wire Joint Area at 50°C. 250×

originating at notches formed around the periphery of the joints, and through void formations existing along intermetallic compound interfaces as a result of the Kirkendall effect. All of these failure mechanisms were postulated on the basis of metallographic studies.

Colteryahn et al. have, by etching techniques and microprobe analyses, identified three phases existing in the gold to aluminum bonds, namely, Au_4Al, Au_5Al_2, and Au_2Al. Without the sophisticated microprobe facility and with no knowledge of the etching techniques or findings of these investigators, this study relied principally on the examination of well-polished surfaces, color of the phases, the characteristics of the phases under polarized light, and very high magnification examinations up to 3000×. A knowledge of the equilibrium phase diagram of the Au-Al system and of the crystal structures of reported phases was, of course, essential.

Figure 6 is a photomicrograph of a joint cross section in the as-polished condition and with hardness indentations in the various regions. Two distinct phases can be seen between the soft gold button and the hard silicon wafer. When this surface is viewed under polarized light (Figure 7), it can be seen that the phase in contact with the gold was isotropic, the cubic phase Au_4Al, and the phase in contact with the silicon substrate was quite anisotropic, the hexagonal Au_5Al_2 phase. In this instance no Au_2Al was noted and only very minute amounts of the pink phase $AuAl_2$ were observed at the aluminum interface. The $AuAl_2$ which existed in this cross section was not easily seen at 600×.

Another joint, which was found to be broken, is shown in the composite photomicrograph of Figure 8 which was photographed at 3000× in the as-polished condition. A gold region is indicated by 1, a small zone of Au_4Al by 2, a large region of Au_5Al_2 by 3, and a small layer of possibly Au_2Al by 4. Under this layer (4) and to the right, massive pink $AuAl_2$ interspersed with remaining aluminum was observed.

The thickness of the aluminum film (extreme left) was determined to be 5.7 microns, and some of the pink $AuAl_2$ phase existed at the aluminum joint interface on this side.

A large void was observed under the massive Au_5Al_2 along the silicon substrate. All the observations disclosed that this joint was improperly made and would be expected to have very poor integrity.

The existence of periphery notches and voids and of interfacial porosities was noted on many of the joint cross sections that were prepared. Manifestations of these types of potential failure origin locations are illustrated by Figures 9 and 10.

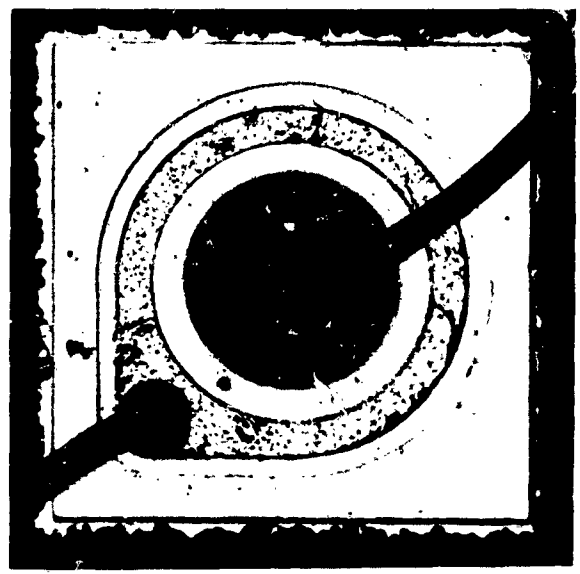


Figure 5

Photograph of Silicon Transistor Showing Gold Lead Wire to Aluminum Film Joints. 100×

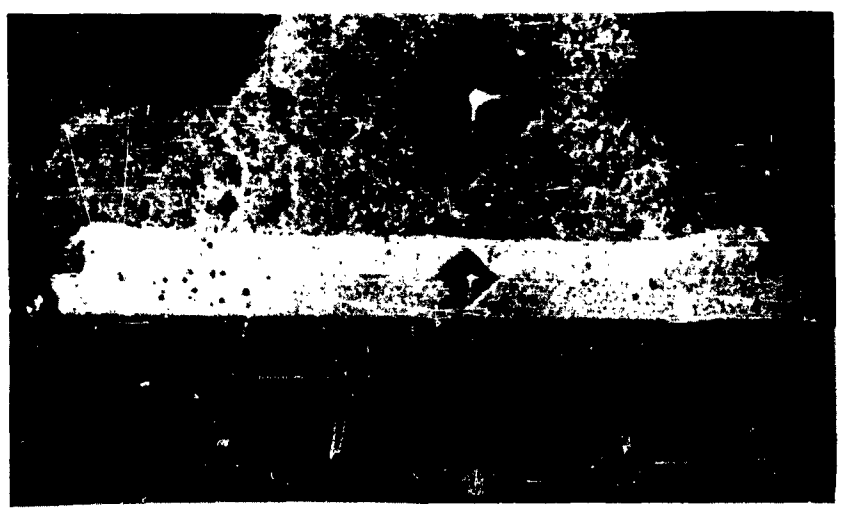


Figure 6

Cross Section Through Gold to Aluminum Joint on Silicon Substrate. Bright Field Illumination. 1000×

Figure 7

Cross Section Through Gold to Aluminum Joint on Silicon Substrate. Polarized Light. 1000x

Figure 8

Composite Photomicrograph Through a Broken Gold to Aluminum Joint on Silicon Substrate. 3000x

Figure 9

Cross Section Through Gold to Aluminum Joint on Silicon Substrate Showing Peripheral Notch. 3000×

Figure 10

Cross Section Through Gold to Aluminum Joint on Silica Substrate Showing Minute Porosities Along Au_4Al and Au_5Al_2 Interface. 3000×

As can be noted from the photomicrographs presented in this section, the existence of brittle phases, phase identifications, periphery notches, and interface voids was revealed and established on as-polished surfaces and in some instances at very high magnification.

(3) Taper Section Studies of Gold, Molybdenum, and Silica Films on Silicon Substrates and of a Gold to Aluminum Joint

Since the limit of resolution of the finest metallographic equipment is on the order of 0.25 microns, it is necessary to magnify thin film thicknesses mechanically. This is performed by providing a cross section of the film planes at such an angle to the planes that the cosecant of the angle would be sufficiently large to render the films sufficiently optically thick that they could be measured.

However, there are limitations in this approach which influence the accuracy of the measurements and must be taken into consideration. It is very difficult to mount and prepare specimens and maintain a preconceived angle, and thus it is better to determine the angle from the specimen geometry after the specimen has been prepared. The accuracy of the method of determination of the angle has tremendous influence on the accuracy of the measurements of the films. The following methods were used in the determination of thicknesses of gold, molybdenum, and silica films on silicon substrates.

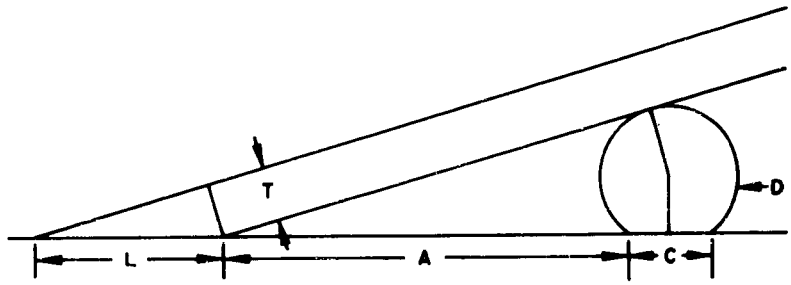
Method 1. A wafer was placed face down in the bottom of a potting mold with one end supported by a length of 29-gage copper magnet wire running transversely across the end of the wafer. The unit was then vacuum mounted in Maraset epoxy resin. After grinding and polishing, the determination of the angle was made by the following measurements which could be ascertained optically.

Let A = distance from the wafer surface intercept to the nearest intercept of the wire,
C = width of the exposed section of wire,
D = diameter of the wire,
L = width of the exposed film section,
T = true thickness of the film.

Then

$$T \simeq L \sin \left[\tan^{-1} \left(\frac{\sqrt{D^2 - C^2}}{2A + C} \right) + \sin^{-1} \left(\frac{D}{2A + C} \right) \right]$$

This method is reasonably accurate for the specimen as a whole. However, slight relief effects from polishing dissimilar materials can alter the angle at the point where the measurements are made.



Method 2. A simpler and perhaps more locally accurate method is as follows. The wedge of translucent epoxy formed by the surface of the specimen and the polished surface of the mount will exhibit an interference pattern as shown in Figures 11 and 12. In Figure 12 a P-N junction between the molybdenum and the silicon is visible. No explanation for the junction's being visible in the as-polished condition was obtained. Since a black band exists at every one-half wavelength (i.e., 0, 1/2, 1, 1-1/2, etc.), then the following relation holds if the wavelength of the incident light is known:

Let T = true thickness,
L = width of the exposed film in microns,
N = number of extinctions per distance D (microns),
λ = wavelength of incident light.

Then

$$T = L \sin \left[\tan^{-1} \left(\frac{N\lambda}{2D} \right) \right]$$

The angles and thicknesses of the films on the gold-molybdenum integrated circuit were determined to be as follows:

Method 1. The angle was 5-degrees 43 minutes, with a magnification factor of 10.039. Thickness values (in microns) were:

Figure 11

Taper Section Through SiO_2, Mo, and Au Layers Deposited on Silicon Wafer. 500×

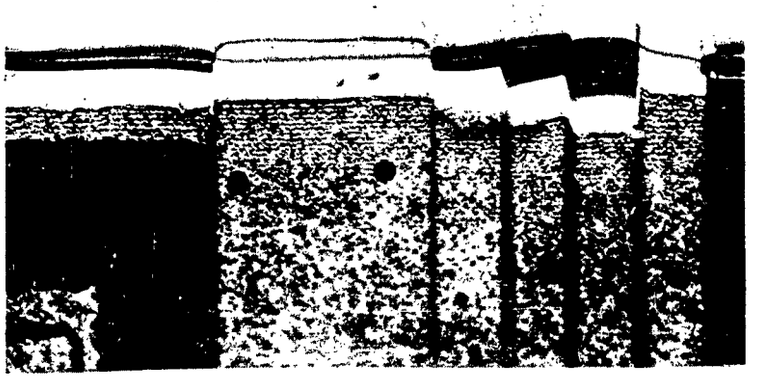


Figure 12

Taper Section Through SiO_2, Mo, and Au Layers Deposited on Silicon Wafer. Note P-N Junction Visible in As-Polished Condition. 500×

Au - 1.253
Mo - 0.148
Mo over Si - 0.128
Junction depth - 0.690

Method 2. The angle was 6 degrees 10 minutes, with a magnification factor of 9.309. Thickness values were:

Au - 1.342
Mo - 0.160
Mo over Si - 0.134
Junction depth - 0.752

It is reemphasized that metallography should play an increasing and continuing role in the study of failures and of the ultimate production of highly reliable electronic components. To be fruitful, however, greater care in the preparation of specimens and the development of new preparation techniques is of utmost importance. Metallographers need to learn the function of electronic components and to understand the materials problems of the electronic engineer; and for this learning and for technique development, he must be afforded more time.

REFERENCES

1. L. E. Colteryahn and D. D. Shaffer, "Characterization of Failure Modes in Gold-Aluminum Thermocompression Bonds," Paper 16B.2, WESCON/65, Session 16B: A Failure Mechanism of Gold/Aluminum Integrated Circuit Bonds, San Francisco, August 24-27, 1965.

2. L. E. Colteryahn and J. L. Kersey, "Failure Mechanisms and Kinetics of Intermetallic Formation," Paper 16.B.3, WESCON/65, Session 16B: A Failure Mechanism of Gold/Aluminum Integrated Circuit Bonds, San Francisco, August 24-27, 1965.

ANALYSIS OF SEVEN SEMICONDUCTOR METALLURGY SYSTEMS USED ON SILICON PLANAR TRANSISTORS

William H. Gianelle
Grumman Aircraft Engineering Corporation
Bethpage, Long Island, New York

I INTRODUCTION

Based upon an intense study of referenced materials and the experience of the author, two metallurgical systems are recommended for improved reliability of silicon planar transistors. These comprise one each for SiO_2 protected transistors with extended and nonextended contacts. A third system for glassed transistors would have been recommended; however, this system is considered proprietary.

The body of this paper is divided into 4 sections: Section II discusses wedge ball versus thermocompression bonding, Section III defines an electrical failure, Section IV discusses the failure mechanisms and/or potential problems of the seven metallurgical systems listed below:

Land Metallization	Wire	Thermo-compression Bond	Contact System	Manufacturers
Al	Au	wedge	NE*	IBM, GE
		ball	NE	Fch, TI
Au	Au	wedge	NE	Motorola
Al	Ag	wedge	NE	IBM
Ag	Au	wedge	NE	IBM
Al	Al	wedge	NE	Sylvania
CrAg	Au	ball	E**	Philco
MoAu	Au	ball	E	TI

* Nonexpanded contact system
** Expanded contact system

And Section V defines the recommended systems.

II WEDGE VERSUS BALL THERMOCOMPRESSION BONDS

A. Wedge Bonds

Wedge bonding is accomplished by using a very small chisel whose tip is appropriately shaped to apply pressure to a small wire properly located on the metal film contact which has been heated to the bonding temperature (Figure 1). This type of bond requires individual registration of the stripe, wire, and bonding tool during bonding. The end result is a bond formed by a deformed and weakened wire.

If the chip is not parallel with the header, three possibilities exist: (1) further pinching of the wire at the neck, (2) a poorer contact, or (3) a cracked chip. Any or all of these possibilities could go undetected through the manufacturers screening process.

The usefulness of wedge bonding is limited to a small geometry devices.

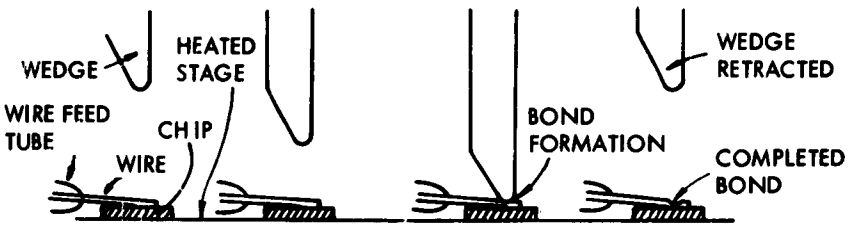


Figure 1

Wedge Bonding Technique

B. Ball Bonds

Bonding of this type is effected by threading the wire to be bonded through a quartz or tungsten carbide thick-walled capillary; the capillary tapers to a small annulus of a few mils diameter on the outside. The tube is mounted in a suitable mechanical fixture so that it can be moved vertically to apply pressure to the ball formed on the end of the wire by passing a flame over it and melting it. The wire fed from a roll is free to move in the capillary. Bonding is effected by lowering the capillary annulus onto the ball and applying pressure while at the same time applying heat (Figure 2). The capillary is then withdrawn an appropriate distance and the flame is again used to sever the wire. The

header is withdrawn from the bonder and the bonded leads, which are free at one end, are welded with a microwelder to the gold plated kovar stems on the header.

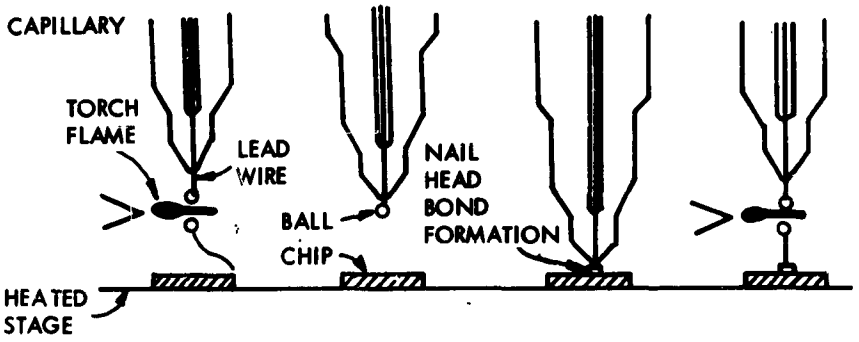


Figure 2
Ball Bonding Technique

With the advent of the metal-over-oxide technique, ball bonding to high-frequency small-stripe transistors no longer presents a problem. This takes the edge off the wedge bonds one-time advantage--that is, bonding directly to small geometries.

Ball bonding[1] is the most popular method of lead attachment in silicon devices. This type of attachment predominates in the industry mainly because of its high speed, economy, versatility, strength, and reliability[2,3]. Gold wire is almost ideally suited for this technique[4].

III ELECTRICAL FAILURE CRITERIA

Intermetallic problems can be detected by monitoring the forward voltage drops of each junction during life

1. O. L. Anderson, H. Christensen and P. Andreatch, Journal of Applied Physics, vol. 28, p. 923 (1957).

2. I. A. Lesk, Semiconductor Products, I, p. 32 (1964).

3. E. Keonjian, Microelectronics, p. 298 (McGraw-Hill, New York 1963).

4. A. Cohn, Semiconductor Products, VII, p. 18 (1964).

testing. A 50 mv increase at 100 ma for both the emitter-base and/or collector-base diode forward voltage drop was considered a failure. Any device which exhibited this increase was submitted to a failure analysis group. Figure 3 shows how the forward voltage drop increases with an intermetallic problem. As can be seen, the problem is more easily detected at higher current levels.

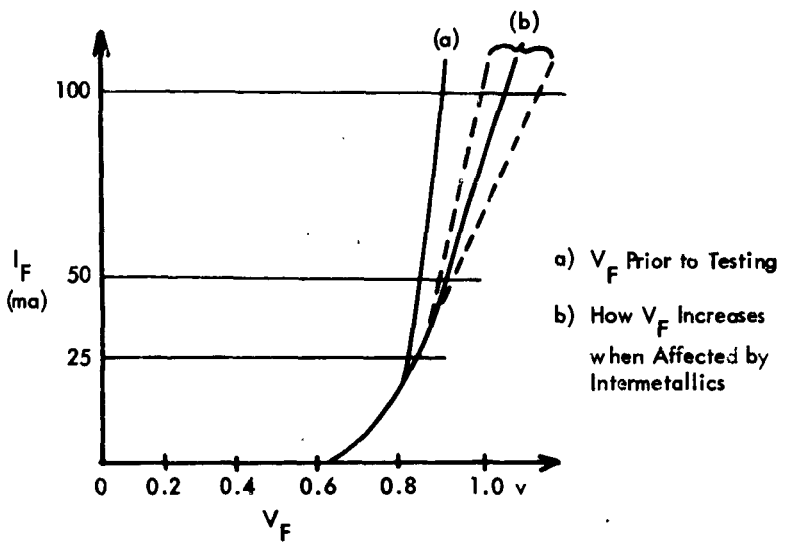


Figure 3

Forward Voltage Failure Criteria

IV FAILURE MECHANISMS AND/OR POTENTIAL PROBLEMS OF THE SEVEN METALLURGICAL SYSTEMS

Seven metallization systems of various manufacturers have been evaluated. They were packaged in TO-5, TO-18, or TO-46 cans.

A. Aluminum Metallization - Gold Wire

Initially it was felt that the purple phase, $AuAl_2$, was responsible for the failure of bonds on silicon planar transistors using the gold-to-aluminum system. This is not the case. The results of a Bell Laboratories experiment in 1962 indicate that the problem is with the tan phase Au_2Al.

The results of this experiment[5] are as follows:

Properties of $AuAl_2$ & Au_2Al

Property	$AuAl_2$	Au_2Al
Color	Purple	Tan[6]
Electrical Conductivity	High	Poor
Mechanical Strength	Strong	Brittle
Conditions of Formation	Heat	Heat, silicon as catalyst

The author's test experience on devices with 3000, 7000, or 12,000Å of aluminum metallization is summarized below. The results verify the tan phase being the problem.

3000Å Group. The tan phase was not generated by either a 1000-hour 150°C or 300°C bake on two separate groups of units with 3000Å of aluminum. The purple phase and black growth appeared around the periphery of the bonds in the 300°C group as shown in Figure 4. However, there were no significant increases in the forward voltage drops in either group. The black growth is formed by gold diffusion from the bond, out over the aluminum film. The rate of diffusion is both time and temperature dependent.

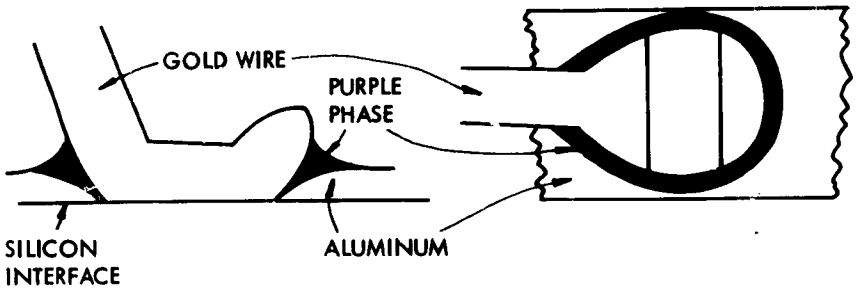


Figure 4

Purple Phase Forms On Au Wire Wedge Bonded To 3000Å Of Al After Being Subjected to 300°C for 1000 hours

5. John Haenichen, Motorola Integrated Circuit Design Course, Sec. 6, p. 8 (1963).

After the 300°C, 1000 hour bake, this group was subjected to ten, 2000 g, 0.5 millisecond blows followed by ten, 3000 g, 0.2 millisecond blows. The force was applied in the y_1 direction. There were no opens.

7000Å Group. Both the white and tan phases appeared in some units subjected to 150°C for 1000 hours. The white phase was assumed to be Au_4Al. Not enough was present to chemically verify this assumption. The white phase appeared near the gold lead, the tan near the silicon interface, as shown in Figure 5.

12,000Å Group. The tan phase was generated on units subjected to 150°C for 1000 hours. The emitter-base and/or collector-base forward voltage drops increased 50 mv or more. When sectioned, the tan phase appeared as shown in Figure 6. The results of a pull test experiment on the remainder of this lot are as follows:

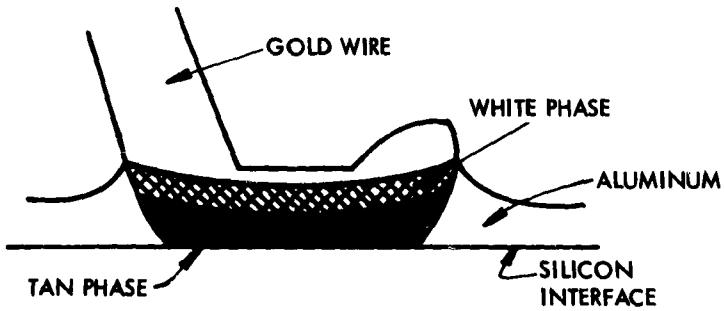
before bake	150 hrs. 150°C	350 hrs. 150°C
6.0 gms	6.0 gms	4.5 gms

Before the bake the wire broke. At 350 hrs. the bonds started to break due to the tan phase. Another group was subjected to 300°C storage. After 350 hours there were no bonds.

The following is proposed as the mechanism for increased forward voltage drops. Since the interface between the relatively pure gold and the tan phase is a region of high dislocation density, the gold diffuses from this area into the pure aluminum lands, producing porosity and voids, increasing the forward voltage drops, and eventually opening.

Grain Boundary Problem. There is a grain boundary problem using small diameter (0.4, 0.5 mil) gold wire as a means of interconnecting the chip to the header posts. After a 300°C bake, the wire will slip into sections, as shown in Figure 7. This is one reason why the largest possible diameter gold wire should be used for interconnections. The other is greater pull strength.

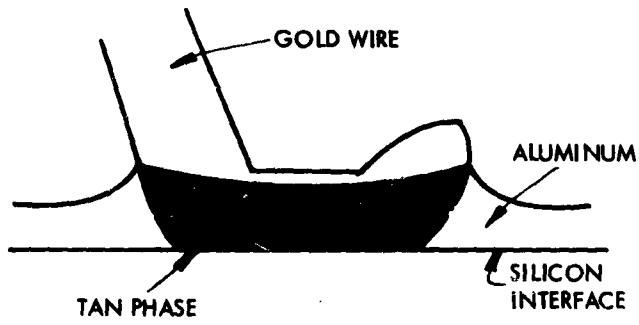
Conclusions. Intermetallic formation can be eliminated under the bonds by controlling the amount of aluminum metallization on the chip.



Side View

Figure 5

White And Tan Phases Form On Au Wire Wedge Bonded To 7000Å Of Al After Being Subjected To 150°C For 1000 Hours



Side View

Figure 6

Tan Phase Forms On Au Wire Wedge Bonded To 12,000Å Of Al After Being Subjected to 150°C For 1000 Hours

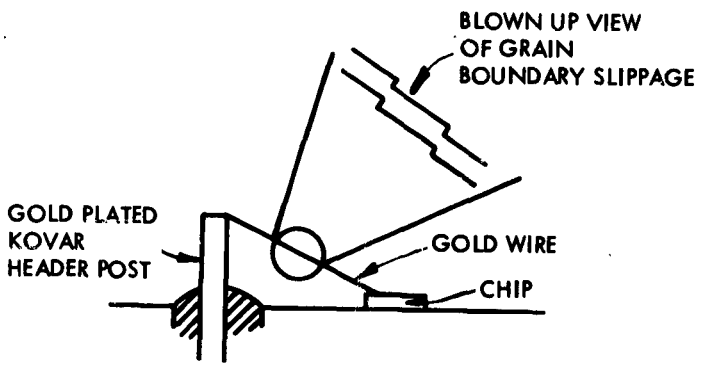


Figure 7

Grain Boundary Slippage

B. Gold Metallization - Gold Wire

Gold metallization results in the following three problem areas:

- The low Au-Si eutectic temperature (370°C) makes the gold metallized units far more susceptible to damage as a result of hot spots than the Al-Si (577°C) system.
- Gold does not reduce the residual oxide on the silicon surface, which is necessary to make a good ohmic contact.[6]
- Gold diffuses into silicon very rapidly during the metallization of the chip and again during the wire bonding operation, softening the breakdown voltages. Depending upon how the devices are being life tested or used, they could end up shorted.

Two manufacturers used gold metallization on the chip of their PNP silicon planar transistors. However, aluminum could have been used as a land contact material. To prevent the formation of a nonohmic contact on the base, an N+ enhancement diffusion (usually phosphorous) must be made prior to the application of the aluminum,[5] resulting in the structure illustrated in Figure 8.

---

6. B. Selikson and T. A. Longo, Proceedings of the IEEE, Vol 52, p. 1638 (1964).

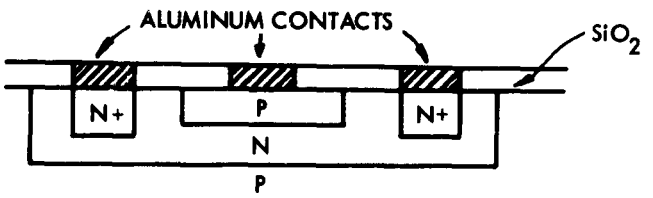


Figure 8

N+ Enhancement Diffusion, PNP Transistor

C. Aluminum Metallization - Silver Wire

When used as a lead material, silver has been observed to migrate under the ambient conditions described below. Such migration caused shorting between emitter and base due to the very close spacings involved. Because of this migration, intermetallics formed which affected the bonds as is evidenced in the following data:

before bake	150 hrs. 150°C	350 hrs. 150°C	150 hrs. 300°C
6.5 gms	2.5 gms	2.5 gms	no bond

The numbers above are the medium values of pull strengths. Before the bake the wire broke; after the bake the leads broke at the bond due to the intermetallics formed. More than one manufacturer has tried this combination experimentally. None, however, uses it.

D. Silver Metallization - Gold Wires

Silver metallization does not lead to a satisfactory solution. Silicon has a strong affinity for oxygen and is therefore always covered with a thin film layer. The silver metallization frequently does not penetrate the thin oxide to alloy into the silicon, and it pulls away after a short time at certain conditions, as is evidenced in the following data. Another problem was formation of intermetallics which weaken the bond.

before bake	150 hrs. 150°C	350 hrs. 150°C	150 hrs. 300°C
4.0 gms	*1.0 gms	*0.5 gms	no bond

* stripe material lifted away

The number above is the median value of pull strength. Initially some of the bonds broke, reducing the median value. More than one manufacturer has tried this combination experimentally. None, however, uses it.

E. Aluminum Metallization - Aluminum Wire

Three phases (purple, tan and white) have been detected between the aluminum wire and the gold-plated kovar header posts on two groups of diode quads over a two year period prior to testing. These units were subjected to 150°C for 1000 hours. The forward voltage increased 50 mv or more, indicating an exodus of gold at the gold intermetallic interface. This will be reflected in the bond strengths which are initially weak. In conclusion, the disadvantages of this system are:

- Intermetallics still form at the posts. Silicon is introduced into the aluminum in order to extrude small diameter wires required for interconnecting the chip to the header posts.[5] This must act catalytically in accelerating the rate formation of the compounds found at the posts.

- Bond strengths are initially weak. Initial bond strengths are reported to be about 0.5 gram[6] which is some 16X lower than the bond strength of an average ball bonded 1 mil gold wire.[7]

- Mechanical problems are associated with wedge bonding. They are discussed in Section II.

This system has only one advantage. It is useful on very small and very large geometry devices.[8]

F. Chromium - Silver Metallization - Gold Wire

This is a high temperature system tried by Philco on their 2N2710 transistors. A high-temperature lead-oxide silicate glass was used as the passivation layer. The devices had no aluminum between the Cr and Si interface. The advantage of this system is the 960.5°C melting point of the CrAg deposit. This type of metallization is far less susceptible to damage as a result of hot spots than the

7. J. A. Cunningham, Solid State Electronics, VIII, p. 735 (1965).

8. S. S. Baird, Conference on Reliability of Semiconductors and Integrated Circuits, p. 11 (1964).

9. Philco Planar Reliability Booklet, p. 3 (1963).

prior systems discussed in this paper. However, some of the problems of this system are:

- There is no mutual solid solubility and no compounds are formed between Cr and Ag. However, Ag does form a continuous series of solid solutions with Au.
- Silver migration will lead to whisker growth under certain conditions.
- Black bonds and lands will develop if the units are capped in a room ambient atmosphere and then subjected to high temperature stabilization bake. The compound formed is silver sulfide.
- A number of intermetallic compounds between silicon and chromium are reported.[10]

G. Molybdenum - Gold Metallization - Gold Wire

This system was selected as the best[11] for an expanded contact PNP fast switching transistor developed by Texas Instruments under contract to the U. S. Signal Supply Agency (Contract no. DA-36-039-AMC-00145(E)). The systems evaluated were: chromium-gold, chromel-gold, vanadium-gold, cobalt-gold, molybdenum-aluminum, nickel-aluminum and molybdenum-gold.

An N+ enhancement diffusion was required in the base prior to the molybdenum metallization.

Molybdenum and gold do not form intermediate phases, and the mutual solid solubility of gold and molybdenum is exceedingly small.[10] Devices that were heated at 300°C for 10 days and sectioned indicated no movement or mixing at the Mo-Au interface.[7]

During the past year a wide variety of accelerated aging environments have been placed on approximately 900 silicon transistors and IC's constructed with the Mo-Au expanded contacts. Open problems have never been observed, and there have been no failures which could be attributed to the contact system.[7]

---

10. M. Hansen, Constitution of Binary Alloys, p. 69, p. 217 (McGraw-Hill, New York 1958).

11. Defense Documentation Center for Scientific and Technical Information. Cameron Station, Alexandria, Virginia, p. 111, Publication AD443686.

The gold wire to aluminum land system could not be used. Although aluminum adheres very well to SiO_2, gold does not. In the bonding operation the gold wire would penetrate the aluminum and would be expected to make a good contact to SiO_2.[7]

Another obvious advantage of a bi-metal system is the extra metal at the step.

V RECOMMENDED SYSTEMS

A. Nonexpanded Contact, NPN or PNP, SiO_2 Transistors

- For NPN transistors, the gold wire to aluminum land system is recommended, keeping the following 3 points in mind.

  (1) Keep the amount of aluminum to a minimum. This will eliminate the formation of intermetallics under the bonds.

  (2) Use the largest possible gold wire. An example of the pull strengths of large diameter gold wire is as follows:

diameter	pull strength
1.5 mil	7 to 14 grms
2.0 mil	14 to 21 grms

  (3) Use a ball thermocompression bond.

- For PNP transistors, the same system is recommended. However, an N+ enhancement diffusion is required in the base prior to the aluminum metallization to achieve a good ohmic contact.

- The reliability of the bonds on transistors or other silicon devices using nonexpanded aluminum contacts is excellent, since the bonding wire penetrates the aluminum and becomes firmly attached to the underlying silicon.[8]

B. Expanded Contact, NPN Or PNP, SiO_2 Protected Transistors

The gold wire to molybdenum-gold land system recently developed by Texas Instruments is recommended. Again an N+ enhancement diffusion is required in the base of the PNP transistors prior to metallization. The largest gold wire and ball bonding should also be used.

SECTION II
TEST, ANALYSIS, AND CORRELATION

A TECHNIQUE FOR CONTROLLABLE ACCELERATION AND PREDICTION OF DEGRADATION MECHANISMS OF ELECTRONIC PARTS

T. Walsh

General Electric Co.
Missiles and Space Division
Philadelphia, Pennsylvania

M. Rocci

Rome Air Development Center
Griffiss Air Force Base, New York

As the required reliability of electronic parts increases together with the growth in functional complexity of electronic systems, the need for data for use in the selection, application and life prediction of parts becomes more acute. At the same time, both the delay and cost of obtaining such data by real time testing ranges from high to prohibitive. In addition, the improved reliability of parts today makes it difficult to determine life characteristics by conventional use level stress tests without large sample sizes and prolonged test times. Hence, the concept of accelerated testing is one of the most promising approaches to the solution of, or significant relief from the problem.

Accelerated Tests are tests which are designed to induce failures or significant parameter trends of the items on test in relatively short periods of time. The results from such tests may be used to predict the time to failure or parameter trends for similar items at use conditions provided the basic assumptions used in the test design and the analysis are correct. In reliability terminology, accelerated tests provide information regarding wearout failures.

There has been much effort sponsored by RADC during the past three years on the development of accelerated test techniques. Some of the present effort includes work on silicon epitaxial transistors, silicon planar integrated circuits, thin-film passive networks, high power silicon planar devices, snap-action switches, crystal

canned relays and soldered and welded connections.

Limited work has also been performed by other agencies. The most significant past work occurred in the Air Force Minuteman and Advent programs. The "A" program of Minuteman and the present "B" programs are more qualitative with both using high stress testing to identify predominant failure mechanisms which can be traced to the material, process or design. Under the Advent Program, high reliability diodes, transistors, resistors and capacitors were tested to determine and specify procurement methods, screening techniques, and life evaluation criteria for electronic parts. The information thus generated was to provide assurance that the flyable hardware would meet the program objective of a 3-year life in a non-maintainable space environment. Accelerated step-stress and constant-stress tests appeared to offer the most promising approach for solution to the problem. The test program was started in May of 1962 and continued until September, 1963, at which time the parts on life tests at various applied stresses and environmental conditions had accumulated 6,000 to 10,000 hours each. The parts were then placed at dormant conditions of 25C and zero power for 10 months and were re-activated to the same prior test conditions in June, 1964, as a requirement of Contract AF30(602)-3415. This one year program was jointly sponsored by RADC and NASA and it is with the results of this contract that this paper is concerned. The life tests will be continued until at least 25,000 hours of test time have been accumulated.

GENERAL DESCRIPTION OF TEST PROGRAM

If accelerated tests are to provide useful information for long life designs, the rate of degradation of the part at high stress levels must be related to the rate of degradation which would normally exist at lower "use" stress levels. The design of accelerated tests, therefore, must rely on knowledge and assumptions concerning degradation rates and continuity of failure mechanisms across the stress ranges of the test. It is because of this dependence upon knowledge of failure mechanisms that this program consists of two distinct yet interrelated activities. The first is the design, performance and analysis of accelerated tests and the second is the definition of physical models which describe the degradation of the parts. This paper is devoted primarily to a discussion of the test program and the use of data to predict the life characteristic of the parts.

Figure I shows in graphical form the seven-phase test plan used during this program. There are three step-stress tests and four constant-stress tests. It was the original intent of the program to provide a method which could correlate the results of step-stress tests with constant-stress tests. Thus, a technique would be available whereby analysis of short time step-stress tests could effectively predict the results of long term low level constant-stress tests.

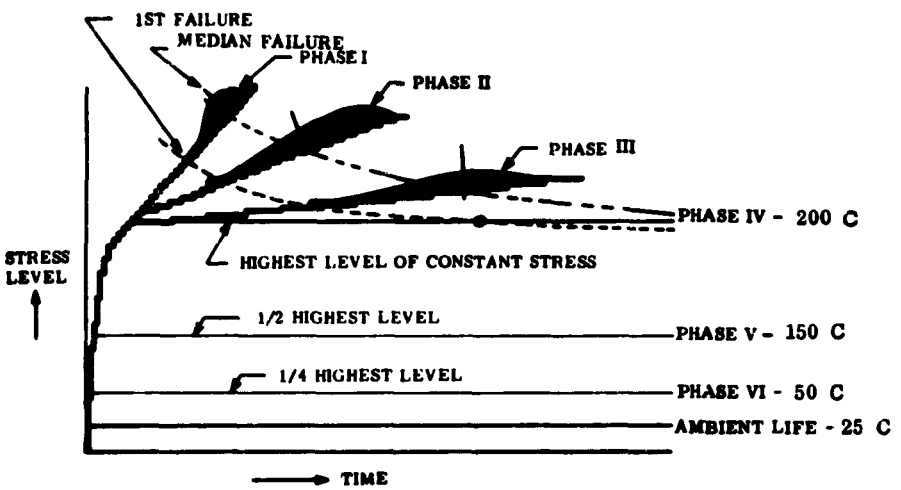


FIGURE I

GENERAL TEST PLAN

Each test phase of the program utilized sample sizes of 25 or 50 parts. The parts are fully screened high reliability parts purchased to General Electric R-series specifications. The selection of stress levels and rates of increase of stress for the various phases was based upon expected part parameter degradations, measurement accuracy limitations and the capabilities of laboratory stressing equipment. The parts placed on test are shown in Table I.

The step-stress tests associated with the program have been completed. The constant-stress tests have accumulated over 20,000 hours of testing with the high level constant-stress tests being 70% completed. No failures

R-NUMBER	DESCRIPTION	CLOSEST COMMERCIAL DESIGNATION
Resistors		
R2016	Ni Cr Metal Film, 1/8 Watt	XLT, IRC
R2048	Tin Oxide Film, 1/8 Watt	NF60, Corning
Capacitors		
R2045	Fixed Glass Dielectric	CYFR, Corning
Diodes		
R2008 P5, P10	Silicon, Regulator	IN751A, 758A, CDC
R2010 P1	Silicon, VHF	IN251, CDC
R2011 P1	Silicon, Computer	IN643A, CDC
R2013 P1	Silicon, Low Current	IN647, PSI
Transistors		
R2004 P1	NPN, Silicon, 3 Watt	2N1613 FSC
R2005 P1	PNP, Silicon, 2 Watt	2N1132 FSC
R2026 P1	NPN, Silicon, 2 Watt	2N708 FSC
R2050 P1	NPN, Silicon, 4 Watt	2N657 FSC
R4041 P1	PNP, Silicon, 1 Watt	2N869 FSC

TABLE I

DESCRIPTION OF PARTS TESTED

have occurred over this period of time on the low-level tests; however, the measured part parameters are in most cases showing definite degradation trends.

The final report on contract AF30(602)-3415 is being issued by RADC and it contains a complete discussion, analysis and display of the data. The following examples are given to illustrate the analysis method used for combining the results of the various test phases into a stress versus expected life chart which may be used in the design of long life equipment.

DISCUSSION OF RESULTS

In general, it may be stated that the lifetime of a device is the time during which no measurable parameters exceed established limits. The cessation of "life" may be observed in either of two ways: first, it may occur at a measurable rate of degradation when a measurable part parameter exceeds established limits, as in the case of resistance drift, of a resistor; or second, it may occur catastrophically without indication of measurable parameter variations when the strength of the device degrades to the applied stress level (environment stress plus operating stress), as in dielectrics which fail suddenly due to dielectric breakdown. Thus, two general techniques were used for the analyses of the data. Both are discussed below and data are shown to illustrate each technique. Recommendations are also given for the design and analysis of future accelerated test programs.

CASE I DEGRADATION OF PART PARAMETER

For degradation type failures, the rate of reaction of the degradation mechanism will govern the rate of change of the part parameters. The approach taken here was to assume that the mechanism causing the drift of the parameter is chemical in nature, thus the relationship of reaction rates and temperature expressed by the Arrhenius or Eyring Equations was used to develop models to describe the degradation. A film resistor is selected to illustrate the analysis technique.

Briefly, the approach to testing the film resistor considers thermal stress as the independent variable and some function of resistance as the dependent variable of degradation. Thermal energy, precisely controlled and measured, was selected as the stress to cause deterioration of the resistor's active element (i.e., the resistive film). The necessity of uniformly applying thermal energy resulted in the design of special mounting fixtures and the use of ovens with accurate temperature control. The resistors were tested with a voltage applied which was 50 percent of the manufacturer's power rating to accentuate the effect of any structural defects or imperfections of the resistor. A function of resistance ($|\Delta R|/R$) was selected as the damage sensitive parameter which would linearize the degradation of the resistor in time.

The R-2048(1/8-watt) is a film resistor which has as as its basic active resistive element a semiconducting tin-oxide film doped with antimony to provide varying values

of resistivity. After film deposition on the ceramic substrate, the ends of the resistor blank are coated with conducting material. Heat treatment provides stabilization of the oxidation reaction and stress relief. The blanks then have a helix cut through the resistance film to provide the final resistance; leads are attached and the resistor is hermetically sealed in glass.

The resistors on long time constant-stress tests continue to exhibit a uniform degradation. Figure 2 shows the average cumulative damage of the 100,000 ohm resistors for Phases IV (200C), Phase V (150C) and ambient (25C) and 15,000 hours of test time. The relation of damage and time appears to be linear and lines were drawn with the same slope.

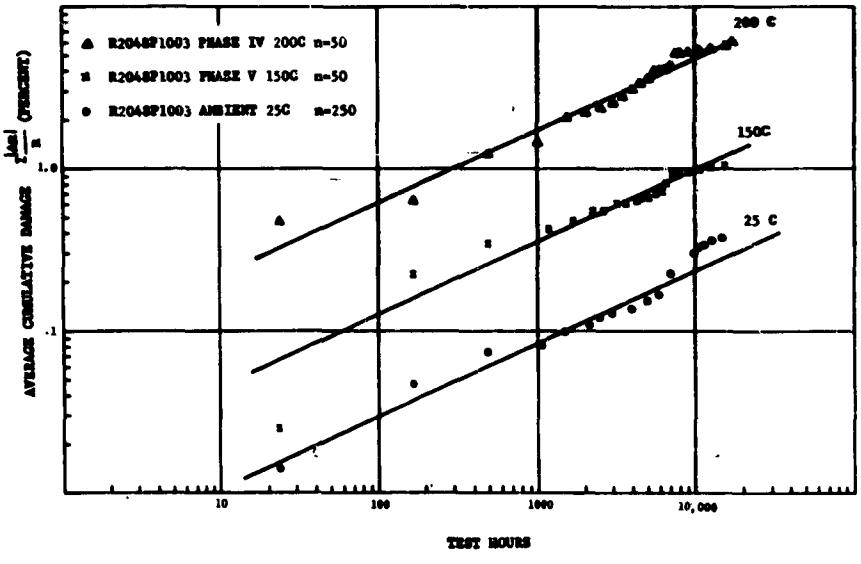


FIGURE 2

CUMULATIVE DAMAGE VERSUS TIME FOR THE R2048P1003 RESISTOR

Figure 3 is a log-normal plot of the accumulated

damage for individual resistors for three time periods - 500, 5,727, and 15,448 hours. It is apparent from this plot that the distribution of accumulated damage in time is remaining consistent over 15,000 hours. The fact that there have been no catastrophic failures during the tests and the obvious homogeneity of the parts makes the analysis of the results more meaningful.

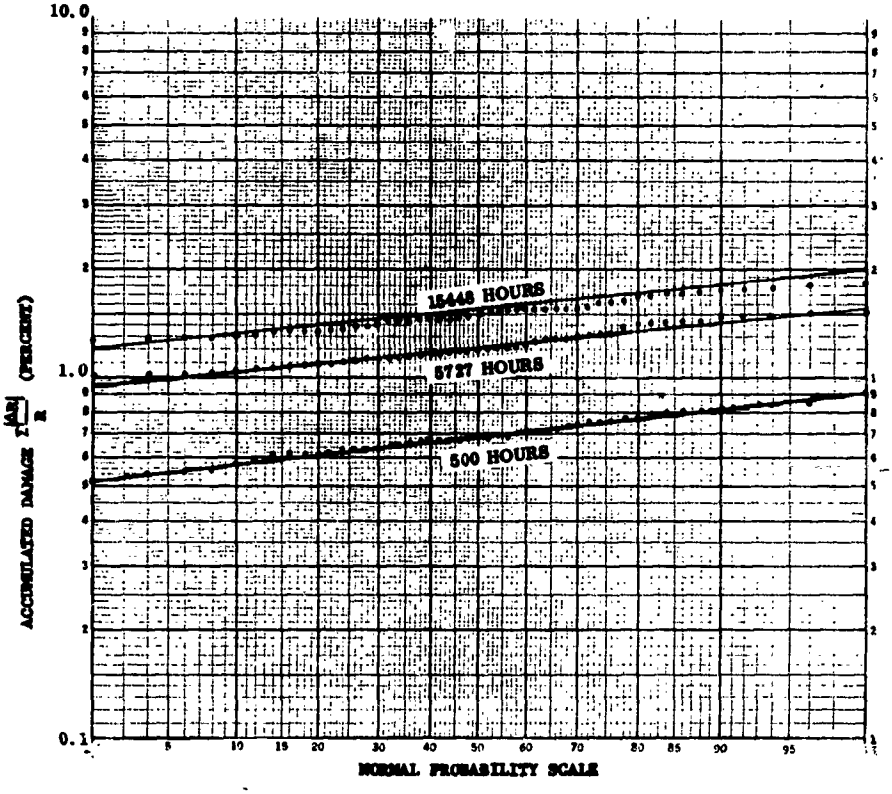


FIGURE 3

CUMULATIVE DISTRIBUTION OF DAMAGE FOR FIXED TIME
R2048P1003(150C)

The accumulated damages in time for the step-stress and constant-stress tests are plotted in Figure 4. Phase I was performed using 12-hour steps; Phase II using 60-hour steps and phase III using 300-hour steps. Lines of constant damage values were drawn using the data of the three step-

stress tests. The lines are parallel to each other indicating that the rate of damage accumulated for the times shown are similar. The ½% and 1% accumulated damage points for the 200C and 150C constant-stress tests are shown and connecting lines are drawn. The data is now in a format which may be used by design engineers making long-life designs.

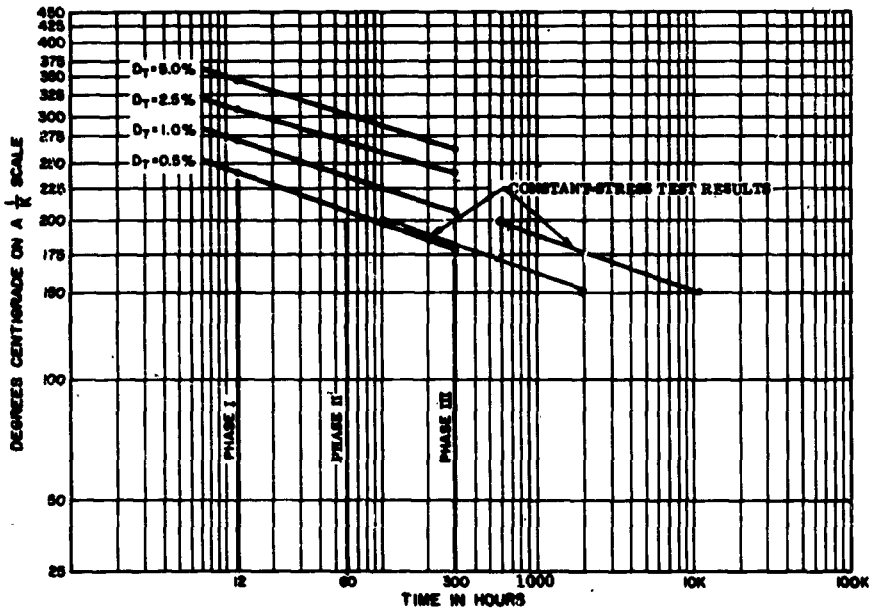


FIGURE 4

R2048P1003 LIFE DATA CHART

If the step-stress constant damage lines are extended to longer time periods, and compared with the constant-stress test results, they will show a conservative estimate of accumulated damage. This is as expected since the total damage plotted for any step of the step-stress tests is the accumulated damage of all the previous steps. Mathematical relationships between step-stress and constant-stress test results have been derived; however, we have found it simpler to use the step-stress results for information regarding the consistency of failure mechanisms and the selection of the stress levels for the constant-stress tests and not for life predictions. At least three constant-stress tests are recommended to be used in the construction

of the life data charts.

From the results of this type of test where a damage sensitive parameter can be measured, we believe the performance of two step-stress tests and three constant-stress tests to be a reasonable approach to the initial design of accelerated tests and the development of life data charts. A technique which may be used for the analysis of combined accelerated test results to generate life data charts is as follows:

1 - Perform two step-stress tests to failure using a time ratio of 1 to 5. For instance a twelve-hour step may be used for the first test and a 60-hour step for the second test. The data are then plotted using relationships developed from degradation models devised as a result of a physics of failure investigation. Figure 5 shows a plot of resistor data which was generated during this program. The fact that the data fit lines drawn parallel at the same slope provides evidence that the failure mechanism producing the damage remains consistent throughout the stress range shown; therefore, the high level constant-stress tests may be designed with some assurance of a constant degradation mechanism.

2 - The step-stress data of Figure 5 are replotted on the life data format of Figure 4. Stress levels for the three high level constant-stress tests may then be selected and estimates of the time required to perform the test are possible. Consideration must be given here to the capability of the test facility to control the stress accurately and measure the part parameters reliably.

3 - Analysis of the data from the three different constant-stress tests could take the following form:

a) Make a plot of the median failure points on a stress versus test-time chart as shown in Figure 6. This line may then be extended to lower levels of stress. This will provide an estimate of the median time to failure at use stress levels. Additional percent failure points may be estimated in the same manner.

b) Plot the times to failure or times to reach

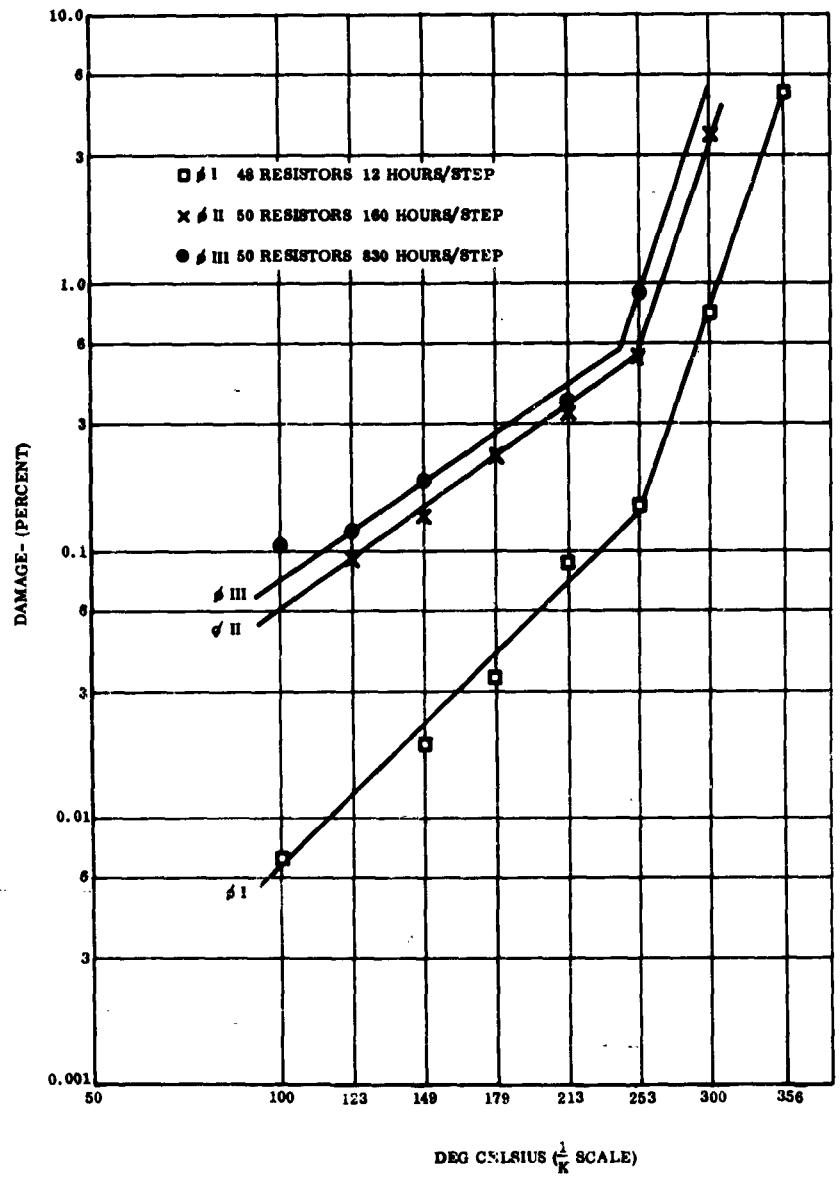


FIGURE 5

CUMULATIVE DAMAGE VERSUS RECIPROCAL OF ABSOLUTE TEMPERATURE
R2016P1003, STEP-STRESS TESTS

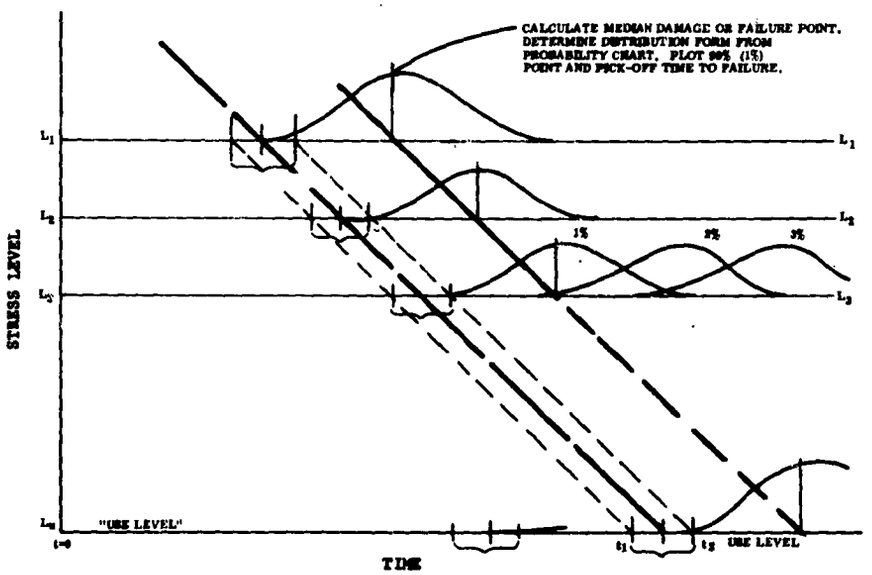


FIGURE 6

TECHNIQUE FOR PLOTTING CONSTANT-STRESS DATA

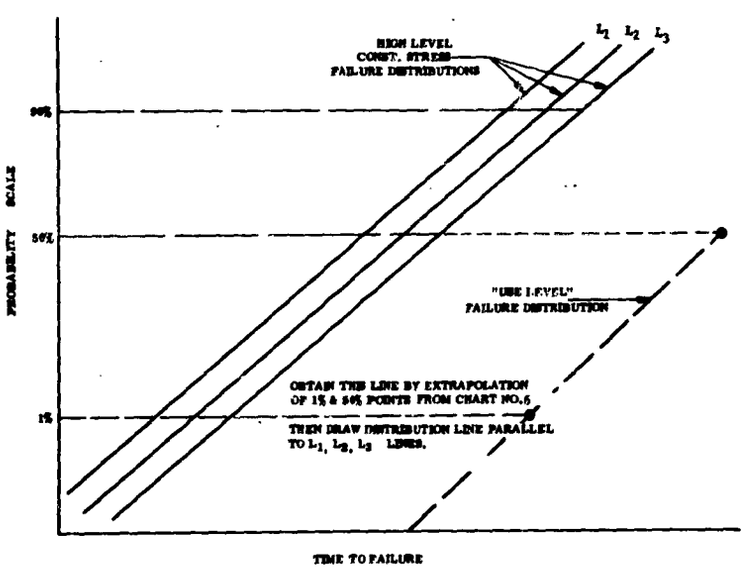


FIGURE 7

ECHNIQUE FOR ESTIMATING FAILURE DISTRIBUTION AT USE LEVEL

some level of damage on suitable probability paper as shown in Figure 7. If the data plots are parallel, then it may be said that the failure distribution is essentially constant for the parts at the stress levels used in the tests. If they are not parallel, then the failure distribution is changing in time and additional tests and analysis may be necessary.

c) If the lines are parallel, another parallel line may be drawn at a use stress level by using the median time and some other percentage failure point estimated from Figure 6. An estimate of the life distribution of the parts using the selected failure criteria is then available for long life designs.

CASE II DEGRADATION OF ULTIMATE STRENGTH

In the case where a damage sensitive parameter which is consistently degrading in time cannot be found, it is necessary to define an arbitrary failure criteria for the parts on test. In capacitor testing, this is usually a catastrophic failure or dielectric breakdown. We use this approach also for plotting the results of semiconductor tests. A damage function which would describe the parameter degradation for all semi-conductors on test, in a consistent form, has not as yet been fully developed, although some of the test results show very consistent parameter trends.

It was anticipated that surface effects would dominate the degradation of the parts on test and various models of surface behavior were used as the basis for the selection of the parameters likely to be degrading. Thus, the parameters selected are primarily surface related measurements. The electrical power applied was the same for all tests of the same part type. Thermal energy was used as the stressor during the step-stress and the constant-stress tests. The parts are removed from stress and allowed to cool for eight hours prior to measurement of parameters.

The transistor selected for this illustration is a silicon, NPN, planar, surface-passivated, medium power, high-voltage transistor. It has a maximum power dissipation of four watts at 25C case temperature, zero watts at 200C case temperature; and a maximum collect-to-emitter voltage rating of +100 volts. It is in a TO-5 case, dry-nitrogen

filled. It is designated as the R2050P1 and its closest commercial equivalent is the 2N657.

Quantities of 50 parts each were subjected to Phase IV (200C), V (150C) and Ambient Life (25C) tests. The estimated average junction temperatures are 250;200;and 175C, respectively. The parameter trends are as shown in Table II.

	I_{CBO}	h_{FE}	BV_{CBO}	BV_{EBO}	BV_{CEO}
Phase IV T_j = 250C - 8300 hrs	INC.	DEC.	INC.	Stable	Stable
Phase V T_j = 200C - 16000 hrs	SLT. DEC.	DEC.	INC.	Stable	Stable
Ambient Life T_j = 175C - 16000 hrs	SLT. DEC.	DEC.	INC.	Stable	Stable

TABLE II

CONSTANT STRESS TEST PARAMETER TRENDS-R-2050P1

The failure times of Phase IV have been plotted on log-normal probability paper in Figure 8. A straight line was fitted to the data to obtain estimates of the 2 percent and 50 percent failure times. These points were then plotted on Figure 9 together with the Phase I and Phase II step-stress test results.

The step-stress and constant-stress data points appear to line up. The 2% failure point for the phase V test is expected to occur between 20,000 and 30,000 hours. There have been no failures in 16,000 hours of test at phase V conditions; however, degradation in Beta is occurring and it appears likely that the first failure could occur in the time period estimated.

A technique which could be applied to analyze accelerated test results of parts which do not exhibit consistent degradation is as follows:

1. Perform two step-stress tests as discussed under Case I. The selection of the failure criteria is extremely important. Figure 10 shows a typical plot of data generated during this pro-

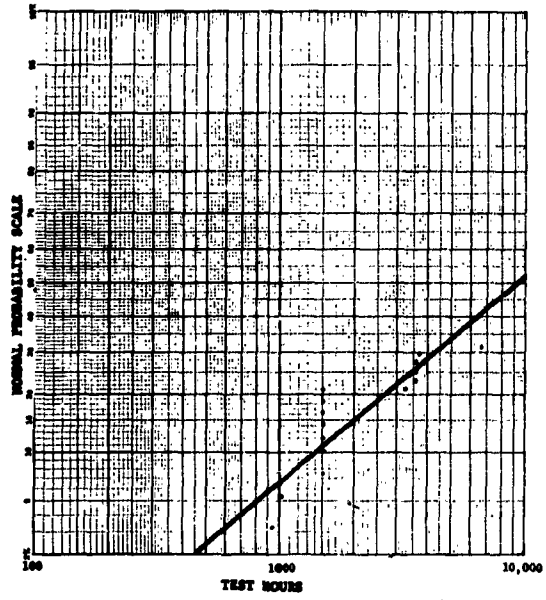


FIGURE 8

LOG NORMAL PROBABILITY PLOT OF PHASE IV FAILURES, R2050P1

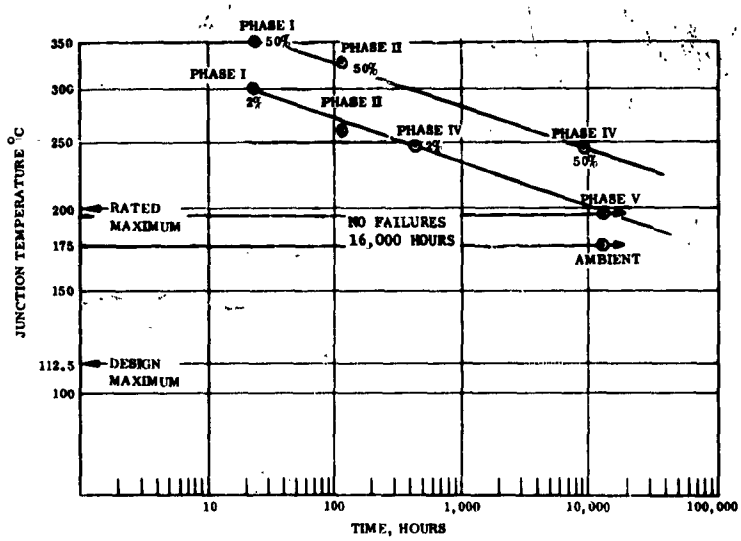


FIGURE 9

LIFE DATA CHART, R2050P1

gram for a diode. The step-stress data in this case are plotted in a time to failure versus a probability scale format rather than the damage versus stress form of Case I.

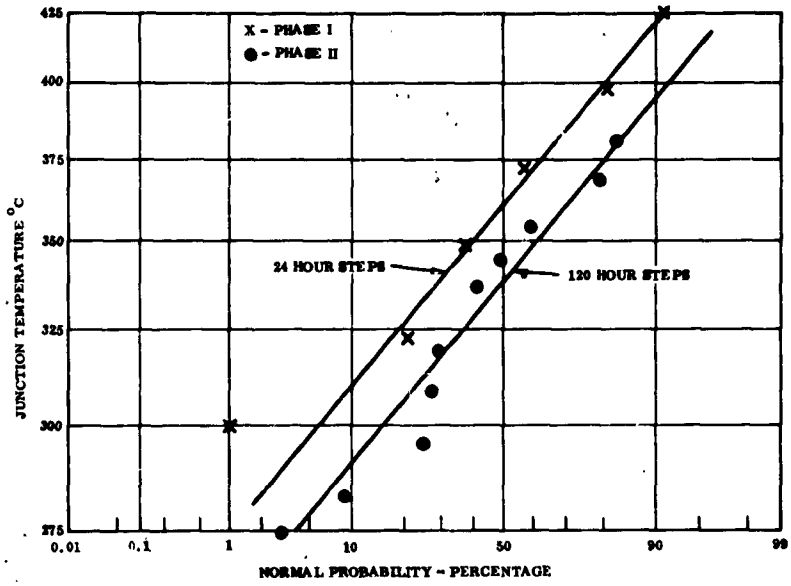


FIGURE 10

STEP-STRESS RESULTS, R2008P5

2. The step-stress data may then be re-plotted on the life data format of figure 4 and analysis and presentation of the constant-stress test results would be similar to that shown for Case I.

CONCLUSIONS

Accelerated tests yield data which when properly analyzed can provide estimates of the expected occurrence of failure of homogeneous parts in time. The methods of analysis shown here may be used to reduce the information into a format which can be used in the design of long-life hardware.

CUMULATIVE DEGRADATION MODEL AND ITS APPLICATION TO COMPONENT LIFE ESTIMATION

Hiroshi Shiomi

Electrotechnical Laboratory, Ministry of
International Trade and Industry
Nagata-cho, Tokyo, Japan

1. Introduction

It is quite important to find a general degradation model for prediction and acceleration of component life under any stress condition, both for constant and variable stress conditions. For this purpose, the reaction theory gives a most powerful basis. Under the assumption of cumulative degradation, a general degradation model which is essentially applicable not only to electrical components but also to mechanical elements has been developed.[1)2)] The validity of the model has been investigated by experiments of some electronic components. Applications of this principle to component life estimation are also illustrated.

2. Reaction Model and Acceleration Factor

2.1 Reaction Model

One of the most popular degradation models is the reaction model. A physical strength (characteristic parameter) $u(t)$ and concentration of the reaction N are connected with time t and a reaction rate K by some functional relationship f and Ψ respectively.[3)]

$$f(u(t))=\Psi(N)=Kt \tag{1}$$

The reaction rate K is given by Eyring,[4)5)]

$$K=AT\exp(-B/T)\exp(S(C+D/T)) \tag{2}$$

where A,B,C,D are constants, T is temperature and S is stress. The approximate expression of K is the well known Arrhenius equation. For example, exponential decay $u(t)=N(t)/N_0=\exp(-Kt)$ is obtained from a reaction equation $dN/dt=-KN$ and power (diffusion) type decay $u(t)^2=(\Delta N)^2=Kt$

is deduced from a reaction equation $dN/dt=-K/(N_O-N)$.

The relationship between life and stress factor K is obtained by putting t=L when the strength u(t) decays down to FL (Failure Level, known or unknown).

$$L=f(FL)/K \tag{3}$$

2.2 Failure Rate Acceleration Factor and Lifetime Acceleration Factor

Equation (1) shows that degradation is uniquely determined by the product of K and t which are mutually independent, and this relationship is interpreted as the transformation equation between stress and time.

For a constant stress condition, the lifetime acceleration factor A_L is given by:

$$A_L=L_r/L=K/K_r \tag{4}$$

where suffix r means the reference condition.

On the other hand, failure rate acceleration factor A_λ is not always equal to AL. If the failure distribution is given by the Weibull distribution $R(t)=\exp(-t^m/t_o)$, A_λ is expressed by:

$$A_\lambda=t_{or}/t_o=\lambda(t)/\lambda_r(t)=(K/K_r)^m \tag{5}$$

because $\lambda(t)$ and L are given by $\lambda(t)=mt^{m-1}/t_o$ and median life $L=(t_o\ln 2)^{1/m}$ respectively. From equations (4) and (5),

$$A_\lambda=A_L^m \tag{6}$$

The following is the example of condenser life data. The so-called 5th life law is well known for condensers under direct current voltage stress. The reaction rate and A_L are expressed as follows:

$$K=V^k A T\exp(-B/T) \tag{7}$$

$$A_L=(V/V_r)^k(T/T_r)\exp(B\Delta T/\overline{T}^2) \tag{8}$$

For convenience, A_L is approximately expressed:

$$A_L=(V/V_r)^k 2^{\Delta T/\theta_r} \tag{9}$$

where $\overline{T}^2=TT_r$, $\Delta T=T-T_r$ and $\theta_r=(\overline{T}^2\ln 2)/B$ which is an equivalent temperature rise to reduce lifetime by one half.

In the case of failure rate, those acceleration coefficients are all multiplied or divided by m, that is, k,B,θ_r for A_L are replaced by $mk,mB,\theta_r/m$ for A_λ. We must be careful not to mix them up in the

expression of life test data. The life data of mylar condenser applied direct current voltage show that the slope of lnL:lnV plot is k=4.95 and the slope of $\ln t_0$:lnV plot is mk=3.62. The slope ratio 3.62/4.95=0.73 coincides well with the Weibull shape parameter m=0.72~0.77 obtained from independent experiments.

3. Cumulative Degradation Model and Generalized Miner's Equation[1)2)]

To find a general degradation model, we make the following fundamental assumptions: (1) Degradation follows a f=Kt type reaction model, and (2) Degradation is linearly accumulated independently of the past stress path. This means degradation proceeds one way without hysteresis or recovering effect.

If the component is put under the stress-time state (K_i,t_i) successively, the total degradation is determined by the net accumulation of those k_it_i products.

$$f=\sum K_it_i \tag{10}$$

If stress is continuously applied with time to the component,

$$f=\int K(t)dt \tag{11}$$

Those general expressions include the case of step stress method and constant stress test as a special case, that is, we get from (10), $f=t_s\sum K_i$ for step stress and from (11), $f=K\int dt=Kt$ for constant stress. Another example of expression $f=\int K(t)dt$ is the progressive stress method applied to condenser life[6)7)] in which time is artificially introduced into K such as V=αt of equation (7).

The stress-time state (K_i,t_i) is applied successively until the life end occurs at the nth state when parameter u reaches the failure level FL.

$$f(FL)=\sum_1^n K_it_i \tag{12}$$

The expected mean lifetime is obtained from $\sum_1^n t_i$. In the case of the step stress method, this becomes:

$$f(FL)=t_s\sum_1^n K_i=t_sK_n\gamma_n \tag{13}$$

where γ_n is a correction factor.

We can get a more convenient expression putting $(1/L_i)=K_i/f(FL)$ in equation (12).

$$\sum_1^n (t_i/L_i)=1 \tag{14}$$

For metal and alloys, the relationship between logarithm of mean cycles to fatigue N and applied alternative stress S is often expressed by S-N curve. When n_i cycles at stress S_i are applied successively from i=1 to i=n, until failure at the nth step, the expected total fatigue cycles $\sum n_i$ is estimated from the so-called Miner's equation, $\sum (n_i/N_i)=1$ which just corresponds to equation (14).[8] M. Miner introduced this equation rather intuitively assuming the network absorbed at failure, $W=\sum w_i$, is related to $(w_i/W)=(n_i/N_i)$. The expression (14), however, can be applied to much more general cases, including the Miner's rule of mechanical fatigue life. We may call equation (14) the generalized Miner's equation. It is reported that $\sum(n_i/N_i)$ is not always equal to unity at fatigue life and this value depends on material, shape of item, treatment, and stress application path. But, as a first approximation equation, it is often used as a preliminary design tool and for life estimation criterion in the mechanical engineering field.

4. Experimental Verification of Linear Accumulation Degradation

The linear accumulation assumption means that the same applied stress-time set always results in the same state of degradation, independently of stress application path. Deviations from the rule $\sum (t_i/L_i)=1$ are also caused by the invalidity of assumption of a unique and monotonic degradation reaction.

Checking experiments have been conducted for some components to which step-up, step-down, cyclic, random and constant stress are applied and their degradation behaviors are observed. According to those experiments, accumulation of degradation is influenced by kind of stress, stress path direction, past stress history, degree of component degradation, mixing of decay modes, change of drift pattern, and stabilization process.

4.1 Example 1: Degradation of Transconductance of Electron Tube

A. Degradation Pattern

It was found that transconductance gm of electron tube follows well an exponential decay pattern[11], $gm(t)/gm_0=u(t)=\exp(-Kt)$. One example of obtained K value under constant stress is shown in Fig. 1. Strictly speaking, the K value which reflects mainly decay of emission and growth of interface resistance of oxide cathode is not constant throughout life, but its pattern is changing quite regularly with time and stress. From the macroscopic viewpoint, it might be said the exponential pattern is well held for gm decay.

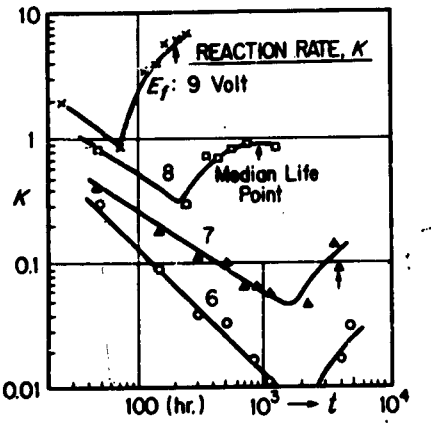


Figure 1

Variation of Reaction Rate K with Time

B. Effect of Stress Path on Degradation Accumulation

Degradation amount during each step for various stresses was observed and its hysteresis effects were investigated. Obtained results are summarized as follows:

(1) Monotonic step-up stress:

When the K value of each step is almost the same as the value of the corresponding constant stress, then it is expected the linear accumulation rule is applicable to a step-up stress experiment. Fig. 2 shows the comparison of K values of constant stress, monotonic step-up stress, and random stress. The calculated mean $t_s \sum(1/L_i)$ values under several conditions of step-up stress are quite near to 1 (Table 1). It seems that small deviation from unity is due partly to sample variation.

(2) Monotonic step-down stress:

As stress decreases, a big recovering effect is observed as shown in Fig. 3. The total degradation amount is much smaller than the degradation of step-up stress even though the total applied stress-time set is the same.

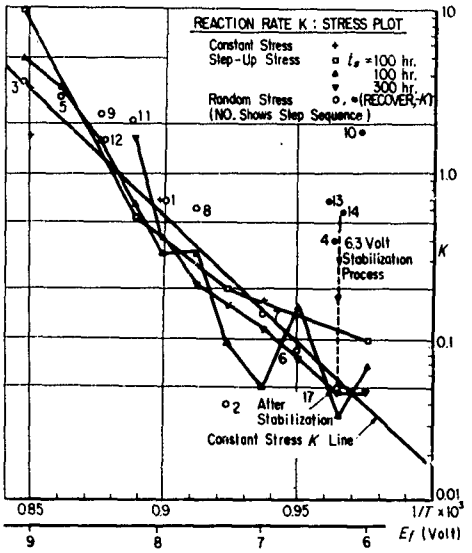


Figure 2

Reaction Rate K under Various Stress

sample	stress step $\delta=\Delta(1/T)\times10^3$	time step $t_s\times10^3$Hr.	$\sum^n K_1t_1$*	$t_s\sum^n(1/L_1)$
A	0.0125	0.3	0.765	1.1
	0.0125	0.1	0.89	1.28
	0.00625	0.3	0.70	1.01
	0.00625	0.1	0.72	1.04
B	0.0125	0.3	0.57	0.82
	0.0125	0.1	0.653	0.95
	0.0125	0.1&0.05	0.692	1.00

* theoretical value is $Kt=\ln(1/0.5)=0.693$, & $t_1=t_s$

Table 1

$t_s\sum^n(1/L_1)$ **Value for FL(Failure Level)=50% of** Gm_o

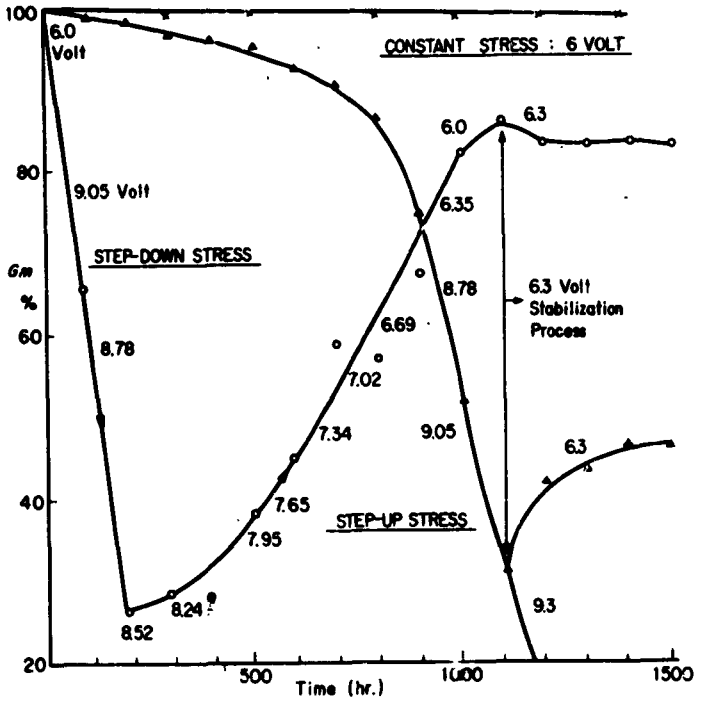


Figure 3

Comparison of Step-up and Step-down Stress

(3) Random and cyclic stress:

When the component is not so degraded and remains in an earlier state, the K value is almost the same as the corresponding constant stress value. In general, when the present stress is decreased down from the previous stress it recovers, that is, when $\Delta u = u_i - u_{i+1}$ is negative, then the calculated K value, $\Delta u/(\pi \Delta t)$, is also negative (• mark in Fig. 2). On the other hand, when stress increases up from the previous stress, degradation is accelerated and K value becomes greater than the constant stress value.

The recovering amount and the amount of excessive degradation are both approximately proportional to the stress difference between the present stress and the previous stress, that is, $\Delta(\Delta gm/gm_0)$ or $\Delta K \propto /\Delta$ stress/. On an average, their effects are canceling out each other as long as stress variation is well balanced and then it tends to degrade the same amount by application of the same stress-time set. Fig. 4 is one example of those experiments. Decay curves of step-up, cyclic and random stress coincide well at expected points.

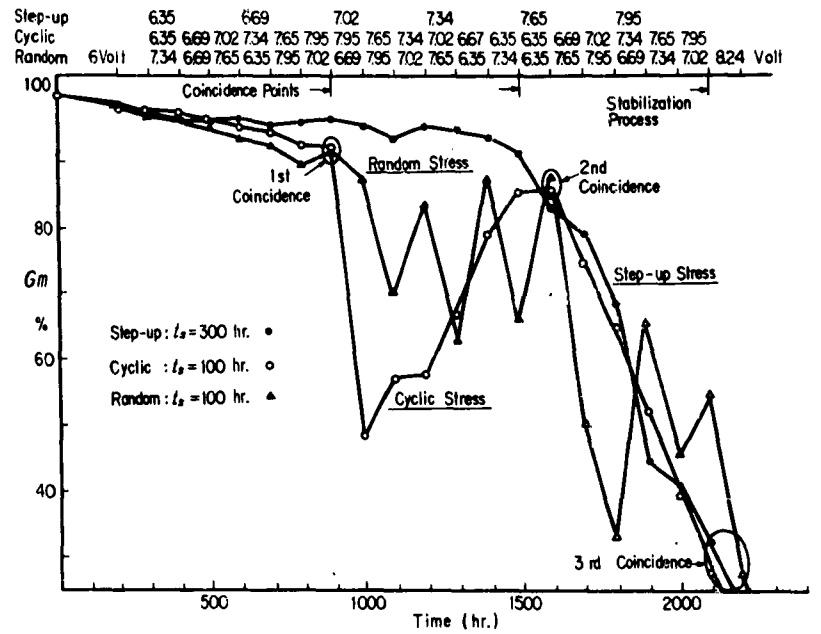


Figure 4

Degradation Curves under Various Stress Conditions

(4) Stabilization process:

It should be emphasized that a stabilization process is necessary to check the true degradation state. The stabilization at a reference condition wipes off hysteresis effects due to past stress history and it brings observed K value back to the corresponding constant stress value (see Fig. 2).

The time interval necessary to stabilize the component greatly depends on the degree of degradation. The smaller the decay, the greater and more rapid is the extent of recovery. The degraded sample loses its recovering force and it takes more time to restore.

The reaction model assumes essentially one-directional concentration flow being independent of actual dimension. But, in a degraded sample, positional uniformity of concentration is lost. The variation of concentration localization caused by external stress change will produce the hysteresis effect. Monotonic step-up stress may not disturb this one-directional model and, accordingly, the K value is kept the same as the steady value.

4.2 Example 2: The Case of Ge. Alloy Transistor

At the earlier period of stress application, a reversible parameter change caused by equilibrium transition of the surface state is usually observed. When the sample is removed from stressed state to the reference condition, say room temperature, the parameter recovers reversibly to its initial value. The detailed experiments and physical explanations for reversible effect of transistor are reported by J. Partridge, et al.[12)14)] The recovering time to initial value differs very much from sample to sample depending on its surface state. Meanwhile, the irreversible true degradation due to surface oxidation superposes on the reversible drift.

For the pnp Ge. alloy transistor with rather accumulated surface state, temperature rise brings its surface to more accumulated state on s(surface recombination velocity)-φ_s(surface potential) diagram. On the other hand, oxidation moves the surface state in reverse direction, that is, less accumulation. Coexistence of those two modes causes rather complex drift behavior. To distinguish reversible and irreversible change and to detect true degradation, a stabilization process is also required.

The shape of h_{fe} drift and the time to equilibrium during stabilization greatly depend on transistor type, surface state (manufacturer), and degree of degradation. The reversible change at initial period shows longer time constant (from 10 hr. to more than 10^3 hr.), but, as the true degradation dominates, the time to equilibrium becomes shorter (less 100 hr., sometimes less than 0.1 hr.). Finding reasonable and common stabilization time to any sample is impossible. However, it is still important to set up the reference condition with minimum stabilization time in accordance with each lot behavior.

Some examples of h_{fe} transition drift of alloy transistors are shown in Fig. 5. In some step stress experiments, it is found that the total net decrease of h_{fe} is less than the ultimate total decrease under constant stress. This discrepancy comes from the accumulated transition effects and inadequacy of stabilization.

Fig. 6 shows h_{fe} hysteresis drift under cyclic temperature stress. Degree of hysteresis depends on transistor type, manufacturer, and past stress history. In sample A, remarkable hysteresis is observed. Its dependency on the previous stress is almost the same as the case of transconductance, but the regularity of dependency is not so apparent. It is worthwhile to note the stabilization time of sample A is much longer than that of sample B (70 hr.~less than 0.1 hr. for sample A and 1 hr.~less than 0.1 hr. for sample B). The slower response and time lag to the outer stress change is the main reason of this hysteresis drift. However, the resultant net degradation measured after stabilization seems almost the same in spite of different sequence of applied stress as far as the same stress-time set is applied.

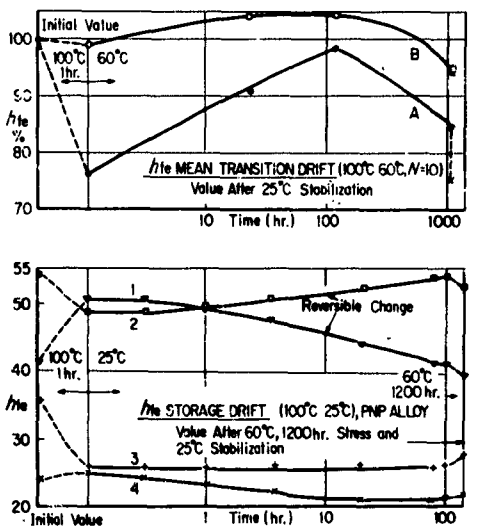


Figure 5 : Transition Drift of h_{fe}

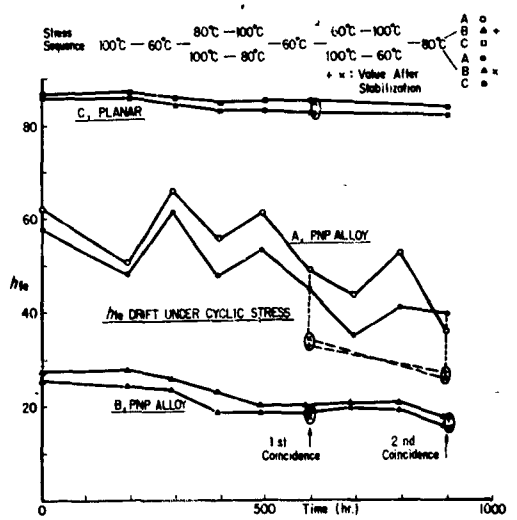


Figure 6 : h_{fe} Drift under Cyclic Stress

4.3 Example 3: The Case of Composition Resistor

Composition resistors under thermal and power stress show also two modes of resistance change: (1) initial increase caused by ceramic curing, and (2) ultimate decrease caused by carbonization or separation of composite elements. The latter change finally ends in short failure or open defect by cracking.[15][17]

The cyclic variation of those stresses usually causes hysteresis effect. The tendency is similar to that of gm case. For sample A in Fig. 7, stress decreasing results in resistance recovering for both temperature and power. For sample D, the curing process is so remarkable that the power effect on hysteresis is just the reverse of temperature action. Thus, the matrix acceleration life test by temperature and power is inapplicable to sample D. In the earlier stage, subtle mixing of drift modes is generally observed depending on the combination of stress factors: temperature, power, and moisture.

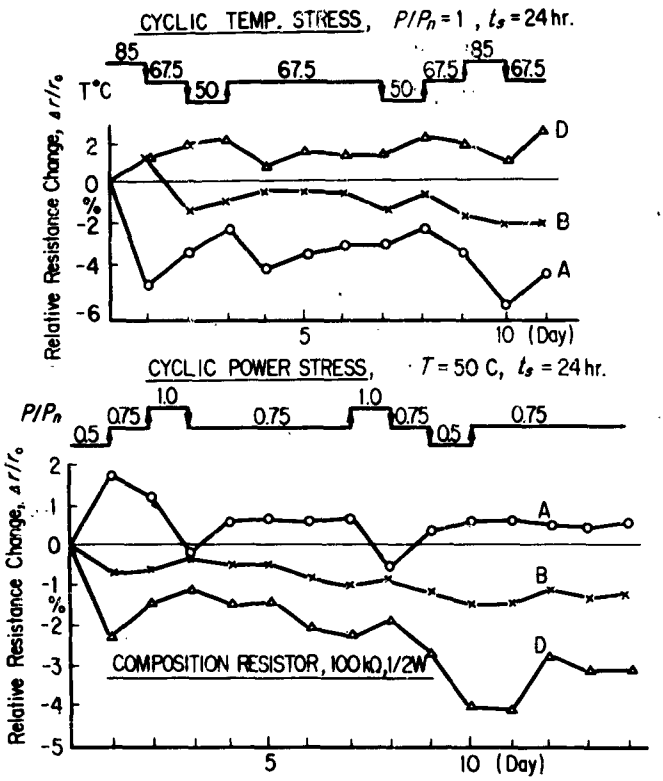


Figure 7

Resistance Change Caused by Cyclic Stress

5. Some Defects of Conventional Step Stress Method

The step stress method is quite useful for the purpose of preliminary survey of life test design, marginal check, screening, and comparison of life quality of different lots. However, there are some defects such as testing time, stress range, and accumulation error of hysteresis effects.[18][19]

(1) Total testing time:

The total testing time of step stress method is not so much shorter. If the step stress test is stopped at the nth step of 50% failure, the testing time is nt_s where t_s is the step time interval, $t_1=t_s$. Moreover, we have to repeat the experiment at three different t_s levels in order to assure accurate extrapolation of lnL:stress plot.

Table 2 shows the comparison of testing time required for constant stress and step stress methods assuming $u(t)=\exp(-Kt)$ type degradation of gm. We must note that the total testing time, 5.24×10^3 hr. for step stress method of three t_s levels at FL=80%, is about two times longer than 2.55×10^3 hr. of constant stress method of three stress levels (7,8, and 9 volt). The conventional life test at higher constant stress level sometimes will give rather more accurate information both for life time distribution and degradation type within shorter period.

Constant Stress			Step Stress*		
Stress Level (Volt)	Testing Time ($\times 10^3$hr.) FL=50%	FL=80%	t_s	Testing Time (n) ($\times 10^3$hr.) FL=50%	FL=80%
6	22.2	7.2	0.5	3.8 (8)	2.54(6)
7	6.0	1.9	0.3	2.63(9)	1.86(7)
8	1.6	0.53	0.1	1.04(11)	0.58(9)
9	0.35	0.12			
total testing time	30.2 (6,7,8,9 volt)	9.75	total testing time	6.5	5.25
	8.0 (7,8,9 volt)	2.55			

*Initial Step: 6 volt, stress step $\delta=\Delta 1/T \times 10^3$= 0.0125, activation energy term $B=34 \times 10^3$.

Table 2

(2) Error due to hysteresis effect:

The range of choice of equal time interval t_s is practically limited within a relatively narrow range. If we choose a larger t_s, it is not economical for the purpose of early prediction. And, if we take a shorter t_s, the error caused by hysteresis may increase. The sub-mode of degradation which usually appears in earlier periods at lower stress will also cause estimation errors. Sometimes, the degradation effect of t_s is not significant in comparison with stress effect. In those cases, the estimation of lnL:stress line by the extrapolation of ln t_s(or ln $t_s\delta$):stress plot is quite erroneous (see later examples shown in Figs. 8 and 9).

6. Application of $\sum(t_i/L_i)=1$ Rule for Life Estimation

To avoid cumulative hysteresis error of step stress, it is desirable to adopt an experiment with fewer stress steps. For this purpose, the cumulative degradation rule is effectively applicable both to degradation and catastrophic failure estimation.

The simplest case is a two-level life test in which shorter life L_H at the higher stress is known in advance and it is required to know L_L at the lower stress level. The estimation equation is given by:

$$(t_L/L_L) + (t_H/L_H) = 1 \text{ or } L_L = t_L L_H/(L_H - t_H) \qquad (15)$$

where t_H, t_L are actual stressing times at the higher and the lower stress levels, respectively, and the total testing time until life end is $\sum t_i = t_H + t_L$. We place samples in a lower stress level during t_L and then transfer them to a higher stress level, for example. Using the a priori knowledge of L_H and observed t_H value, L_L is easily obtained from the above equation. When t_L (or t_H) is fixed, the estimation accuracy of L_L is checked by taking account of sample variation of t_H (or t_L) and certainty of L_H value in equation (15). For the case of $(t_H/L_H)=(t_L/L_L)$, the total testing time is equal to $(L_H+L_L)/2$. If t_H/L_H is chosen to be greater than 0.5, testing time is much more reduced but accuracy decreases. An example of this method for gm life is illustrated in Fig. 8. This principle is, of course, applicable to the experiment of more than two stress levels.

The advantage of this method is saving of samples and testing time. We can make economical use of samples and data of other truncated life tests halfway through the tests to get further information by putting those samples into the intended stress condition. Moreover, this rule is also successfully applicable to other problems such as: (1) prediction of component life (electronic and mechanical) for preliminary design, and (2) checking the equation to make sure whether the linear accumulation rule holds or not.

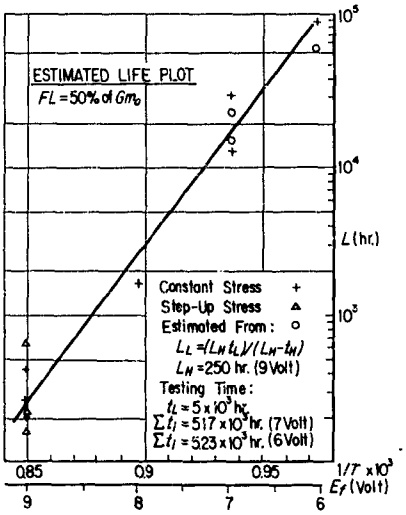


Figure 8

Estimated Median Life: Stress Plot by Various Methods

7. Applications of Cumulative Degradation Principle

7.1 Cumulative Degradation: Stress Plot Method[2)11)20)]

The conventional step stress method needs at least two or three step stress experiments at different t_s levels. These methods use only information of mean stress value and t_s value to get lnL:stress plot. In order to avoid hysteresis error of the step stress method and further shorten the testing time, we can make use of the knowledge of the degradation pattern $f(u)=Kt$, which is ascertained by a constant stress experiment in advance. Knowing the degradation pattern, it is possible to evaluate the dependence of K on stress or the slope of lnL:stress plot by a one-step stress process experiment. This method is useful for drift failure evaluation.

In brief, procedures of the method are as follows:

(1) Conduct one-step stress experiment of time interval t_s,

(2) Plot logarithm of net parameter change $\ln u_n$ or logarithm of parameter change increment $\ln \Delta u_i$ for each step against stress,

(3) Combining the knowledge of reaction pattern $f(u)=Kt$, that is, $\Delta \ln u / \Delta \ln t$ and the slope of $\ln u$:stress plot; $\Delta \ln u / \Delta$stress, find the slope of lnK:stress; $/\Delta \ln K / \Delta$ stress$/ = /\Delta \ln t / \Delta$stress$/$,

(4) F... lnL:stress by drawing a straight line of the obtained slope passing through the known $\ln t_s$ or lnL. If experiments are repeated at other t_s levels, estimation accuracy will be improved.

Fig. 9 shows an example of composition resistor for which resistance change $(\Delta r/r_0)=u$ follows approximately a diffusion model or power model with exponent $\beta=0.5$ (exact experimental value, 0.46 on average): $u=(Kt)^{\beta}$ or $(\Delta r/r_0)= Kt$. The obtained slope of $\ln \Sigma(\Delta r_i/r_0)=\ln(t_s K_n)^{\beta}$ n versus 1/T plot is $10\sim11\times10^3$. Then, the estimated activation energy term B is $(10\sim11\times10^3)/\beta$, that is, $20\sim22\times10^3$ which coincides with $B=20\sim23\times10^3$ obtained independently from constant stress life tests.

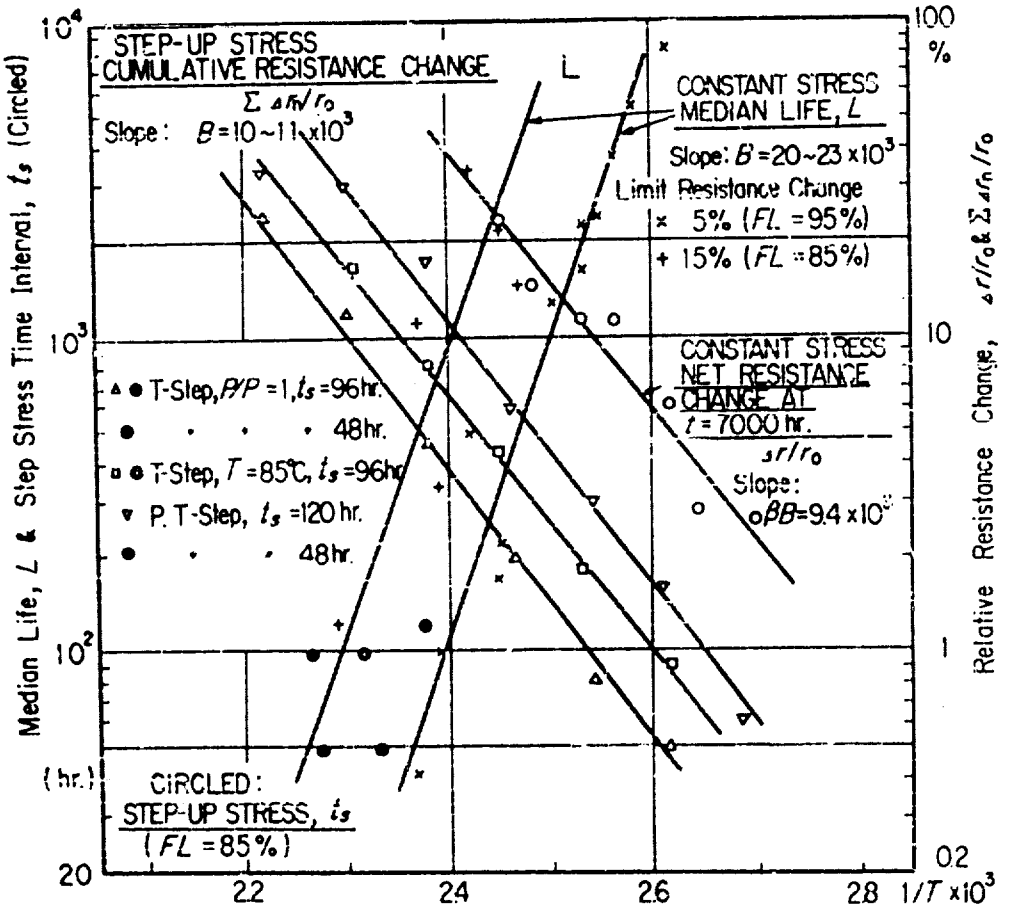


Figure 9

Cumulative Drift and Life versus 1/T Plot

When hysteresis due to a sub-degradation mode has occurred, it is desirable to omit the inaccurate data from earlier steps and use only the reliable information of higher stress. As long as we use the cumulative data, slight variations and errors in the plot do not greatly affect the slope estimation. Further, to avoid uncertainty caused by adopting an unreliable t_s value, use ordinary median life of constant stress.

7.2 Cumulative Failure: Stress Plot Method

A. Introducing Accumulation Rule into Failure Distribution

For simplicity, it is assumed here that component failure follows the Weibull distribution with shape parameter m and scale parameter t_o.

$$R(t)=\exp(-t^m/t_o)=\exp(-\ln 2(t/L)^m) \tag{16}$$

where median life, $L=(t_o \ln 2)^{1/m}$.

If the life is defined as mean life $L=t_o^{1/m}\Gamma(1+1/m)$, $(\ln 2)^{1/m}$ of equation (16) will be replaced by $\Gamma(1+1/m)$. Now, introducing the accumulation principle into equation (16), we get

$$R_n=\exp(-\ln 2\ (\Sigma(t_i/L_i)\ )^m) \tag{17}$$

$$R_n=\exp(-(\ (\ln 2)^{1/m}/f(FL)\ x\Sigma K_i t_i\ )^m)$$

$$=\exp(-(\ \textcircled{H}\ (m,FL)\ x\Sigma K_i t_i)^m) \tag{18}$$

where $\textcircled{H}\ (m,FL)=(\ln 2)^{1/m}/f(FL)$

When it is a random failure, $\textcircled{H}\Sigma K_i t_i$ would be replaced by the sum of (failure rate x time), $\Sigma\lambda_i t_i$. For the step stress, putting $t_i=t_s$:

$$R_n=\exp(-(t_s\ \textcircled{\theta}\ \Sigma^n K_i)^m) \tag{19}$$

$$\ln\ln(1/R_n)=m\ln t_s\ \gamma_n\ \textcircled{\Phi}+m\ln K_n \tag{20}$$

where γ_n is a correction factor.

Two typical expressions of γ_n are given as follows:

(1) Thermal step stress with stress step $\delta=\Delta(1/T)$.

The reaction rate is approximately expressed by Arrhenius type, $K=\lambda_T\exp(-B/T)$ for thermal stress. The correction term is given by $\gamma_{T\cdot n}=(1-\exp^{-n\delta})/(1-\exp^{-B\delta})$.[18] As this value asymptotically approaches to $\gamma_{T\cdot\infty}=1/(1-\exp^{-B\delta})$, we can consider it almost constant except initial steps. When $B\delta\geq 3$, it is almost unity.

(2) Voltage Step Stress with an Equal Ratio $v=(V_{i-1}/V_i)$.

The condenser life is accelerated by increasing applied direct current voltage V, and it is expressed by $(L_r/L)=(V/V_r)^k$. If V is increased by an equal voltage increment ΔV, ΣK_i or ΣV_i^k takes rather complex form.[7] Then, we choose each step voltage so that the voltage ratio of successive steps is a constant, $v=(V_{i-1}/V_i)$. For this case, $\gamma_{V\cdot n}$ takes a similar form as $\gamma_{T\cdot n}$.

$$\sum_{1}^{n} K_i = V_n^k (1+v^k+v^{2k}+\cdots\cdots+v^{(n-1)k}) = V_n^k \, \gamma_{V \cdot n} \qquad (21)$$

where $\gamma_{V \cdot n} = (1-v^{nk})/(1-v^k)$

When we choose $v=0.91=1/1.1$ and assuming the exponent $k=5$, we get $\gamma_{V \cdot \infty} = 1/(1-v^k) = 2.29$. The difference of γ_{V4} and $\gamma_{V \cdot \infty}$ is only about 10%.

For thermal stress, from equation (20), $K = \Lambda_T \exp(-B/T)$ and $\gamma_{T \cdot n}$, we have

$$\ln\ln(1/R_n) = m\ln t_s \oplus \gamma_{T \cdot n} \, \Lambda_T \, - mB/T_n \qquad (22)$$

The slope of $\ln\ln(1/R_n)$: $1/T_n$ plot equals to mB.

For voltage stress, from equation (20), $K = \Lambda_V V^k$ and $\gamma_{V \cdot n}$, we have

$$\ln\ln(1/R_n) = m\ln t_s \oplus \gamma_{V \cdot n} \, \Lambda_V + mk\ln V \qquad (23)$$

The slope of $\ln\ln(1/R_n)$: $\ln V$ plot is mk.

B. Estimation Procedures of lnL: Stress Plot

The principle of this method is quite similar to the cumulative drift method for drift failure life estimation mentioned in section 7.1. Analogical comparison of those methods is illustrated in Fig. 10. Instead of using the knowledge of time dependency of the drift pattern, we can now make use of the knowledge of failure time distribution shape. Using the information of distribution shape, such as Weibull shape parameter $m = \Delta\ln\ln(1/R)/\Delta\ln t$ and the obtained slope of $\ln\ln(1/R)$:stress plot from one step stress experiment, that is, $\Delta\ln\ln(1/R)/\Delta$stress, we can find the slope of lnL:stress, $\Delta\ln t/\Delta$stress or $\Delta\ln K/\Delta$ stress.

Assuming the failure follows Weibull distribution with known shape parameter m, estimation procedures to get lnL:stress plot are as follows:

(1) Conduct one step stress experiment. The equal step interval for thermal stress is $\delta = \Delta(1/T)$ and, for voltage stress, the equal voltage ratio is $v = (V_{i-1}/V_i)$.

(2) Plot $\ln\ln(1/R_n)$: $1/T_n$ for thermal stress and also plot $\ln\ln(1/R_n)$: $\ln V_n$ for voltage stress. We can make use of the ordinary Weibull chart by replacing t or lnt of the abscissa by 1/t or lnV.

(3) Read the slope of plot, that is, $mB = \Delta\ln\ln(1/R)/\Delta(1/T)$ for thermal stress and $mk = \Delta\ln\ln(1/R)/\Delta\ln V$ for voltage stress. Then, we get the activation energy term $B = \Delta\ln t/\Delta(1/T)$ and voltage acceleration exponent $k = \Delta\ln t/\Delta\ln V$ from known m.

(4) Draw a line with the slope B passing through $\ln t_0 \gamma_{T\cdot n}$ on the lnt:1/T diagram for thermal stress. This line is the estimated lnL:1/T plot. For voltage stress, draw the line with slope k through $\ln t_s \gamma_{V\cdot n}$ on the lnt:lnV diagram to get lnL:lnV plot. If the t_s value is unreliable, use the L value known from past experiments or obtained from a higher stress level life test.

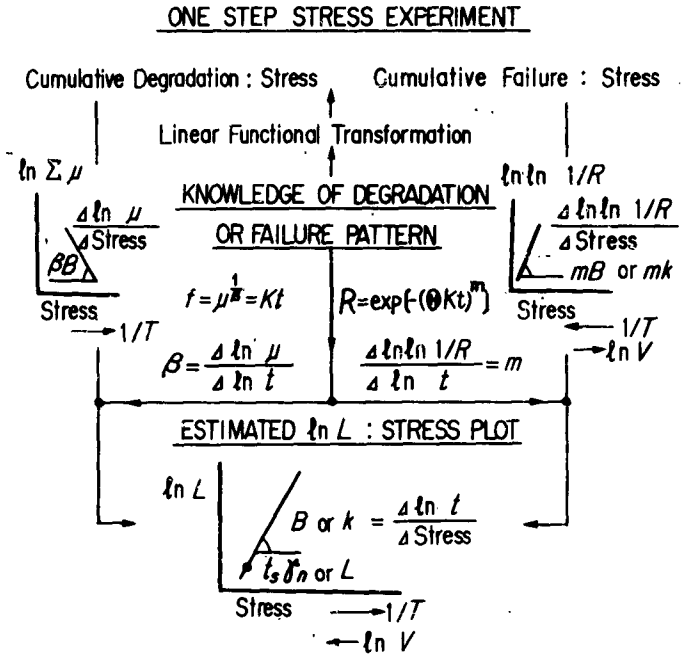


Figure 10

Comparison of Cumulative Degradation and Failure Methods

C. Actual Examples

(1) Thermal step stress experiment for composition resistor.

Figure 11A shows $\ln\ln(1/R_n):1/T_n$ plot for drift failure and short failure of composition resistor. The allowable resistance changes for drift failure are 10% and 15% (mark o and x in the figure). The Weibull shape parameter m is previously known to be about 1.45. The slope of $\ln\ln(1/R_n):1/T_n$ is about 13 15 x 10^3, then B is estimated as 9 10.3 x 10^3. This is quite near to the value B=10 x 10^3 (0.86ev) obtained from usual constant stress life tests.

(2) Voltage step stress experiment for mylar condenser.

The shape parameter m has been obtained as m=0.75 on average (Figure 11B). The slope of $\ln\ln(1/R_n)$:lnV plot is found to be

mk=3.8. The estimated k value is 3.8/0.75=5.2 which shows good coincidence with k=5.3 obtained from constant stress experiments.

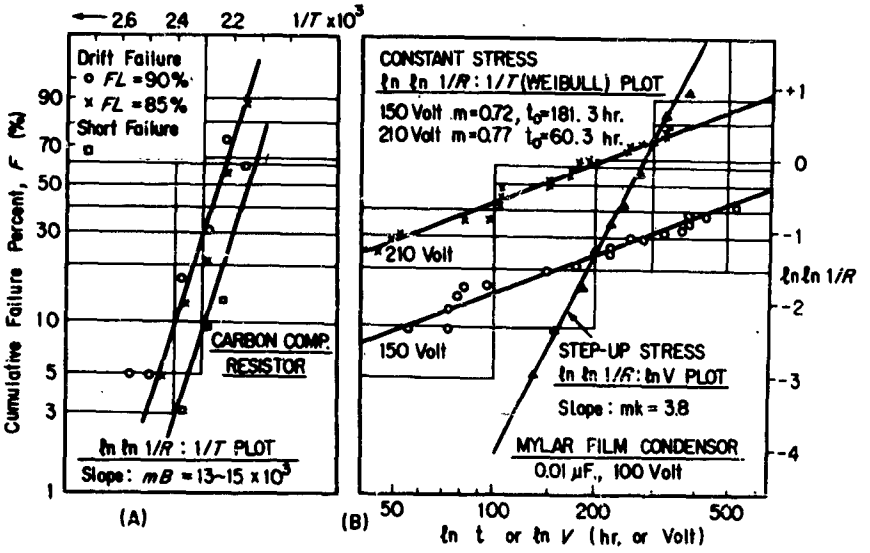


Figure 11

Examples of Cumulative Failure: Stress Plot Method

D. Additional Comments to Applications of the Method

Regarding applications of this cumulative failure method, some remarks to be emphasized are as follows:

(1) The rapid check of life characteristic variation is possible by the modification of one step stress experiment. First, put the samples under constant stress specified and plot conventional Weibull plot, i.e., $\ln\ln(1/R)$:$\ln t$ plot to make sure of shape parameter m and after a certain percentage of failure (say, 50%) has occurred, transfer the remaining samples to a step-up stress experiment, successively, to check mB or mk by observing $\ln\ln(1/R)$: stress plot. We can check shape parameter m and mB or mk on the same Weibull paper through a single experiment.

(2) The failure distribution to which this principle is applicable is not only limited to the Weibull distribution with a constant shape parameter. When the Weibull shape parameter m is not constant but the plot transfers in parallel by the amount $\Delta\ln t$, keeping its shape unchanged as stress varies, the life acceleration factor A_L is related to $\Delta\ln t$ by $\ln A_L = \Delta\ln t$. Thus, as far as acceleration of failure pattern is possible, this method is applicable. If the Weibull plot is not straight, the obtained $\ln\ln(1/R)$:stress plot will also vary nonlinearly with B or k times slope.

In the case of lognormal distribution, for example: $-dR/dt=f(t)=(\sqrt{2\pi}\sigma)^{-1}\exp(-\ln(t/L)^2/2\sigma^2)$, $E(t)=\exp(\ln L+\sigma^2/2)=L\alpha$, $V(t)=\exp(\ln L^2+\sigma^2)(\exp\sigma^2-1)=(L\alpha)^2(\alpha^2-1)$, the above accumulation rule could be introduced into t/L of the exponential term as long as a shape factor $\sigma^2=\ln\alpha^2$ is kept constant for stress acceleration.

(3) It is worthwhile to note that the equation similar to $(L_T/L)=(V/V_T)^k$ for condenser life mentioned above is often found in other fields.[21][22] This equation is generally expressed by:

$$(L_T/L)=(S/S_T)^k \qquad (24)$$

where S is stress

The endurance life of ball bearings, rupture strength of steel, and filament life of light bulbs follow this relationship, hence, the method explained here is applicable not only to condensers but also to these materials.

8. Conclusion

A. The degradation accumulation principle and stress-time transformation of reaction theory offer us some useful estimation tools of component life.

(1) The difference between failure rate acceleration factor and life acceleration factor is pointed out. Given the Weibull shape parameter m, two factors are related by $A_\lambda=A_L^m$.

(2) The generalized cumulative degradation rule, $\sum_i(t_i/L_i)=1$, is introduced. Using this rule, estimation and prediction of component life under successive different stress levels are possible both for drift and catastrophic failure. The advantage of the method is economic evaluation of component life by combining its past knowledge of life and stress history.

(3) Assuming the knowledge of drift pattern or shape of failure time distribution, lnL:stress plot is estimated from the information of accumulated drift or failure versus stress plot obtained by one step stress experiment.

9. Acknowledgement

The author wishes to express his appreciation for the assistance provided by Mr. T. Harada and Mr. S. Shiota of the laboratory.

10. References

1. Shiomi, H., "Fundamental of Thermal Stress Aging Methods and Degradation Failure Model," Bul, Electrotech. Lab., Vol. 27, No. 2, pp. 81-105, 1963.
2. Shiomi, H., "Cumulative Law of Degradation and Generalized Miner's Equation for Component Life Estimation," Bul. Electrotech. Lab., Vol. 27, No. 12, pp. 901-910, 1963.
3. Dakin, T. W., Electro-technology, Dec., pp. 123-130, 1960.
4. Eyring, H. E., "The Theory of Rate Process," McGraw-Hill, 1941.
5. Thomas, R. E., and Gorton, H. C., "Research Toward a Physics of Aging of Electronic Component Parts," 2nd Symp. on Physics of Failure, 1963.
6. Kimmel, J., "Accelerated Life Testing of Paper Dielectric Capacitors," 4th National Symp. on R&QC, pp. 120-134, 1958.
7. Endicott, H. S., Hatch, B. D., and Sohmer, R. G., "Application of the Eyring Model to Capacitor Aging Data," IEEE Tr., CP-12, No. 1, pp. 34-41, 1965.
8. Miner, M. A., "Cumulative Damage in Fatigue," Proc. ASME, Vol. 67, pp. A-159-164, 1945.
9. Gatts, R. R., "Cumulative Fatigue Damage with Progressive Loading," Paper No. 62-WA-292, ASME, 1963.
10. Freudental, A. M., and Heller, R.A., "Accumulation of Fatigue Damage," Fatigue in Aircraft Structure, Academic Press, 1956.
11. Shiomi, H., "Life Expectancy and Thermal Stress Aging Methods for u(t)=exp(-Kt) Type Decay," Bul. Electrotech. Lab., Vol. 27, No. 5, pp. 361-380, 1963.
12. Wahl, A. J., and Kleimack, J. J., "Factors Affecting Reliability of Alloy Junction Transistors," Proc. IRE., pp. 494-502, April 1956.
13. Partridge, J., and Borofsky, A. J., Symp. on Cleaning and Materials Processing for Electronics and Space Apparatus, ASTM, 4th Pacific Area National Meeting, October 1963.
14. Partridge, J., "On the Extrapolation of Accelerated Stress Conditions to Normal Stress Conditions of Ge. Transistor," 2nd Symp. on Physics of Failure, 1963.
15. Tada, H. Y., Aspects Affecting the Reliability of a Carbon Composition Resistor," IEEE Tr., CP, pp. 67-79, June 1963.
16. Porter, D. C., "Failure Analysis of Electronic Parts," IEEE Tr., AS, Vol. 2, No. 2, pp. 328-337, April 1964.
17. Chadwick, G. F., "The Effect of Moisture on Molded Composition Resistor," IEEE Tr., CP-10, No. 3, pp. 113-115, September 1963.
18. Howard, B. T., and Dodson, G. A., "High Stress Aging to Failure of Semiconductor Devices," 7th National Symp. on R&QC, 1961.
19. Go, H. T., "Improving Semiconductor Reliability," Electronic Ind., pp. 110-113, February 1963.
20. Shiomi, H., "Degradation Failure of Logarithmic and Power Decay and Life Estimation," Bul. Electrotech. Lab., Vol. 27, No. 12, pp. 881-900, 1963.
21. Haviland, R. P., "Engineering Reliability and Long Life Design," pp. 33, D. Van Nostrand, 1964.
22. Robinson, E. L., "Effect of Temperature Variation on the Long-time Rupture Strength of Steels," Trans., ASME, pp. 777-781, July 1952.

THE APPLICATION OF FAILURE ANALYSIS IN PROCURING AND SCREENING OF INTEGRATED CIRCUITS

Jayne Partridge, Eldon C. Hall

and

L. David Hanley

ABSTRACT

The procedure for the testing, screening, and lot rejection of integrated circuits for the Apollo Guidance and Navigation computer is described. The procedure, based on a knowledge of failure modes, failure mechanisms and contributing causes to failures in the manufacturing of devices, attempts to increase the reliability of integrated circuits. This is accomplished by screening and analyzing weak devices and using the generated data to quantitatively assess the lot for acceptance, rework or rejection. The technique, which is primarily aimed toward high-usage high-volume devices, was developed after extensive testing of many tens of thousands of integrated circuits. The process documents included in the appendix contain stress test procedures, classification of failure modes, numerical rejection limits per class of failure modes, internal visual rejection criteria, and leak test procedures.

To emphasize the need for the described technique, data is presented showing variations among vendors and variation among procurement lots shipped from a single vendor. The contributing factors to the variations are discussed.

A discussion of the evolution of the process documents is presented. The ultimate goal of the documents is the elimination or minimization of detected failure modes. Failure studies have shown that some failure modes are screenable with high confidence whereas attempts to screen other types of failure modes merely decrease the life of the device. In the latter case, the detection during short term stressing of devices which exhibit long time dependent failure modes is a low probability event. After one-hundred-percent nondestructive testing, sample destructive testing, failure analysis and failure mode grouping, the classes of failure modes in a lot are then weighted in accordance with screenability and detectability. Failure of a lot to fall within the acceptable limits will instigate action as to whether the lot will be rescreened, resubmitted to tighter acceptable limits, or whether a portion of the lot or the entire lot will be rejected. The decision for lot or sublot rejection is based on the traceability of the nonscreenable failure modes to a critical manufacturing process. The approach presents a continuous monitoring procedure for qualification of parts and vendors, and creates an incentive on the part of the vendor to eliminate causes of failures.

THE APPLICATION OF FAILURE ANALYSIS IN PROCURING AND SCREENING OF INTEGRATED CIRCUITS

Jayne Partridge, Eldon C. Hall

and

L. David Hanley

Instrumentation Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

INTRODUCTION

Small-sample stress testing as applied to device procurement has dubious application for systems with increased high reliability goals. As a result, one-hundred-percent nondestructive stress testing has become fashionable. Unfortunately, even the testing and stressing of entire lots, by itself, cannot assure the elimination of nonscreenable failures. In an attempt to fill in this gap, an approach will be presented, using semiconductor integrated circuits as an example, whereby the frequency of field failures can be decreased beyond the point presently possible through one-hundred-percent testing alone. This approach is based on the knowledge of failure modes, failure mechanisms, and contributing causes during device manufacturing, all of which can be applied to the screening and acceptance criteria of procurement lots. The technique was developed after extensive testing, data analysis, and failure analysis of integrated circuits.

There were two major factors which aided the study and development of the screening and lot acceptance procedures. One was the decision to use only one Nor gate, as shown in Fig. 1, for all logic functions in the Apollo Guidance and Navigation computers. This resulted in high volume procurement, an absolute necessity for establishing proven low failure rates of any new device in a short period of time. The second was the choice of an extremely simple circuit which aids an effective screening process. The accessibility of the circuit elements enables quick detection and diagnosing of insidious failures without extensive probing as required with more complicated circuits. For some failures, as for example those which are induced by surface conditions, it is desirable to be able to study the characteristics of the integrated circuit components without opening the package.

It became immediately obvious that small-sample stress testing could not guarantee that each purchased lot would meet the Apollo

integrated circuit failure rate requirements. The MIL-S-19500D statistical sampling procedure was both not applicable nor practical. Furthermore, as long as all failure modes were not completely screenable, one-hundred-percent screening alone was not sufficient to attain the required high reliability goals. A study of the various failure modes of integrated circuits created the dilemma whereby some of the failure modes were easily screened by standard screening techniques and others only occasionally detected. No assurance could be made with any reasonable confidence that the devices with these troublesome defects had been removed from the lot. To overcome this problem, lot acceptance criteria were established which would identify with high confidence those lots in which insidious failure modes were not prevalent and screening had been adequate. Providing an effective failure mode detection system, the procedure for lot acceptance is based on one-hundred-percent nondestructive tests and sample destructive testing. All the failures generated from the testing are completely analyzed. The failure modes are then classified by groups and compared to the acceptance criteria. It must be emphasized that the lot is accepted or rejected not only because of the number of failures but also on whether the failure modes generated were nonscreenable or insidious and long-time dependent.

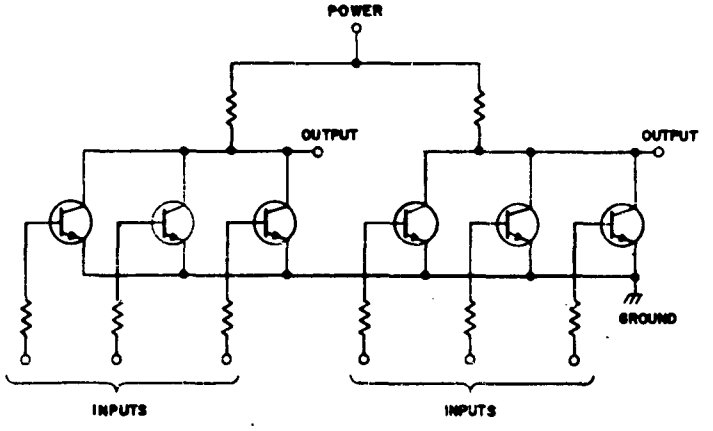


Figure 1

Schematic of the silicon monolithic dual three input nor gate.

VENDOR SELECTION & FLIGHT QUALIFICATION PROCEDURE

To assist the understanding of the lot acceptance procedures, a general discussion of the semiconductor part vendor selection and flight qualification procedures will be given as performed for the Apollo Guidance and Navigation computer.

The process begins with an assessment of the vendor's ability to supply devices, the institution of component standardization in designs, and the preliminary study of device failure modes. A block diagram of this preliminary evaluation which precludes any production procurement is given in Fig. 2. The qualification procurements which supply parts for the qualification testing and engineering evaluations established the manufacturer's device processing. One of the indirect results of the initial procurements is the early detection of new failure modes. The conclusions of the failure analyses are then fed back to the manufacturer who in turn attempts corrective action. This cyclic procedure is continued until the most obvious problems have been eliminated. Additionally, the early detected failure modes coupled with past experience are utilized to design the qualification testing.

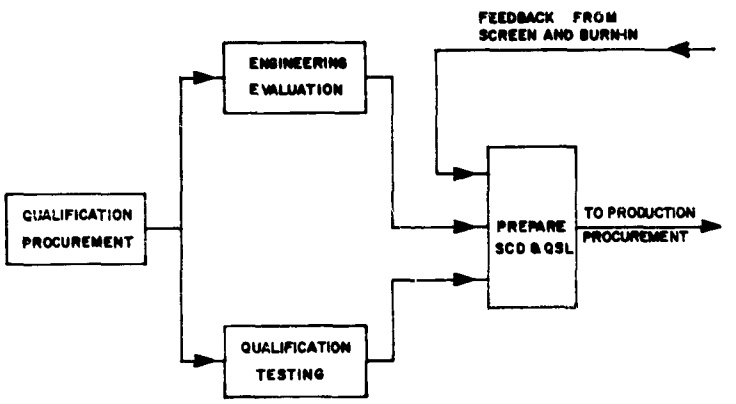


Figure 2

Block diagram of the vendor selection procedure.

The formalized qualification testing begins when the vendors have supplied devices representative of their finalized manufacturing process. It is extremely important that all qualification and engineering testing be performed on devices fabricated from the identical process used to supply computer production devices. The qualification tests subject the devices from various vendors to the extremes of and, to a limited extent, beyond usage conditions in an attempt to detect failure modes which could occur in normal applications.

The engineering evaluations are performed simultaneously with the qualification procedures to determine device speed, fanout capabilities, noise immunity, and operating temperature range. From this evaluation, the optimum computer design is developed. It is at this time that tests are conducted to determine the electrical parameters which will insure proper device operation in every usage mode and to establish the logical design rules for the computer.

The qualification and engineering evaluations determine those vendors who are capable of supplying the semiconductor part and who do not exhibit any gross reliability problems. The qualification tests alone are insufficient to determine the ability of a vendor to control his process, but large-volume production procurement data fed back from screen and burn-in supplies extensive vendor history.

Utilizing the data generated during the engineering evaluations and qualification tests, the specification control document (SCD) is prepared. The SCD is the document to which production parts are bought. Based on the qualification by vendors, the qualified suppliers list (QSL) is formed which specifies the vendors from whom the production parts shall be procured.

Once the SCD and QSL have been released, production procurement may begin. Figure 3 pictures the general flow of parts and data as required for flight qualification. The devices procured by lots proceed through the screen and burn-in (S&BI) test sequence.

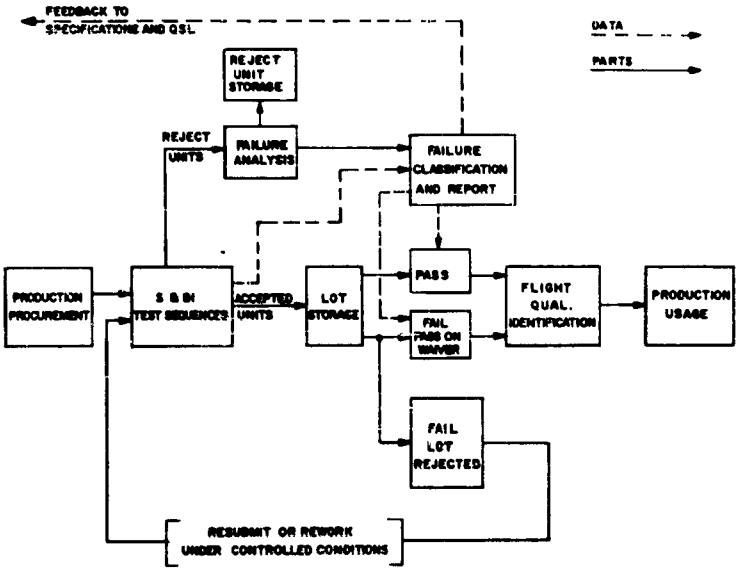


Figure 3

Block diagram of the flight qualification procedure.

Upon completion of screen and burn-in the lot is stored until failure analysis is completed. All failed units are catalogued, analyzed,

and classified to complete the lot assessment, followed by a written report. If the lot passed, all the devices that passed all tests can be identified as a flight qualified part with a new part number and sent for production usage. Only the semiconductor part with the flight qualification part number can be used in flight qualified computer assemblies. From failure analysis, rejected parts proceed to reject storage where they will be available for future study if required. In the event that the lot failed because of circumstances not completely defined through failure classification, the lot can be flight qualified by waiver. The waiver must be authorized by NASA and will accompany the computer. In certain limited cases, parts from a failed lot may be resubmitted for rescreening.

The accumulated data from the screen and burn-in procedure and failure analysis are utilized to further evaluate the vendor production capability and his device quality and reliability. This in turn affects a vendor's continued status as a qualified supplier.

LOT ACCEPTANCE SPECIFICATIONS

The specifications which control the implementation of the process described in Fig. 3 for the dual Nor gate, are given in the appendix. These documents will be briefly described to summarize the salient features.

A. ND 1002248

The Apollo Guidance and Navigation Specification, ND 1002-248, is the central document on which each procured lot qualification is based. This document specifies the procedures required for lot acceptance resulting in flight qualified parts. In particular, ND 1002248, specifies the details of:

1. The operational and environmental stress test procedures and sequence commonly referred to as the screen and burn-in procedure. The screen and burn-in procedure was designed to detect failure modes which could occur during the normal stress and environmental application of the device.

2. The electrical parameter tests to be performed during the screen and burn-in procedure. The tests as defined were determined during the engineering evaluation and were chosen to detect failures and assure proper computer operation.

3. Definitions of failures. Failures have been defined as catastrophic, several categories of non-catastrophic, induced, and inspection failures.

4. Allocation of failures. The conditions are defined for removal from the screen and burn-in procedure of failures which are to be forwarded to failure analysis.

5. Classes of failure modes. Failure modes are classified according to screenability and detectability. This

classification will be discussed in detail in a later section.

6. Maximum acceptable number of failures per class of failure mode for all 100% electrical parameter test stations.

7. Maximum acceptable number of failures for nonelectrical tests and all sample electrical parameter tests.

8. The report required for each flight qualified lot. The report must contain the complete history of the lot with the specific data and analysis required for flight qualification.

9. Data and failed parts storage. In order to assure traceability and future analysis should field failures occur, the conditions of data and failed parts storage are given.

10. Contractual requirements to implement lot qualification.

B. ND 1002257

The Apollo Guidance and Navigation specification, ND 1002-257, defines the rejection criteria for internal visual inspection of silicon monolithic integrated circuits. This specification was included in the appendix because of the affect of the criteria on lot acceptance. ND 1002257 serves a dual purpose in that it is applied by the device manufacturer during a one-hundred-percent preseal inspection for removal of defective parts, and by the customer on a sample basis as a destructive test for lot acceptance. Some of the problems in a lot may only be detected by destructive internal visual inspection. Certain failure modes can only be observed after the sealed and branded device has been exposed to operational and environmental stresses.

The internal visual inspection criteria were defined after most, if not all, of the failure modes of silicon monolithic integrated circuits were determined. No device is rejected merely on the basis of aesthetics. Devices are rejected only when a fault which contributes to a known, potential failure may be visually observed. The rejection criteria of ND 1002257 do not attempt to reject all of the visually observed faults contributing to failure, because of the difficulty of precisely or quantitatively defining faults which are subject to individual interpretation. It has been our approach that ultimate reliability will be improved by rejecting to major, easily observed defects rather than by rejecting to a long complicated list of qualitatively defined defects.

C. ND 1002246

The Apollo Guidance and Navigation Specification, ND 1002-246, which was written after a series of correlation tests, states the procedures for leak testing of flat packages. It was determined

that the standard fine and gross leak tests are insufficient for detecting the entire range of leakers for all flat package designs. This specification was included because lack of adequate hermeticity abets some of the insidious time-dependent failure modes.

FAILURE MODE CLASSIFICATION AND ACCEPTANCE NUMBERS

The essence of the lot acceptance procedures as specified in ND 1002248 is contained in the classification of failure modes and associated acceptance numbers. Although any failure which occurs in a critically highly reliable system is undesirable, failure modes may be grouped in accordance with available methods of elimination. A previous report[1] lists the failure modes detected in silicon monolithic integrated circuits along with some contributing causes and stress dependencies. An updated list of the detected failure modes is given in ND 1002248, section 4.2.2. These failure modes which were generated after one-hundred-percent testing may be classified in the following manner:

1. Special Cases of Noncatastrophic Failures (Group 0)

These include devices which do not meet the electrical specifications at incoming electrical tests, devices which drift out of the electrical specification limits during stress testing but do not exceed a given percentage drift, and devices which never exceed the electrical specification limits during stress testing but which exceed a given percentage drift. These failures do not necessarily impede computer operation.

2. Screenable Failure Modes (Group I)

The contributing causes and stress dependencies of these failure modes are sufficiently well known so that the failure modes are screenable to a high confidence through electrical and stress testing.

3. Nonscreenable but Detectable Failure Modes (Group II)

Failure modes which are classified in this manner usually exhibit intermittency, serious surface problems, or require severe stressing for screenability. These failure modes are not screenable at normal, nondestructive stress levels, but are generally detected in finite amounts at normal, nondestructive stress levels if they are insidious within the lot.

4. Nonscreenable Difficult-to-Detect Failure Modes (Group III)

These failure modes are generally long accumulative time dependent at nondestructive stress levels. As a result, detection of such failure modes during short-term nondestructive testing becomes a low probability event. Also placed in this category are failure modes which are nonelectrically detectable, as for example, a chip detached from the header where the header does not provide electrical contact and the bonds or lead wires are not broken.

5. New Failure Modes or Failure of Unknown Causes (Group IV)

If such failure modes occur, studies must be performed to determine the screenability and the destructive effects of the various stress tests.

The failure mode classifications described above were listed by groups in order of decreasing screenability and detectability. Table I of ND 1002248, section 4.2.3, gives the maximum allowable percent failures for each failure mode classification for each lot processed through screen and burn-in. As the degree of screenability and detectability decreases, the failure mode classification is more heavily penalized. The Group II and III failure modes are more heavily penalized after the incoming electrical test, because it is of more concern when these devices fail after they were known to be good. Note that a maximum limit is still placed on the screenable failure modes. This is done for two reasons. First, one can never be one-hundred-percent confident that all potential failures have been eliminated. Second, the limit sets a guard against a multiplicity of failure modes which is indicative of poor workmanship and sloppy control. The limits set on the Group 0 failure modes are, in general, a guard against careless testing.

The percentages of Table I were based on a screen and burn-in study of over 200,000 Nor gates. Lots which have not passed the limits of Table I have exhibited reliability problems predicted by the results of screen and burn-in and failure analysis.

Table II of ND 1002248, gives the limits of the leak tests and sample tests of the screen and burn-in procedure. Theoretically, package leaks are screenable, but the limits guard against poorly executed leak test procedures. All the sample tests performed, with the exception of the emitter-base back bias and physical dimensions tests, are considered destructive. The additional handling required by the shock and vibrations testing have induced failures in the flat package. The emitter-base back bias test which is performed to detect surface instabilities will indicate the surface problems of the entire lot. Although the limit set on the internal visual inspection appears loose, the limit reflects problems in interpretation and subjectivity.

METHODS OF LOT REJECTION AND ACCEPTANCE

Because the provisions of ND 1002248 do not explicitly define, at this time, the various methods of possible lot rejection, a discussion will be presented here.

1. Lot Rejection

Any large lot of a semiconductor device fabricated from a continuously operating production line does not necessarily consist of homogeneous product and is certainly a function of start-to-finish yield. However, there are some failure modes which when detected are known to be prevalent in the entire lot. An example of such a failure mode is interconnect corrosion which is caused primarily by the presence of excess oxygen and moisture although thinning of metalization and heat are aggravants. The primary causes of this failure mode are usually traced to improper wash and dry techniques

of the unsealed assembled device, device storage and device sealing. Another example requiring lot rejection are failure modes due to surface instabilities. Surface problems depend on the variabilities of most of semiconductor processing and, if not eliminated, are known to be prevalent in many lots of semiconductor devices. Lot rejection is necessitated when failure modes whose causes originate in the production line affect the entire lot.

2. Sublot Rejection

Since each lot of integrated circuit Nor gates is composed of many diffusion lots, sublot rejection is possible if failure modes are related to the diffusion sublots, and traceability is maintained after dicing and subsequent assembly. It is the intent of ND 1002248 to reject devices with the possibility of insidious failure modes and accept devices with excellent reliability potential. The assumption that sublots are handled nearly identically implies that if one unit exhibits insidious or long-time dependent failure mode, many undetected devices will contain that failure potential. Failure modes whose causes originate in the variations of diffusion, oxidation, metalization, and etching may not be insidious in all diffusion lots. Examples of such failure modes are contained in Section 4.2.2 of ND 1002248 and are a1, b3, b8, c2, and c4.

3. Rescreen and Rework

Some of the listed failure modes are amenable to rescreen by the stressing which triggers or selects out failures. Most of these failure modes are generally in the Group 0 and Group I category. If this resubmittal procedure is applied, tighter acceptable limits should be met.

There are also certain applicable screen procedures which are not part of the screen and burn-in procedure because of their lack of universal application, or their effectivity or lack of destructivity have not been proven. For example, x-raying of TO-47 package integrated circuits has shown to be an excellent screen procedure for excess lead length and leads shorting to one another. However, the technique is useless for devices which employ aluminum leads, (aluminum is transparent to x-rays) and the applicability to flat packages has not been proven.

As effective nondestructive screen procedures are developed, use of such procedures on an individual or universal basis may be instituted.

4. Waivers

Since no specification is perfect, unforeseen contingencies of the specification may cause lots to fail for reasons other than device faults and poor stressing procedures. In these events, waivers become necessary.

VARIATION AMONG MANUFACTURERS AND LOTS

The previous sections have dealt with a description of the system incorporating the lot acceptance procedure by failure modes. Data will

now be presented indicating the need for such an acceptance procedure because of the variations among manufacturers and lots of a single manufacturer.

Table I is a summary of reliability data accumulated up to October 1964 for the single Nor gate in a TO-47 package. This data has been previously discussed in detail.[1] The data is presented here to show the extreme differences in reliability performance among manufacturers. The screen and burn-in procedure is as described in ND 1002248 except that Y_2 centrifuge, emitter-base back bias, vibration, and shock testing were not performed. The electrical failure definitions during screen and burn-in were any inoperable devices or any device exceeding the electrical specifications. The percentages include approximately 0.05 to 0.1% combined induced failures and testing errors. The initial qualification results are also included in Table I where the failure definition was an inoperable device. The extreme differences among the manufacturers is also reflected in the failure modes generated during both the initial qualification and screen and burn-in. For the data in Table I Manufacturer A rarely exhibited the nonscreenable, and/or long-time dependent failure modes while both Manufacturers B and C consistently did. The inoperable failures generated at computer use conditions for Manufacturers B and C were of the nonscreenable, long-time dependent failure modes.

TABLE I

A summary of vendor reliability evaluation.

VENDOR	INITIAL QUALIFICATION % FAILURES	SCREEN & BURN-IN % FAILURES		FAILURE RATES AT USE CONDITIONS 90% CONFIDENCE
		TOTAL *	POST STRESS **	
A	5	1.8	0.3	0.0018%/10^3 hrs (0 FAILURES)
B	26	3.8	1.7	0.3%/10^3 hrs (2 FAILURES)
C	58	5.0	2.5	1.8%/10^3 hrs (26 FAILURES)

* TOTAL = ALL ELECTRICAL AND MECHANICAL FAILURES

** ELECTRICAL FAILURES AFTER INCOMING ELECTRICAL TESTS

It is interesting to note that the same devices used to generate the data of Vendor A of Table I have since exhibited a failure rate of 0.0018%/10^3 hours at 90% confidence as of 30 August 1965 with no operational failures. The same devices of Vendors B and C have not improved their failure rates because additional failures have occurred.

The data of Table I once again points out the fact that there are differences in the quality and reliability of devices produced among different manufacturers. Even though the general technical procedures of designing and building semiconductor devices are well known throughout the industry, the approaches to production control, problem detection and elimination, and process refinement varies widely among manufacturers. It is necessary for a qualified manufacturer not only to minimize the number of failures but also to maintain process control such that a multiplicity of failure modes does not occur, and long-time dependent failure modes do not exist. If it is not the ultimate goal of a manufacturer to design and build reliability into his device, effective improvement can never be realized through device evaluation, stress testing and quality control.

The more subtle differences in quality and reliability may be observed in variations of lots shipped from one manufacturer. The data of Fig. 4 indicates the numerical variations for the single Nor gate in a TO-47 package from one qualified manufacturer. Here, only the inoperable failures are plotted and induced failures and testing errors have been eliminated from the data. These single Nor gates were exposed to the screen and burn-in procedure as described in ND 1002248 except that Y_2 centrifuge, emitter-base back bias, vibration and shock testing were not performed. Each point represents a shipment lot of 2000 to 5000 Nor gates. Figure 4a shows the percent catastrophic failures at the incoming electrical tests. Figure 4b shows the percent catastrophic failures which were generated after stressing with incoming catastrophic failures removed. There are fewer points plotted in Fig. 4b than in Fig. 4a, since some lots not used for flight hardware were not exposed to screen and burn-in.

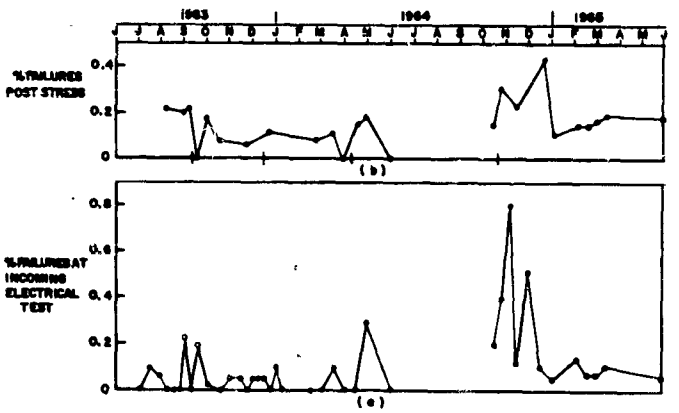


Figure 4

Vendor's performance through screen and burn-in vs time.

Examining only the numerical differences among lots of the same device shipped from the same vendor, Fig. 4 shows that the percent catastrophic failures developed an average stabilized region of 0 to 0.1% at incoming electrical test and 0.1 to 0.15% at electrical tests after stressing. Most high points above these levels have been correlated to events occurring at the manufacturer. The high points prior to October 1963 represent the tail end of the manufacturer's learning curve. The high point at late April 1964 may have occurred due to reallocation of line personnel in anticipation of line shut down. There was no buying of the three input Nor gate between June and October 1964, so that the line producing the integrated circuit was temporarily discontinued. As a result when the production line was reinstated, several lots after October 1964 indicated a new region of instability. At that time rapid feedback to the manufacturer from the customers resulted in subsequent decrease of catastrophic failures during the screen and burn-in procedure.

One aspect of problem areas which the data of Fig. 4 does not indicate are the failure modes generated during screen and burn-in. With very few exceptions, the lots prior to June 1964 exhibited only the less troublesome or screenable failure modes and usually each lot would exhibit only one predominant failure mode. This was not the case for lots shipped after October 1964. These lots exhibited a variety of failure modes including the nonscreenable type, but they were waivered because of some uncertainty of the lot reject levels. It is interesting to note that several of these lots have already exhibited failures after screen and burn-in. Confidence in the reject levels has since been established.

It might appear that the high points of Fig. 4 represent only a small percentage of catastrophic failures. But once again we must be reminded of the very low failure rates that must be achieved. Referring again to Table I it is seen that Manufacturer A developed a total of 0.3% failures after 100% stress tests were performed. Looking only at the catastrophic failures due to device faults, the percentage becomes 0.2 to 0.25%. This sample of devices exhibited one predominant failure mode of a screenable type and subsequently proved that a failure rate of $0.0018/10^3$ hours at 90% confidence is attainable. In accordance with the data of Table I and the failure modes generated, the lots with the larger number of failures of Fig. 4 do not represent a negligible percentage fallout in light of the required reliability goals.

One might ask how a manufacturer can achieve excellence in reliability performance, and then for a short period of time relax his control. The reasons are encompassed in a "state-of-the-art" process where incomplete knowledge of all the variables or insufficient control of all the variables (including the human variable) causes inadvertent changes. As a "state-of-the-art" device approaches excellence in performance, the recipe for producing the device becomes critical.

UPGRADING RELIABILITY THROUGH THE LOT ACCEPTANCE SPECIFICATIONS

The lot acceptance specifications provide a direct and indirect means of upgrading component reliability. In addition to rejecting unacceptable lots, it is readily seen that the lot acceptance specifications present a formal means of continuously monitoring a manufacturer. Other than occasionally rejecting a lot, the procedure provides extensive vendor history with time. If the manufacturer shows a consistent

degrading of performance, he is then eliminated as a qualified source until he shows proven recovery. This is an effective means of maintaining a list of reliable vendors.

On the positive side, the screen and burn-in evaluation process sets a procedure for rapid dynamic feedback of information both to the components manufacturer and to the customer analysis, testing, and reliability groups. This information, in turn, has the potential of eliminating failure mechanisms. After failure analysis, the customer may find some failure modes are eliminable by allowing controlled process changes, by modifying specifications, or by refusing to qualify certain designs (i.e., package, device design, or metalization patterns) which have been found to contribute to failures. The manufacturer receives from the customer extensive data and failure information which increases the incentive to study the problems, find the failure causes and, consequently, eliminate or control them. Both manufacturer and customer are roused to study screen procedures because of the failure modes which defy present screening or are difficult to detect.

The lot acceptance specifications themselves are open to constant study and revision. The appended specifications represent a first approach toward accomplishing an assurance of needed reliability goals and have already shown their effectivity. However, as more data is accumulated, as semiconductor processing and screening procedures reach new levels, and as failure mechanisms become better known, modifications must be employed. Since the procedure must be realistic with respect to needed failure rates, this does not necessarily mean that all acceptable limits will be tightened. On the contrary, the limits may be loosened and failure modes reclassified as more assurance of process control and screenability is developed. Tightening of the limits becomes necessary if failures occur during field use, indicating that the rejection criteria are insufficient.

The entire approach requires an intimate cooperation between the customer and vendor with resultant understanding of the problems of both. The mutual cooperation is necessary to achieve success of the mission.

COST

The approach presented by ND 1002248 has the potential of quantitatively defining reliability cost. By establishing limits for the number of failures per failure mode for lot acceptance, it has been shown that failure rates of 0.0018% per thousand hours are achievable. The approach also makes possible future correlation to field failure rates by the formalized procedure of the lot acceptance specification.

The cost of applying ND 1002248 may occur in one of two ways. The customer, buying from reliable component manufacturers, may absorb the cost by buying an excess of units and not using the rejected lots or sublots of parts in high reliability equipment. The component manufacturers may absorb the cost by guaranteeing the device will meet the specification and thus accept the return of rejected lots. For high volume usage devices, where adequate competition is possible, the latter approach appears to be the least expensive and most desirable. The component manufacturer is incentively induced to increase his yield to

the acceptance specification while adequate competition and low unit price prevents a component manufacturer from attempting to screen out failures. Attempts by manufacturers to "test in" reliability does not remove nonscreenable and insidious long-time dependent failures which is one of the primary reasons for lot rejection by failure mode. The actual added cost of applying ND 1002248 lot rejection to the manufacturer has been shown to be equivalent to the cost of standard Group B sample testing.

In any event, the cost of assuring the success of space missions is finite and justified. However, for any program, the cost required to increase component reliability must be weighed against the cost of retrofit due to field failures. In short, the most effective positive method of building an economically reliable system is to build the system with reliable parts.

ACKNOWLEDGEMENT

The authors wish to acknowledge Raytheon Co., Space and Information Systems Division, for implementing the process specifications for semiconductor parts. In particular, we greatly appreciate the help and cooperation of the Reliability and the Screen and Burn-in Groups.

REFERENCE

1. Partridge, J., Hanley, L. D., and Hall, E. C., "Progress Report on Attainable Reliability of Integrated Circuits for Systems Application", MIT Instrumentation Laboratory Report E-1679, presented at the Symposium on Microelectronics and Large Systems, Nov. 18, 1964, Washington, D.C.

APPENDIX A

Apollo G & N Specification
ND 1002248

PROCESS SPECIFICATION
SPECIAL CONDITIONING OF NOR GATES
(Flat Packs)

1. SCOPE

1.1 PURPOSE

This specification establishes the minimum requirements for the acceptance of integrated circuit nor gates for use in flyable deliverable end items. The procedures described herein shall be performed by the G&N Industrial Contractor as part of incoming inspection, screen, and burn-in.

2. APPLICABLE DOCUMENTS

2.1 EFFECTIVE ISSUES

The following documents of the issue in effect on the date of this document form a part of this specification to the extent specified herein.

SPECIFICATIONS

Military

MIL-STD-750	Test methods for semiconductor devices

Apollo G&N

ND 1002246	Leak test procedures for nor gates.
ND 1002257	Internal visual rejection criteria for integrated circuits.

DRAWINGS

Apollo G&N

1006321	Specification control drawing for dual nor gate (flat packs).

REPORTS

MIT/IL E-1679	Progress Report on Attainable Reliability of Integrated Circuits for System Application.

3. REQUIREMENTS

3.1 GENERAL

The provisions of this specification shall be applicable to all phases of acceptance of integrated nor gates to the extent specified herein. Specific requirements or provisions not covered by this specification shall be as specified on the applicable drawing or purchase order. In the event of conflict between the requirements of the applicable drawings, this specification and other documents cited herein, the requirements of the applicable drawings and this specification shall govern in that order.

3.2 PROCESS CONTROL

The process covered by this specification shall be controlled in accordance with the process control provisions of 4.2.

3.2.1 Lot Control

Each lot (6.2.1) of up to 5000 units as supplied by the vendor in compliance with 1006321 shall be identified and maintained by the contractor throughout the test sequence, 3.3.1.

3.2.2 Serialization

All units of a lot shall be serialized by the G & N contractor. A unit shall be identified by the lot number and the unit serial number.

3.3 TEST PROCEDURES

3.3.1 Test Sequence

Each lot of nor gates shall be subjected to tests in the following sequence:

a. External visual inspection (Test #1).

b. Physical dimension, lead tension, and fatigue inspection (Test #2).

c. Electrical test (test #3).

d. Thermal cycle test (test #4).

e. Helium leak test (test #5).

f. Nitrogen bomb, oil bubble tests (test #6).

g. High temperature bake test (test #7).

h. Centrifuge Y_1 test (test #8).

i. Continuity open and short test (test #9).

j. Centrifuge Y_2 test (test #10).

k. Electrical test (test #11).

m. Propagation delay (test #12).

n. Emitter base back bias (test #13).

o. D.C. current gain measurement (test #14).

p. Operation life test (test #15).

q. Electrical test (test #16).

r. Vibration test (test #17).

s. Shock test (test #18).

t. Continuity test (test #19).

u. Internal visual inspection (test #20).

A flow diagram of the above sequence is attached.

3.3.1.1 Removal of Failures

Catastrophic failures only shall be removed from the test sequence at the point of detection and subjected to failure analysis. The point in the test sequence 3.3.1 where the failure was detected must be recorded and a set of electrical readings as specified in Paragraph 3.3.2.3 shall be performed. All electrical failures at the end of the test sequence 3.3.1 shall be subjected to failure analysis except the incoming marginal failures as defined in 6.2.3 (a), may be returned to the vendor.

3.3.2 Tests

3.3.2.1 External Visual Inspection

Each lot of nor gates shall be subjected to an external visual inspection in accordance with MIL-STD-750, method 2071, and Specification Control Drawing 1006321 with additional requirements to be negotiated with the vendor, and to be included in the purchase order.

3.3.2.2 Physical Dimension, Lead Tension, & Fatigue

A sample of 10 nor gates shall be subjected to the physical dimension examination of MIL-STD-750, method 2066, and Specification Control Drawing 1006321. Five of the 10 units shall be subjected to the lead tension and lead fatigue tests. Test #2, as specified below. The five units subjected to the lead tension and lead fatigue tests are to be forwarded to test 20 of the test sequence 3.3.1. The remaining 5 units shall be forwarded to test 3 of the test sequence 3.3.1.

(a) Lead Fatigue. Leads shall be capable of withstanding the following test: The unit shall be held in a vertical position with a 2 ounce weight suspended from the lead to be tested. Two cycles of bending shall be performed. A cycle consisting of moving the body of the unit, 45 degrees from the vertical in one direction, and back 45 degrees to the original position. No mechanical damage shall be evidenced after the test.

(b) Lead Tension. Each lead shall be capable of withstanding an axial pull of 1 pound for a period of 30 seconds. No mechanical damage shall be evidenced after the test.

3.3.2.3 Electrical Test (Test #3 of 3.3.1)

The entire lot shall be subjected to electrical test as described in Specification Control Drawing 1006321 with the limits as specified. The test to be performed on all base currents, I_B, all output voltages. V_O, both output currents, I_O, both collector emitter threshold currents, I_{CEX} and both collector emitter sustaining voltages, V_{CEO} sust. D.C. current gain h_{FE} shall be measured on only one transistor of each gate and RL shall be measured on one gate only.

3.3.2.4 Electrical Test (Test #11 and #16 of 3.3.1)

The electrical test shall be the same as performed for 3.3.2.3 except that the maximum limits as defined in Specification Control Drawing 1006321 shall be raised 4% the minimum limit decreased 4% and the V_{CEO} sust. test will not be performed.

3.3.2.5 Thermal Cycle Test

The units shall be subjected to thermal cycle consisting of 3 cycles of the following: The units shall be stabilized for 30 minutes minimum at +150°C ± 5°C in an oven. They shall then be transferred to an oven operating at -65 ± 5°C in less than 10 seconds. The units shall stabilize for not less than 30 minutes and then be returned to the ±150 ± 5°C oven in less than 10 seconds transfer time.

3.3.2.6 Helium Leak Test

The helium leak test shall be performed in accordance with ND 1002246 using a rate of 5×10^{-8} cc/atm/sec, as the upper limit.

3.3.2.7 Nitrogen Bomb, Oil Bubble Tests

The nitrogen bomb, oil bubble tests shall be performed in accordance with ND 1002246. The nitrogen bomb test shall be performed first.

3.3.2.8 High Temperature Bake Test

The high temperature bake test shall be performed in accordance with MIL-STD-750, method 1031, except the temperature shall be 150 ± 5°C and the time shall be 168 ± 8 hours.

3.3.2.9 Centrifuge Y_1 Test

The centrifuge Y_1 test shall be performed with an acceleration of 20,000g in accordance with MIL-STD-750, method 2006. Plane Y_1 is defined as a force attempting to push the internal lead wires toward the bottom of the device.

3.3.2.10 Continuity Open and Short Test

The continuity test shall be performed to detect open bonds and shorts between leads, leads and case, and leads and chip. At test #9 check 100% at test #19 check 77 units from test #18.

3.3.2.11 Centrifuge Y_2 Test

The centrifuge Y_2 axis test shall be performed with an acceleration of 20,000g in accordance with MIL-STD-750, method 2006. Plane Y_2 is a force opposite to Y_1 as defined in paragraph 3.3.2.9.

3.3.2.12 Propagation Delay (Test #12)

Propagation delay shall be performed according to Specification Control Drawing 1006321 on a sample of 200 units* from each lot.

3.3.2.13 Emitter Base Back Bias Test

The emitter base back bias test shall be performed on 200 units* as follows. Each base input shall be connected to minus 6 volts with respect to common emitter via a 10K series resistor in each base. The units shall be operated with voltage applied at a temperature of 25°C ± 5°C for a period of 72 hours.

3.3.2.14 Beta Measurement

The D.C. current gain measurement shall be performed in accordance with Specification Control Drawing 1006321 on the same transistors as measured at 3.3.2.4.

3.3.2.15 Operation Life Test

An odd number of units (Gates) shall be connected in series with the output of the last unit supplying the input to the first unit, thus forming a "Ring" oscillator with 8 vdc ± 5% applied continuously to the power terminals of all units in the circuit, ("Ring") oscillation must occur at the initiation of the test. This test will be performed on all units for a period of 168 hours ± 8 hours. The ambient temperature shall be maintained at 25 ± 10°C.

---

* When samples are selected for tests #12, #13, & #17, the sample shall be representative of all diffusion sub-lots included in the shipment lot.

3.3.2.16 Vibration Test

The vibration test shall be performed on random sample of 77 units.* The vibration test shall be performed in accordance with MIL-STD-750, Method 2056, 30g's from 5 to 2000 cps. limited to 0.12 inch double amplitude, 3 cycles, 15 minutes per cycle minimum.

3.3.2.17 Shock Test

The shock test shall be performed on the same units tested in 3.3.2.16. The shock test shall be performed in accordance with MIL-STD-750, Method 2016, 1500 g's, 0.5 m sec, 5 blows in all axis directions, 30 blows total.

3.3.2.18 Internal Visual Inspection

The internal visual inspection shall be performed on the 82 units from test 3.3.2.17 and 3.3.2.3 and in accordance with ND 1002257.

4. QUALITY ASSURANCE PROVISIONS

4.1 GENERAL

In order to assure proper control of the acceptance process covered by this specification, the contractor shall meet all the requirements specified herein and shall provide continuous audit of the acceptance process to assure compliance with the requirements of this specification.

4.1.1 Inspection

The contractor, through his quality assurance or control agency shall be responsible for the performance of all inspection requirements and tests specified herein.

4.2 FAILURE CRITERIA

4.2.1 Failure Analysis

All nor gates failing in the electrical tests specified in 3.3.2 (test #3, 9, 11, 14 and 16) except the non-catastrophic failures as defined in 6.2.3 (a,b,c, and e), shall be subjected to a failure analysis sufficient to identify cause and mode of failure. For failure definitions refer to Section 6.2.

4.2.2 Failure Modes

After failure analysis all failures detected at test 3, 9, 11 and 16 of test sequence 3.3.1, except induced failures and non-catastrophic failures as defined in 6.2.3 (a, b, c, and e), shall be classified as to the following failure modes which are described in MIT/IL report E-1679.

* When samples are selected for tests #12, #13, & #17, the sample shall be representative of all diffusion sub-lots included in the shipment lot.

a. Class A failure modes (class A failure modes are generally of a type readily weeded out during screen and burn-in).

1. Open bonds due to poor metalization adhesion to the silicon dioxide.

2. Open bonds due to underbonding.

3. Open bonds due to gold-aluminum eutectic formation.

4. Open bonds due to overbonding.

5. Opens due to nicks and cuts in the bonding wire.

6. Leads shorting to the edge of the chip or leads shorting to the package lid.

7. Open interconnects detected only during test 3.3.2.3 due to only scratches with no evidence of metalization corrosion at the open.

8. Shorts due to metalization scratching and smearing.

9. Shorts induced by the collector to emitter sustaining voltage test of paragraph 3.3.2.3.

10. Failures due to cracked chip.

11. Opens due to the thinning of lead wire due to poor bonding procedure.

12. Non-catastrophic failures due to surface instabilities that are not included in 6.2.3 (a, b, c and e).

b. Class B failure modes (Class B failure modes are of a type less readily detected during screen and burn-in as compared with Class A).

1. Shorts resulting from leads touching any other leads and shorts resulting from leads touching metal interconnects.

2. Opens in the interconnect due to the gold-aluminum eutectic formation at the neck of an interconnect.

3. Shorts through the silicon dioxide due to poor oxide dielectric strength.

4. Shorts through the oxide because the bonds are too close to the chip edge.

5. Shorts, intermittent or otherwise, due to particles in the package.

6. Shorts, intermittent or otherwise, due to free lead material and fixed extra leads or lead material.

7. Catastrophic failures due to surface instabilities.

8. Opens in interconnect at oxide steps detected during test 3.3.2.3.

c. Class C failure modes (Class C failure modes are of a type which are time dependent and/or are not easily detected during screen and burn-in).

1. Opens in the interconnect due to corrosion.

2. Opens in the interconnect detected after test 3.3.2.3 at oxide steps.

3. Opens in the interconnect detected after test 3.3.2.3 at scratches.

4. Any failures due to electrically insulating or electrically high resistance layers forming at the silicon oxide window between the metal contact and the silicon or between the layers of metal.

5. Die separated from package header.

4.2.3 Failure Mode Grouping

4.2.3.1 Classification

Following the classification of failure modes from a lot the electrical failures will be divided into Group I - IV below and the percentage failure for the lot in each group shall be determined. Group 0 contains special cases of non-catastrophic failures.

a. Group 0. Test 3

Non-catastrophic failures as defined in 6.2.3 (a) and propagation delay failure of Test #12.

b. Group 0. Test 9, 11, 16

Non-catastrophic failures as defined in 6.2.3 (b, and c).

c. Group I. Class A failure modes.

d. Group II. Class B failure modes.

e. Group III Class C failure modes.

f. Group IV Any failure, except induced failures, not listed in Section 4.3.2 or any failure for an unknown cause.

TABLE I

Test Number (See Para. 3.3.1)	*Maximum Percent of Failures				
	Group 0	Group I	Group II	Group III	Group IV
3	0.5%	0.3%	0.08%	0.04%	0%
9, 11, and 16	1.0%	0.3%	0.04%	0.02%	0%

* For shipment lots of from 4000 to 5000 units, use the same number of allowable failures as applied to lots of 5000. For smaller lots, use percentages as shown in Table I. If the number of unit failures allowable is calculated to be a mixed number, a combination of an integer and a fraction, use the integer only.

TABLE II

Test Number (See Paragraph 3.3.1)	Maximum Percent of Failures or Maximum Allowable rejects
2	1 defective unit, Physical Dimensions
2	1 defective unit, Lead fatigue and tension
14	10 units
5	2.0%
6	2.0%
19	1 unit
20	8 units

4.3 REJECTION CRITERIA

4.3.1 Lot Rejection

The failures of a shipment lot shall be classified as specified in 4.2.2 such that the failures can be identified with the groupings specified in 4.2.3. The maximum allowable percentages of failures from test number 3, 9, 11, and 16 of paragraph 3.3.1 according to the failure mode

groupings of 4.2.3 are given in Table I. The maximum allowable percentages of failures from test numbers 2, 5, 6, 14, 19, and 20 of paragraph 3.3.1 are given in Table II where the failure definitions are given in 6.2. Failure to meet any one of the maximum allowable percentages of Tables I and II or failure to comply with the test sequence of 3.3.1, the test procedure 3.3.2, the flight qualification requirements of 4.4, or the data requirements of 4.5 shall be cause for lot rejection.

4.3.11 Conformance to ND 1015404

Disclosure of any violation of previously agreed to contractor-supplier ND 1015404, "Critical Process" list without prior notification automatically fails the entire lot. Notice of such deviation must be made by the contractor to MIT/IL within 24 hours of disclosure.

4.3.2 Sub-lot removal

If the reason for shipment lot rejection can be assigned to failure modes which are traceable to a diffusion sublot(s), the entire diffusion sub-lot(s) shall be removed from the shipment lot and the provisions of paragraph 4.3.1 shall be reapplied to the remainder of the shipment lot. For example, failure modes which are traceable to diffusion sub-lot(s) are described in paragraph 4.2.2 sub-paragraphs, a 1, b 3, b 8, c 2, and c 4.

4.4 FLIGHT QUALIFICATION

4.4.1 Flight Qualified Hardware

A nor gate is flight qualified when the lot, of which the nor gate is part, is not rejected according to 4.3 and the nor gate does not fail any test of the sequence of paragraph 3.3.1.

4.4.1.1 Failure Traceability

Any nor gate failure detected in qualified flight hardware must be traceable to a lot as identified in paragraph 3.2.1 and to the unit serial number.

4.4.2 Qualification Report

Two copies of a report justifying the acceptance or rejection of a lot as flight qualified shall be forwarded to MIT/IL, prior to use in deliverable end items. The report shall include the following:

a. A summary of screen and burn-in data.

b. A detailed list of the screen and burn-in results which includes the number of failures at each test station of test sequence 3.3.1.

c. A failure report of all electrical failures, by unit serial number including induced failures, as specified in 4.3.1 which includes:

1. Photographs of each category of each photographable failure. A minimum of two photographs of each category is required where more than one exists.

2. Analysis of each failure.

3. Classification of each failure according to 4.3.2.

4. All electrical test data of each failure.

d. Number of failures in the failure mode groups according to 4.2.3.

e. Date of purchase.

f. Total number of ordered parts.

g. Date code of parts received.

h. Lot identification number.

i. Allocation of all parts from the lot updated to the date of issue of qualification report indicating the number of units which passed screen and burn-in, the number of failed units, the number of induced failures and the number of units removed from the lot for any other reason.

j. Vendor supplied Table I and Table II, sub group 1 and 3 test data.

k. A report by lot of all internal visual inspection failures (test #20 in test sequence 3.3.1) by unit serial number which includes

1. Photograph of each failure category detected.

2. Classification of each failure according to ND 1002257.

l. A list of all process changes allowed by the contractor in accordance with 1015404, paragraph 3.3.2.2.

4.5 DATA

4.5.1 Data Storage

Incoming inspection, screen, and burn-in data shall be maintained and stored by lot number and unit serial number for three years.

4.5.2 Cataloging

Nor gates failing the tests specified herein with the exception of the external visual inspection and leak tests shall be cataloged and stored by lot number and serial number for three years. The devices must be readily accessible for future reference.

4.5.3 Lot Storage

Units submitted to the contractor as part of the Quality Demonstration Test shall be stored by contractor's lot number for three years. The devices must be readily accessible for future reference.

5. PREPARATION FOR DELIVERY

This section is not applicable to this specification.

6. NOTES

6.1 INTENDED USE

This process conditions nor gates used in Apollo Guidance and Navigation Equipment.

6.2 DEFINITIONS

6.2.1 Lot

A shipment lot is defined as a group of nor gates submitted by a vendor in compliance with 1006321.

6.2.2 Catastrophic Failures

A catastrophic failure is defined as any device which fails the electrical tests of Table II of Specification Control Drawing 1006321 by twice the maximum or one half the minimum limits of that table. •

6.2.3 Non-Catastrophic Failures

A non-catastrophic failure is any device which cannot be classed as a catastrophic failure by definition 6.2.2 but which fails according to the definitions described below:

a. A non-catastrophic failure at test 3.3.2.3 exceeds the limits of Table II of Specification Control Drawing 1006321 and does not become a catastrophic failure during test sequence 3.3.1.

b. A non-catastrophic failure at test 3.3.2.4 exceeds the limits described in 3.3.2.4 but changes parameters from 3.3.2.3 to 3.3.2.4 by less than ± 15% for base current and output voltage, ± 10% for output current or ± 20% for collector emitter threshold current. For collector emitter threshold current of less than 100 namps, an initial reading of 100 namps is assumed.

c. A non-catastrophic failure at test 3.3.2.4 does not exceed the limits described in 3.3.2.4 but changes parameters from 3.3.2.3 to 3.3.2.4 by more than ±15% for base current and

output voltage, ± 10% for output current, or ± 20% for collector emitter threshold current. For collector emitter threshold currents of less than 100 namps, an initial reading of 100 namps is assumed.

d. A non-catastrophic failure at test 3.3.2.4 exceeds the limits described in 3.3.2.4 and changes parameters from 3.3.2.3 by more than ± 15% for base current and output voltage, ± 10% for output current, or ± 20% for collector-emitter threshold current. For collector-emitter threshold currents of less than 100 namps, an initial reading of 100 namps is assumed.

e. A non-catastrophic failure at test 3.3.2.14 exceeds a change in D.C. current gain by ± 15% in test sequence 3.3.1 test number 11 and test number 14.

6.2.4 Induced Failures

An induced failure is a catastrophic failure which through failure analysis can be proven to be caused by exceeding the stress limits of Specification Control Drawing 1006321.

6.2.5 Leak Test Failures

A failure at test 5 of paragraphh 3.3.1 is failure to meet the leak rate therein. A failure at test 6 of paragraph 3.3.1 is the failure to meet the criteria as specified in ND 1002246.

6.2.6 External Visual Inspection Failures

A failure at test 1 of paragraph 3.3.1 is failure to meet the visual and mechanical examination criteria of Specification Control Drawing 1006321 and additional requirements negotiated with the applicable vendor.

6.2.7 Physical Dimension Failures

A physical dimension failure of test 2 of paragraph 3.3.1 is a package which does not meet the physical dimension criteria of Specification Control Drawing 1006321.

6.2.8 Lead Tension, Lead Fatigue Failures

The lead tension, lead fatigue failures of test 2 of paragraph 3.3.1 is a partial or complete severing of a lead from the package.

6.2.9 Propagation Delay Failures

A failure at test 12 of paragraph 3.3.1 is a failure to meet the criteria of Specification Control Drawing 1006321.

6.2.10 Failures

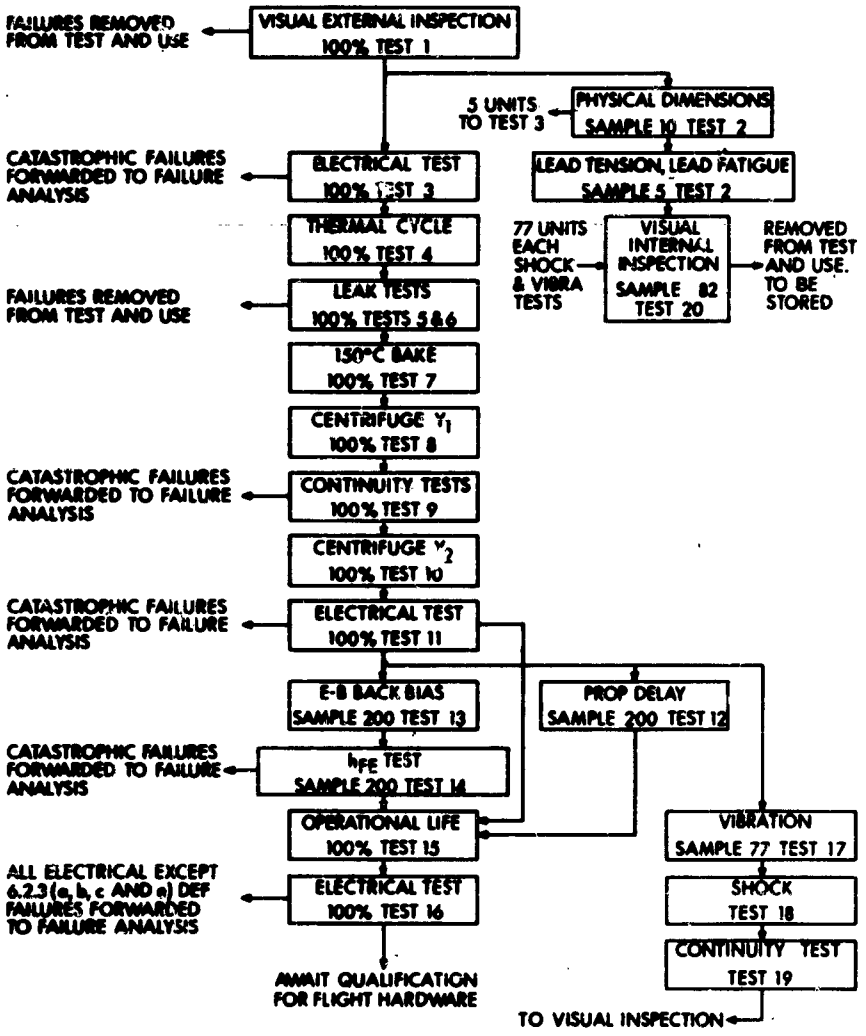
Failure of a unit in one or more tests will be charged as a single failure. A unit which could be classed by several failure modes shall be classed in the highest alphabetical mode as listed in 4.2.2. A unit which meets the definition of 6.2.3 (a,b and c) shall be counted in Group 0 only.

6.2.11 Internal Visual Inspection Failures

A failure at test # 20 of paragraph 3.3.1 is a failure to meet the criteria of ND 1002257.

6.2.12 Continuity Failure

A failure of the continuity test is the detection of an open or a short.



FLOW DIAGRAM FOR THE TEST SEQUENCE 3.3.1

APPENDIX B

Apollo G&N Specification
ND 1002257

INTERNAL VISUAL REJECTION CRITERIA FOR INTEGRATED CIRCUITS

1. SCOPE

1.1 This specification defines the internal visual rejection criteria covering surface imperfections, cleaniness, workmanship, and design as it applies to silicon planar integrated circuits for use in the Apollo Guidance Computer.

2. REQUIREMENTS

2.1 INSPECTION

Inspections shall be performed with at least the minimum microscope powers as specified herein. When the minimum microscopic powers are not sufficient to determine if the herein specified faults are present, higher magnifications shall be used. When powers of 80 or more are employed, a columnated light source applied through the objective lens shall be used.

3. REJECT CRITERIA

3.1 SCRATCHES

A scratch is defined as any tear in the metalization caused by instruments such as tweezers, probes, vacuum pickups, scribing tools, etc. Inspection for scratches shall be performed at a magnification of 150 power minimum. The following constitutes rejects for scratches:

a. Any device which exhibits a scratch which reduces the width of the undisturbed metalization to less than 0.4 mils and which exposes silicon dioxide anywhere along the scratch is a reject. (Refer to Fig. 1).

b. Any device, which exhibits a scratch over or along an oxide step (when the oxide step intersects all but 0.4 mil or less of the interconnect) and which reduces the width of the undisturbed metalized conducting path to any contact to less than 0.6 mil, is a reject. (Refer to Fig. 2).
NOTE - It is assumed that scratches over oxide steps are electrically open at the step.

3.2 METALIZATION CORROSION

Inspection for metalization corrosion shall be performed at a magnification of 150 power minimum. Any device which exhibits any junction area covered only by unthermally oxidized silicon is a reject.

3.4 VOIDS

A void is defined as any region in the metalization where silicon dioxide is visible within the designed areas of the metalization and the silicon dioxide exposure is not caused by a scratch. Inspection for voids shall be performed at a magnification of 80 power minimum. The following constitutes rejects for voids:

a. Any device which exhibits voids at an interconnect which reduces the width of the undisturbed metalization to less than 0.6 mils is a reject (refer to Fig. 3).

b. Any device which exhibits voids at a pad or fillet which leaves the pad or fillet less than 50% of its designed area is a reject (refer to Fig. 4).

c. Any device, which exhibits voids over an oxide step (when the oxide step intersects all but 0.6 mil or less of the interconnect) and which reduces the width of the undisturbed metalized conducting path to any contact to less than 0.75 mil, is a reject (refer to Fig. 5).

3.5 MISALIGNED CONTACTS

The alignment of the metalization contact to the silicon shall be inspected at a magnification of 80 power minimum. The metalization shall make contact to the silicon over at least one half the area of the applicable window contact. Any device which does not meet this requirement is a reject.

3.6 CRACKS IN THE DIE

Inspection for die cracks shall be performed at a magnification of 80 power minimum. Any die which exhibits cracks in the active circuit, metalization, or bond areas is a reject. Any die which exhibits cracks 1 mil in length or greater which point toward the active circuit metalization or bonds is a reject (refer to Fig. 6).

3.7 BOND PLACEMENT

The placement of bonds shall be inspected at a magnification of 80 power minimum. For ultrasonic or bird beak bonds, the word "bond" refers to the tool impression. Bond placement shall be viewed directly from above. The following constitutes rejects for bond placement (refer to Fig. 7):

a. Any bond which is placed such that silicon dioxide is not visible between the outer periphery of the bond contact area and any other bonding pad or a silicon oxide edge (unthermally oxidized or "raw" silicon) shall cause the device to be rejected.

b. Any ultrasonic bond which has less than 75% of the bond area in contact with the metalized pad shall cause the device to be rejected.

c. Any bird beak or ball bond which has the wire at the base of the bond outside the boundaries of the pad or any bond with less than 50% of the bond area in contact with the metalized pad shall cause the device to be rejected.

d. Any bond which is located in the fillet area and the longest distance between the bond periphery and edge of fillet is less than 0.4 mils shall cause the device to be rejected.

e. Any bond contact area made on the interconnect shall cause the device to be rejected.

3.8 DAMAGED LEADS

Leads shall be examined for damage at a magnification of 80 power minimum. Any lead which exhibits nicks, cuts, crimps or scoring which cut into or deform the wire by more than 25% of the original diameter shall cause the device to be a reject.

3.9 LEAD WIRES

Lead wires shall be inspected at a magnification of 20 power minimum. The following constitutes rejectes for improper lead placement and lead dress:

a. Leads which exhibit sufficient excess length such that there exists the capability of shorting to another lead, edge or surface of the die, or to the bottom or top of the package without deforming the diameter of the lead shall cause the device to be rejected.

b. When viewed from above, leads which cross one another or which cross any metalization which is discontinuous with the pad to which the lead is bonded shall cause the device to be rejected.

c. Lead material greater in length than 2 wire diameters that is fixed only on one end, as for example "pigtails", shall cause the device to be rejected.

3.10 CONDUCTING PARTICLES

Inspection for conducting particles shall be performed at a magnification of 20 power minimum. Any device which contains loose or easily removable electrically conducting segments of material which are not part of the device design is a reject. Electrically conducting material shall include any material of sufficient conductance to cause device failure of any electrical specification by shorting contacts.

3.11 WEDGE BONDS

Wedge bonds shall be inspected at a magnification of 20 power minimum. Wedge bonds which are made at the post such that the diameter of the wire at regions where the wire does not make metallurgical contact to the post is constricted to less than 1/2 of the normal wire diameter shall cause the device to be a reject.

4. PAD, FILLET, AND INTERCONNECT AREAS

The pad, fillet, and interconnect areas shall be negotiated by the purchaser with the vendor prior to procurement.

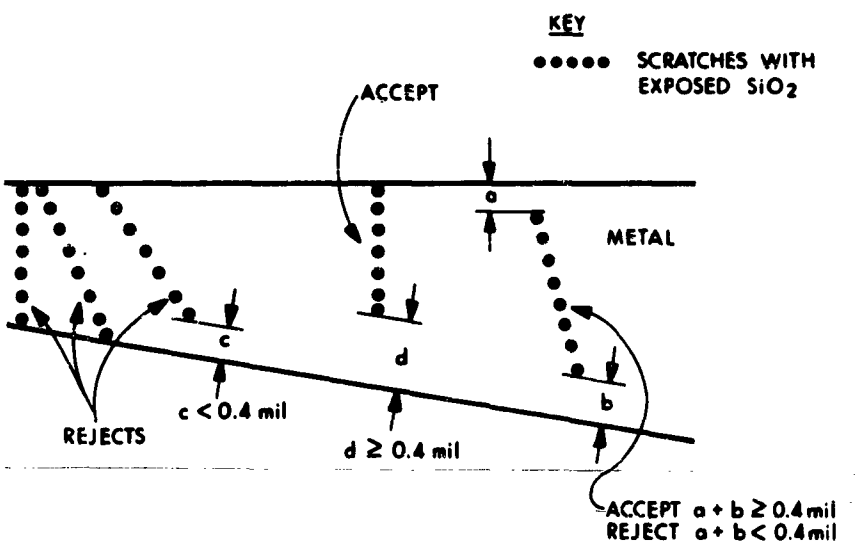


Fig. 1. Examples of acceptable and rejectable devices for scratches. Note that only those scratches which expose silicon dioxide somewhere along the scratch shall cause rejection.

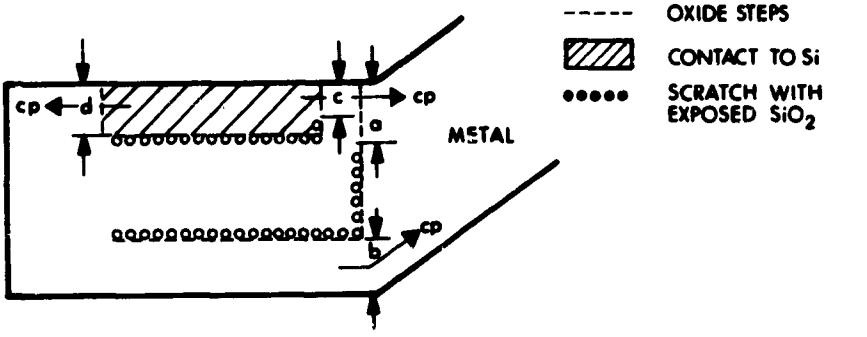


Fig. 2a. ACCEPT, a + b AND c + d ≥ 0.6 mil.
REJECT, a + b OR c + d < 0.6 mil.

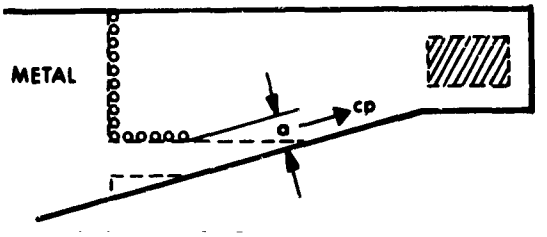


Fig. 2b. ACCEPT, a ≥ 0.6 mil.
REJECT, a < 0.6 mil.

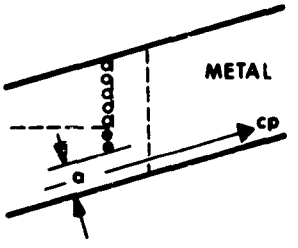


Fig. 2c. ACCEPT, a ≥ 0.4 mil.
REJECT, a < 0.4 mil.

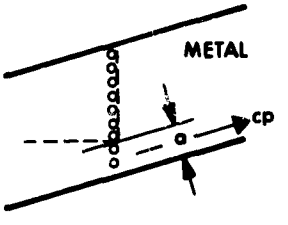


Fig. 2d. ACCEPT, a ≥ 0.4 mil.
REJECT, a < 0.4 mil.

Fig. 2. Examples of acceptable and rejectable devices for scratches at or along oxide steps.

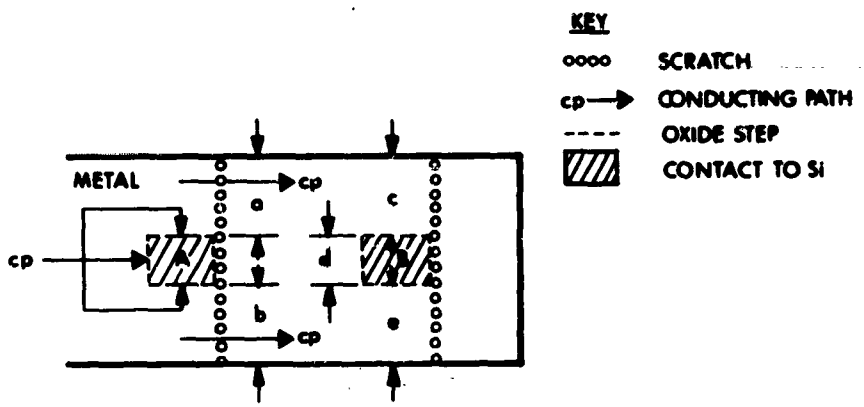


Fig. 2e. ACCEPT, a + b ≥ 0.4 mil.
REJECT, a + b < 0.4 mil SINCE THE CONDUCTING PATH TO CONTACT B IS AFFECTED.

ACCEPT, c+d+e ≥ 0.6 mil SINCE THE CONDUCTING PATH TO CONTACT B ≥ 0.6 mil AND NO OTHER CONTACT IS AFFECTED.
REJECT, c+d+e < 0.6 mil.

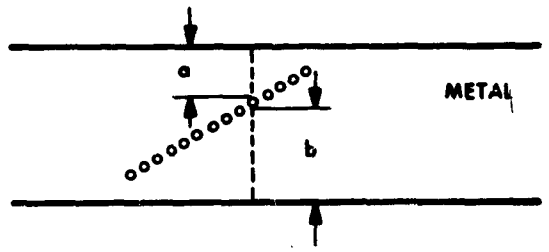


Fig. 2f. ACCEPT SINCE THE SCRATCH DOES NOT EXPOSE SiO_2 AND a + b ≥ 0.6 mil.
REJECT, a + b < 0.6 mil.

Fig. 2 cont. Examples of acceptable and rejectable devices for scratches at or along oxide steps.

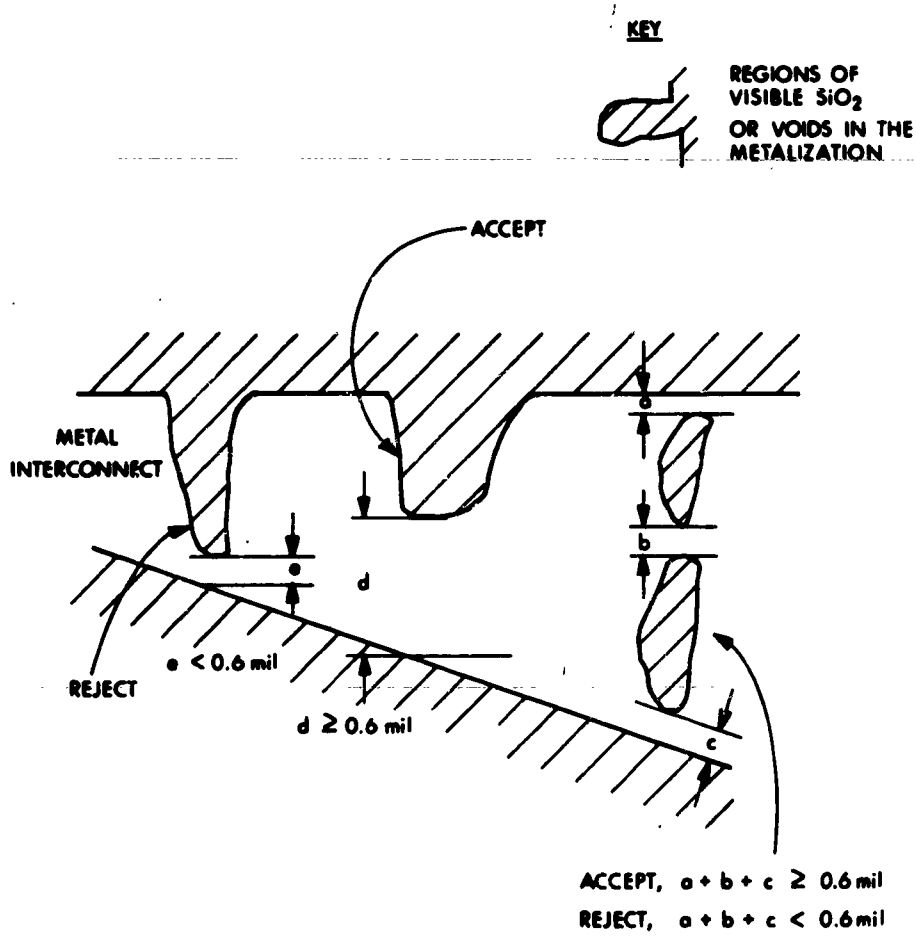
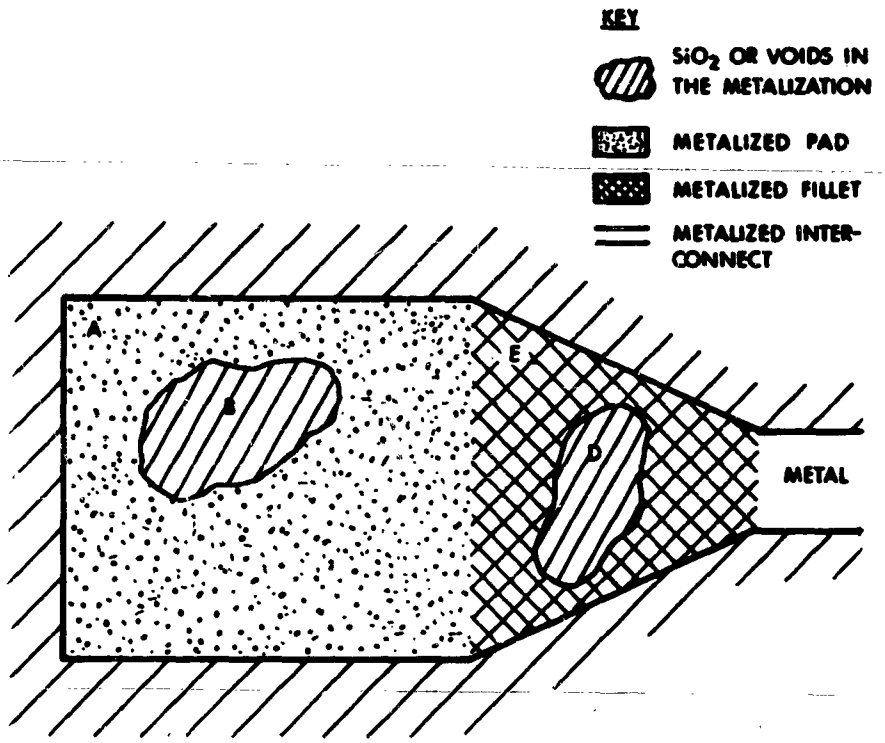


Fig. 3. Examples of acceptable and rejectable devices for voids in the metal interconnect. Note that the conditions for accept or reject are the same as for scratches (Fig. 1) except that the minimum acceptable distance is larger.



ACCEPT IF AREA B < 1/2 AREA A AND IF THE REQUIREMENTS OF PARAGRAPH 3.7b, c, d ARE MET.

REJECT IF AREA B ≥ 1/2 AREA A.

ACCEPT IF AREA D < 1/2 AREA E.

REJECT IF AREA D ≥ 1/2 AREA E.

Fig. 4. Examples of acceptable and rejectable devices for voids in the metal pad and fillet.

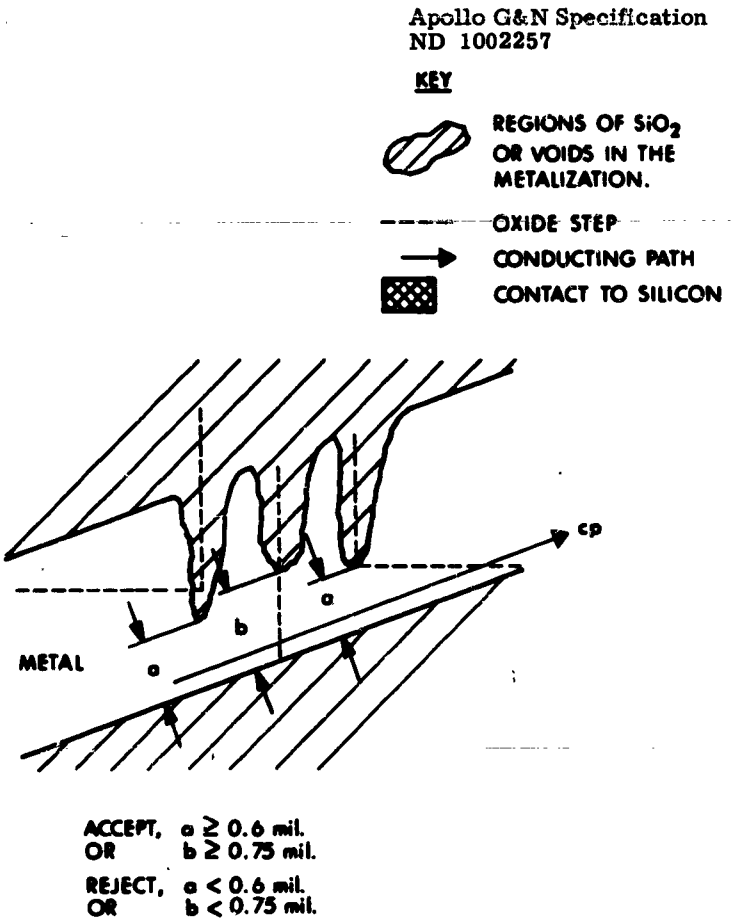


Fig. 5. Examples of acceptable and rejectable devices for voids at an oxide step. Note that the conditions for accept or reject are the same as for scratches at oxide steps (Fig. 2) except that the minimum acceptable distance is larger. For more examples of voids at oxide steps refer to Fig. 2, substituting the appropriate numbers.

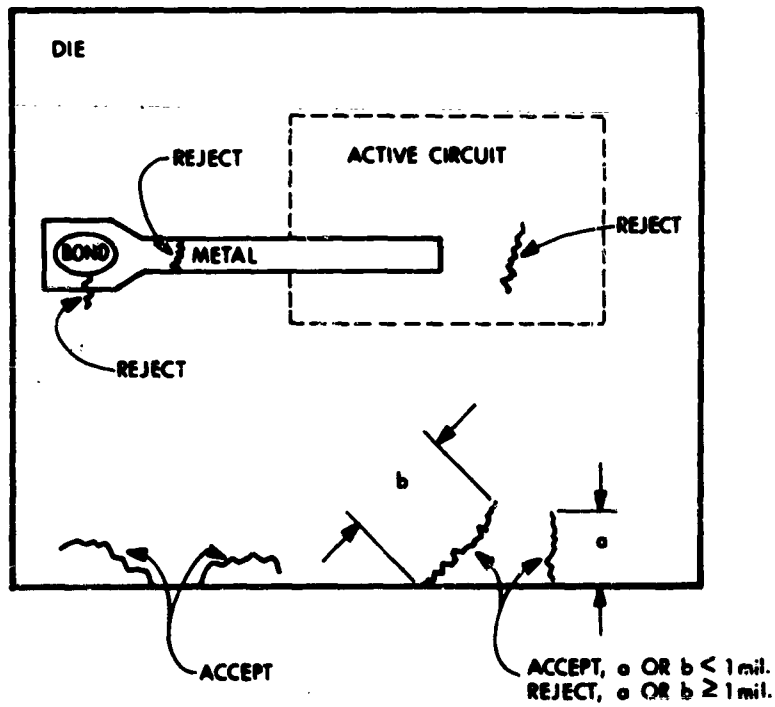


Fig. 6. Examples of acceptable or rejectable devices for cracks in the die.

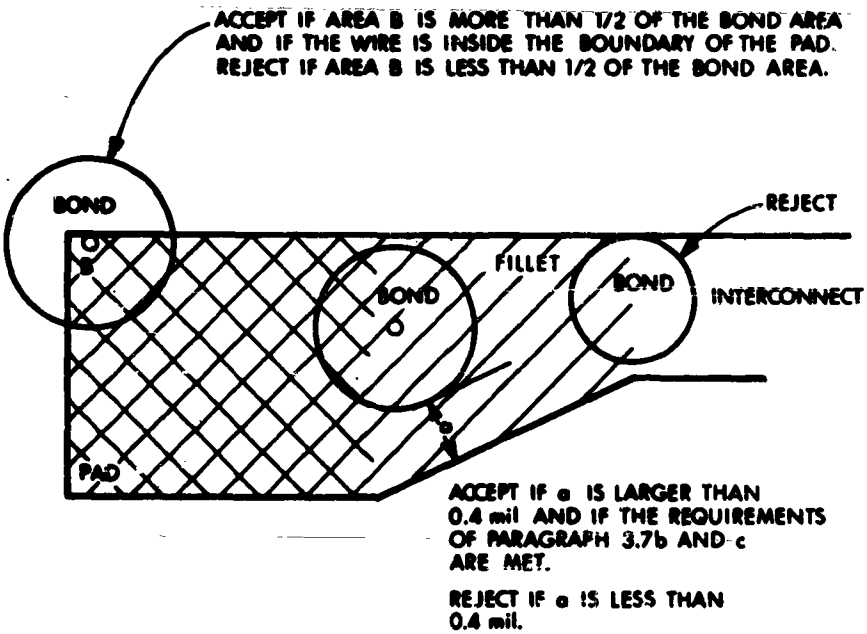


Fig. 7. Examples of acceptable and rejectable devices for bond placement. Pad, fillet, and interconnect areas are to be negotiated.

APPENDIX C

Apollo G&N Specification
ND 1002246

LEAK TEST PROCEDURES FOR NOR GATES

1. SCOPE

1.1 PURPOSE

This specification establishes the procedures for leak testing of Nor gates in a flat package and the rejection criteria for the leak tested flat packages.

2. APPLICABLE DOCUMENTS

2.1 EFFECTIVE ISSUES

The following documents of the issue in effect on the date of this document form a part of this specification to the extent specified herein.

Specifications

Military

MIL-STD-202C	Test methods for electronic and electrical component parts.

Drawings

Apollo G&N

1006321	Specification control drawing for dual nor gates (flat-packs).

3. REQUIREMENTS

3.1 GENERAL

The three leak tests specified herein are required to detect all nor gate flat pack leakers up to 5×10^{-8} cc/sec. The leak tests shall be performed in the following order.

3.2 HELIUM OR RADIFLO LEAK TESTS

3.2.1

The helium or radiflo leak tests shall be performed in accordance with MIL-STD-202C method 112, condition C, to the limits specified in 1006321.

3.3 NITROGEN BOMB

3.3.1 Materials

3.3.1.1 Isopropyl Alcohol, Reagent Grade

3.3.2 Apparatus

3.3.2.1 Pressure Vessel

A pressure vessel capable of storing flat packages and capable of maintaining 150 psi of nitrogen for 20 hours shall be used. The vessel must be constructed such that the packages can be removed from the vessel under pressure to the alcohol bath within a time period of no longer than three minutes.

3.3.2.2 Alcohol Bath Container

A container of approximately 4 inch diameter and 1/2 inch depth shall be used.

3.3.2.3 Binocular Microscope

A binocular microscope capable of magnification of 7 to 10 X shall be used.

3.3.3 Procedure

Flat packages shall be subjected to a nitrogen gas pressure of 150 psi for 10 to 20 hours. The flat packages will then be removed from the pressure vessel and placed in an alcohol bath such that the top of the package is under a 3/8 to 1/2 inch depth of alcohol and the alcohol bath container shall be under a binocular microscope of 7 to 10 X magnification. The time interval from beginning of depressurization to examination of the packages under the microscope shall be no longer than three minutes. The flat packages shall then be examined through the binocular microscope in groups of no more than 25 per person observing and no package body shall rest on another package body. The examination procedure shall consist of the following:

> The entire group of 25 packages shall be examined for a continuous period of fifteen minutes.

The criteria of a failure are the following:

> Observation of a continuous or intermittent stream of bubbles emanating from package leak producing areas during any examination period.

Note: Some packages will immediately emit a stream of bubbles then stop bubbling. Others will emit an intermittent stream of bubbles while others will not emit a stream of bubbles until a time period of minutes has elapsed.

3.4 HOT GLYCERINE BUBBLE TEST

The hot glycerine bubble test shall be performed, testing units in accordance with MIL-STD-202C method 112, condition A, with the following exceptions:

1. Glycerine shall be used instead of mineral oil.

2. The failure criteria shall be the observation of a growing bubble emerging from a sealed area, instead of observation of a continuous stream of bubbles emanating from the specimen. The observation of a growing bubble constitutes a leaker of greater than 10^{-5} cc/sec. Note that small non-growing bubbles may immediately appear upon insertion of the specimen into the glycerine due to trapped air in external package voids.

LIFE PREDICTIONS OF DIFFUSED GERMANIUM TRANSISTORS BY MEANS OF POWER STRESS

W. C. GIBSON

BELL TELEPHONE LABORATORIES
MARION AND VINE STREETS
LAURELDALE, PENNSYLVANIA

Introduction

Accelerated stress techniques[1,2] have gained general acceptance as a means of rapidly predicting semiconductor device life. The technique was initially employed using temperature as a stressing agent. Later, because of the belief that normal operating failure mechanisms may be enhanced by the applied voltage (field), power was also used as a stressing agent. In this case the thermal resistance of the device being evaluated was used to determine the power necessary to realize a junction temperature equivalent to storage temperatures. Little attention was paid to the manner in which the power was applied.

Results obtained from power stress experiments, therefore, frequently did not agree with those previously arrived at via temperature stress.[3] Different failure modes were observed and different life predictions resulted. These differences have caused some skepticism as to the validity of the use of power as a stressing agent, and in fact, the use of accelerated testing as a means of determining device reliability under operating conditions.[4]

Considerable effort has been expended to justify the use of both temperature and applied power as the stressing agents for silicon devices.[5] In general, however, life predictions of germanium transistors are based on temperature stress data only.[2,6] Mainly, this results from the fact that germanium has a lower intrinsic temperature than silicon. Consequently, the power dissipation range over which one can power stress a germanium device without introducing catastrophic thermal failures is relatively narrow. The range also depends on the particular device design.

The objective of the work reported herein was to show the feasibility of power stressing germanium devices to obtain meaningful life predictions. The general approach followed in an attempt to achieve the objective was to:

A) Insure that failure modes often observed when using power stress techniques[4] were not the result of inadequate testing or aging equipment;

B) secure a "safe" range over which the device could be operated without introducing abnormal failures;

C) stress sample groups of devices based on these findings and predict the device life;

D) compare the prediction with temperature stress life predictions of this and a similar device;

E) comment on "in service" results.

Device Used

An epitaxial germanium transistor designed for operation in the low microwave region was used for this study. The device is of the diffused base, alloyed emitter, mesa variety (Figure 1). SiO_2 is deposited on the device wafer for surface passivation and to protect it from possible particle inclusion during encapsulation. It is encapsulated in a TO-18 package. An encapsulation ambient of nitrogen, oxygen and helium is used. A BaO getter is included to minimize moisture content. The measured thermal resistance of the device is 1.3°C/mw.

Special Precautions

The transistor has a relatively large gain-bandwidth product. Consequently, it is very susceptible to oscillations in standard test sets. The feedback required to initiate and sustain the oscillations is provided by the measuring circuit and the package.[7] These oscillations are of sufficient magnitude to drive the device to destruction. Failure, due to oscillations, is recognized by a surface irregularity between the base and emitter contacts. This irregularity results from excessive localized heating which causes the base to collector, and usually, the base to emitter junction to short. Because of the speed with which this type failure occurred it was not directly proven the result of oscillations. It was concluded, however, from the fact that elimination of oscillations in the test set eliminated this type failure during testing. It will be shown later [Section [Power Stress Evaluation, Part C]] why oscillations cause the failure.

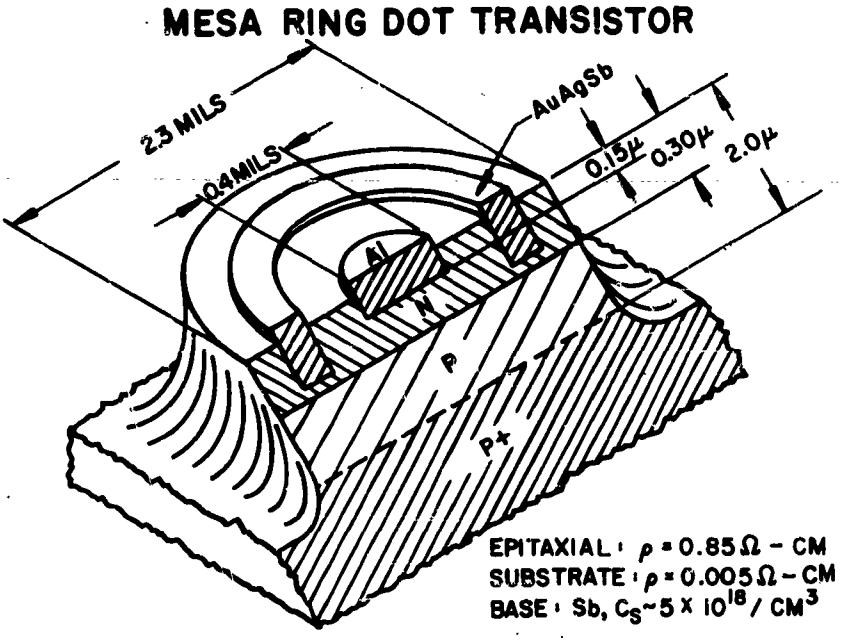


Figure 1

A second problem resulted from the small gold wires required to connect the emitter and base regions of the device to the header terminals (0.2 mils dia). The wires are easily burned open circuit by relatively low energy transients. For short pulses ($\leq 10^2$ μsec) the energy required for burnout is approximately 4.4×10^{-5} joules.[8] Opens as a result of transients are easily distinguished from those resulting from poor contacts or wirebonds by the characteristic ball which forms at the melted end of the wire. In general, the transistor element is not damaged, but is protected by the fuse-like behavior of the wire.

The two conditions discussed above made it necessary to carefully design and construct test sets specifically for measuring this device. Oscillations were eliminated simply by minimizing long lead-wire runs which provided feedback loops, and by inserting inductive ferrite beads around the base terminal at the test socket. This was done by trial and error and required adjustment as the gain of the device was improved.

The new test set design also incorporated a low pass filter between the test device and the test set circuitry. In this way transients from power lines, power supplies and switches of sufficient magnitude and duration to cause failure were eliminated. In addition, static charges from test set operators were dissipated by use of grounding bracelets. The same approach used to eliminate oscillations and transients in the test sets was used also in designing the aging facility. No catastrophic shorts or opens resulting from transients were observed for evaluation samples tested and power stressed using this equipment.

Determination of "Safe" Operating Region

Background

Sikora and Miller[5] have done extensive work to determine the reasons for the differences between semiconductor device life predictions based on temperature stress results and those based on power stress results.[3,4] Their general conclusion was that much of the discrepancy could be eliminated if devices were operated in the region below h_{FB} (common base current gain) equal unity. If devices are operated above this region, corrections must be made for power dissipated in the emitter and for changes in the thermal resistance. Although their work was done using diffused silicon mesa transistors, it is reasonable to suppose that the same conditions would apply to germanium transistors.

Experimental Approach

Ten transistors, randomly chosen from production devices, were used to generate an h_{FB} equal unity curve. The devices were placed in a specially constructed test circuit, Figure 2, and the V_{CB} (collector-base voltage) was varied for fixed values of $|I_C|$ (collector current) until h_{FB} equalled unity; i.e., until I_B equalled zero. During the tests, the collector terminal was monitored to insure that no oscillations occurred. This was necessary because the base current usually reverses when the device oscillates, erroneously indicating an h_{FB} greater than unity.

The results of these tests are illustrated by Figure 3 which contains a plot of the distributions of $|V_{CB}|$ versus the various set values of $|I_C|$ for the parameter h_{FB} equal unity. The plot exhibits a definite nonlinearity throughout, as the 175mw isopower curve illustrates. The power capability of the device is seen to decrease drastically as $|V_{CB}|$ is increased. The plot shows that misleading power stress results might be expected if the device is operated above a $|V_{CB}|$ of about 8 volts to supply power greater than about 150mw.

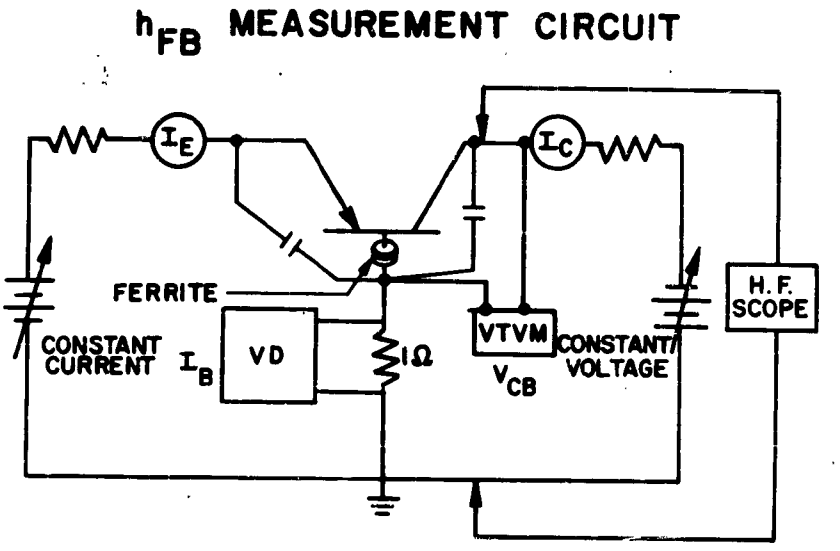


Figure 2

Dominant Failure Mechanism for Operation in $h_{FB} >$ Region

The reason for the decrease in the device power capability with an increase in $|V_{CB}|$ can best be understood by reviewing the device geometry shown on Figure 1, and the device impurity profile shown on Figure 4.

The three epitaxial layer impurity profiles shown on Figure 4 were measured using differential capacitance-voltage techniques. The base impurity profile was calculated using the complementary error function and the nominal measured sheet resistance and junction depths. The depth of the alloyed emitter was calculated using the weight of aluminum evaporated over the device emitter area and the temperature of alloy. The difference in the three epitaxial layer profiles shown results from variations in the epitaxial layer thicknesses. It is mainly this variation that causes the relatively wide distributions in $|V_{CB}|$ for a given $|I_C|$ at h_{FB} equal unity as shown on Figure 3.

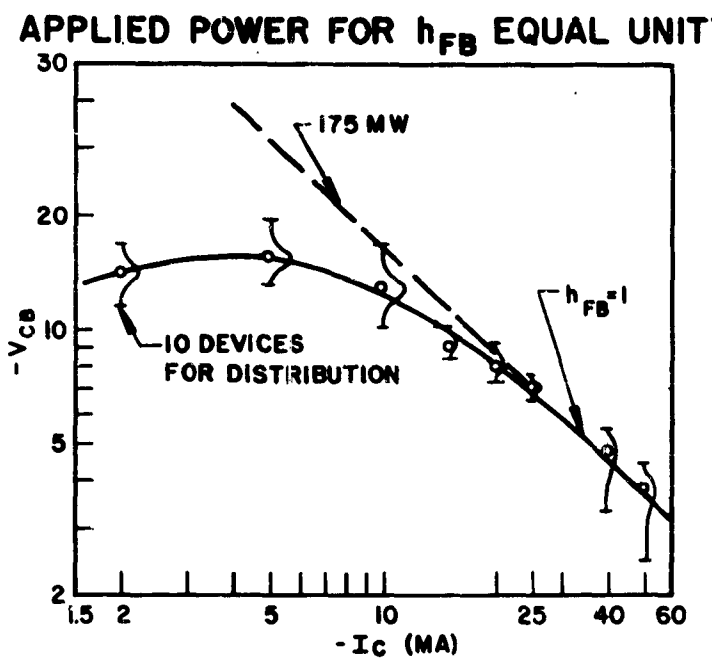


Figure 3

Three different values of $|V_{CB}|$ required to extend the collector depletion region to the epitaxial layer-substrate interface for three different devices are shown on Figure 4. When the collector side of the depletion region reaches this point, additional increases in $|V_{CB}|$ will result in a rapid extension of the depletion region into the base. Using the average of the impurity concentrations shown on Figure 4, and again applying depletion layer equations, one will find that for a V_{CB} = -4 Vdc 0.126μ of the base region and 0.9μ of the collector region have been depleted. Thus about 2/3 of the base thickness is swept free of majority carriers at this low voltage. Any further increase in $|V_{CB}|$, therefore, would result in a severe increase in the transverse base resistance which causes an increased voltage drop in the semiconductor between the base and emitter and particularly under the emitter. Consequently, the emitter starts to severely debias at the center which causes extreme current crowding beyond this $|V_{CB}|$.[9-12]

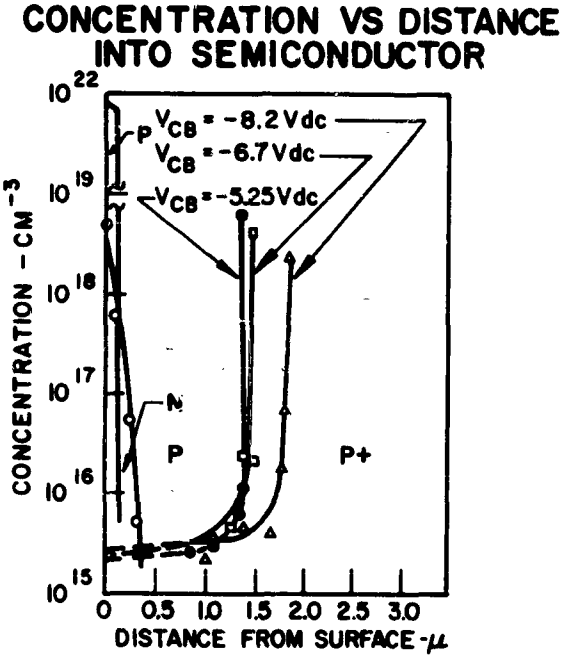


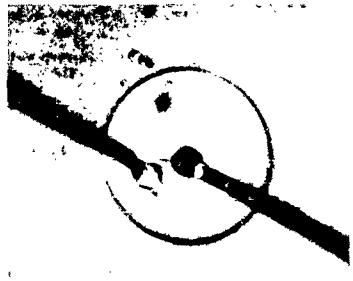
Figure 4

For this reason, if one chooses to maintain a constant I_E and step $|V_{CB}|$ to increase the power stress level, as $|V_{CB}|$ increases beyond about 8 volts, severe emitter current crowding will occur. For the same amount of current (constant I_E) passing through the much reduced emitter area, the emitter current density will become excessively high; i.e. greater than 15,000 amp/cm^2 at I_E = 10ma. The result is thermal runaway, intrinsic conduction, and finally failure due to local melting.

A failure caused by this mechanism is recognized by a surface irregularity which results from the melted region. The irregularity is the same as was observed for devices that failed as a result of sufficiently high energy oscillations. Figure 5 is a photomicrograph of such a failure.

The melt occurs between the base ring and the emitter dot adjacent to, or under, the base wire bond for this device. It occurs here because this is the line of least resistance during the extreme emitter current crowding. The melt usually originates at the base gold wire bond and extends towards the emitter during this period since the Au-Ge eutectic temperature (356°C) is the lowest melting point of the structure. It is postulated that failures of this nature resulting from oscillation occur when the oscillations are of sufficient amplitude and frequency to cause an excursion of operating voltage beyond the "safe" voltage during the negative portion of the cycle. When the RMS power dissipated in the "unsafe" region is sufficient to cause thermal runaway, failure occurs in a manner similar to the DC case.

CURRENT DENSITY FAILURE



BIAS SETTING

V_{CB} = -8.6 VDC

I_C = -14.8 MADC

P_T = 127 MW

Figure 5

By eliminating the capability of the test sets and the aging facilities to oscillate, and by selecting points within the "safe" operating region, these type failures will be eliminated. For example if one chooses a constant $|V_{CB}|$ of 5 volts and steps the current, the device can be safely stressed to about 150mw. (By "safe" it is meant that only normal operating failure mechanisms will be

accelerated). As stated previously, the "safe" operating region will vary from device to device depending on the design. This is true regardless of whether the device is silicon or germanium.

Power Stress Evaluation

Criteria of Failure

The failure criteria used for this evaluation are as follows:

1. $I_{CBO}(V_{CBO}= -10Vdc) > 3\mu Adc$(initial $I_{CBO} \leq 1.5\mu Adc$)
2. $\Delta h_{FE}(I_C= -5mAdc, V_{CB}= -5Vdc) > 25\%$(initial $h_{FE} \geq 8$)
3. $\Delta C_{ob}(V_{CBO}= -5Vdc, f=100KC) > +25\%$(initial $C_{ob} \leq 0.6pf$)

(I_{CBO}= collector to emitter current, base open)

(V_{CBO}= collector to base voltage, emitter open)

(C_{ob}= collector to base capacitance, emitter open)

These characteristics were selected on the basis of performance of this device and a similar device (2N559) during temperature stress experiments, and because, by correlation, their stability will determine the performance of the device in the field. The allowed drift in I_{CBO}, however, is more restrictive than would be required for good field performance. The allowed range in h_{FE} is sufficient, however, to cause field failures.

The collector capacitance was monitored for initial samples tested in the hope that channels could be detected. This proved ineffective, however, since, in general, increases in I_{CBO} resulted from general collector-base junction softening rather than from channeling.

Experimental Approach

Five sample groups of devices were randomly selected from regular production lots. Two of the sample groups were set for long term, low level fixed stress aging. The remaining three sample groups were set for intermediate to high level power step stress[1,2] aging. Table I, below, lists the pertinent information related to the sample groups.

TABLE I

Power Stress Summary

Sample	Sample Size	Variable	Operating Pt.	Applied Power	Equiv. Temp*
1	69	Time	$\|I_C\|$=10mAdc, $\|V_{CB}\|$=5Vdc	50mw	90°C
2	52	Time	$\|I_C\|$=15mAdc, $\|V_{CB}\|$=5Vdc	75mw	123°C
3	20	Stress 4 hr Steps	$15mA \leq \|I_C\| \leq 30mA$	75-150mw	
4	20	Stress 20 hr Steps	$\|V_{CB}\|$ = 5Vdc	Δ=25mw	123-220 C
5	20	Stress 100 hr Steps			

*$T_J = \Theta_{JC} P_A + T_A$, Θ_{JC} = thermal resistance = 1.3°C/mw,

T_A = temperature of ambient = 25°C.

As the table indicates, sample 1 and 2 were held constant at 50 and 75mw and measured periodically. Samples 3, 4 and 5 were stressed beginning at 75mw and increased to 150mw in increments of 25mw. They were at these stress levels for periods of 4, 20 and 100 hours, respectively. The equivalent junction temperature was arrived at using a measured Θ_{JC} of 1.3°C/mw (measured at I_C = 5mA and V_{CB}=-5V). This is justified since the operating points did not exceed the point where h_{FB} equals unity; therefore, the thermal resistance should not vary significantly from the measured value.

Analysis of Results

The data obtained from these tests were plotted to determine the failure distribution of each group. Samples 1 and 2 were plotted using coordinates of log time versus percent failure whereas samples 3, 4 and 5 were plotted using coordinates of linear applied power versus percent failure. The data points were easily fitted by a straight line in each case. Table II, below, contains a tabulation of the time of failure, the applied power and the equivalent junction temperature for the five sample groups power stressed. The 10 and 50% failure points are included.

TABLE II

Power Stress Failure Rates

	10%			50%		
Sample	Time of Failure	Power	Equiv. T_J	Time of Failure	Power	Equiv. T_J
1	$1x10^4$hrs*	50mw	90°C	$1.5x10^5$hrs*	50mw	90°C
2	$3.5x10^3$	75	123	$5x10^{4*}$	75	123
3	4	105	162	4	141	208
4	20	83	133	20	119	180
5	100	70	126	100	122	184

*Extrapolated values

The failure modes observed were increased I_{CBO} and decreased h_{FE}. Eighty-eight percent of the failures occurred because of high I_{CBO} and 12% because of low h_{FE}. No catastrophic failures such as described previously were observed, although in isolated cases (because of variations in gain) the devices were operated within the range which separates the "safe" from the "unsafe" operating region. No significant differences were observed between these devices and the others as far as failure mode and frequency of failure is concerned. These results, therefore, indicate that the first two objectives of this program had been met. That is, the capability to safely test and power stress the device without causing catastrophic failures.

The values of time and equivalent junction temperatures listed in Table II are used to construct a 10% and 50% acceleration curve. The curves are shown on Figure 6 along with an extrapolated 1% acceleration curve. Coordinates of linear reciprocal degrees Kelvin and log time in hours are used. The points obtained from the five sample groups used describe reasonably well defined straight lines. The activation energy obtained from the slope of the lines is about 1.6 ev. From these curves the extrapolated 10% and 50% life at the rated temperature (100°C) is about 8×10^3 and 1×10^6 hours, respectively. At 30mw, the maximum power dissipated by the device in the field at 25°C, the expected life is about 6×10^6 and 4×10^8 hours for 10% and 50% failure, respectively.

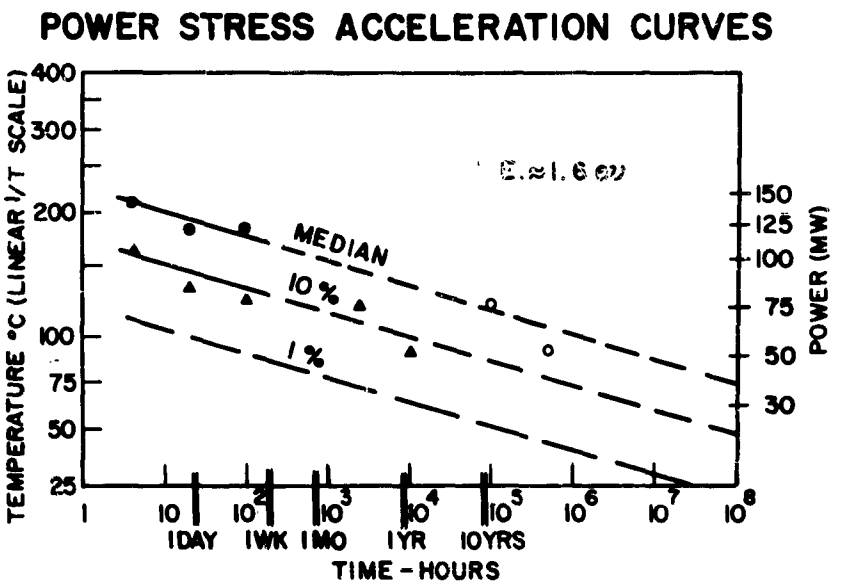


Figure 6

Comparison of Results With Other Reports

For a general comparison, the activation energy and median life at 100°C are listed in Table III along with data reported by Dodson and Howard[1] and by Peck[2], who used temperature step stress to determine expected life of a similar device (2N559).

TABLE III

Report	Activation Energy, ev	Median Life at 100°C	Encapsulation Process
Peck	1.0	10^4	vacuum baked
Dodson & Howard	0.9	3.5×10^5	vycor getter
This Data	1.6	10^6	SiO_2 coated wafer barium oxide getter

The differences in results listed in Table III are attributed to improvements made in the techniques used to fabricate this type device. For example, Peck's data was obtained using devices which were vacuum baked prior to encapsulation and then encapsulated without a moisture getter. Dodson and Howard's work was done after the vycor getter had been introduced. As stated previously, the device wafer used here was coated with SiO_2 prior to encapsulation and then encapsulated with a BaO getter.

Although these data were necessarily obtained using small samples, the use of five such samples and both short term and long term power aging lend confidence to the results. A comparison of the results with temperature stress data obtained for this device lends additional support.

Long term fixed temperature and short term step temperature experiments were conducted. Figure 7 is a plot of the resulting median acceleration curve and the power stress median acceleration curve shown previously on Figure 6. For purposes of comparison, the median temperature acceleration curve was drawn parallel to the median power stress acceleration curve. As seen, the line provides a reasonable fit to the points. The activation energy obtained from the two acceleration curves is, therefore, considered the same, and since the observed failure modes for both temperature and power stress were alike, the failure mechanism is also considered the same.

It is interesting to note the difference in levels between the median power and the median temperature acceleration curve. Recall that this difference resulted even though no abnormal failure modes[4,5] were observed among the failures. Therefore, even after extensive precautions are taken to eliminate all but normal failure modes, the predicted life at a given temperature may be significantly less when power stressed than when temperature stressed. Consequently, it must be concluded that the associated electric field enhances the failure mechanism. For any particular device design, therefore, the electric field (applied voltage) should be kept constant (and in a "safe" region) during power stress experiments. This will insure a constant electric field stress for each sample group of the experiment. The actual electric field enhancement of the failure mechanism, however, was not determined during this experiment.

The mechanism which caused failure of this device during power and temperature aging is believed to be diffusion of gold from the base contact through the 0.15μ base layer to the collector-base depletion region, or migration of gold or silver or both under the SiO_2 and down the mesa to the collector base junction. Such migration has been observed. The gold diffusion through the base probably takes place along strain lines set up in

the base layer when the emitter region is alloyed. These beliefs are supported by the 1.6 ev activation energy resulting from these tests.

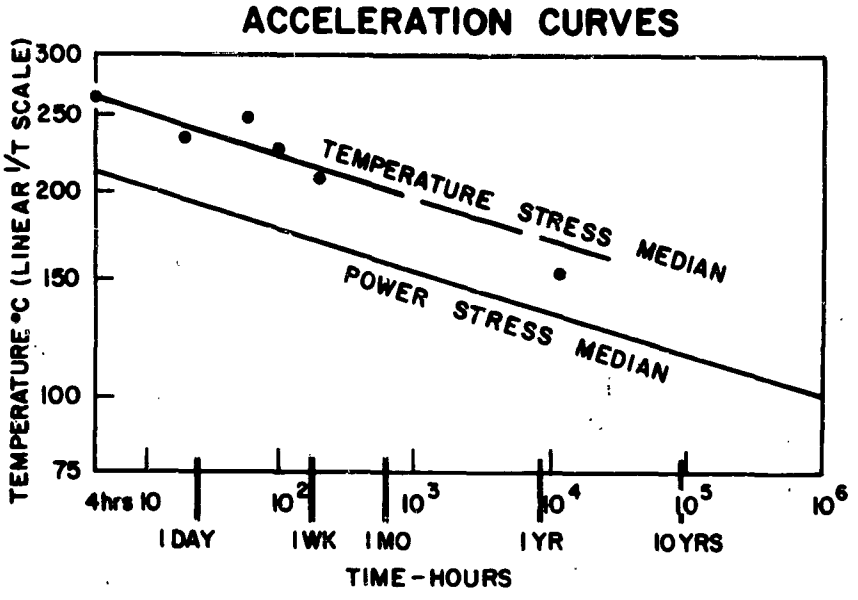


Figure 7

"In Service" Results

Twenty thousand of these high frequency germanium transistors have been in service for over a year. To date, from these twenty thousand devices, only eight "in service" failures have been observed. These eight, through failure analyses, were traced to failure of another component in the system. The characteristic molten region shown in Figure 5 appeared on each device. This type failure is not normally observed when the device is operated at rated conditions. It would only be expected to occur if the emitter current density is greatly increased as discussed earlier.

An extrapolated 1% power acceleration curve was plotted, Figure 6, in an effort to determine when "in service" failures should start occurring. Under maximum "in service" operation conditions (30mw(max); 64°Cmax) the curve predicts that 1% of the devices (200) should have failed after 15 months of operation. To date, however,

after about six months of actual "in service" operation no known normal failures have occurred. This was expected, however, for two reasons. First, devices which would have failed the $I_{CBO} > 3\mu Adc$ end point will not necessarily appear as field failures, and secondly, few, if any, of the devices are operated at maximum ratings.

Conclusions

The conclusions drawn from the work reported herein can be summarized as follows:

A) Meaningful predicted life expectancy of germanium devices can be determined from power stress results if test sets and aging facilities do not introduce failures, and if the "safe" operating region of the device is predetermined and operation is kept within this region;

B) predicted semiconductor device life based on power stress results are not necessarily the same as when based on temperature stress results. It was found not to be the case for the germanium device used for this work even after extensive precautions were taken to eliminate abnormal failure mechanisms. Power stress results yielded about 2 magnitudes shorter life for a given temperature than did temperature stress results. The activation energy in both cases is considered the same, about 1.6 ev;

C) the magnitude of the difference between the temperature stress extrapolated median life and the power stress extrapolated median life predictions depends on the device design. For any given design, voltage, i.e., the electric field, should be held constant. This will insure a constant field enhancement of the failure mechanism and the same transverse base resistance for each sample group;

D) temperature and power stress techniques should be used during any initial device reliability evaluation program to determine the difference in magnitude of the two acceleration curves. If a consistent difference occurs, (i.e., same activation energy but different life prediction) the easier temperature stress technique can then be used with the necessary correction.

ACKNOWLEDGMENTS

The author wishes to extend his thanks to Messrs. G. A. Dodson and L. E. Miller for their many helpful suggestions during the preparation of this manuscript.

BIBLIOGRAPHY

1. G. A. Dodson and B. T. Howard, "High Stress Aging To Failure Of Semiconductor Devices", Proceedings of Seventh National Symposium On Reliability And Quality Control, pp 262-272 (January 1961).

2. D. S. Peck, "A Mesa Transistor Reliability Program", The Solid State Journal, pp 25-30 (November/December 1960).

3. C. H. Zierdt, Jr., "On The Importance Of Operating Life Tests As Compared To Storage Tests of Transistors", The Solid State Journal, pp. 21-27 (September 1961).

4. J. Partridge, "On The Extrapolation Of Accelerated Stress Conditions To Normal Stress Conditions Of Germanium Transistors", Physics of Failure in Electronics, Vol. 2, pp. 208-225 (September 1963).

5. G. C. Sikora and L. E. Miller, "Application Of Power Step Stress Techniques To Transistor Life Predictions", Physics of Failure in Electronics, pp 30-41, Vol. 3, (1964).

6. G. A. Dodson, "Step-Stress Aging Of Diffused Germanium Transistors - A Process Study", Bell Telephone Eng. Service on Transistors, 3rd Interim Report, Contract, DA36-039 SC85352, pp. 12-22, (February 28, 1961).

7. H. F. Cooke and R. R. Webster, "Packages For Small-Signal VHF and UHF Transistors", Electrical Design News, Vol. 9, No. 11, (October 1964).

8. S. J. Anastasio, W. E. Beadle, BTL-LD, private communication.

9. N. H. Fletcher, "Some Aspects Of The Design Of Power Transistors", Proceedings of the IRE, Vol. 43, pp. 551-559 (May 1955).

10. N. H. Fletcher, "Self Bias Cut-Off Effect In Power Transistors", Proceedings of the IRE, Vol. 43, p. 1669 (November 1955).

11. N. H. Fletcher, "The High-Current Limit For Semiconductor Devices", Proceedings of the IRE, Vol. 45, pp. 862-872 (June 1957).

12. R. Emeis, A. Herlet and E. Spenke, "The Effective Emitter Area Of Power Transistors", Proceedings of the IRE, Vol. 46, pp 1220-1229, (June 1958).

FAILURE MECHANISMS OF ELECTRONIC COMPONENTS

H. F. Church and B. C. Roberts

The Electrical Research Association
Leatherhead, Surrey, England

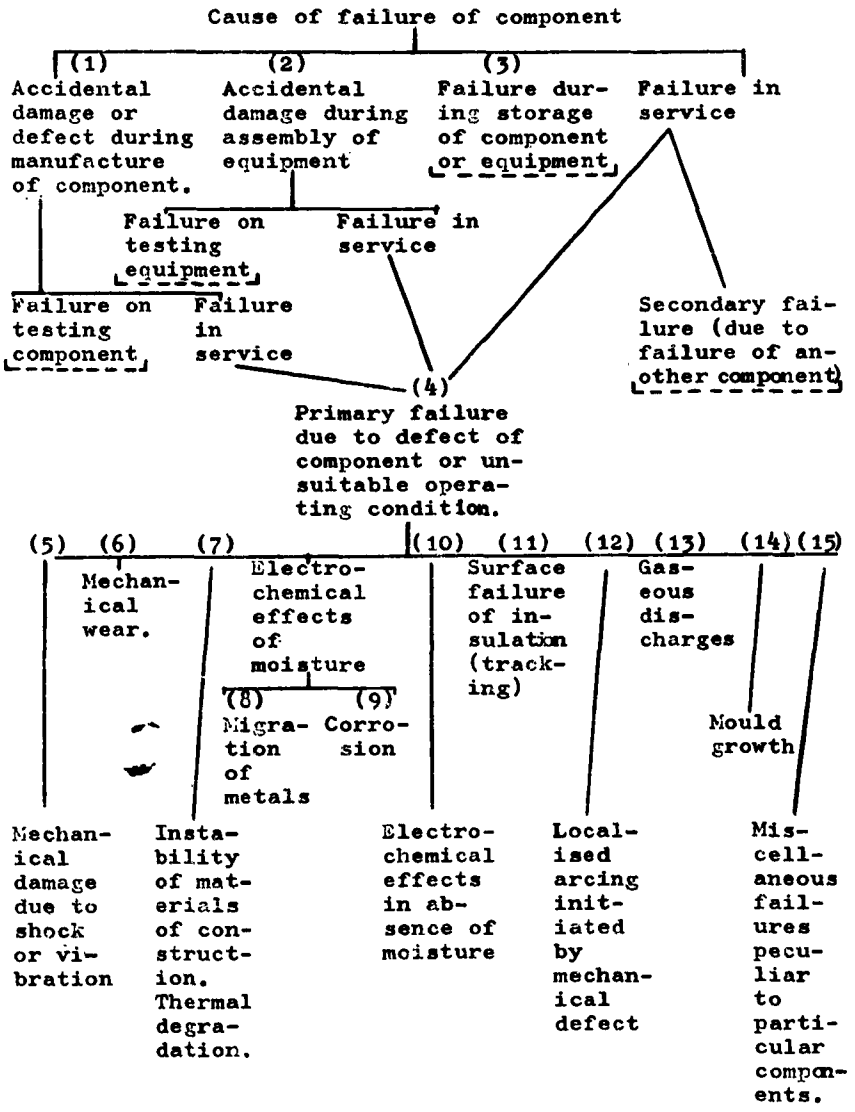
Introduction

The Electrical Research Association has for many years been investigating the causes of failure of electronic components. Long term tests under fixed conditions have been made on most types of resistor and on impregnated paper capacitors; short term tests have been made on resistors under changing conditions. The effects of storage on a wide range of components have been investigated. Components derived from AGREE testing and from field service have been examined. This work falls within the Materials Department of E.R.A., whose general concern is with the behaviour of insulating materials and dielectrics under electrical stress. The work is supported and guided by member companies and by the Ministry of Aviation.

This paper is a review of E.R.A.'s work on mechanisms of failure in components. An attempt has been made to assemble the many different mechanisms into a few groups and to point out the relationships between these groups where they exist. The chart in Fig.1 shows these groups and relationships, and can be used also as a table of contents for the paper. The number against each group refers to the relevant section in the text.

This programme of research has occupied many years, and the types of failure observed do not necessarily occur to the same extent in current production, the manufacturers having profited by the results of the research as they became available.

NOTE: Section 4 is missing in this paper.



Classification of mechanisms of failure in Electronic Components

Fig. 1.

1. Accidental damage or defect during manufacture of component.

Examples of failure due to defects in manufacture are drawn from life tests of resistors and capacitors, from AGREE testing of equipment and from field service. The latter two sources also provided a few failures of other components.

In film resistors, scratches in the film and flaws in the substrate have led to failure by local overheating and eventual arcing at the discontinuity. The consequences of this defect are dealt with in detail in Section 12.

Poor cutting of the spiral groove has occasionally been found in metal film resistors. Fig.2 shows a partial bridge across the groove where the cutting tool nearly left the surface. This defect caused instability on load and a high noise level.

Figure 2

Bridge across groove in metal film resistor, caused by cutting tool leaving surface.

A rare defect is the cracking of the brass end cap of a resistor which led to electrochemical attack in a moist environment. Examples are given in Section 9.

Silver, which is used by some resistor manufacturers for making connections, is occasionally applied accidentally to the resistive film. Here it causes an unstable path for the current in parallel with the film, and, in moist conditions, can migrate to other parts of the resistor (see Section 8). An example of careless application of silver is given in Fig.3 together with an electro-

graphic print in Fig.4. In Fig.5 is shown a speck of silver bridging a groove. The silver was applied before the groove was cut, but evidently survived the passing of the cutting tool and was pressed into the bottom of the groove, making a fairly permanent conductive bridge. The alternate heating and cooling during the subsequent life test caused the bridge to open and the resistance to rise by about 10%.

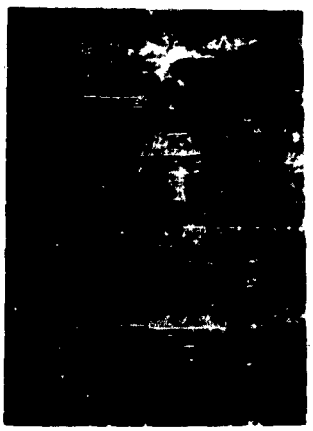


Figure 3.

Figure 4

Figure 5

Fig.3: Silver accidentally applied to body of film resistor.
Fig.4: Electrographic print, made with potassium chromate of silver shown in Fig.3. The dark areas correspond to the light, silver, areas of Fig.3.
Fig.5: Silver bridging groove in film resistor. Silver in groove was not removed by cutting tool.

Fig.6 is an X-ray photograph of two film resistors with ceramic substrate which failed during cyclic load. One of the leads in each resistor had been pushed too far into the ceramic tube during assembly. Heating, or a momentary surge of voltage probably initiated an arc which led to a short circuit.

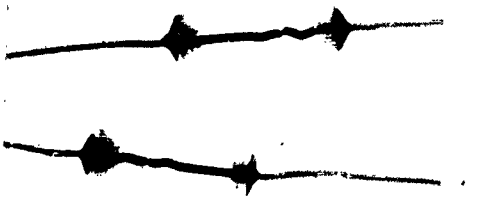


Figure 6

X-ray photograph of 2 film resistors, showing leads causing short circuit

Inadequate sealing against moisture has been a frequent cause of failure among resistors on test. The electrochemical effects of moisture penetration are dealt with in Sections 8 and 9. Generally the resistors with a rigid protective coating of moulded plastic or synthetic resin admit moisture eventually between the case and the leads. Resistors protected with lacquer may eventually admit moisture through the sides as well as through the ends. Thermal cycling either in service or on test may open cracks through which moisture can enter. In one type tested at E.R.A. moisture evidently entered at the ends while the environment was cool and moist, but became trapped when load was applied. Failure was very rapid.

In paper and plastic film capacitors, conducting particles can cause local electrical stress concentrations which lead to breakdown of the dielectric. The effect of ionizable impurities introduced during manufacture into impregnated capacitors is dealt with in Section 10.

A gross manufacturing defect in a silvered mica capacitor is shown in Fig.7. The rectangular deposits of silver were applied asymmetrically to opposite faces of the mica, so that at one corner they nearly overlapped the mica and touched. Breakdown took place at this corner. Fig. 8 shows a crack in a sheet from another silvered mica capacitor which led to breakdown.

Figure 7

Silver applied asymmetrically to mica in capacitor, causing short circuit at lower left corner.

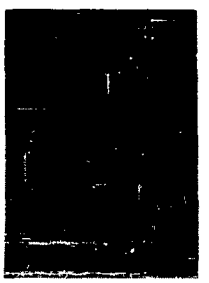


Figure 8

Crack in sheet of silvered mica capacitor, causing short circuit.

Fig. 9 shows part of an aluminized plastic film capacitor, in which external connection was made by spraying the edges with zinc. Some of the zinc had penetrated too far and, at another part of the capacitor, made contact with the aluminium of the wrong electrode.

On a small AF transformer a solder tag was incorrectly positioned and because of vibration eventually touched a vacuum tube screening can.

Figure 9

Penetration of sprayed zinc into metallized plastic capacitor. The invading zinc appears as the streaks on the left-hand side.

A tap on the inside of a toroidally wound RF transformer was carelessly soldered and, under vibration or heat cycling, touched an adjacent turn and put the circuit out of tune.

In a silicon Zener diode the wire connecting the crystal to a termination was severed but made intermittent contact. Also a short piece of the same kind of wire was found loose in the case. Thus by appropriate knocking or shaking it was possible to make the device function correctly, have an open circuit, or a short circuit.

2. Accidental damage during assembly or repair of equipment.

Serious damage at this stage would usually be detected by pre-service testing. The fault illustrated in Fig.10 did, however, escape detection at this stage. The photograph shows a wire-wound resistor, originally coated with lacquer, with three adjacent turns evidently damaged by a sharp object such as a screwdriver. The wire was flattened at the point of impact and broken at one point. The wire must have developed a hot spot at the point of damage and eventually burnt out.

Figure 10.

Damage to wire-wound resistor during assembly. Extreme left wire is broken. Two others are damaged.

Soldering is a hazard to thermally sensitive components, which can result in failure after the equipment has been put into service. An example of this failure is provided by the small silvered mica capacitors mentioned in Section 5. The terminal wires were partially detached from the mica stack during soldering, and became disconnected during service.

3. Failure during storage of component or equipment.

For the past four years E.R.A. has been engaged on a programme of component storage. Five identical sets of components are under different storage conditions. The range of components includes fixed resistors and capacitors, potentiometers, relays, plugs and sockets, transformers, potted circuits, and semiconductors. The storage conditions are

(1) Laboratory condition, protected only from dust;

(2) Laboratory temperature, doubly wrapped in polythene and desiccated;

(3) Laboratory temperature in dry air;

(4) 0°C, desiccated, as in (2);

(5) -20°C, desiccated, as in (2).

The results which have so far emerged from these tests are as follows :-

Carbon composition resistors have lost moisture and fallen in value in the dry conditions at laboratory temperature (Nos. 2 and 3 above). This is not a serious effect, since it takes place in any case when the resistors are loaded, and is reversible.

Some electrolytic capacitors have also lost moisture in the dry conditions at laboratory temperature (Nos.2 and 3 above). This loss is accompanied by loss of capacitance and increase in loss tangent. Whether the effect is eventually reversible has not yet been determined, but if the capacitors were put into service straight from dry storage they would be defective.

Some paper-foil and metallised-paper capacitors in the first storage condition (only protected from dust in the laboratory) have absorbed moisture with consequent increase of capacitance and increase in loss tangent. The sealing against moisture is evidently inadequate. These would be classed as defective.

Among the other types of resistor and capacitor, and all the other components, there are no significant differences which can be attributed to the storage conditions. It is now proposed to load the components for 10,000 hours to discover if the storage conditions have had any delayed effect upon performance.

In addition to the components in the above programme, E.R.A. has examined components which have failed in Service stores. The behaviour of one type of carbon composition resistor is described in Section 7.2. Large numbers of vitreous enamelled wire-wound resistors were also found to be open circuited after storage in an un-

heated brick building. The failure was caused by moisture penetrating to the resistance wire down minute cracks in the glaze[1]. Contact EMF's within the resistor then initiated electrochemical erosion of the wire. The pattern of crazing was revealed by an electrographic technique using photographic emulsion pressed on to the resistor body. Fig.11 shows two prints obtained by this method.

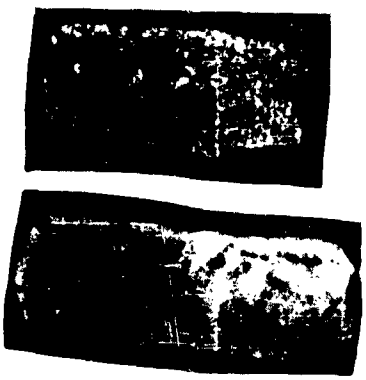


Figure 11

Pattern of crazing in vitreous-enamelled wire-wound resistors. Print was made electrographically.

5. Mechanical damage due to shock or vibration.

AGREE tests were carried out on batches of two British communication equipments, one for use in Army tanks, the other for use in R.A.F.aircraft. The equipments were repeatedly put through cycles of rapid temperature change and operated periodically, when hot, with vibration. Components were replaced as they failed. E.R.A. examined all the faulty components except vacuum tubes, and found that most failures had been caused by vibration. The more interesting examples of this type of failure are given below.

Several resistors and small capacitors and many of one type of RF choke had broken leads resulting from inadequate support. A few cable-forms became detached for the same reason. In one small metallised paper capacitor, one lead was pulled off by the vibration. In an electrolytic capacitor an internal connection was broken because bending stresses were transmitted through the rubber end seal.

A high proportion of small silvered mica capacitors failed by loss of capacitance or instability because the terminal wires which were crimped around the ends of the mica stack were loosened during soldering and later detached by the vibration. Thus there were three causes of failure: the capacitor was very easily damaged during soldering; this weakness was not recognised and countered by use of a heat shunt during assembly; vibration further weakened the terminal connections.

In a large electrolytic capacitor an open circuit was found between a solder tag and the rivet which attached it to the insulated top and connected it internally to the capacitor foil. Vibration probably loosened an already poor contact.

A trimmer capacitor with air dielectric failed because the two sets of vanes touched. The stationary set of vanes were supported by a single pillar which depended for stability on how tightly it was held in a hole in the ceramic base. Vibration caused failure in an inherently weak component.

A voltage sensitive relay failed because the plastic base, on which the whole of the inner assembly was mounted, broke.

Several sealed relays caused trouble in a keying circuit because their contacts did not open and shut simultaneously. Vibration had put them out of adjustment.

A very high proportion of vibrators failed. The fault in every case was due to breakage in the same place of one or both of the leads joining the operating coil to base pins. The breakage occurred at a point in the wire where it underwent oscillatory bending during vibration. Re-routing the wire from coil to pin might have prevented this failure.

Several neon indicator lamps became short circuited because the electrodes were not rigid enough to withstand the vibration.

P.T.F.E. retains its shape and excellent electrical properties even at a temperature of 250°C. It has however a low coefficient of friction and cannot be relied upon to grip smooth metal parts firmly under vibration. This short-coming led to the loosening of the reed of a coaxial relay and the central members of several coaxial plugs and sockets.

6. Mechanical wear.

E.R.A.'s main contribution to the subject of mechanical wear in components has been the study of contact

erosion in light-duty electromechanical relays[2]. One investigation was made of the way in which erosion at break is affected by the circuit inductance[3]. Below about 0.2μH the transfer is from anode to cathode, and the amount is independent of the inductance. At higher inductances the direction and amount of transfer depends on the type of discharge.

The volume of transfer is not the only measure of its effect on contact operation; equally important is the shape of the pip and crater formed as a result of the transfer. A tall pip is more likely to cause mechanical locking than a flatter pip of the same volume[4]. E.R.A. investigations have shown that pip steepness increases with the number of operations, and decreases with the radius of curvature of the contact surfaces. The use of nearly flat contacts will therefore delay failure. Pips formed by high-conductivity metals like gold and silver are steeper than those formed by poorer conductors.

By correct choice of circuit inductance and contact material it is possible in many cases to reduce erosion to a satisfactorily low rate. E.R.A. is also, however, investigating more radical ways of reducing erosion by the use of special contact metals and contacts of dissimilar metals.

7. Instability of materials

Some materials used in the construction of components may be chemically unstable and slowly degrade until the characteristics of the component are seriously impaired or even until failure occurs. Chemical instability is more serious at high temperatures and is therefore more likely in components such as resistors which generate heat during operation.

7.1. Thermal instability

An example of thermal instability was an early type of metal oxide film resistor (no longer made) which was protected by an unsuitable lacquer coating. This coating gradually carbonised, became electrically conductive, and caused a reduction of the resistance value by shunting the resistive element. Resistors of 47 KΩ (2 watts) and 100 KΩ (1 watt) fell rapidly in resistance after about 7,000 hours on full load. 47 KΩ (1 watt) resistors endured 15,000 hours before the deterioration of the varnish affected the resistance. 100 Ω (1 watt) and 200 Ω (2 watt) were unaffected even after 15,000 hours on load.

Carbon composition resistors also degrade thermally when appreciably loaded. The resistive element in this case consists of a rod or layer of synthetic resin loaded

with fine carbon particles. During long continued operation under load, the resin decomposes at the high operating temperature and shrinks, thus increasing the number of contacts between carbon particles and causing the resistance to fall[1]. At constant voltage a fall in resistance results in an increase in current and eventual thermal instability. This effect may become serious after 10,000 hours' operation under full load but the phenomenon is not an electrochemical one. The same effect occurs if this type of resistor is heated excessively in an oven without voltage.

7.2. Effects of moisture

Moisture, even in the absence of voltage, can adversely affect components. In damp atmosphere, carbon composition resistor elements absorb water, the synthetic resin binder swells, and the resistance value increases. In some cases a doubling of resistance occurred from this cause after a few years' storage[1]. On the other hand, electrolytic capacitors which rely on a moist conductive electrolyte may slowly dry out if operated at too high temperature or maintained in exceptionally dry conditions, and, as a consequence, fall in capacitance and increase in loss tangent. Some capacitors fell in capacitance by more than 30% in two years in a desiccated store at room temperature (See Section 3).

7.3. Incompatibility of materials

Sometimes materials used in the construction of components are chemically or physically incompatible. An example of chemical incompatibility occurred in a wire-wound potentiometer. The ceramic former contained a number of cavities in the surface which were filled by the manufacturer with magnesium oxychloride cement. The presence of this material was revealed by its alkalinity and its fluorescence in ultraviolet light. This in time corroded the resistance wire and caused an open circuit. In the manufacture of tin oxide resistors it is particularly important that the protective coating is chemically neutral since the resistivity of the film is changed, at high operating temperature, by both oxidizing and reducing environments.

In the construction of capacitors, electrically active impurities should not of course be present in the capacitor unit but it is also important that they should be absent from materials within the container, since they may slowly diffuse into the unit and cause an increase in the leakage current and dielectric loss[5]. Initial tests on the capacitor may not reveal their presence.

8. Migration of metals in the presence of moisture.

Silver is well known to migrate electrolytically in moisture films on insulating surfaces under d.c. stress. Its use in components is therefore a potential hazard; in resistors it can shunt parts of the resistive element, and in capacitors cause a short circuit. The hazard can be reduced by adequate protection against moisture.

Fig. 12 shows part of a metal film resistor in which one branch of a dendritic growth of silver nearly bridged the spiral groove. It was not clear how the speck of silver on the edge of the groove originated: it may have been accidentally applied during manufacture.

Figure 12

Dendritic growth of silver across groove of film resistor

In the porous lacquer coating mentioned in Section 9 not only silver but also compounds of copper, derived from the end cap, were found distributed along the body of the resistor and lodging in the spiral groove. Fig. 13 shows two craters in the lip of a copper end cap made by electrochemical attack. In Fig. 14 is shown a dendritic growth of a copper compound derived from the brass end cap of another resistor.

Figure 13

Electrochemical erosion of rim of end cap of film resistor. 3 small pits are visible. Cap was of silver-plated copper.

Figure 14

Dendritic growth of copper compound from brass end cap of wire-wound resistor, along vitreous enamel coating.

One type of resistor examined employed a turned brass end cap. The cap had resisted electrochemical attack even when uncoated with lacquer. In one case, however, the cap had evidently been cracked while being forced on to the body of the resistor. The metal exposed in the crack was severely attacked, and the products of attack spread down the body. This preferential attack in a crack has also been observed in a silver-plated copper end cap.

9. Electrolytic corrosion in the presence of moisture.

Failures due to erosion of the conducting film have been studied at E.R.A. in cracked carbon, tin oxide and metal film resistors. The conditions conducive to this erosion are

(1) A relative humidity of 95 - 100%.

(2) A direct potential across the resistor which is small enough not to drive moisture off by heating.

(3) A porous or poorly fitting coating.

Warmth or thermal cycling encourages the effects. Erosion proceeds wherever moisture connects parts of the film which are at different potentials.

In cracked carbon films erosion takes place at the anode where oxygen is released. The film is converted to carbon dioxide and removed. If a water droplet connects the film on opposite sides of a groove, erosion can be very rapid.[1] [6]

In tin oxide films erosion takes place at the cathode where nascent hydrogen is released and reduces the film to metallic tin. The tin thus produced is thinly deposited and too poorly attached to the substrate to be an effective substitute for the oxide from which it is derived. Fig.15 shows a point of erosion. The large light area is film and the large dark area part of the spiral groove. The dark semicircular area is metallic tin, and the small light areas bare substrate, from which the tin was probably stripped when the lacquer was removed.

Figure 15

Electrochemical erosion of tin oxide film.

In metal film resistors erosion takes place at the anode. Fig.16 shows erosion on one side of the groove in a resistor with a ceramic substrate.

Figure 16

Erosion of nickel-chromium film in moisture with d.c. erosion is on right (anode) side of groove.

Carbon, tin oxide, and metal film resistors of similar value and wattage from which the protective coating had been removed were immersed in distilled water and a small direct potential was applied. The carbon and metal film were eroded at about the same rate, but the metal oxide was eroded at less than one hundredth of this rate.

Although tin oxide is in itself much more resistive to erosion than carbon or metal the performance of all three types depends strongly on the effectiveness of the protection against moisture. The most effective (though not perfect) protection for general purpose resistors seems to be a combination of several layers of lacquer and a rigid outer case.

Open circuiting of vitreous-enamelled wire-wound resistors at the base of cracks or pores in the vitreous coating during storage in damp atmosphere has already been considered in Section 3. In this case, the electrolysis originates from small contact EMF's. This erosion process is greatly accelerated if a small d.c. voltage is applied between the terminations[1].

When a d.c. potential difference exists between adjacent windings (e.g., in a transformer) or between a single winding and a metal case (e.g., in transformers and chokes) corrosion of the wire can occur in moist situations if the insulation contains harmful hygroscopic impurities. Such impurities may be present in the wire covering, in interlayer insulation, in adhesive tapes used in the component and in wire sleeving. Spinning lubricants used with fibrous insulating materials are a serious cause of electrolytic corrosion. Tests are available for checking

the liability of materials to cause corrosion[7][8], and for the ionic content of insulation materials[5][9].

10. Electrochemical effects in the absence of moisture

Electrochemical corrosion owing to moisture films in components such as thin-film resistors, transformers, and chokes, has already been considered in Section 9. Electrolysis can also occur under dry conditions in insulation if impurities are present which dissociate in the material. These effects are more serious at high temperatures because of the increased ionic mobility and they are especially important under d.c. stress.

One example of this is the migration of sodium ions under d.c. stress in ceramics and glass at high temperatures. Ceramic or glass formers for wire-wound resistors must be of very low sodium content. If not, sodium migrates to the cathode end of the resistive element and can cause corrosion of the wire or film or detachment of the film.

Paper capacitors are thoroughly vacuum dried, impregnated and sealed during manufacture, but nevertheless may fail quickly under d.c. stress by an electrochemical process if impurities are present which dissociate in the dielectric. One class of impurity which has serious effects, especially with Aroclor impregnants, is soldering fluxes used when making connections or sealing capacitor cans. Any fluxes used should be removed before impregnation takes place. Rosin flux has been known to reduce the mean life of capacitors by 100 times[10]. The following evidence confirms the electrochemical nature of failure in low-voltage capacitors :- Under long maintained d.c. stress visible changes occur which are dependent on polarity. For example, with some impregnants, a polymer which is fluorescent in ultra-violet light, forms at the cathode whilst with aluminium electrodes the anode becomes oxidised. With metallised paper capacitors the anode may be gradually eroded away (with loss of capacitance) whilst the cathode remains intact. The rapid decrease of life with increase of temperature of paper capacitors is consistent with this deterioration process.

Some nickel-chromium film resistors sealed in a ceramic case were found to have eroded on full load (d.c.) at 70°C. Erosion had taken place mainly on the positive side of the spiral groove on the resistor. This erosion was evidently electrochemical in nature, although water was absent. The "electrolyte" in this case consisted of rosin in a viscous organic solvent used in soldering the connections.

11. Surface failure (Tracking)

Surface breakdown of organic insulation, known as tracking, can occur on exposed surfaces between electrodes maintained at a potential difference.[11] It can occur with less than 200 volts r.m.s. In electronic apparatus it may occur in moulded plugs and sockets and on exposed terminal boards. It is initiated by leakage currents in surface contamination; sparking in the surface films carbonizes the insulating material until a carbon track grows across between the electrodes. Phenolic mouldings and laminates, especially if the filling or reinforcement is cellulosic, are particularly susceptible to tracking, melamine and alkyd materials are fairly resistant, whilst silicones are almost immune.

12. Localised arcing initiated by mechanical defects

Open circuits have been found in film resistors at scratches in the film and flaws in the substrate. In the period of service before the open circuit occurred localised overheating was probably caused at the defect, which led to degradation of the film and eventually to rupture and arcing. Further damage was caused by the arc itself.

Examples of scratches in metal oxide resistors are shown in Figs. 17 and 18. In Fig.18 a scratch crossed the film track at three places and caused an open circuit at one of them. The dark material at the open circuit consisted of remnants of the film and of the protective lacquer. In a resistor similar to the one shown in Fig.18 an open circuit was caused at an existing scratch by overloading the resistor while it was under observation through a microscope. An alternating potential was applied, and, after the open circuit was created, the film was burnt away at equal rates on each side of the gap.

Figure 17
Scratches in film of a tin oxide resistor. No open circuit was caused.

Figure 18
Scratch and open circuit in film of tin oxide resistor.

In Fig.19 is shown another oxide film resistor with a flaw in the ceramic substrate causing an open circuit. Fig.20 shows an electrographic print of this resistor made in a reagent sensitive to metallic tin. The dark area shows the distribution of tin near the flaw. It will be seen that one edge of the dark area on the print corresponds closely to the flaw in the ceramic, and that the film has been burnt away on one side only of the flaw. It is thought that the tin oxide film was chemically reduced to tin by carbon derived from the plastic case by decomposition in the heat of the arc across the original flaw.

Figure 19.

Flaw in ceramic substrate of oxide film resistor.

Figure 20.

Electrographic print of resistor in Fig.19, made with cacotheline, showing extent of metallic tin. The left edge of the dark (tin) area corresponds to the shape of the flaw in Fig.19.

Another example of a discontinuity in the thin film of a resistor is shown in Fig.21. The conducting film is here nickel-chromium and the substrate ceramic.

Figure 21

Crack in surface of metal film resistor causing open circuit. The light bands at each side of the photo are the grooves.

13. Failure due to gaseous discharges

If components operate at high alternating voltage (greater than 250 V r.m.s.) discharges (corona) may occur in air or other gas included in the insulation or in the surrounding air. The gaseous ions in the discharge bombard the insulation, form breakdown channels and eventually produce complete breakdown[12]. Materials vary in their resistance to discharges: in general inorganic materials are more resistant than organic materials. This mechanism of failure cannot occur in components operating at low voltage, and even at high voltage it is slow if the applied stress is unidirectional. Discharges not only lead to breakdown but cause electrical disturbances. Dust or moisture on a component such as a transformer can reduce clearance distances and initiate discharges.

Absence of gaseous discharges in electronic components can be established in suitable cases with the E.R.A. Discharge Detector [13].

14. Mould Growth

Most component specifications include a test for susceptibility to mould growth. Common moulds are dependent for their growth on the presence of moisture and oxygen and on the existence of a suitable nutrient substance which may be a material used in the construction of the component or a contaminant such as grease or dirt. In moist air the thread-like mycelia of moulds are somewhat conductive electrically and may cause undesirable surface leakage over insulation. In addition, the host material may be badly degraded and chemically harmful by-products may be formed. In order to avoid mould growth the proper choice of materials and conditions of use are

important.

15. Miscellaneous failures peculiar to particular components

Occasionally failures occur which do not fall into any general classification. Three examples will be mentioned.

(i) Some resistors consisting of a fine wire winding coated with an organic varnish developed many open circuits during use. This was found to be due to gradual embrittlement of the varnish which cracked and imposed sufficient strain on the wire to break it in tension. Fig.22 shows three points of breakage, all under cracks in the lacquer.

Figure 22

Breakage of thin wires in brittle lacquer
The dark bands are the unbroken wires

(ii) High-K ceramic capacitors, in which the electrodes were fired-on silver, failed at the edge of one of the electrodes. This was because the applied high frequency stress was highly concentrated at the sharp edge and resulted in local heating, due to dielectric loss. Rapid degradation of the ceramic at the edge followed.

(iii) In polystyrene film capacitors it is important to avoid any material containing grease or oil. This, if in contact with the film plasticises it, and causes distortion and the formation of holes. Accidental sources of grease during manufacture have been found to be human hairs and dandruff. These must be scrupulously avoided since one such inclusion can lead to failure.

16. Discussion

The authors believe that high reliability of electronic components is attainable only if the causes of failure are understood and eliminated in advance by correct design and production control. A specified level of performance in short term tests cannot ensure reliability. It is essential to control every manufacturing process so that faults will not develop during service.

This paper attempts to summarise all known failure mechanisms (physical, chemical, electro-chemical and mechanical) which can affect conventional passive electronic components. The conclusions reached are based on many years of component and equipment testing and long experience of "trouble-shooting" in this field. Although the list of failure processes is long, failures are rare in the best components, provided they are properly used.

Very recent electronic devices, especially microcircuits of various types, inevitably have not yet received this long detailed treatment but many of the conclusions of this paper are of a basic nature and are likely to apply also to these new developments.

References

(1) H.F.Church: "Long Term Stability of Fixed Resistors" Proc.I.E.E., Vol.107, pt.B, pp.377-385; July, 1960.
also I.R.E. Trans. on Component Parts, Vol.CP-8, No.1, pp.31-40; March, 1961.

(2) W.Nethercot: "Contact Wear in Light Duty Electromagnetic Relays"; Co-operative Electrical Research (Journal of E.R.A.), No.15, pp.2-5; October, 1961.

(3) J.Riddlestone: "The Variation with Current and Inductance of Metal Transfer Between Platinum Contacts". I.E.E. Monograph No.133, 1954. (also as Proc.I.E.E., Vol.102, pt.C, p.29).

(4) J.Pullen: "Electrical Contact Erosion; A theoretical study of Pip and Crater Shape". (E.R.A. report Ref.G/T145, 1961.)

(5) Z.Krasucki and H.F.Church: "The Detection of Ionic Impurities in Insulation". Proc.I.E.E., Vol.109, pt.A, Supplement No.3; June, 1962.

(6) I.D.L.Ball and R.G.J.Butler: "High Stability Carbon-Film Fixed Resistors. D.C.Failure in Moist Atmospheres". E.R.A. report Ref.Z/T124; 1959.

(7) A.G.Day: "Testing for Electrolytic Corrosion Liability of Electrical Insulation". British Plastics, Vol.34, pp.491-494; September, 1961.

(8) British Standard 3924:1965. "Specification for Pressure Sensitive Adhesive Tapes for Electrical Purposes".

(9) British Standard 2689:1956. "Recommended Methods for Estimating the Effect of Deleterious Substances in Fibrous Electrical Insulating Materials (as amended 1960).

(10) H.F.Church: "Factors Affecting the Life of Impregnated Paper Capacitors", Proc.I.E.E., Vol.98, Pt.III, No.52; March, 1951.

(11) N.Parkman: "Electrical Breakdown by Tracking", Proc.I.E.E., Vol.109, pt.B, Supplement No.22, pp.448-453; 1962.

(12) J.H.Mason: "The Deterioration and Breakdown of Dielectrics Resulting from Internal Discharges", Proc.I.E.E., Vol.98, pt.1, No.109; January 1951.

(13) G.Mole and F.C.Robinson: "A Versatile Discharge Detector", Co-operative Electrical Research, No.21, p.8; April, 1963. (Journal of E.R.A.)

ACCELERATED AGING AND FAILURE MECHANISM ANALYSIS OF THIN TANTALUM FILM R-C NETWORKS*

M. Walker
A. McKelvey
G. Schnable
M. Sharp

Philco Corporation
Lansdale Division
Lansdale, Pennsylvania

ABSTRACT

Accelerated stress testing of tantalum thin-film resistor-capacitor networks was conducted to establish a means of predicting use level failure rates and to determine the major modes of failure associated with the networks. Also, a rigorous study was made of the physical mechanisms of failure.

Accelerated stress tests were performed simultaneously at a number of conditions involving elevated voltage and temperature as well as at use voltage and temperature levels. The relationships between accelerated stress failure rates and actual failure rates were investigated. The resulting modes of failure include: resistance changes and thermal runaway for the resistors, and leakage current increases and dielectric breakdown for the capacitors.

Failure mechanism analysis resulted in significant information on the causes of the observed modes of failure. In conjunction with this study, information on the basic structure of sputtered tantalum and anodically formed Ta_2O_5 was acquired through the use of electron diffraction and electron microscopy techniques. Plots of the rate of interaction of tantalum resistors with various gases were made at elevated temperatures. In air, the reaction rate was found to be parabolic. The effect of substrate ion migration on resistors was also studied. Ta_2O_5 dielectric conduction mechanism studies indicated ionic conduction. Several methods for the location of defects in the Ta_2O_5 capacitor

* The work reported in this paper was supported in part by the United States Air Force, Rome Air Development Center, under Contract No. AF 30(602)-3287.

dielectric films were evaluated; the best results were obtained with a copper plating technique. A detailed study of the defective regions indicated that crystallization of the amorphous Ta_2O_5 is the principal reason for localized defects.

It was concluded that the reliability of tantalum thin-film resistors is very good, and the failure mechanisms are well enough understood to permit prediction of resistor reliability under various stress conditions. Conservatively rated thin-film capacitors are reasonably reliable, but represent the limiting factor in R-C network reliability.

INTRODUCTION

The work described here was performed to develop techniques for accurate and economical accelerated testing of resistor-capacitor networks formed from tantalum and tantalum oxide films, and for analysis of the physics of failure of such networks. Accelerated testing, as governed by physics of failure considerations, can be reliable and useful only if one of the following conditions exists:

1. A single or predominant mechanism accounts for the greater part of the failures, and this mechanism can be accelerated;
2. Failure is caused by more than one significant known mechanism, and sufficient information regarding reactions rates as a function of stress level and type is available to permit weighing the relative effect of each for various conditions of stress.

The construction, testing, and statistical failure study of tantalum film R-C networks are described, as well as the study of the physics of failure mechanisms and their associated rates. Since the device under study is less complex in structure than many electronic circuits and systems studied with similar objectives in the past, it is probable that the work will yield accelerating techniques that will be widely applicable, and will provide a basic understanding of all the significant failure mechanisms. Early work on this project was previously reported by M. Walker in a paper on tantalum thin-film circuits[1].

In addition to information on accelerated and use-level R-C network failure distribution, and tantalum resistor and capacitor failure models, the program thus far has resulted both in the development of a number of novel techniques useful in film failure mechanism analysis, and in

considerable information in the area of basic tantalum and Ta_2O_5 film structure.

NETWORK DESCRIPTION AND FABRICATION

The devices tested are networks consisting of an 850-pf capacitor and a 5000-ohm resistor on the same oxidized silicon substrate. The resistor is U-shaped, 10 mils wide and 250 mils long. It consists of a sputtered tantalum film with a thermally grown oxide passivating coating. The capacitor consists of a sputtered tantalum lower electrode and a dielectric formed by anodizing selected areas of the lower electrode. Resistor terminations, conductors, and capacitor upper electrodes are vacuum deposited aluminum. The capacitor and resistor are connected internally, and three external leads are provided so that either component can be tested independent of the other. The networks are sealed in dry nitrogen in TO-5 packages. Figure 1 is a sketch of the cross section of a completed R-C network and illustrates the various film layers involved. The process used to fabricate similar devices has been described by Walker and Sharp[2].

Similar networks are also made on Corning #7059 glass substrates coated with 500 Å of tantalum oxide.

The preliminary tests of the program were conducted on resistors and capacitors sealed individually in TO-5 packages. Their construction was generally similar to that of the components of the R-C networks except that chromium-gold terminations and upper capacitor electrodes were used instead of aluminum.

R-C Network Accelerated Testing

The network test circuit (Figure 2) was designed to maintain a constant voltage across the network capacitor regardless of leakage current variations. This was done by means of a Zener diode voltage clamping network. This clamp, however, would not be in effect in the event of a drastic increase in leakage current as would occur with a capacitor short. The test circuit was designed to allow initial hourly, and eventual daily, monitoring of resistor current and capacitor voltage to determine the catastrophic failure time as accurately as possible. The devices were removed from the circuit for parameter measurement at 5, 25, 125, 625 hours and will be removed for parameter measurements every 1000 hours thereafter for units under the lower stress conditions. The parameters measured included capacitor value, dissipation factor, voltage at

1 na and 25°C, current at 15 volts and 100°C, and resistance at 30°C. Controls held at room temperature and no stress were also measured at these intervals and indicated a change of approximately ±0.06% for resistors and ±0.1% for capacitors after 625 hours. These figures represent a combination of shelf life change and variations in measurement accuracy.

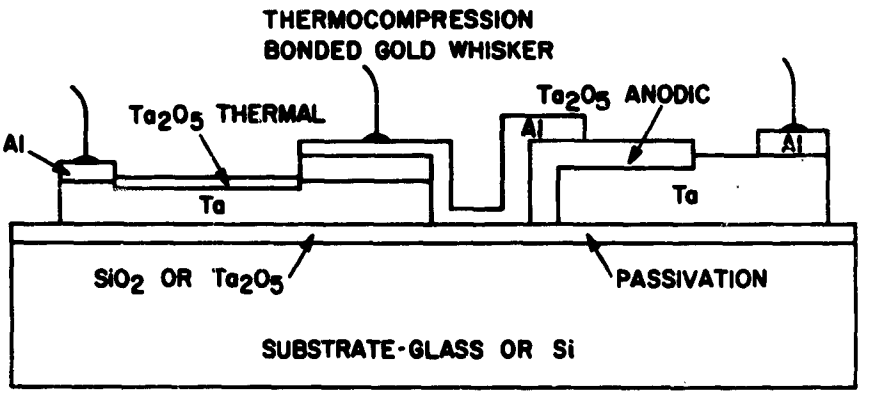


Figure 1

Sketch of cross section of R-C network illustrating film layers involved.

To limit the amount of damage caused by capacitor shorting which would restrict subsequent analysis, capacitor current is limited by means of a series resistor, R_{LE}. This resistor is varied to allow a maximum leakage current for the particular test temperature involved. This is required to allow for the higher operating leakage currents associated with elevated capacitor temperature.

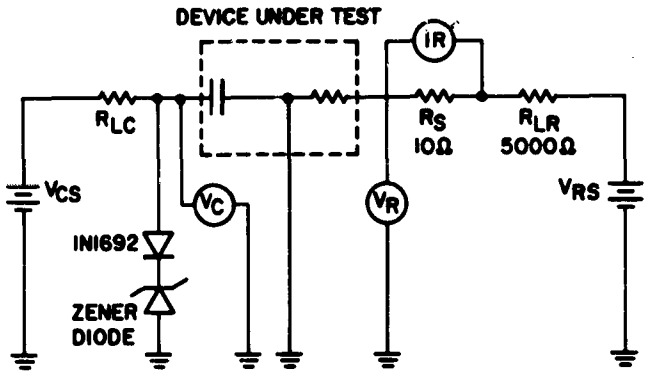


Figure 2

Network Test Circuit

There are two ways to effect capacitor heating. If the network resistor is not under load, all of the heating is effected by ambient heating. However, when the resistor is in operation, the high thermal conductivity of the silicon substrate results in the capacitor being heated as a result of the resistance heating. The temperature which the capacitor reaches as a function of power applied to the resistor is determined by means of an infrared thermal scanner, known as the Philco Thermal Plotter. The hotspot thermal resistance of the resistor to air can also be determined in the same manner.

The accelerated test matrix (Figure 3) was designed to permit variation of capacitor voltage stressing from 0 to 35 volts and of capacitor temperature stressing from 25°C to 225°C. Resistor power stressing varied from 0 to 1440 mw, and temperature stressing from 25°C to 256°C. Since earlier work had proven that film capacitor failure rates are considerably higher than film resistor failure rates, more attention was given to designing the matrix to judicious acceleration of capacitor failure. The matrix contains capacitor test conditions which include the rated use level of 15 volts and 125°C and the rated absolute maximum level of 15 volts and 175°C. The conditions were also selected to permit variation of temperature stress while voltage was held constant as well as variation of the voltage stress with constant temperature. This was done to acquire information relating failure to stress level.

	CAPACITOR BIAS				
	0V	10V	15V	25V	35V
NO. IN GROUP	20	10	10	10	10
CAPACITOR T	25°C	25°C	25°C	25°C	25°C
RESISTOR PWR.	0	0	0	0	0
RESISTOR T	25°C	25°C	25°C	25°C	25°C
NO. IN GROUP	20	-	20	20	20
CAPACITOR T	125°C	-	125°C	125°C	125°C
RESISTOR PWR.	0	-	755mw	0	755mw
RESISTOR T	125°C	-	146°C	125°C	146°C
NO. IN GROUP	20	-	20	20	20
CAPACITOR T	175°C	-	175°C	175°C	175°C
RESISTOR PWR.	0	-	1094mw	0	1094mw
RESISTOR T	175°C	-	200°C	175°C	200°C
NO. IN GROUP	20	20	20	-	-
CAPACITOR T	225°C	225°C	225°C	-	-
RESISTOR PWR.	0	0	1440mw	-	-
RESISTOR T	225°C	225°C	256°C	-	-

Figure 3

Accelerated Test Matrix

The 625-hours data included in Figure 4 showed no catastrophic film failures at conditions of 0 voltages and accelerated temperature stress or of 25°C and accelerated voltage stress. A number of failures occurred during conditions when both stresses were accelerated, with over 50% film failures at the 35-volt, 175°C condition, and 25% film failures at the 225°C, 15-volt condition. These were also a result of capacitor dielectric breakdown. The conditions resulting in the highest number of failures were selected as the accelerated stress conditions to be compared with the use and absolute maximum stress levels for determination of acceleration factors. At this time, insufficient failures have been achieved in the low stress levels to allow any but approximate acceleration factor calculations. A few failures occurred because the ball bonds lifted off the aluminum pads. Examination indicated the lack of metal fusion, probably caused by insufficient heating at the time of bonding. Since only actual film

failures are considered to merit extensive study, these factors were not included in the statistical analysis. No actual film resistor failures occurred under any of the test conditions.

VOLTAGE	CAPACITOR TEMPERATURE			
	25°C	125°C	175°C	225°C
0	0	0	0	2 NETWORKS HAD TC BONDS OPEN AFTER 625 HOURS
10	0	-	-	1 OPEN TC BOND-125 HRS. 1 C SHORT, 225-250 HRS. 1 C SHORT, 550-625 HRS. 3 NETWORKS HAD OPEN TC BONDS AFTER 625 HRS.
15	0	1 C SHORT DURING 25 HR. READING	1 C SHORT ON 5 HR. READING 2 C SHORTS, 125-150 HRS.	3 C SHORTS AT 25 HR. READING 1 C SHORT, 125-150 HRS. 1 C SHORT, 325-350 HRS. 1 C SHORT, 550-625 HRS.
25	0	1 C SHORT, 200-225 HRS.	1 C SHORT, 250-275 HRS.	-
35	0	1 C SHORT IN 1ST HR. 1 C SHORT, 125-150 HRS. 1 C SHORT, 250-275 HRS. 1 C SHORT, 500-525 HRS. 1 NETWORK LOST AN EXTERNAL LEAD, 625 HR. READING	1 C SHORT 1ST HR. 1 C SHORT 2ND HR. 1 C SHORT 5TH HR. 1 C SHORT AT 5TH HR. READING 1 C SHORT, 5-25 HRS 1 C SHORT DURING 25 HR. READING 2 C SHORTS, 25-50 HRS. 1 C SHORT, 200-225 HRS. 1 C SHORT, 300-325 HRS. 1 C SHORT, 400-425 HRS. 2 C SHORTS, 425-500 HRS.	-

Figure 4

Summary of voltage and temperature stress results after 625 hours of testing.

No large shifts (>5%) in resistance or capacitance were noted at any of the test conditions. Therefore the shifts were not included in the statistical failure analysis as failures. The shift data was related, however, to stress type and level and was used in the examination of minor R-C network failure mechanisms, such as resistor oxidation and ion migration. Significant shifts, however, were noted in the capacitor dielectric resistance, particularly under high temperature, 0 voltage conditions, where 100°C leakage current increases of 30X to 40X were observed. In some cases, particularly under bias, these shifts were reversed and often resulted in net decreases. These changes in leakage current had little observed effect on the first order capacitor parameters such as capacitance and dissipation factor. This shift therefore was also studied with respect to dielectric conduction mechanisms, in the analysis of the major failure mode, which was capacitor dielectric breakdown.

Failure distribution analysis was made, through the use of the Weibull distribution function[3], of the capacitor catastrophic failures:

$$F_x = 1 - \exp \frac{-(x-y)\beta}{\alpha} \qquad (1)$$

where: F_x = the cumulative fraction failed at any time, x,
x = a variable parameter, for this case, = t, (time),
y = location parameter, usually (0),
α = scale parameter, and
β = a shape parameter.

Substituting t + 0 for x + y, and differentiating (1) yields

$$\frac{d(F_x)}{dt} = \frac{\beta}{\alpha} t^{\beta-1} \exp - \frac{t\beta}{\alpha}, \qquad (2)$$

since, from (1), $1 - F_x = \exp - \frac{t^\beta}{\alpha}$

$$\frac{d(F_x)}{d_t} / 1 - F_x = \frac{\beta}{\alpha} t^{\beta-1} = \text{failure rate.} \qquad (3)$$

The value of β therefore is valuable in defining failure rate as a function of time, since

$\beta = 1$ indicates a constant failure rate,
$B < 1$ indicates a decreasing failure rate with time,
$B > 1$ indicates an increasing failure rate with time.

Equation (1) can be manipulated to the form

$$\ln \ln \left(\frac{100}{100 - \% \text{ Failure}}\right) = \beta \ln t - \ln \alpha \qquad (4)$$

This relationship is graphically represented on Weibull plotting paper (Figure 5). The shape parameter, β, now becomes the slope of the plot. The fact that β is independent of time and therefore results in a straight line Weibull plot is a good indication that a single or predominant mechanism is operative over the time examined. Also, plots resulting in similar β values, which indicate that their failure rates are similar functions of time (equation

(3)), are probably derived from failures caused by a similar mechanism. This is so because it is a necessary (although not sufficient) condition that failure rates resulting from similar mechanisms are the same function of time. Therefore, similar β values for both low and high level stress testing are an indication of a similar failure mechanism being operative and a good indication of time failure acceleration.

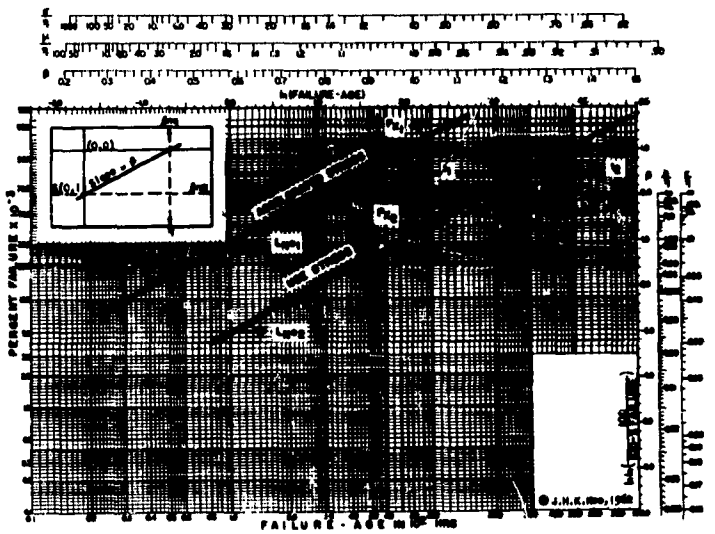


Figure 5

Failure - Age in 10^3 hrs.

The scale parameter, α, is the ordinate intercept on the Weilbull plot and a function of F_x at time, t = 1, and it is also time independent. The ratio of the α values from two plots achieved at an accelerated (α_1) and a use (α_2) condition therefore can be used to calculate the ratio of the cumulative fractions failed at time, t = 1, and thus is an acceleration factor. However, there is some question as to the utility of such a factor since, although it can be used for cumulative failure ratio calculation in any time t = 1 to t = total hours of accelerated testing for plots having similar slopes or β values, the number of failures incurred at the low stress level during the time of an accelerated test (<1000 hours) will, in many cases, be so small as to result in a low confidence level for the

low stress information, and therefore make the factor difficult to validate.

A more useful ratio is the ratio of (t_2), or time, to achieve a specified cumulative fraction failed at the accelerated stress level to (t_1), or the time required to achieve the same cumulative fraction failed (Equation (1)) from 1 we get the fraction surviving, or

$$FR = (1 - x) = \exp\left(\frac{-t^{\beta}}{\alpha}\right) , \tag{5}$$

or, $\ln FR = \frac{-t^{\beta}}{\alpha}$

At equal cumulative percent failed (or equation fraction surviving), the following relationship will apply between groups at two stress conditions:

$$\frac{t_2^{\beta_2}}{\alpha_2} = \frac{t_1^{\beta_1}}{\alpha_1} \tag{6}$$

where β is the same for the two stress conditions,

$$\frac{t_2}{t_1} = \frac{\alpha_2^{1/\beta}}{\alpha_1^{1/\beta}} \tag{7}$$

which is the acceleration factor used in the analysis of test data acquired in this program.

Weibull plots of the capacitor catastrophic failure distribution incurred at the 35-volt, 175°C condition, and the 15-volt, 225°C condition are shown in Figure 6. With both conditions resulting in failure distributions having β of 0.45 indicating decreasing failure rates and similar failure mechanisms. The plots also result in an α value of 95 for the 35-volt, 175°C condition, and 122 for the 15-volt, 225°C condition. When more failure data has been achieved at the low stress levels, the above shape and scale parameters will be compared with those parameters derived from the low stress data, provided the β values are similar, to arrive at acceleration factors.

Figure 7 shows the conditions used to validate the derived acceleration factors. This is done by testing larger quantities of networks than were tested in the initial phase of the program, at the two selected accelerated stress conditions and the two use-level conditions. The devices are selected from different lots than those used to

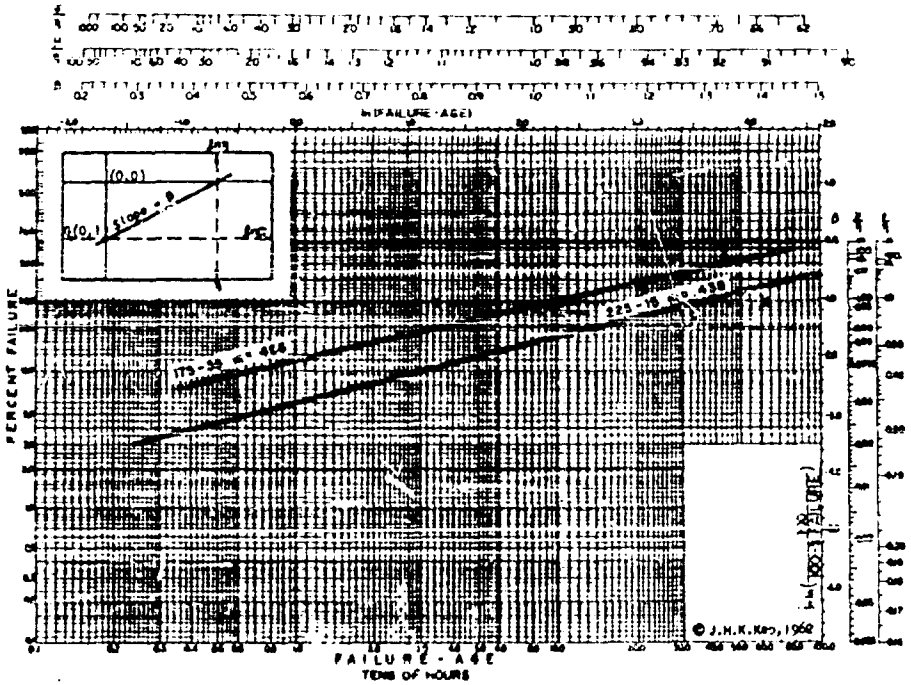


Figure 6

Failure-Age Tens of Hours

CAPACITOR VOLTS		CAPACITOR TEMPERATURE				
		25°C	125°C	175°C	200°C	225°C
0	NUMBER OF NETWORKS	20				
	RESISTOR POWER	0				
	RESISTOR HOT SPOT	25°C				
15	NUMBER OF NETWORKS		300	100	40	50
	RESISTOR POWER		755mw	1100mw	1265mw	1430mw
	RESISTOR HOT SPOT		146°C	200°C	223°C	256°C
25	NUMBER OF NETWORKS			20		
	RESISTOR POWER			1100mw		
	RESISTOR HOT SPOT			200°C		
35	NUMBER OF NETWORKS			50		
	RESISTOR POWER			1100mw		
	RESISTOR HOT SPOT			200°C		

Figure 7

Conditions used to validate the derived acceleration factors

derive the acceleration factors. It is therefore conceivable that the actual failure rates will differ from those of the originally tested devices, due to uncontrollable process variations. However, if true acceleration has been achieved, and the same basic failure mechanism is still operative in the validation sample, the acceleration factors derived in the validation test should be similar to those derived in the initial test. The validation test includes, in addition to larger quantities of units at the four specified conditions, small sample quantities under similar conditions of either voltage or temperature stress as the four specified. This is done to acquire more data relating both catastrophic failure and parameter shift to stress, as was done for the resistor oxidation shown later in this paper.

Tantalum Thin-Film Resistor Failure Model

There are two major modes of failure in thin-film resistors: parameter shift and catastrophic failure, the latter including opens and shorts. Some of the mechanisms to which these modes have been attributed in the literature are:

1. Oxidation
2. Precipitation
3. Evaporation
4. Chemical Corrosion
5. Agglomeration
6. Internal Stress Relief
7. Thermal Fatigue
8. Contamination
9. Mechanical Fatigue.

In the study of tantalum thin-film failure, many of these mechanisms can be excluded because:

1. Precipitation obviously does not apply in the case of a single constituent film,
2. Evaporation and chemical corrosion do not apply because of the refractory and inert nature of tantalum film,
3. Agglomeration and internal stress relief do not apply, based on knowledge acquired through practical experience or from authoritative external sources,
4. Thermal and mechanical fatigue do not apply because of the use of thin films (<2500Å). This is supported by program testing.

The remaining mechanisms, oxidation and contamination, were found to apply in the case of tantalum film. In the course of the initial study, more were uncovered which are combinations and extensions of these two.

The failure mechanisms believed predominant in tantalum film failure and studied in this program are:

1. Oxidation,
2. Constriction,
3. Crystallization,
4. Thermal Runaway,
5. Ion Migration.

Mechanisms 2, 3, and 4 ultimately involve oxidation.

The primary cause of tantalum film resistance change, particularly at elevated temperatures, then, is oxidation. This change occurs in the form of a resistance increase caused by the formation of tantalum-oxygen or tantalum nitrogen compounds or solid solutions having much higher specific resistivities than the tantalum itself. Measurable increases in resistance (0.2%) occurred in 650 hours at 175°C. A median 0.7% increase occurred in 650 hours at 225°C. The only type of failure of which oxidation would be a direct cause would be of the drift variety, where a device having a resistance value at the edge of a specification might exceed the specification due to a slight resistance increase. However, almost every catastrophic failure mode mentioned here has oxidation as its indirect cause.

Thermal Runaway

Thermal runaway results from the combination of two other resistor failure mechanisms: (1) oxidation, and (2) constriction. It occurs when current is passed through an area which is constricted enough to cause a local increase in resistance, thereby creating a hot spot, which results in increased local oxidation and further constriction. This cycle continues until the substrate in the area of the constriction either cracks or vaporizes, causing an open. This type of failure is highly dependent on both the resistor constriction and the substrate material, since they are related to power dissipation. In fact, the existence of a resistor hot spot without an actual constriction but due to a substrate having poor thermal conductivity can cause this type of failure. The final result of a thermal runaway on a 5 mil wide resistor failure is shown in Figure 8.

Figure 8

Thermal runaway failure.

Crystallization

The majority of the catastrophic failures of resistors encountered in the initial test phase involving individual resistors were related to transparent areas in the immediate vicinity of the gold termination. Figures 9A and 9B are photographs of a typical instance. Close inspection indicated that these areas resulted from the high temperature stabilizing step, which caused complete opens during power aging testing. Talysurf measurements showed the transparent areas to be raised rather than depressed, indicating an additive reaction. Because of the transparent nature of the material, it was assumed that there was a high rate of oxidation in the immediate areas of the termination, since tantalum pentoxide is transparent. As this occurred mainly during the high temperature stabilization bake, when the entire resistor surface was at the same temperature, the preferential oxidation in the termination area must have been due to a reduction of effectiveness of the protective self-limiting oxide film.

The existence of impurity nucleated crystallization has been reported by Vermilyea[4] and recognized by many others in bulk silicon and tantalum oxides. We have had indications of this phenomenon in capacitor failure analysis and recognize it as a potential basic failure mechanism.

The ease with which gold diffuses into many materials, including tantalum, leads to the suspicion that the existence of gold in the tantalum adjacent to the termination

could act as a nucleation center for crystal growth. Substitution of aluminum for the gold-over-chromium terminations has eliminated the termination transparent areas in the R-C networks.

Figure 9

Photomicrographs of resistor affected by crystallization:
(a) resistor termination area, and
(b) structure of crystallized area.

Oxidation

The curve of resistivity versus time in air at elevated temperature (Figure 10) indicates three areas of varying resistance stability (or curve slopes) with time. The initial and final parts of the curve, relative to the center portion, indicate unstable films. Stability is obtained in the film resistors by:

1. An initial 1-hour stabilization step,
2. Working with as thick a film (>600 Å) as possible.

This has the effect of placing the resistors in the central or most stable portion of the curve, minimizing the slope (a process control measurement) in that part of the curve.

In the oxidation study, R-versus-time curves were plotted at a number of different temperatures and the films were studied at various intervals of oxidation.

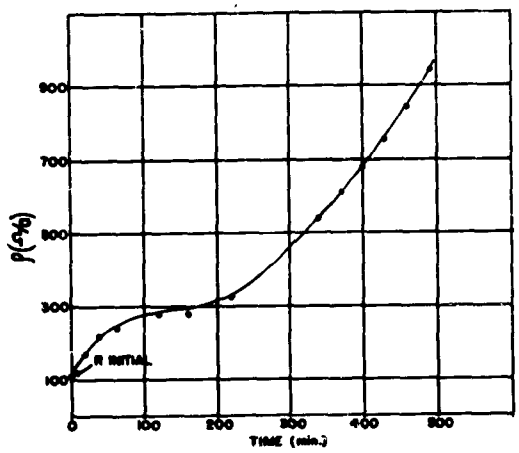


Figure 10

Sheet Resistivity vs. Time at 420°C in air.

Initial exposure of the film to high temperature in an air ambient resulted in a rapid increase in resistance and an immediate increase in the negative temperature coefficient. Examination of the film at this point indicated none of the surface interference color which is associated with surface oxide formation. Furthermore, Talysurf measurement indicated no significant increase in thickness, also normally associated with surface oxidation. A plot of the growth of oxide above the original surface is given in Figure 11.

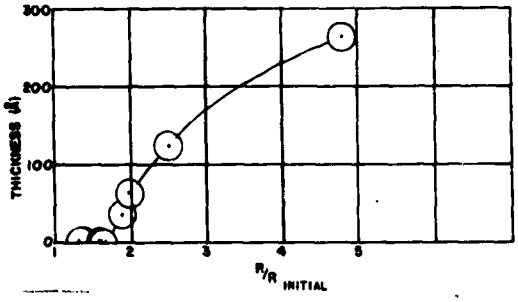


Figure 11

Growth of Oxide Above Original Surface

It is believed that the initial increase in resistance is caused by oxygen dissolution into the tantalum and the formation of a tantalum-oxygen solid solution having a relatively high specific resistivity. The oxygen is believed to come from one or all of the following sources:

1. External, resulting in a process rate limited by the diffusion of O_2 through the film,
2. Grain boundary occlusion as a result of the sputtering process,
3. Substrate occlusion.

The oxygen dissolution theory has been postulated by Kofstad[5] and Bassechès[6].

In the second period of film oxidation, a considerable decrease in the resistance increase rate is noted, no further increase in negative TCR is indicated, a surface interference color begins to form, and thickness increases have been measured. In this area the tantalum oxidation rate is limited by the formation of a dense surface oxide layer, which in turn limits oxygen diffusion and results in a self-reduction of growth rate. Since the oxide is a nonconductor, oxide growth results in a calcuable increase in resistance, given by

$$\Delta R = \frac{\rho_1}{w(T_o - T_{oxide})} - R_o, \qquad (8)$$

where R_o = initial resistance,
1 = line length,
w = line width,
ρ = specific resistivity of the Ta film,
T_o = initial film, and
T_{oxide} = thickness of Ta converted to surface oxide.

This equation approaches infinity as the thickness of the tantalum converted to oxide approaches the initial thickness. For this reason, the inherent stability of a stabilized film is a function of the initial film thickness. This also explains the increased slope encountered at the end of the resistance-versus-time curves.

Similar experiments were conducted to observe surface reaction effects in various other media, namely dry oxygen, dry nitrogen, and vacuum. The activation energy obtained from Arrhenius plots was similar in each case - running between 30,000 and 35,000 calories per gram mole. The

reaction rate was 0.4 to 0.5 as great in nitrogen as in oxygen or in air (Figure 12).

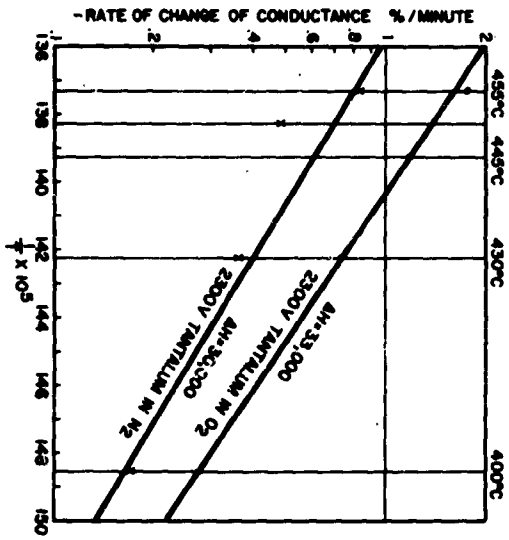


Figure 12

Plots of reaction rate of tantalum in dry O_2 and in dry N_2.

Resistance Change by an Ion-Migration Mechanism

Typically, a gradual increase in resistance is observed on tantalum thin-film resistors when these resistors are life-stressed by temperature, power, or a combination of the two. This increase in resistance is generally attributed to oxidation. However, it was also observed during the RADC accelerated stress tests on resistors that, in certain cases, a negative change in resistance occurs. U-shaped resistors on oxidized silicon substrates on power stress showed initial declines in resistance, followed by the expected increase. The decreases were more rapid and ended sooner at the higher power stress. The maximum decline was about 2% regardless of the magnitude of the stress. It was proposed that this effect could be explained by an ion migration mechanism whereby mobile ions from regions on the substrate surface adjacent to the resistors are caused to migrate and accumulate along the resistor length under the influence of an applied electric field.

To study this ion migration phenomenon proposed as a failure mechanism in resistors, a pattern of adjacent, parallel resistors having various spacings was used as a

vehicle. The ion migration experiments involved applying a d-c potential between adjacent resistors and comparing the resistance values of these resistors before and after the application of the electric field.

Two phenomena which might be expected as the ultimate result of ion migration are:

1. An increase in the resistance of the resistor biased positively, and
2. A decrease in the resistance of the resistor biased negatively.

Figure 13 represents the layout of the resistors in the test vehicle. Individual resistors were used as controls so that orginary oxidation effects produced by ambient temperature could be taken into account when analyzing the resistance change of the biased resistors.

Figure 13

Ion Migration Test Pattern

In preliminary experiments, a 65-hour application of 215 volts at 25°C between Ta resistors 2 mils apart resulted in a 3% increase in resistance of the resistor which was biased positively. The negatively biased adjacent resistor, and control resistors on the same chip which were not biased, remained essentially unchanged. The experiment was performed in room air (50% relative humidity). Microscopic examination of the biased resistor pair indicated the side of the positively biased resistor facing the adjacent resistor became discolored. It was concluded that migration of ions, presumably arising as a result of absorbed moisture on the

surface or by transfer of oxygen ions, from the Ta_2O_5 covered glass, to the metal film, produced an anodization process along the inside edge of the positively biased resistor, resulting in the observed resistance increase.

Experiments at 280°C with 215 volts applied for 24 hours between resistors separated by 0.5 mil have resulted in decreases in the resistance of the negatively biased resistor relative to the adjacent positively biased resistor, or to unbiased resistors on the same chip.

The possibility of the observed resistance decreases being due to annealing effects was ruled out by the controls, which are unbiased resistors on the same chip. Thus it appears that cation migration can reduce resistance values of resistors.

The following data support the logical assumption that accumulated ions can be redistributed by reversing the polarity of the applied field. Two sealed R-C networks, which were part of a group of devices subjected to an RADC acceleration power stress test (85 V, 1440 mw), were evaluated for ion migration. These units had shown a negative change in resistance of ≈2% after five hours, and then a gradual increase in resistance approaching their initial values. The same bias of 85 volts which the units had undergone during accelerated stressing was applied, but in the reverse direction. After one hour, the resistances were measured and the field was again reversed. The results are shown below.

	UNIT # B4-135	UNIT # B4-108
Initial Reading	4774Ω	5010Ω
Polarity Reversed		
1 hour at 85 V	4722Ω	4993Ω
Change	-1.1%	-0.33%
Polarity Reversed		
1 Hour at 85 V	4740Ω	5005Ω
Change	+0.4%	+0.24%

A significant negative change in resistance was observed in one hour. These resistors are U-shaped and the adjacent parallel portions of the resistor are 10 mils apart. Applying a potential of 85 volts produces an electric field of 8.5 V/mil between the resistor legs near the terminals, and lower fields near the middle of the resistor.

Experiments were also performed, using autoradiographic techniques with radioactive sodium, to demonstrate that sodium ion migration, particularly on surfaces of "alkali-free" glass, is a significant factor which limits the stability of resistors. These experiments clearly demonstrated that sodium ions, when present on the surface of the Corning #7059 glass of R-C networks, migrate in an electric field.

Work dealing with the autoradiographic technique employed to show the ion migration phenomenon was done by Dr. S. S. Choi[7] of the Ford Scientific Laboratory, Blue Bell, Pennsylvania.

Capacitor Thermal Runaway

Capacitors fabricated on either oxidized silicon substrates or Ta_2O_5 coated glass substrates showed a high incidence of dielectric failures and catastrophic increases in leakage current on elevated temperature plus bias stressing. The failure mode is characterized by current runaway. Once a high localized current starts, the Ta_2O_5 dielectric in the path of the current is heated, lowering the resistance of the dielectric and causing current to increase further until a failure occurs. A tantalum thin-film capacitor with low resistance in series with it that exhibits this type of failure usually ends as an open as sufficient current passes through the capacitor to burn open some part of the tantalum under the anodized dielectric. A higher resistance in series with the capacitor prevents the lower electrode from opening and so leaves the final failure a shorted dielectric.

Any contributory factor which produces a temporary or permanent weak point in the capacitor dielectric can initiate a thermal runaway type of failure. It is felt that this failure mode is caused by localized impurity nucleated crystallization of the dielectric during anodization. Application of high fields and temperature accelerate the nucleation and the growth of the already nucleated crystals.

Dielectric Degradation by Ionic Conduction

Capacitors showed large increases in leakage current after stressing. On zero bias stressing in the temperature range of 125°C to 250°C, the median value of leakage current increased 40 times over its initial value. The leakage current eventually leveled off, with the leveling off occurring earlier for high temperature conditions. A bias voltage on the capacitor inhibited the leakage current

increase. The current decreased if voltage was high enough. Approximately 15 volts bias inhibited leakage current change at 125°C. Approximately 25 volts bias inhibited leakage current change at 175°C.

To illustrate these changes, a group of tantalum thin-film capacitors were baked for one hour at temperatures of 151°C, 175°C, 207°C, 227°C, and 270°C. Leakage current was read as a function of applied voltage before the series of bakes, and then after cooling to room temperature subsequent to each bake. The resulting data were converted into plots of logarithm of leakage current versus reciprocal temperature to estimate activation energy for the mechanism causing degradation of leakage current.

The construction of the capacitors was exactly that used for the R-C networks, except that tin solder was used in header bonding instead of gold-tin eutectic, and the 200°C pre-sealing bake and post-seal leak test were eliminated to diminish exposure to elevated temperatures before the test.

Arrhenius plots were used to estimate the activation energy for the mechanisms causing changes in room temperature leakage current. The estimation of activation energy was complicated by the fact that room temperature leakage current does not increase indefinitely with baking temperature, but reaches a peak at some baking temperature and then proceeds to decline. Two straight line segments were estimated from the leakage current data, and activation energies for both the mechanism causing increased leakage current and the mechanism causing decreasing leakage current were estimated. The precision of the estimates of the slope of the log I versus 1/T plots was poor, especially at the high temperatures where leakage current declined.

The estimated temperature at which leakage current reached a peak after one hour of baking varied between 192°C and 211°C. Leakage current increased with one-hour bakes at increased temperatures below the 192°C to 211°C range, and decreased with one-hour bakes at increasing temperatures above this range. Shorter baking times presumably would increase the temperature at which the current peaks, and longer baking times presumably would lower it.

The activation energy for the mechanism causing increasing leakage current was calculated as 22,000 calories/mole from the average of twelve Arrehenius plots. The calculated values ranged from 15,000 to 28,000 calories/mole.

The calculated activation energy for the mechanism causing decreasing leakage current varied from 7,000 to 17,000 calories/mole. The average value obtained was 11,500 calories/mole. Activation energies for the leakage current appear slightly higher at lower voltages.

Measurements of at least four capacitors which had previously been subjected to temperature and bias stressing indicate the existence of at least two separate portions in the log I versus V curve. When I was plotted versus V, a straight line, which extrapolated to the origin, was obtained between 0.5 volt and 2 to 10 volts, depending on the particular capacitor measured, as illustrated in Figure 14. Thus leakage was ohmic in the low voltage region. The slopes of the lines at low voltage are interpreted as representing resistance, at various temperatures, due to ionic conduction. A plot of log of R versus reciprocal of absolute temperature was approximately linear indicating that the Arrhenius equation is obeyed (Figure 15). The slope indicates an activation energy of 17 kcal/mole, which is interpreted as the activation energy for ionic conduction in the Ta_2O_5 dielectric.

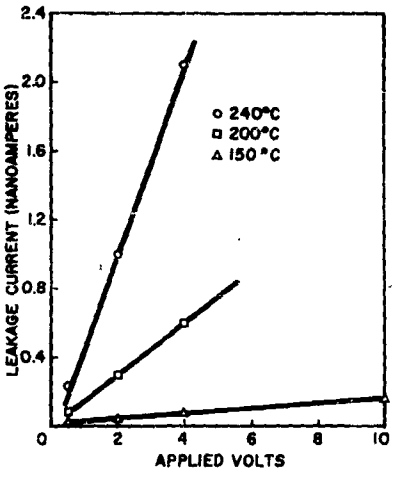


Figure 14

Plots of leakage current vs. applied voltage and temperature.

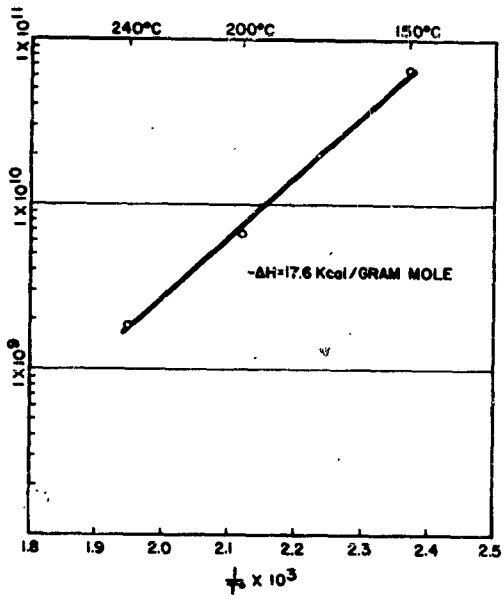


Figure 15

Plot of log of R versus reciprocal of absolute temperature.

A series of recorder plots have been made of capacitor leakage versus time under various bias conditions at 218°C. Figure 16 graphically shows the change in anodic leakage current with time.

In another test, the 10-volt bias was applied in the reverse direction at 218°C. After the current had fallen to about 20% of its original value and the polarity was reversed, the leakage current measured 0.1 ma. In 12 minutes the leakage current gradually increased to a peak of about 2 ma. In an additional 4 minutes, the leakage current fell to about 1.2 ma.

The above changes in leakage current with time are considered typical of what could be expected as a result of ion migration in the Ta_2O_5. The observation that anode leakage current increased dramatically both on temperature and temperature-plus-voltage stressing was made earlier in this section. Analysis of the stress data has also shown that at various stress levels, a small but detectable increase in capacitance was seen on groups stressed at

125°C, 35 V, and 225°C, 15 V. This kind of change in capacitors has been discussed by Vermilyea[8] and Smyth[9]. The changes are caused by oxygen diffusion from the Ta_2O_5 layer into the tantalum lower electrode. A thin layer of the Ta_2O_5 at the tantalum interface becomes so depleted of oxygen that it is in effect a semiconductor instead of an insulator, causing the effective dielectric thickness to decrease. As a consequence, an increase in capacitance results.

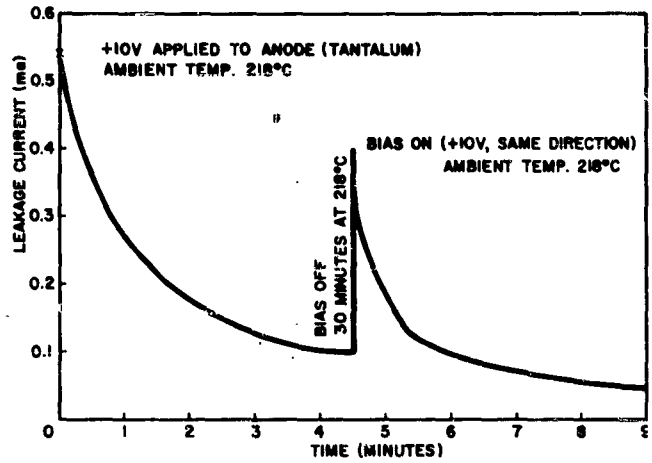


Figure 16

Change in anodic leakage current with time

Location and Identification of Dielectric High Current Density Areas

Efforts have been made to develop and evaluate means for detecting defects in anodic Ta_2O_5 films. When devices have failed during accelerated life tests, the defective regions must be precisely located to permit detailed examination, as under the high magnification of an electron microscope.

Several techniques were tried in attempts to locate the defective regions of shorted or near-shorted thin-film capacitors. Using a Thermal Plotter and a 0.3 mil resolution lens, successive scans were made across the upper electrodes of biased capacitors. No hot spots were

detected on any of the six shorted devices scanned. It is believed that if any hot spots exist, they are too small, relative to the field of view, to be detected.

However, defects in these same units were subsequently located by a copper plating technique. This technique consists of the following steps: The tantalum metal is made cathodic in an electrolytic cell. An acidic aqueous cupric sulfate solution is used. Several volts are applied to the cell between the anodized tantalum sample and a relatively large platinum cathode. The current passed is less than 5 mA with a typical capacitor structure.

The copper plating technique is particularly useful because it is very sensitive and provides a permanent visual indication of the location of current-carrying defects in the Ta_2O_5. On extended plating, the defect areas become covered with more or less hemispherical particles of copper, the diameter of which is an indication of the magnitude of the current flowing through the particular defect. The use of copper plating to show damaged regions on a dielectric was discussed by Fresia, et al[10]. The novel feature of our application of copper plating is to use the electroplated dot as a mask during subsequent chemical etching so as to produce an identifiable region after the copper is removed.

This sequence of steps produces a capacitor structure in which the defects are clearly delineated, with each defect being located in the center of a plateau of thicker Ta_2O_5.

The delineation of defects in anodic Ta_2O_5 films by anodic plating was compared with that revealed by copper plating (cathodic). It was concluded that the copper plating technique offered considerably greater sensitivity than anodic plating methods which used solutions containing Pb++ of Mn++ to deposit PbO_2 or MnO_2 at the anode. A major problem with anodic plating is an apparent tendency for the defective regions in the Ta_2O_5 film to anodize, with the result that the total anodic current drops with time after application of potential. By contrast, the cathodic copper plating process can be carried out over a long period of time without any decrease in magnitude of the cathodic current, thus enabling readily visible copper dots to be obtained and enhancing defect detection.

The technique of encircling defective regions by using the copper electroplated dot as a mask and etching with an acidic fluoride solution to dissolve unmasked Ta_2O_5 has permitted a more detailed study of the defects in the capacitor dielectric. Devices which were weed-out failures because of high leakage, but were not shorts, were used for these studies.

Upon removal of the copper mask from the marked area, the sample was re-etched in a solution of 30 ml of hydrofluoric acid (49% HF by weight) and 70 ml of water for five to ten minutes. A crystalline Ta_2O_5 hexagonal structure appeared within the originally marked area. This structure was distinguishable because the dissolution rate of crystalline Ta_2O_5 is lower than that of amorphous Ta_2O_5 in the same solution. Figure 17 shows a localized defect in the capacitor dielectric before and after etching.

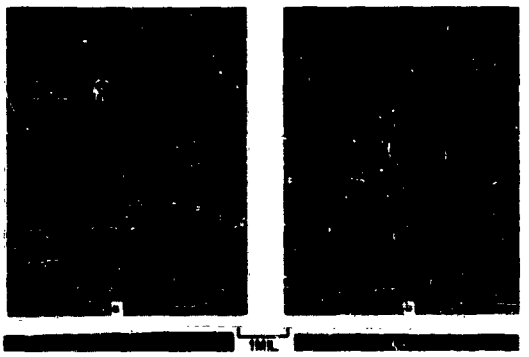


Figure 17

Localized Crystallization of Dielectric in Capacitor Failure
(a) before chemical etching, and
(b) after chemical etching.

A schematic of the processing steps used to isolate and investigate defects in the capacitor dielectric is shown in Figure 18.

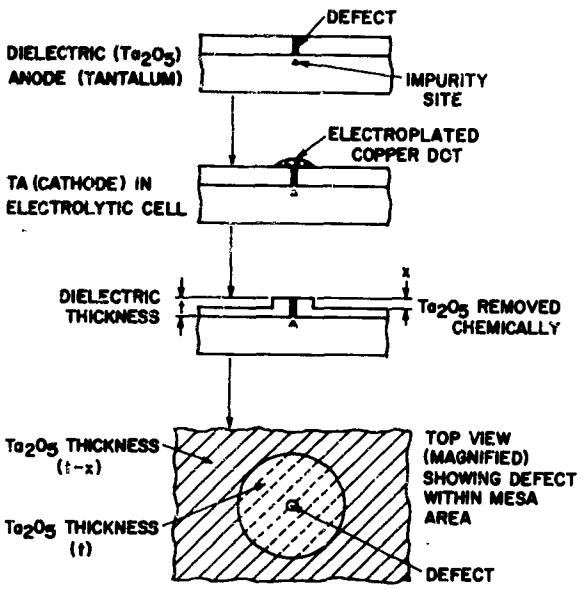


Figure 18

Procedure for isolating and investigating capacitor dielectric effects.

Tantalum, Tantalum Oxide Crystallographic Studies

Crystallographic studies performed by Holloway and Schlacter[11] produced some interesting information about the structure of sputtered tantalum films. The purpose of this study was to determine the compositional structure of the tantalum metal electrode used in the R-C networks. This was necessary to establish a norm for comparison with Ta layers in devices which fail. A further aim was to establish the structural consequences of changes in sputtering conditions.

Tantalum was sputtered onto carbonized microscope grids to a thickness of 300 Å at four voltages: 1000 V, 1450 V, 2300 V, and 3500 V.

High magnification transmission micrographs of the samples were taken.

Initial results indicated a relation between sputtering voltage and grain size, and indicated voids, which had been suspected because sputtered Ta films were appreciably less dense than bulk Ta. (The geometry of the Ta grains and of the voids may explain the variation of electrical conductivity of films with sputtering conditions[12]. Figure 19 shows several transmission micrographs of tantalum films deposited at various cathode sputtering voltages.

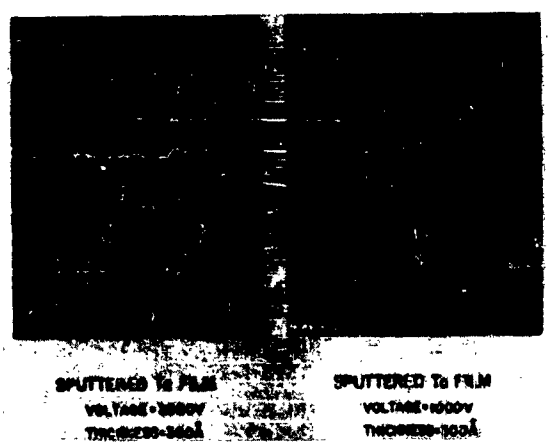


Figure 19

Transmission Electron Micrographs

Conclusions

It is concluded that the reliability of tantalum thin-film resistors is very good and the failure mechanisms are sufficiently well understood to permit predictions of resistor reliability under various stress conditions. The major failure mode for tantalum film R-C networks has been shown to be capacitor dielectric breakdown. Statistical analysis of failures due to this mode indicates a failure rate which is a single function of time over the time examined, which is, in turn, indicative of a single or predominant basic failure mechanism.

The development of a unique technique for dielectric defect location has resulted in relating high-current-carrying dielectric areas to Ta_2O_5 crystallization.

A study of the minor failure modes has resulted in defining tantalum oxidation as the basic mechanism for eventual resistor catastrophic failure. The similarity of calculated activation energies for reaction in air, oxygen, and nitrogen indicates a basic rate-limiting mechanism - probably diffusion.

Parameter shifts in both resistors and capacitors were traced to ionic conduction from ionic contamination probably introduced during processing.

The study indicates a high degree of reliability for the aluminum terminated tantalum film resistor and that the R-C network reliability limitations are keyed to the capacitors. The major capacitor problem was found to be strikingly similar to that reported for discrete capacitors[13].

REFERENCES

1. Walker, M. J., "Tantalum Thin Film Circuits", presented at the University of Wisconsin Microelectronics Symposium, May 1963.

2. Walker, M. J. and M. Sharp, "Thin Film Accelerated Life Tests", Rome Air Development Center, Research and Technology Division, U. S. Airborne System Command, Griffiss Air Force Base, New York, Technical Documentary Report No. RADC-TR-65-137 (1965).

3. Goode, H. P. and J. H. K. Kao, "Sampling Plans Based on the Weibull Distribution", Proceedings of the Seventh National Symposium on Reliability and Quality Control (1961).

4. Vermilyea, D. A., "Nucleation of Crystalline Ta_2O_5 During Field Crystallization", J. Electrochem Soc., 104, pp. 542-546 (1957).

5. Kofstad, P., "The Oxidation of Tantalum at 300°C to 500°C", Central Institute for Industrial Research, Oslo, Norway, ASTIA Report #273 214 (1962).

6. Basseches, H., "The Oxidation of Sputtered Tantalum Films", Journal of the Electrochemical Society, 109, 475-478 (1962).

7. Choi, S. S., "Radiochemical Study on Lateral Ion Migration in Insulating Substrate for Thin-Film Microcircuits", paper prepared for presentation at Fourth Failure Conference, Chicago, Ill., Nov. 1965.

8. Vermilyea, D. A., "Some Effects of Annealing Ta_2O_5 at High Temperatures", Acta Metallurgica, 5, 113 (1957).

9. Smyth, D. M., "Heat Treatment of Anodic Oxide Films on Tantalum", J. Electrochem. Soc., pp. 1264-76 (1963).

10. Fresia, E. J. and J. M. Eckfeldt, "Failure Modes and Mechanisms in Solid Tantalum Capacitors", Physics of Failure in Electronics, Vol. 2, pp. 483-97

11. Holloway, H. and M. Schlachter, Ford Scientific Laboratory, Blue Bell, Pennsylvania.

12. Wagner, S. R. and M. Walker, "A 2500 Ω/□ Tantalum Thin Film Resistor For Silicon Integrated Circuits", International Electron Devices Meeting, sponsored by IEEE, Washington, D. C. October 20-22, 1965, paper 10.1.

13. Didinger, G., "Tantalum Capacitor Accelerated Life Testing", Evaluation Engineering, Sept. - Dec. 1964.

SECTION III

DEVICE PHYSICS

A TRANSIENT COMPONENT IN THE BREAKDOWN VOLTAGE OF SILICON P-N JUNCTION RECTIFIERS

H. C. Gorton

Battelle Memorial Institute
Columbus, Ohio

ABSTRACT

The Electronic Component Reliability Center at Battelle Memorial Institute is involved in a continuing study of physical mechanisms and processes in electronic parts pertinent to device reliability. In a study of avalanche breakdown in diffused silicon p-n junctions, an exponential decrease with time has been observed in the onset voltage for microplasma conduction. The time constant of this decrease is inversely proportional to the ratio of average on- to off-time of the microplasma pulses, and the magnitude is temperature dependent, peaking at approximately 260° K. The theory is advanced that, during the avalanche pulse, the acquisition rate of electrons by ionized donor impurities is increased, decreasing the ionized donor concentration, thus changing the field intensity and hence the breakdown voltage. In effect, compensating donor sites tend to fill during the microplasma pulses and empty in the intervals between pulses. Thus, $\Delta N_D(t) = r_f \tau_{on} - r_e \tau_{off}$, where r_f and r_e are the filling and emptying rates, respectively, and τ_{on} and τ_{off} are the relative average on- and off-times of the microplasma pulses.

At temperatures where the Fermi level is several kT from the energy level of the trapping sites, the occupation index of the traps is relatively unaffected by injected carrier density. However, where E_F is within a few kT of E_T, the occupation density of the traps is a strong function of the injected carrier density. An increase in the magnitude of the transient voltage effect of a factor of 10 is observed as the Fermi level is made to approach 0.24 eV above the valence band--the energy level associated with copper donor atoms. This effect may be accounted for by a decrease in the concentration of ionized copper of the order of 10^{13} cm^{-3}.

INTRODUCTION

One of the objectives of the ECRC Reliability Physics Research Program[1] is to develop techniques for obtaining from terminal measurements on electronic parts information on the physical processes and properties upon which the useful life of the parts may depend. Specific to this objective, a study of microplasma conduction in reverse-biased silicon rectifiers has been carried out. At the onset of microplasma conduction, the avalanche current is intermittent, and particular attention has been focused on a study of the physical properties which influence the turn-on and turn-off mechanisms of the microplasma. Devices used in this study were commercial medium-power, controlled-avalanche rectifiers with breakdown voltages between 750 and 1000 volts.

CALCULATED BREAKDOWN VOLTAGE

The usual expression for the breakdown voltage of a step junction[2] is:

$$V_B = \frac{\epsilon_o k}{2q}\left(\frac{1}{n_n} + \frac{1}{p_p}\right) E_B^2, \qquad (1)$$

where ϵ_o = dielectric permittivity of vacuum,
k = dielectric ratio,
q = electron charge,
n_n, p_p = majority carrier concentrations on either side of the p-n junction,
E_B = critical field.

The value of the critical field is given as 4.4×10^5 volts/cm[3]. However, the notion of critical field is only an approximation, and the value of the field

---

(1) The member companies of the ECRC include Bell Telephone Laboratories, Inc., General Electric Company, General Motors Corporation, International Business Machines Corporation, International Telephone and Telegraph Corporation, Lockheed Aircraft Corporation, National Aeronautics and Space Administration, U. S. Air Force (Rome Air Development Center), U. S. Army (Electronics Command), U. S. Navy (Bureau of Weapons), and Westinghouse Electric Corporation.

(2) R. M. Ryder, Transistor Technology, Vol. 1, Edited by H. E. Bridgers, J. E. Scaff, and J. N. Shive, D. Van Nostrand Company, New York, 202 (1958).

(3) A. Goetzberger, B. McDonald, R. H. Haitz, and R. M. Scarlett, "Avalanche Effects in Silicon p-n Junction. II. Structurally Perfect Junctions", J. Appl. Phys., 34, 1591-1600 (June, 1963).

at which an ionizing current may be sustained may depend somewhat on the values of n_n and p_p. In our rectifiers $n_n \gg p_p$, and the average value of p_p (see Figure 7) is 2.3×10^{14} cm^{-3}. Thus, the calculated value of V_B is 2000 volts. The observed value, however, is 749 volts, suggesting that breakdown is occurring through a localized region where p_p has been increased by about a factor of 3.

Since the majority carrier concentration is the difference between the ionized n-type and p-type impurity concentrations, Equation (1) may be expressed as

$$V_B = \frac{6.25 \times 10^{17}}{N_A - N_D} . \qquad (2)$$

Of practical interest is the fact that the breakdown voltage can be influenced by compensating impurities or defects which may have sufficiently high diffusion rates to constitute reliability hazards.

TRANSIENT VOLTAGE EFFECT

Consider an n^+- p junction containing compensating donor sites with a reverse bias insufficient to sustain an avalanche current. The current in the space-charge region will be comprised principally of carriers generated within the space-charge region. Because of the high field, the generated carriers will be quickly swept out, and their probability of becoming trapped in the depletion layer is very low.

Consider now the junction in the avalanche mode. Because of the very high current density, appreciable numbers of carriers may be trapped at the donor sites although the trapping probability is low. However, the degree of trapping is strongly temperature dependent, as will be discussed later. Thus, the value of $N_A - N_D$ will tend to be higher under the high current condition of avalanche conduction. Contrarywise in a p^+- n junction in the avalanche mode, compensating acceptors in the space-charge region in the n-type base will tend to capture electrons and become negatively ionized, decreasing the value of $N_D - N_A$.

At the onset of breakdown, the avalanche current is intermittent. Thus, the net ionized impurity concentration tends to oscillate between the pre- and postbreakdown equilibrium values. If the average pulse length and repetition rate are short with respect to the time constants for the filling and emptying of traps, the concentration of ionized compensating impurities will depend on the relative on-times, τ_{on}, and off-times, τ_{off}, of the microplasma pulses. Thus, in an n^+- p junction in the avalanche mode,

$$\Delta N_D(t) = r_f \tau_{on} - r_e \tau_{off}, \qquad (3)$$

where r_f and r_e are the filling and emptying rates, respectively. For the case of a model containing a single trap level, the transition from the prebreakdown occupation density to the postbreakdown occupation density of trapped carriers is expected to be an exponential process. Haitz has shown this to be the case in diodes with shallow traps and breakdown voltages of approximately 20 volts which were cooled to 77° K[4]. The time constant which he quotes for the transition from prebreakdown to postbreakdown occupation density is 7.2 microseconds. This time constant was obtained under conditions in which the on-time of the microplasma pulse was long with respect to the filling rate of the traps.

In the circumstance where the pulse length is short with respect to the filling rate of the traps, of course, much longer times would be required for equilibrium density to be reached, and the time constant would depend strongly on the ratio of the average on- to off-time of the microplasma pulses. The trapping rate would also increase exponentially with the nearness of the energy level of the trap to midband. In our rectifiers, which contain relatively deep traps, the filling rates were relatively slow, and τ_{on} was not long with respect to the time constant of the traps. Time constants of the breakdown voltage decay of hundreds of seconds were observed at room temperature.

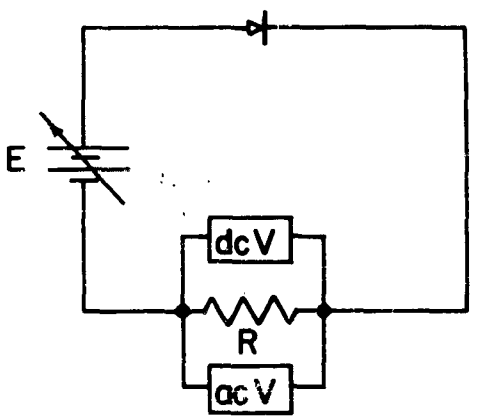
EXPERIMENTAL PROCEDURE AND DISCUSSION

The rectifiers to be studied were placed in a simple circuit, as shown in Figure 1. Temperature was monitored by packing a thermocouple into a hole drilled in the rectifier stud with dental gold foil. The rectifiers were attached to copper heat sinks which were placed in constant-temperature baths.

The intermittent current associated with the onset of microplasma conduction was monitored on a Model 310B Ballantine VTVM, which has a frequency cutoff of 6.7 mc. As the applied voltage is increased from the threshold of avalanche conduction to a value sustaining continuous conduction, the average duty cycle of the microplasma pulses increases from zero to infinity. The response of the a-c voltmeter will be a maximum at an average duty cycle of one half. Since the voltmeter response is also proportional to pulse height, which increases with applied voltage, the a-c voltmeter readings give only qualitative readings of the pulse current. However, a given value of the a-c voltmeter reading corresponds to a given average ratio of on- to off-time.

---

(4) R. H. Haitz, "Variation of Junction Breakdown Voltage by Charge Trapping", Phys. Rev., 138, 260-267 (April, 1965).



E = variable voltage power supply
R = 100 ohms
d-c V = d-c vacuum tube voltmeter
a-c V = a-c vacuum tube voltmeter

Figure 1

Circuit for Measuring Transient Voltage Effect

Because of the superposition of the characteristics of two or more microplasmas which are simultaneously conducting, only those rectifiers were studied that demonstrated a single microplasma breaking down at an appreciably lower voltage than any others. A typical separation between first and second microplasmas is shown in Figure 2, which is a plot of the noise envelope of the microplasma pulses as registered by the a-c voltmeter, superimposed on the direct current-voltage characteristic of the rectifier.

The transient component in the breakdown voltage is observed by switching the applied voltage from a prebreakdown value to a value in the range of intermittent conduction. In the n^+- p rectifiers, as shown in Figure 3, the voltage necessary to maintain a fixed ratio of on- to off-time of the microplasma pulses, and thus a constant microplasma current, decreased exponentially with time. The variable plotted against time after the application of the voltage is the difference between the voltage at time t and its final stabilized value. Thus, extrapolation of the curve to the origin gives the total amount of voltage change, and the slope of the curve represents a time constant for the process. The exponential decay curve may be represented by the empirical expression:

$$V_t - V_\infty = (V_o - V_\infty)\, e^{-t/\tau}, \qquad (4)$$

where t = time from application of voltage,
V_t = voltage at time t,
V_o = voltage at $t = 0$,
V_∞ = voltage at $t = \infty$,

$$\tau = \frac{t_1 - t_2}{\ln V_1/N_2}.$$

The curve with the time constant of 75 seconds was obtained by maintaining a peak reading on the a-c voltmeter. Thus, the average duty cycle of the microplasma pulses was maintained at a constant value of one half. The curve with a time constant of 910 seconds was obtained by maintaining the a-c voltmeter reading at a constant value of one half that of the peak reading on the leading edge of the noise envelope, i.e., $\tau_{on} < \tau_{off}$. Thus, as the relative on-time is decreased, the rate of the transition to a final steady-state condition is decreased. This situation is demonstrated in Figure 4 in which the decay time constant is plotted against the initial voltage required to sustain a given ratio of on- to off-time. The curve is seen to be roughly exponential.

The observation of a transient voltage decreasing with time is consistent with a decreasing value of N_D in Equation (2). Thus, during the microplasma pulse, electrons are captured by the traps at a rate proportional to the number of unoccupied traps and to their capture cross section. During the pulse interval, the traps release the captured electrons at a rate proportional to the number of trapped electrons and to the time constant of the traps. The rate of release of electrons from the traps is assumed to be slow with respect to the rate of capture. By the time the circuit has restored itself to a state which permits the initiation of the next microplasma pulse, not all of the captured electrons have been released; the number of ionized donors, N_D, is less than at the beginning of the preceding pulse; and V_B is accordingly less. This process continues as the traps continue to fill until the number of electrons released from the traps during the pulse intervals equals the number captured during the pulses.

The Voltage Transient in p^+- n Rectifiers

Utilizing p^+- n silicon controlled-avalanche rectifiers with breakdown voltages in the range from 750 to 1000 volts[5], similar experiments were performed. The sign of the voltage transient was positive, indicating the presence

(5) Courtesy of Standard Telecommunication Laboratories, Harwell, England.

of compensating acceptors in the n-type base region. The time constants were of the same order as those observed with p-type base devices. However, whereas the magnitude of the voltage transient in devices with p-type bases was in the range from 0.5 to 10 volts, that in devices with n-type bases was as high as 100 volts!

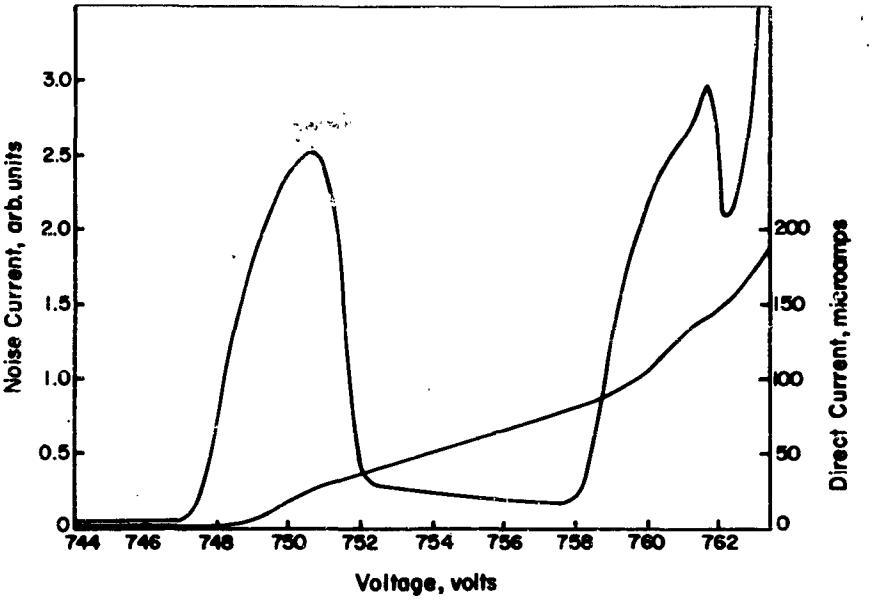


Figure 2

Direct Current and Noise Current
in the Avalanche Region

Temperature Dependence of the Breakdown Voltage Transient

The breakdown voltage transient was measured at several fixed temperatures between 195 and 320° K. Surprisingly, the magnitude of the voltage transient ($V_o - V_\infty$) did not change monotonically with temperature. Rather, as shown in Figure 5, a maximum was observed near 260° K. The maximum at this temperature was observed in every case on each of the rectifiers measured (temperature effect studies were made only on the n^+- p devices).

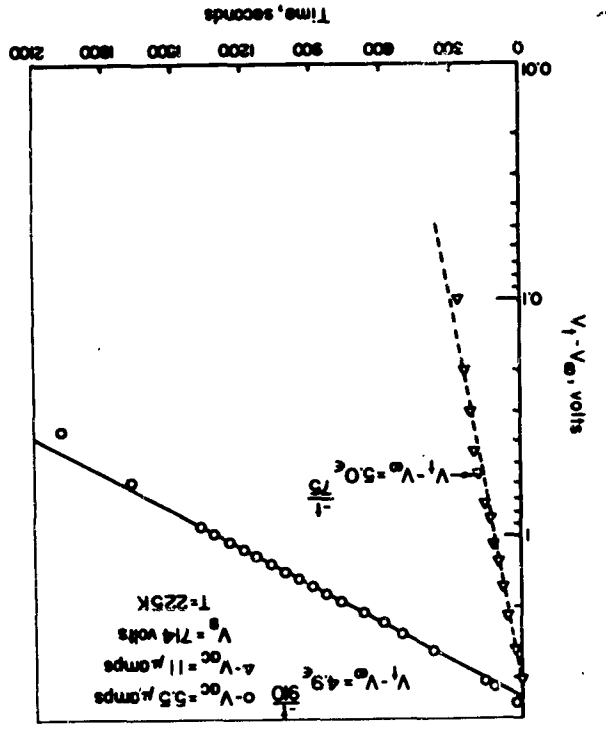


Figure 3

Decay of the Transient Component of Breakdown Voltage--Ratio of On- to Off-Time of the Microplasma Pulses as a Parameter

Figure 6, which is a plot of the temperature dependence of the Fermi level in the silicon comprising the rectifiers, shows that at 262° K, the Fermi level is positioned at 0.24 eV above the valence band, which is the energy level of the copper donor in silicon. $E_F(T)$ was calculated for the case of a single acceptor level comprising 2×10^{14} borons/cm^3 and also with the addition of a donor level at 0.24 eV with a concentration of 2×10^{13} cm^{-3}. It is seen that the presence of a compensating donor with a concentration up to 10 per cent of the majority carrier concentration has a negligible effect on the position of the Fermi level above 200° K.

Realizing the difficulties associated with invoking the concept of the Fermi level in a region of high injected hole and electron densities or in the region of high fields, both of which obtain during avalanche conduction, the

observed effects nevertheless suggest trapping phenomena consistent with the relationship between Fermi level and trap depth under equilibrium conditions.

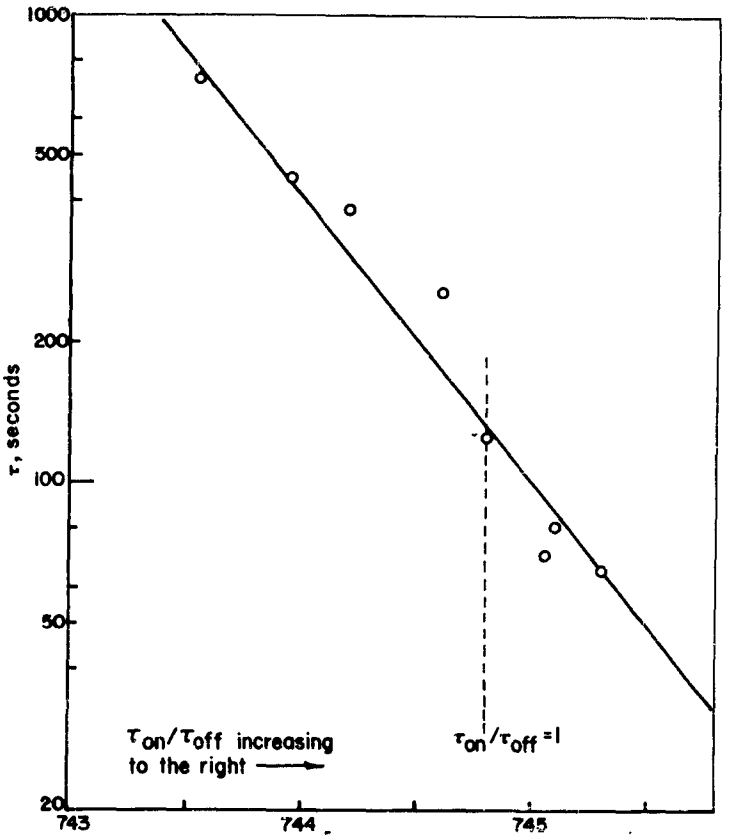


Figure 4

Dependence of the Transient Voltage Time Constant on the Relative Average On-Time of the Microplasma Pulses

Consider a quasi Fermi level for electrons, E_n, in the space-charge region. With the increase in carrier density in the space-charge region associated with microplasma conduction, E_n would shift toward the conduction band by an amount ΔE. In the temperature region where E_n would be several kT above the trap depth E_T, the value of the Fermi function and thus the value of N_D^+ (the ionized donor concentration) given by

$$N_D = N_T \left[1 - \frac{1}{e^{(E_T - E_n)/kT} + 1}\right], \qquad (5)$$

where N_T is the total number of trapping sites, would change but little from the equilibrium value, $N_D \approx 0$. Also, in the temperature range where $E_n << E_T$, a small change in E_n would have little effect on the value of the occupation index, and thus a large change in current density would have only a small effect on the equilibrium value, $N_D \approx N_T$. However, in the temperature range corresponding to Fermi energies near E_T, a small change in E_n would produce a marked change in the occupation index, and thus the value of N_D would depend strongly on the injected carrier density.

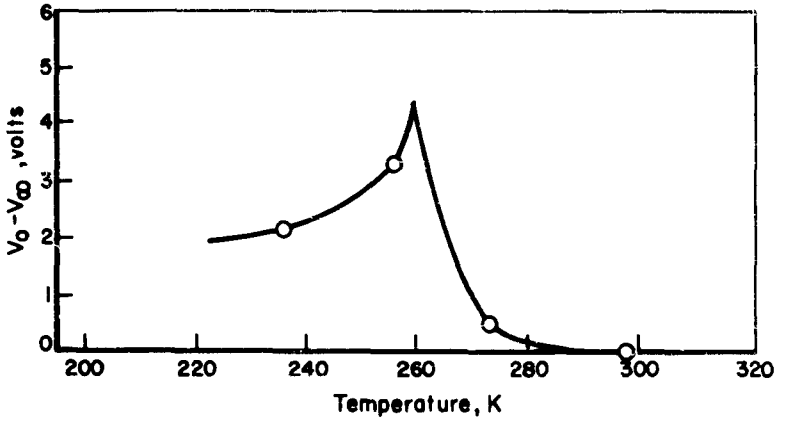


Figure 5

Temperature Dependence of the Magnitude of the Breakdown Voltage Transient - $V_{ac} = 1/2\ V_{ac}(max)$, $\tau_{on} < \tau_{off}$

Thus, at temperatures several kT above and below 262° K, the magnitude of the voltage transient is relatively small, indicating little change in N_D under the different conditions of high and low injected current density. However, near 262° K, where $E_n \approx 0.24$ eV, the magnitude of the voltage transient has been observed to be a factor of 10 or more

greater than at other temperatures, suggesting that the occupation density of the 0.24 eV level at 262° K depends strongly on the injected carrier density.

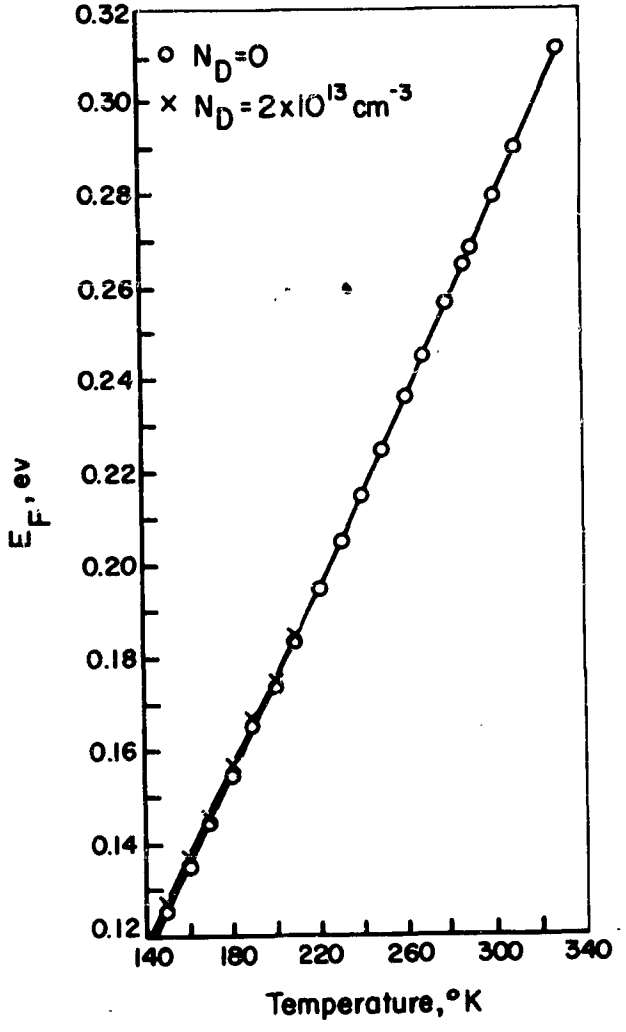


Figure 6

Temperature Dependence of Fermi Level in Silicon with p_o = 2 x 10^{14} borons/cm^3 and with 0.24 eV Donor Concentrations of 0 and 2 x 10^{13} cm^{-3}

In order to investigate further the possible correspondence between the position of the Fermi level and trapping effects in the space-charge region under intermittent avalanche conduction, measurements of the temperature dependence of the voltage transient were repeated after exposing the rectifiers to various doses of reactor neutrons. The energy levels in the bottom half of the band gap associated with the defects from various types of nuclear radiation are shown in Table 1.

Energy Level, eV	Type	Rediation	Remarks
0.31	donor	β	Largely responsible for lifetime reduction in n-type material[6]
0.30	--	β	Found in p-type silicon[7,8]
0.29	--	neutron[9]	--
0.27	acceptor	β, neutron[10]	Acceptor member of defect pair with lattice spacing < 50 Å[10,11]
0.05	--	(7)	--

Table 1

Radiation-Induced Energy Levels
Between E_v and 1/2 E_g

(6) G. K. Wertheim, "Energy Levels in Electron-Bombarded Silicon", Phys. Rev., 105, 1730-1735 (1957).

(7) D. E. Hill, "Electron Bombardment in Silicon", Phys. Rev., 114, 1414-1420 (1959).

(8) D. E. Hill and K. Lark-Horovitz, "Energy Levels Introduced into Silicon by Electron Irradiation", Bull. APS, Ser. 2.3.2, 142 (March 27, 1958).

(9) C. A. Klein and W. D. Straub, 3rd Semiannual Radiation Effects Symposium, Lockheed Aircraft Company, 2 (October 28-30, 1958).

(10) G. K. Wertheim, "Neutron Bombardment Damage in Silicon", Phys. Rev., 111, 1500-1505 (1958).

(11) H. Roth and V.A.J. Van Lint, 2nd Conference on Nuclear Radiation Effects on Semiconductor Devices, New York (September 17-18, 1959).

The result of massive irradiation is to drive the Fermi level to midband; however, the radiation level of 8.6×10^{12} neutrons/cm^2 (our maximum fluence) was not sufficient to compensate the p-type material in the rectifiers by as much as 1 per cent of the boron concentration of 2×10^{14} cm^{-3}, which is the limit of the concentration change detectable from measurements of the net ionized impurity concentration profile. Plots of the net ionized impurity concentration profiles across the base region of the rectifier before and after irradiation are shown in Figure 7. These plots are obtained from measurements of the voltage dependence of the junction capacity(12). (Charge carrier lifetimes, however, were degraded from 8 microseconds to less than 0.5 microseconds by a radiation fluence of 2.6×10^{12} neutrons/cm^2.)

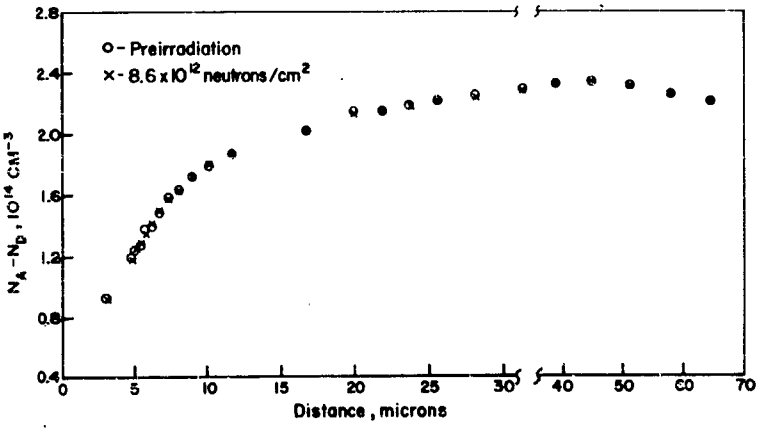


Figure 7

Net Ionized Impurity Concentration Profile Across Base Region of Rectifier Before and After Exposure to 8.6×10^{12} Reactor Neutrons/cm^2

(12) H. C. Gorton, "Studies of Impurity Profiles in Silicon p-n Junction Rectifiers", Physics of Failure in Electronics, Vol. 3, Edited by Goldberg and Vaccaro, Cato Show Printing Company, 355-364 (1965).

The acceptor level at 0.27 eV is the simple vacant lattice site which has a small activation energy for diffusion and is relatively mobile at room temperature[13]. It would be expected either to have disappeared at the surface or to have complexed with an interstitial impurity. Thus, the only available sites introduced by irradiation are at 0.29 and possibly 0.30 eV. Therefore, the temperature excursion was extended to 320° K to position the Fermi level at 0.30 eV, and indeed an increase in magnitude of the voltage transient was observed. (The effect of the introduction of levels at 0.29 or 0.30 eV is not confirmed, however, since measurements were not taken at this temperature prior to irradiation.)

A comparison of the temperature dependence of the voltage transient before and after exposure to 8.6×10^{12} neutrons/cm^2 is shown in Figure 8. In addition to an increase in the magnitude of the voltage transient at 320° K above the value observed at room temperature, a significant increase in the magnitude of the peak at 262° K was also observed. Such an effect is not unexpected, however, since copper exists in silicon in both interstitial and substitutional sites--the interstitial site being electrically inactive and predominent in concentration over the electrically active site by a factor of 1000 to 1[14].

From Equation (2), one may obtain

$$N_{D_\infty} - N_{D_o} = \frac{6.25 \times 10^{17} \, (V_\infty - V_o)}{V_o V_\infty} , \qquad (5)$$

or

$$\Delta N_D \approx \frac{6.25 \times 10^{17} \, \Delta V_B}{V_B^2} . \qquad (6)$$

Thus, the change in occupation density of the trapping centers, due to the increased carrier density in the space-charge region during microplasma conduction, may be calculated from Equation (6). Assuming values of 4 volts and 6 volts for the magnitude of the breakdown voltage transients at 262° K, as shown in Figure 8, values are obtained of $N_D \approx 5 \times 10^{12}$ cm^{-3} and 7.5×10^{12} cm^{-3} before and after irradiation, respectively. Assuming a defect introduction

---

(13) H. Y. Fan, Symposium on Radiation Effects in Semiconductors, International Conference on the Physics of Semiconductors, Paris (1964).

(14) C. B. Collins and R. O. Carlson, "Properties of Silicon Doped with Iron and Copper", Phys. Rev., 108, 1409-1414 (1957).

rate of 10 vacancy-interstitial pairs per neutron per cm, and a radiation level of 8.6×10^{12} neutrons/cm^2, a 6 per cent efficiency of the vacant lattice sites coalescing with interstitial copper atoms is required--not an unreasonable value.

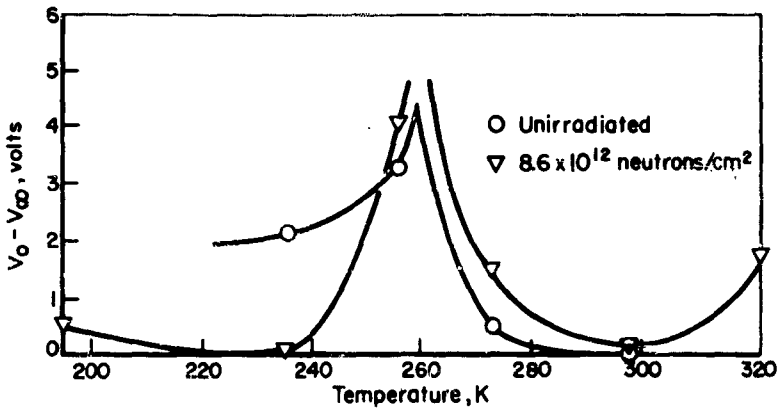


Figure 8

Temperature Dependence of the Magnitude of the Breakdown Voltage Transient Before and After Exposure to 8.6×10^{12} Reactor Neutrons/cm^2

CONCLUSIONS

A correlation between the magnitude of the transient component of the breakdown voltage and the coincidence of the Fermi level with the energy level of generation-recombination centers has been demonstrated. The transition of copper atoms from interstitial to substitutional sites by reaction with radiation-induced lattice vacancies and a value for the concentration of substitutional copper have been deduced from an analysis of the temperature dependence of the transient voltage effect. It is suggested that, by ordering the position and range of the Fermi level as a function of temperature by appropriately doping the silicon, the transient voltage effect could be used as a tool to investigate both chemical impurities and structural defects in silicon.

ACKNOWLEDGEMENT

Grateful acknowledgement is made to Gordon Kramer for helpful discussions and to D. A. Kaiser and W. L. Mefferd for their careful experimental work. Appreciation is extended to the member companies of the ECRC and to Standard Telecommunications Laboratories, Ltd., Harlow, England, for permission to publish.

ELIMINATION OF FORWARD-BIASED SECOND BREAKDOWN BY RESISTIVE BALLASTING OF SILICON POWER TRANSISTORS

Daniel Stolnitz

Radio Corporation of America
Somerville, New Jersey

INTRODUCTION

Failure of power transistors by second breakdown often results from severe non-uniformities in current density which may be initiated, for example, by lower emitter turn-on voltages in portions of the device. These, in turn, can arise from structural variations which may be systematic, as in the case of voltage drops along contact metallization, or random as in the case of diffusion-front spikes. The latter effect is simply illustrated by a glance at an I-V equation for transistor current in terms of emitter-base potential:

$$I = \frac{qA\, D_n n_p}{W} e^{qV_{BE}/mkT}$$

For fixed I: $$\ln W + \ln (\text{const.}) = \left(\frac{q}{mkT}\right) V_{BE}$$

Taking differentials: $$\left(\frac{mkT}{q}\right)\frac{\Delta W}{W} = \Delta V_{BE}$$

This equation shows how the consequences of base-width variations (ΔW) are magnified when the base is narrow, as is the case with high f_T transistors, or when high collector biases are applied.

For m=1, kT/q = 26 mv., a $\Delta W/W = 20\%$ leads to a ΔV = 5.2 mv. This is a small difference in terms of the magnitude of V_{BE}, typically about 600 mv., but it can give rise to significant non-uniformities of current.

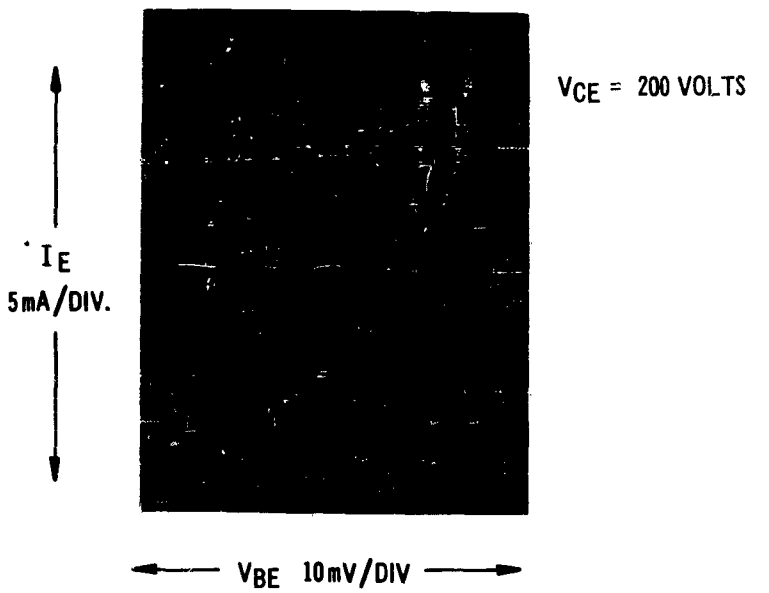


Figure 1

I_E vs. V_{BE} For Two Sites of Si Power Transistor

Fig. 1 shows the separation of I-V curves for two small, isolated sites on a large power transistor. (The loop results from the return trace of the oscilloscope.) Current variations translate into the thermal variations that are frequently displayed with phosphors or paints, and these in turn aggravate the initial current difference in a process that can become regenerative and finally unstable.

A solution to the inevitable non-uniformities of current density consists of dividing the emitter region into a large number of small, discrete areas and placing a limiting, debiasing resistance in series with each of them. Similarly, one might use a continuous, distributed resistance in series with an undivided emitter. In either case, an excess of current to any sub-region produces a larger voltage drop across the resistance in series with that area and reduces the forward bias applied. The principle is a simple one and is encountered in the parallel operation of power transistors, voltage regulator tubes, and recently silicon controlled rectifiers. Fig. 2 schematically shows the approach.

To reason quantitatively about the resistances required and their effect, one might examine the I-V equation for emitter current to the various sites. This can also be done graphically by load lines as seen

in Fig. 3. The I_E vs. V_{BE} characteristics must be obtained under the actual conditions of temperature and collector voltage that will apply. Then the maximum permissible site current I_{max}, and the minimum desired site current I_{min} are assigned. The intersection of these abscissae with the extreme cases of the characteristics will determine a load line that corresponds to an appropriate resistance. (See Appendix).

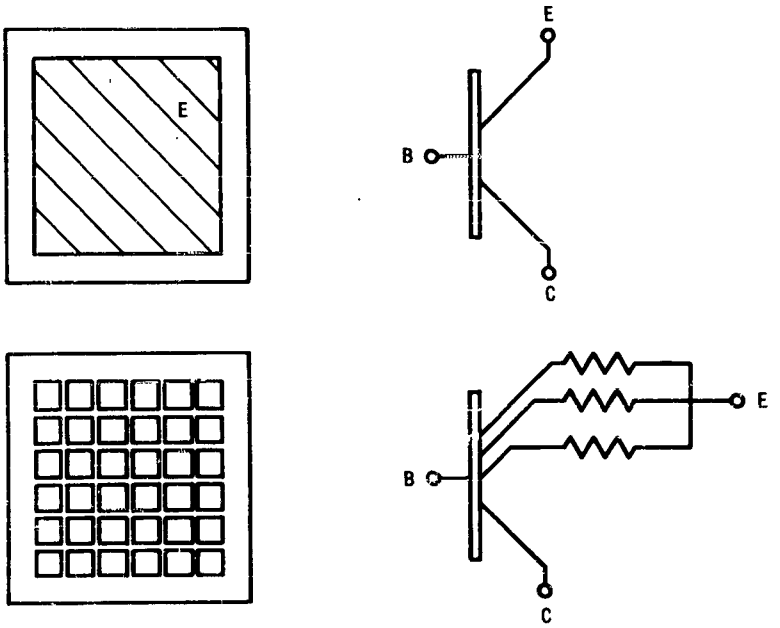


Figure 2

Principle of Ballasting Resistors

APPARATUS

To experimentally study the effect of resistive ballasting on forward-biased second breakdown, the following apparatus was constructed. First, a conventional triple-diffused silicon power transistor which employs an interdigitated contact geometry was modified as shown in Fig. 4.

Each of the eight emitter fingers (0.120" x 0.010") was divided into eleven segments (0.008" x 0.007") before diffusion and metallization, creating eighty-eight electrically discrete emitters. Base and

collector were left unaltered. Gold wires two-mils in diameter were then bonded to forty-eight of these emitter sites and taken out to two tiers of contacting clips. Each emitter clip lies in series with an external, variable resistance and a current sampling resistor which was held to 25 milliohms so as to be negligible. (Fig. 5) This arrangement allows us to simulate the use of any desired resistance uniformly in series with each emitter site, or to vary the limiting resistor independently to each site. Keeping the rheostats external allows the resistance value to be known accurately without the complication of changes with temperature.

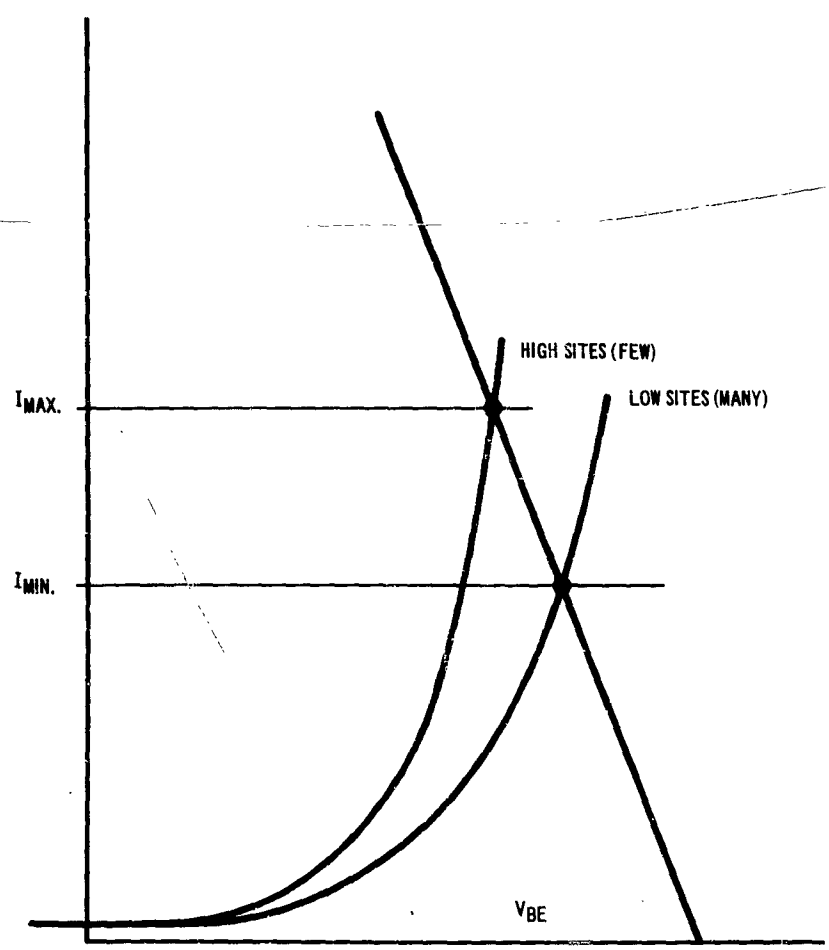


Figure 3

Load Line Determination of Ballast Resistance

Figure 4

Experimental Modification of Conventional Transistor

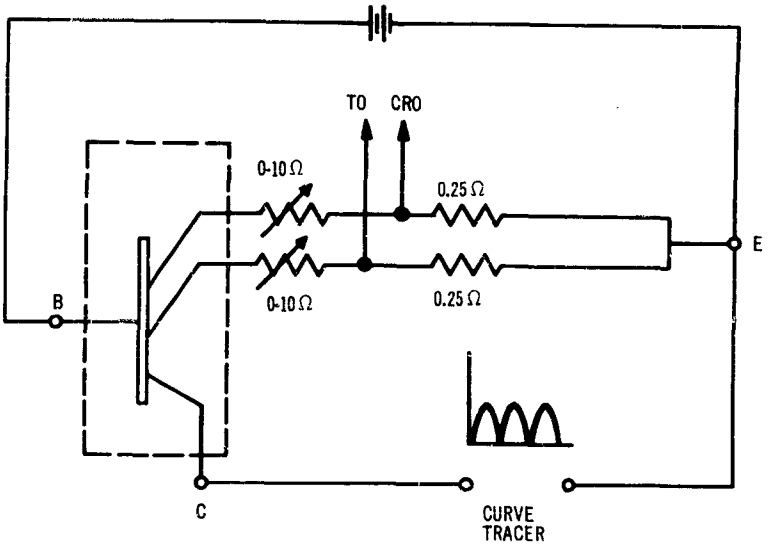


Figure 5

Biasing and Measurement Circuit

The collector was reverse-biased and pulsed by a Tektronix curve tracer. The emitter was forward-biased and driven from a D.C. current source. While some non-destructive test methods have been recently developed, the work reported here employed the condition of incipient second breakdown. The criterion used was the characteristic arching-back of the thermally-induced hysteresis loop that appears in the collector waveform. The results obtained correlate quite well with those found when units are taken completely into breakdown.

MEASUREMENTS

To demonstrate the improvement obtainable with ballasting resistors, Table 1 gives the variation of second-break-free power derived through resistors for twenty-four emitter sites operating at a collector voltage of 200 volts. The power column, labelled $P_{C.T.}$, is simply the product of maximum collector voltage and maximum collector current, as read from the curve tracer. Since all measurements are of the relative, before-and after type, this product is an adequate index. For undistorted waveform the continuous collector dissipation is calculated to be ~60% of the product $P_{C.T.}$

$V_{CE(max)}$	$I_{C(max)}$	$P_{C.T.}$	R
200 v	150 ma	30 w	0 Ω
200 v	250 ma	50 w	1.0 Ω
200 v	1000 ma	200 w	3.0 Ω
200 v	1000 ma	200 w	5.1 Ω
200 v	1000 ma	200 w	10.0 Ω

Table 1

Change in Second-Breakdown-Free Power with Ballast Resistance

A six-fold increase is observed from zero to 3 Ω. Five to ten ohms gives no significantly greater power, but provides a more closely uniform current distribution. For the various resistances in Table 1, each of the photographs of Fig. 6 shows the emitter current distribution among four of the twenty-four sites, including the worst, all superimposed on a single photograph. The abscissa is a time axis, with the distance between the triangular peaks corresponding to one complete cycle of collector sweep. The five photos reading from upper left to lower right correspond to the five values of resistance reading down the table.

The limit on power is set by the thermal resistance of the experimental unit. The factor of power increase is not really the important measure, as this can simply reflect an initially poor unit.

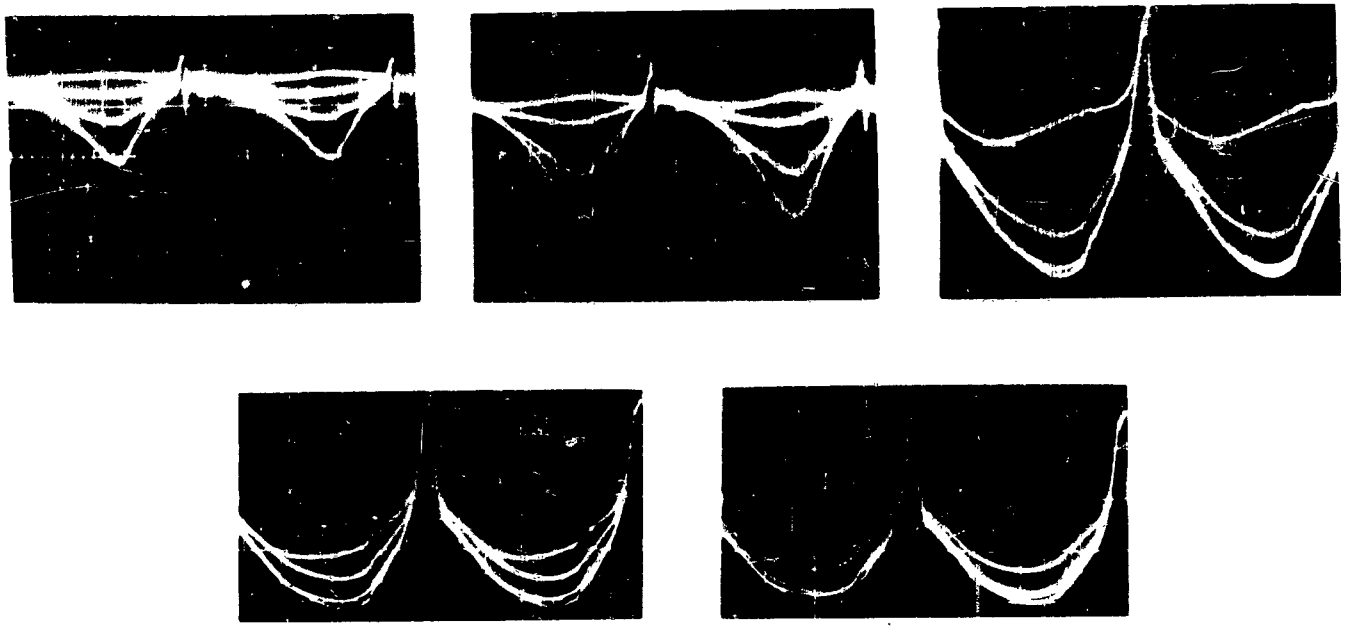


Figure 6

Change in Current Distribution with Ballast Resistance

What matters more is that dissipated power be independent of collector voltage and determined by thermal limits only. At a collector voltage of 150 volts, data similar to that obtained at 200 volts was found. 3 Ω was still required to secure the full power (200 w) but there was a 50% increase in the power obtained with 1 Ω . At a collector voltage of 100 volts, 1 Ω was sufficient to obtain the full power. It is seen that at lower collector voltages less debiasing resistance is required to prevent forward-biased second breakdown.

Thus for this experimental unit under the conditions of bias employed 3 Ω in series with each of twenty-four emitters ensures our obtaining the full second-breakdown-free power independently of collector voltage over the range zero to 200 volts.

It should be clear that the effective resistance placed in series with the emitter of the over-all unit is the parallel combination of all the ballasting resistors. Thus 3 Ω to each of 24 sites represents about a tenth of an ohm to the whole unit.

CURRENT HOGGING AND BASE-SOURCE IMPEDANCE

The process of forward-biased second breakdown can begin with the non-uniform turn-on of different emitter regions, but a central feature of the regeneration that leads to breakdown is the drawing away of current by the early turn-on sites from the others. This phenomenon, called "current-hogging", is clearly displayed in Fig. 7. Here the emitter current to four sites is shown as one site (the top trace) approaches second breakdown. Observe the rising hump in the traces of the three other sites at that region of the cycle in which the worst site draws the most current. These "humps" correspond to a reduction in emitter current for these three sites, since the magnitude of I_E increases as the waveform moves toward the bottom of the photograph.

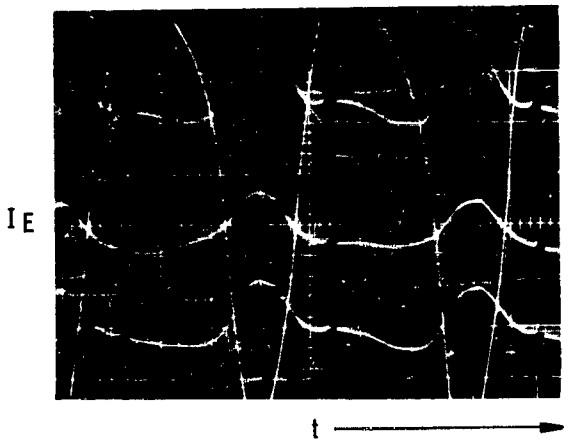


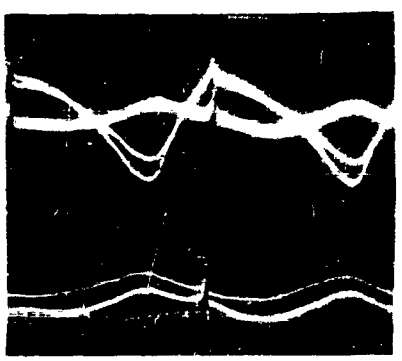
Figure 7

Current-hogging by Second-Breakdown Prone Site

A solution to this aspect of regeneration is to again divide the emitter region, and make the base generator a constant-voltage source rather than a constant-current source as it is typically. To demonstrate this, a transistor with only six emitter sites was used in order to observe instantaneously the current to all active sites. The results are the same when twenty-four or forty-eight sites are used.

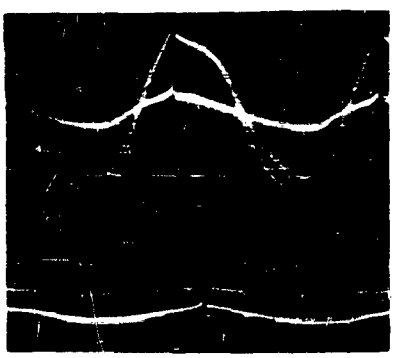
The photographs of Fig. 8 show the I_E distribution among the six sites for constant-current and constant-voltage base supply. In the constant-current case, observe the rising humps (decrease in current) in four traces as the leading two sites increase their current. By contrast, in the photograph of the constant-voltage distribution on the right, although the same two sites are each drawing perhaps four times the current of the others, there is no decrease in any site current over the cycle.

CONST. I



VERT. = I_E, HORIZ. = t

CONST. V



VERT. = I_E, HORIZ. = t

Figure 8

Emitter Sites Current Distribution for Constant-Current and Constant-Voltage Base Drive Generator

A possible effect on second breakdown of eliminating current-hogging in this way is offerred with the aid of Figs. 9, 10. In Fig. 9 for constant-voltage we see the I-V curves for both the worst site and the large number (m) of average sites. The intersection of these characteristics with the bias voltage ordinate gives the respective currents drawn. If the situation is stable nothing further occurs. But if the temperature-current situation is not stable for the worst site, then it will begin to draw more current and attain higher temperatures. This is shown for the worst site by the intersection of the bias-voltage ordinate with the I-V characteristics drawn for higher temperatures. The process is terminated when the worst site achieves some temperature T_F at which it draws a maximum current that causes it to be destroyed (I_{FAIL}).

The entire process occurs without any change in V_{BE}, and thus both the bias and the current to all the other sites remains unchanged. At failure, the total current drawn by the transistor is the sum of the failure current to the worst site plus m-times the current drawn by the average sites.

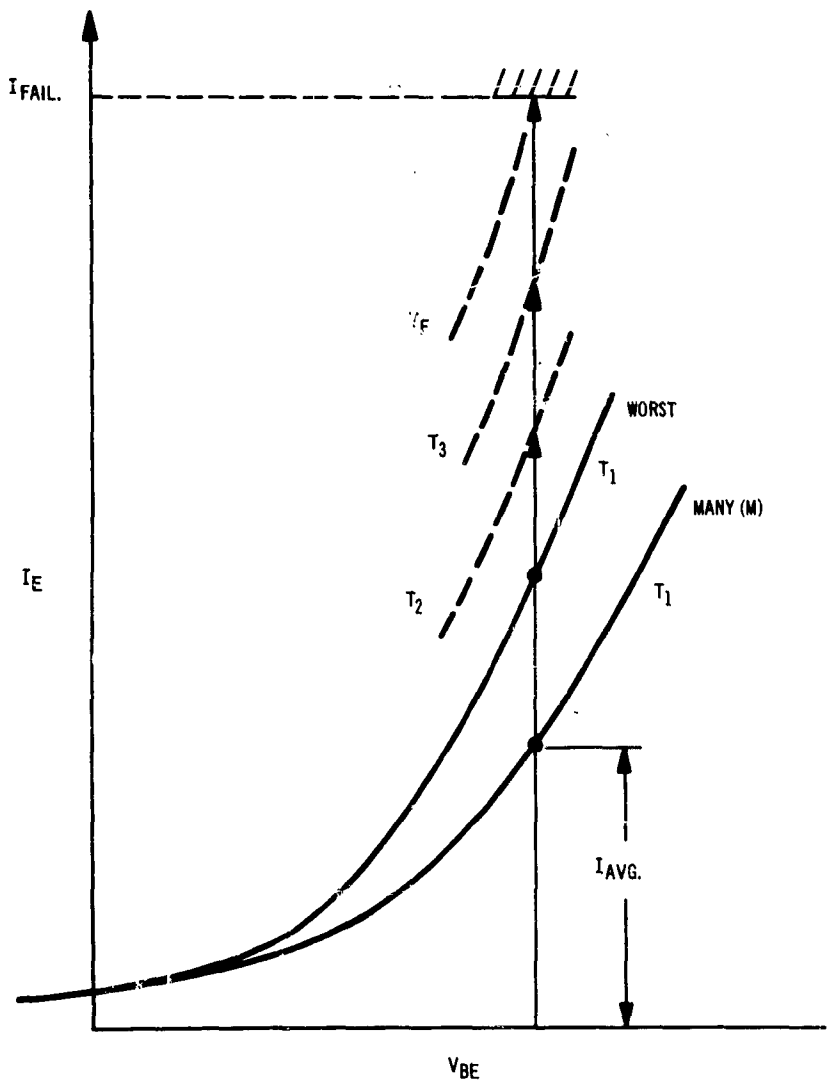
In Fig. 10 for the constant-current case the same initial I-V curves are shown. The process to be described is a dynamic one, but must be outlined by several discrete steps. The initial condition of bias is the ordinate "1" whose intersection with the characteristics gives the initial currents. If the worst site is unstable under this bias then it will tend toward higher currents and temperature. Its static impedance decreases and thus the load impedance presented to the base generator is reduced. The generator voltage decreases slightly as shown by ordinate 2. This reduces the average current to the many sites by some increment and makes available m-times this increment to the worst site. With this additional current the worst site can rise to some higher current and temperature shown by the intersection of ordinate 2 with the I-V characteristic for temperature T_2.

Since the worst site will be unstable at T_2, its impedance will decrease further and the process described will repeat, moving the voltage ordinate from 2 to 3. Fig. 10 shows this final change as producing failure by second breakdown. At failure the total current drawn by the transistor is the sum of the failure current to the worst site plus m-times the reduced current drawn by the average sites. This reduction in average current could lead to lower second-breakdown power under the constant-current condition.

Lastly, it should be clear that the impedance of the base generator must not be confused with that of the collector generator. As Scarlett has pointed out, the latter can have no effect on breakdown.

APPENDIX

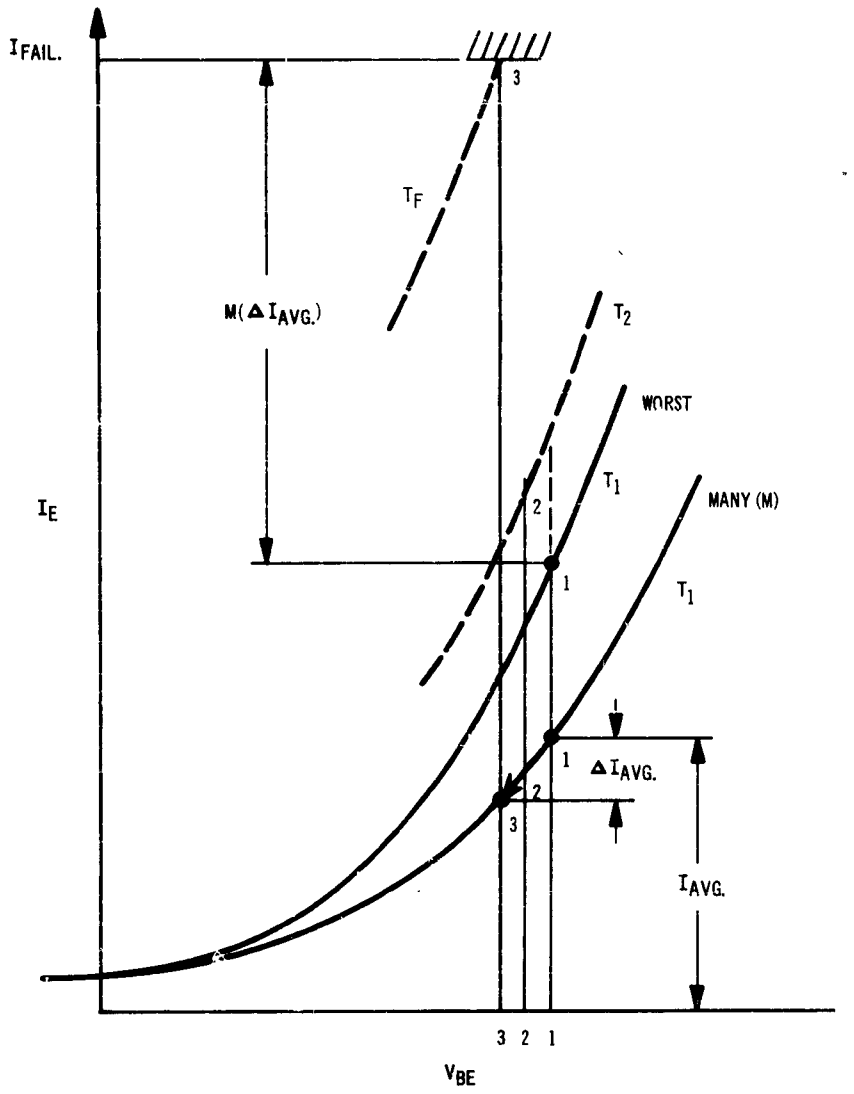
In thinking about the magnitude of resistance required for protective ballasting, one must keep in mind the effect of resistors on the current distribution as a function of temperature. In Fig. 11 (drawn for I-V curves at the same temperature) the initial current distribution is shown as the intersection of the ordinate "I" with the two I_E-V_{BE} characteristics. The re-distribution by ballasting is indicated by the intersection of a load-line (R_1 or R_2) with the same curves. It is seen that both a low resistance (R_1) or a high one (R_2) can produce fairly similar results. In Fig. 12 I-V curves for the high site at elevated temperature have also been included, since the heating of this site is an important feature of instability. Now it is seen qualitatively that as the temperature rises the high resistance provides much more favorable current distribution compared to the low resistance.



$I_{TOTAL} = I_{FAIL} + (M) I_{AVG.}$

Figure 9

Constant-Voltage Base Supply



$$I_{TOTAL} = I_{FAIL} + (M)(I_{AVG.} - \Delta I_{AVG.})$$

$$= I_{FAIL} + (M) I_{AVG.} - (M)(\Delta I_{AVG.})$$

Figure 10

Constant-Current Base Supply

plus m-times the reduced current drawn by the average sites. It is this reduction in average current which may account for the smaller breakdown power seen under the constant-current condition.

Lastly, it should be clear that the impedance of the base generator must not be confused with that of the collector generator. As Scarlett has pointed out, the latter can have no effect on breakdown.

APPENDIX

In thinking about the magnitude of resistance required for protective ballasting, one must keep in mind the effect of resistors on the current distribution as a function of temperature. In Fig. 11 (drawn for I-V curves at the same temperature) the initial current distribution is shown as the intersection of the ordinate 'I" with the two I_E-V_{BE} characteristics. The re-distribution by ballasting is indicated by the intersection of a load-line (R_1 or R_2) with the same curves. It is seen that both a low resistance (R_1) or a high one (R_2) can produce fairly similar results. In Fig. 12 I-V curves for the high site at elevated temperature have also been included, since the heating of this site is an important feature of instability. Now it is seen qualitatively that as the temperature rises the high resistance provides much more favorable current distribution compared to the low resistance.

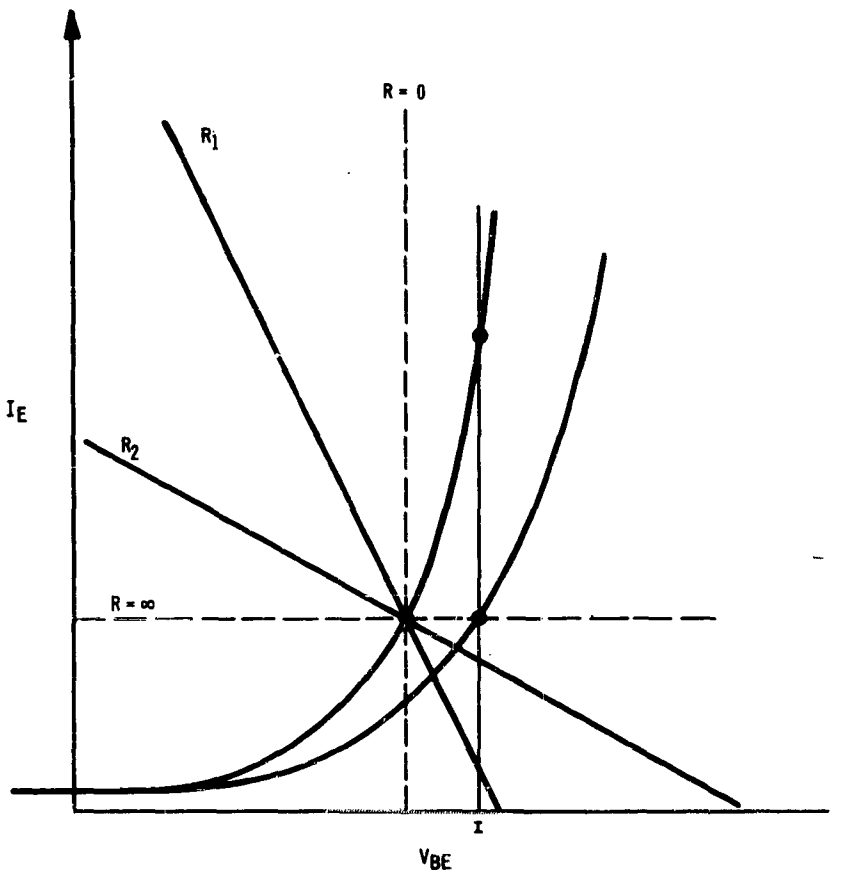


Figure 11

Effect of Ballast Resistors on Current Distribution

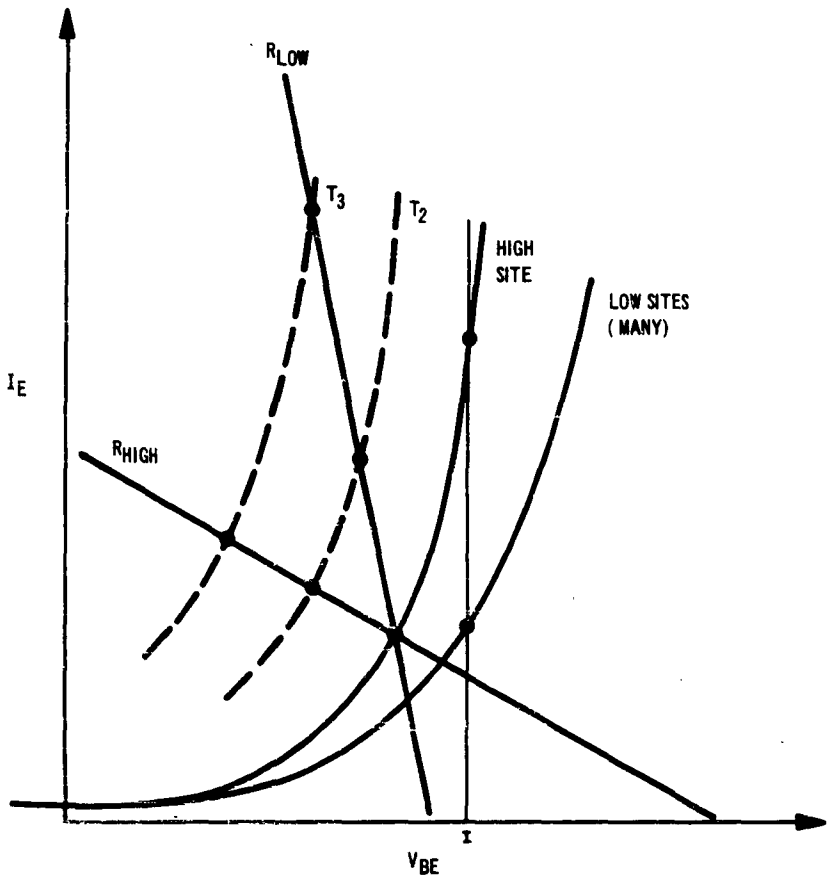


Figure 12

Effect of Ballasting on Temperature Dependence of Current Distribution

FAILURE OF LARGE-AREA EPITAXIAL-DIFFUSED SILICON DEVICES

T. L. Chu
Westinghouse Research Laboratories
Pittsburgh, Pa. 15235

and

P. J. Kannam
Westinghouse Semiconductor Division
Youngwood, Pa. 15697

ABSTRACT

The structural perfection of the epitaxial-diffused silicon films is one of the most important factors determining the performance and reliability of large-area epitaxial-diffused devices. The characteristics of these devices are affected by structural imperfections generated during the epitaxial growth and the diffusion processes.

The commonly observed imperfections in epitaxial silicon films are dislocations, stacking faults, growth pyramids, and polycrystalline inclusions. They all have adverse effects of varying degrees on device characteristics. Except those propagated from the substrate, structural imperfections in epitaxial silicon films can be essentially eliminated by using proper growth techniques.

The diffusion of high concentrations of dopants into epitaxial silicon films has been shown to generate dislocations and to transform stacking faults into other types of imperfections. Furthermore, localized areas of high dislocation density in epitaxial-diffused silicon films, due to contamination and improper handling techniques, have been shown to be responsible for the softness of large-area epitaxial-diffused junctions. The characteristics of these junctions were found to be improved considerably by using a gettering technique.

The epitaxial growth technique has been shown to be a valuable substitute for diffusion in reducing the failure of certain large area devices.

I. Introduction

Epitaxial silicon films are used extensively in the manufacture of low power devices, and have become increasingly important in the fabrication of large area devices. In many cases, the diffusion technique is also used for the further processing of epitaxial silicon films. The reliability of epitaxial-diffused devices is affected by the quality of the epitaxial material, the parameters of the diffusion process, the effectiveness of surface passivation, and subsequent mechanical operations such as contacting, lead attachment, encapsulation, etc. It is likely that the first three factors will ultimately limit the performance of large-area epitaxial-diffused devices. Unless these factors can be controlled reproducibly, the reliability of the device cannot be predicted. Thus, an investigation of the effects of imperfections in epitaxial silicon and diffusion parameters on device characteristics is a necessary approach to attack reliability problems in large-area epitaxial-diffused devices.

The requirements of structural perfection for large area devices are more stringent than those for small area, low power devices. While imperfections in epitaxial or epitaxial-diffused material only reduce the yield for low power devices, one major imperfection in a silicon wafer would cause the failure of the large area device. In this paper, the structural imperfections in epitaxial silicon and diffused silicon are correlated with the parameters in epitaxial growth and diffusion processes. The effects of these imperfections on the performance of devices and possible techniques of minimizing their occurrence are discussed. It is also shown that the epitaxial growth technique could be used in place of the diffusion technique for reducing the failure of certain large area devices, such as power transistors.

II. Experimental

II.1 Epitaxial Growth

The epitaxial growth of silicon was carried out by the pyrolysis of silane, using hydrogen as a diluent, on surfaces of heated silicon substrates in a flow system. The silicon substrates were of {111} orientation, 0.01 ohm-cm n-type, and were of 1" diameter. After mechanical lapping and chemical polishing in the usual manner, the substrates were supported on a susceptor in a quartz reaction tube, and the susceptor was heated externally by an rf generator. Prior to the deposition process, the substrates were chemically etched in situ with anhydrous hydrogen chloride (1) or water vapor (2). Subsequently, silicon films were grown using phosphine or diborane as a dopant for n- or p-type material, respectively.

II.2 Diffusion

The diffusion of boron into n-type epitaxial silicon films was carried out at 1150°C in a flow system using boron tribromide as a diffusant. Phosphorus oxytrichloride was used as an n-type diffusant, and the diffusion temperature was 1100°C.

II.3 Evaluation

Structural imperfections in epitaxial and epitaxial-diffused silicon films were investigated by chemical etching and optical microscope techniques. The Sirtl etch (100 g. anhydrous CrO_3, 200 ml.H_2O,

and 200 ml. 49% HF) (3) is very useful for this purpose. The intersections of imperfections with the surface exhibit faster etch rates than the adjacent crystal and are revealed as grooves, pits, etc.

The electrical characteristics of epitaxial or epitaxial-diffused junctions were evaluated by measuring their reverse current-voltage relations at room temperature. When the large area junctions (5 cm^2) exhibited high reverse currents at low voltages, these specimens were made into mesa structures by photolithographic techniques. Each mesa was about 20 mm^2 in area, and the reverse characteristics of these mesa junctions were than measured to isolate defective regions.

A number of power transistors with an active junction area of 4 cm^2 were fabricated by (a) the double diffusion technique using n-type epitaxial films deposited on low resistivity n-type substrates and (b) the phosphorus diffusion using epitaxial n- and p-films deposited on low resistivity n-type substrates. The voltage capabilities and gains of these transistors were measured by standard techniques.

III. Results and Discussion

III.1 Imperfections in Epitaxial Silicon Films

Structural imperfections are often observed in epitaxial silicon films (4). They could propagate from the substrate, originate at the substrate-film interface, or develop during growth. The most common imperfections are dislocations, stacking faults, growth pyramids, and polycrystalline inclusions.

Dislocations in bulk silicon crystals have been studied with respect to their effects on the properties of silicon and the electrical characteristics of silicon devices (5). The most direct influence of dislocations on device properties is related to the recombination of minority carriers and breakdown phenomena. Prussin (5b) has compared the reverse characteristics of a number of diffused mesa devices which were all identically prepared and differed from one another only on the basis of the dislocation density. In areas where dislocation arrays were present, the mesa devices experienced a decrease in the reverse voltage to a fraction of the voltage in the adjacent areas. The distribution of soft junctions also had a direct correlation with dislocation density. The indirect effects of dislocations in silicon, such as the enhancement of the diffusion of dopants along dislocations and the precipitation of impurities at dislocations, have been shown to be important factors for device failure. The diffusion enhancement along dislocations results in the formation of non-planar junctions, and the precipitation of impurities is a predominant cause of the soft reverse characteristics of p-n junctions.

The dislocations in epitaxial silicon films are the results of the propagation of nearly all dislocations in the substrate and the improper growth conditions. For example, the work damage and foreign impurity particles on the substrate surface could increase the dislocation density in the grown film. Furthermore, if the substrate was not adequately supported, the deformation of the substrate during the growth process could produce arrays of dislocations.

Stacking faults in epitaxial silicon films usually initiate at the substrate-film interface and propagate along {111} planes. When the substrate is of {111} orientation, the intersections of the stacking fault planes with the grown surface are revealed, by chemical etching, in the form of equilateral triangles, incomplete triangles, or straight lines. Experimental results indicated that structural defects at the substrate surface, foreign impurity particles on the substrate surface, and local stress developed during the growth process are predominant causes of faulting.

Stacking fault planes in epitaxial films are not expected to have direct harmful effects on device characteristics, since the atomic arrangement in the region of stacking fault planes is similar to that in twin planes and twin planes exhibit no observable effects on device properties. However, stacking faults always terminate at partial dislocations. For example, the adjacent faulting planes in stacking faults with triangular etch figures intersect at stair-rod partial dislocations with 1/6 <110> Burger's vectors. Thus, stacking faults may have similar effects on device characteristics as dislocations. Finch et al. concluded that p-n junctions in epitaxial silicon containing stacking faults did not exhibit anomalies which can be attributed directly to faulting (6). On the other hand, preferential microplasma breakdown at the stair-rod dislocations, presumably due to the precipitation of silicon dioxide, metals, etc. have been observed upon application of a reverse bias (7). The effects of stacking faults on the subsequent processing of epitaxial films such as diffusion and alloying were studied, using films containing high concentrations of stacking faults ($10^5 cm^{-3}$ or higher). Chemical etching and optical examinations of specimens subsequent to boron diffusion indicated that a major portion of the stacking faults has undergone rearrangement (cf. Section III.2). Chemical staining of the cross-sectioned surface showed no enhancement of diffusion along the remaining stacking fault planes or their intersections at a junction depth of 25 μ. Furthermore, stacking faults were also found to cause no localized deep alloy penetration into silicon.

Triangular growth pyramids, observed on epitaxial silicon films of {111} orientation, possess flat faces which are tilted from the underlying surface at small angles. No low angle boundaries greater than 0.5° are present at the edges of the pyramid. However, other structural imperfections such as stacking faults or polycrystalline inclusion may be present at the apex of the pyramid. Impurities on the substrate surface are presumably responsible for the formation of growth pyramids. Besides the associated imperfections, growth pyramids are undesirable for device purposes because of the non-planarity of the surface.

Polycrystalline inclusions in epitaxial silicon films usually originate at the substrate-film interface, owing to the presence of foreign particles which generated random nucleation. They have the most harmful effects on devices. Grain boundaries in the inclusions act as mobility scattering centers and carrier recombination centers. Junction devices containing inclusions in the depletion region exhibit high reverse currents and are incapable of supporting voltages more than 50V. The diffusion of impurities in silicon is also strongly enhanced along inclusions.

In summary, structural imperfections in epitaxial silicon films are undesirable because of their adverse effects on device characteristics. Except the imperfections propagated from the substrate, the formation of other imperfections can be attributed mainly to contaminants in the reactant mixtures, mechanical damage in the substrate, and foreign impurities on the substrate surface. By observing proper cleanness precautions and using suitable *in situ* etching techniques and high purity reagents, these imperfections can be essentially eliminated. Thus epitaxial silicon films possess great promise for the fabrication of large area devices. Conventional silicon crystals have been shown to exhibit various types of chemical and structural defects (8), and epitaxial filicon films can be superior in these aspects.

III.2 Diffusion

Diffusion of dopants into silicon is perhaps the most widely used fabrication technique. The surface concentration of several dopants in silicon and the junction depth can be controlled reproducibly to a high degree. However, the diffusion technique has the inherent disadvantage that dislocations are introduced into silicon due to the lattice mismatch between the silicon and the dopant atoms (9). The diffusion of a high concentration of boron or phosphorus into a shallow surface layer of silicon produced slip lines due to the stress from dopant atoms in the silicon lattice. The presence of silicon phosphide precipitate has been detected in the surface layer of silicon after the diffusion of a high concentration of phosphorus (10). The introduction of dislocations and second phase material during the fabrication of p-n junctions has serious implications for the quality of devices.

Diffusion-induced imperfections in epitaxial silicon films were investigated by using optical examinations and microplasma observations. When the substrate and the epitaxial film differed greatly in dopant concentrations, the diffusion of dopant during the growth process resulted in high concentrations of dislocations in the interface region. For example, when a boron-doped substrate of 0.01 ohm-cm resistivity was used for the deposition of a 20 ohm-cm n-type epitaxial film of 200μ thickness, the grown material just above the interface showed a high density of etch pits arranged in arrays along $\langle 110 \rangle$ directions. This is illustrated in Fig. 1 where the etch figure of the surface 2 to 4μ above the interface is shown. The grown material above this region and the substrate below this region had few dislocations. Thus, the observed etch pattern resulted from the dislocations induced by the diffusion of boron from the substrate into the grown film. Dopants with smaller misfits in silicon, such as arsenic, gallium, or aluminum could reduce considerably the dislocations at the epitaxial interface.

A guard ring test junction, with cross-sectional configuration shown in Fig. 2, was used to evaluate the effects of phosphorus diffusion on the breakdown characteristics of p-n junctions (11). The guard junction is much deeper than the central junction and thus has a higher breakdown voltage. The central junction should breakdown uniformly except for regions of imperfections or local resistivity variations. To permit the observation of the light generated in the junction, the central junction was one micron or less in depth. Epitaxial silicon films used for the diffusion process were of 0.1 ohm-cm p-type and differed widely in stacking fault densities (1 cm^{-2} to 6 x 10^4 cm^{-2}).

In nearly all cases, microplasmas were observed under a reverse bias at the periphery of the central junction, the region labeled A in Fig. 2. Figure 3a shows the top view of a completed structure, and Fig. 3b illustrates the microplasmas observed. These microplasmas presumably occurred at dislocations generated during the formation of the guard junction by diffusion, and these dislocations would intersect the junction and served as sites for localized breakdown.

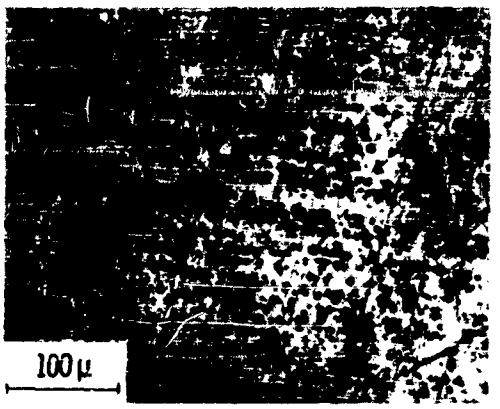


Figure 1.

Etch Figure of Surface Near the Interface of Epitaxial Silicon Grown on Low-Resistivity Boron-Doped Substrate (Sirtl etch, 3 min)

When the guard ring test junction fabricated from epitaxial films containing a high concentration of stacking faults was subjected to a reverse bias, no microplasmas related directly or indirectly to stacking faults were observed. When the epitaxial film was etched before diffusion, however, there was a one-to-one correspondence between microplasma breakdown and stacking faults. Detailed examinations indicated that stacking faults were transformed during diffusion into other imperfections and that the nature of the transformation was dependent on whether the epitaxial film had been etched prior to the diffusion process. This phenomenon is rather complex and will not be further discussed here.

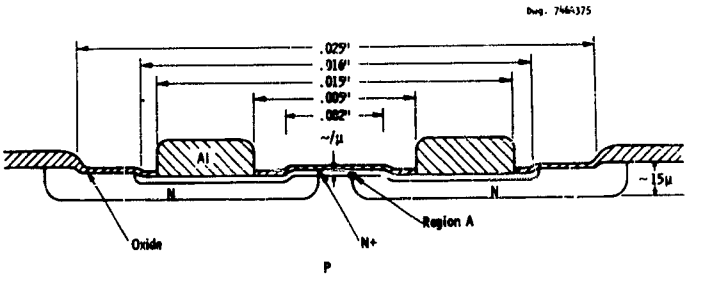


Figure 2.

Cross Section of Guarded Junction Structure

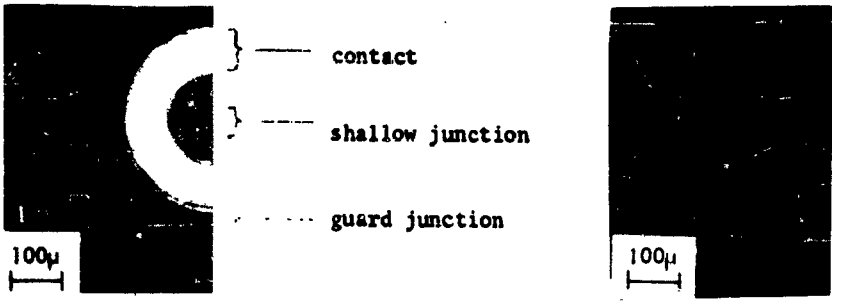


Figure 3.

(a) Top View of a Guarded Junction Structure (left)

(b) Microplasmas From Guarded Phosphorus-Diffused Junction (right)

Besides imperfections produced at high dopant concentrations due to the lattice mismatch, the processing techniques used for diffusion may also introduce imperfections into silicon. The processing techniques are particularly important for the formation of shallow junctions. For example, the formation of "pipes" through the diffused layer has been attributed to contamination (12). The effects of processing parameters, such as the cleanness of the specimen surface and the handling of specimens, on the properties of large-area shallow junctions were investigated in detail in connection with our program on power transistors. The transistors were fabricated by the double diffusion technique using 10 ohm-cm n-type epitaxial silicon films deposited on low resistivity n-type substrates. The depth of the base diffusion was 5 to 6μ, and the surface concentration of boron in silicon was approximately 10^{18} cm^{-3}. The collector junctions, approximately 5 cm^2 in area, sometimes exhibited high reverse currents at low voltages, 50 mA or higher at 100V or lower. To correlate the electrical characteristics with imperfections in epitaxial-diffused junctions, the specimens with high reverse currents were made into 19 mesas, and the junction area of each mesa was about 20 mm^2. It was found that mesas with poor characteristics invariably had localized areas with etch figures of various shapes associated with high concentrations of dislocations. Several examples are shown in Fig. 4 and 5. The etch figures shown in Fig. 4 were most likely formed by contaminants on the specimen surface before or during the diffusion process. Contaminants such as dust particles, solvent residues, impurities in the carrier gas of the diffusants, etc., could react with silicon at the temperature used for the diffusion process, and the stress resulting from such reactions would be sufficient to generate dislocations. The etch figures shown in Fig. 5 were apparently due to tweezer or probe marks introduced into the specimen by various handling processes before diffusion, and the stresses caused by these damages were relieved during the diffusion process by the formation of dislocation networks. These observations clearly illustrate the importance of foreign contaminants and handling damage as quality-detracting factors in large-area epitaxial-diffused devices. Imperfections introduced by these factors are more serious in large area devices than in low power devices.

To further investigate the effects of process-induced imperfections on the electrical characteristics of epitaxial-diffused junctions, specimens with poor characteristics were subjected to a gettering process using phosphorus pentoxide at 900°C for 1/2 hr. (13). The reverse characteristics of mesas in a large area epitaxial-diffused specimen before and after the gettering process are summarized in Table I, and the configuration of the mesa structure is shown in Fig. 6. This specimen had unusually poor characteristics because of the presence of many localized areas of imperfections. Since the reverse characteristics of most of the mesas were found to show considerable improvements after the gettering process, the observed softness of the junctions is due partly to the presence of metallic precipitates. The tendency for heavy metals to precipitate in the damaged area is therefore more harmful than the dislocations themselves or the non-planar junctions which they sometimes cause.

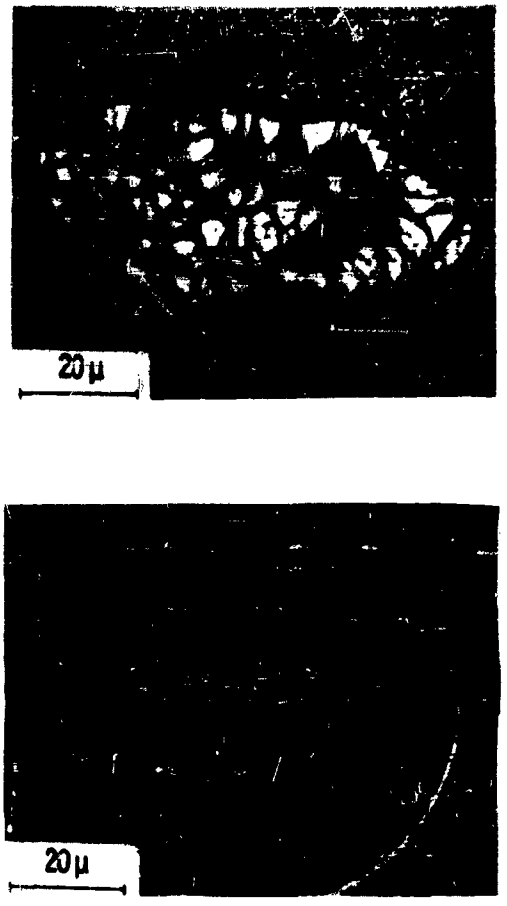


Figure 4.

Etch Figures Showing Clusters of Dislocations in Epitaxial Silicon Specimens after Boron Diffusion, Sirtl Etch, 2 min.

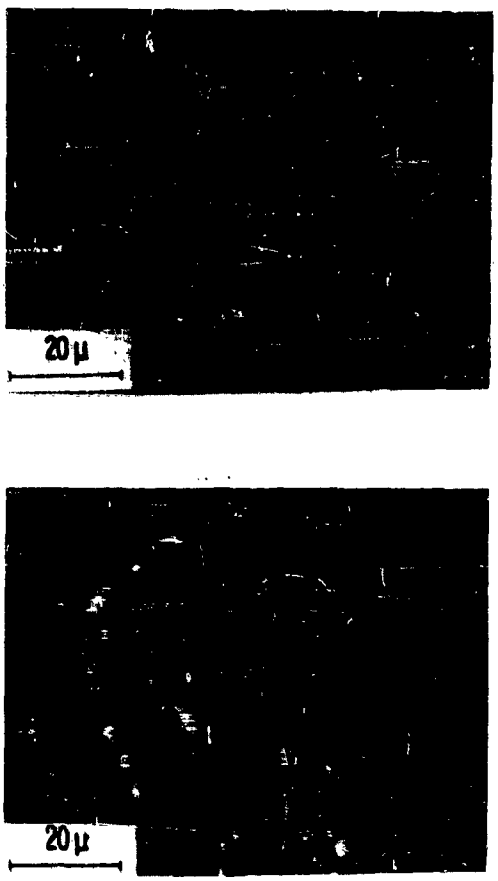


Figure 5.

Etch Figures Showing Deep Scratches and Associated Dislocations in Epitaxial Silicon Specimens after Boron Diffusion, Sirtl Etch, 2 min.

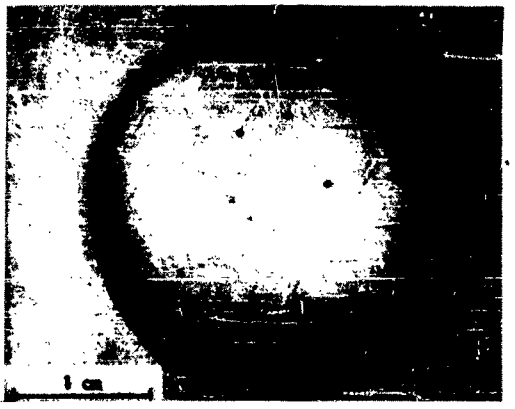


Figure 6.

Mesa Structure Used for Evaluating Quality of Small Regions of a Large Area Epitaxial Structure. The Black Dots in the Mesas Are Regions Where Destructive Breakdown Occurred Because of a Localized Defect Area.

III.3 Epitaxial Growth

In spite of the extensive applications of diffusion in the fabrication of devices, it is difficult to produce large area shallow junctions free from process-induced dislocations by the diffusion technique in a routine manner. On the other hand, the epitaxial growth technique in which the junction structures are produced during the crystal growth process is more attractive in many instances. In the power transistor under consideration, the preparation of the base region by the epitaxial growth technique simplifies considerably the fabrication technique and could also provide better electrical characteristics of the collector junction. Experimentally, a low resistivity p-type film of a few microns thickness, the base region of the transistor, was deposited subsequent to the deposition of medium resistivity (5 - 20 ohm-cm) n-type epitaxial silicon film, the collector region, on low resistivity n-type "carrier" substrates. Large area epitaxial p-n junctions prepared in this manner usually exhibited good electrical characteristics; the characteristics of mesas in a specimen are given in Table I for comparison. The large area epitaxial p-n structures, after emitter diffusion, have produced transistors with reasonably good characteristics, as shown in Table II.

Table I

IV Characteristics of Mesas in Diffused and Epitaxial P-N Junctions

Mesa	Diffused Junction		Epitaxial Junction
	After Boron Diffusion	After Gettering	(Without Gettering)
1	100V, 80mA	100V, 20mA	200V, 1mA
2	30V, 80mA	80V, 5mA	200V, 1mA
3	80V, 80mA	65V, 5mA	200V, 1mA
4	200V, 15mA	NO CHANGE	200V, 2mA
5	200V, 5mA	NO CHANGE	200V, 1mA
6	10V, 90mA	70V, 90mA	200V, 1mA
7	30V, 90mA	50V, 90mA	200V, 2mA
8	200V, 15mA	NO CHANGE	200V, 1mA
9	200V, 15mA	NO CHANGE	200V, 1mA
10	200V, 8mA	NO CHANGE	200V, 5mA
11	90V, 80mA	SHORT	200V, 3mA
12	40V, 30mA	100V, 50mA	200V, 4mA
13	100V, 90mA	SHORT	200V, 3mA
14	160V, 50mA	SHORT	200V, 1mA
15	40V, 50mA	100V, 15mA	200V, 2mA
16	90V, 80mA	100V, 10mA	180V, 1mA
17	6V, 80uA	80V, 90mA	180V, 2mA
18	36V, 90mA	100V, 50mA	180V, 1mA
19	32V, 90mA	100V, 10mA	180V, 2mA

Table II

Summary of Characteristics of Large Area Epitaxial Base Transistors

	V_{CE}	V_{CB}	h_{FE} at 10A
1.	170V at 1mA	175V at 1mA	15
2.	120V at 10mA	130V at 10mA	33
3.	120V at 20mA	130V at 20mA	15

IV. Summary and Conclusions

Structural imperfections in epitaxial and epitaxial-diffused silicon films are presumably the ultimate limitations on the reliability of large-area epitaxial diffused devices. They are more serious in large area, high power devices than in low power devices since one major imperfection in a wafer would cause the failure of the large area device.

The commonly observed imperfections in epitaxial silicon films are dislocations, stacking faults, growth pyramids, and polycrystalline inclusions. Their principal causes of formation and effects on device properties are summarized in Table III. All these imperfections have adverse effects of varying degrees on device characteristics, the polycrystalline inclusions being the most harmful. Except those propagated from the substrate, structural imperfections in epitaxial films can be essentially eliminated by using proper growth techniques thereby minimizing the failure of devices.

Structural imperfections in diffused silicon could be inherent from the diffusion process or introduced by processing techniques, as summarized in Table III. Optical examinations and microplasma effects have been used to study the effects of dopant diffusion in epitaxial silicon films. The process-induced imperfections, mostly in the form of localized areas of high dislocation density, have been observed in large-area epitaxial-diffused silicon films due to contaminations and improper handling techniques. These imperfections have been shown to be responsible for the softness of p-n junctions by providing preferential sites for the precipitation of metallic precipitates. The cleanness of the diffusion process and the proper handling of specimens are therefore important in reducing the process-induced imperfections.

The epitaxial growth technique may be used in place of diffusion in reducing the failure of certain diffused devices. It has been shown that in the fabrication of power transistors, the use of epitaxial base has improved the characteristics of the collector junction and simplified the fabrication process.

ACKNOWLEDGEMENTS

The authors wish to thank E. D. Wolley and R. Stickler for their contributions in microplasma studies. The major portion of the research reported in this paper was supported by the George C. Marshall Space Flight Center of the National Aeronautics and Space Administration under Contracts NAS8-11432 and NAS8-5335.

Table III. Imperfections in Epitaxial and Diffused Silicon

Imperfection	Principal Cause	Important Effects	Seriousness
Dislocations, epitaxial silicon	Propagation from substrate Deformation of substrate Contaminants on substrate surface	Decrease of lifetime, mobility, and reverse voltage Enhancement of dopant diffusion Preferential precipitation of impurities	Concentrations $>10^4 cm^{-2}$ usually have noticeable effects on devices, directly or through secondary effects Relatively easy to control
Dislocations, diffusion-induced	Stress from lattice mismatch	Localized breakdown Preferential precipitation of impurities	Noticeable effects on shallow junction devices Difficult to eliminate
Dislocation clusters mechanically induced	Handling damage Surface contamination	Soft, low-voltage junctions Localized current conduction, hot spot phenomena, second breakdown	Very serious effects on devices Can be eliminated by careful techniques
Stacking faults	Defects at substrate surface Impurities on substrate surface Stress during growth	Transformation to other defects during diffusion No direct effects on junction properties	Probably not serious at $<10^3 cm^{-2}$ Secondary effects at high conc. can be serious Relatively easy to elimin.
Growth pyramids	Impurities on substrate surface	Non-planar junction	Relatively serious Easily eliminated
Polycrystalline inclusions	Impurities on substrate surface	Harmful because of grain boundaries Soft, low-voltage junctions	The most serious defects in epitaxial silicon Can be eliminated
Precipitates	Exceeding solubility limits Chemical interaction Impurities in reactants and etchants used for surface preparation and epitaxial growth	Local shorting Sites for second breakdown General degradation of junction characteristics Change in diffusion profile	Serious defect of varied origins can be minimized by using pure reactants and proper handling techniques

REFERENCES

(1) G. A. Lang and T. Stavish, "Chemical Polishing of Silicon With Anhydrous Hydrogen Chloride," RCA Rev., 24, 488-498 (1963).

(2) T. L. Chu and R. L. Tallman, "Water Vapor as an Etchant for Silicon," J. Electrochem. Soc., 111, 1306-1307 (1964).

(3) E. Sirtl and A. Adler, "Chromsäure-Flusssäure als spezifishchs System zur Ätzgrubenentwicklung auf Silizium," Z. Metallk., 52, 529-531 (1961).

(4) See, for instance,
(a) T. B. Light, "Imperfections in Germanium and Silicon Epitaxial Films," in Metallurgy of Semiconductor Materials (AIME Metallurgical Society Conferences, Vol. 15), J. B. Schroeder, Ed., Interscience, New York, 1962, p. 137-158.

(b) G. R. Booker and R. Stickler, "Crystallographic Imperfections in Epitaxially Grown Silicon," J. Appl. Phys., 33, 3281-3290 (1962).

(c) T. L. Chu and J. R. Gavaler, "Imperfections in Vapor Grown Silicon," In Metallurgy of Advanced Electronic Materials (AIME Metallurgical Society Conferences, Vol. 19), G. E. Brock, Ed., Interscience, New York, 1963, p. 209-225.

(5) See, for instance,
(a) W. E. Taylor, W. C. Dash, L. E. Miller, and C. W. Mueller, "The Role of Dislocations in Device Properties," in Properties of Elemental and Compound Semiconductors (AIME Metallurgical Society Conferences, Vol. 5), H. C. Gatos, Ed., Interscience, New York, 1960, p. 327-335.

(b) W. Shockley and A. Goetzberger, "The Role of Imperfections in Semiconductor Devices," in Metallurgy of Semiconductor Materials (AIME Metallurgical Society Donferences, Vol. 15), J. B. Schroeder, Ed., Interscience, New York, 1962, p. 121-135.

(6) R. H. Finch, H. J. Queisser, G. Thomas, and J. Washburn, "Structure and Origin of Stacking Faults in Epitaxial Silicon," J. Appl. Phys., 34, 406-415 (1963).

(7) H. J. Queisser and A. Goetzberger, "Microplasma Breakdown at Stair-Rod Dislocations in Silicon," Phil. Mag., 8, 1063-1066 (1963).

(8) J. W. Faust, Jr., H. F. John, and R. Stickler, "The Role of Micro-Defects in Silicon Starting Materials as Quality Reducing Factors in Semiconductor Devices," This Symposium.

(9) See, for instance
(a) H. J. Queisser, "Slip Patterns on Boron-Doped Silicon Surfaces," J. Appl. Phys., 32, 1776-1780 (1961).

(b) S. Prussin, "Generation and Distribution of Dislocations by Solute Diffusion," J. Appl. Phys., 32, 1876-1881 (1961).

(10) P. E. Schmidt and R. Stickler, "Silicon Phosphide Precipitates in Diffused Silicon, "J. Electrochem. Soc., 111, 1188-1189 (1964).

(11) R. Stickler, E. D. Wolley, and T. L. Chu, "Effects of Phosphorus Diffusions in Epitaxial Silicon Layers," J. Electrochem. Soc., 112, 260C (1965).

(12) A. Goetzberger, "Impurity-Induced Pipes Through Diffused Layers in Silicon," Solid State Electronics, 5, 61-70 (1962).

(13) S. W. Ing., Jr., R. E. Morrison, L. L. Alt, and R. W. Aldrich, "Gettering of Metallic Impurities From Planar Silicon Diodes," J. Electrochem. Soc., 110, 533-537 (1963).

A LIMITATION TO THE STEP STRESS TESTING CONCEPT FOR INTEGRATED CIRCUITS

W. Shurtleff and W. Workman

Texas Instruments, Inc.
Dallas, Texas

INTRODUCTION

The concept of step stress testing, briefly stated, is to accelerate time-dependent failure modes by increasing stress levels in discrete steps until a large percentage of the test sample fails in a relatively short time span. The failure modes observed are assumed to be active at lower levels and would have caused failures after extended times. "Time compression" by the utilization of step stress techniques is desirable; however, the failure modes observed must be carefully screened to determine whether or not they would have occurred at lower stress levels. If, indeed, it can be shown that the failure modes observed would not occur at lower levels, no useful information has been obtained. Simple laboratory tests may be used to determine the onset of failures that would not occur at normal "use" conditions. It is the intent of this paper to describe certain examples for determining upper stress limits for the step stress testing of integrated circuits.

The word stress denotes the act of subjecting a device to mechanical, thermal, and/or electrical factors which cause the device to deviate from its equilibrium condition. The particular examples chosen in this paper will be restricted to the simultaneous application of thermal and electrical stress. Thermal-electrical stress can be stepped in three obvious ways:

(1) increase electrical power dissipation with ambient temperature fixed,
(2) increase ambient temperature with power dissipation held constant, and
(3) simultaneous increase of ambient temperature and power dissipation.

To determine the useful maximum limits for the preceding tests, consider the following stress limitations:

(1) bar temperature must not be increased beyond the eutectic point of the materials composing the units,
(2) bar temperature must not be such that thermal biasing of components occurs,
(3) reverse voltages across p-n junctions must be less than the reverse breakdowns of these junctions unless current flow is limited to prevent violation of items (1) and (2), and
(4) the operating test circuit must be transient free to prevent transient latch-on of p-n-p-n action.

Eutectic Limitation

When the eutectic point of the materials comprising the unit is surpassed, the unit will become a molten mass resulting in device destruction (one example is the gold-silicon eutectic which occurs at 370°C); hence, bar temperatures should not exceed or reach this temperature. This obvious limitation is often violated by designers of step stress tests.

Thermal Biasing Limitation

Thermal biasing refers to thermally generated transistor action which occurs when the impurity carrier concentration of a doped material becomes "swamped" by the thermal generation of carriers. Figure 1[1] gives the behavior of the resistivity as a function of temperature for n type (phosphorous) silicon. Figure 2[1] gives this information for p type (boron) silicon. As may be observed, the resistivity of both n and p type silicon increased with temperature to a maximum value, then decreases as temperature is further increased. The region of negative slope is the intrinsic range of the impurity semiconductor and is caused by thermal generation. The materials having the lower acceptor or donor concentration are more susceptible to the onset of the intrinsic region. For example, an n type material with 10^{13} phosphorous atoms/cubic centimeter will begin to decrease in resistivity when its temperature exceeds 100°C. A p type material with 10^{20} boron atoms/cubic centimeter will surpass 1000°C before carriers produced by thermal generation "swamp" out impurity carrier concentration.

To illustrate how the resistivity behavior with temperature produces thermal biasing consider the following example. Suppose three regions are juxtaposed such that an n type region of 10^{15} atoms/cubic centimeter is sandwiched between two regions of p type doping of levels 10^{16} and 10^{13} atoms/cubic centimeter respectively. Further suppose that their cross sectional areas are identical and the 10^{16} and

10^{13} regions are the same thickness while the 10^{15} region is very thin in comparison (effectively of zero resistance). Place a bias of 15 volts across these three layers with 10^{16} being the most positive terminal. These regions may be viewed as a resistor voltage divider where the potential of the n region with respect to the 10^{16} p region is

$$V_{10^{16} \longrightarrow 10^{15}} = - \frac{\rho_{10^{16}}}{\rho_{10^{13}} + \rho_{10^{16}}} \qquad 15$$

At 0°C, the ratio of resistivities in the preceding equation may be determined from Figure 2 to be .001; hence, the potential of the n region with respect to the 10^{16} p region is -.015 volts. The effective transistor is turned "off". At 200°C, the ratio of resistivities is approximately .04 and the potential would be -.6 volts. The effective transistor would be biased "on". This thermally generated transistor action will be referred to as thermal biasing.

Reverse Voltage Limitation

Reverse voltages that exceed breakdown voltages of p-n junctions coupled with unlimited current produce high power dissipation leading to increase in bar temperature. This increased temperature may result in thermal biasing of other components and/or eutectic formation.

Transient Limitation

Any four layer device may serve as a p-n-p-n rectifier[2] and transients may act as a gating signal causing "latch-on" and device destruction due to increased bar temperature.

Further Clarification of the Preceding Stress Limitation

To further clarify the preceding stress limitations, a discussion of their effect upon integrated circuit components and their parasitic components will be presented. This discussion will be restricted to triple diffused, monolithic integrated circuits that utilize reverse p-n junction isolation. Also, the above discussion of the eutectic point limitation is adequate and will not be discussed further.

Component Transistors of Integrated Circuits

A typical diffusion profile of an n-p-n transistor is shown in Figure 3. The basic substrate material is p type (boron doped) silicon into which successive n (phosphorous), p (boron), and n+ (phosphorous) diffusions are made to form the collector, base and emitter regions of the transistor. The electrical equivalent circuit as shown in Figure 4 may be described as two transistors; namely, an n-p-n transistor and a p-n-p transistor. The upper three layers form the

n-p-n, and the lower three layers form the p-n-p. The two interjacent layers are common to both transistors.

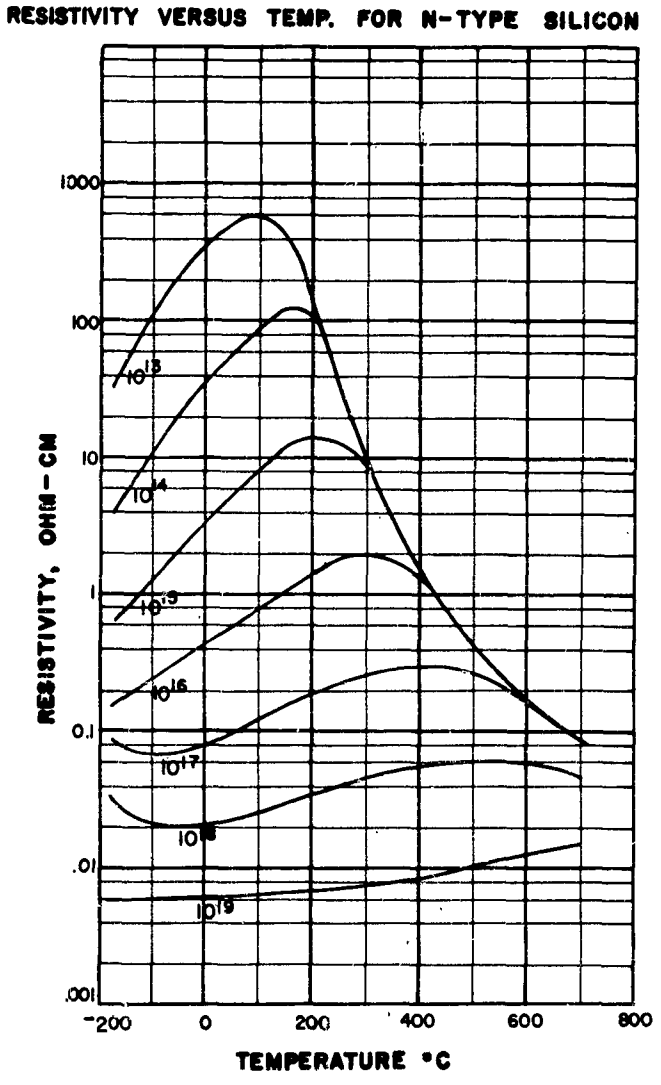


Figure 1

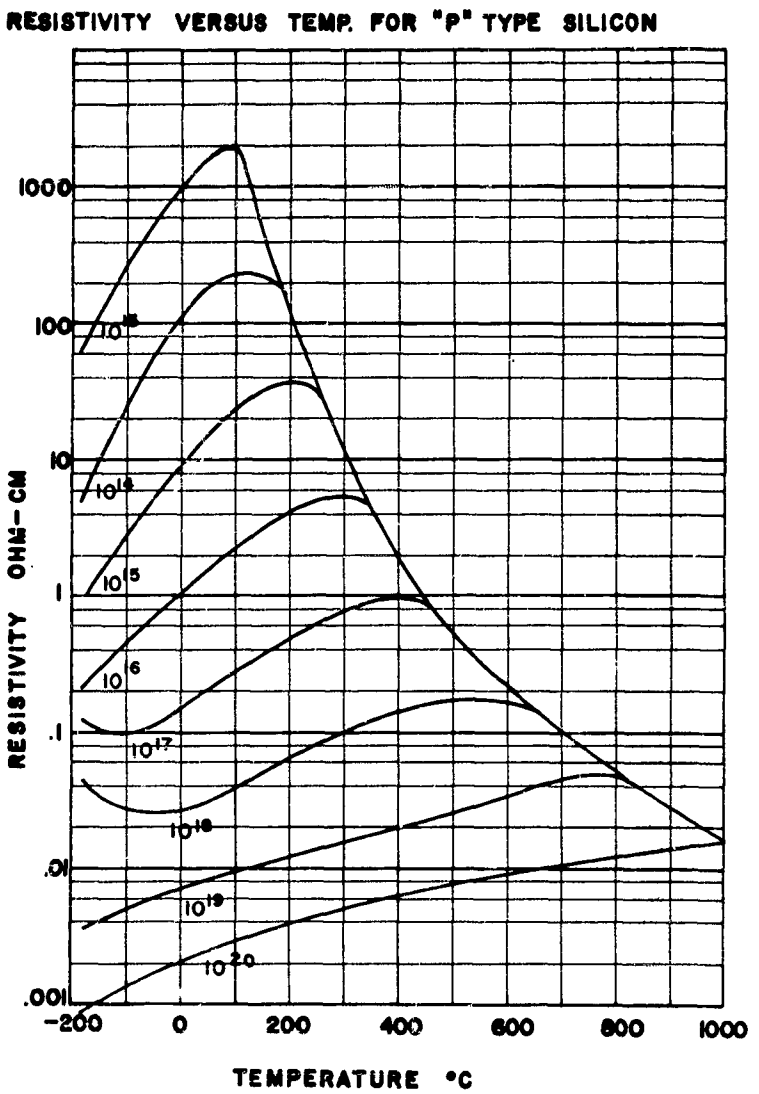

RESISTIVITY VERSUS TEMP. FOR "P" TYPE SILICON
RESISTIVITY OHM-CM
1000
100
10
1
.1
.01
.001
10^{13}
10^{14}
10^{15}
10^{16}
10^{17}
10^{18}
10^{19}
10^{20}
-200
0
200
400
600
800
1000
TEMPERATURE °C

Figure 2

TRANSISTOR

TYPICAL DIFFUSION PROFILE

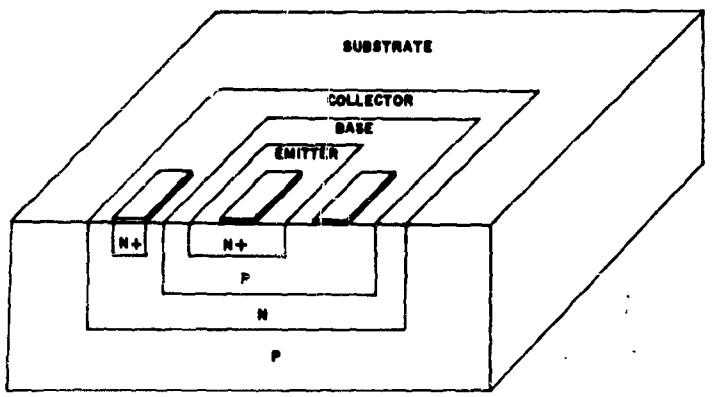


Figure 3

TRANSISTOR

PARASITIC EQUIVALENT

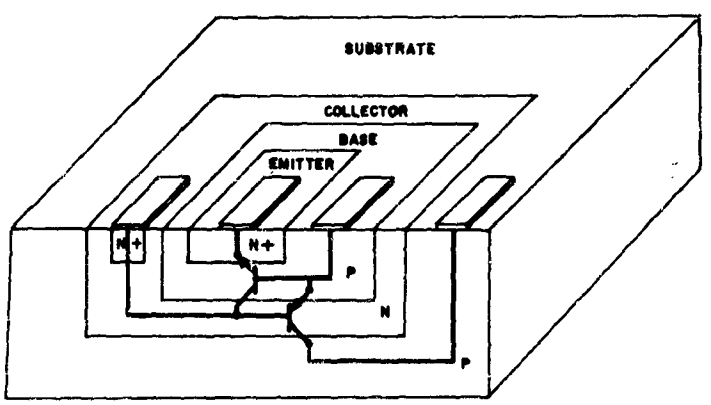


Figure 4

Thermal Biasing Limitation of Integrated Circuit Transistors

Examination of the equivalent circuit shown in Figure 4 reveals that transistors whose bases are floating; i.e., no external electrical connection to the base, are particularly susceptible to thermal biasing. The hypothetical transistor discussed above is a direct analogy to this situation. The parasitic p-n-p transistor of an n-p-n transistor is also susceptible if the collector of the n-p-n (the base of the parasitic p-n-p) is floating. However, general practice by designers of step stress tests is to turn-off components (reverse-bias stress) or turn them on into their saturation condition (power dissipation stress). This prevents thermal biasing as bases are generally biased externally and not allowed to float. Also, the doping level of the emitter, base and collector regions of the n-p-n are high and the intrinsic region lies above the eutectic point of the contact in silicon systems. This is not true, however, in general for the parasitic p-n-p transistor.

Reverse Voltage Limitation for Integrated Circuits Transistors

In designing a step stress test, consideration should be given to the fact that resistivity shifts with temperature; hence, reverse breakdown voltages shift to lower values when the silicon is thermally driven into its intrinsic region. Driving any of the junctions of the transistor into avalanche and not limiting current will produce high power dissipation resulting in localized eutectic formation and/or thermal biasing of neighboring components. Hence, failure to observe this limitation will lead to the meaningless information that a unit can indeed be melted.

Transient Limitation for Integrated Circuit Transistors

Examination of the equivalent circuit shown in Figure 4 reveals that the transistor and its parasitic transistor are not equivalent to the two transistor analogy of a p-n-p-n controlled rectifier. The basic difference lies in the mode of bias application; namely, the polarity is reversed. Hence, the transistor is in general not directly subject to transient latch-on of p-n-p-n action.

Component Diodes of Integrated Circuits

A typical diffusion profile of a diode is shown in Figure 5. The basic substrate material is p type (boron doped) silicon into which successive n (phosphorous), p (boron), and n+ (phosphorous) diffusions are made to form the diode isolation tank, its cathode, and its anode. By comparing Figures 3 and 5, it is easily seen that the diffusion profile of a diode and a transistor are identical.

Hence, the equivalent circuit of the diode shown in Figure 6 is the same as that of the transistor shown in Figure 4. This basic structure may be modified into the following diodes:

(1) a diode with low reverse breakdown and normal forward voltage characteristic,
(2) a diode with a high reverse breakdown and normal forward voltage characteristic, and
(3) a zener diode.

Type (1) diode is formed by either electrically connecting the diode isolation tank to the cathode or allowing the tank to "float". The reverse breakdown is the same as the reverse emitter-base breakdown voltage of the n-p-n transistors. Type (2) diode is formed by using the diode isolation tank as a cathode. Hence, its reverse breakdown is the same as the collector-base breakdown of the n-p-n transistors. A zener diode may be formed by adjusting the resistivity of the isolation tank until it is equivalent to that of the final n diffusion. By making electrical contact to the n regions, a zener diode with essentially an emitter-base reverse breakdown under both voltage polarity conditions is formed.

DIODE
TYPICAL DIFFUSION PROFILE

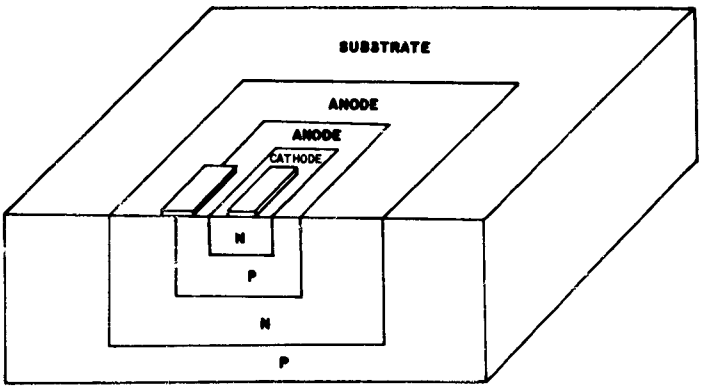


Figure 5

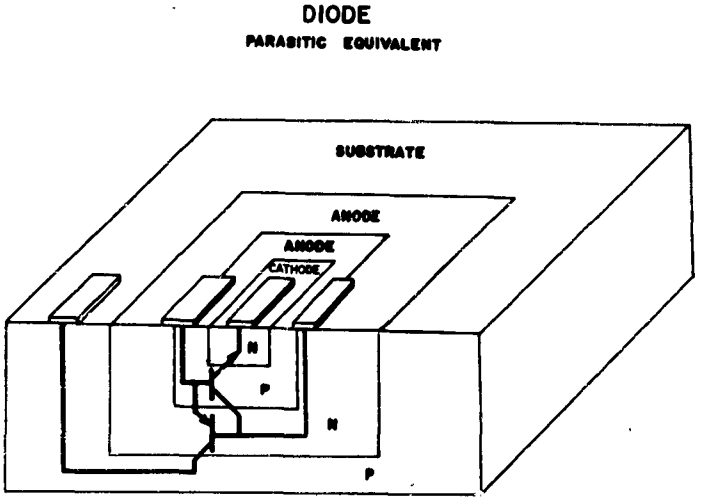


Figure 6

Thermal Biasing Limitation for
Integrated Circuit Diodes

Examination of the equivalent circuit shown in Figure 6 reveals that diodes whose isolation tanks are floating are particularly susceptible to thermal biasing. The hypothetical transistor discussed above is a direct analogy to this situation. Diodes whose isolation tanks are ohmically connected to their anodes are not subject to thermal biasing.

Reverse Voltage Limitation for
Integrated Circuit Diodes

The remarks about the application of this limitation to transistors are applicable and need no further discussion.

Transient Limitation for
Integrated Circuit Diodes

The diode, as the transistor, is not in general subject to the transient latch-on of p-n-p-n action.

Component Capacitors of Integrated Circuits

A typical diffusion profile of a capacitor is shown in Figure 7. The order and type of diffusions are identical to those for the formation of an n-p-n transistor. The diffusions differ only in regard to their placement. The final n+ diffusion is not centered in the p diffusion;

rather, it overlaps the n diffusion and leaves a section of the p diffusion exposed as shown in Figure 7. The n+ and n diffusions form one plate of the capacitor, and the p diffusion serves as the other, i.e., the p-n/n+ junction formed is reverse biased, and the resultant depletion layer forms the capacitor.

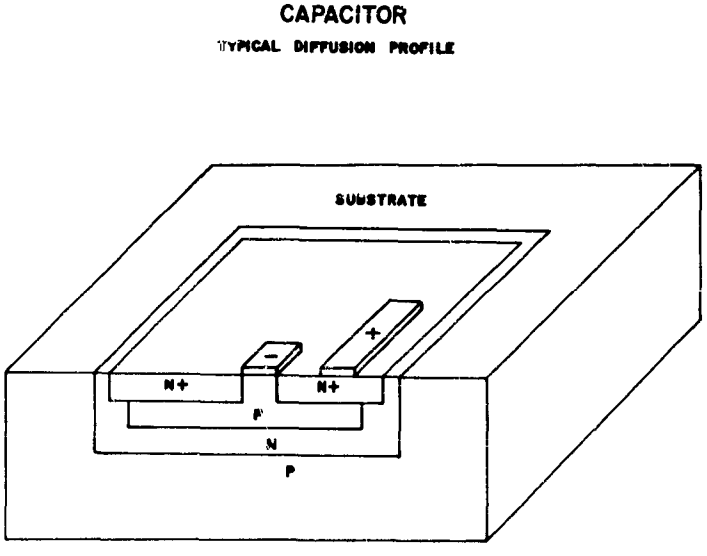


Figure 7

The equivalent circuit of the capacitor is given in Figure 8. It can be represented by a p-n-p transistor. The n+ and n diffusions serve as the base, the p type substrate as the collector, and the p diffusion as the emitter.

The capacitor is not sensitive to the above discussed limitations of transistors and diodes. Thermal biasing can be neglected since the junctions of the capacitor are formed from heavily doped material; therefore, the intrinsic temperature range is high. Transient latch-on can be ignored as the device is three layered. Reverse voltage limitations must be observed, but the heavy doping levels prevent shifting of reverse breakdown voltages.

Component Resistors of Integrated Circuits

Resistors are formed by two successive diffusions, n type and p type, respectively, into a p type substrate. See Figure 9. The n diffusion forms the resistor isolation tank and the narrow p diffusion forms the resistor.

CAPACITOR

PARASITIC EQUIVALENT

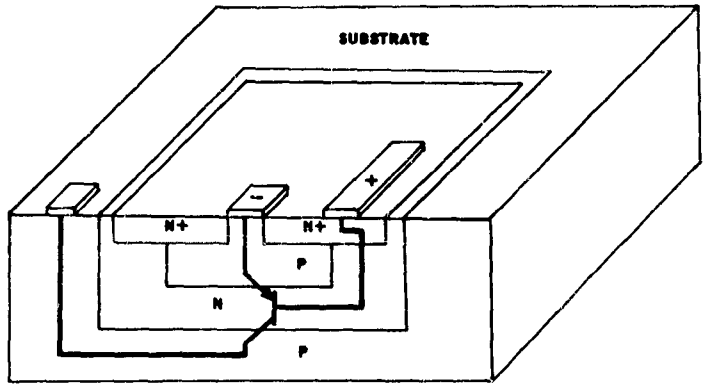


Figure 8

RESISTOR

TYPICAL DIFFUSION PROFILE

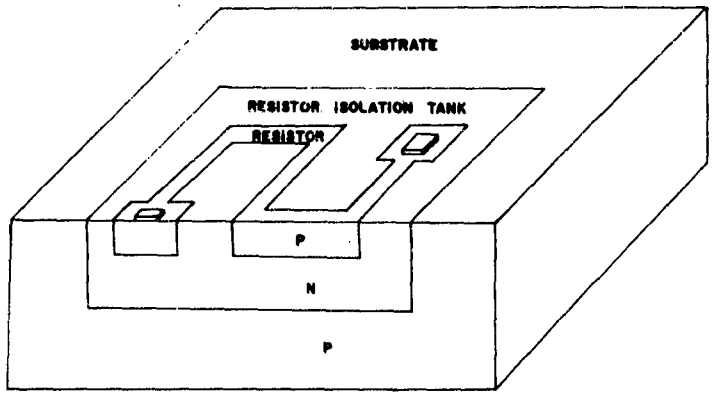


Figure 9

The electrical equivalent of the resistor (refer to Figure 10) is a dual emitter p-n-p transistor with a resistor connecting its emitters. The base of the p-n-p is the resistor isolation tank, and its collector is the substrate. The resistor is not directly subject to p-n-p-n latch-on as it is a three layer device. It is sensitive to the reverse voltage limitation in the same manner as the transistor and diode. Hence, the remarks made earlier are again applicable. The resistor is exceptionally sensitive to thermal biasing. The resistor isolation tank may or may not be externally biased. Consider first the case where it is not biased. Here the hypothetical transistor discussed above is again a direct analogy; namely, the elevation of temperature shifts the silicon to its intrinsic range and causes the voltage of the tank to move toward that of the substrate. As this happens, it becomes forward biased with respect to the resistor, and the equivalent p-n-p transistor turns-on leading in many cases to device destruction. If the tank is externally biased, this tendency is reduced but not eliminated because it is of large geometry and hence, is difficult to maintain at a particular bias. In other words, voltage gradients across the tank are generated by the application of temperature. These gradients allow thermal biasing of the resistor isolation tank leading to turn-on.

RESISTOR

PARASITIC EQUIVALENT

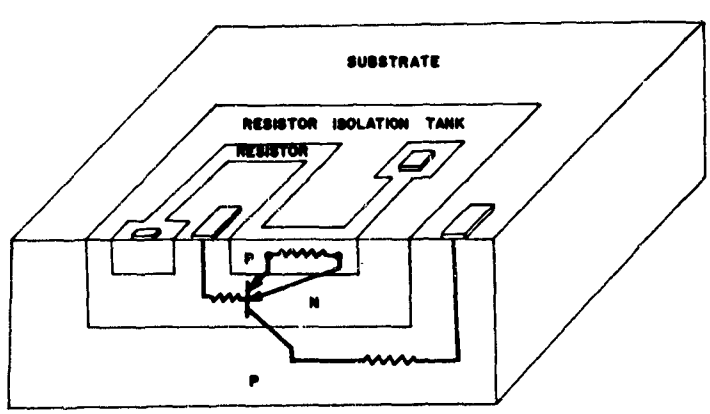


Figure 10

Two examples of devices that failed because their temperature capabilities were exceeded will now be discussed.

The SN355 was subjected to an operating step stress test. This test was performed by applying constant power biasing with increased ambient temperature at each stress level. The biasing circuit is shown in Figure 11. The components within the dotted line comprise the integrated circuit while the outside components furnish bias. Voltage nodes are also included.

SN355 DRIVER SWITCH
OPERATING STEP
STRESS SCHEMATIC

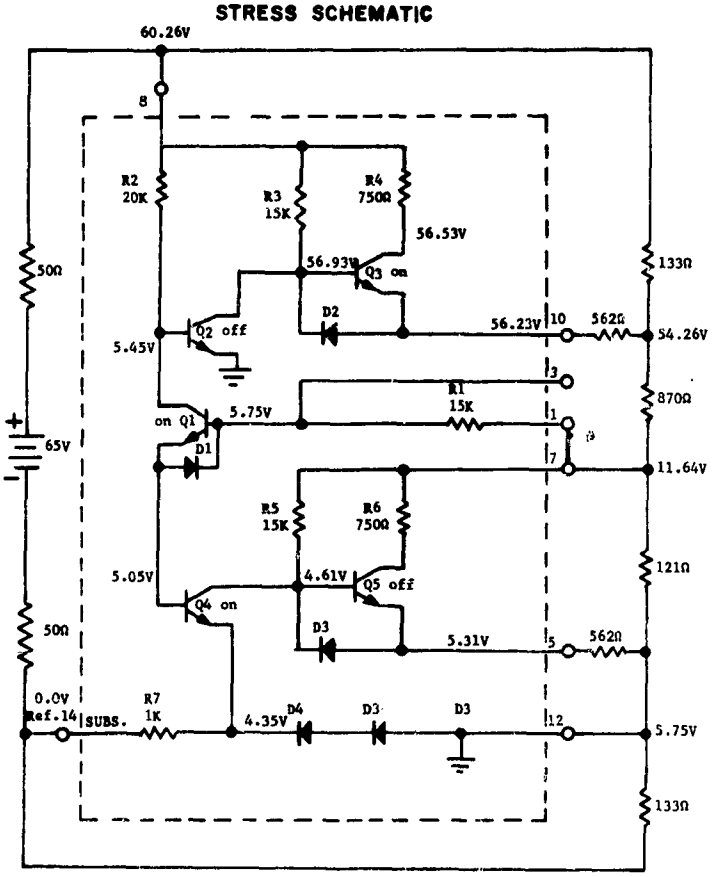


Figure 11

Figure 12 graphs cumulative percent failure as a function of bar temperature for the step stress test. Bar temperature was determined by the equation

$$T_b = T_a + (P)\quad(\theta j),$$

where T_b = Bar temperature in °C
T_a = Ambient temperature in °C
P = Power dissipation in mw
θj = Thermal conductivity in °C/mw.

The graph shows a sharp increase in the rate of failure after the bar temperature exceeded 200°C.

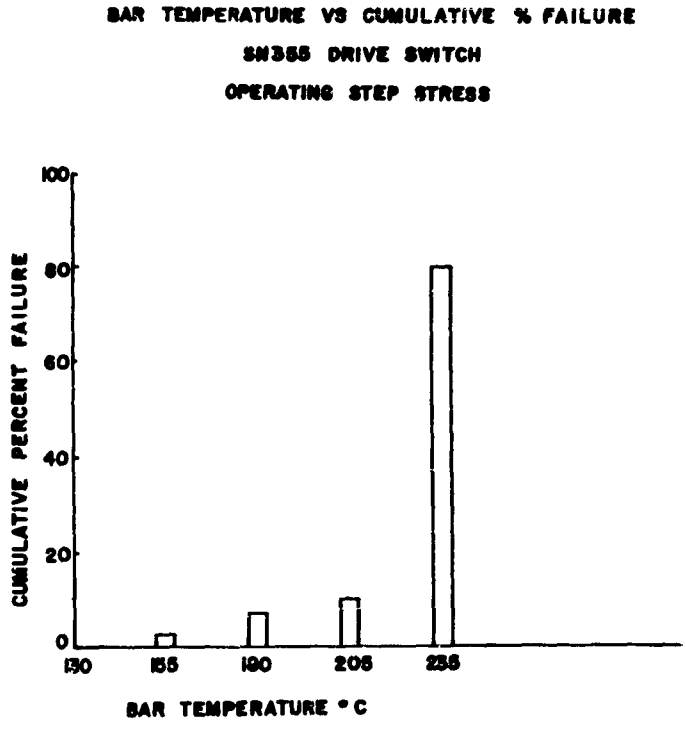


Figure 12

Examination of the failures revealed melted evaporated leads in all cases on the pin 8 side of the R2, R3, and R4 resistor. Figure 13 is a photomicrograph of the bar with arrows to indicate the location of the open leads. The

photomicrograph also reveals the resistor isolation tanks are not biased on this bar design.

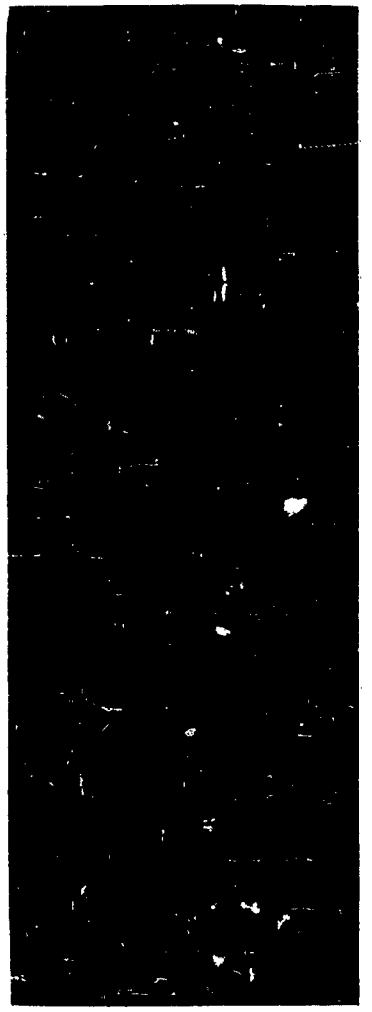


Figure 13

PHOTOMICROGRAPH OF SN355 BAR

A study of the operating circuit indicates that sufficient current to melt the leads could not have passed through the surface resistors, but instead passed through the resistor window and down to the substrate. This theory was verified by opening the evaporated leads across the resistors from pin 8 and observing that when the bar temperature was increased above 200°C, the evaporated leads melted on the pin 8 side of the resistors. Material resistivities and temperatures attained correlate well with the thermal generation theory discussed earlier.

It was concluded that this device failed due to a thermally generated transistor action when subjected to bar temperatures above 200°C and that meaningful step stress information cannot be obtained above 200°C.

The SN341 was subjected to an operating step stress test which was performed by applying constant power biasing with increased ambient temperatures at each stress level. The biasing circuit is shown in Figure 14. Components outside the dotted line supply bias for the integrated circuit within the dotted line. Voltage nodes are included to show the operational conditions.

Figure 15 graphs the cumulative percent failure versus the bar temperature for the step stress test. Bar temperature was determined by the equation used in the previous example. The graph shows a marked increase in rate of failure as the bar temperature surpasses 300°C.

An examination of the failed devices revealed melted leads at several places on the bar (see Figure 16). These melted leads are associated with the 5.9K resistor, the 1.0K resistor, the output transistor, and the substrate. Since the melted leads are associated with only the above mentioned components, a study of the diffusion profile of these components is desirable. Figure 17 is the diffusion profile of the 5.9K resistor, the 1.0K resistor and the output transistor in their biased condition. Figure 18 has included in the profile, a parasitic equivalent. Observe that thermally generated transistor actions at the two resistor diffusions has created a current path through the substrate material to its bias tap. This current path has caused a voltage gradient to exist across the substrate material. The voltage gradient has raised the substrate potential beneath the collector of the output transistor to a p-n diode above the "on" V_{CE} of the collector which allows substrate current to flow into the collector of the "on" transistor. The failure indicators are melted evaporated leads along the current paths.

It is concluded that meaningful step stress information cannot be obtained if the device is operated at bar temperatures above 300°C.

SN341 NAND GATE
OPERATING STEP
STRESS SCHEMATIC

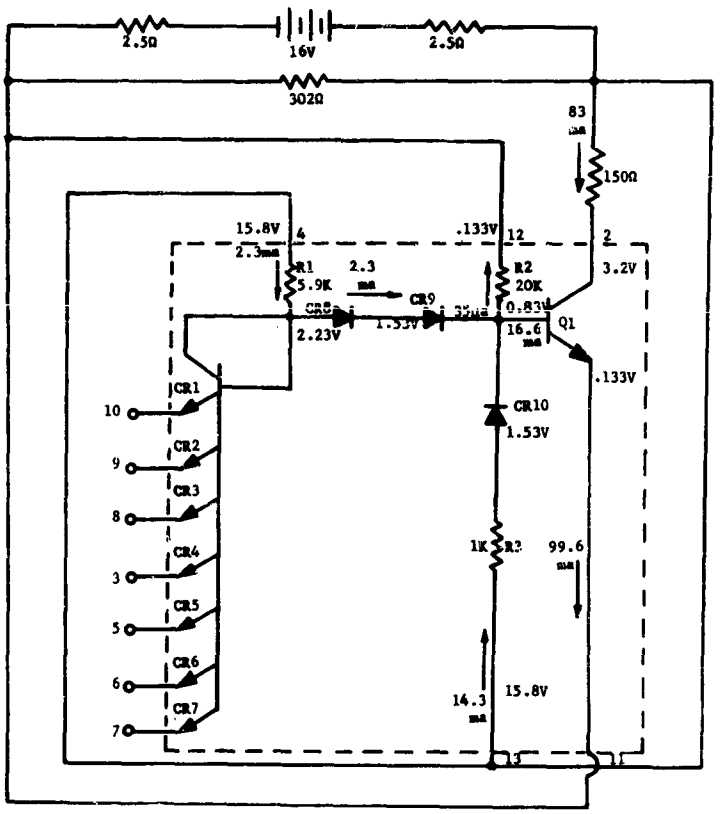


Figure 14

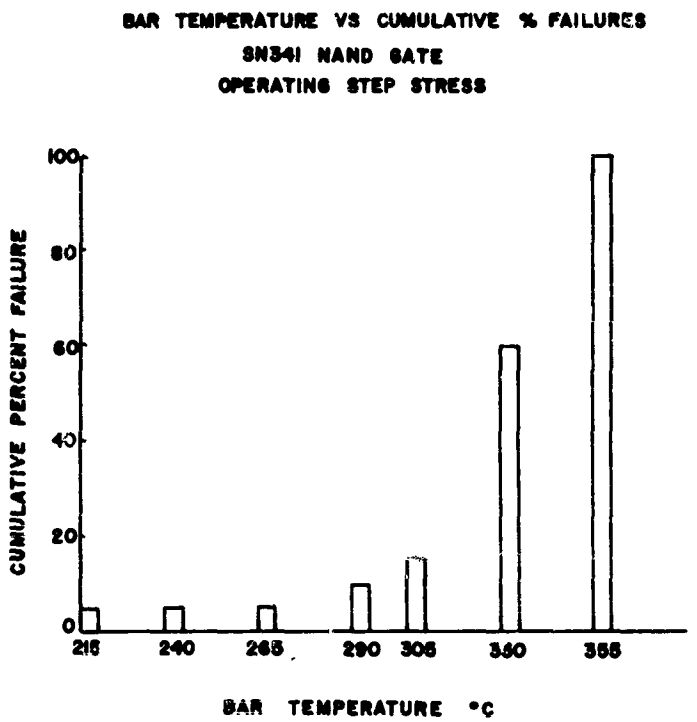

BAR TEMPERATURE VS CUMULATIVE % FAILURES
SN341 NAND GATE
OPERATING STEP STRESS
CUMULATIVE PERCENT FAILURE
100
80
60
40
20
0
215
240
265
290
305
350
355
BAR TEMPERATURE °C

Figure 15

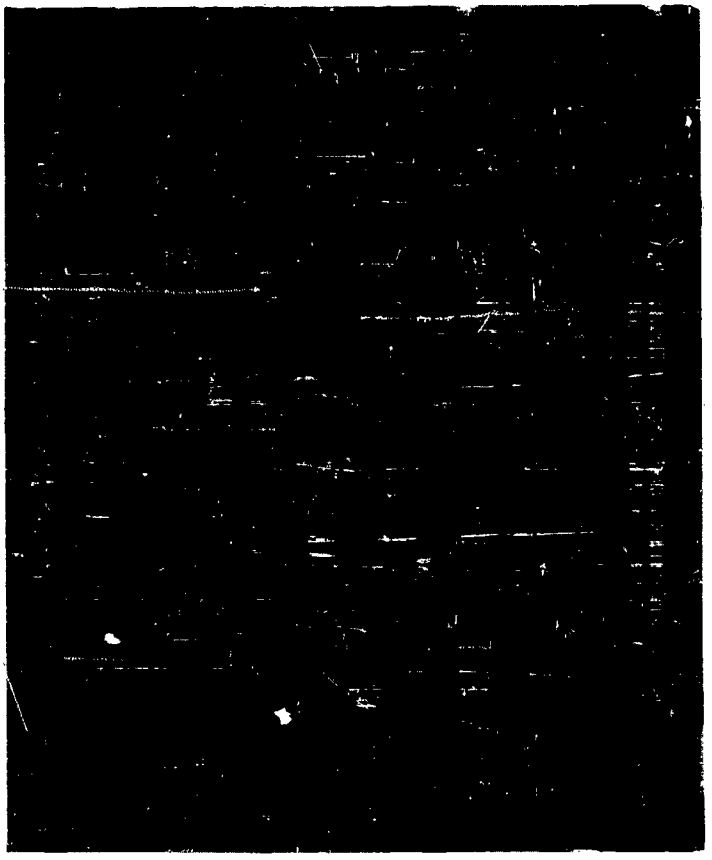


Figure 16

PHOTOMICROGRAPH OF SN341 BAR

SN341 NAND GATE
EQUIVALENT PARASITIC COMPONENT LAYOUT

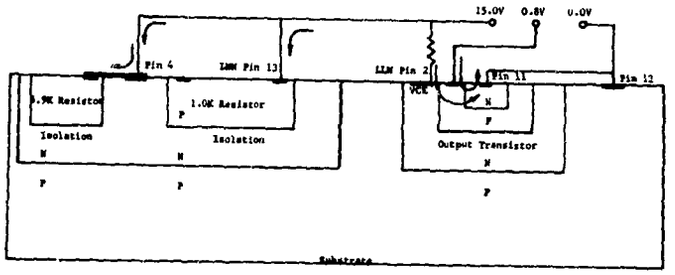


Figure 17

SN341 NAND GATE
EQUIVALENT PARASITIC COMPONENT LAYOUT
AND ILLUSTRATED PARASITICS

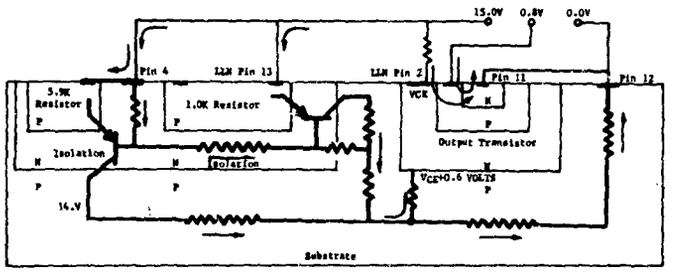


Figure 18

CONCLUSION

Reliability information for integrated circuits is extremely difficult to obtain. Process improvements and major breakthroughs in the "state of the art" have pushed the reliability of integrated circuits extremely high. Reliability engineers must depend on accelerated life testing to obtain the required reliability information. There are limits to accelerated life testing which, if exceeded, preclude the possibility of obtaining useful data. Simple laboratory tests, such as those described above, on a few devices can reveal the onset of these limitations and allow the maximum information to be obtained.

REFERENCES

1. Runyan, W. R., "Silicon Semiconductor Technology", McGraw-Hill, Pp. 167 and 168.

2. Adams, J. and Workman, W., "Semiconductor Network Reliability Assessment", Proc. IEEE, Vol. 52, No. 12, December, 1964, Pp. 1624-1635.

THERMOPHYSICS OF SILICON POWER TRANSISTORS*

David A. Peterman

Central Research Laboratories
Texas Instruments, Inc.
Dallas, Texas

INTRODUCTION

This paper describes a study of the fundamental thermophysics of silicon epitaxial power transistors with emphasis on obtaining quantitative agreement between analytical calculations of junction temperature variations and experimental temperature profiles obtained with an infrared microradiometer. These temperature distributions are a prime manifestation of thermal-electrical interactions which affect the performance and reliability of the device. Obtaining quantitative agreement requires an accurate model representing the heat conduction problem, a realistic energy dissipation distribution for correct boundary conditions, and accurate experimental temperature profile data.

A commercial NPN silicon epitaxial power transistor was chosen for detailed study in this analysis for several reasons: (1) Its common comb-type structure shown in Figure 1 is typical of geometries used for most high-frequency power transistors. The wafer is 100 x 100 mils. The dotted line indicates the emitter diffusion outline and the cross-hatched areas are emitter and base contact metalization. (2) The 9-mil emitter finger width allows detailed temperature variations across the emitters to be observed by the infrared microradiometer, which has a resolution of 1.5 mil.

* This work is being supported by Rome Air Development Center, U.S. Air Force under Contract AF 30 (602) 3727.

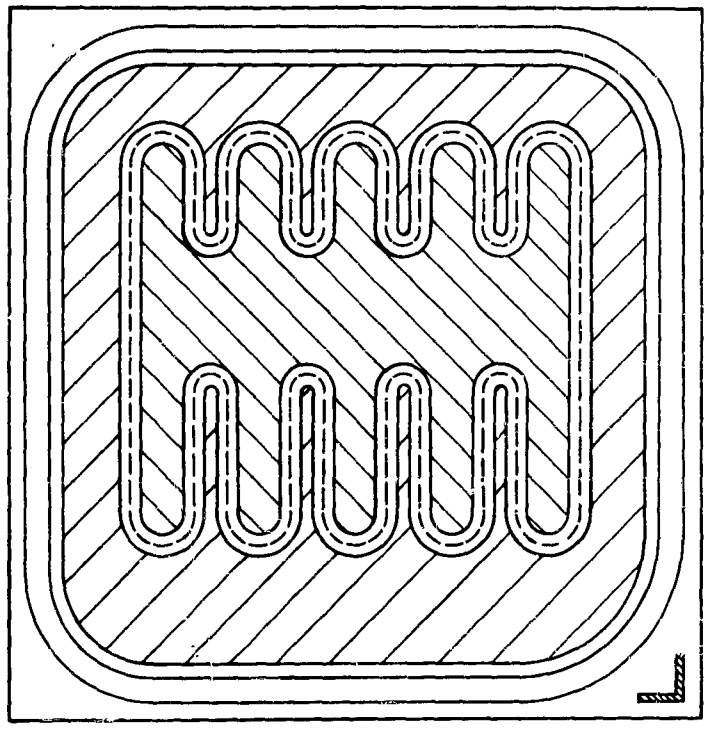


Figure 1

Surface Geometry of Test Transistor

(3) Analytical results and understanding resulting from this study should be extendable to devices designed for use at either higher frequency (smaller geometry) or higher power (larger geometry). (4) Volume production insures applicability of the results and availability of detailed information on device characteristics and design.

INFRARED PROFILE EXPERIMENTS

The infrared microradiometer used in this study consists basically of an InSb photodetector that is sensitive to

infrared radiation in the 2 to 5.6 μ range. Radiation is collected from a 1.5 mil spot and focused on the detector by a 15X Beck reflective objective lens. Amplification is achieved by chopping the radiation below the objective and using a phase-sensitive ac amplifier. A reference chamber is provided into which the detector looks when the mirrored chopper blade is interrupting the source radiation. The detector and reference chamber are maintained at 77°K by liquid N_2 dewars. This instrument has a sensitivity of 0.5°K and an absolute accuracy of about 1.0°K at 300°K for a black body. Subjects are mounted on an X-Y stage which has an automatic X drive with a manual Y adjustment. A position signal is obtained from a multi-turn potentiometer connected to the drive mechanism. Infrared profiles are obtained by recording the amplifier output (infrared signal) vs. the position signal on an X-Y recorder.

In order to convert the output signal of the instrument to actual temperature, the emissivity of the device surface must be known at every point. This poses a serious problem when the surface temperature of a semiconductor device is desired. Generally, the emissivity varies on the surface from less than .05 on the metal contact areas to an apparent value of .6 to .8 on the silicon. The silicon actually is not opaque to radiation in the 2 to 5.6μ region so that the radiation signal originates from within the device. This situation not only distorts the thermal picture but introduces unknown variations which preclude quantitative measurement of surface temperature. It is essential, therefore, to use some method of controlling surface emissivity.

Several coatings were studied to determine their capability to provide uniform high-emissivity surfaces. Figure 2 shows reflectance measurements on three sample coatings. This data was obtained by applying the coating to a silver mirror surface and measuring the reflection spectrum with a Perkin-Elmer Model 337 Infrared Spectrophotometer. Since emissivity in this case is one minus reflectance, a 1 percent reflectance corresponds to an emissivity of .99. Although 3M Black gives a higher emissivity, it requires a 4-mil thickness to obtain uniform coverage. Krylon, on the other hand, also gives better than .99 emissivity and only requires a 0.2 mil thickness. Although both have very high resistivities, 3M Black is at least an order of magnitude better than Krylon. Parsons has several strong absorption lines in the 4 to 6μ region causing it to have the lowest emissivity; therefore, in device testing, 3M Black has been used to obtain contour maps and Krylon has been used to obtain detailed temperature profiles.

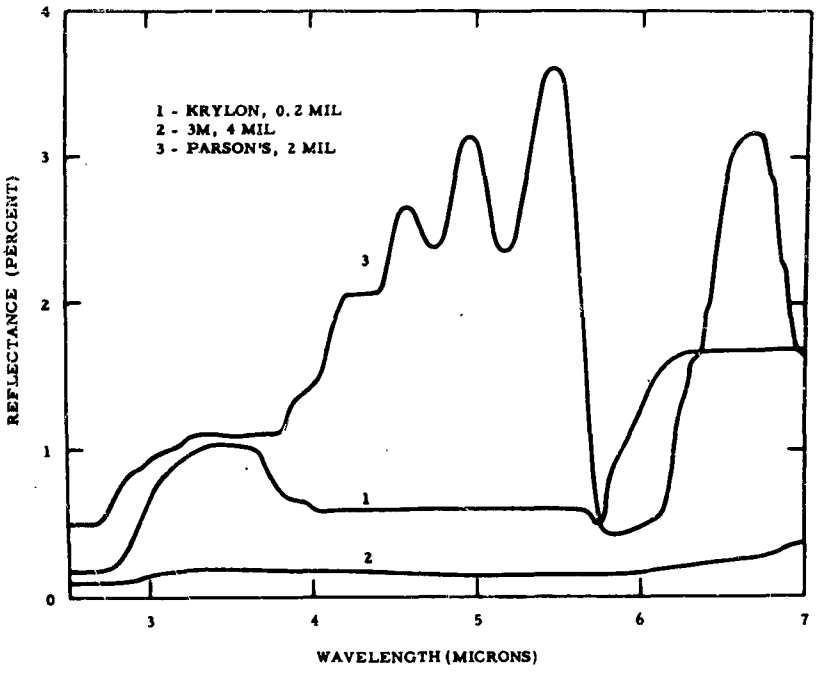


Figure 2

Infrared Reflectance of Emissivity Control Coatings

Transistors were tested at steady state conditions using a common-emitter circuit with conventional power supplies. Thermocouples were mounted on the header adjacent to the chip to record the true stud temperature during operation. Continuous scans were made at 5-mil intervals across the chip. Isothermal contour maps are obtained from this data by plotting the location of incremental temperatures on a map of the transistor surface. Figures 3 and 4 show typical isothermal contour maps obtained with 3M Black coating for two test conditions. Figure 3 was obtained with I_C = 5 amps and V_{CE} = 10 volts. This map indicates nearly uniform heat dissipation across the active area of the device. The longer emitter fingers produced proportionately larger heating than the shorter ones. The maximum temperature of 125°C occurs near the center on the long emitter side. In general, this isothermal map is quite reasonable and

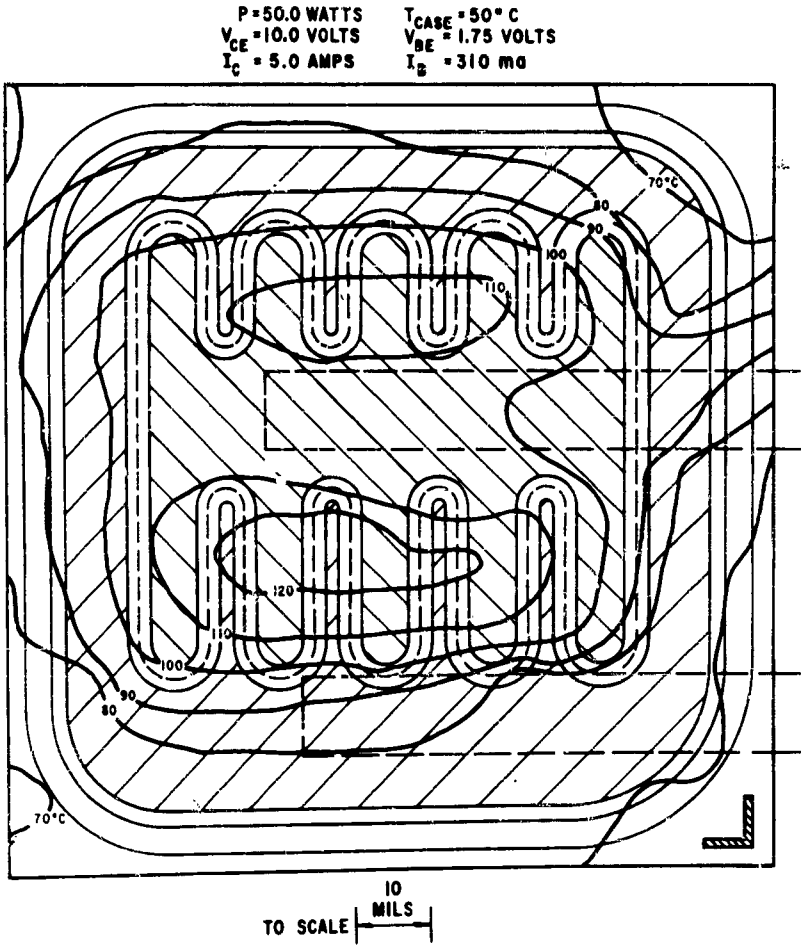


Figure 3

Isothermal Contour Map of Transistor Surface

indicates approximately uniform current distribution over the emitter fingers. Figure 4 was obtained with I_C = 2 amps and V_{CE} = 25 volts. These parameters produced a significant change in temperature distribution. The maximum

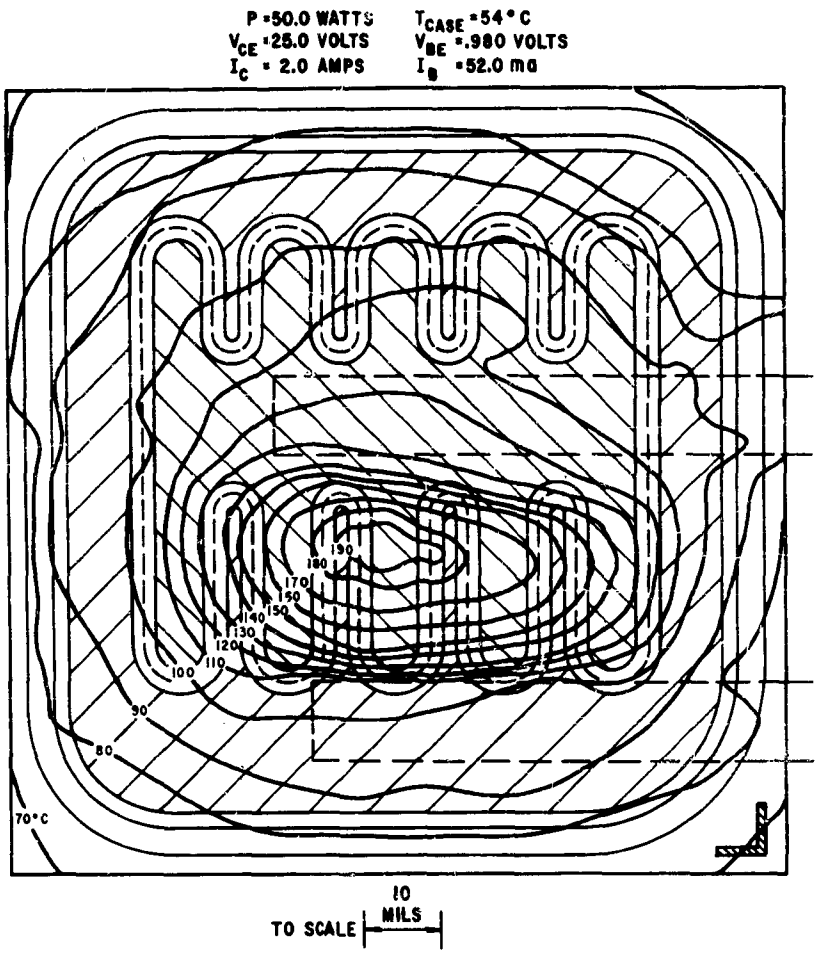


Figure 4

Isothermal Contour Map of Transistor Surface

surface temperature near the center on the long emitter side has now increased to 195° C, while the outer areas of the device remain at a lower temperature. Although hot spot formation has previously been predicted and qualitatively

observed, the contour maps shown in Figures 3 and 4 are quantitative and were obtained at the same total power dissipation. This indicates a mechanism other than power level causing current concentration, resulting in local heating and hot spots. Contour maps also were obtained with constant collector-emitter voltage (V_{CE}), constant collector current (I_C), and constant base emitter voltage (V_{BE}) to determine the relative importance of these parameters. At sufficiently high voltages ($V_{CE} \geq 20$ v), hot spots developed in each case as power level increased. The increase in current concentration with increasing I_C for the constant V_{CE} case is particularly significant in that higher I_C values did not result in more uniform distribution as predicted from Figures 3 and 4. Tests with varying case temperature showed that case temperature level was not a significant factor in the development of hot spots.

Several devices were deliberately destroyed after obtaining temperature contour data. Examination of these devices showed local melting of the aluminum contacts in regions exactly corresponding to the maximum temperature regions observed with the infrared microradiometer.

ENERGY DISSIPATION STUDIES

In order to correctly interpret the experimental results and obtain a valid model for the thermal characteristics, it is necessary to understand the mechanisms and location of energy dissipation in transistors. Energy is dissipated in a transistor in several ways. Potential energy is lost by the electrons as they come to thermal equilibrium in the depletion region near the collector-base junction. This energy is absorbed by the lattice in the thermalization process. A similar change of energy distribution occurs in the emitter. In this case, however, the lattice must give up energy to thermalize the electrons in the emitter because of the loss of high energy electrons at the emitter base junction due to injection into the base region. Energy is also dissipated in the base region due to sheet resistance. The actual energy exchange between electrons and lattice will therefore be a complex function of the operating current and voltage values and will be distributed through the active region of the transistor. When the base current (I_B) is much smaller than the collector current (I_C) (i.e., $I_C \approx I_E$), the net energy dissipation to the lattice can be approximated by $I_C V_{CE}$. Assuming that this is the case and that the active region is very near the surface of the transistor, the problem reduces to one of obtaining the

local variation of collector (or emitter) current density across the surface of the device.

For this type of structure, a significant portion of the base current flows parallel to the emitter-base junction. This parallel component of base current produces an IR drop in voltage which debiases the center of the emitter with respect to the periphery, causing a large portion of the emitter current to be carried at the periphery. Power transistors, therefore, are usually characterized by a series of long slender emitters (usually called fingers). In this case, a two-dimensional model can be used to characterize heat conduction in such devices.

Wilcox[1] has solved the heat conduction equations for the case of an infinite number of emitters with uniform energy dissipation across the emitter surface and a constant substrate temperature (T_o). He also gives an approximate solution for the surface temperature variation across a single emitter. This solution is an infinite series and is given by:

$$T_s = T_o + \frac{q}{k}\left[\frac{1}{6} + \frac{3}{\pi^2\left(\frac{w}{b}\right)} \cdot \sum_{n=1}^{\infty} \frac{1}{N^2} \sin\frac{1}{3}n\pi\left(\frac{w}{b}\right) \cos\frac{1}{3}n\pi\left(\frac{x}{b}\right) \tanh\frac{n\pi}{3}\right] \quad (1)$$

where T_s = surface temperature,
T_o = constant substrate temperature,
R = thermal conductivity of semiconductor material,
q = energy dissipation rate per unit length of emitter,
w = half-width of the emitter,
b = thickness of semiconductor chip,
x = distance normal to emitter finger.

Since the equations and boundary conditions of this problem are linear, a solution for a finite number of emitters can be obtained by superposition of a series of the solution in equation (1). The results of these calculations are compared in Figure 5 with an experimental surface temperature

[1] Wilcox, W.R., "Heat Transfer in Power Transistors" IEEE Transactions on Electron Devices, Vol. ED-10, Number 5, p. 308-313, September 1963.

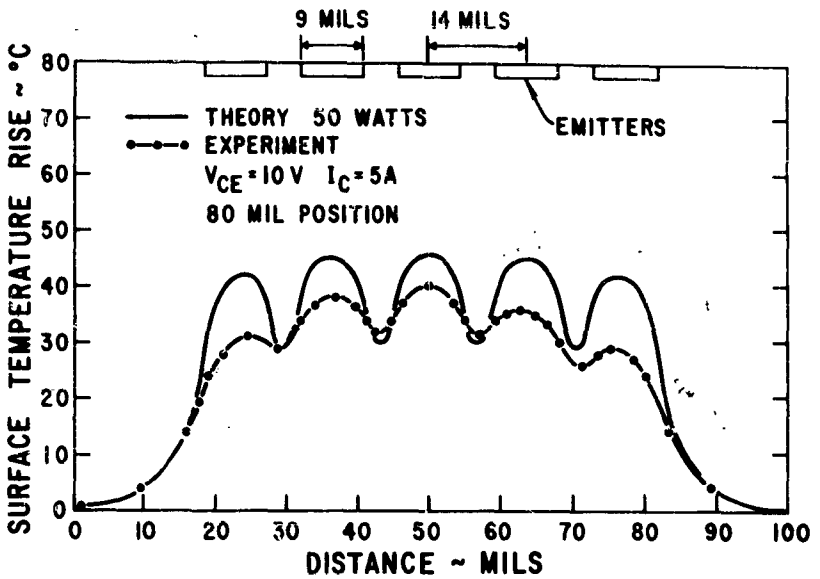


Figure 5

Temperature Profile Across Emitter Fingers

profile obtained with the infrared microradiometer. The profile was obtained at the end of the emitter fingers, most nearly representing the condition of uniform energy dissipation across each emitter.

Although the agreement is good, several differences between theory and experiment point to improvements that can be made to improve the analytical model other than including variation of energy dissipation. The experimental curve indicates a more rapid decrease in peak height than predicted by the calculations. This indicates that the isothermal boundary condition on the header side of the chip should be modified to account for the actual temperature variation in the header. The difference in smoothness probably results from the fact that, in the actual device, energy is not dissipated on the surface as assumed in the analytical model but near the collector-base junction. The coating of paint may also need to be accounted for in the analysis. Some smoothing could also be due to 3-dimensional

spreading effects at the end of the emitters. The analytical solution, however, does yield a valid representation of the temperature variation in the junction region of the transistor.

THERMAL IMPEDANCE

It has been common practice to indicate the heat dissipating performance of semiconductor devices by defining thermal impedance (θ_{JC}) as the temperature difference between the heat dissipating junction (T_J) and the case temperature (T_C) divided by the power being dissipated (P). There are several difficulties in using this parameter. Since device area and thickness are not used, θ_{JC} can only indicate relative performance of devices with identical geometries. Definition of a unique, meaningful case temperature is often difficult. The most significant problem, however, deals with defining what is meant by junction temperature. Junction temperature is not a unique quantity, but depends on location and operating condition as demonstrated in Figures 3 and 4 (assuming that the surface temperature variation correctly indicates the temperature variation in the junction region which is just below the surface).

It is easy to see, therefore, why junction temperature measurements based on temperature dependent device parameters give θ_{JC} data that is difficult to interpret unequivocally. Often, the measurements are made at conditions not at all representative of operating stress conditions thereby yielding no correlation of failure rate with stress level. Infrared microradiometry and thermal analysis can be very valuable in defining test procedures which yield θ_{JC} values that can be used with confidence for reliability screening. The major requirement for a definition of junction temperature is that a unique physical quantity is obtained which truly indicates the state of thermal stress on the device. Figure 6 shows experimental data obtained with the microradiometer on the test transistor. With the case temperature held at 100°C, the collector current was set at various current levels. At each level, the collector-emitter voltage was increased until the point of maximum surface temperature was at 200°C. The line drawn through these points represents the operating conditions for the maximum value of T_J = 200°C. A constant power line (30 watts) and the maximum dc operating limit (Joint Electron Device Engineering Council Registration Data) for this particular device are also shown for comparison. The JEDEC limit line was established prior to these experiments by displacing to the left a line representing the experimentally determined

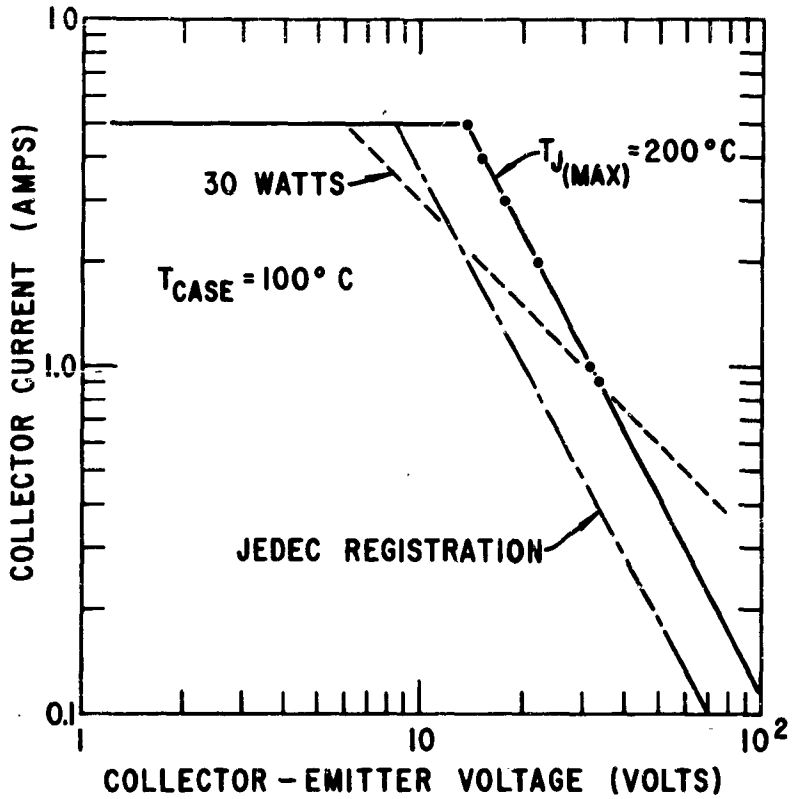


Figure 6

Variation of Maximum Surface Temperature

failure zone. It therefore represents an average safe operating limit for this device type. Since the constant maximum T_J line is parallel and to the right of the JEDEC line, it appears that the maximum value of T_J is a direct indication of thermal stress level. This quantity is not only easily defined physically, but can logically be expected to indicate critical thermal stress.

Having identified the maximum value of T_J as the critical junction temperature, a corresponding thermal impedance can be defined as

$$\varphi = \frac{T_{J_{max}} - T_C}{P} \qquad (2)$$

where P is the power dissipation and T_C is the case temperature. Both P and T_C must be appropriately defined for a particular device. Electrical θ_{JC} measurement techniques can be directly evaluated by their ability to indicate the critical thermal impedance (φ) and its variation with operating condition.

CONCLUSION

This paper has described an analysis of the thermal characteristics of silicon power transistors aimed at understanding temperature-dependent failure mechanisms. The infrared microradiometer has been shown to be a powerful analytical tool which can be used to validate and guide theoretical efforts. A simple analytical solution for the temperature profile of a transistor surface has yielded good agreement with the experimental profiles. These results demonstrate how a difference in surface temperature leading to a thermal instability can occur even in an ideal device with uniform energy dissipation. Experimental isothermal contour maps clearly demonstrate the accelerating effect of V_{CE} on the formation of a hot spot in the test transistor. A realistic thermal impedance (φ) has been defined based on the maximum temperature in the junction region which can be used to evaluate indirect measurements of thermal impedance (θ_{JC}).

SECTION IV

SURFACE EFFECTS

ACCUMULATION AND DECAY OF MOBILE SURFACE CHARGES ON INSULATING LAYERS AND RELATIONSHIP TO RELIABILITY OF SILICON DEVICES*

W. Schroen**

International Telephone and Telegraph Corporation
ITT Semiconductors Product Laboratories
Palo Alto, California

ABSTRACT

Reverse biased diodes of various geometrical dimensions, oxide preparation techniques and breakdown voltages were used to measure surface contact potential differences and surface ion distributions in various ambients. The experimental data of ion accumulation were compared to complementary error function solutions expected for a distributed capacitance and sheet resistance. After bias removal the accumulated positive and negative charges were allowed to decay. Significant differences of decay times could be observed determined by the sign of the charge and the chemical preparation of the sample. The influence of mobile surface charges on device characteristics, particularly on the diode breakdown voltage, is described. The relationship between the surface breakdown voltage and the external bias applied to a metal strip located geometrically above the p-n junction is established. Finally, it is shown that a "freezing" of the surface charges can change and stabilize the diode V-I characteristics considerably.

I. INTRODUCTION

The behavior of mobile surface ions on insulating layers has been investigated by a number of authors. Brown[1] proposed ion motion along the surface across a p-n junction to form an underlying inversion layer. Atalla, Bray and Lindner[2] explained the observed instabilities

* Research sponsored by the U. S. Air Force, Rome Air Development Center, Rome, New York.

** Present address: Central Research Laboratories, Texas Instruments, Inc., Dallas, Texas.

of thermally oxidized silicon junctions in wet atmospheres by the motion of ions on the surface of the oxide under the influence of an applied electric field. The ions, in order to become mobile, require the presence of water or of organic molecules of large dipole moment. Essential features of Attala's model, namely that the ions can set up fields of the order of 10^6 V/cm in the oxide layer, producing potential differences of the order of half the reverse bias applied to the p-n junction, could be verified by Shockley, Queisser, Hooper and Schroen.[3,4] They measured the large contact potential differences between the charged oxide surface and a metal reference electrode using a vibrating capacitor probe ("Kelvin probe"[4,5]). In the present paper, investigations are reported which demonstrate that surface treatments and ambient variations can drastically change the charge distribution and the associated drift.

Furthermore, it was the aim of the present investigations to study the influence of external fields - supplied either by mobile surface ions or by an external potential to a metal layer - on the breakdown voltage of oxide protected p-n junctions. It is well known that high voltage planar junctions generally exhibit breakdown voltages considerably lower than those predicted[6] from impurity density profiles (in addition, the onset of breakdown at corners plays an important role[7]). It is expected that a better understanding of the influence of external fields on the breakdown voltage will contribute to mitigate and finally eliminate device reliability problems due to charge motion and surface breakdown influences.

Historically, as early as in 1948, Shockley and Pearson[8] discussed the concept of using an external electric field normal to the surface of a semiconductor to control the carrier density near the surface. In 1956, Garrett and Brattain[9] established a theoretical model for the influence of surface fields on avalanche breakdown. Later, Kahng and Atalla[10] pointed out that a structure, consisting of a reverse biased p-n junction superseded by an insulating layer combined with a metal control, can be used to enhance as well as deplete charge near the surface of the semiconductor. Using the insulated-gate field effect transistor, Hofstein and Heiman[11] reported the control of avalanche breakdown at the surface of a p-n junction. Nathanson[12] observed a decrease of the breakdown voltage with increasing gate field of the p^+n high field triode. The variation of the breakdown voltage of an n^+-p control ring diode with various gate potentials was investigated in detail by Schroen, Hooper and Queisser.[13,14] Part of their results were applied by Shockley and Hooper[15,16] in the surface-controlled avalanche transistor. Using a p^+-n control ring diode, Castrucci and Logan[17] found an increase in breakdown voltage when the control ring was biased negatively, producing a field that opposed the field of the immobile charge. Their measurements showed an optimum value of the control-ring voltage, at which breakdown voltage was a maximum. The effect of increasing the breakdown voltage by a metallic contact that is connected to the

diffused region and extends over the oxide-protected junction edge has been reported by Clark and Mack,[18] Castrucci and Logan,[17] and Schroen.[19]

The following Sections 2 and 3 describe at first briefly the diode structures and the experimental equipment used, then the results of surface ion accumulation and decay. The complementary error function distribution is found to be a good representation of the measured values. Section 4 deals with the influence of surface ions and external potential on device reliability. The diode breakdown voltage is investigated as a function of time and of the bias on a control electrode. Finally, some methods are briefly described for immobilizing the surface ions.

2. EXPERIMENTAL ARRANGEMENT

The specimens used for the investigations consisted of diffused n^+p or p^+n planar diodes with various protecting oxides. The preparation of the diodes was performed in the following sequence. (A) For p-type starting material: Resistivity 1 Ω-cm, oxidation in oxygen plus steam at 1200° C for 30 minutes, resulting in an oxide thickness ≈ 0.6 μ window opening, P_2O_5 phosphorus predeposition at 1050° C for 30 min. (B) For n-type starting material: Resistivity 5 Ω-cm, oxidation and window opening as above, B_2O_6 boron predeposition at 950° C for 30 min, followed by a diffusion at 1300° C for 45 minutes in wet oxygen.

For some experiments, the oxide was removed and new oxides were deposited: (1) Method: Pyrolytic decomposition of ethyl silicate at 850° C for 1 hour according to

$$SI(OE_T)_4 \longrightarrow SiO_2 + 2[(E_T)_2O] \qquad (1)$$

where $E_T = CH_3CH_2$, resulting in an oxide thickness ≈ 1 μ. Decomposition at 500° C resulted in a thickness of ≈0.2 μ. Some samples received an additional phosphorus glass layer using $POCl_3$ at 1050° C for 20 min. (2) Method: Oxidation in oxygen plus steam at 1200° C for 30 min, resulting in an oxide thickness ≈ 0.6 μ. (3) Method: Oxidation in wet oxygen at 1200° C for 45 min, resulting in an oxide thickness ≈ 0.6 μ. (4) Method: Oxidation in dry oxygen at 1300° C for 1 hour, resulting in an oxide thickness ≈0.3 μ.

After oxide windows were opened for contacts, the samples were rinsed in trichloroethylene and then heat treated at 600° C for 30 min. in vacuum (pressure 10^{-6} Torr, residual gas consisted almost exclusive of pure N_2). Samples prepared by this procedure are referred to as "untreated" because they showed surface ion motion on a very reduced scale.

For the investigations of the ambient influence, the samples and the measuring equipment were enclosed in a metal can so that the ambient type and relative humidity could be well controlled. The moist ambient was obtained by passing dry nitrogen through a flask of heated

water, the relative humidity being determined by the nitrogen flow rate and water temperature. The humidity was monitored by a relative humidity indicator located in the ambient flow line.

Bias was applied to the diodes as shown schematically in Fig. 1 for an n+p diode. The electric field fringing through the oxide layer disturbs the originally randomly distributed surface ions so that positive charges accumulate over the p-substrate and negative charges over the n-layer. The resulting ion distribution is "frozen" by flowing dry N_2 over the sample and measured using the vibrating capacitor technique.[4, 5] The "Kelvin probe" is positioned close to the diode surface (see Fig. 2) and scanned across the junction under various conditions of diode bias and ambient. The probe used consisted of a fine wire tip (diameter 25 or 75 μ) which had been gold plated to minimize changes in its work function. The measured contact potential difference V_K between the probe and the specimen surface is that of the bucking voltage required to produce a null condition on the oscilloscope.[4, 5]

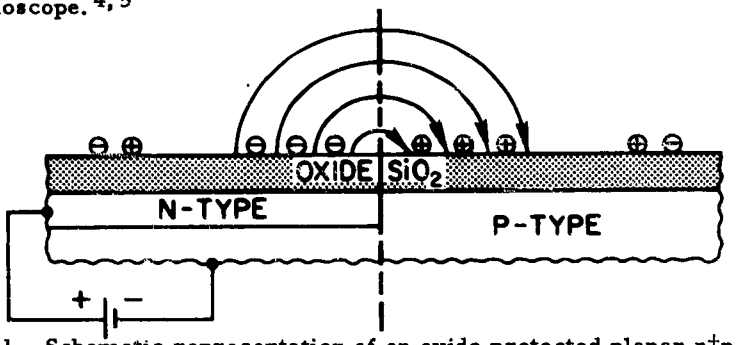


Fig. 1. Schematic representation of an oxide protected planar n+p diode. The electric field due to the reverse bias is fringing through the oxide layer and causes separation of mobile surface ions.

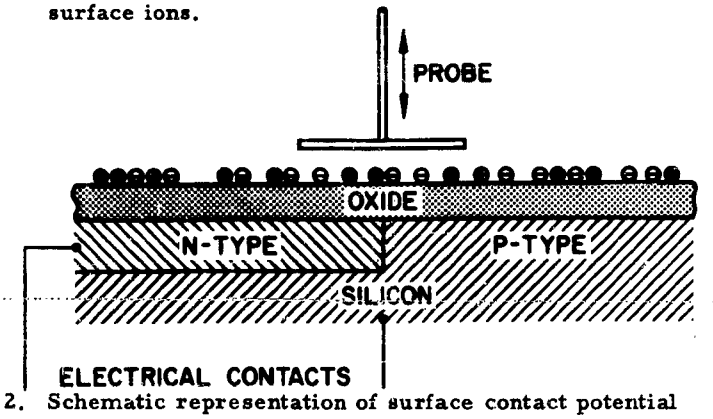


Fig. 2. Schematic representation of surface contact potential difference measurement. Ions located on outer surface of oxide covering p-n junction.

3. ION ACCUMULATION AND DECAY

3.1 Complementary Error Function Distribution

A p^+n diode with ethyl silicate oxide was reverse biased with a 22.5 V battery (terminal voltage 24 V). The measurement was carried out first in dry N_2 ambient, and then repeated in 50% relative humidity. The results are shown in Fig. 3. The curve marked "dry nitrogen" shows the contact potential distribution prior to surface charge motion. The ambient was then changed to 50% relative humidity and, with bias applied, the surface was allowed to "drift" for 12 minutes, after which time the diode was again flushed with dry nitrogen. The diode was then scanned with the Kelvin probe to measure the distribution of the accumulated ions, and the curve of Fig. 3 marked "measured, 50% RH" was obtained. It will be noticed that the change in potential due to the motion of surface charge is not equal on both sides of the junction. Over the n-material, the drift did not go past 1500 μ, whereas on the p^+ side, an appreciable disturbance is evident past 2500 μ. The reason for this behavior is the fact that edge effects of the geometry become important over the diffused area when the disturbance occurs over a distance which is an appreciable fraction of the diode diameter.

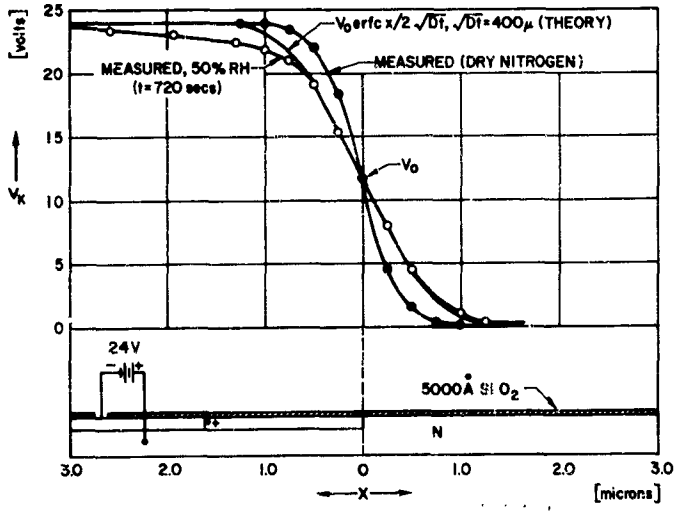


Fig. 3. Surface potential vs. distance from the junction of a p^+n diode with ethyl silicon oxide.

It is quite informative to compare the potential distribution obtained in Fig. 3 to the complementary error function solution expected for a distributed capacitance, C_a, and sheet resistance R_{sq},

which give the equation

$$\partial^2 V_K/\partial x^2 = C_a R_{sq} \partial V_K/\partial t \qquad (2)$$

or

$$C_a \dot{V}_K = (1/R_{sq}) \partial^2 V_K/\partial x^2, \qquad (2a)$$

The complementary error function solutions are available in standard diffusion tables, and can be readily compared to the measured potential distribution. Figure 4 shows this comparison. Plotted is the logarithm of the contact potential difference V_K vs. the probe position x, where x = 0 is chosen at the junction and V_0 is the potential directly above the junction, equal to half the applied bias.

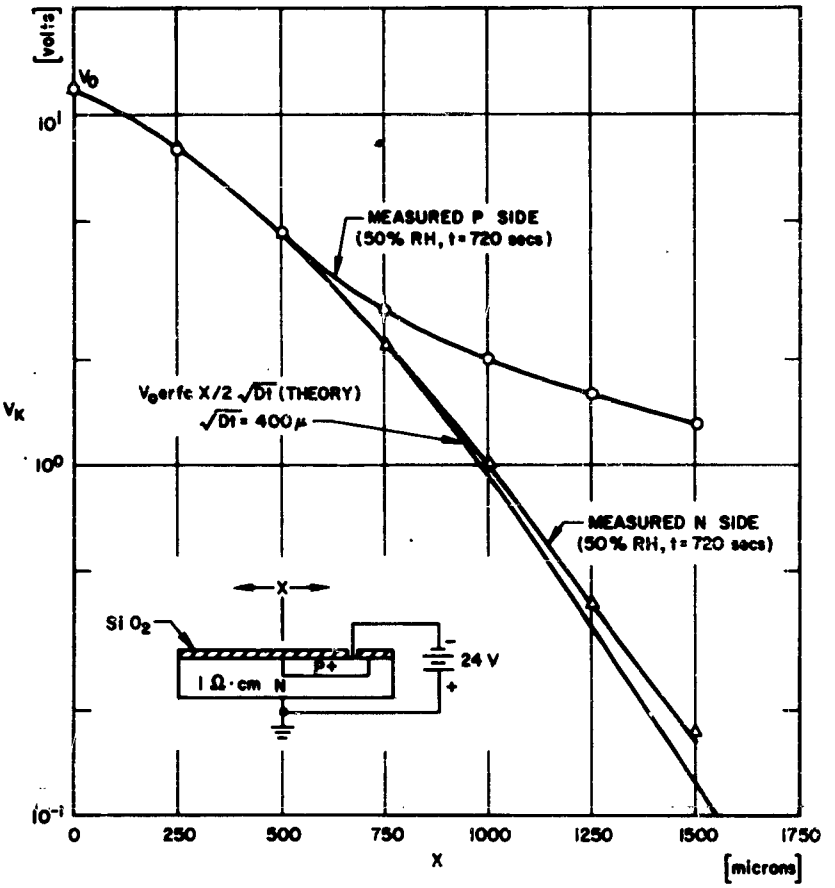


Fig. 4. Logarithm of contact potential difference vs. distance from the junction. Also plotted is a theoretical complementary error function solution for potential distribution.

Three curves are shown, the contact potential distribution measured over the n-side of the junction, the contact potential measured over the p-side, and the complementary error function which gave the best fit to the measured potential distribution. It should be mentioned that the potential measured over the p-side should actually be plotted upward from $V_0 = 12$ V at $x = 0$, since here the potential is increasing with distance. For the purpose of comparison, however, both curves are plotted together. The curve labeled $V_0 \operatorname{erfc} x/2\sqrt{Dt}$ is a complementary error function plot, obtained from diffusion tables, with $\sqrt{Dt}$ chosen equal to 400 μ for the best fit. It can be seen for x up to 500 μ that the potential distribution on both sides of the junction is a perfect fit to the error function. For $x > 500\ \mu$ the measured potential over the p-side does not fit the error function because of diode edge effects mentioned earlier. The measured potential over the n-side, however, is an excellent fit to the error function between 750 and 1500 μ. Since the data for the n-side of the diode does fit the theoretical curve so well, one can accurately determine D, the diffusion coefficient of charges on the oxide surface. For $\sqrt{Dt} = 4 \times 10^{-2}$ cm, and $t = 720$ seconds, $D = 2.2 \times 10^{-6}$ cm^2 sec^{-1}. This value of D is in excellent agreement with the value determined by Shockley, et al.[4] studying charge motion on glass surfaces.

The sheet resistance of the oxide surface can in turn be found by the relationship

$$R_{sq} = 1/(D \quad C_a) \tag{3}$$

where R_{sq} is the sheet resistance in ohms per square, D the diffusion coefficient in cm^2 sec^{-1} and C_a the capacitance per unit area in F cm^{-2}, of the oxide layer. The diffusion coefficient D was calculated above to be 2.2×10^{-6} cm^2 sec^{-1}. The capacitance per cm^2 of the SiO_2 layer was 6×10^{-9} F, so that $R_{sq} = 7.6 \times 10^{13}$ ohms per square.

3.2 Ion Accumulation and Decay as a Function of Chemical Treatment

Previous investigations[13] have shown that chemical treatments of the diode surface cause charges to compile in different quantities in the same time interval and to decay with different time constants. An acceleration of the decay was particularly pronounced if the sample was treated with water. On the other hand, benzene with its very small dissociation constant ($pK_a > 60$) and lack of electric dipole moment was found to decay very slowly. In order to get more information about ion accumulation and decay, detailed measurements were performed which are reported in the following paragraphs.

Bias of 50 V was applied to the sample for 10 minutes. It can be seen in Fig. 5 that for the untreated sample the amounts of charges accumulated over the n and p sides are almost equal. The sample was then boiled in water for about 10 minutes and baked at 150° C for 1 hour. (Without the bakeout process the decay of the surface ions

was too rapid to be measured accurately). Ion accumulation was achieved by applying bias of 50 V across the diode for only 1 minute. The ion decay is also shown in Fig. 5. There is a very fast decay of the surface ions, and, in addition, a difference of the decay rate for the ion type. Accumulated negative ions decay fast to an equal distribution while positive charges slow down their decrease.

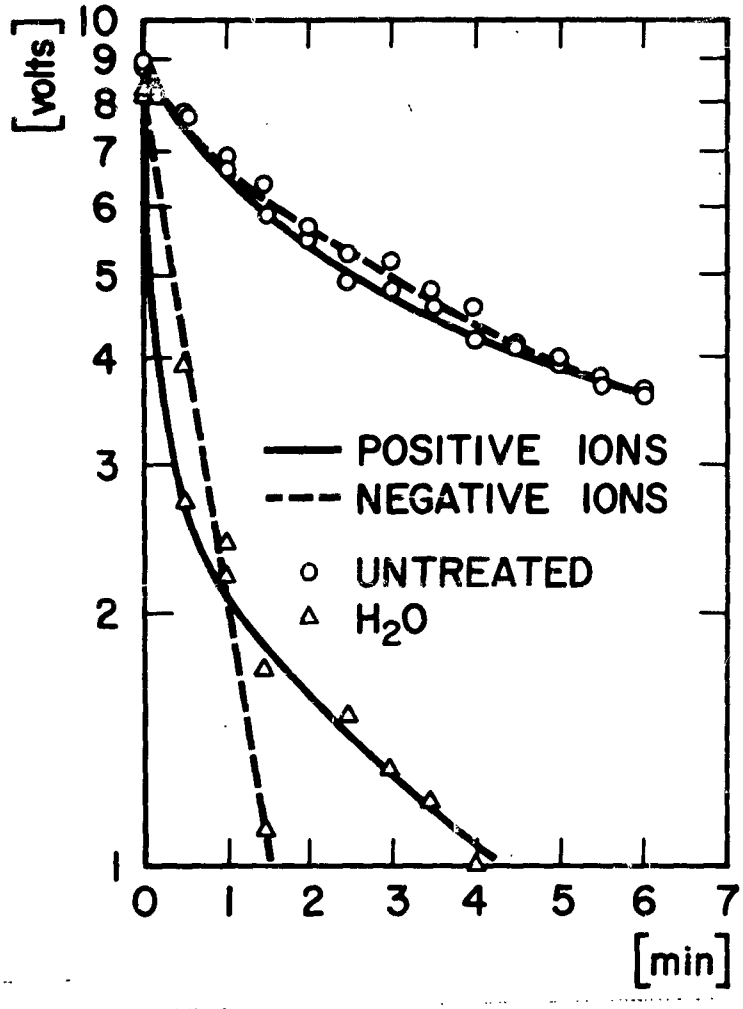


Fig. 5. Effects of surface treatment on the decay of charge accumulated on the oxide. Shown is the decay with time of the maximum negative and positive potential observed on the oxide above the n-type and p-type silicon after bias removal. Untreated and treated (H_2O) states are shown.

As another example of inorganic treatment, Fig. 6 shows the curves after a rinse in 5% H_2SO_4. The diode was submerged for 5 minutes and then dried with dry nitrogen. Reverse bias of 50 V was applied for 10 minutes. As can be seen from Fig. 6, the amount of accumulated charges is almost equal in the untreated and the treated case. There is, however, a pronounced difference in the decay times for the treated and the untreated samples, but not as drastic as in Fig. 5. The decay rates of positive and negative charges are similar.

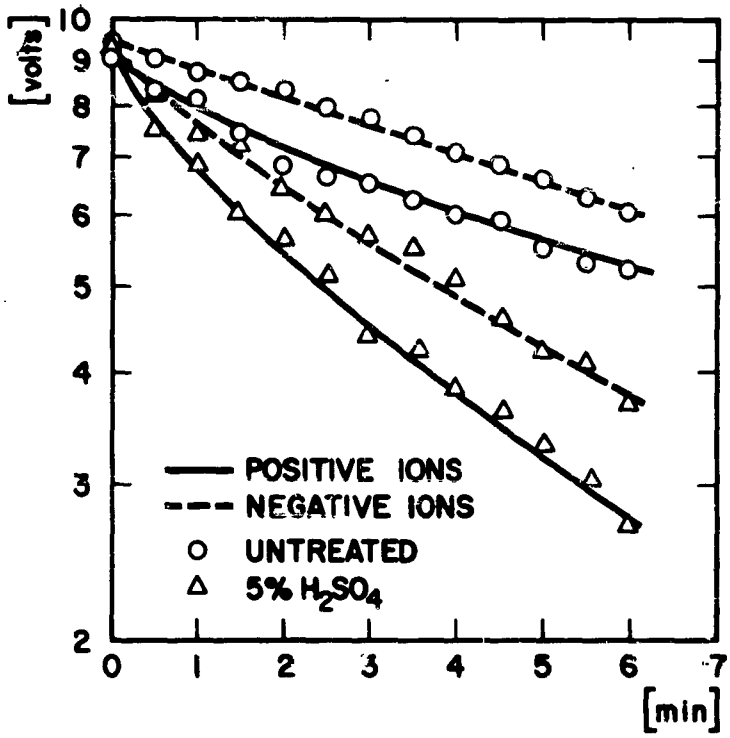


Fig. 6. Effects of surface treatment on the decay of charge accumulated on the oxide. Shown is the decay with time of the maximum negative and positive potential observed on the oxide above the n-type and p-type silicon after bias removal. Untreated and treated (5% H_2SO_4) states are shown.

In Figs. 7 and 8 two examples of an organic treatment are shown, benzene and methanol. In both cases the sample was submerged for 5 minutes in the organic solvent and dried in dry nitrogen. A bias of 50 V was applied for 10 minutes in laboratory atmosphere. As can be seen, the accumulation of charges is almost equal in all

cases. In comparison to Figs. 5 and 6, the decay of the ions is slower (particularly for the C_6H_6-treated sample), the decay rates being very similar to those of the untreated samples. The results of Figs. 7 and 8 are in agreement with the experimental findings of the previous report.[13] Benzene has a slow surface ion motion because of small dissociation constant and the lack of electric dipole moment. Methanol, on the other hand, has an appreciably higher dissociation of $pK_a = 16$ and thus can provide protons and negative charged CH_3O^- methoxyl ions. There is also an electric dipole moment for methanol of 1.7 Debye. Methanol films, therefore, increase the surface conductance and the ions decay with a faster decay rate. Similar results were found for nitromethane.

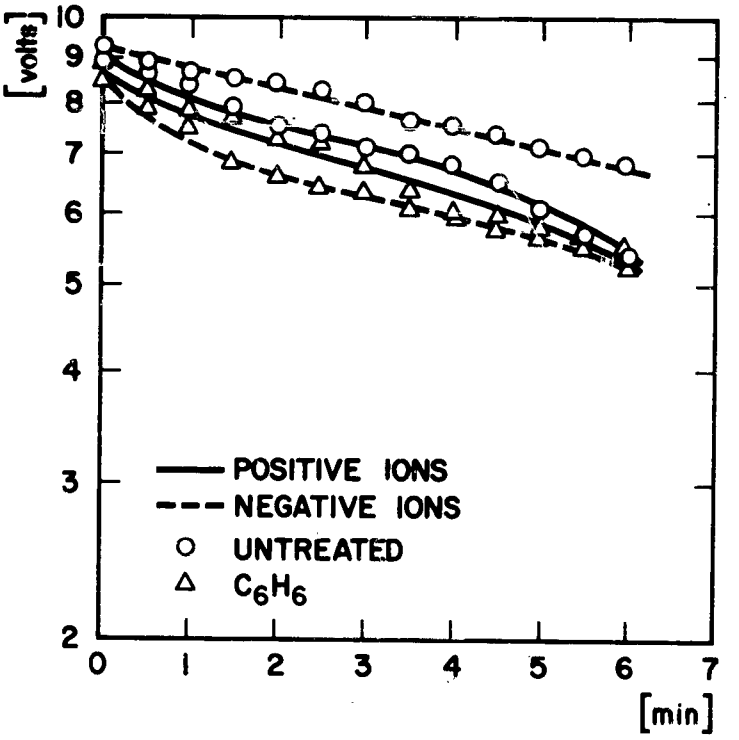


Fig. 7. Effects of surface treatment on the decay of charge accumulated on the oxide. Shown is the decay with time of the maximum negative and positive potential observed on the oxide above the n-type and p-type silicon after bias removal. Untreated and treated (C_6H_6) states are shown.

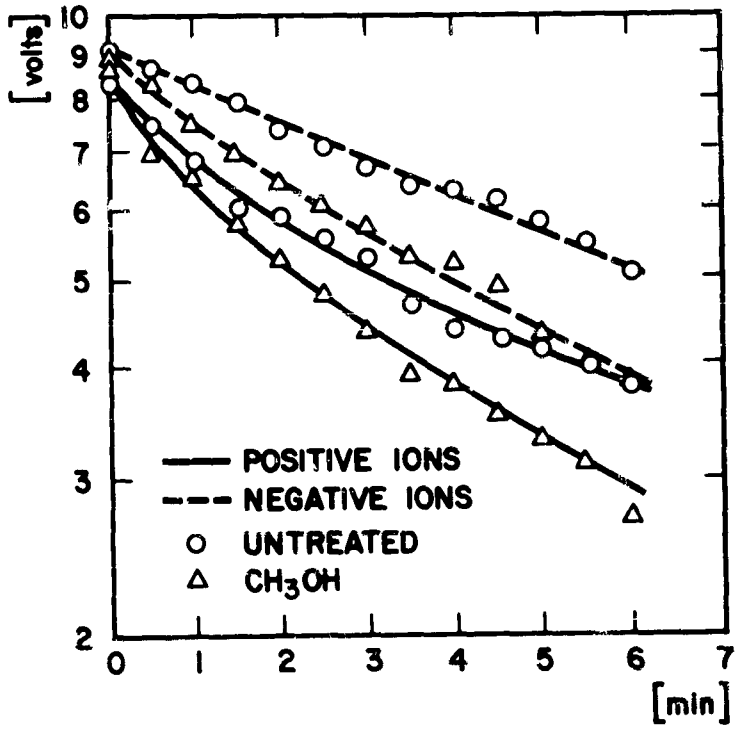


Fig. 8. Effects of surface treatment on the decay of charge accumulated on the oxide. Shown is the decay with time of the maximum negative and positive potential observed on the oxide above the n-type and p-type silicon after bias removal. Untreated and treated (CH_3OH) states are shown.

4. SURFACE CHARGE INFLUENCE ON DEVICE CHARACTERISTICS

It is obvious that these widely differing charge distributions and decay rates can seriously affect the reliability of semiconductor devices. The presence of mobile charges on the oxide surface influences the performance of the junctions and the surface layers of the silicon beneath the oxide.

4.1 Surface Charge, Oxide Contact Potential and Diode Leakage Current

Figure 9 shows an oscillogram of a reverse biased n^+p diode exhibiting distinct channeling characteristics. A narrow channel

region was located between the diode and substrate edge by measurement of the oxide contact potential while the diode was reverse biased. The figure shows the diode under two conditions, "A" during exposure to room atmosphere and "B" after exposure to moist atmosphere. Three distinct characteristics of each curve can be seen. First, considerable channel current flows with very low reverse bias. Second, the channel current saturates and becomes voltage independent. This is referred to as the channel "pinch-off" condition. Third, the normal avalanche breakdown occurs. All three conditions are determined by the surface. It can be seen that introducing a moist ambient has doubled the channel current at "pinch-off" and nearly doubled the breakdown voltage V_B.

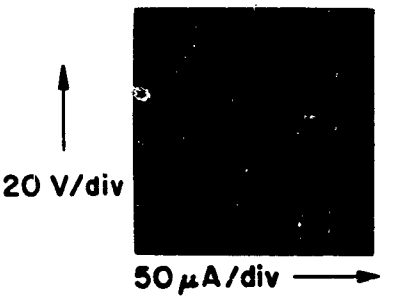


Fig. 9. V-I characteristics of n^+p planar diode with channel. Condition "A" in room atmosphere, "B" after 100% relative humidity. Accumulated surface charge increases channel strength and surface breakdown voltage.

In Fig. 10 is plotted the contact potential in the channel region vs. the diode leakage current, with 20 V reverse bias applied. With these conditions, the leakage current was maintained in the channel "pinch-off" region. It is seen from the plot that an almost linear relationship holds. Figures 9 and 10 together show a definite correlation between breakdown voltage, leakage current, and oxide surface potential.

Condition B of Fig. 9 represents a strong accumulation of surface charges. The increased density of positive charge on the oxide over the p-region enhances the channel conductance and therefore the pinch-off voltage. The breakdown voltage is also increased in agreement with the model proposed by Garrett and Brattain,[9] which explains surface breakdown by field enhancement through surface charges. From the model it follows that high leakage ("channel") is

linked to high breakdown voltage, whereas a low breakdown voltage is associated with low channel leakage.

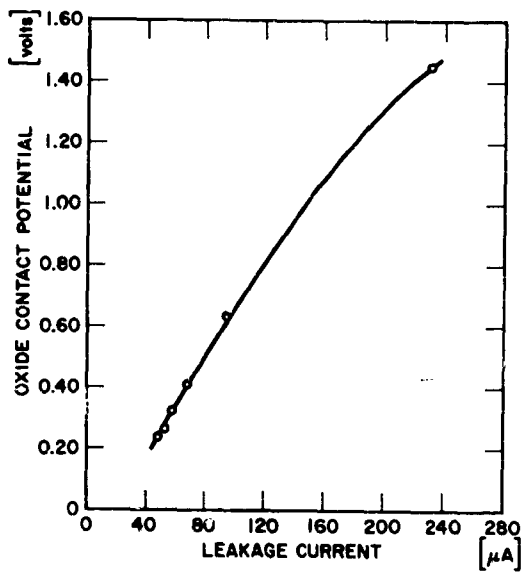


Fig. 10. Oxide contact potential vs. diode leakage current

Figure 11 shows an experimental arrangement to study time effects; an aluminum gate ring is evaporated around the n^+p junction. A saturation current of $I_o \approx 3.7\ \mu A$ is flowing across the junction at a voltage well below V_B. At the time t = 0 a potential of -100 V is applied between gate and substrate. It can be seen from the chart recorder curve in Fig. 11 that the current drops to 1 μA with a very fast time constant. This can be interpreted as a narrowing of the channel by the gate potential, resulting in a reduction of channel current. After this initial drop the current decreases further, at first with a long time constant, the origin of which is not fully understood, and then with an almost exponential time constant. This time constant can be readily explained as an effect of redistribution of surface ions by the potential applied to the gate ring. After 120 seconds, only a very small channel current remains persisting. The bias at the gate is then discontinued. As can be seen from Fig. 11, the very low channel conductivity remains for a certain time; later on, the gradually redistributing surface ions enhance the channel current. This increase in current is approximately exponential. Figure 11 demonstrates clearly that mobile surface ions have a considerable influence on the electrical characteristics of the diode.

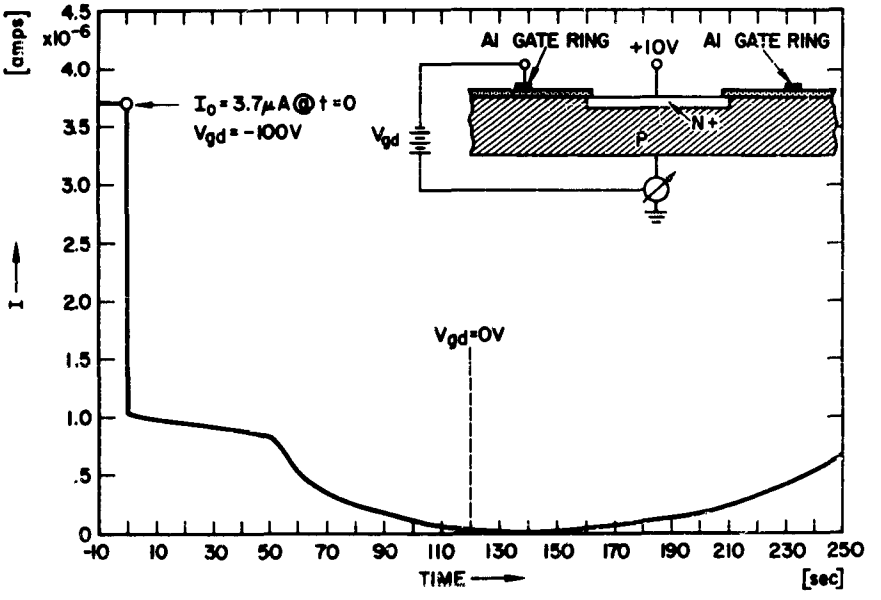


Fig. 11. Influence of surface ions on saturation current I of diode surrounded by metal ring; I plotted vs. time.

4.2 Surface Charge, Oxide Contact Potential and Breakdown Voltage

The increase in breakdown voltage of reverse biased planar diodes ("walk out" effect) has been correlated with changes in oxide contact potential difference in the region directly over the junction. A diode with a typical breakdown voltage of ≈ 150 V before "walk out" effect occurred, was rinsed with methanol to neutralize surface charges, dried, and the Kelvin probe then positioned over the oxide on the weakly doped substrate, close to the junction, as shown in Fig. 12. The diode was then connected to a curve tracer, and reverse bias applied sufficient to produce avalanche breakdown. Avalanche breakdown voltage was measured at a reverse leakage current of 20μA. The diode was allowed to remain biased in the laboratory ambient of ≈30% relative humidity during which time the breakdown voltage increased from 160 V to nearly 400 V. At the same time the contact potential of the oxide was measured using the vibrating capacitance probe, as indicated in the inset. Plotted in Fig. 12 is the breakdown voltage as a function of contact potential difference. Although not indicated in the figure, contact potential was of positive polarity. Over a wide range, an approximately linear dependence with the breakdown voltage prevails. This change in V_B with surface charge accumulation may be understood by the theory of Garrett and Brattain;[9] a description based on the superposition of surface field and junction field has been presented by Shockley and Hooper.[15,16]

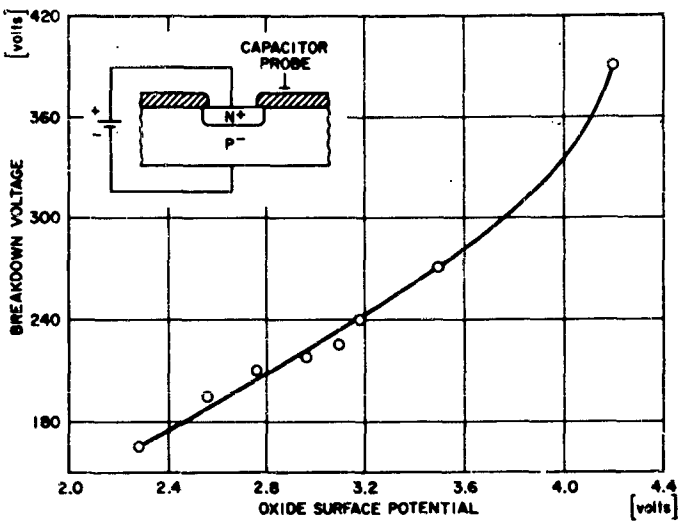


Fig. 12. Plot of breakdown voltage vs. oxide contact potential for diode exposed to moist ambient.

In order to show the time constants involved in the above described effect, an n^+p diode ($V_B \approx 210$ V) was reverse biased at a relative humidity of ≈50%. Figure 13 shows the result of the experiment. V_B is plotted as a function of time. The observed long-time change in V_B is typical for surface ion motion under the field set up by the reverse bias. In this experiment, V_B increased from 210 to 270 V. The increase continued for almost 7 seconds.

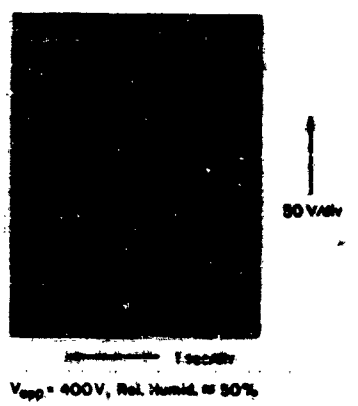


Fig. 13. Surface ion influence on breakdown voltage V_B of n^+p diode. V_B plotted vs. time. V_{app} = 400 V. A load resistance of 1 MΩ is in series with the diode; the dynamic impedance of the diode is about 0.1 MΩ.

4.3 Control Electrode

In order to replace the field of mobile surface charges by an easily controllable surface field, an aluminum gate ring over the oxide was evaporated around the planar junction (distance 10, 30, or 100 μ from the junction). Bias V_g was applied to the gate in order to study the influence of this additional surface field on the diode breakdown voltage. The externally applied electric field penetrates the semiconductor surface and extends into the space charge layer. This field modifies that produced by reverse bias, V_d, across the p-n junction, thus changing the voltage across the junction at which avalanche breakdown occurs. The plot of Fig. 14 shows the effective breakdown voltage, namely the surface breakdown voltage, V_s, as a function of the voltage, V_g, applied to the field electrode; also plotted is the ratio V_s/V_b, where V_b denotes the final constant maximum voltage ("corner breakdown" voltage) which is close to the predicted bulk breakdown voltage. It is seen that the breakdown voltage of the diode can be controlled as dc voltage is applied to the metal ring; by application of negative voltage (polarity required to produce an accumulation layer over a p-type substrate). V_s is observed to decrease, and conversely, application of positive voltage, i. e. polarity of the same sign as the conductivity type of the weakly doped material, causes an increase of V_s.

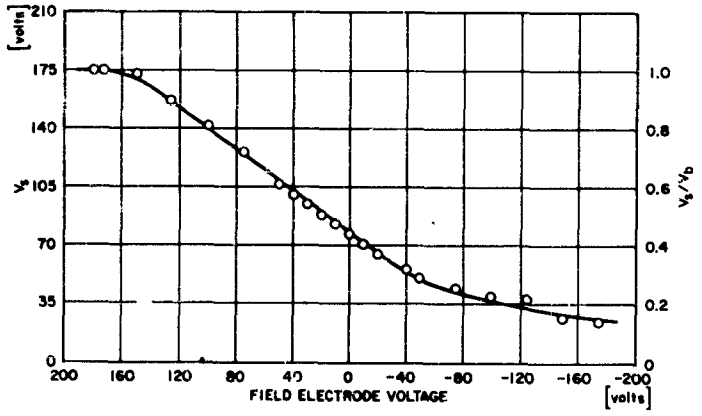


Fig. 14. Surface breakdown voltage. V_s, vs. field electrode voltage, V_g. Also shown is the ratio of surface to bulk breakdown voltage, V_s/V_b.

Three regions of the curve in Fig. 14 can be distinguished. First, the center portion of the curve shows a strictly linear dependence of V_s on V_g extending from positive to zero and small negative values of V_g. This linear relationship can be understood by the theory of the "surface controlled avalanche transistor (SCAT)" developed by

Shockley.[15,16,20] Briefly, the condition for avalanche breakdown can be approximated[21] by the condition that the maximum field F in the avalanche source region (located usually close to the intersection of p-n junction and silicon/oxide interface) reaches or exceeds the breakdown value F_B,

$$F \geq F_B. \tag{4}$$

The corresponding voltage at breakdown will be denoted as $V_B = V_s$. In first approximation, F can be written by superposition as

$$F = F(V_g) + F(V_d). \tag{5}$$

For wide space charge layer widths (in Fig. 14: for positive and small negative field electrode voltages), the effect of V_d is less than the effect of V_g because of the larger distance avalanche region/drain, and because of the shielding action of the gate electrode. The total voltage V acting over the width of the avalanche source to produce the internal field F can be written as

$$V = V_g + V_d/\mu. \tag{6}$$

At breakdown yields, according to Eq. (4):

$$V \geq V_B = V_s \tag{7}$$

μ is the amplification factor defined as

$$\mu = \frac{\partial V_d}{\partial V_g} > 1 \text{ at } I_d = \text{const.} \tag{8}$$

μ describes how strongly V_g influences the field F compared with the influence of V_d on F, and has to be calculated from potential theory.

The second portion of the curve in Fig. 14 shows a curvature at high negative voltages leading to an asymptotic decrease of V_s. The high negative values of V_g cause a narrowing of the space charge layer width so that it becomes comparable to the oxide layer thickness. There μ approximates 1, and the influence of V_d on F exceeds the effect of V_g. Finally, V_s becomes almost independent of changes in V_g.

The third portion of the curve is the flattening at high, positive values of V_g which leads to a constant maximum voltage V_b (in Fig. 14 at 175 V) determined by corner breakdown of the diffused layer. A subsequent decrease of V_s at still higher V_g values, as found by Castrucci and Logan,[17] has not been observed. Curves similar to the one shown in Fig. 14 have recently been measured by Brown.[22]

In order to achieve an equipotential layer over the oxide surface and hinder surface ion motion, a thin metal layer was evaporated. As material, gold was selected. The layers were 100 to 200 Å thick (diameter 350 μ). The gold layer was at the same potential as the n^+ plug. While the breakdown voltage of the planar diodes without the gold layer was about 400 V, V_s after gold evaporation was about 1400 V. In addition, the V-I characteristic was hard. It was interesting to observe how this considerable increase of V_s by a factor of 3.5 occurred. On an oscilloscope, 900 V of the breakdown voltage appeared instantaneously, the additional 500 V occurred with a drift of a short time constant (≈ 1 sec). The first increase is caused by the "SCAT" effect described above, the second one probably by the push-out of the surface channel due to an Atalla-type effect.[2,13,17]

4.4 Stabilization of Device Characteristics by Heating of Sample or Freezing of Surface Charges

n^+p and p^+n diodes have been subjected to repeated heat treatments at 600°C for 30 minutes in vacuum (pressure 10^{-6} Torr, residual gas pure N_2). Ion accumulation and decay curves have then been measured using the techniques described in Section 3. The result for an n^+p diode is shown in Fig. 15. While the ion decay curves before treatment are similar to those of Figs. 5 through 8, the curves measured after the second heat treatment show a considerable reduction of the amount of ions accumulated and a decrease of the decay time constant. Finally, after the third heat treatment almost no mobile surface ions are left. The diodes exhibited hard and stable V-I characteristics. This stabilization persisted not only as long as the measurements were performed in dry ambient, but also for several hours at a relative humidity of 30%. Heat treatments were also performed at 200°C with similar results.

It is well known that charges which are normally mobile on an insulating surface can be made immobile in a number of ways. Atalla[23] succeeded in freezing surface charges on oxide covered p-n junctions by actually freezing them, i.e. lowering the device temperature below 0°C. Surface treatments with organic compounds[24,25] have proven to be quite effective in reducing ion drift on oxides in moist ambients; the reason for this being that large organic radicals "tie up" the OH^- groups of the surface water layer, thereby reducing the surface conductivity. Silicone rubber has also been employed to reduce the effects of charge motion on high voltage diodes.[26]

It has been found that covering the surface of a diode with commercial KPR (Kodak Photo Resist) has a marked effect on the stability of the V-I characteristics in a moist ambient. For illustration, Fig. 16 shows the influence of KPR treatment on surface ion accumulation and decay. While the ion decay curves before KPR application behave in the familiar fashion, no ion accumulation and hence no ion decay is observed after the KPR treatment. The surface ions

appear immobilized by the KPR layer. Reverse bias was applied for three days without deterioration of the V-I characteristic or any observable influence of moist ambient.

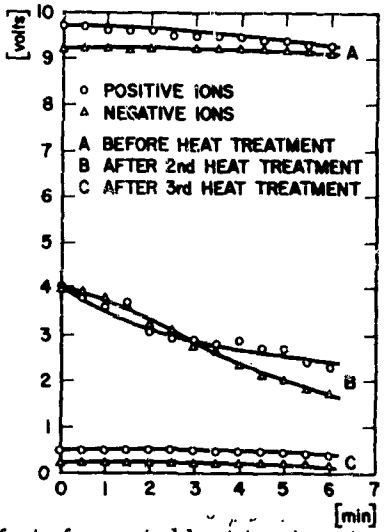


Fig. 15. Effect of repeated heat treatment on surface ion accumulation and decay.

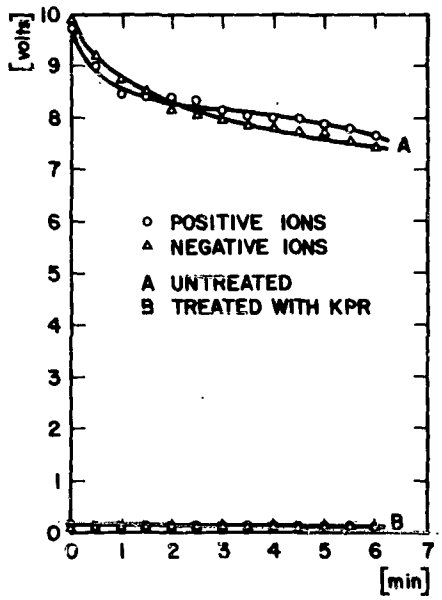
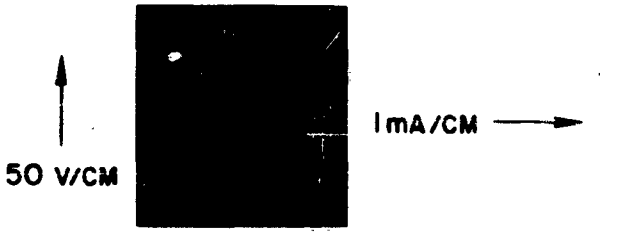
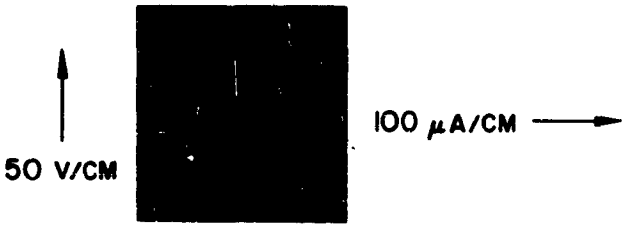


Fig. 16. Effect of KPR treatment on surface ion accumulation and decay.

Covering the diode surface with a masking wax also improved the stability of the V-I characteristics. To illustrate this fact, a p-n junction diode was heated on a hot plate, and the diode surface covered with molten apiezon masking wax. The diode V-I characteristics were photographed under two conditions. First, reverse bias was applied while the wax was in a molten state; second, the wax was allowed to harden before reverse bias was applied. Figure 17 shows these two conditions.



WAX APPLIED WITH BIAS ON



WAX APPLIED WITH BIAS OFF

Fig. 17. V-I characteristics of planar diode with "frozen in" surface ion distribution.

The upper photograph shows the characteristic when bias was applied throughout the waxing procedure, and the lower photograph shows the reverse characteristic upon applying bias after the wax had hardened. When the wax was melted with bias applied, "walk out" occurred, and breakdown voltage increased from 130 V to approximately 300 V. In addition, channel current increased from 20 μA to nearly 1 mA. The magnitude of reverse bias applied during the time the wax was soft was the essential factor. The upper curve was "frozen" with 300 V bias applied, hence a large amount of charge separation, channeling and increased V_s. The lower curve, on the other hand, was "frozen" with no bias, hence no charge separation, higher surface fields, and reduced V_s. Once the wax had hardened, it was found that both V-I characteristics were very stable, and were impervious to moisture or other ambients except for those that dissolve the wax. This diode was stored in the high V_s ("walked out")condition

and retested periodically over a one month period. A slow drift back to the "pre-frozen" condition was observed, which took nearly a full month, rather than seconds as observed without surface protection. Although the surface charges were not permanently "frozen in", this experiment is clear evidence that the V-I characteristics of a p-n junction can be tailored by fixing the surface ion distribution in a manner similar to that just described.

5. SUMMARY

Diodes of various geometrical dimensions and breakdown voltages were diffused into p- as well as -n type material. Various oxide preparation techniques were employed. Quantitative measurements of the potential distribution on the surface determined by the Kelvin vibrating condensor method were performed. The measurements were carried out in dry and humid ambients. The contact potential distribution observed after bias had been applied for a certain length of time was compared to the complementary error function solution expected for a distributed capacitance and sheet resistance. From the theoretical curve fitting the experimental data best, the diffusion coefficient of the charges on the oxide surface could be estimated.

After bias removal the accumulated positive and negative charges were allowed to decay. Significant differences of decay times could be observed determined by the chemical preparation of the sample. In particular, a treatment with water caused peculiar differences of the behavior of positive and negative charges.

In the last part of this paper the influence of mobile charges on device characteristics is described. The relationship between oxide contact potential and diode leakage current was established. The vibrating condensor probe was also used to determine the correlation between surface charges, oxide contact potential and diode breakdown voltage. The time dependences of the leakage current and the breakdown voltage were investigated. Particular emphasis was put on a study of the influence of an external bias applied to a gate electrode on the breakdown voltage. Finally, methods for stabilizing the device characteristics by heating of the sample or freezing of mobile surface charges were investigated. It has been shown that a treatment with silicone rubber, KPR, and black wax, as well as the evaporation of a thin metal layer enhance the reliability of the planar diodes.

Acknowledgements

The author wishes to express his sincere thanks to Dr. W. Shockley for many stimulating discussions. He also gratefully acknowledges the very valuable assistance of Mr. W. W. Hooper (now at Fairchild Semiconductors, Palo Alto, California),

Mr. R. Stephens, Mr. R. D. Woodruff, and Mr. P. G. G. van Loon, who prepared the samples and performed the measurements. Mr. W. W. Hooper also helped in the evaluation and interpretation of the results.

References

1. W. L. Brown, "n-Type Surface Conductivity of p-Type Germanium", Phys. Rev. 91, 518-527 (1953).

2. M. M. Atalla, A. R. Bray and R. Lindner, "Stability of Thermally oxidized Silicon Junctions in Wet Atmospheres", Proc. Inst. El. Engr. (London) 106, Pt. B, Suppl. 17, 1130-1137 (1960).

3. W. Shockley, H. J. Queisser, and W. W. Hooper, "Charges on Oxidized Silicon Surfaces", Phys. Rev. Lett. 11, 489-490(1963).

4. W. Shockley, W. W. Hooper, H. J. Queisser, and W. Schroen, "Mobile Electric Charges on Insulating Oxides with Application to Oxide Covered Silicon p-n Junctions", Surf. Science 2, 277-287 (1964).

5. W. Schroen, W. W. Hooper, and R. D. Woodruff, "Vibrating Capacitor Probe for Contact Potential Measurements", to be published.

6. D. P. Kennedy and R. R. O'Brien, "Avalanche Breakdown Characteristics of a Diffused p-n Junction", IRE Trans. ED-9, 478-483 (1962).

7. G. Gibbons and J. Kocsis, "Breakdown Voltage of Germanium Plane-Cylindrical Junctions", ED12-4, 193-198 (1965).

8. W. Shockley and G. L. Pearson, "Modulation of Conductance of Thin Films of Semiconductors by Surface Charges", Phys. Rev. 74, 232-233 (1948).

9. C. G. B. Garrett and W. H. Brattain, "Some Experiments on, and a Theory of, Surface Breakdown", J. Appl. Phys. 27, 299-306 (1956).

10. D. Kahng and M. M. Atalla, Paper presented at IRE-AIEE Solid-State Device Research Conf., Pittsburgh, Pa., June 1960.

11. S. R. Hofstein and F. P. Heiman, "The Silicon Insulated-Gate Field-Effect Transistor", Proc. IEEE 51, 1190-1202 (1963).

12. H. C. Nathanson, "A High Field Triode", Solid State El. 8, 349-363 (1965).

13. W. Schroen and W. W. Hooper, "Failure Mechanisms in Silicon Semiconductors", Final Report, RADC-TR-64-524 (304 pages) (1965).

14. W. Schroen, W. W. Hooper, and H. J. Queisser, "Failure Mechanisms in Silicon Semiconductors", Quarterly Report, RADC-TDR-64-153 (59 pages) (1964).

15. W. Shockley and W. W. Hooper, "The Surface Controlled Avalanche Transistor", IEEE Trans. ED-12, 535 (1964).

16. W. Shockley and W. W. Hooper, "The Surface-Controlled Avalanche Transistor", WESCON Techn. Papers 8, Pt. 3, Paper 12.1 (5 pages)(1964).

17. P. P. Castrucci and J. S. Logan, "Electrode Control of SiO_2-Passivated Planar Junctions", IBM J. 8, 394-399 (1964).

18. L. E. Clark and E. B. Mack, "Fabrication of High-Voltage Oxide-Passivated Silicon Junction Structures", Electrochem. Soc. Extended Abstr., Electr. Div. 13, 117 (1964), Abstr. No. 52.

19. W. Schroen, "Reliability Physics Studies on Transistors", Quarterly Report, RADC-TR-65-141 (90 pages) (1965).

20. R. M. Scarlett, "New Concepts for Solid State Microwave Generators", Quarterly Report No. 6 for Contract DA 36-039-AMC-00001(E) (26 pages) (1964).

21. W. Shockley, "Problems Related to p-n Junctions in Silicon", Solid State El. 2, 35-67 (1961).

22. G. A. Brown, "A study of Silicon Surface Junction Reverse Characteristics Using Surface Potential Controlled Diodes", J. Electrochem. Soc. 112, 69C (1965).

23. M. M. Atalla, private communication.

24. J. A. Cunningham, L. E. Sharif, and S. S. Baird, "The Formation of Organo-Silicon Surfaces", Electrochem. Techn. 1, 242-246 (1963).

25. C. T. Hall, "Surface Stabilized High Voltage Diodes", Microsemicond. Corp. Rep. 13.2 (6 pages) (1962).

26. W. Schroen, K. Hubner, J. Beaudouin, and R. M. Scarlett, WESCON Techn. Papers 8, Pt. 3, Paper 12.2 (10 pages) (1964).

MECHANISMS OF CHANNEL CURRENT FORMATION IN SILICON P-N JUNCTIONS

D. J. Fitzgerald and A. S. Grove

Fairchild Semiconductor
Research and Development Laboratory
Palo Alto, California

ABSTRACT

The physical mechanisms of the formation of channel currents associated with reverse biased P-N junctions are studied in detail. It is demonstrated that a channel characteristic results when a site of very high carrier generation rate is connected by an inversion layer to the P-N junction. Such sites are shown to be due either to a fault associated with the field-induced junction between the inversion layer and underlying silicon, or to the breakdown of this junction.

1. INTRODUCTION

Typical reverse characteristics of silicon planar diodes are shown both on linear and log-log scales in Figure 1a. Characteristics such as these are referred to as hard characteristics. Also shown in Figure 1 are the reverse characteristics corresponding to the two most common reverse-bias failure modes. The first of these, the soft characteristic, has been shown by Goetzberger and Shockley (1) to be the result of the presence of defects within the depletion region of the junction. The second, the channel characteristic, is the subject of this paper.

It is well known that surface inversion of one side of the junction plays an important role in channel current junction failures. Therefore we shall begin by briefly reviewing the mechanisms of inversion layer formation on thermally oxidized silicon surfaces.

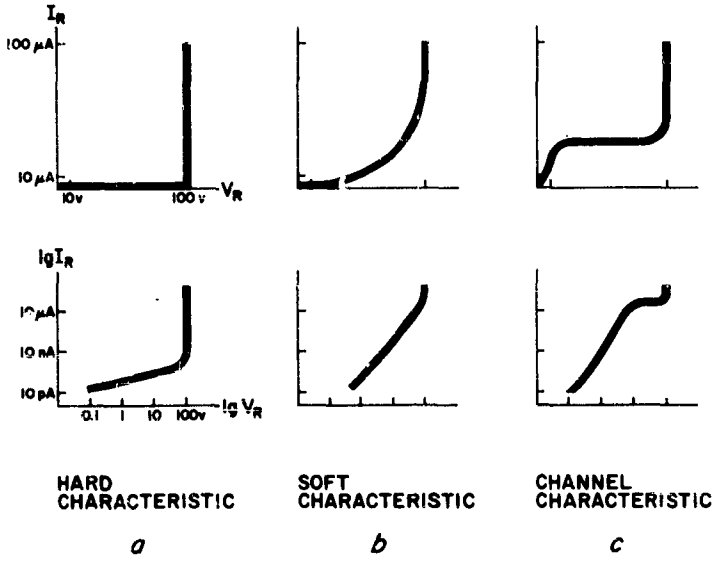


Figure 1

Comparison between a hard reverse characteristic and the reverse characteristics corresponding to the two most common reverse-bias failure modes. Note from the log-log characteristics that the reverse currents of the failed junctions are orders of magnitude larger than the reverse current of the hard junction.

2. INVERSION LAYER FORMATION

A clean thermal oxide is characterized by a positive surface state charge which is fixed at the oxide-silicon interface (2). The density of this charge per unit area, Q_{ss}, is generally of the order of 10^{11}-10^{12} electronic charges/cm^2. If its magnitude is large and the doping concentration of the silicon low, the positive surface state charge will induce an inversion layer on P-type material as illustrated in Figure 2a.

A second type of positive charge which can be present in SiO_2 is due to alkali ion contamination (Figure 2b). Such charge can migrate readily in the oxide at elevated temperatures. The kinetics of this migration have been studied in detail by Snow et. al. (3) using MOS capacitors.

An additional type of positive charge which can exist within the oxide is a positive space-charge that develops upon exposure to ionizing radiation (4). This is illustrated in Figure 2c.

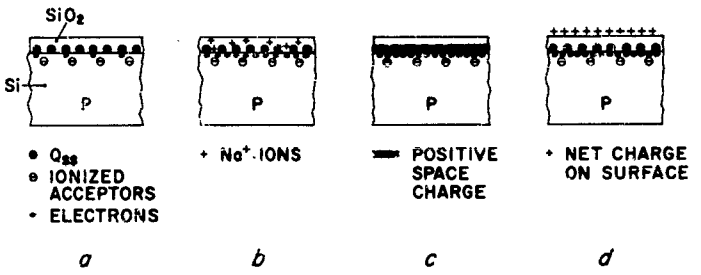


Figure 2

Schematic representation of mechanisms of inversion layer formation on thermally oxidized silicon surfaces. The mechanisms illustrated are due to (a) surface state charge, Q_{ss} , (b) Na+ ions within the oxide, (c) positive space charge resulting from ionizing radiation and (d) net charge on the outer surface of the oxide.

In Figure 2d an inversion layer caused by a net positive charge on the outer surface of the oxide is shown. Such a charge can readily arise in the presence of an electric field if moisture is present in the ambient (5). Unlike the surface state charge Q_{ss}, and ionic contamination within the oxide, the net charge on the oxide can be either positive or negative.

In the absence of a metal-over-oxide gate, the net charge due to the above mechanisms, $\sum Q$, will induce an equal and opposite charge, Q_s, in the semiconductor. The absolute value of Q_s corresponding to the onset of inversion, $|Q_s(inv)|$, is given in Figure 3 as a function of surface impurity concentration, C_s. For a P-type semiconductor, inversion occurs when $\sum Q$ is positive and exceeds $|Q_s(inv)|$ given by Figure 3 for the surface concentration of the material. For an N-type semiconductor inversion occurs when $\sum Q$ is negative and larger in magnitude than $|Q_s(inv)|$

Another way of inducing an inversion layer is simply by application of a voltage to the gate of an MOS structure. By varying the voltage, the condition of the surface under the gate can be varied in a controlled manner from strong inversion to strong accumulation.

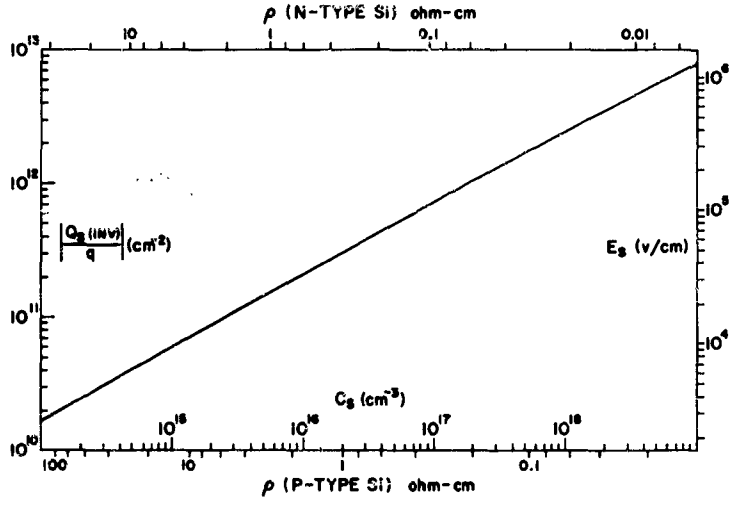


Figure 3

The magnitude of the total charge induced in the silicon at the onset of inversion, $|Q_B(inv)|$, as a function of the surface impurity concentration, C_B. This relationship was obtained by setting $\phi_S=\phi_F$ in Equation (5) of Ref 2. A scale giving the corresponding surface field, E_s, is also included; this field is given by Gauss's law, $E_S=Q_B(inv)/K_S\epsilon_0$.

3. CARRIER GENERATION

Any of the mechanisms of inversion layer formation reviewed in the last section, or a combination of them, can result in inversion of the surface of one side of a junction. In Figure 4 a P^+N junction having a metal-over-oxide gate covering all of the junction perimeter and extending out over part of the N region is shown. By applying a negative voltage to this gate, the surface of the N region under the gate can be inverted. A field-induced junction then exists in parallel with the original metallurgical junction as illustrated.

It is commonly assumed in the literature that when conditions are such as illustrated in Figure 4 a large channel current inevitably results. The fact that channel currents can be orders of magnitude greater than the reverse current of a corresponding hard junction (see Figure 1) is usually attributed to the existence of a high concentration of thermally activated carrier generation centers close to the oxide-silicon interface. The validity of this model in the case of large channel currents is highly doubtful as is shown by the following arguments:

From theoretical considerations the density of such centers required for a particular channel current can be calculated if both the area and the depletion region width of the field-induced junction are known. Values obtained in this manner are many orders of magnitude too high for the model to be at all feasible. For example, for those channel currents associated with P^+ regions (which we will consider in detail later) one gets the following values (6):

Required density of surface generation centers $\approx 10^{17} cm^{-2}$, i.e. $\sim$100 per surface atom

Required density of bulk generation centers $\approx 10^{23} cm^{-3}$, i.e. $\sim$10 per bulk atom.

Experimentally one can observe the change in reverse current of a gated diode, such as that illustrated in Figure 4, as the condition of the surface under the gate is varied from strong accumulation to strong inversion. Typically this change is very small (as will be shown in a later figure) in disagreement with the predictions of the model.

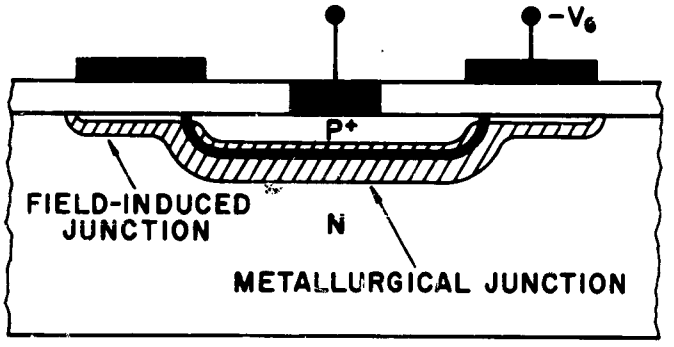


Figure 4
A field-induced junction in parallel with a metallurgical junction.

Thus it appears that a channel current involves not only an inversion layer but also some site of very high carrier generation rate which can supply carriers to this inversion layer, i.e.,

CHANNEL CURRENT = INVERSION LAYER + CARRIER GENERATION SITE.

The remainder of this paper will be devoted to the consideration of such carrier generation sites. It will be shown that the generation mechanisms fall into two classes: (i) mechanisms involving defects in the depletion region of the field-induced junction, and (ii) mechanisms involving breakdown of the field-induced junction. Since the excess currents which can flow in metallurgical junctions also involve either a defect within, or breakdown of the junction, there is a close analogy between these two types of large excess currents. A defect in the depletion region of a metallurgical junction results in a soft characteristic (1), whereas a defect in the depletion region of a field-induced junction results in a channel characteristic. In the latter case the carriers must flow through the series resistance associated with the inversion layer and, because of this, current saturation occurs as in an MOS transistor. In the same way, when the field-induced junction breaks down, this series resistance gives rise to a saturating channel characteristic instead of the customary breakdown characteristic of metallurgical junctions.

4. FAULTS ASSOCIATED WITH THE FIELD-INDUCED JUNCTION

We will first consider those carrier generation mechanisms which involve some fault or defect associated with the field-induced junction. These mechanisms will be demonstrated by controlled experiments using gated diodes. The diodes initially had hard reverse characteristics, even when the voltage applied to the gate was such that the surface under the gate was strongly inverted. A carrier generation site was then deliberately created resulting in a large channel current. We will study the first experiment in some detail, even though it is very simple, because it serves to bring out the significance of the concept contained in the channel expression given above.

4a. Carrier Generation At Scribe Lines

The first experiment involved a gated N^+P diode having the structure shown in the inset to Figure 5. Note that the gate did not extend to the edge of the die. For $V_G \leq 0$, none of the P surface of this device was inverted. The reverse current initially was in the 10 to 100 pA range and increased only $\sim$ 5% upon inversion of the surface under the gate. The experiment was as follows: A corner was scribed off of the die in such a way that part of the scribe line passed through the gate but not through the metallurgical junction. (This is indicated schematically by the section line AA drawn through the cross-section of the device in Figure 5). Then with a positive gate voltage, part of the surface of the P region could be inverted all the way from the metallurgical junction to the new scribe line. The reverse characteristics for $V_G \leq 0$ (surface not inverted) were unchanged by the scribing; the curve-tracer display was the zero current trace shown in Figure 5. However, for V_G sufficiently positive for the surface under the gate to be inverted a large channel current flowed. The particular case shown in Figure 5 represents a six orders of magnitude increase in I_R with inversion of the P region.

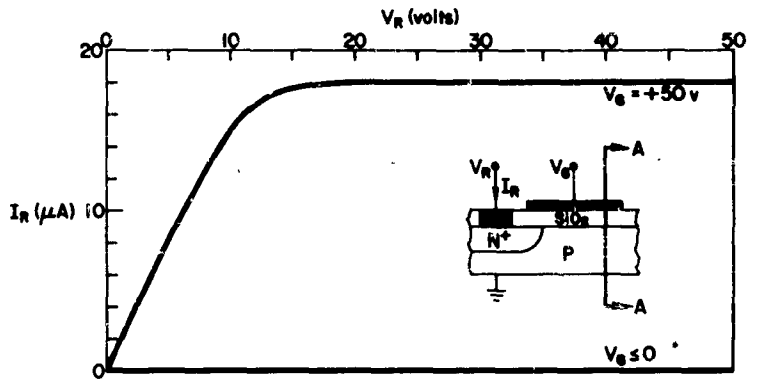


Figure 5

Reverse current-voltage characteristics after scribing along Section AA. ($C_B = 1x10^{16}$ boron atoms/cm^3, $Q_{ss}/q = 2x10^{11} cm^{-2}$, $x_o = 0.75\mu$).

Let us now consider this experiment in terms of the channel expression given in the last section. Initially, a channel current did not flow when the surface of the P region was inverted; this implies that a carrier generation site was not present. Since after scribing a channel current did flow whenever the surface of the P region was inverted, it is concluded that scribing along line AA resulted in the creation of a carrier generation site. This is not unreasonable since the carrier generation rate associated with the severely damaged region along the scribe line is expected to be extremely high.

The same result as above is obtained by extending the inversion layer all the way to the original scribe lines. This can be shown very easily with the same gated structure by blowing moist air on the surface of the die while the gate is positively biased. As shown schematically in Figure 6, the oxide surface then tends to charge up to the gate potential (5), and so the inversion layer spreads outwards from the gate. If V_G is greater than a critical value (which is a function of the charge within the oxide, the oxide thickness and the surface concentration of the P region), the inversion layer will eventually extend from the metallurgical junction to a scribe line, and a channel current will then result.

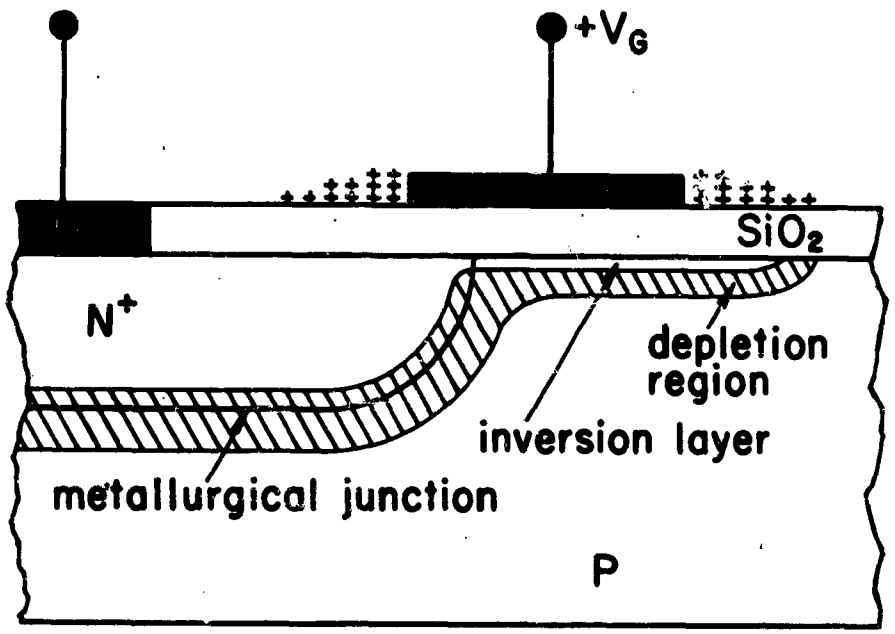


Figure 6

Illustration of the role of oxide surface conduction in inducing an inversion layer. When the inversion layer reaches a defective region (e.g., a scribe line), channel current results.

It follows from these controlled experiments that channel currents can occur in production devices whenever the surface of a device becomes inverted all the way from a metallurgical junction to an edge of the die. This fact accounts for the very significant reliability improvement which results by surrounding a junction with a diffused channel-stop region, since it is much less likely that an inversion layer will then extend all the way to a scribe line. However, a channel-stop does not prevent the formation of the other types of channel currents which will be considered in the remainder of this paper.

4b. Carrier Generation At Surface Defects

A junction with an annular channel-stop can obviously experience channel current failure if there is a carrier generation site between the metallurgical junction and the channel-stop region. To demonstrate this a carrier generation site was again deliberately

created in a gated N^+P diode, but one which had an annular P^+ channel-stop added as shown in the inset to Figure 7. The initial reverse characteristics are shown dotted for V_G = - 100v, 0, and +100v. (100v applied to the gate of this device corresponds to a net charge induced in the silicon of magnitude $3x10^{12}$ electronic charges/cm^2). Note, as previously pointed out, that there is only a very slight change in reverse current as the condition of the surface of the P region under the gate is varied from strong accumulation to strong inversion.

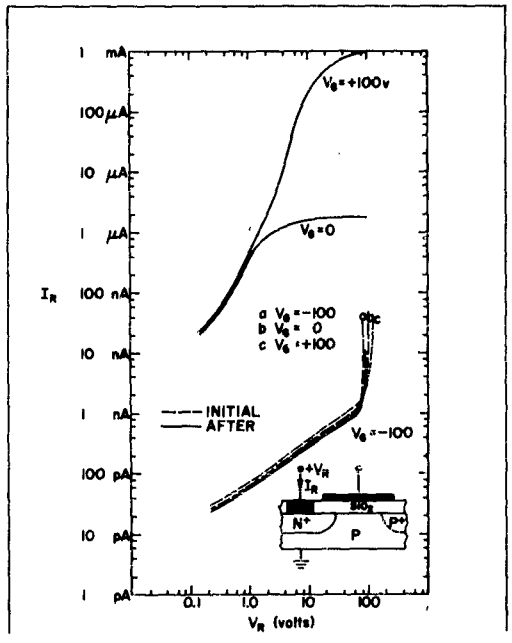


Figure 7

Reverse current-voltage characteristics before (dotted) and after (solid) electrostatic damage to the oxide-silicon interface under the gate. (C_B=$5x10^{15}$ boron atoms/ cm^3, Q_{ss}/q=$5x10^{11}$ cm^{-2}, x_o=0.75μ).

The technique used to create a carrier generation site in this device was to partially discharge a capacitor between gate and substrate. (This technique is based on the well known fact that MOS devices can be permanently damaged if their gates pick up stray electrostatic potentials). The insulating properties of the oxide were unchanged by the discharge, but the channel characteristics shown solid in Figure 7 resulted when the P^- surface was inverted. In terms of the channel expression, a carrier generation site was created. The carrier generation site might possibly be aluminum which was shot

through some weak point in the oxide as a result of a localized oxide breakdown during the partial discharge of the capacitor. This would give rise to a region of extremely high carrier generation rate which could effectively short the inversion layer to the underlying silicon. It is interesting to note here that, as early as 1958, Bray and Lindner (7) pointed out that defects resulting from localized oxide breakdown could play a significant role in channel current failures of passivated silicon junctions.

A small percentage of the devices of the type shown in Figure 7 were found to have initial reverse characteristics similar to those shown solid. The percentage of such devices depended very strongly upon processing history. Since the incidence of defects which result in soft junctions (1) is highly process-dependent, it is concluded that these devices had processing-induced defects between the metallurgical junction and the channel-stop.

The channel currents just considered are of the same class as those channel currents having carrier generation sites at scribe lines, because in each case the carrier generation mechanism involves a fault associated with the field-induced junction. We will now consider a completely different class of channel currents, a class in which the carrier generation mechanism involves an intrinsic property of the field-induced junction rather than a fault. This property is junction breakdown.

5. BREAKDOWN OF THE FIELD-INDUCED JUNCTION

The controlled experiments which were considered in the previous section were performed on N^+P diodes. For these devices inversion was induced on the P side of the junction. Analogous results are obtained with P^+N diodes if the N region is inverted. However, since the surface state charge and ionic contamination within the oxide are both positive, inversion of N material is less likely to occur in production devices. The class of channel currents which we will now consider involves inversion of the surface of a P^+ diffused region. If the concentration of positive charge within the oxide is high, this can take place in many types of production devices, for example P-N-P transistors having P^+ annular channel-stop regions, or N-P-N transistors where the P^+ region involved is the base.

Controlled experiments were performed on devices which were initially hard. The structure used was the gated P^+N diode shown in Figure 8. The oxide under the gate of this structure was purposely contaminated with sodium ions using the method of Snow, et. al. (3). This method involves a rinse in a dilute solution of sodium chloride prior to metallization. By applying a positive gate voltage at an elevated temperature to the finished devices, Na^+ ions from the NaCl trapped at the metal-oxide interface can be driven down to the oxide-silicon interface, and then "frozen" into position by cooling under bias.

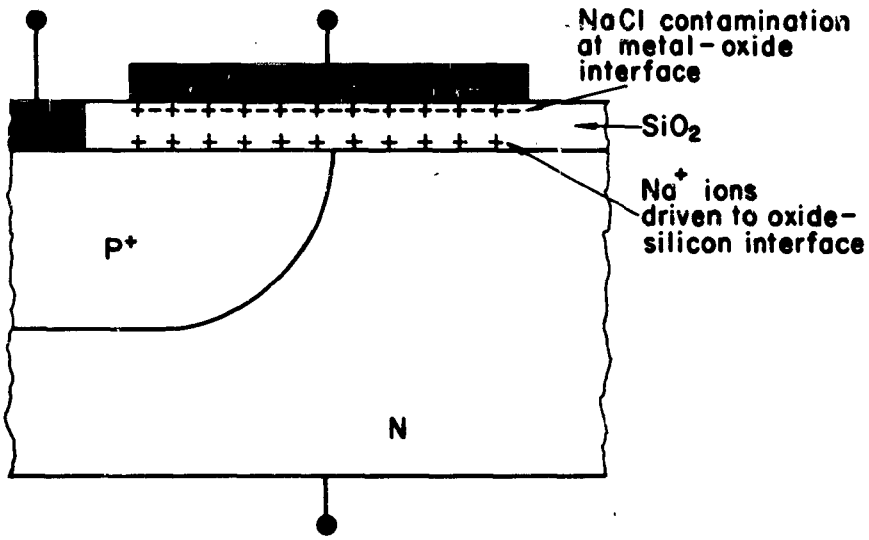


Figure 8

Schematic cross section of the experimental structure used to study channel currents associated with P+ regions. NaCl contamination was deposited at the metal-oxide interface prior to metallization. Na+ ions could then be driven down to the oxide-silicon interface on the finished devices. (C_B=5x10^{15} phosphorus atoms/cm^3, Q_{ss}/q=3x10^{11}cm^{-2} x_o=0.7μ).

The devices used in the experiments initially had reverse currents of less than 1 nA. It was found that a channel current ALWAYS resulted when the density of Na+ ions near the oxide-silicon interface exceeded a critical value. This is shown in Figure 9, where lg I_R (measured at V_R=10v) is plotted as a function of Q_s, the charge per unit area induced in the silicon by the positive charge within the oxide. Q_s was varied by gradually driving Na^+ ions to the oxide-silicon interface and its value was determined from the shift of MOS CV characteristics along the voltage axis (3).

The experimental data of Figure 9 were obtained using several diodes from a device run having a P^+ surface concentration of 4x10^{18} boron atoms/cm^3. The value of Q_s at the onset of inversion of a P+ surface of this concentration (obtained from Figure 3) is indicated in the figure; note that channel currents appear when the magnitude of Q_s exceeds this value. Thus the onset of channel current flow corresponds to the formation of a field-induced junction as shown in Figure 10.

A very large increase in zero-bias junction capacitance occurs along with the onset of channel current flow. This is to be expected on the basis of the representation of Figure 10 because the

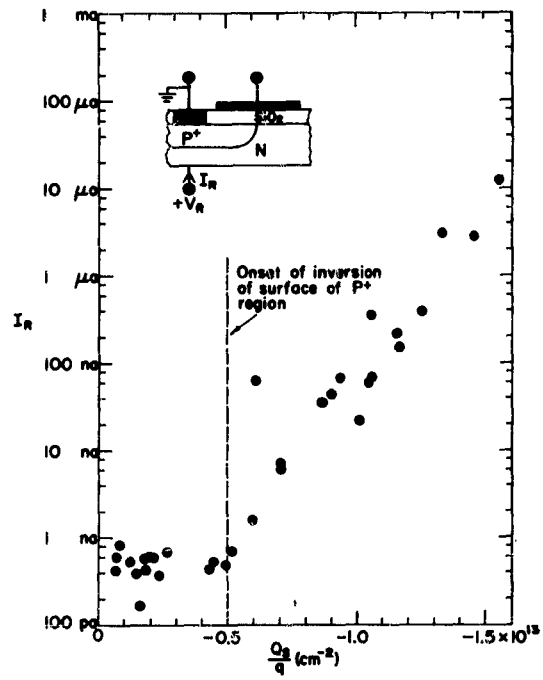


Figure 9

Reverse current at V_R=10 volts as a function of the charge per unit area induced in the silicon, Q_s. Experimental data were obtained using several diodes from a device run having a P^+ surface concentration of $4x10^{18}$ boron atoms/cm³.

capacitance of the field-induced junction is in parallel with the capacitance of the metallurgical junction. Since the field-induced junction area is known, a value for its depletion region width can be estimated from this increase in capacitance. This value is ~100 Å and is in reasonable agreement with the theoretically calculated value of the maximum width of the depletion region corresponding to inversion of the P^+ surface (2).

Having accounted for the inversion layer term of the channel expression, we will now turn our attention to the carrier generation mechanism responsible for supplying carriers to this inversion layer. It is possible to determine this mechanism by studying the channel characteristics in detail.

Typical characteristics are shown in Figure 11. The saturation value of the channel current is again determined by the properties of the inversion layer, and can be modulated by varying V_G as shown. The shape of the low current part of the characteristic is determined by the carrier generation mechanism. Note that this part looks very much like the breakdown characteristic of a junction. Experimental evidence that it is in fact the breakdown characteristic of the

field-induced junction between the inversion layer and the P^+ region has been presented in detail by the authors elsewhere (6). This evidence will now be briefly reviewed.

Breakdown of the field-induced junction of Figure 10 is determined by the surface concentration of the P^+ region. Thus if the low current part of the channel characteristic is indeed the result of breakdown of the field-induced junction, then its shape should in turn depend on the P^+ surface concentration. This is in fact the case as can be seen in Figure 12 where the low current parts of channel characteristics are plotted with P^+ surface concentration as parameter. Also shown in this figure are reverse characteristics of narrow alloyed silicon tunnel junctions of Chynoweth et. al. (8) with bulk doping as parameter. Note the qualitative agreement between the characteristics.

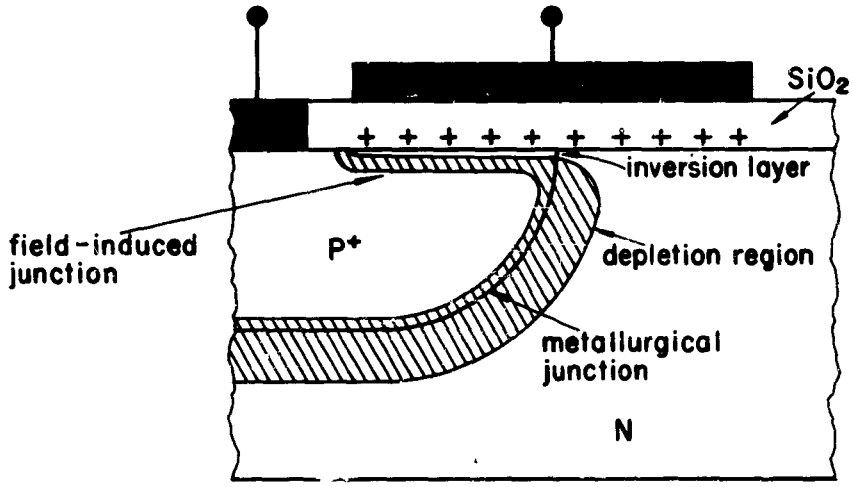


Figure 10

The field-induced junction in parallel with the metallurgical junction following the onset of a channel current associated with a P^+ region.

The characteristics of Figure 12 can be quantitatively compared by plotting the breakdown voltages of the junctions of Chynoweth et. al. (8) and the offset voltages of the channel current characteristics (both measured at 1μA) vs. doping concentration. Such a plot is shown in Figure 13. There is good quantitative agreement over three orders of magnitude of both voltage and doping concentration. This agreement indicates that it is breakdown of the field-induced junction that determines the low-current part of the channel characteristics.

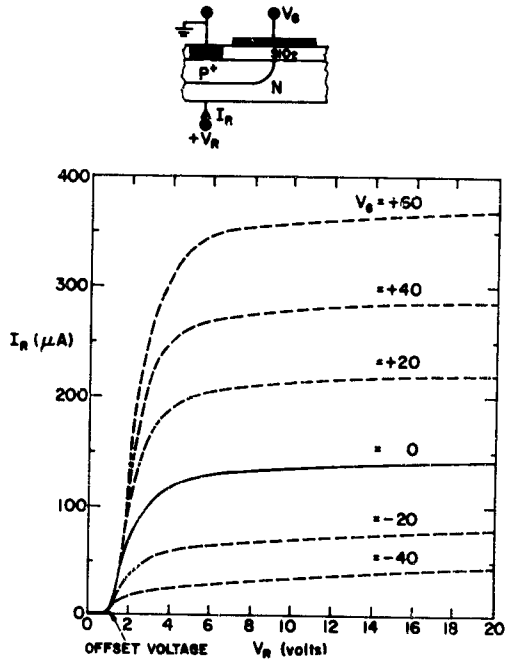


Figure 11
Channel current-reverse voltage characteristics as a function of gate voltage at T = -66°C for a device having a P^+ surface concentration of $4x10^{18}$ boron atoms/cm^3.

It would be expected for the low offset-voltage cases (~ 1 V and less) that the breakdown mechanism is tunneling. The fact that these voltages decrease with increasing temperature shows that this is indeed the case. Furthermore, the value of Q_s at the onset of channel current flow corresponds to a surface field $E_s \approx 10^6$ v/cm as can be seen from Figure 3. This value and the value of ~100 Å given above for the depletion region width of the field-induced junction are both of the correct magnitude for tunneling to occur. In the greater than 10 volt offset cases, the temperature dependence is opposite to that of the lower offset cases indicating that the breakdown mechanism is avalanche. These observations thus confirm that breakdown of the field-induced junction determines the low current part of channel characteristics associated with P^+ regions, i.e., the carrier generation mechanism involved is breakdown of the field-induced junction.

In the above controlled experiments care was taken to maintain a uniform density of Na+ ions at the oxide-silicon interface. However, in actual production devices, the concentration of positive charge within the oxide may vary radically from point to point. Such variations can in turn affect the channel characteristics as the following experiment demonstrates.

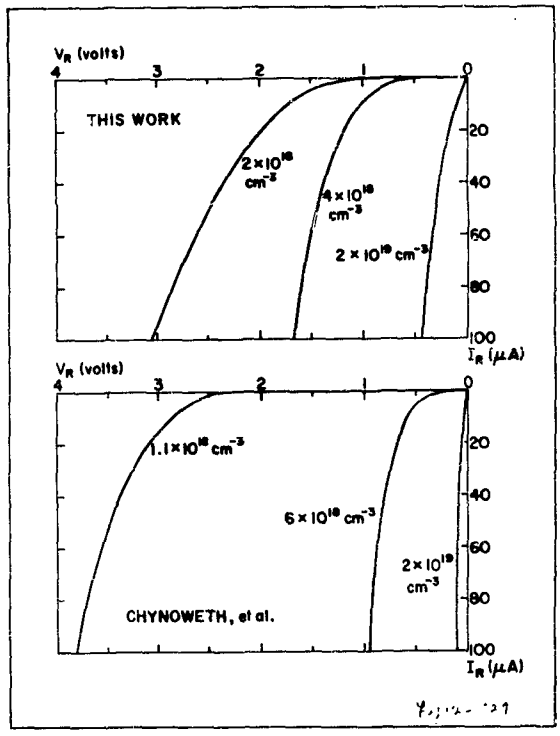


Figure 12

Comparison between the low current parts of channel characteristics and the reverse characteristics of the narrow alloyed silicon tunnel junctions of Chynoweth et. al. (8) at T = 25°C. The concentration values designate the estimated P+ surface concentrations and the substrate concentrations, respectively.

In Figure 14 the reverse characteristics of a gated device following channel current failure are shown dotted. The inset labeled "Before" schematically shows the originally uniformly distributed Na+ ions and the resulting inversion layer. The junction was then reverse biased well into avalanche for about a minute. The reverse characteristics after avalanching are shown solid. Note that a channel current did not flow until $V_R \approx 20v$. The explanation of this is indicated schematically in the "After" inset. Because of the localized heating of the oxide in the vicinity of the junction during avalanching and the action of the junction fringing field, the Na+ ions were rearranged to the configuration shown in this inset. Thus after avalanching, the inversion layer was isolated from the junction (9) until depletion region (1) spread far enough into the P^+ region to touch depletion region (2).

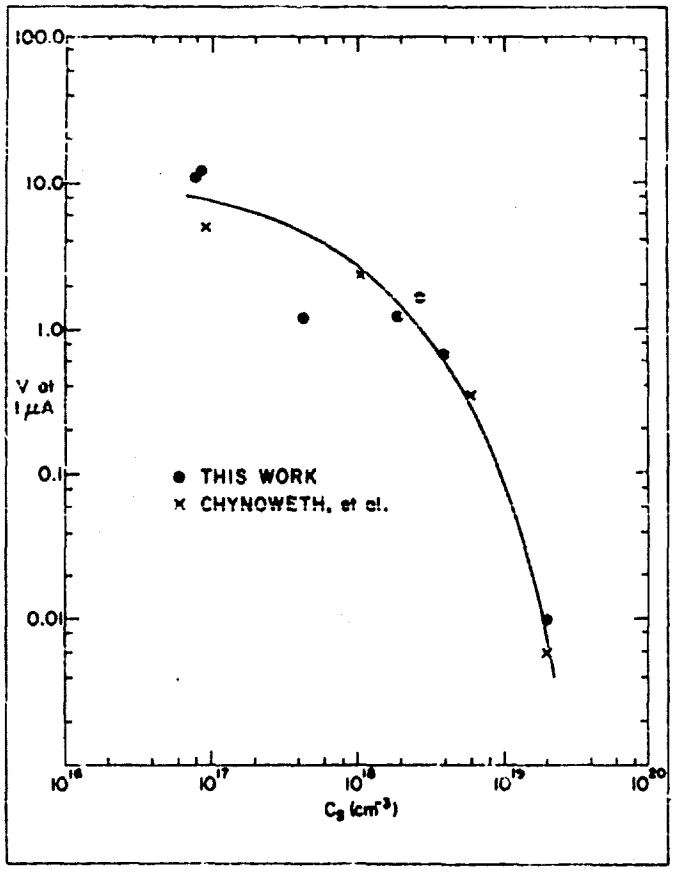


Figure 13

Offset voltages of channel characteristics and breakdown voltages of the narrow alloyed silicon junctions of Chynoweth et. al. (8) (both measured at 1 μA) as a function of P+ surface concentration and bulk concentration, respectively, at T= 25°C. Each point from this work represents a separate device run. Diode to diode variation in each device run was negligible.

The validity of the above model was demonstrated by measuring the zero-bias capacitance of the junction. Prior to driving Na+ ions down to the oxide-silicon interface the zero-bias capacitance of the device of Figure 14 was 9.0 pf; this value represents the capacitance of the metallurgical junction alone. Following the onset of channel current flow, the capacitance increased to 47.0 pf, the additional capacitance being that of the field-induced junction. After avalanching in the manner described, the capacitance returned to 8.3 pf indicating that the inversion layer was now isolated from the metallurgical junction at zero applied junction bias, corresponding to the picture in the "After" inset.

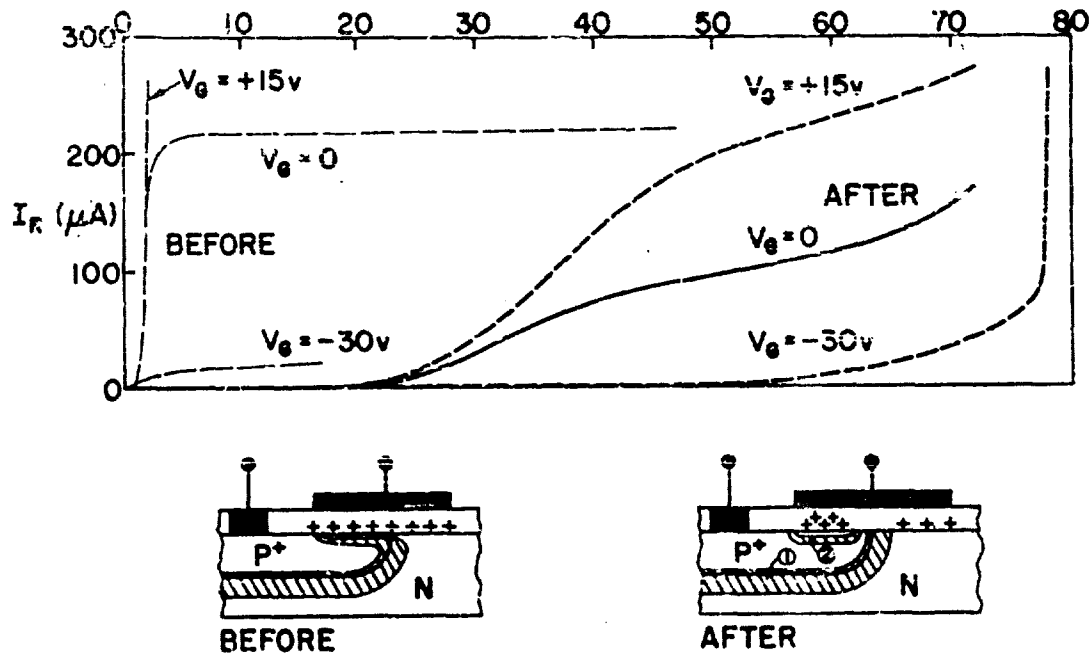


Figure 14

The effect of the fringing field of a junction on channel characteristics . The original channel characteristics are shown dotted; the characteristic after avalonching are shown solid.

6. CONCLUSION

It has been demonstrated by this work that the simple expression

CHANNEL CURRENT = INVERSION LAYER + CARRIER GENERATION SITE

provides the basis for understanding channel currents. Whereas the physical mechanisms which give rise to inversion layers have been discussed in detail in the literature , the mechanisms responsible for carrier generation have been all but neglected. It has been shown in this study that these mechanisms involve either (i) a fault or defect associated with the field-induced junction (the junction between the inversion layer and underlying silicon) or (ii) breakdown of the field-induced junction. Since the excess currents which can flow in metallurgical junctions also involve either a defect within, or breakdown of the junction, a complete analogy can be made between these two types of large excess currents. The difference in the shape of characteristics is due to the additional series resistance associated with the inversion layer along which the channel current must flow. This series resistance gives rise to the distinctive saturating characteristic of channel currents.

ACKNOWLEDGEMENT

The authors wish to thank E. H. Snow, J. W. Kelley and G. E. Moore for many stimulating discussions concerning this work.

REFERENCES

(1) A. Goetzberger, W. Shockley, "Metal Precipitates in Silicon P-N Junctions", J. Appl. Phys., Vol. 31, pp. 1821-1824, Oct. 1960.

(2) A. S. Grove, B. E. Deal, E. H. Snow, C. T. Sah, "Investigation of Thermally Oxidized Silicon Surfaces Using Metal-Oxide-Semiconductor Structures", Solid-State Electronics, Vol. 8, pp. 145-163, Feb. 1965.

(3) E. H. Snow, A. S. Grove, B. E. Deal, C. T. Sah, "Ion Transport Phenomena in Insulating Films", J. Appl. Phys., Vol. 36, pp. 1664-1673, May 1965.

(4) H. L. Hughes, R. R. Giroux, "Space Radiation Affects MOS FETS", Electronics, Vol 37, pp 58-60, Dec. 1964; A. G. Stanley, "Effects of Electron Irradiation on Metal-Oxide-Semiconductor Transistors", Proc IEEE, Vol 53, pp. 627-628, June 1965; E. Kooi, "Influence of X-Ray Irradiations on The Charge Distributions in Metal-Oxide-Silicon Structures", Philips Res. Repts, Vol 20, pp306-314, 1965; A. J. Speth and F. F. Fang,"Effects of Low-Energy Electron Irradiation on Si-Insulated Gate FETs", Appl. Phys. Letters, Vol 7, pp. 145-146, Sept. 1965.

(5) W. Shockley, W. W. Hooper, H. J. Queisser, W. Schroen, "Mobile Electric Charges on Insulating Oxides with Application to Oxide Covered Silicon P-N Junctions", Surface Science, Vol. 2, pp. 277-287, 1964.

(6) A. S. Grove, D. J. Fitzgerald, "The Origin of Channel Currents Associated with P^+ Regions", to appear in IEEE Transactions on Electron Devices, 1965; presented in part to the IEEE Solid State Device Research Conference, Princeton, New Jersey, June 1965.

(7) A. R. Bray, R. Lindner, "Stability of Reverse-Biased Thermally-Oxidized Silicon Junctions in Wet Ambients - Experiment and Measurements", Chapter 2 of Fourteenth Interim Technical Report, Engineering Services on Transistors, Bell Telephone Laboratories, Inc., 1958.

(8) A. G. Chynoweth, W. L. Feldmann, C. A. Lee, R. A. Logan, G. L. Pearson, P. Aigrain, "Internal Field Emission at Narrow Silicon and Germanium P-N Junctions", Phys. Rev., Vol. 118, pp. 425-434, April 1960.

(9) J. R. Mathews, W. A. Griffin, K. H. Olson, "Inversion of Oxidized Silicon Surfaces by Alkali Metals", Jour. Electrochem. Soc., Vol 112, pp 899-902, Sept. 1965.

EFFECT OF AMBIENT ON BREAKDOWN OF SILICON P-N JUNCTIONS

John F. Carroll

Rome Air Development Center
Griffiss Air Force Base, New York

Introduction

The occurrence of microplasma breakdown considerably below bulk breakdown in reliable devices has been considerably reduced through improved processing controls, screen tests, and burn-in tests. Field plates, guard rings, and design of junctions so that breakdown is in the bulk can, in many instances, be employed to prevent surface breakdown through design of the device. There still remains, however, those devices in which these design techniques cannot be employed or those in which these design techniques can themselves be the cause of high leakage or premature breakdown due to defects. It is the purpose of this paper to show that surface effects can significantly influence preferential breakdown near the surface.

Light emission at breakdown and its change with surface conditions can provide significant information on the junction characteristics and surface condition of planar silicon devices. Extensive studies have been made by various workers on field emission, avalanche breakdown and microplasma breakdown. Most studies have been concerned with the limiting situations of low voltage tunneling, uniform avalanche, large numbers of metal precipitates, etc. In silicon planar high reliability devices, deviation from ideal behavior is often caused by a relatively low number of defect sites and their interaction with surface fields.

The assumption made in analyzing data from experimental studies of junction and surface interactions and reliability test programs can be verified in many instances through observation of the light emission during reverse bias. The light emission from the junction is dependent on the distance of the source of light from the surface, type of defect, position of the defect with respect to the metallurgical junction, surface condition and the applied field.

Light emission near the surface appears white and as the breakdown increases in depth the color becomes deep red due to the absorption properties of the silicon. Light from near the surface can be observed at a leakage current in the order of 1 microampere through the breakdown region whereas light from a few microns below the surface required reverse currents in the milliamp range and are difficult to locate. Emission from greater than 3 microns will not be observed at reasonable current levels.

To locate the light emission from a small defect it is necessary to apply a voltage considerably greater than the turn-on voltage due to the small amount of light emission at turn-on when the plasma has a high resistance after breakdown. If the defect breakdown voltage is close to the bulk breakdown voltage and exhibits a high resistance characteristic it will not be observed since the applied voltage will be limited to the value of the bulk breakdown voltage. This is why even with a strong channel condition light emission from breakdown in a channel is seldom observed.

The limiting series resistance associated with various defects after breakdown is usually greater than 100k ohms for point defects and less than 25k ohms at the corner of a rectangular diffusion. In most commercial devices the resistance after breakdown will be less than 1000 ohms with an applied bias which is a few volts above breakdown.

Small point defects (less than 0.5 micron) which cause breakdowns with a high limiting resistance can have either a sharp breakdown or a soft breakdown. The characteristics appear to be governed by the type of defect, location with respect to the surface, surface condition and the field in the depletion layer at the defect. If it is in the bulk junction well below the surface, it will be unaffected by surface changes.

A strong channel in the low resistivity region will cause a low soft breakdown with white uniform light emission in the region where the metallurgical junction meets the surface. A strong channel on the high resistivity side of a deep diffusion will force the breakdown to be far below the surface with a high sharp breakdown. If there is a low voltage breakdown region in the channel, high leakage at low voltage or an offset type I-V characteristic will be observed. When sufficient voltage can be applied to the channel junction, light emission may be observed in the channel. The observation of light emission outside the depletion layer spread of the metallurgical junction is seldom observed due to the high series resistance of the channel.

The above described light emission provides a good picture of the junction at the surface in the extreme condition of strong accumulation or inversion. The low voltage and high current reverse characteristics provide sufficient information to assess the reliability of devices with these characteristics. Most devices will not be in either of these extreme conditions but rather in some intermediate state. The characteristics will be determined to a large degree by small defects, either where the junction meets the surface, or in the channel, since gross defects can be screened by visual inspection. If the defect is well below the surface, it will usually be stable and can be screened according to the required specification with a high degree of confidence.

A similar defect at the surface has the possibility of a changing current-voltage characteristic with time, depending on bias, temperature and ambient. The initial condition of the surface, particularly in the region where the junction meets the surface, will cause anomalous responses to stress conditions. For example, a defect may be active and the stress can cause it to become inactive, or the converse can be true, depending on location of the defect and initial surface condition. It is necessary therefore to understand the role of defects and their interaction with the surface to effectively design and interpret reliability tests of devices.

Experimental Observations

Diodes available were surveyed for an isolated premature breakdown with the following characteristics (1) near the surface (2) light emission at about ½ of bulk breakdown voltage of the junction (3) no significant other premature breakdown (4) high leakage at low voltage and (5) resistance after breakdown of greater than 50k ohms. A plasma on a large area (74 x 122 mils) p^+-n device with 1.0 ohms - cm material, 7 micron diffusion, and steam grown oxide was found which satisfied these criteria. A photograph of the plasma is shown in figure (1). The diameter of the plasma is about 3.2 microns and the P-region is on the left side. A 40x objective with an NA of .65, a 5x eyepiece, a 10 minute exposure and Polaroid 3000 film was used. The most intense region of the plasma is approximately 1.1 microns from the edge of the emission on the p region side. The plasma had a bluish center and was very bright at 58 volts. The original photograph was taken at 310X with 10 minute exposure.

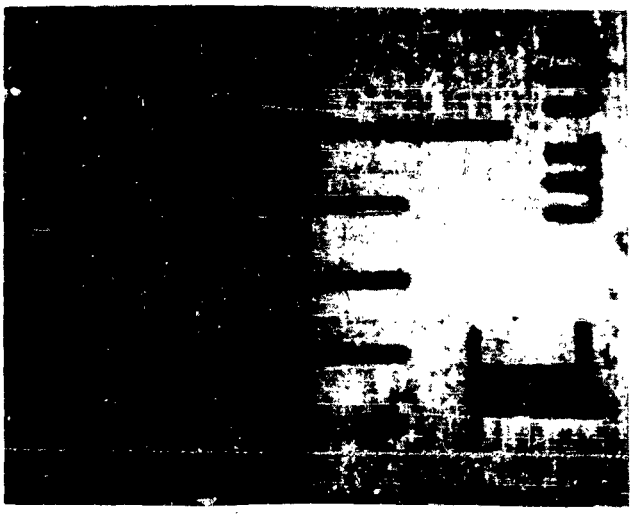
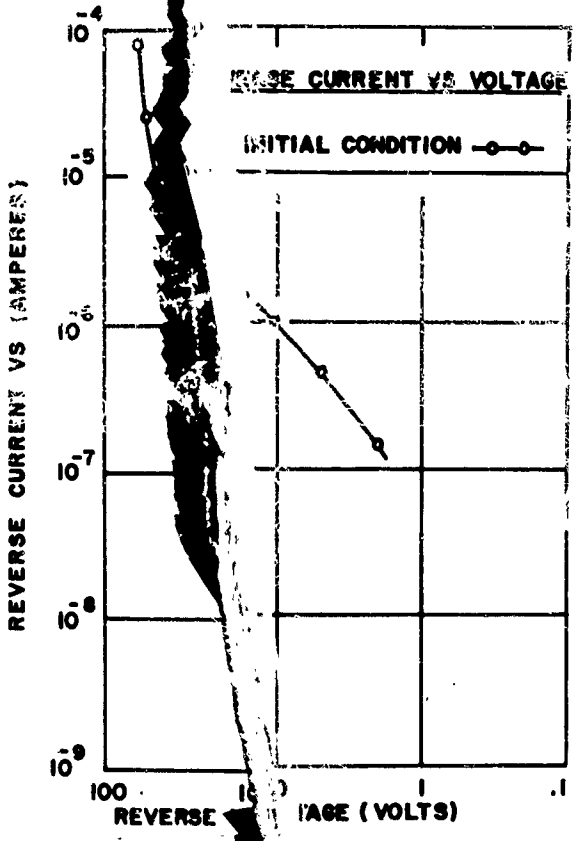


Figure 1

Photograph of plasma in initial condition.

The I-V curve fo… the device with the plasma in this condition is shown in Figure (2). … can be seen that the current tends to saturate at 5 volts and the slo… is resistive up to breakdown and that the breakdown is soft. Li… emission could be observed at 35 volts. The limiting resistance aft… breakdown was 50k ohms and no break or noise in the curve could be ob… ed on the curve tracer up to 58 volts. Careful optical exam…at… (400x) of the plasma site revealed no obvious mask defect, scratc… imperfection of the oxide in that area. The light emission was th… one distance from the step in the oxide as higher voltage breakdo… , indicating that it was not due to diffusion at a pin hole in the … lue or a large diffusion spike.



Fi… 2.

Current versus reverse voltage plo… diode with plasma in initial condi…

A plot of photo response versus voltage is shown in Figure (3). A 5 micron spot size, chopped light source and a Princeton Research Associated lock-in amplifier were used. The photo response was small and was consistent with measurements made on similar point sources of light emission at the surface of other devices. The plot is at the point of maximum response, which was at the point of initial high emission, or slightly on the p side of that point. The response fell off rapidly in all directions from the point of maximum response. The flat portion below breakdown shows that the size of the light spot is much larger than the breakdown region. No attempts were made to correct for impedance of the diode, light spot intensity and size, or ΔVB effects.

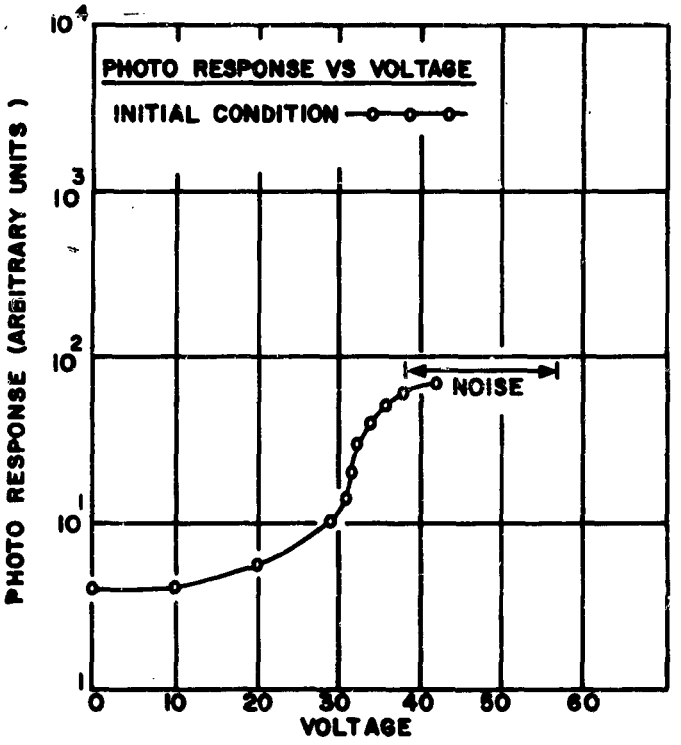


Figure 3.

Photoresponse versus voltage at plasma site.

In the process of making photo response measurements the plasma became dim and difficult to locate. The voltage at which light emission could be observed was about 42 volts. The photo response increased and the location of the maximum response moved from the site of the plasma to about 10 microns from the plasma in the direction normal to junction as shown in Figure (4).

After the position of maximum photo response was passed the response fell off slowly as the light spot moved away from the junction. Along lines perpendicular to the line through the plasma and the maximum photo response the photo response fell off rapidly with distance. The general characteristic of the maximum response versus voltage is shown in figure (5). It was established that moisture had been the cause of the change. Return towards the initial condition was observed when the leads were shorted. This was verified by increasing brightness of the plasma at a given voltage, increase in leakage, and change in photo response maximum and position.

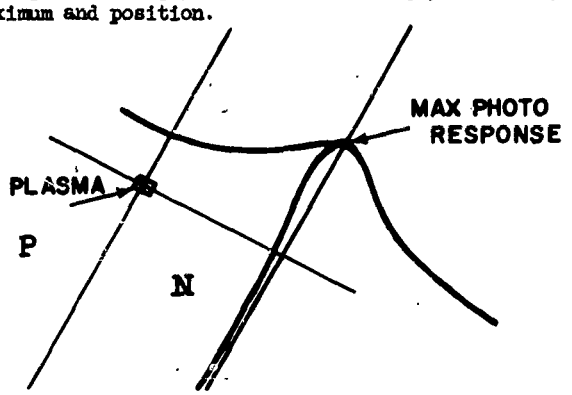


Figure 4.

General shape of photo response after reverse bias and water vapor. Maximum photo response has moved from the plasma cite and is localized 10 microns from p-n junction.

It was observed that when the plasma was in the dim condition, a diffuse red streamer appeared after breakdown of the plasma site in a direction normal to the junction. This is shown in Figure (6) where the plasma on the right is the light emission characteristic in the initial condition and the one on the left is the same plasma after drift caused by moisture on the surface. Both photograph are at 58 volts reverse bias. The distance from the intense area in the original photograph to the edge of light emission on the substrate side is about the same distance as the length from the center of the bright spot to the end of the streamer in the drifted condition which ia about 2 microns. Photographs were taken at various reverse voltages using time exposure to determine the effect of voltage on the streamer and the sensitivity of the film on the photographs.

Figure (7) is a composite of four photographs of the plasma at the following reverse voltages and time exposures: (1) 45 volts, 10 minutes (2) 50 volts, 10 minutes (3) 50 volts, 30 minutes and (4) 55 volts, 10 minutes. In the reproduction of the 45 volt photograph the exposure was three times that of the other three photographs so its relative intensity in the actual photograph was much less than it appears in the figure.

From the photographs at 50 volt reverse bias the length of the streamer has hardly changed while the width is significantly effected with an increase from 10 minutes exposure to 30 minutes exposure. It can be seen that at 55 volts reverse bias the width has increased.

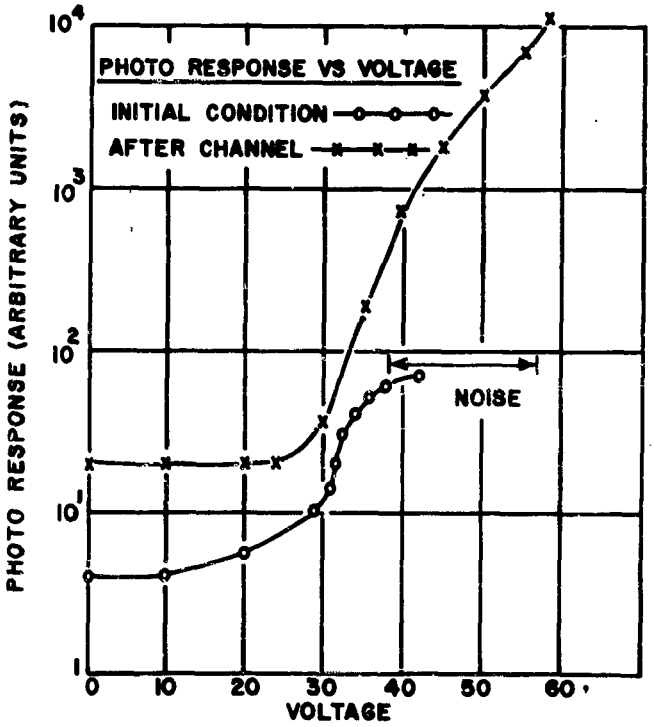


Figure 5.

Photoresponse versus voltage at plasma site showing change in maximum response due to reverse bias and water vapor.

With water vapor and reverse bias at 58 volts the leakage at breakdown was reduced from the 350 micro amp in the original condition to 75 micro amps. The resulting I-V plot is shown in Figure (8) with the original I-V plot. It is apparent that the low voltage leakage has decreased and the softness of the curve is revealed. Another diode from the same wafer had a leakage of 3 nanoamps at 10 volts and 40 nanoamps at 40 volts. Surface breakdown of the type described above was not observed on this device.

Figure 6.

Two photographs of same plasma. Photograph on right is the same as Figure 1 and the one on the left is after reverse bias and water vapor. Original photographs were taken at 310x and 58 volts reverse bias with a 10 minute exposure.

Figure 7.

Four photographs of the same plasma after reverse bias and water vapor. From the lower left and going clockwise the reverse voltages and time exposures are (1) 45 volts, 10 minutes (2) 50 volts, 10 minutes (3) 50 volts, 30 minutes and (4) 55 volts, 10 minutes. Original photographs were taken at 310x.

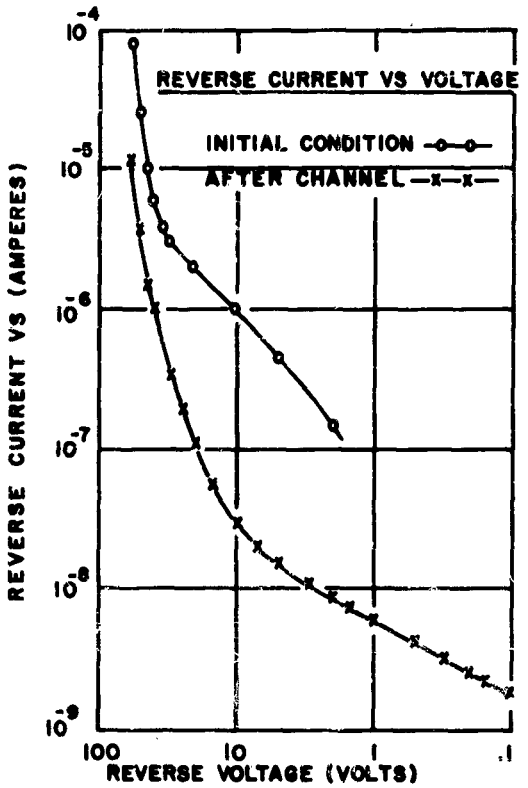


Figure 8.

Current versus reverse voltage plot of diode with plasma in initial condition and after reverse bias with water vapor.

Discussion

The model for the observed behavior is as follows:

a. An accumulation situation existed at the surface in the initial condition.

b. At the point where light emission appeared there was a defect at or near the surface of the silicon. The defect was probably a metal precipitate as evidenced by the soft character of the I-V curve and the deep diffusion.

c. When reverse bias was applied, breakdown with bright emission took place between the defect and the surface.

d. When the diode was exposed to high humidity and reverse bias, the accumulation was reduced, and a channel was formed to the edge of the chip due to the increased mobility of contaminant on the surface.

e. The breakdown now took place from the defect to the bulk depletion layer.

The depletion region in the accumulated condition is shown in Figure (9). At low voltage, field emission takes place from A to B with a high series resistance which does not allow the voltage to increase rapidly across the depletion region near the defect. From the diameter of the plasma a large photo response would be expected except for the high limiting resistance and the very narrow depletion region. A $\frac{1}{M}$ plot of the photo response data showed the breakdown is at 35 volts which is the voltage at which light was observed. The saturation of the photo response at 40 volts shows that the depletion region at the defect is completely broken down.

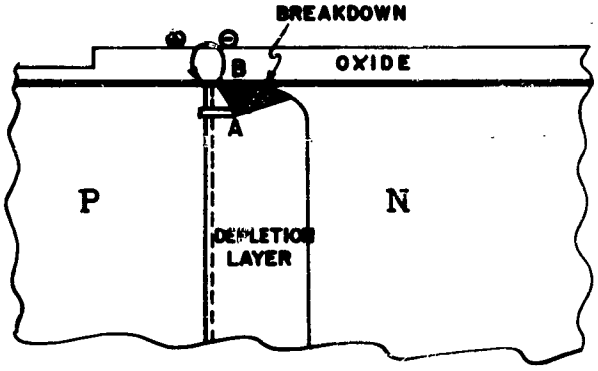


Figure 9.

Depletion layer at plasma site in accumulated condition.

The application of reverse bias and water vapor to the sample eliminated the accumulation condition at the plasma site and caused a channel condition. The change was probably due to increased mobility of contaminant on the surface since reversible changes could be produced with reverse biasing at breakdown and shorting of the terminals at room temperature.

The space charge region in the channel condition is shown in Figure (10). The breakdown no longer takes place from the defect to the surface, but rather is from the defect at A to the bulk at C. The change in light emission due to the change in surface conditions is consistent with the I-V curve after drift and the photoresponse. The leakage current at low voltage decreased, indicating much less field emission and therefore a wider space charge region at low voltage. This is confirmed by the increased photo response at low voltage. The $\frac{1}{M}$ plot of the data gave a breakdown voltage of 40 volts and light emission could be observed at 42 volts.

The softness of the I-V curve at breakdown and the high turn-on probability shows that field emission is still present. The increased depletion width however, resulted in a lower field for a given voltage. Photo response continues after breakdown due to the increase in field which provides a greater active volume in the space charge region. The increase in photo response to a maximum at 10 microns from the position of the initial turn-on of the plasma shows that the breakdown is limited by field region at the defect. Figure (11) depicts this situation when the light is at a distance from the plasma site. Electrons will move away from the metallurgical junction while holes will drift toward the junction. This changes the field in the transition region between the depletion layer of the junction and the channel such that a greater collection area for breakdown from the defect is provided.

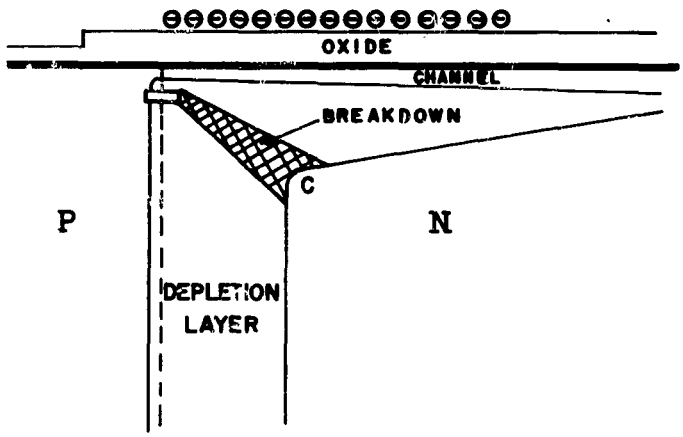


Figure 10.

Depletion layer at plasma site in channel condition.

Conclusion

It is concluded that small defects near the surface with high limiting resistance can cause significant change in the reverse current at both low voltage and near breakdown with a change in surface condition. They can be distinguished from large defects or normal breakdown of the junction by their high limiting resistance after breakdown. If their breakdown is near bulk breakdown or above, light emission will not be observed. Alternatively, if the defect is in a channel, it will not produce visible emission unless it is close to the metallurgical junction due to the series resistance of the channel.

Point defects in the region of the metallurgical junction provide a means of determining the field configuration on the basis of their light emission characteristics after breakdown when the surface condition is changed by the drift of ions in the oxide or charge motion on the surface.

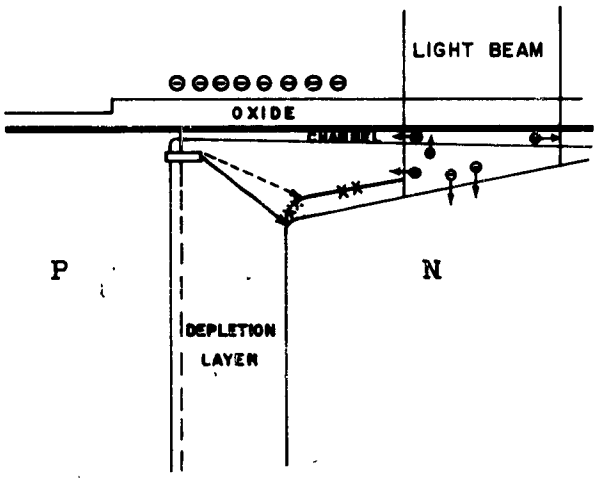


Figure 11.

Depletion layer at plasma site with channel and light at position of maximum photo response.

SURFACE LEAKAGE OF DIELECTRICS

L. Fedotowsky, P. Ho* and K. Lehovec

Research & Development Laboratories
Sprague Electric Company
North Adams, Massachusetts

Abstract

The surface leakage resistance, R_s, of silicon oxide, and its bulk leakage resistance, R_b, in humid ambients are determined from charging and discharging curves of MOS capacitors. Using these data we derive the extension of creep of surface charges from an electrode to which a d.c. voltage is applied. This creep which may induce breakdown phenomena in oxide-covered silicon microcircuits is retarded by reducing thickness and bulk resistance of the oxide.

Introduction

Creep of surface charges adjacent to a metal electrode on an insulator has been studied by Shockley et al.[1] by means of a vibrating reed electrometer. They analyzed the charge motion in terms of a distributed network consisting of the insulator surface resistance per square, R_s, and of the insulator capacitance per unit area

$$C = \epsilon\epsilon_o/L \qquad (1)$$

This paper describes the derivation of the surface resistance R_s, and the insulator bulk leakage resistance, R_b, from the time dependence of charging and discharging curves of MOS structures. These data will be used to predict the extension of spreading of surface charge from a charged electrode. It is well known that the field generated by a surface charge on silicon oxide may influence adversely junction properties in a microcircuit utilizing

* Now at Cornell University, Ithaca, New York

the underlying silicon substrate.[2][3] Thus the analysis of charge creep has some bearing on long range stability of microcircuits.

Mathematical Analysis

The conservation of charge on the oxide surface during creep leads to the distributed equivalent circuit of Fig. 1. Here R_s represents the surface resistance per square, R_b the bulk resistance per unit area and C the oxide capacitance per unit area. The left side termination of the circuit represents the boundary of an electrode on the oxide. The common ground is the silicon substrate. By applying a voltage V_o to this electrode for the time t, the following voltage distribution arises:

$$V(x,t) = V_o \exp\left[-(D\tau)^{-1/2}\right]\cdot \operatorname{erfc}\left[\frac{x}{2(Dt)^{1/2}} - \frac{t}{\tau}\right]^{1/2}$$

$$+ V_o \exp\left[+(D\tau)^{-1/2}\right]\cdot \operatorname{erfc}\left[\frac{x}{2(Dt)^{1/2}} + \frac{t}{\tau}\right]^{1/2} \qquad (2)$$

where

$$D = (R_s C)^{-1} \qquad (3)$$

and

$$\tau = R_b C \qquad (4)$$

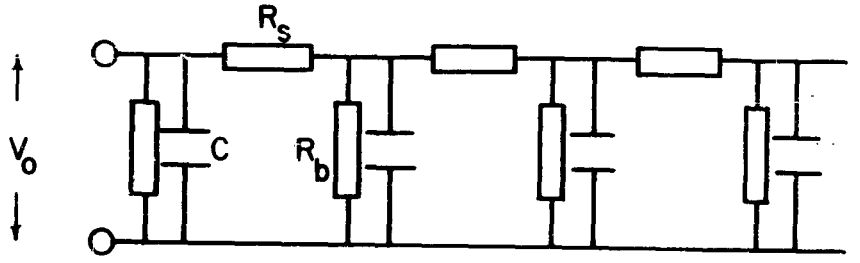


Figure 1

Equivalent circuit for the calculation of charging and discharging currents due to surface charge migration on the oxide

The current flowing into the network of Fig. 1 is the charging current

$$i_c = -(S/R_s)\cdot\partial V/\partial x]_{x=0} \qquad (5)$$

where S is the length of the rim of the electrode at x = 0. Inserting (2) into (5) one obtains the charging current as function of time

$$i_c = (SV_o/R_s)\cdot[(\pi Dt)^{-1/2}\cdot\exp(-t/\tau) + (D\tau)^{-1/2}\ \mathrm{erf}(t/\tau)^{1/2}] \qquad (6)$$

Thus the charging curve i_c vs. t can be analyzed in terms of R_s and R_b, after suitable correction for the contributions to the charging current from the capacitance under the electrode to the oxide.

After a charging time t_o, the electrode can be shorted to ground. The resulting discharge current from the oxide surface can be obtained by the superposition principle

$$i_d(t') = i_c(-V_o, t') + i_c(V_o,\ t = t' + t_o) \qquad (7)$$

where t' is the discharge time. Thus the analysis for R_s and R_b can also be based on the time dependence of the discharge current which has the advantage that corrections for the leakage current under the electrode are less important.

The equations (2) and (6) were based on a one-dimensional charge flow such as provided by a line contact, or else, a circular contact of radius r_o, large compared to the average extension of the charge distribution, i.e., assuming that

$$r_o \gg L_D = \sqrt{2\ t_o/R_sC} \qquad (8)$$

The surface charge flow for small circular contacts [$r_o \ll L_D$] can be expressed in terms of Bessel functions[4] in the case that $\tau = \infty$, and the transformation of Danckwertz[5] can be used to obtain the solution for finite τ.

Experimental

Charging and discharging measurements have been made on MOS structures of the two electrode configurations shown in the upper left of Fig. 2. In wet ambients (relative humidity of about 50% at room temperature) it was found that these currents are nearly in proportion to the rims of the electrodes indicating the dominance of surface effects.

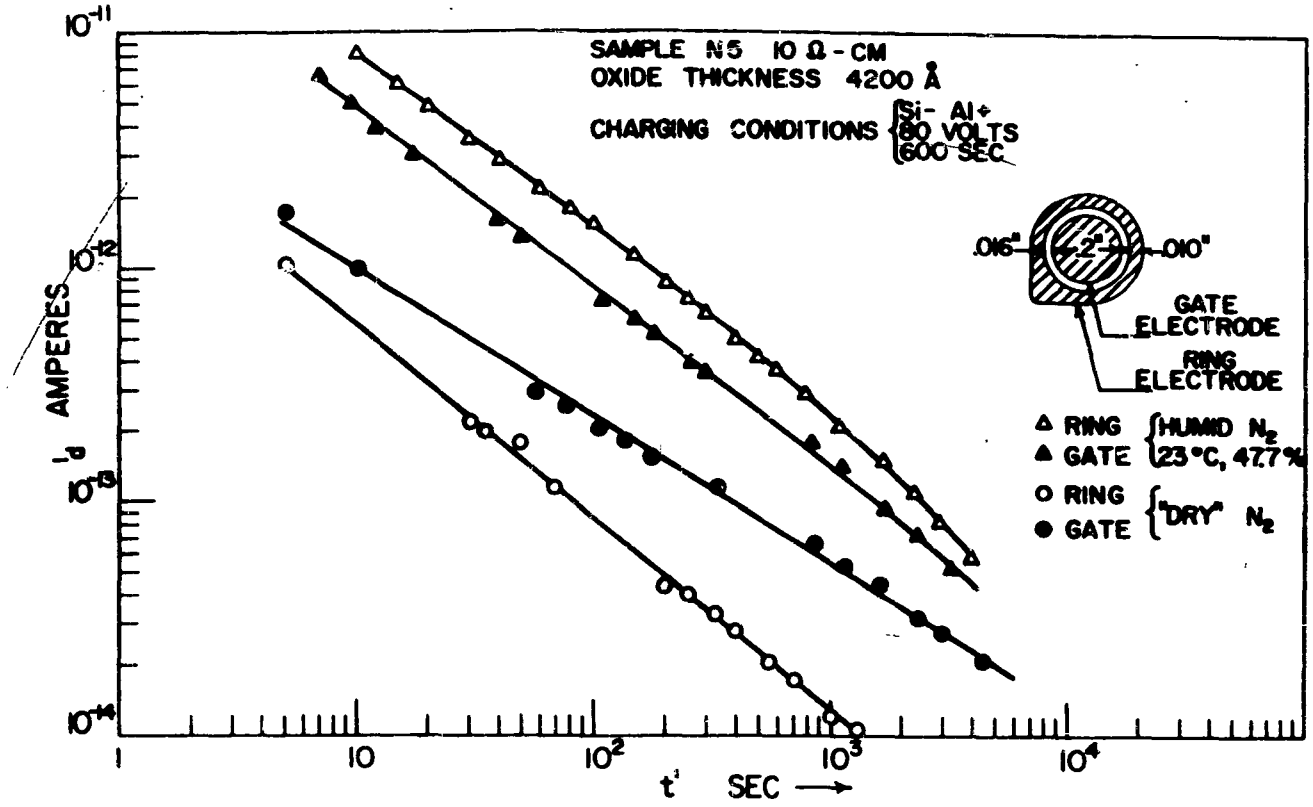


Figure 2
Discharge currents as functions of time for the two electrode configurations shown at the upper right and for a "dry" and "wet" ambient

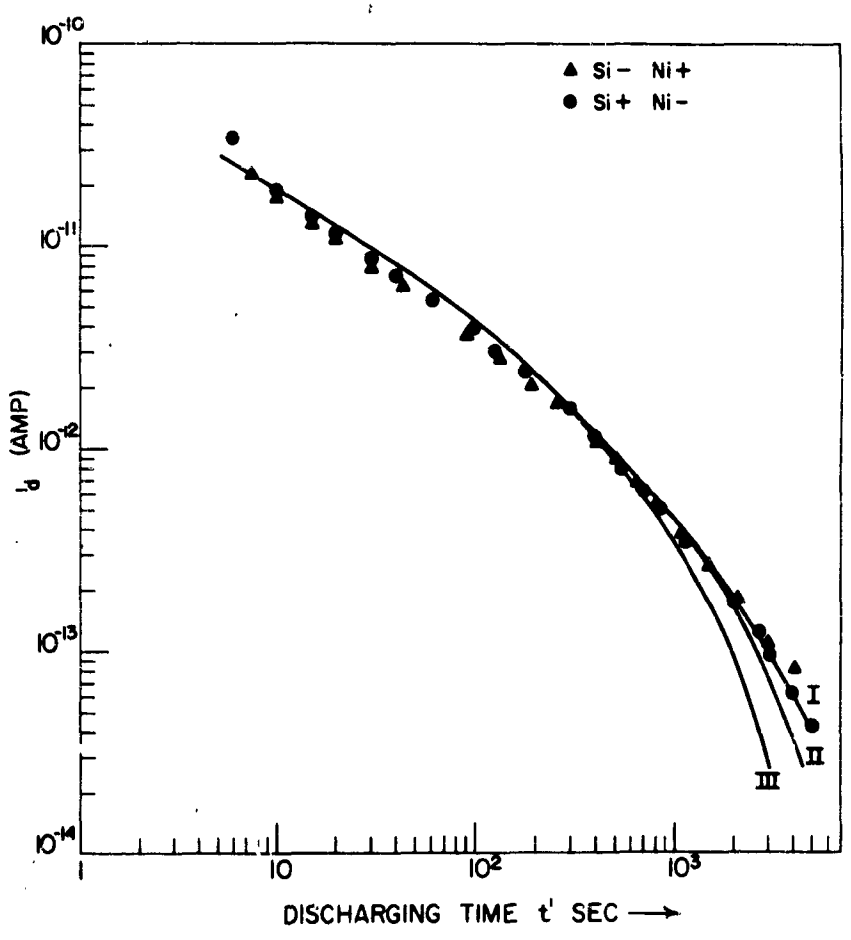


Figure 3

Comparisons of experimental discharging curves (corrected for bulk effect under the electrode) with theory (full lines) assuming $R_s = 4 \times 10^{16}$ ohms/square and I: $R_b = 2 \times 10^{12}$ ohms/cm^2; II: $R_b = 1.2 \times 10^{12}$ ohms/cm^2 and III: $R_b = 4 \times 10^{11}$ ohms/cm^2. N-sample 9 ohm-cm; oxide thickness 4150 Å; Ni-electrodes: charging at V_o = 80 volts for t_o = 600 sec.; 47.7% relative humidity at 23°C

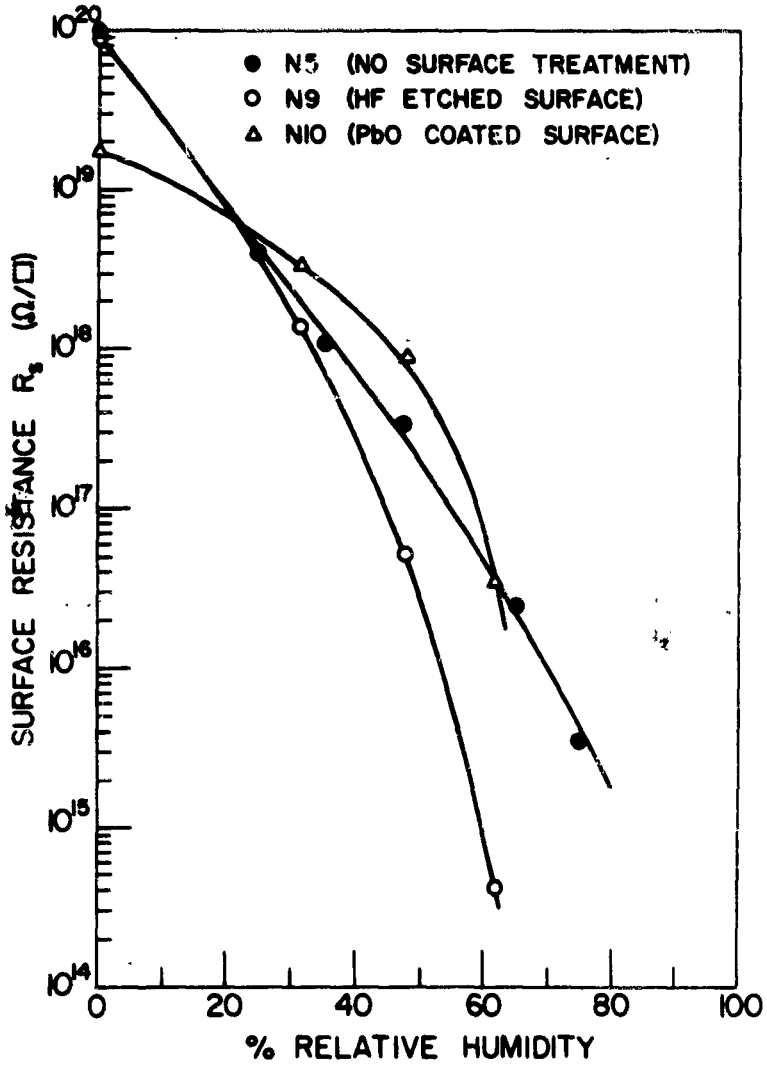


Figure 4

Surface resistance, R_s, as function of ambient humidity at room temperature for three oxide surfaces. Steam grown silicon oxides with no etch before Al-electrode application (·), an HF etch before Al-electrode application (o), and silicon oxide overlaid with 300 Å lead oxide (Δ)

Further experiments have shown that the charging and discharging currents are in proportion to the charging voltage which indicates that R_s and R_b in Fig. 1 are voltage independent quantities. Figure 3 shows the fitting of a discharge curve by Eq. (7), using the values of R_s and R_b indicated in the figure. Good fitting was always possible for metal electrodes which have been charged negatively, but not always for electrodes charged positively. The reason for this phenomenon is not known at present.

The theoretical prediction on the effect of charging time on discharge current $[i_d\, t_o^{1/2} = f(t'/t) \text{ if } R_b \approx \infty]$ is found to be valid. Figure 4 shows surface leakage resistances as functions of ambient humidity.

The values of R_b determined by our method should not be greatly affected by pinhole-shunts through the oxide, since such shunts would be in series with a large spreading resistance at the oxide surface.

Implication for Microcircuit Failure

Figure 5 shows an example for a prototype of a microcircuit failure resulting from charge migration across the overlaying oxide surface. The right hand contact in Fig. 5 is charged positively to provide a bias in the blocking direction to the junction between the underlying n-region and the p-type bulk of the silicon wafer. This p-n junction meets the $Si\text{-}SiO_2$ interface at the distance Λ from the rim of that contact. Creep of positive charge across the outer oxide surface from the contact over the distance Λ generates a field

$$\Delta F = V(\Lambda, t_o)/L \qquad (9)$$

in addition to whatever field already may exist at the silicon surface where the p-n junction meets the oxide. Such an additional field is known to have an adverse effect on the junction leakage current.[2] Eqs. (9) and (2) can be used to calculate the induced field as function of potential V_o, time of its application t_o, creep distance Λ, oxide thickness L, surface resistance R_s and bulk resistance R_b.

The field induced at the junction will be less for smaller oxide thicknesses, because the "diffusion constant" D of the charge creep is decreased [Eq. 3] and this effect over-compensates the increase of the proportionality constant, L^{-1} in Eq. (9). Furthermore, this induced field will be less if the oxide leakage resistance is decreased since charge leakage through the oxide detracts from the surface spreading of charge. Thus thinner oxides of larger bulk leakage are preferable for slowing down or preventing the adverse surface creep effect under discussion.

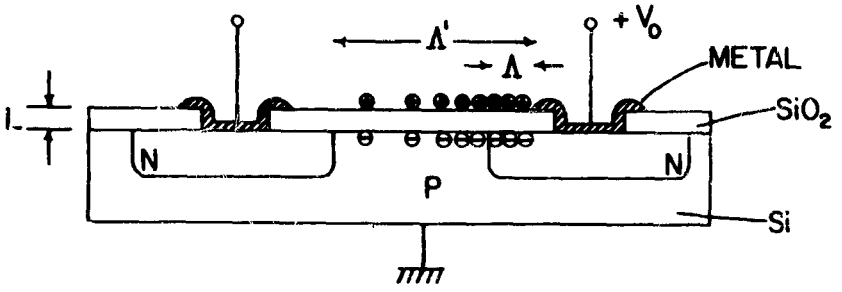


Figure 5

Cross section through a silicon based microcircuit. Creep of charge at the outer oxide surface from the positively biased contact over the distance Λ may affect the underlying p-n junction, and creep over the distance Λ' may "short" the two n-regions by an inversion layer

For a rough estimate of the time required for the junction failure visualized in Fig. 5, we may equate the distance Λ to a "diffusion length" (neglecting bulk oxide leakage) to obtain

$$t_o = \Lambda^2/2D = \Lambda^2 R_s \epsilon\epsilon_o/(2L) \qquad (10)$$

Considering an oxide thickness of $L \simeq 4000$ Å, a surface leakage resistance of $R_s \simeq 10^{18}$ ohms per square (as is typical for many types of silicon oxides under ordinary ambient conditions, see Fig. 4) and a distance $\Lambda \simeq 4$ mils $\simeq 10^{-2}$ cm, we obtain $D = 10^{-10}$ cm^2/sec, and $t_o \simeq 5$ days. Thus the phenomenon discussed here may contribute to "failure under load on life test". In a "dry" ambient, R_s values of 10^{19}-10^{20} ohms/square occur (Fig. 4) and the life time under load would then be extended to periods ranging from two months to two years, for the creep distance Λ = mils used above.

It is possible to treat cases of intermittent load either mathematically or using equivalent circuits (with τ-and R_sC-time constants scaled down conveniently). Note that the

"life time t_δ" of Eq. (10) increases with the square of the distance which must be travelled by the surface charge.

The left portion of Fig. 5 has been added to demonstrate a different failure type arising from charge creep in the outer oxide surface. If the surface charge extends over the distance Λ', an n-type inversion layer may be induced bridging the p-gap between the two n-regions and thus interconnecting them electrically. Before the inversion layer spreads across the entire width Λ', it may already significantly increase the capacitance of the p-n junction at the right of Fig. 5. This effect of a spreading surface charge occurs also in MOS-capacitors.[6]

It is known that silicon oxide and the silicon-silicon oxide interface usually contain an excess positive charge. Migration of a negative charge on the oxide surface would partially compensate the influence of this negative charge and may, therefore, improve underlying device properties, viz., the helpful effect of a negatively biased guard-ring over the oxide of a p-n junction.[2] Thus the failure mechanism visualized by us should be expected preferably or even exclusively near positively biased electrodes. Since surface charge migration is enhanced by large bias voltages (i.e., for p-n junctions biased in the blocking direction), the failure mechanism discussed by us should be encountered particularly with collector electrodes to n-regions of n-p-n transistors or for n diodes in a p-type substrate biased in the blocking direction.

References

1. Shockley, W., Hooper, W. W., Queisser, H. J., and Schroen, W., "Mobile Electric Charges on Insulating Oxides with Applications to Oxide-Covered Silicon p-n Junctions," Surface Science, 2, pp. 277-87 (1964).

2. Castrucci, P. P., and Logan, J. S., "Electrode Control of SiO_2-Passivated Planar Junctions," I.B.M. Jnl. of Research & Development, 8, pp. 394-399 (1964).

3. Metz, E. D., "Silicon Transistor Failure Mechanisms Caused by Surface Charge Separation," Physics of Failure in Electronics, 2, pp. 163-172 (1964).

4. Carslaw, H. S., and Jaeger, J. C., "Conduction of Heat in Solids," Oxford, p. 335 (1959).

5. Danckwertz, P. V., "Absorption by Simultaneous Diffusion and Chemical Reaction into Particles of Various Shapes and into Falling Drops," Trans. Faraday Society, 47, pp. 1014-1023 (1951).

6. Hofstein, S. R., Zaininger, K. H., and Warfield, G., "Frequency Response of the Surface Inversion Layer in Silicon," Proc. of the IEEE, 52, pp. 971-972 (1964).

A SIMPLE TECHNIQUE FOR THE DIRECT OBSERVATION OF TEMPERATURE DISTRIBUTION IN MICROELECTRONIC STRUCTURES

Donald W. Howarth

Guidance and Control Systems Division
Litton Systems, Incorporated
Woodland Hills, California

Introduction

The physical breakdown and failure of elements of electronic circuitry has long been related to localized power dissipation and resultant high temperatures. This is an increasingly serious problem as the state of the art in microelectronics tends towards smaller and smaller devices that are being packaged in tighter mechanical configurations.

In the course of the design and package development of a number of hybrid circuits, a simple technique for directly observing temperature distribution in microelectronic structures has been evolved. The technique is best suited to planar structures and consists of coating the surface of the device with a thin layer of a birefringent organic material which has a sharply defined, relatively low temperature melting point. Power is applied to the device in slowly increasing steps. As local temperatures at the surface of the device reach the melting point of the material coating the surface, a line which separates the liquid and solid phases of the material is easily observed through a microscope, using polarized illumination. This line traces an isothermal contour on the surface of the device. The line is sharply defined and changes position on the surface as heat dissipation in the device is changed.

The technique is simple to apply, provides direct observation of temperature as a function of power dissipation over an entire surface with extremely fine resolution, and provides information to allow heat flow and thermal impedances in the structure to be inferred.

Choice of Coating Material

Conduction is the principle mechanism of heat transfer in microelectronic structures since these structures are usually packaged in such a manner as to minimize convection heat transfer. Therefore, a coating material needs to be chosen whose melting point is near room temperature so that free convection heat transfer that is induced

by the temperature at the surface of the unpackaged microelectronic specimen is minimized. In addition, while the line which separates the liquid and crystalline phase in a thin layer of material is observable under proper conditions, it is observable to the point of being spectacular if the crystalline material is birefringent and is viewed under polarized light. Many materials are suited to the application of this technique, possessing the useful properties of low melting point and birefringence, being also chemically passive to the surfaces upon which they are to be coated.

Phenyl salicylate, sold commercially as salol, was the material chosen by us. It melts at 43°C which is usually less than 20°C above room temperature. It is non toxic and does not react with the materials upon which it is to be coated such as aluminum, nichrome, copper, gold, and silicon so that non destructive testing of the particular microelectronic structure can be accomplished. It is birefringent in its crystalline form, and supercools so that it is easily seeded in order to influence the path of crystal growth through the liquid layer. It is very soluble in ethanol, benzene, and acetone(1).

Coating Technique

The ultra clean surfaces of planar microelectronic structures are not easily wetted by phenyl salicylate. A simple technique was devised however which allows a uniform thin layer of this material to be coated on these surfaces. First a thin glass (microsheet) plate is cut to cover as large a region of the surface as is to be viewed. Phenyl salicylate is melted and a small quantity of this melt is picked up in a capillary pipette (made by drawing a fine tip on a 2 x 100 mm "melting point" capillary tube). The tip of the pipette is brought into contact with the interface between the specimen microcircuit and the glass cover plate. Surface tension draws the liquid phenyl salicylate from the pipette into this interface until the region under the plate is completely filled. Should there be insufficient material in the pipette to fill the entire region it may be refilled and re-introduced to the interface as many times as is necessary. Should there be an excess of material so that spillage occurs and the cover plate tends to float, the excess may be soaked up with a piece of analytical grade filter paper. In either event the phenyl salicylate will remain a liquid at temperatures below its melting point, due to the supercooling phenomenon, so that the region of interest can be satisfactorally filled with the liquid material. This liquid layer is then "seeded" from one corner or one edge of the cover plate with a minute quantity of crystalline material. Crystal growth procedes in a nearly radial pattern from the seed point through the layer.

Test specimens of planar microcircuits are coated at room temperature. There is no requirement for preheating of the surface to be coated so as to inhibit premature crystal formation. If the supercooled layer of phenyl salicylate is accidentally seeded during the introduction of more material, or while picking up excess material, a few seconds of exposure to a 250 watt infra red sun lamp at a distance of six inches will remelt the entire layer. It can then be immediately re-seeded. Crystal growth is relatively fast

(1) Handbook of Chemistry and Physics pp C191

through the supercooled layer of liquid, travelling across a 0.20 x 0.40 inch cover plate in a few minutes. The crystalline layer is transparent and is typically 0.005 - 0.015 mm thick. This coating technique is easily handled at low magnification (10X to 20X) under a well lighted stereo microscope as shown in figure 1.

Figure 1

Coating Surface of Microcircuit with Phenyl Salicylate

Handling of Specimens

Planar microelectronic structures on which the technique can best be applied, specifically thin film resistor networks on glass and ceramic substrates, screened resistors on ceramic substrates, and integrated circuits, are small and are fragile, particularly when some parts of their protective packaging have been stripped away.

Special care must be taken in handling these devices and their interconnect leads during the coating and viewing phases. Small boxes of molded plastic are easily modified to become ideal test fixtures for handling the specimens. The modification consists of drilling a series of holes in each end of the bottom section of the box, through which are brought as many conductors as are needed to energize the specimen microcircuit. The ends of these conductors are turned up in the box to form terminal posts to which the leads of the circuit are attached. The conductors extend through the end walls of the box, where they are fastened with an epoxy cement, and are fanned out for convenience in attaching power connections.

The fixtures shown in figure 2 are thin wall boxes of 1 1/4 x 2 1/8 x 1/2 inch dimensions having five 20 gauge copper wire conductors in each end. Specimen structures are supported below the plane of the box hinge by a pad of polyfoam and their leads are soldered to the upturned ends of the copper conductors. The specimens are mounted in the fixture prior to coating and remain there throughout the experiment. This fixture may be reused, or may become a permanent container for the coated specimen so that later experimentation can be carried on with a minimum of setup time.

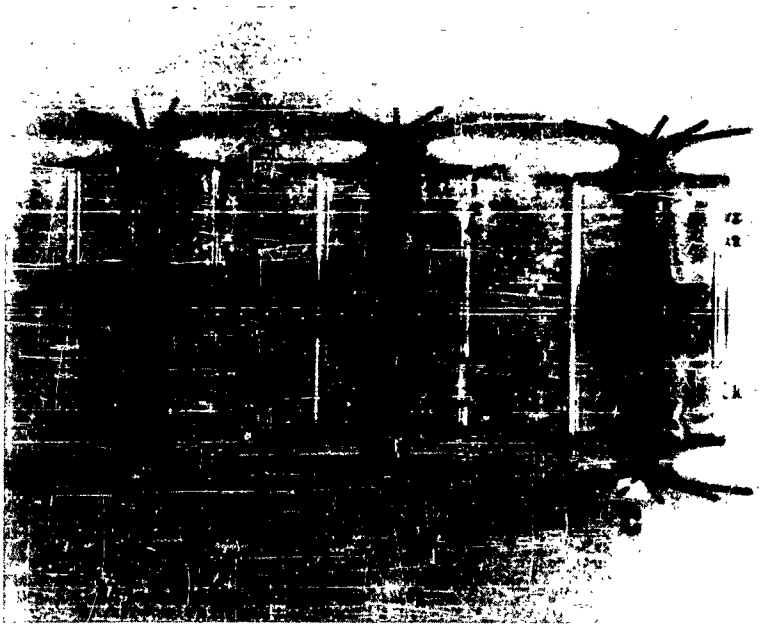


Figure 2

Test Fixture for Viewing Coated Microcircuit
(a) thin film resistor on alumina, (b) thin film resistor on glass (c) integrated circuit amplifier

Observation of the Specimens

The circuit element in the microelectronic structure which is to be observed, typically a resistor or group of resistors in a thin film or screened network or a resistor or transistor in an integrated circuit, is connected to a power supply and other suitable instrumentation to measure input voltage and current. Power is applied in slowly increasing steps while the circuit element, or the surface of the structure directly over this element, is viewed at low magnification under polarized light. This surface initially appears as an array of transparent multicolor crystalline material through which the detail of the circuit structure is seen.

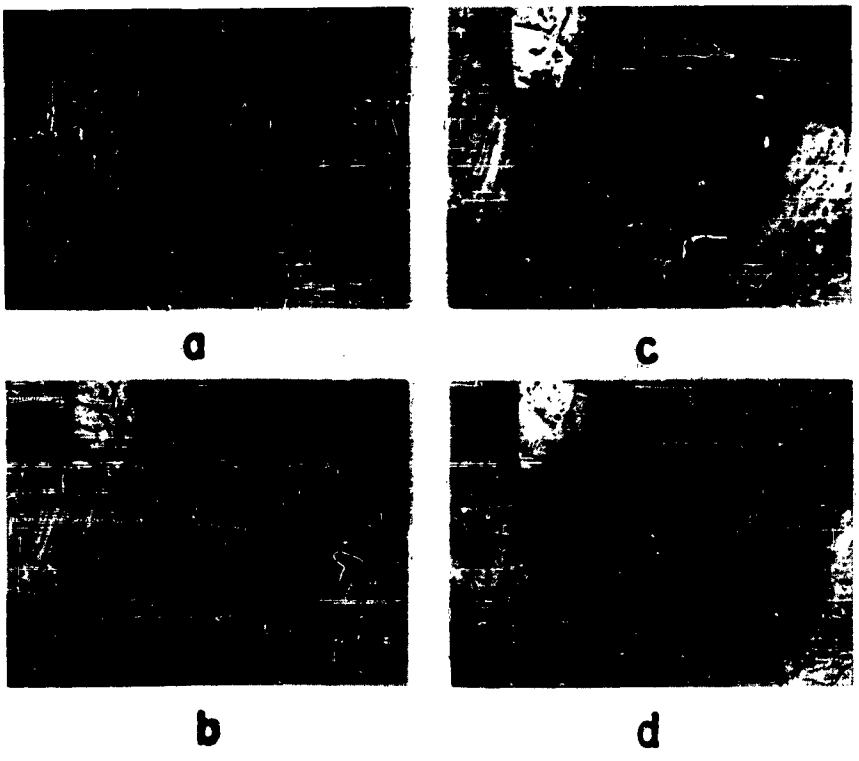


Figure 3

Photomicrographs of coated thin film resistor on glass substrate at various levels of power dissipation
(a) zero (b) 16.9 m watts (c) 20.1 m watts
(d) 26.5 m watts

As power is increased dissipation in the circuit element raises the local temperature to the melting point of the phenyl salicylate layer, and local melting in the layer at the surface of the structure is detected by an abrupt change in color of the crystalline material over that point. With further increase in power melting continues up through the layer with continuous changes in color over a larger area until the melting reaches the lower surface of the glass cover plate. At this point the melted material takes on a single unchanging color while the crystalline material surrounding it continues to change color as the melting progresses. These color patterns, superimposed on the surface of the microelectronic structure, delineate the 43°C isotherm for the particular level of power dissipation being observed as shown by the series of photomicrographs in figure 3. The process as described above seems to be dynamic, with constantly changing color. This however is not the case, particularly for thin film resistor networks on glass and alumina substrates. The color changes proceed almost in steps. The temperature distribution with each change in power level reaches equilibrium rapidly - in seconds - so that the color patterns are for all practical purposes static for each power level.

Application of Technique

The technique for coating planar microelectronic structures with a birefringent organic material and viewing, under polarized light, the changes in color pattern with changes in heat dissipation, provides us with a method of observing directly the temperature distribution in those structures. A set of 43°C isotherms may be drawn, one for each power level used during the observation, and from these contours heat flow and local thermal impedance may be inferred. The technique is applied here to four planar structures in order to demonstrate its usefulness.

The first example, seen in figure 3, is a series of photomicrographs of a thin film resistor on a glass substrate. The resistor material is nichrome, having a sheet resistivity of 250 ohms/square, and the conductor material is aluminum. The photo of 3b was taken at a dissipation of 16.9 m watts after local melting had begun, and shows a small solid color area in the center of the resistor, surrounded by several interference fringes, the outer most of which touches the long edges of the resistor and nearly reaches the aluminum conductors at the ends of the resistor. This outer band delineates the 43°C isotherm and represents a 17°C rise above room ambient (26°C). In figure 3c, at 20.1 m watts of dissipation, the outer band is outside the long edges of the resistor and is touching the aluminum conductor on the lower end. At the upper end however, the band follows the aluminum edge only a short way, bending back into the resistor area near the lower corner. This implies a local change in heat flow caused either by discontinuous contact of the nichrome film with the aluminum pad along that edge or by the influence of the gold ribbon lead which is seen to be bonded to the aluminum pad at the extreme right of the picture. Note the detail with which this local anomaly can be observed. Figure 3d, at 26.5 m watts of dissipation, shows the bands to be confined tightly around the melted region as this region moves out from the dissipating element. The expected

elliptical shape of the pattern is distorted along the lower right edge, again, influenced by the same local anomaly.

These pictures allow us to estimate gross thermal impedance of the structure surrounding the thin film resistor. The estimate is biased by the assumption that all heat flow is confined to the plane of the substrate, and that heat flow perpendicular to that plane arising from conduction or convection heat transfer through the surface is negligible. This is a fair assumption for free convection heat transfer when the temperature of the resistor is close to room ambient temperature. It is also a fair approximation of plastic encapsulated thin film resistor substrates in actual use, which can be seen by comparing the relative thermal conductivities of phenyl salicylate and (cover) glass to encapsulant materials such as diallyl pthalate or silicone molding compound.

Material	Thermal Conductivity watts cm^{-2} $^{o}C^{-1}$ cm	Reference
Alumina	.2 - .4	(1)
Diallyl Pthalate	.002 - .004	(3)
Glass	.006 - .008	(2)
Silicone Molding Compound	.001 - .002	(3)
Phenyl Salicylate	.0013	(4)
Silicon	1.2 - 1.4	(5)

Table 1

Thermal Conductivities of some materials associated with the application of this technique.

Dissipation, temperature, and thermal impedance are related to each other in our simple planar model by,

$$T_{mp} - T_a = Z_{th}\, q \qquad (1)$$

where T_{mp} = melting point temperature

T_a = ambient temperature

q = rate of heat flow (heat dissipation)

Z_{th} = thermal impedance from T_{mp} to T_a

Figure 4 superimposes the melting point isotherms for Figure 3 bcd on the region of the resistor network from which our specimen was taken. The values of thermal impedance shown in the figure were calculated from equation 1 using $(T_{mp} - T_a) = 17^{\circ}C$ and values of q as shown in 3 bcd. The large thermal impedances associated with the specific isotherms are easily understood when it is realized that the specimen circuit was well insulated from the ambient, both by long leads and by the plastic fixture itself, a condition that would be much improved in actual use by shorter leads and by the package being in direct contact with a printed circuit structure that would provide conduction paths. The local (differential) thermal impedance between isotherms, however is quite realistic, being biased only by the assumptions made in formulating the planar model.

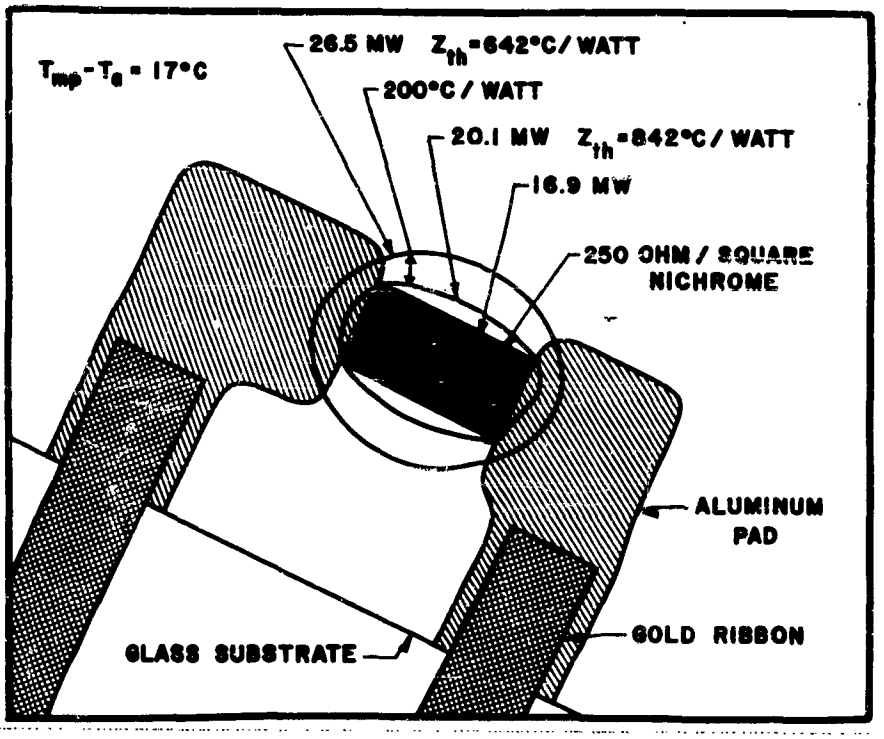


Figure 4

43° C Isotherms superimposed on thin film resistor of figure 3.

The second example is seen in figure 5, a series of photomicrographs of a thin film resistor on an alumina substrate. The resistor is 200 ohm/square nichrome and the conductor material is copper. This series is of interest in that it shows clearly the influence of the square corner on current distribution, which can be inferred directly from the color patterns. Figure 5b shows the beginnings of a hot spot on the inside of the corner. Figures 5cd show the necking down of current flow in the corner.

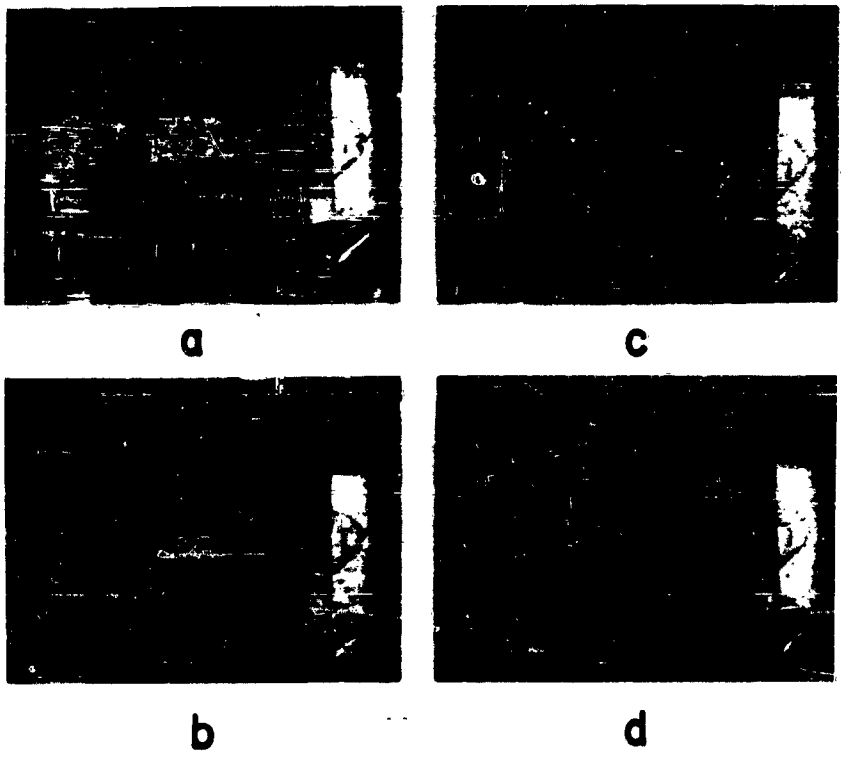


Figure 5

Photomicrographs of coated thin film resistor on alumina substrate at various levels of power dissipation
(a) zero (b) 51.3 m watts (c) 58.6 m watts (d) 75 m watts

In the third example, the series of photomicrographs of figure 6 look at the temperature distribution around a defect in a thin film resistor. The resistor is 200 ohm/square nichrome on an alumina substrate. The resistor value is nominally 22K ohms, the defect involving only a few per cent of the total resistor length. The three strips visible in the photos are all part of this resistor. The beginning of melting is seen in 6b as color bands which surround the narrowest point in the defective center strip. Continued melting around this defect and the beginnings of melting in the defect of the lower strip are seen in 6c. Note in 6d that the hot spot in the center strip has now caused melting in the upper strip.

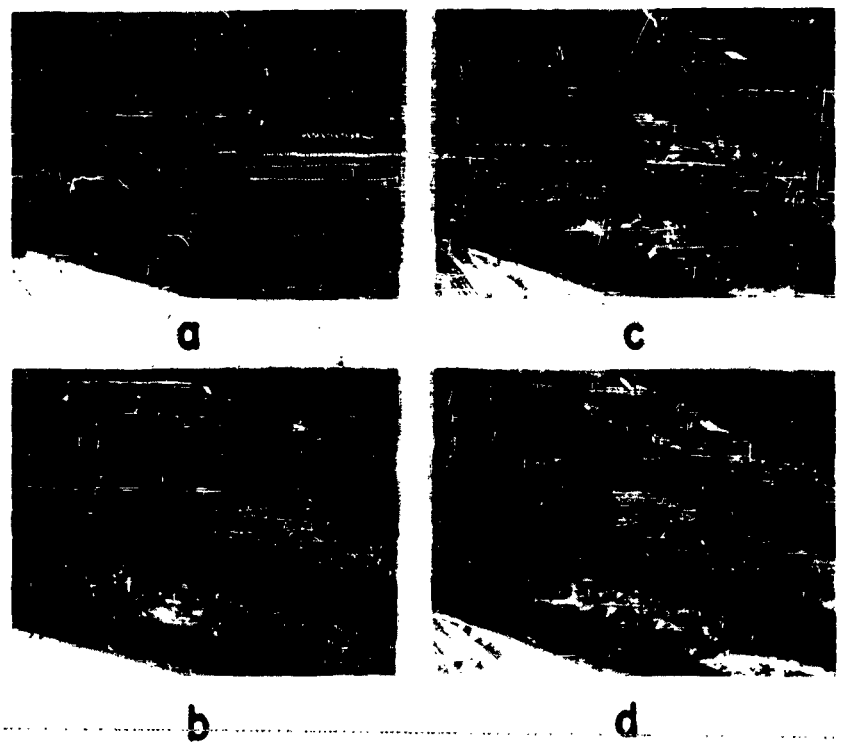


Figure 6

Photomicrographs of coated thin film resistor with process defect, at various levels of power dissipation
(a) zero (b) 25.5 m watts (c) 59.8 m watts (d) 64.2 m watts

A final example is taken from an integrated circuit amplifier. The integrated circuit, because of the high thermal conductivity of silicon around room temperature, proves to be the stickiest type of planar microcircuit upon which to use this technique. The silicon surface does not support much of a temperature gradient due to this high thermal conductivity so that the color pattern changes rapidly over a very narrow range of dissipation. However if care is taken to shield the surface of the specimen and its glass cover plate from air currents, and if current limiting precautions are taken in the power supply, a satisfactory set of isotherms can be obtained. The series of photos of figure 7 show three levels of dissipation in a diffused resistor, nominally 4.3 K ohms.

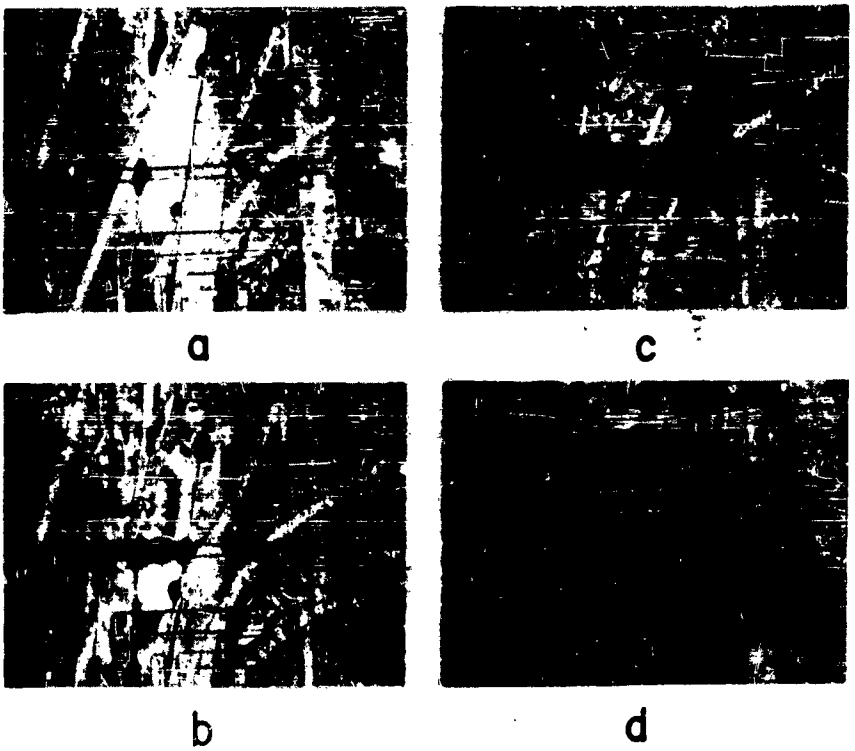


Figure 7

Photomicrographs of coated diffused resistor in integrated circuit amplifier at various levels of power dissipation (a) zero, (b) 64 m watts (c) 65.5 m watts (d) 66.7 m watts

The instrument used by us to observe and record these color changes is a Reichert "Zetopan POL" research polarizing microscope, set up for 35 mm photomicrography, as shown in figure 8. The test fixture is attached to a standard microscope slide with double back adhesive tape so that specimen may be manipulated with the mechanical stage of the microscope. A 100 watt zirconium arc lamp which is integral with the microscope frame provides incident (bright field) illumination of sufficient intensity for photographic exposures of 1/2 to 1 second to be made on ASA 100 speed film when both polarizing filters are in use. The range of magnification used is 40X to 100X.

Figure 8

Research polarizing microscope as setup for viewing and photographing coated microcircuit

Conclusions

This melting point technique has so far been used to evaluate temperature distribution and heat flow and to estimate thermal impedance in thin film resistor arrays and in integrated circuit structures. Design parameters such as line width and spacing have been studied to determine their effects on temperature distribution in thin film resistor substrates. It has been possible to compare evaporation through metal mask with photo etch processing of aluminum - nichrome thin film resistor arrays, both on glass and ceramic substrate materials. Conductor patterns and lead breakout have also been investigated in order to evaluate their influence on local heat flow at the boundaries of the package. The evaluation of heat flow through jumpers and package leads of mounted integrated circuit structures is seen to be a simple extension of the idea.

A number of applications to studies of device failure have been found. In thin film resistors, the thermal effects of irregularities in line width, caused by mask imperfections or substrate scratches, have been studied. In semiconductor devices, localized current flow and excessive power dissipation, caused by non-uniform junction properties of four-layer interaction, can be found and related to fabrication or design deficiencies.

Acknowledgments

The author wishes gratefully to acknowledge the many helpful comments and suggestions of Messr's F. Bartels, N. Grossman, and J. Mark of the Applied Research Laboratory of Guidance and Control Systems Division, Litton Systems, Inc., and the encouragement of its director, Dr. A. Stevenson, throughout the course of this work.

References

1. Handbook of Chemistry and Physics, 45th Ed. the Chemical Rubber Co., 1964, pp C191, D30.

2. American Institute of Physics Handbook, McGraw - Hill, 1957, pp 4/68.

3. Insulation Directory/Encyclopedia Issue, (May-June 1965) Lake Publishing Co., Libertyville, Illinois, pp 149-184.

4. Filippov, L. P., "Thermal Conductivity of Organic Liquids", Chem. Abstracts, Vol 56 pp 10936.

5. Neuberger, M., "Silicon Data Sheets", Electronic Properties Information Center, Hughes Aircraft Co., Culver City, Calif., DS-137, May 1964, pp 178-182.

SECTION V

BULK EFFECTS

THE ROLE OF MICRODEFECTS IN SILICON STARTING MATERIALS AS QUALITY REDUCING FACTORS IN SEMICONDUCTOR DEVICES

J. W. Faust, Jr., H. F. John, and R. Stickler

Westinghouse Research Laboratories
Westinghouse Electric Corporation
Pittsburgh, Pennsylvania

SUMMARY

Transmission electron microscope studies have shown that commercial silicon materials have varying concentrations of microdefects which are not detectable by normal evaluation procedures. The size of these defects ranges from less than 100Å up to microns. The following types of defects have been found: small loop-shaped defects, large loop-shaped defects, inclusions of several kinds, and precipitates including those associated with crystallographic defects. The chemical composition of these particles, precipitates, and inclusions are generally uncertain; but some evidence exists that SiC, SiO_2, and possibly Si_3N_4 are present. The widespread occurrence of microdefects in standard or premium quality silicon emphasizes the need for developing improved materials growth techniques and rapid procedures for detecting microdefects. The frequency of occurrence of the different varieties of microdefects differs with the type of material; however, crystals grown by the three principal techniques - float-zone, Czochralski, and crucibleless pulling - did not differ significantly in the degree of the overall microdefect problem.

The effects of microdefects on devices depends strongly on the nature and size of the defect, whether it is conducting (metallic), non-conducting, or purely crystallographic. The size of the defect is important in determining the behavior of the device under electrical stress and during long time operation. Indirect effects, in particular precipitation of fast-diffusing heavy metal impurities on microdefects, are believed to be a significant cause of poor device performance. Qualitative relationships between microdefects and the quality of p-n junction devices have been established. Further work to investigate more quantitatively the effects of specific microdefects on the performance of devices is necessary to establish acceptable concentration limits for microdefects.

INTRODUCTION

Almost from the beginning of modern semiconductor device technology, it was realized that the electrical properties of semiconductor materials and devices were strongly influenced by crystallographic and impurity inhomogeneities. Consequently, considerable research and engineering effort has been invested during the past twelve years or so in the improvement of silicon material quality with the results that silicon is readily available today which is free of such crystallographic defects as dislocations, stacking faults, twins, and growth strains. With some selection, material over a wide resistivity range can be obtained whose radial resistivity on a mm scale is uniform within $\pm 10\%$(1). Thus with an apparently reasonable quality of material assured, much of the research on failure problems in small area and power devices has been concerned with surface effects and defects introduced by processing, which are now and undoubtedly will continue to be major problems. The relative importance of materials and processing defects tends to vary with time, depending on the state of development of the respective technologies. At present it would appear that device technology is again becoming materials limited, particularly with respect to high voltage, power devices.

It is the theme of this communication that most commerical silicon crystals contain varying concentrations of sub-micron size defects which are generally not detectable in the normal evaluation procedures. These small inclusions, precipitates, segregation and associated crystallographic defects, referred to here as microdefects, can seriously affect the properties of materials and devices. At the present state of the silicon crystal art, they are probably more influential in determining device yields, performance and reliability than the material properties normally specified.

Most of the extended crystallographic defects (dislocations, stacking faults, etc.) and the gross impurity segregation effects (both of which are referred to here loosely for comparative purposes as macrodefects) have been eliminated and improvements in process technology are gradually minimizing the effects of defects introduced during fabrication. Thus, microdefects are a major barrier to the further improvement of p-n junction devices. One purpose of this communication is to urge silicon producers to direct additional research efforts to the elimination of microdefects and to set up quality standards for material based on concentration, and possibly size, of foreign particulate matter.

The evidence for the existence of micro-defects in silicon has been supplied mostly by transmission electron microscopy (TEM). The results of a survey of commercial starting materials using TEM and other techniques will be reported here, as well as some qualitative observations of the effects of microdefects on devices.

EXPERIMENTAL TECHNIQUES AND RESULTS

Twenty-five crystals of standard or premium quality were purchased from the four major domestic suppliers in the second half of 1964 especially for this survey of the microdefect problem. Corroborating data are available on a large number of other crystals that were examined in a less systematic manner, before and since that time.

This group of 25 crystals included material grown by the three principal techniques - float-zone, Czochralski, and crucibleless pulling. Both n- and p-type crystals with resistivities up to 200 ohm-cm were examined. In addition to examination by TEM for microdefects, the crystals were characterized with respect to uniformity of radial resistivity and presence of dislocations and other defects revealed by chemical etching.

The techniques described in detail by Booker and Stickler[2] were used in preparing the specimens for TEM. The essential feature of the specimen preparation is the formation of a thin section by chemical jet polishing for electron transmission. The position of the section to be examined can be varied within limits by polishing either from one side only or from both sides of the specimen, as shown in Figure 1. Usually five regions were thinned at different positions on the cross-section and examined to ensure that a representative view of the microdefect situation was obtained.

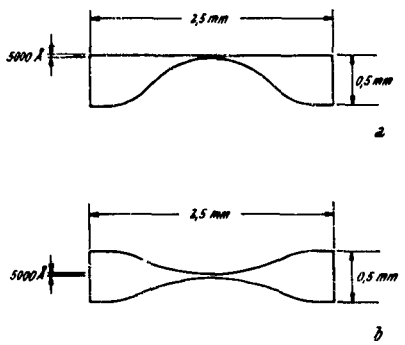


Figure 1.

Cross-section of a specimen for TEM examination:
(a) thinned from one side only
(b) thinned from both sides

Specimens were prepared for optical examination by a two-step procedure consisting of a 2 min. CP4* etch followed by a 3 min Sirtl** etch. This etching technique is very effective for bringing out dislocations and other defects, such as shallow, flat-bottom pits (referred to as "spots" in Table I), line networks, and hillocks. Some of these etching effects can be correlated qualitatively with microdefects as revealed by TEM and, therefore, may be useful for rapid evaluation of silicon materials. A discussion of these etching effects and their correlation with microdefects is not practical here, but will be covered in detail in another communication.[3]

* 120 cc 49% HF, 120 cc acetic acid, 200 cc HNO_3
** 100 g CrO_3, 200 cc H_2O, 200 cc 49% HF

Resistivity measurements were made on a mm scale using standard two-probe techniques[4] and on a micro-scale using special spreading resistance probe techniques.[5] The radial resistivity of silicon crystals when viewed on a mm scale often varies 50% or more, sometimes in a very irregular way, as shown later in Table I for representative samples. Fluctuations are even more noticeable when measurements are made at intervals of a few microns by the spreading resistance probe. Since junction properties, particularly peak voltage, are a function of base resistivity, the performance of devices and the ability to withstand electrical stresses are often seriously affected by non-uniform resistivity. This has been recognized as a problem by device manufacturers and selected material is available with a radial resistivity uniformity of $\pm$10%, as measured by standard two probe techniques. The fluctuations on a micro-scale may, however, be considerably greater than $\pm$10%. The problem of non-uniformity impurity distributions on a macro- and micro-scale is outside the main theme of this communication and will be treated in more detail elsewhere.[3] It should be emphasized, however, that impurity inhomogeneities are a limitation on device performance and that the establishment of standards of acceptability for material used in various types of devices is desirable.

Table I contains data on some representative crystals from the survey group, the data being obtained according to the procedures already described. Typical transmission electron micrographs are shown in Figure 2. Figure 2a shows an example of an inclusion, i.e., a particle so large that it must have been incorporated directly into the crystal from the melt during growth. An example of what is meant by "particles", as noted in Table I (particularly crystals 1, 3, 4, and 6), is shown in Figure 2b. These particles are probably formed by diffusion and precipitation processes. The third general type of microdefect is what is referred to here as a "loop" defect, as shown in Figures 2c and d. Two varieties of loop-shaped defects, distinguished by size, seem to be encountered. The large loop defects are very distinctive and are often the primary microdefect in crucibleless and Czochralski silicon, where it is frequently present in higher concentration in the outer few mm of the crystal.

Although there is a tendency for microdefects to be clustered in areas of a few mm^2 to $1cm^2$, it is not in general necessary to search for or select regions in the examination areas which have microdefects. Although it is not unusual to find 1 or 2 of the 5 examination regions relatively free of microdefects, only rarely are all five regions defect-free. Of the special group of 25 crystals, only three ingots had a microdefect concentration significantly lower than average. Fortuitiously, perhaps, one of each of the three growth types was represented.

Table I. Properties of Representative Commercial Silicon Crystals

	Growth Process	Type	Quoted Resistivity	Resistivity Uniformity*	Optical Examination	TEM Examination
1	float-zone	N	12-13 ohm-cm	10.5 ohm-cm ±20%	high conc. dislocations (8×10^4);"line networks"	high conc. particles and loops
2	float-zone	N	119-151 ohm-cm	140 ohm-cm near edge; center 112 ±5%	very high conc. of dislocations (1.2×10^5);slip; many "spots" near edge	rod-shaped particles or inclusions; sharply defined loops
3	float-zone	P	12 ohm-cm	10.5 ohm-cm ±10%	high conc. of dislocations; "line networks" present	high conc. of particles and loops
4	Czochralski	N	7.5-11.9 ohm-cm	12.8 ±.8 ohm-cm in center 20 mm (35mm crystal)	no dislocations; scattered hillocks and high conc. of "spots" near edge	particles and occasional massive inclusion
5	Czochralski	N	155-180 ohm-cm	100-300 ohm-cm in center	many tiny "spots"	high conc. of small and large loop defects
6	Czochralski	P	7-9.9 ohm-cm	10 ohm-cm ±15% but very irregular	med. conc. of dislocations (1.5×10^4); high conc. of "spots" near edge	some loop-type defects; in some areas a high number of precipitates
7	Crucibleless	N	100-200 ohm-cm	95-125 ohm-cm, higher at edge	no dislocations, but high conc. of spots in a band 2-4 mm around edge	extremely high conc. of large and small loop defects
8	Crucibleless	P	.45-.55 ohm-cm	.45 ohm-cm ±10% but very irregular	scattered, med. conc. dislocation (7.5×10^3); high. conc. of "spots"	many small and some large loop defects
9	Crucibleless	P	10-12 ohm-cm	9.8 ohm-cm ±5%	scattered dislocations in bands; many hillocks;"spots" high near edge	large particles or inclusions; many loop defects

* Two probe technique

Figure 2a.

Inclusion in Czochralski Si

Figure 2b.

Particles in Float Zone Si

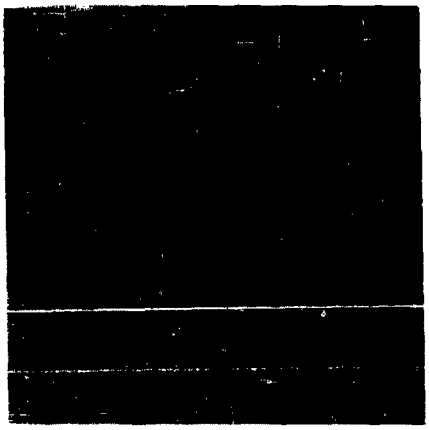


Figure 2c.

Small loops in Czochralski Si

Figure 2d.

Loops in crucibleless Si

The origin and chemical composition of these microdefects are in most cases uncertain. Inclusions are probably heterogeneous in composition and origin. Likely SiC, Al_2O_3, quartz fragments, metal carbides, and general crucible and furnace debris are among the species represented. In many cases the small particles or precipitates are probably SiC. The presence of distinctive "line network" etch effects (Optical Examination Table I) are apparently associated with a high concentration of particles. Carbon is often carried through the process used to prepare the polycrystalline Si which is used to grow single crystal material.[6] It appears that C or SiC may be in solid solution in the Si crystal and subsequently precipitate because of the decrease in solubility as the temperature of the crystal is decreased. In other cases, particualy Czochralski crystals, SiC may precipitate in the liquid near the interface and be incorporated directly into the crystal.

In some cases, the precipitates may be SiO_2. In general, however, the precipitation problems discussed here are not related to the precipitation and type conversion phenomena known to occur when silicon containing a high oxygen concentration is annealed.[7] Many low oxygen content crystals have a very high concentration of particles.

Even less is known about the nature of the loop-shaped defects. It would appear that they involve particulate matter, associated strain fields, and segregation effects; however, there is no positive evidence for crystallographic defects. The small shallow "spots" noted in the Optical Examination Column of Table I can apparently be correlated with these loop defects. As noted previously, the evidence for this will be discussed in detail elsewhere.[3]

Based on the overall results of the crystals examined, the relative frequency of occurrence of the three major types of microdefects is summarized in Table II.

Table II.

Frequency of Occurrence of Microdefects

	Inclusions	Particles	Loop-Shaped Defects
Float-zone	M-R	H	M
Czochralski	M	H	M
Crucibleless	M-R	M	H

H = high frequency, present in at least 1 examination region of perhaps 75% of crystals
M = medium frequency, present in at least 1 examination region of perhaps 50% of crystals
R = relatively rare, present in at least 1 examination region of perhaps 10% of crystals

The data presented in Tables I and II should not be interpreted as indicating that silicon grown by one technique is necessarily superior.

First, the number of ingots systematically evaluated has been too limited to draw any general conclusions. Second, the microdefect situation is bad in all types of material. Third, other factors not directly involving microdefects may determine the preferred type of material for a particular device, e.g., dislocation density, diameter, oxygen content, resistivity, resistivity uniformity, etc.

EFFECTS ON DEVICES

For purposes of a qualitative discussion of the effects on devices, it is convenient to consider three general categories of microdefects: non-conducting particles, conducting particles, and purely crystallographic defects, such as vacancy clusters, collapsed vacancy disks, and dislocation loops. Both conducting and non-conducting particles should be further divided according to size, since larger particles give more pronounced current channeling effects which are likely to lead to a rapid deterioration of IV characteristics under stressed conditions and under long time operation. We have somewhat arbitrarily chosen the dividing line between small and large particles, as far as electrical effects are concerned, to be about 2000Å for conducting particles and about 4000Å for non-conducting particles; but there are no compelling arguments for these choices.

A fairly extensive literature is available on the effects of poisoning of devices by fast diffusing impurities, precipitation at dislocations and stacking faults, and the effects of other inhomogeneities on device characteristics. Furthermore, the importance of inhomogeneities in determining the performance of devices is becoming better appreciated, as evidenced by the results of several investigations on second breakdown phenomena in silicon transistors. Perhaps the work most relevant to microdefect problems as discussed in the previous section is that of Shockley,[8] Kikuchi,[9][10] and Goetzberger.[11][12] Of particular interest is the proposal by Shockley that small insulating particles with a dielectric constant smaller than that of silicon will produce a field concentration in the vicinity of the particle and cause localized breakdown. For example, the field in the vicinity of a small SiO_2 particle was calculated to be a factor of 1.5 greater than at distance from the particle. This type of localized breakdown is often referred to as microplasma breakdown because of the discrete current pulses which are noted when current is measured as a function of voltage.

Emission of light accompanies the breakdown process and furnishes a very convenient way of detecting and establishing the position of these localized breakdown regions, particularly in specially designed test junctions. The highly localized type of breakdown gives an undesirable, "noisy" reverse IV characteristic. The type of precipitate shown in Figure 2b and the SiO_2 precipitates which arise during annealing of oxygen-containing Si are examples of the type of particle which should exhibit localized breakdown as predicted by Shockley. Larger particles, as shown in Figure 2a, probably exhibit more pronounced effects. The disturbance to the crystal is greater; consequently, the tendency to concentrate current in this region is more pronounced and failure under stress or deterioration with time is more likely.

Table III summarizes in a very qualitative way the possible effects of microdefects on p-n junction devices.

Table III

Effects on P-N Junctions of Various Types of Microdefects

Microdefect	Primary Electrical Effect	Implications for Stressed Conditions and Long Time Operation
Small non-conducting particles	lowered breakdown voltage near particle; microplasma effects with sharp current pulses	progressive deterioration of crystal in vicinity of particle with eventual channeling of current
Larger non-conducting particles	high reverse leakage, lowered breakdown voltage; current channeling; larger more noticeable effects than small particles	rapid deterioration of crystal in vicinity of particle; second breakdown in transistors; local destructive breakdown
Small conducting particles	high reverse leakage, soft breakdown	progressive deterioration of IV characteristics and eventual failure of device
Larger conducting particles	high reverse leakage; low breakdown voltage; current channeling in both reverse and forward directions	rapid deterioration of characteristics; hot spot formation; destructive second breakdown, current channeling
Small crystallographic defects	lowered minority carrier lifetime; high reverse leakage, non-uniform switching	non-uniform switching often leads to destruction under stressed conditions

Under many conditions the crystallographic changes and interactions of the microdefects with other impurity atoms which occur during device processing produce a disturbance which is more harmful than the original microdefect. The microdefects of the various types discussed here, as well as the more familiar edge dislocations and stacking-faults, can serve as sites for the precipitation of fast-diffusing impurities (Au, Cu, Fe, Mn, etc.) or a doping impurity which is being diffused. The effects of the fast-diffusing impurities will obviously be seen over a wider region than the slower-diffusing doping impurities.

An example of the interaction of gold with a loop-shaped defect of the type shown in Figure 2d is given in Figure 3. In this particular experiment, gold was diffused for 2 hrs at 1200°C and subsequently slowly cooled. Not only has gold precipitated on the defect, but the nature of the defect itself seems to have been altered. Precipitation of impurities at loop-shaped defects and the strained region around inert precipitates, such as SiC or SiO_2, is probably responsible for many adverse

electrical effects in devices.

Figure 3.

Gold precipitation at a loop defect

Most of our evidence for the effects of microdefects on yields and performance of devices is qualitative in nature and involves after-the-fact examination of devices and material in runs which have shown unexpectedly low yields or poor performance. It is not too unusual to find material which gives poor devices even though the crystal properties are within the normally acceptable ranges and where processing variables have been eliminated by including test slices from proven ingots. Under such circumstances, microdefects of the types described here have been found in a number of cases to be the probable cause of poor yields and/or performance. An example of a somewhat more definitive correlation is provided by the experiments of A. N. Knopp,[13] in which it was found that PNPN devices made on material having a high concentration of loop defects (Figure 2d) in the edge region of the wafers had poor switching and break-over characteristics. If, however, the size of the PNPN device were reduced to such an extent that the active area did not overlap the regions containing the defects, the IV characteristics were normal.

Attempts have been made to study more quantitatively the influence of microdefects on breakdown processes by observing microplasma phenomena in Goetzberger-type, small diffused guard ring p-n junction devices,[14] as described in this Symposium by Chu and Kannam[15] for epitaxial material. Such experiments have not yielded conclusive data. One of the factors which interferes with obtaining conclusive results is the presence of resistivity fluctuations. The development of techniques for isolation and study of the electrical effects of the various types of microdefects needs further investment of research

effort and would be expected to yield valuable results. This information is especially needed to help establish acceptable concentration limits for microdefects in material to be used for different types of devices.

CONCLUDING REMARKS

The magnitude of the microdefect problem in present day commercial silicon and the usefulness of transmission electron microscopy in studying microdefects have, we believe, been conclusively demonstrated. The TEM method is, however, relatively time-consuming. Therefore, X-ray diffraction microscopy, special etch techniques, and other rapid, but somewhat more qualitative, procedures should be worked out in detail. In devising and evaluating these techniques, it will be necessary to correlate the information with that obtained by TEM. The availability of quick, reliable evaluation techniques should provide the incentive for the improvement of standard growth techniques or for the adoption of new material growth techniques.

Further work is necessary to develop techniques to study the effects of specific microdefects on the electrical characteristics of devices. It would be helpful to materials producers and device fabricators to be able to specify maximum permissible concentrations of microdefects for different kinds of devices.

ACKNOWLEDGEMENT

The assistance of R. G. Mazur and R. J. Pfeil in furnishing resistivity data and the help of W. Hughes, Mrs. S. Hillbeck and Mrs. H. Larson in specimen preparation and examination are gratefully acknowledged. This work was supported by the Astrionics Laboratory, George C. Marshall Space Flight Center, under Contract NAS8-11432.

REFERENCES

1. See, for example, C. G. Currin and E. Earleywine, "Advances in Elemental Semiconductors", Semiconductor Products 7, 20-25 (1964).

2. G. R. Booker and R. Stickler, "Method of Thinning Semiconductor Specimens for Examination by Electron Microscopy", Br. J. Appl. Phys. 13, 446-448 (1962).

3. As reported in the final report for NASA Contract NAS8-11432, "A Study of the Relationship of Crystal Defects to Reliability of Epitaxial Silicon Planar Devices". A publication is in preparation covering these particular results.

4. E. Earleywine, L. P. Hilton, and D. Townley, "Measuring the Properties of Semiconductor Grade Materials", Semiconductor Products 8, 17-28 (1965).

5. R. G. Mazur and D. H. Dickey, "The Spreading Resistance Probe - A Semiconductor Resistivity Measuring Technique", Electrochem. Soc. Extended Abstracts of Electronics Div. 12, 148-150 (1963). Accepted for publication in J. Electrochem. Soc.

6. N. Schink, "Concerning the Carbon Content in Semiconductor Silicon", Solid-State Electronics 8, 767-769 (1965).

7. H. Reiss and W. Kaiser, "Solid State Reactions of Oxygen in Si", Properties of Elemental and Compound Semiconductors, Interscience Publishers, New York (1960), p. 103-119.

8. W. Shockley, "Problems Related to P-N Junctions in Silicon", Solid State Electronics 2, 35-67 (1961).

9. M. Kikuchi and K. Tachikawa, "Visible Light Emission and Microplasma Phenomena in Silicon P-N Junctions, I." J. Phys. Soc. Japan 15, 835-848 (1960).

10. M. Kikuchi, "Visible Light Emission and Microplasma Phenomena in Silicon p-n Junctions, II. Classification of Weak Spots in Diffused P-N Junctions," J. Phys. Soc. Japan 15, 1822-1831 (1960).

11. A. Goetzberger and W. Shockley, "Metal Precipitates in Silicon P-N Junctions," J. Appl. Phys. 31, 1821-1824 (1960).

12. A. Goetzberger and R. H. Finch, "Influence of Mechanical Damage on Avalanche Breakdown in Silicon P-N Junctions," Solid State Electronics 7, 543-545 (1964).

13. A. N. Knopp, private communication, to be published with R. Stickler as "Precipitates in Diffused-Alloyed Silicon Devices".

14. A. Goetzberger, B. McDonald, R. H. Haitz, and R. M. Scarlett, "Avalanche Effects in Silicon P-N Junctions," J. Appl. Phys. 34, 1591-1600 (1963).

15. T. L. Chu and P. Kannam, "Failure of Large Area Epitaxial Diffused Devices," this Symposium.

STRUCTURAL DEFECTS AND JUNCTION CHARACTERISTICS IN SILICON TRANSISTORS

E. D. Jungbluth

General Telephone and Electronics Laboratories, Inc.
Bayside, New York

P. Wang

Semiconductor Division
Sylvania Electric Products, Inc.
Woburn, Massachusetts

Introduction

In the past it had been assumed that dislocations do not seriously affect device performance because devices built from dislocation-free material were shown to perform no better than those built from less perfect material. This conclusion, however, neglected the fact that dislocations and other types of lattice defects may be introduced during the fabrication of these devices. For example, Lomer-Cottrell and/or stair-rod dislocations associated with stacking faults are nucleated in epitaxial layers.[1] Dislocations are generated within shallow surface diffused layers in order to relieve the stresses which result from concentration gradients.[2,3] Temperature gradients have pronounced effects on the segregation of impurities and may lead to plastic deformation. Moreover, strains may originate as a result of surface damage,[4] and built-in strains may occur as a consequence of solidification of impurity atoms within the crystal.[5]

This paper presents evidence relating crystal imperfections to processing treatments and demonstrates the influence of process-introduced dislocations on device yield in NPN silicon epitaxial planar transistors. To follow the generation of imperfections in device processing and subsequently to correlate device performance with the presence of imperfections, the non-destructive technique of x-ray diffraction microscopy[6,7] was used to record large-area topographs of whole crystals wafers at each stage in their fabrication. Knowledge of the locations of dislocations, in particular high densities of dislocations, permits one to probe the sample in specific areas in order to characterize the electrical properties of transistor devices coinciding with high defect versus low defect areas. In addition, this method at once establishes which particular processing step introduced dislocations or lattice strains.

The concomitant strain fields associated with lattice perturbations such as those due to dislocations are revealed as localized regions of enhanced x-ray intensities in the x-ray diffraction topographs. The visibility

of dislocations is dependent on the choice of reflecting crystallographic planes. The criteria for x-ray diffraction contrast are similar to those that apply in electron transmission microscopy. The advantages of this technique are:

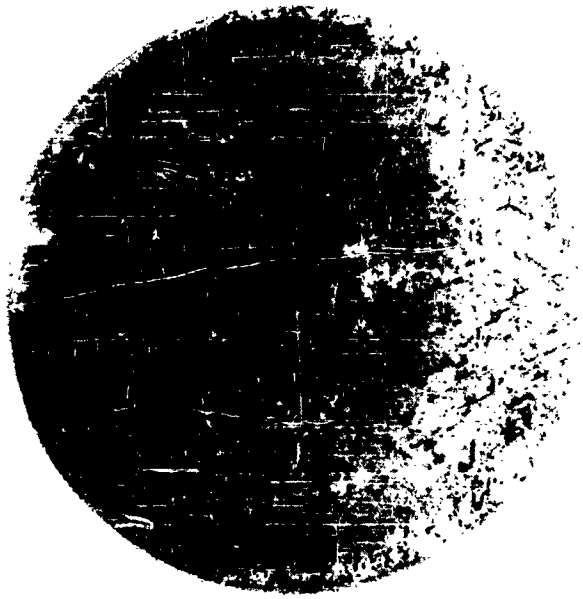
1. Testing in non-destructive.
2. Whole crystal wafers may be examined.
3. Silicon wafers may be examined after each step in the fabrication of devices, up to and including metallization.
4. A single topograph reveals the occurrence of bulk structural defects coincident with transistor areas.

Imperfections in the Starting Material

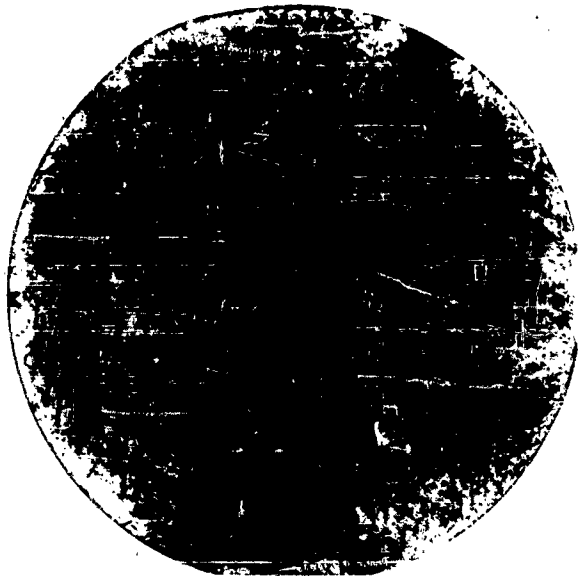
Dislocations in silicon are easily detected by a variety of techniques, but x-ray diffraction microscopy based on the primary extinction effect also detects the effects of segregation and precipitation. Figure 1(a) shows dislocation configurations representative for n-type substrate silicon upon which NPN planar transistors are made. This crystal was solution doped with arsenic to 10^{19} impurity atoms/cm^3. Arsenic substitution atoms are not expected to produce lattice dilation or compression consequently arsenic segregation is not detectable in the x-ray topographs. Segregation of boron in silicon may be detected by this method at doping levels as low as 5×10^{17} impurity atoms/cm^3. Undersized impurity atoms such as boron severely displace the lattice planes in solution-doped silicon as shown in Fig. 2(a). This is a topograph of a silicon wafer that was boron doped to 3×10^{20} atoms/cm^3. Subsequent growth of epitaxial layers upon such substrates may introduce high densities of dislocations, depending on thermal and concentration gradients. Epitaxial growths of doping level and type compatible with the substrate stress the crystal less than non-compatible growths. Figure 1(b) is a topograph of an arsenic-doped (10^{19} atoms per cm^3) epitaxial layer grown on a substrate similar to that shown in Fig. 1(a). Dislocations, located on octahedral (111) planes, are propagated from the substrate through the epitaxial layer. Trigonal symmetry of the dislocation configurations is evidence for slip due to thermal gradients experienced by the wafer in the epitaxial reaction chamber.

To accommodate the mismatch in lattice parameters at interfaces of varying dopant concentration, walls of dislocations aligned along crystallographic <110> directions may be generated. The dislocation configurations in Fig. 2(b) resulted after vapor growth of an arsenic-doped epitaxial layer on a highly boron-doped substrate similar to the wafer shown in Fig. 2(a). The dislocation reactions involved here are similar to those that result from high concentration gradients during boron or phosphorus diffusion in silicon.[8] The dislocation densities depend on the degree of misfit at the interface. Generally, when the doping levels and type between substrate and epitaxial layer are similar, the number of misfit dislocations are few. The high dislocation density in Fig. 2(b) results from a high concentration gradient at the interface.

Occurrence of stacking faults in epitaxial layers may be minimized by surface preparation and cleaning procedures. Stair-rod dislocations at bent stacking faults are known to be electrically active and contribute to localized breakdown.[9] Batsford and Thomas have shown that



(a)



(b)

Figure 1

(a) Arsenic-doped (10^{19} arsenic atoms/cm^3) substrate silicon of moderate dislocation density. (b) Epitaxial silicon growth on a silicon substrate similar to that in Fig. 1(a). Thermal gradients cause rows of dislocations aligned along <110> directions.

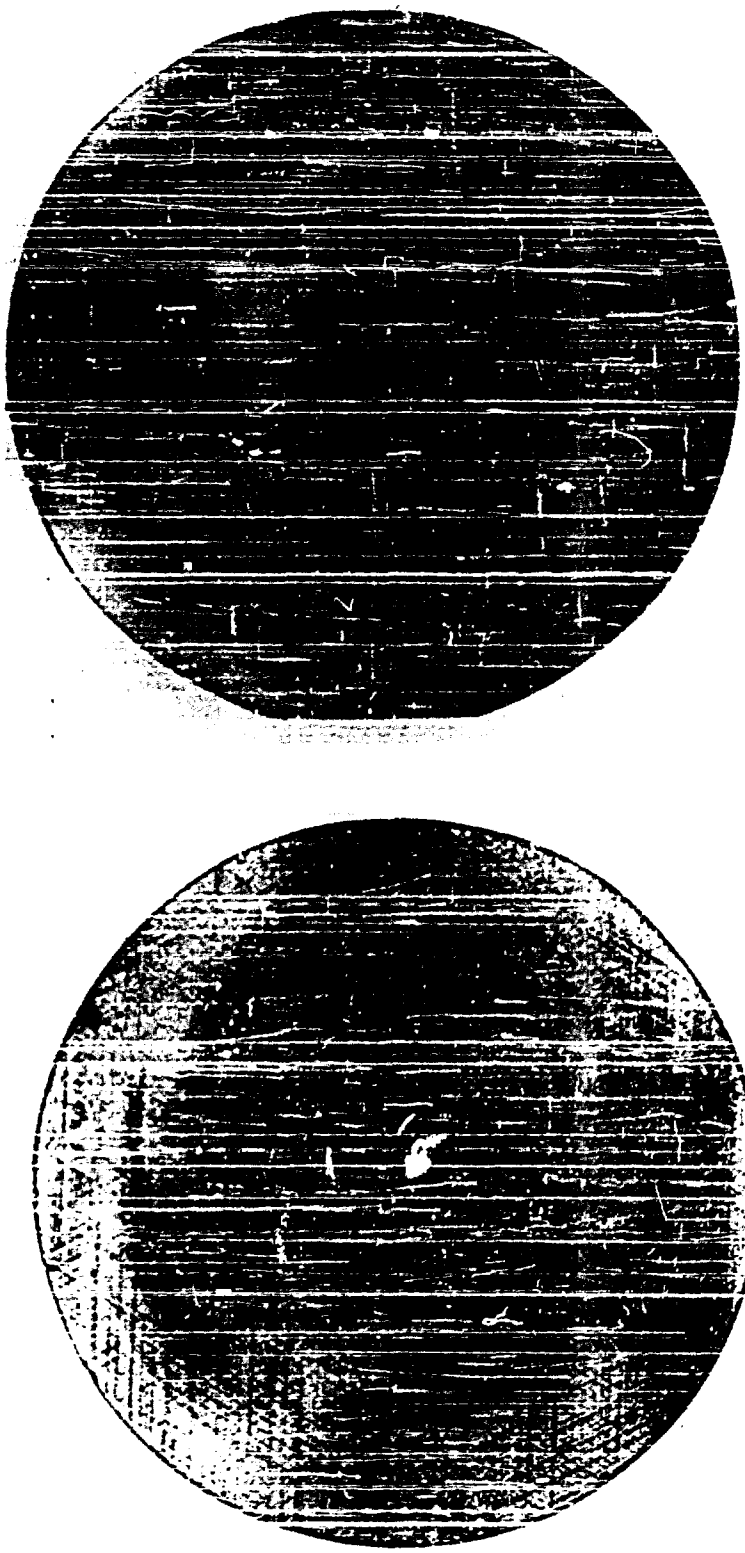


Figure 2

(a) Boron-doped (3 x 10^{20} boron atoms/cm^3) substrate silicon showing a high degree of inhomogeneous strain due to the dopant in a "dislocation-free" wafer. (b) Arsenic-doped (10^{19} arsenic atoms/cm^3) epitaxial silicon growth on a substrate similar to that in Fig. 2(a). <110> dislocation arrays relieve the stress at interfaces of varying dopant concentrations.

metallic impurities are perferentially precipitated on stacking faults.[10] Therefore the effects of stacking faults and their associated dislocations may considerably influence devices made on epitaxial layers. However, stacking fault densities in wafers made into transistor devices were less than 500/cm^2 in this study. Consequently, their influence on reverse breakdown voltage specifications was thought to be minimal in this investigation.

Correlation of Junction Properties and Imperfections

NPN epitaxial planar transistors may be formed on wafers similar to that shown in Figs. 1(a) and 1(b). The junctions are formed by diffusing through oxide masks produced on the wafer surface by standard KPR photoresist and etching techniques. The oxide windows give rise to their own diffraction contrast and are discussed in the next section. X-ray topographs (see Fig. 3) of defects in a silicon wafer after the base diffusion may be used to select areas of specific interest for other tests. The junctions in areas such as A, B and C may now be probed for reverse-breakdown-voltage characteristics. Areas A and B are essentially void of dislocations, while the dislocation density of area C is extremely high. The dislocation networks of area C are aligned in <110> directions and are formed as a result of thermal gradients during the epitaxial silicon growth process.

In terms of meeting reverse-breakdown-voltage specifications (the units were designed for 70-volt reverse breakdown), the junctions in areas A and B had a 75 percent and 56 percent yield respectively, whereas none of the junctions in area C had greater than 70-volt breakdowns. After the emitter diffusion, formed to achieve a 1 μ base width, the yields in areas A and B remain essentially unchanged; however, 10 of the 16 junctions in area C became shorted.

To explain the variations in the reverse breakdown voltages in devices built on the same wafer, several competing processes must be considered. Direct effects of dislocations, such as extra conductivity paths due to "dangling" bonds accepting electrons in n-type material, are probably minimal. However, interaction between dislocations and impurities is known to occur. Impurities of larger or smaller ionic radii fit more easily in regions of dilation or compression in the vicinity of a dislocation. Queisser, et al.[11] estimated that diffusion rates are 10^5 higher in diffusion channels afforded by dislocations along small-angle grain boundaries. As a consequence, selective diffusion and precipitation may occur in heavily dislocated regions. Emitter-collector breakthrough, such as occurs in the highly dislocated area C, may be initiated by diffusion enhancement, i.e., punch-through due to an enhanced diffusion front and a shorted base region due to diffusion pipes along dislocations.

Gettering action resulting from the P_2O_5 diffusion was evident in areas A and B. This suggests that precipitation may be a dominant process even in the absence of dislocations. Silicon phosphide precipitates, approximately 1 μ in the longest dimension, have been detected in diffused silicon by transmission electron microscopy.[12] Although the size of such precipitates precludes their detection by x-ray diffraction microscopy, it is reasonable to assume the existence of these or some other precipitate form in these wafers.

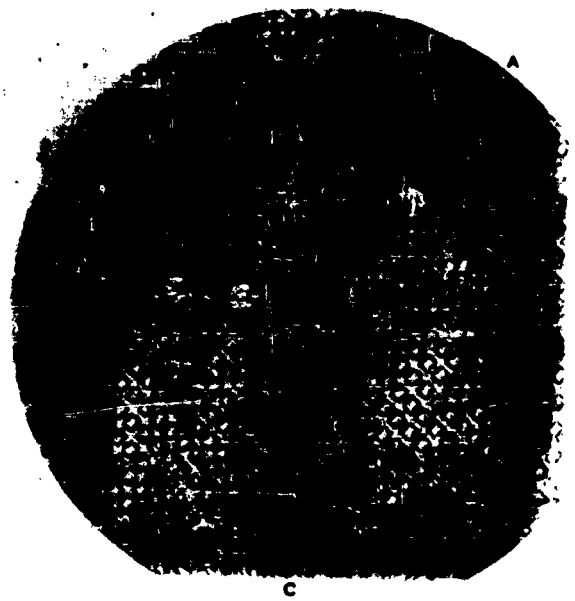


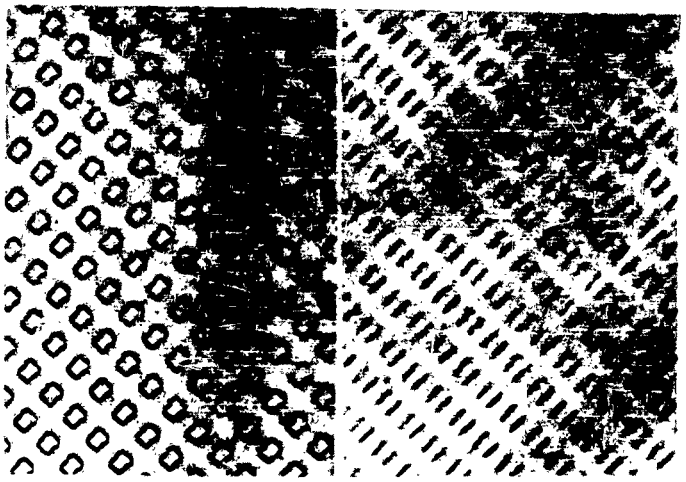
Figure 3

X-ray topograph of silicon wafer similar to Fig. 1 but after base boron diffusion. The areas outlined were selected for measurements of reverse breakdown voltages and correlated with the coincidence of structural defects.

Diffusion Enhancement at Oxide Window Peripheries

Thermal oxidation is commonly employed for passivation and masking. Defects generated due to such standard thermal treatments have not been detected by x-ray methods although recently Queisser and van Loom,[13] and Wilhelm and Joshi[14] have detected by etching techniques the development of stacking faults after steam oxidizing silicon. However, strains due to mismatch in thermal expansion coefficients at silicon-silicon oxide interfaces are evident in x-ray topographs. Figure 4 reveals considerable elastic strain associated with the window openings and the diffraction contrast is clearly dependent on the reflection chosen.[15] In Fig. 4(a), the diffraction vector is vertical while in Fig. 4(b) it is diagonal. Strain associated with grain boundary dislocations enhances diffusion. The migration of impurities, especially during diffusion cycles, toward areas of high strain such as exist at oxide window peripheries might be expected and depend on the nature of the diffusion process.

Emitter diffusions may be more influential in producing structural effects than base diffusions due to the higher surface concentrations and concentration gradients. Quenching after the base and emitter diffusions



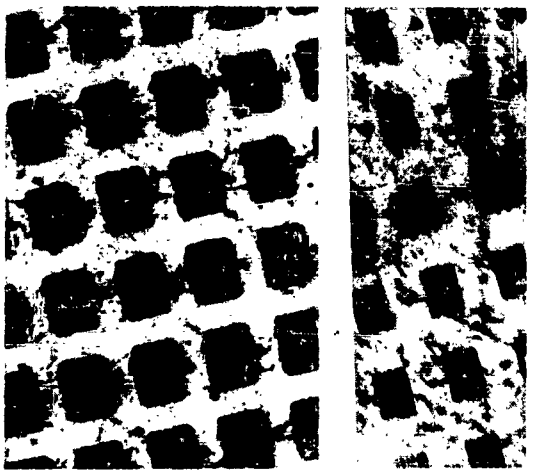
(a) (b)

Figure 4

a) Diffraction contrast due to strain at oxide window peripheries. The otal strain vector has a component in the direction of the diffraction ector which is vertical; consequently, the entire window periphery is isible. (b) The strain vector of the shorter side of the window opening as a component normal to the diffraction vector which is diagonal; consequently, the shorter side of the window periphery is invisible.

s the general practice in the fabrication of silicon high-speed switching ransistors. Joshi and Wilhelm have shown that misfit dislocations in hosphorus-diffused silicon extend in <112> and <110> directions in rder to relieve the stress due to solute lattice contraction of phosphorus.[16] Diffusion induced dislocations were not detected in x-ray opographs after either the base or emitter diffusions (see Fig. 5a). lowever, after removing the oxide layer, residual strain at the emitter inger periphery is present (see Fig. 5b) while no residual strain due to he base diffusion occurs. The rules for interpretation of x-ray diffraction topographs eliminate the possibility of extensive precipitation at he emitter finger periphery and suggest that the diffraction contrast, vident in Fig. 5(b), is due to diffusion enhancement at the emitter inger periphery.

Figure 6 is an optical micrograph of an etched transistor from he wafer displayed in Fig. 5(a). Diffusion-induced dislocations, ligned along <110> directions, are revealed only in the emitter areas. ateral propagation of these dislocations to surrounding regions is rohibited. Diffusion-induced dislocations occur in each emitter area f the same crystal wafer although the dislocation densities vary from mitter to emitter. This suggests that inhomogeneities in the host aterial produce local regions of stress which more easily induce slip



(a) (b)

Figure 5

(a) Enlarged (~60 X) x-ray topograph of the wafer shown in Fig. 3 after emitter diffusion. Note that both the base and emitter oxide window peripheries are visible. (b) Enlarged (~60 X) x-ray topograph showing residual strain due to diffusion enhancement at emitter finger peripheries.

in the presence of concentration gradients. The inhomogeneities in the host material may originate in the starting single crystal, or they may be due to subsequent processing treatments. For instance, in epitaxially grown films, it is known that stacking fault densities vary depending on surface cleanliness. There may be residual KPR in the photoresist process, and subsequently, there may be non-uniform oxide or pinholes in oxide. Dust particles may generate localized growth defects such as growth pyramids, or pinholes in oxide films. The pinholes may cause a non-uniform diffusion profile, defects may be introduced preferentially, and precipitates may form at some defects preferentially.

The sensitivity of x-ray diffraction microscopy is limited to the detection of strain fields that extend 5 to 80 μ around a dislocation depending on dislocation densities. Strain fields extending less than 5 μ would be invisible in the x-ray topographs. The inconsistent observations (see Fig. 5(a) and 6) involving dislocation reactions within emitter areas are explainable if these dislocations have short range (less than 5 μ) strain fields. Such dislocation reactions have been analyzed as due to two partials and a stair-rod dislocation bounding stacking faults. The extension of partial dislocations in distorted silicon has been measured as less than 100 Å. It seems plausible then to assume that an extended dislocation offers little x-ray contrast and consequently is invisible in the topograph.

Figure 6

Optical micrograph (500X) of an etched transistor area of the wafer shown in Fig. 5(a).

Process-Introduced Macroscopic Strains

Surface damage introduced by grinding or scratches strains the crystal so that severe warping and bending of thin wafer sections occur. Thin films of silicon oxide may also elastically bend the crystal so that portions of the crystal are misoriented relative to surrounding crystal areas. N-type (arsenic doped) 10 μ thick epitaxial layers deposited on n-type (arsenic doped) substrates may bend the lattice depending on the substrate surface preparation. Bending in this type of material occurred only when the substrate was mechanically polished.

Since arsenic atoms fit well into the silicon lattice, the bending must occur due to strain originating at damage on mechanically polished surfaces. Diffusion of doping impurities, especially selective area diffusion, likewise introduces bending or flexure in silicon wafers. The extent of this flexure can be measured by scanning crystal x-ray transmission techniques.[17] The flexure becomes more pronounced as the number of diffusion and oxidation steps increases. The strains associated with these flexures may be quite complicated due to multiple diffusions through complicated geometrical masks such as are required for integrated circuits. The ensuing warps or bends make it difficult to record large-area x-ray topographs due to the misorientation of the lattice in the wafers which are 2.5 cm diameter or greater. Large areas of the crystal, which is set for Laue reflection, will not reflect x-rays depending on the degree of elastic or built-in strain. Topographs of the

whole crystal may be made by compensating for the misorientation by superimposing an oscillating motion about the Bragg axis simultaneously with the scanning motion of the crystal. Figure 3 is representative of the quality of the x-ray images which may be obtained by application of this technique.[18] Externally produced stress has been shown by Bernard, et al.[19] to reduce breakdown voltages but no definitive tests have been made to ascertain its significance in our studies.

Conclusions

X-ray diffraction microscopy, due to its non-destructive nature, is unique in that it permits correlating process-introduced defects with electrical measurements. Our work has shown such correlation. Although one can locate the defects, many competing mechanisms contribute to the device failure. Many defects have been revealed by the x-ray technique. However complimentery techniques are always needed to support conclusions deducted from x-ray topographs. Further research efforts will attempt to define which defect is most detrimental and how generation of defects might be avoided.

References

1. G. R. Booker and R. Stickler, "Crystallographic Imperfections in Epitaxially Grown Silicon," J. Appl. Phys. 33, pp. 3281-3290 (1962).
2. S. Prussin, "Generation and Distribution of Dislocations by Solute Diffusion," J. Appl. Phys. 32, pp. 1876-1881 (1961).
3. H. J. Queisser, "Slip Patterns on Boron-Doped Silicon Surfaces," J. Appl. Phys. 32, pp. 1776-1780 (1961).
4. E. D. Jungbluth and P. Wang, "Process-Introduced Structural Defects and Junction Characteristics in NPN Silicon Epitaxial Planar Transistors," J. Appl. Phys. 36, pp. 1967-1973 (1965).
5. D. P. Miller, J. E. Moore and C. R. Moore, "Boron Induced Dislocations in Silicon," J. Appl. Phys. 33, pp. 2648-2652 (1962).
6. A. R. Lang, "Studies of Individual Dislocations in Crystals by X-Ray Diffraction Microradiography," J. Appl. Phys. 30, pp. 1748-1755 (1959).
7. G. H. Schwuttke, "X-Ray Diffraction Microscopy Study of Imperfections in Silicon Single Crystals," J. Electrochem. Soc. 109, pp. 27-32.
8. M. L. Joshi, "Effect of Fast Cooling on Diffusion-Induced Imperfections in Silicon," J. Electrochem. Soc. 112, pp. 912-916 (1965).
9. H. J. Queisser and A. Goetzberger, "Microplasma Breakdown at Stair-rod Dislocations in Silicon," Phil. Mag. 8, pp. 1063-1066 (1963).
10. K. O. Batsford and D. J. D. Thomas, "Defects in Vapour-Grown Silicon," Electrical Communication 38, pp. 354-362 (1963).
11. H. J. Queisser, K. Hubner and W. Shockley, "Diffusion Along Small-Angle Grain Boundaries in Silicon," Phys. Rev. 123, pp. 1245-1254 (1961).

12. P. F. Schmidt and R. Stickler, "Silicon Phosphide Precipitates in Diffused Silicon," J. Electrochem. Soc. 111, pp. 1188-1189 (1964).

13. H. J. Queisser and P. G. G. van Loom, "Growth of Lattice Defects in Silicon During Oxidation," J. Appl. Phys. 35, pp. 3066-3067 (1964).

14. F. J. Wilhelm and M. L. Joshi, "Fault Planes in Steam Oxidized Silicon," J. Appl. Phys. 36, pp. 2592-2593 (1965).

15. See reference no. 4.

16. M. L. Joshi and F. J. Wilhelm, "Diffusion-Induced Imperfections in Silicon," J. Electrochem. Soc. 112, pp. 185-188 (1965).

17. M. Juleff, "Measurement of Flexure in Silicon Wafers by the Lang X-Ray Diffraction Technique," to be published in the International Journal of Electronics.

18. This technique was developed at the GT&E Laboratories as part of the work performed under Contract AF 615 (1685)-7313, sponsored by the Air Force Cambridge Research Laboratories.

19. W. Bernard, W. Rindner and H. Roth, "Anisotropic Stress Effect on the Excess Current in Tunnel Diodes," J. Appl. Phys. 35, pp. 1860-1862 (1964).

THE EFFECT OF PHOSPHORUS DIFFUSION IN THERMAL OXIDES ON THE ELEVATED TEMPERATURE STABILITY OF MOS STRUCTURES

H. G. Carlson, G. A. Brown, C. R. Fuller and J. Osborne

Texas Instruments Incorporated
13500 North Central Expressway
Dallas, Texas

I. INTRODUCTION

At least two papers[1,2] have appeared suggesting that sodium contamination of planar oxides is a major source of device instability when devices are operated under d-c bias conditions at moderately elevated temperatures. This suggestion seems well supported by papers[3,4,5] describing the properties of alkali metals in bulk silica. In addition, at least three references[5,6,7] show how the mobility of the sodium may be used to extract sodium at the negative electrode with the sample at a few hundred degrees Centigrade.

Numerous investigators[8,9,10,11] show how the instability of the silicon surface can be investigated using a simple capacitor formed from a metal dot evaporated on top of a planar oxide, and with electrical connections made to the dot and to a metal film evaporated on the back of the silicon wafer. Measurement of the capacitance-voltage-frequency behavior of the capacitor is sufficient to characterize the underlying silicon surface. In general, it has been proposed that a positive potential applied at the dot with the sample at moderate temperatures causes the sodium to migrate within the oxide to the oxide-silicon innerface, where, by a charge mirroring mechanism, the free sodium ions attract electrons out of the bulk silicon to form an electron-rich region at the silicon surface. This effect is very nearly reversible when caused by moderate electrical fields at a few hundred degrees Centigrade.

It was shown in 1964 that a phosphorus diffusion into the oxide is sufficient to suppress the positive bias instability up to 175°C[12,13]. A mechanism other than the sodium displacement was suggested by these investigators[12].

*Supported in part under RADC Contracts AF 30(602)-3723 and AF 30(602)-3727.

No direct measurement of the sodium contamination and distribution in planar oxides has been reported in the literature to substantiate its presence or to clarify its role in the instability mechanism. Normal chemical techniques which might be applied to the problem are able to be used only when the atoms of interest in a sample accrue to 10^{15-16} . However, much lower limits may be easily detected with the technique of neutron activation analysis. This technique has been exploited in the study of sodium distributions in bulk oxide samples.

This paper applies activation analysis along with accurate sectioning techniques to the study of sodium distributions in simple dried steam oxides, of sodium and phosphorus distributions in phosphorus oxide diffused dried steam oxides, and of sodium and phosphorus distributions in slightly more complex samples to be described. Some data on thermal and electrical stressing of the capacitor structures formed in some of these manners is included for completeness. The stability findings are in essential agreement with the already reported work[12,13].

Some discussion is made of an empirical model for the contaminated planar oxide, and some qualitative arguments are offered for the cause of the various sodium distributions found in this study.

II. EXPERIMENTAL

SAMPLE PREPARATION

Oxidized silicon slices were prepared from a 1-inch diameter, (111) orientation, 4–6 ohm-cm, arsenic doped crystal with an etch pit count of 2680 per square centimeter and a lifetime of 128 μ-seconds. After sawing to 0.020-inch thick slices, the surfaces were lapped with 1800-grit alumina powder to 0.018 inch and then chemically polished to 0.010 inch.

The chemically polished slices were then treated by varying combinations of the following procedures: A. Oxidation in one atmosphere of water vapor at 1200°C, followed by a twenty-minute period at 1200°C in dry (~0.2 ppm H_2O) nitrogen. B. Phosphorus two-stage diffusion: (B-1) Deposition from $POCl_3$ reacted with oxygen in a dry nitrogen diluent at 1200°C for 30 minutes. Solvolysis for 10 minutes in boiling water. Diffusion for 60 minutes in a 5% oxygen, 95% nitrogen mixture at 1100°C. (B-2) Deposition as in B-1 but at 925°C for 30 minutes followed by solvolysis and diffusion as in B-1. C. Etch off the phosphorus glass, reoxidize and dry the oxide as in A.

Processes above will be referred to as A, B-1, B-2, and C in the rest of the paper. Oxidations and diffusions were performed in quartz reaction tubes with 10-ppm sodium contamination enclosed in furnace liners with sodium content from 0.5 to 5%.

Some samples were prepared for evaluation as MOS capacitors using process A followed by B-2. Gold was evaporated over the entire back of the slices and in 30-mil diameter dots on the oxide through a metal mask. After scribing into individual units the capacitors were mounted in a TO-5 header equipped with a special tungsten probe for contact to the gold dot. Ambient during encapsulation was (<3 ppm H_2O) nitrogen.

SECTIONING AND COUNTING TECHNIQUE

Both sodium and phosphorus concentrations were determined by counting incremental etch quantities from the oxide following neutron activation. Since high sensitivity was required, a high thermal neutron flux was necessary. With the Union Carbide reactor facility at Tuxedo, New York, it was possible to irradiate 1-inch diameter silicon slices at a flux of 10^{13} neutrons/cm^2-sec for a period of twenty-four hours. The samples were placed in the reactor one morning, removed the next day, and shipped to the Texas Instruments counting facility for measurements within one half-life of sodium-24 (i.e., 15 hours).

The samples used were described earlier in the paper. After final preparation the samples were sandwiched between zirconium foil disks and then wrapped in aluminum foil. Specific activities to be used for calculation of the sodium and phosphorus present in the samples were determined from counting known quantities of sodium chloride and $Na_3PO_4 \cdot 12H_2O$ irradiated at the same time and location as the silicon slices. In actual practice the sodium and the phosphorus standards had specific activities of 10^{10} atoms/count-minute and 6×10^{11} atoms/count-minute, respectively.

Incremental layers were removed from the oxide using dilute aqueous hydrofluoric acid in minimal volume. The thickness of the layer removed was determined by measurement before and after etching using a Gaertner Model L-119 ellipsometer. The small volume of water used to rinse the slice was combined with the etch volume to total about two milliliters. After counting this etch volume for sodium in a manner to be described, the phosphorus in the etch was prepared for analysis.

It was necessary to separate the phosphorus from small quantities of gold, arsenic, and copper, which comprise the other beta emitters known to be present. First a measured amount of phosphorus carrier was added to the incremental etch solution. After addition of concentrated sulfuric acid, the solution was evaporated to dense SO_3 fumes, and diluted. Additional phosphorus carrier was added, and the white silica precipitate was discarded. A three-stage purification was then performed. First, ten milligrams of gold, copper and arsenic carrier, one gram of ammonium chloride, and one drop of 10% potassium permanganate were added to the above solution. Gold, copper and arsenic were precipitated by hydrogen sulfide. The precipitates were discarded after combining washes. Magnesium chloride solution was added, and the phosphorus present was precipitated as magnesium ammonium phosphate by the addition of ammonium hydroxide. This precipitate was redissolved after

decanting the solution; additional gold, arsenic and copper carrier were added; and the preceding process was repeated two times. The final precipitate of magnesium ammonium phosphate hexahydrate was collected, washed with ammonium hydroxide, methanol and acetone, and dried at 110°C before mounting for counting the beta emission from the phosphorus-32.

Since this process separates the sodium in the solution above the phosphate precipitate, it is possible to measure the phosphorus activity without waiting for the sodium to decay to a low level. Further, elimination of the only significant additional impurities present, gold, copper, and arsenic, yields an unambiguous value for phosphorus concentrations from the beta emission count.

The phosphorus beta emission was counted on a Baird Atomics, gas flow proportional counter using the magnesium ammonium phosphate precipitates from each etch volume. Correction was added to compensate for loss in the precipitation separations. Additional counts were taken over an extended period of time to insure that the material had the decay constant of phosphorus-32.

Gamma radiation from the sodium-24 was analyzed using three different instruments: (1) a Victoreen single-channel analyzer consisting of a model 764-1 scaler and a model 764-2 pulse height analyzer; (2) a Baird Atomics single-channel analyzer composed of a model 215 amplifier, a model 134 scaler, and a model 510 pulse height analyzer; and (3) a Nuclear Data 512 channel multichannel analyzer model number 130 AT. A separate 2- by 2-inch sodium iodide well detector was used with each single-channel counting apparatus. The 512 channel analyzer used a 3-inch sodium iodide detector.

The single-channel analyzers were calibrated using a cesium 137 standard source, and were adjusted to cut off all gamma energies below 1.3 mev and above 4 mev. This region brackets the 1.368 and 2.754 mev gamma peaks of sodium-24. The multichannel analyzer was used to cross check selected data points taken with the Victoreen and the Baird Atomic instruments. Comparison of the data is shown in Table I.

Table I

Sample Number	Etch Step Number	Concentration Na (atoms/cc)	
		Multichannel Analyzer	Single-Channel Analyzer
19N80	1	1.24×10^{20}	1.21×10^{20}
19N73	1	1.16×10^{18}	9.49×10^{17}
19N73	7	1.19×10^{17}	1.15×10^{17}
19N80	7	1.54×10^{16}	1.46×10^{16}

Error in the sodium concentration values found in the experiment could arise from the fast neutron reactions: Mg^{24} (n,p) Na^{24} and Al^{27} (n, α) Na^{24}. These reactions produce significant Na^{24} activity only when the product of the number of atoms present, their reaction cross section (barns), and the fast neutron flux exceeds 10^{24}. The fast neutron to thermal neutron flux ratio in the Tuxedo, New York reactor is about one to three, while that of the Oak Ridge reactor is about one to thirteen. Several samples of smaller dimension exposed in the Oak Ridge reactor gave results comparable to those found after irradiation at the Tuxedo reactor.

TYPICAL SODIUM AND PHOSPHORUS DISTRIBUTIONS

Sodium profiles were obtained on several slices from separate oxidations formed according to process A. While the bulk concentration level in the oxide was found to vary with chemical treatment, as shown in Figures 1 and 2, the general form of the distribution is constant. More than 90% of all sodium found in the samples was contained within a few hundred angstroms of the outer surface of the oxide. A similar increase in concentrations in excess of the bulk level was found at the oxide-silicon interface.

Typical phosphorus and sodium distributions are shown in Figures 3 and 4 for oxides prepared by Process A, B-1 and A, B-2, respectively. It should be noted that the sodium appears to be extracted by the phosphorus glass and is lower in concentration in the oxide region between the glass and the silicon than in the simple oxide.

Samples listed in Table II were prepared by successive combinations of oxide growth (A), phosphorus diffusion (B-1), and etching of the phosphorus glass followed by reoxidation (C). Table II also lists the sodium concentration per square centimeter in the silicon substrates under the oxide. Those samples which did not have a phosphorus glass on the surface had at least thirty times more sodium in the silicon than the diffused oxide samples.

Figure 5 shows the distributions in sample 19N86 from Table II which has undergone a sequence of steps A, B-1, C, B-1, C, B-1. More sodium was found at the outer oxide surface of this sample than for sample undergoing only processes A, B-1. However, the sodium at the interface was removed to a low level. In contrast, sample 19N92, prepared according to the sequence A, B-1, C, B-1, C, gave the sodium distribution shown in Figure 6. This data is similar in most respects to that shown in Figures 1 and 2.

An attempt was made to redistribute the sodium in a simple dried steam oxide by applying an electric field of 1×10^6 volts/centimeter of oxide at 175° C for 180 minutes. The positive potential was applied to a large gold pressure contact on the surface of the oxide, while negative contact was made to the back side of the silicon wafer. The distributions in the oxide for biased and unbiased

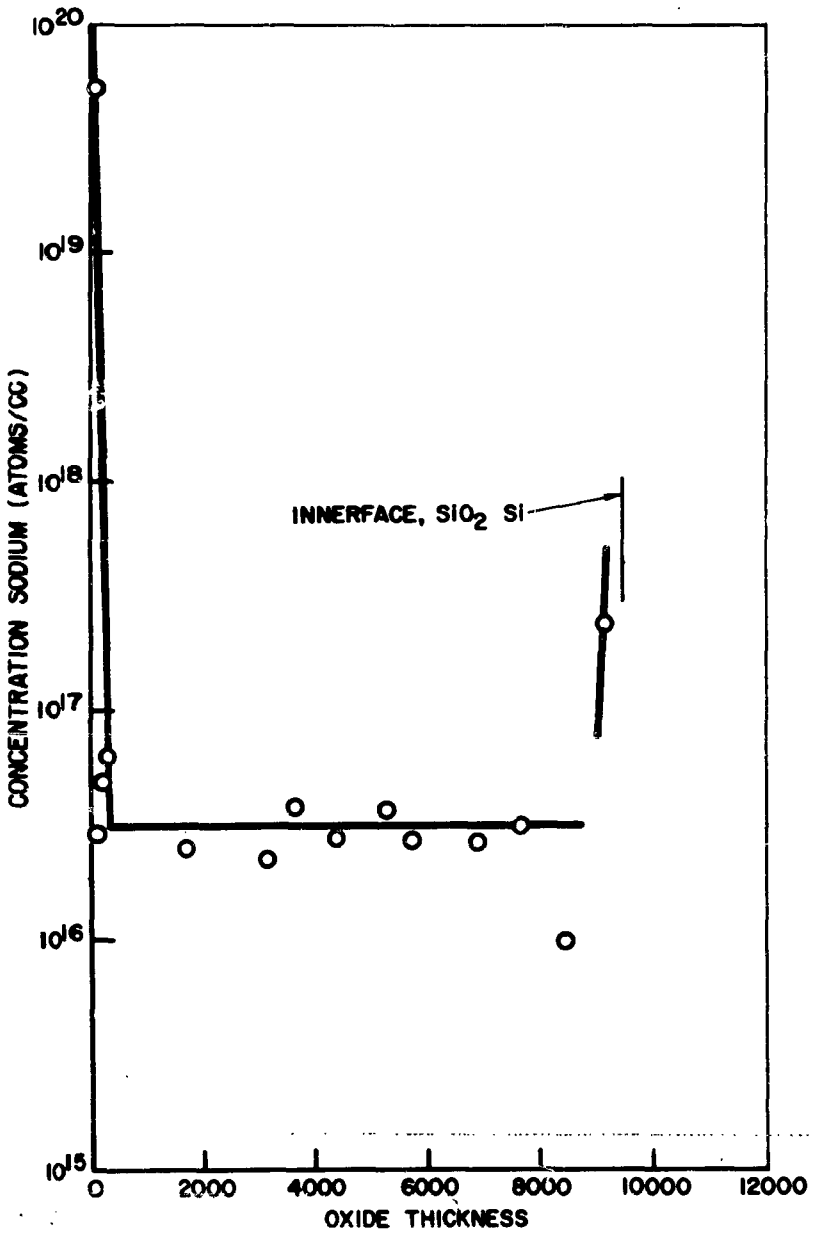


Figure 1

Sodium Distribution in Dry Steam Oxide (Process A)

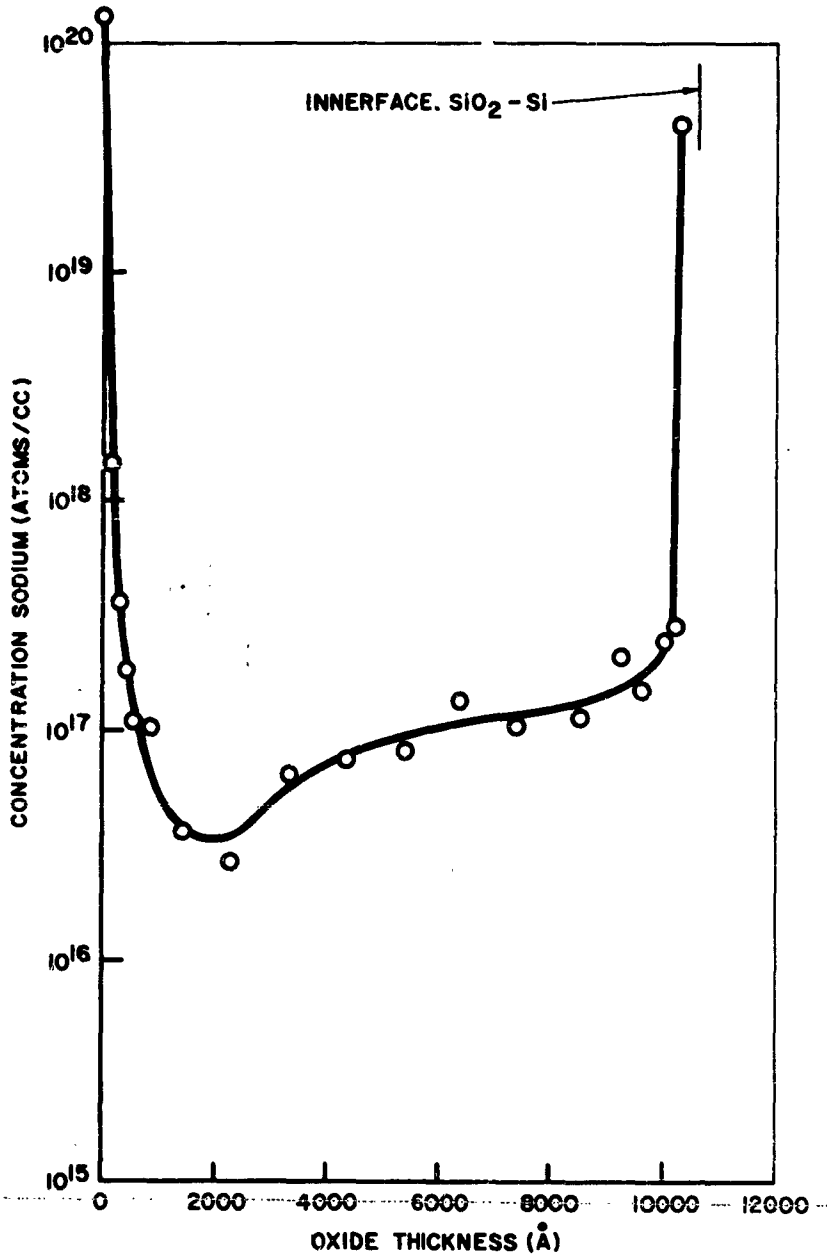


Figure 2

Sodium Distribution in Dry Steam Oxide (Process A)

Table II

Total Sodium in Silicon Substrates

Sample	$Atoms_{Na}/cm^2$	Process*
19N80	1.26×10^{13}	A
19N81	2.80×10^{12}	A
19N73	1.44×10^{12}	A, B-1
19N99	1.14×10^{12}	A, B-1
19N78	1.32×10^{12}	A, B-1, C, B-1
19N84	1.14×10^{12}	A, B-1, C, B-1
19N68	1.05×10^{13}	A, B-1, C, B-1, C
19N96	3.34×10^{12}	A, B-1, C, B-1, C
19N86	1.15×10^{12}	A, B-1, C, B-1, C, B-1
19N98	1.33×10^{12}	A, B-1, C, B-1, C, B-1

* Refer to Sample Preparation discussion.

samples were identical within experimental error. However, the total sodium concentration in the silicon of the biased samples increased thirty fold over the unbiased samples.

On unbiased dried steam oxide samples the sodium concentration in the first micron of silicon was in the 10^{17} atoms/cc range. This decreased to a low level on sectioning into the silicon.

OXIDE STABILITY ASSESSMENT USING MOS CAPACITOR STRUCTURES

Metal-oxide-semiconductor capacitor structures were fabricated to evaluate the stability of the variously prepared oxides. In particular, it was desired to compare the phosphorus treated oxides with those in which no post-oxidation treatment was employed.

The analytical technique was based upon measurement of high frequency MOS capacitance-voltage characteristics of the structures before and after subjecting them to thermal and electrical stress. Quantitative analysis of the

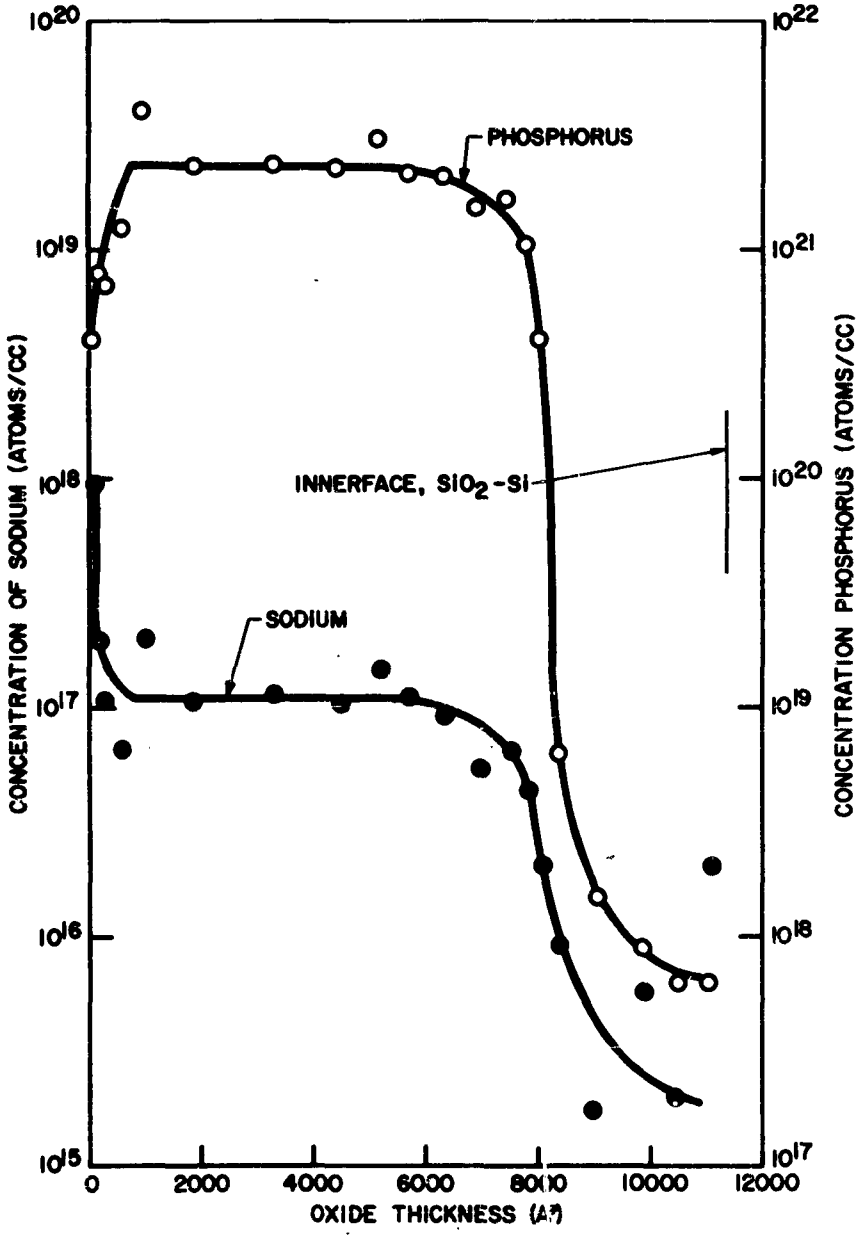


Figure 3

Sodium and Phosphorus Distribution in Diffused,
Dry Steam Oxide (Process A, B-1)

ATOMS SODIUM/CC

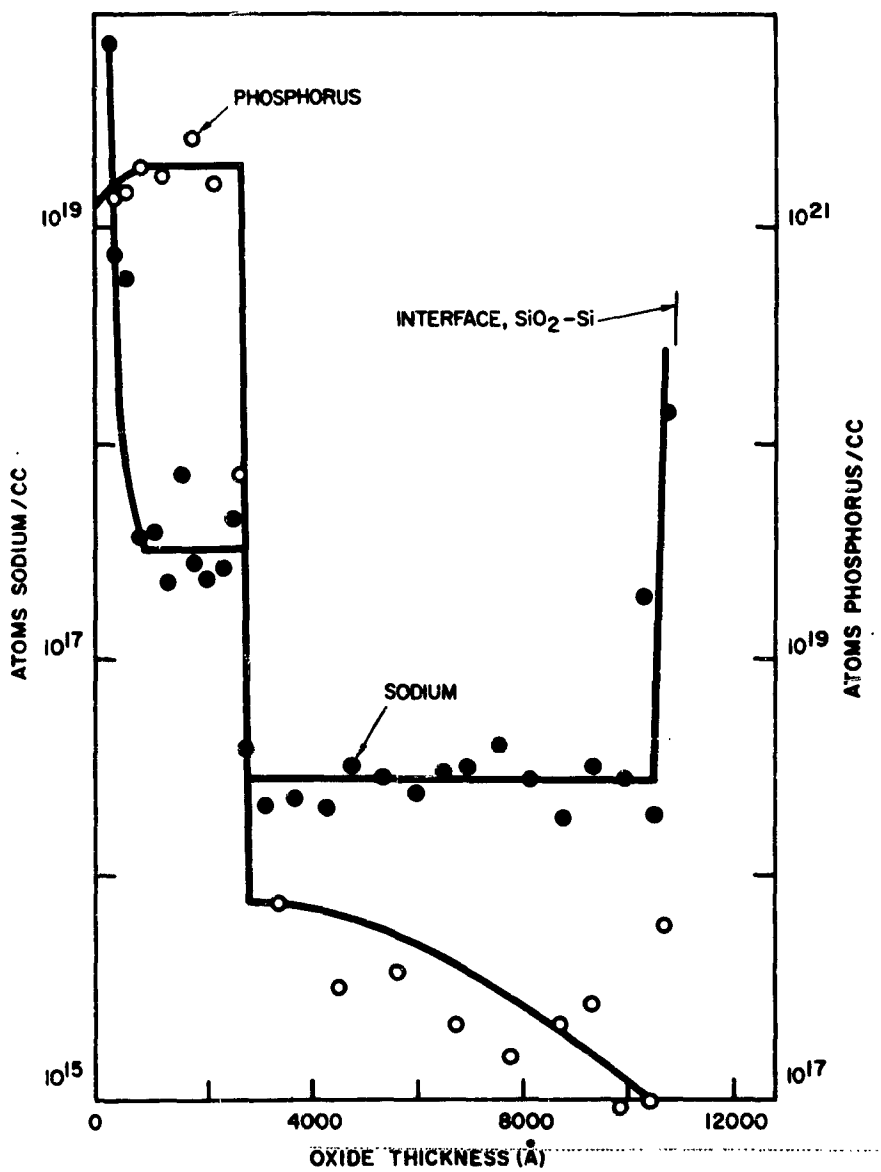


Figure 4

Sodium and Phosphorus Distributions in Diffused,
Dry Steam Oxides (Process A, B-2)

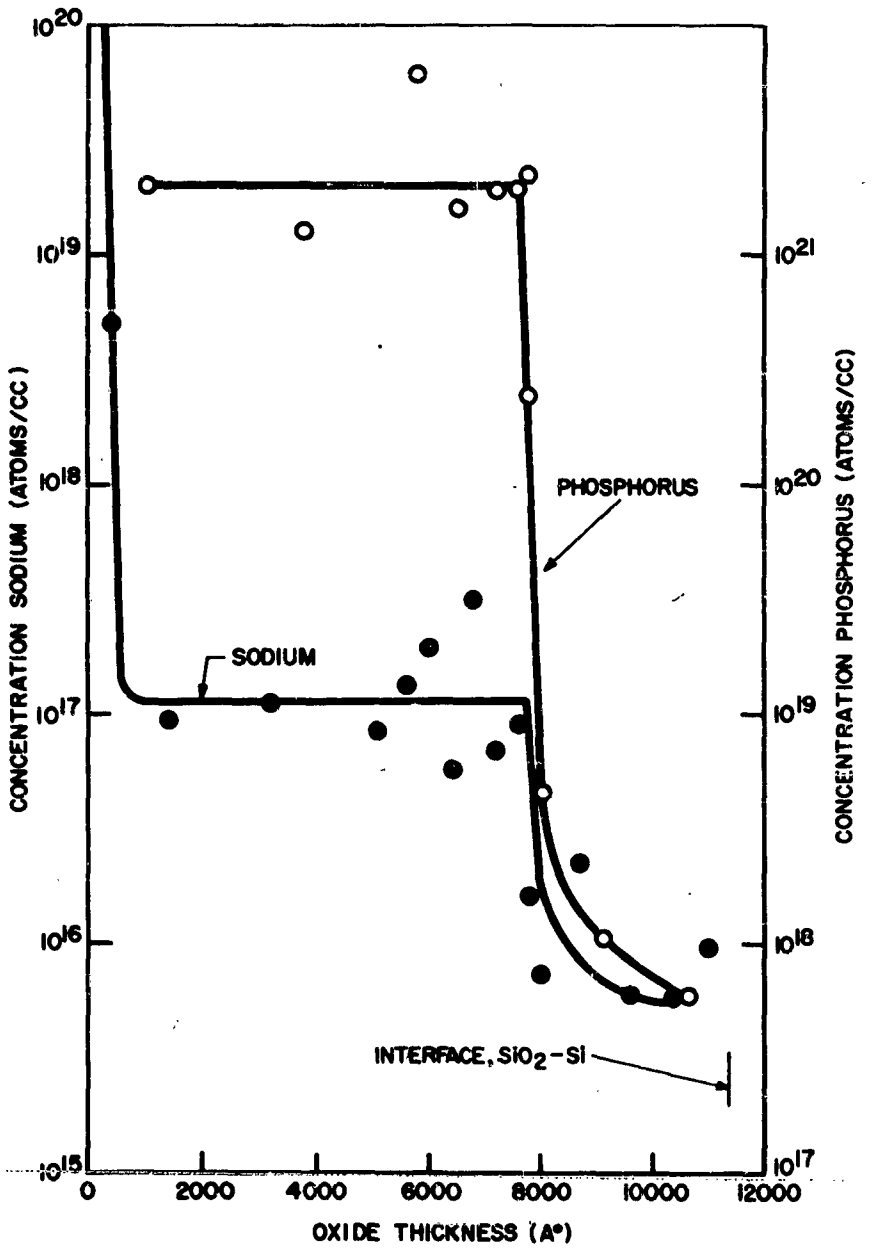


Figure 5

Sodium and Phosphorus Distribution in an Oxide After Several Processes (Process A, B-1, C, B-1, C, B-1)

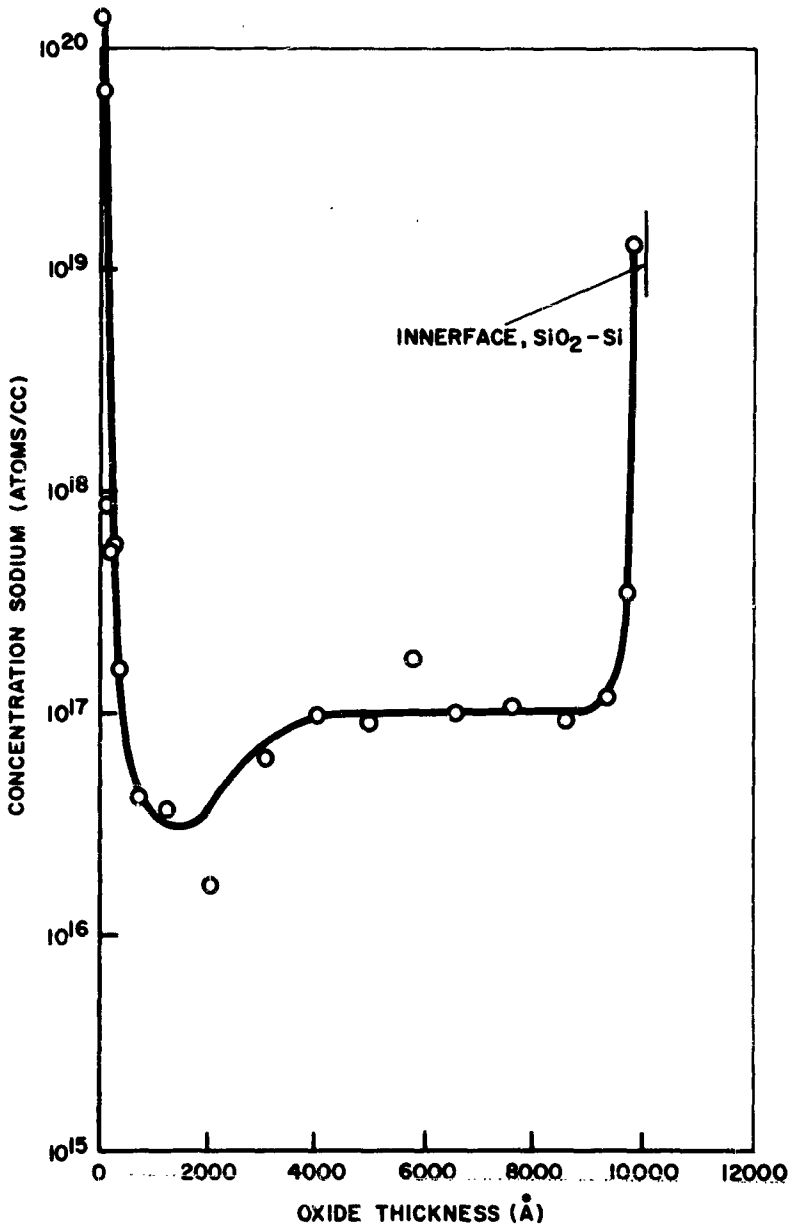


Figure 6

Sodium Distribution in Oxide After Several Processes (Process A, B-1, C, B-1, C, B-1, C)

measured C-V characteristics was accomplished by comparison with theoretically derived curves calculated as described by several workers [8,9,10,11].

It has been shown that the differences between the measured and theoretical capacitance-voltage curves are related to the magnitudes and densities of charge distributions in the oxide-silicon system, and that shifts of the characteristics under stress can be interpreted in terms of charge migration processes.

For this discussion MOS structures have been characterized by the bias voltage, V_{FB}, required to bring the silicon surface to a "flat band" condition. This condition can be determined from the measured C-V curve by theoretical analyses [8,9,10,11] if the oxide thickness and silicon resistivity are known. Since all samples were fabricated on silicon of the same resistivity and all electrical measurements made on oxides of known thickness, a comparison of sample stability can be made directly from the flat band potential. This assumes that the resistivity of the silicon surface is not changed by the processing. With one exception to be discussed below, this condition was satisfied in the experiments.

The exception involves material treated as described in process B-1. The phosphorus profile shown in Figure 3 indicates that an appreciable concentration of phosphorus penetrated the oxide-glass interface and reached the silicon surface. This doped the surface to a higher level, resulting in silicon surface depletion layer capacitance too large to be observed experimentally in this system. For this reason, the comparison of the stability of untreated and phosphorus treated oxides was limited to material treated as in processes A and B-2.

Figure 7 shows the change for samples prepared as in (A, B-2) in flat band voltage, ΔV_{FB}, after various durations under stress at 175°C (T_A) with a positive electric field (E_{ox}) of 10^6 volt/cm in the oxide. The level of stability indicated is so much greater than that of the non-phosphorus diffused oxides that it was not possible to compare both types of samples at this stress level. The C-V characteristics of the process A oxides were so strongly shifted after 5 minutes at the above stress level that they could not be meaningfully analyzed. By lowering the stress temperature to 125°C it was possible to observe the growth of the instability in these samples. This is also shown in Figure 7, where again ΔV_{FB} is plotted versus stress time with $E_{ox} = 10^6$ v/cm and with $T_A = 125°C$. It is seen that the phosphorus treated material is much more stable at 175°C than the untreated material at 125°C.

Material processed by phosphorus deposition, glass removal and reoxidation as described in process C was analyzed for MOS stability after up to four repetitions of the etch-regrowth cycle. In this experiment, flat band voltages, V_{FB}, were measured initially and again after 200 hours at the 175°C,

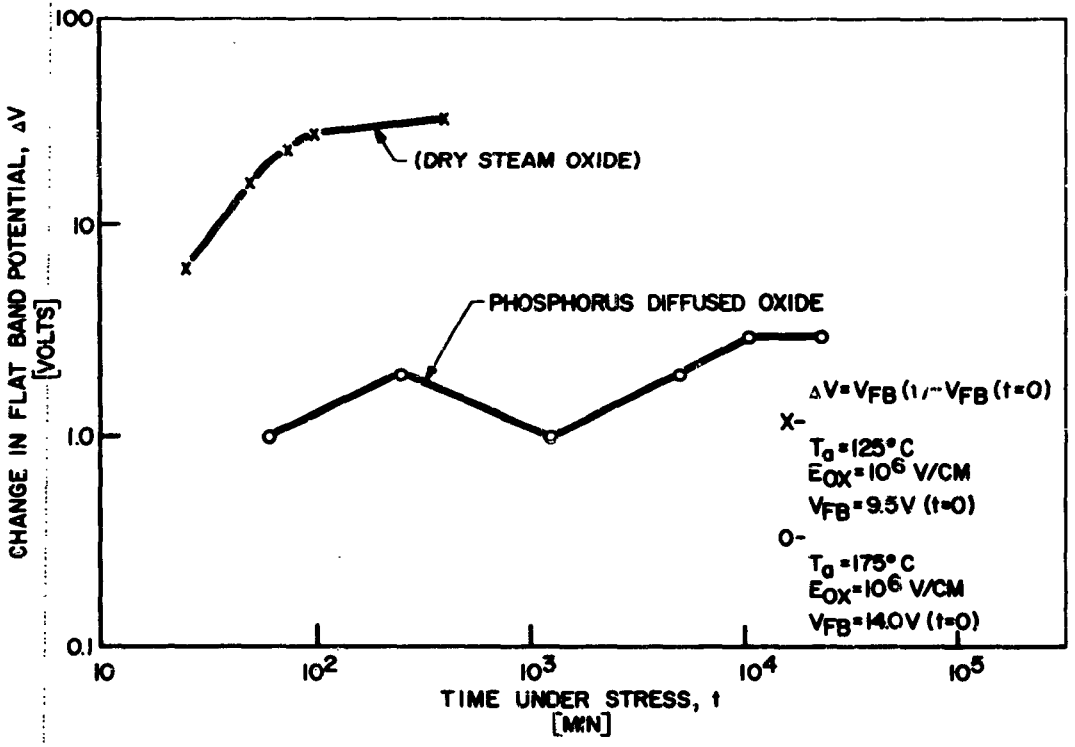


Figure 7

Oxide Stability Assessment From MOS Capacitance — Voltage Characteristics

10^6 v/cm stress condition. All the samples exhibited stabilities similar to that found for the simple phosphorus samples (Figure 7), but no significant improvement in stability with repeated etch-back-regrowth cycles was observed.

III. DISCUSSION

Figure 8 shows a Fischer-Herschfelder model proposed for a planar oxide formed on a (111) orientation silicon slice. On a local basis (a few atomic spacings) it is impossible to join every silicon crystal surface atom via oxygen linkage to other silicon atoms which are members of siloxane chains within the oxide. For the same spatial reasons that prevent total inclusion of all silicon crystal surface atoms in the oxide linkage, only a partial number of the silicon surface atoms uncombined with the oxide can terminate in simple linkages such as SiOH or SiONa. The remaining silicon atoms have to relocate, much in the same manner as on a clean surface to minimize the surface energy of the system. All of the surface silicon atoms which do not bond directly into the oxide are thought of as the origin of electronic surface states.

Because of the random arrangement of siloxane chains in the oxide, sites occur where the silicon-silicon distance is too great for direct oxygen bonding. At such positions reaction with water or with sodium oxide may occur to terminate the siloxane chains in pairs of SiOH or SiONa members. Chains at such positions may also be terminated by SiOH-SiF pairs.

Chains may be modified by introducing phosphorus atoms which can act as replacements for silicon in the siloxane linkage. Because of the differences in bonding angles, covalent radii and electronegativity of silicon and phosphorus, this replacement is suspected to lead to generally shorter chain lengths. This means there will be a larger number of non-bonding oxygen sites in the glass than in normal planar oxide. Since both water and sodium were present during formation of the oxide, it is postulated that the equilibrium ratio of sodium concentration in the phosphorus glass to that in the undiffused oxide is proportional to the ratio of the non-bonding oxygen concentration in the two phases.

The high concentration of sodium at the outer surface of the oxide is attributed to termination of surface silicon atoms by an SiONa structure. Since the steam oxide was "dried" at high temperatures in dry nitrogen (~0.2 ppm H_2O), a competition would arise between water and sodium oxide in the drying ambient. While no measurements were taken of a sodium distribution prior to drying it is expected that the outer surface would be more populated with SiOH terminals and have a lower net sodium level than on the dried samples.

Measurements by the BET[14] method have been made of the surface area of oxides prepared by various techniques[15]. The particular oxide with the sodium distribution shown in Figure 1 has an actual surface area of three to four times the geometric area. Since the thickness of oxide layers removed was

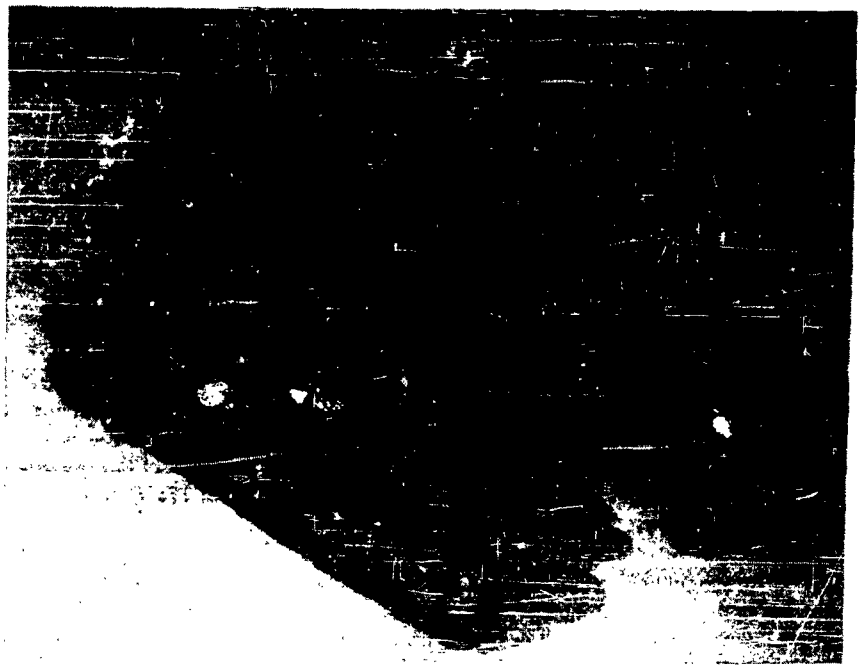


Figure 8

Molecular Model for Planar Oxide on Single Crystalline Silicon

In this model the atoms are identified as follows:

No.	Element
1	Silicon
2	Oxygen
3*	Sodium
4	Hydrogen
5	Phosphorus
6	Fluorine

* Sodium size and bond spacing are not exact.

determined by a difference in total thickness before and after etching, this measurement would not necessarily reflect any porosity at the outer surface. Hence, the physical surface is postulated to extend a short distance below the geometric outer surface.

From measurement with the BET method[14,15] the physical silicon surface after chemical polish has from one to two times the geometric area. At close approach (few hundred angstroms) to the innerface, the oxide thickness measurement has questionable meaning for calculation of concentration of sodium per cubic centimeter of oxide. In particular, if a large amount of the sodium in the oxide innerface region is bonded to the silicon crystal in SiONa units, it becomes meaningless to express the concentration of a monolayer as a bulk concentration. It is interesting to note that if the total number of sodium atoms within 100Å of the innerface on stable samples is computed, this value is very nearly the minimum number of states measured on the MOS capacitor samples (i.e., 2 x 10 inch states/cm^2). Few measurements[16,17] have been made of the properties of sodium in silicon. In particular the contribution of the sodium, which was found to move into the silicon under positive bias at 175°C in the experiment reported in this paper was not recognized.

It should be further noted that the samples stabilized by phosphorus diffusion had appreciable quantities of sodium between the innerface (Si-SiO_2) and glassy phosphorus phase. A simple charge mirroring model is not sufficient to explain all effects. In fact the most likely postulate is that the glassy phase serves as a barrier to diffusion of the sodium present at the outer surface of the glass through the oxide to the innerface or into the silicon. The significant diffusion mechanism may be a replacement of sodium in the oxide bulk by sodium liberated by electrolysis at the outer electrode. Considerable effort will still be required to rigorously support this postulate and others offered in discussion of this work.

REFERENCES

1) A. G. Revesz, IEEE Trans. Electron Devices ED-12, 97 (1965).

2) E. H. Snow, et al., J. Appl. Phys. 36, 1664 (1965).

3) J. Vermeer, Physica XXII, 1257 (1956).

4) A. E. Owen and R. W. Douglas, J. Soc. Glass. Tech. 43, 159T (1959).

5) E. H. Snow and P. Gibbs, J. Appl. Phys. 35, 2368 (1964).

6) H. George, Vitreous Silica and its Manufacture, July 28, 1959, applied for December 2, 1955, U. S. Patent No. 2897126.

7) P. F. Spremulli, Methods for Electrolyzing Glass, Filed October 31, 1962, U. S. Patent No. 3174919.

8) L. M. Terman, Solid State Electronics, 5, 285 (1962).

9) R. Lindner, Bell System Tech. J., 41, 803 (1962).

10) A. S. Grove, et al., Solid State Electronics, 8, 145 (1965).

11) K. H. Zaininger and G. Warfield, IEEE Trans. Electron Devices ED-12, 179 (1965).

12) D. R. Kerr, et al., IBM J. Res. Developm., 8, 400 (1964).

13) D. P. Seraphim, et al., IBM J. Res. Developm., 8, 400 (1964).

14) S. Brunauer, P. H. Emmett and E. H. Teller, J. Am. Chem. Soc., 60, 309 (1938).

15) D. E. Meyer and H. G. Carlson, to be published.

16) J. O. McCaldin, M. J. Little and A. E. Widmer, J. Phys. Chem. Solids, 26, 1119 (1965).

17) J. O. McCaldin and A. E. Widmer, Proc. IEEE, 52, 301 (1964).

RADIOCHEMICAL STUDY ON LATERAL ION MIGRATION IN INSULATING SUBSTRATES FOR THIN FILM MICROCIRCUITS*

S. S. Choi

Philco Applied Research Laboratory
Philco Corporation
Blue Bell, Pennsylvania

INTRODUCTION

It has been shown that a u-shaped tantalum thin-film resistor on oxidized silicon substrate goes through an initial decrease in resistance value on power stress. Also, when a dc potential is applied to the adjacent parallel tantalum thin-film resistors on 7059 glass substrate, a relative decrease in the resistance value of the negatively biased resistor results.[1]

Similar results are reported on the resistance changes of parallel tantalum nitride film resistors on soft glass and other types of substrate.[2]

These phenomena which are distinct from the usual positive change in resistance, noted in tantalum and most other metal oxide thin-film resistors, together with other anomalous changes, have been attributed to the migration of mobile alkali ions, especially sodium ions, in or on the glass substrate under the influence of an electrical field.[1-5] And it has often been suggested that the ion migration causes instability and other undesirable effects on various types of semiconductor devices.[6,7]

While the evidence for this is not conclusive, the results are convincing. And evidently the quantity of such mobile ions in these substrates and devices which influence the characteristics of the devices can be extremely small, in view of the fact that some of the matrices used were very pure in regard to sodium content.

*This work supported in part by Contract #AF30(602)-3287 Project #5519, Task #551902, with Rome Air Development Center, Griffiss Air Force Base, N. Y.

To improve the stability of such devices and the reliability of their performance, there is need for a more complete understanding of the mechanism of ion migration phenomena in these devices and substrates. However, such study is hindered greatly by the lack of suitable techniques which offer direct and quantitative results. Thus, an analytical procedure which has high sensitivity of detection is required to trace the motion of ions in these media under various operational and accelerated test conditions.

The analytical technique is being developed to facilitate the study of ion migration in insulating substrate. Use is being made of radioactive isotopes to trace the motion of minute quantities of supposably mobile ions, and the radiations emitted from these isotopes are being detected by means of autoradiographic techniques. The distribution and the concentration of the tracer isotopes can be thus determined before and after the aging or power stress, or both. The results can be correlated with the test conditions used, and the mechanism of ion migration may be thus evaluated.

Preliminary results indicate the potential value of this technique. And some of the results obtained thus far are presented in this paper.

BACKGROUND

Radioactive isotopes are widely used today in various fields of research. One of the most important applications of these isotopes is their use as tracers. Since the physical and chemical properties of radioactive isotopes and stable isotopes can be considered identical (except an isotope of the elements of low atomic numbers), radioactive isotopes can be used without interference to the process being carried out. Because of the high sensitivity of detection radioactive isotopes have proved useful in numerous analytical procedures.

Photographic emulsion has been used for the identification and the measurement of radiations from the radioactive isotopes. It also provides a record of the space relation of many nuclear events with the object under study. The two or three dimensional image on or in the photographic emulsion produced by the radiation from the radioactive isotopes is called an autoradiogram.

Various types of photographic emulsions or plates can be used for autoradiographic study. The particular emulsion selected for any given program is determined by the type of radiation to be recorded, the level of radiation and the degree of resolution required. The problem under study is ion migration in the dimension of microcircuits and this requires a high resolution of image if a quantitative analysis is to be made. One of the important factors determining the resolution of the autoradiogram is the size of the silver halide grain in the emulsion, which varies greatly with the type of emulsion.

For instance, in X-ray emulsion there are approximately 6×10^6 grains/cm^3 while in nuclear emulsion there are approximately 10^{13} crystals/cm^3. The range of the diameter in X-ray emulsion is from 0.5 to 3μ and that in nuclear emulsion is 0.1 to 0.4μ with a mean diameter of about 0.2μ. Therefore, nuclear emulsion is chosen for use in this initial study since it will provide a higher resolution. The relevant information in various nuclear emulsions is given in Table I.

Table I

Characteristics of Nuclear Emulsion

		Diameter of Silver Halide Grain	Diameter of Silver Grain after Development
Kodak	NTB_3	0.34μ	0.6μ
"	NTB_2	0.26μ	0.6μ
"	NTB	0.29μ	0.6μ
"	NTA	0.22μ	0.6μ
"	NTE	0.06μ	0.06~0.12μ
	after centrifuge	(300~500Å)	(800-1400Å)
"	AR10	same as NTB_2	resolution 2μ
"	AR50	Corse	" 15μ
	NUC307	0.07μ	difficult to handle
	Ilford L4	0.15μ	0.2 ~ 0.45μ

Sodium has two usable radioactive isotopes Na^{24} and Na^{22}. The relevant nuclear data is given in Table II.

Table II

Radiation Data of Na^{22} and Na^{24}

Radiation / Isotopes	Electron Captune (EC,) Beta Energy, (MeV) Abundance	Gamma Energy, (MeV) Abundance
Na^{22} (Halflife:2.6yrs)	EC (11%) B^+ 0.54 (89%) B^+ 1.83 (0.06%)	1.28 (100%)
Na^{24} (Halflife:15 hrs)	B^- 1.39 (100%)	1.37 (100%) 2.76 (100%)

As will be seen in the procedure described in the next section of this paper; the problem under study requires considerable duration of pretreatment of the test specimen before actual autoradiographs can be made. Thus, preference was given to Na^{22} for its long half life with

suitable radiation. Furthermore, Na^{22} is produced by Mg^{24} (d,2)Na^{22} reaction and is obtainable in much higher specific activity than that of Na^{24}. This reduces the minimum amount of sodium which has to be introduced to the specimen to obtain a meaningful autoradiogram. Moreover, the maximum energy of positron from Na^{22} is much less (0.54 MeV) compared to that of beta ray from Na^{24} (1.39MeV), so with sodium 22 a higher resolution and higher photographic response can be obtained on nuclear emulsion. For all these reasons sodium 22 was chosen to be used in this study.

EXPERIMENTAL

Test Specimen and Procedure

Shown in Figure 1 is an experimental area of interest of a discrete unit of the resistor sets on the test vehicle. At least four such discrete units are contained in each test vehicle, which usually measured 1 x 1 cm^2~1 x 2 cm^2.

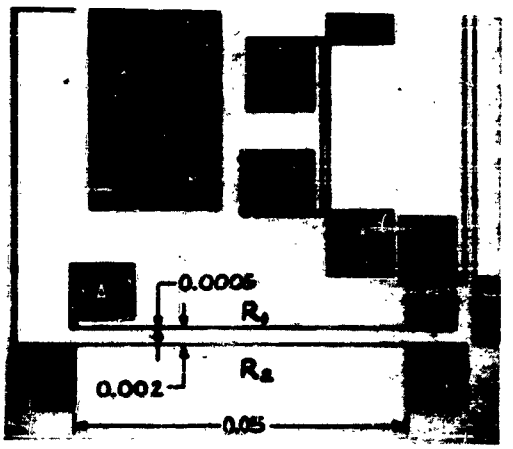


Figure 1

Parallel Resistor Test Pattern (X34)

Alkali-free glass (Corning #7059) was used as a substrate material; a tantalum film, 700Å thick, sputtered at 2300 volts was used as the resistor material. Details of the method of fabrication of the test vehicle have been given elsewhere.[1] In this investigation adjacent parallel resistors, R_1 and R_2 in Figure 1, were used as the test pattern. Each resistor is 1/2 mil in width, 50 mils in length, and they are spaced 2 mils from each other.

All the resistors were stabilized at 400°C for 1 hour in air before use. The typical sheet resistivity of 7059 glass substrate was in the order of 10^{11} ohms per square at 250°C in room air. Leakage current between the two parallel resistors was in the order of 10^{-12} amperes at the same temperature.

After all the necessary electrical measurements were made at 250°C in air, the following procedure was adopted for the autoradiographic study of the migration of sodium ion in the insulating substrate.

1. Place approximately 6×10^{-2} micro-curies (μc) of Na^{22} C ℓ solution onto two of the usable parallel resistor test patterns of adjacent units.

   a. About 1/2 cm^2 or less of the surface area should be covered by the radioactive solution.

   b. The parallel resistor test pattern of the rest of the unit in the same test vehicle is left untouched as a control.

2. After the radioactive solution has dried, place the test specimens individually into a platinum box and heat to 400°C for about 1 hour in air.

3. Measure the radioactivity of the specimen.

4. Rinse the specimen once with D. I. water to remove foreign matter and, also, the excess of easily removable sodium chloride from the surface of the specimen.

5. Measure the radioactivity of the rinsed specimen and check the amount of sodium activity remaining with the test specimen.

6. Measure the resistance of each resistor, and the leakage currents between the adjacent parallel resistors of test patterns, maintaining the same conditions as before, such as temperature, bias, etc.*

7. Note and discard the units in which the resistor or resistor patterns have been damaged or have produced anomalous effects during these processes.

---

* To be sure that the introduction of radioactive sodium did not affect greatly the electrical characteristics of the test pattern.

8. Strictly maintaining the safe light conditions, expose the radioactive test specimens to the appropriate nuclear emulsion plate by holding together the radioactive surface of the specimen and the emulsion side of the plate for a predetermined duration under preset conditions.

9. At the end of the exposure, separate the nuclear track plate from the test specimen and process photographically as per instructions given in the instruction manual for the particular nuclear plate.

10. Measure the radioactivity of the test specimens again, to detect any loss of activity due to contact with the nuclear plate.

11. Apply the potential for a given duration between two adjacent parallel resistors of test pattern at given temperature (220°C).

12. Quickly cool the specimen after biasing. Re-expose to another nuclear plate of the same type.

    a. Quick cooling will freeze the radioactive atoms in their positions.

13. Photographically process the exposed nuclear plate as before.

14. Compare the photomicrograph of the first and second autoradiograms to observe the changes, if any.

15. Measure the activity of the specimen [a] again and mount onto a slide glass which was previously coated with gelatine. [b]

    a. To ensure that subsequent autoradiograms will be made from the test specimen which retains the same distribution and the same amount of radioactivity as in the previous steps.

    b. To ensure good wet adhesion of the emulsion when the autoradiographs are processed.

16. Apply stripping film directly onto sample surface as per instruction under the specified safe light conditions, and store in a light-tight box for a given duration under the preset conditions.

17. Process the autoradiographic stripping film on the sample specimen without removing it from the specimen, as per instruction.[8]

18. Prepare photomicrograph of the autoradiogram viewed on the sample.

    a. The change in density* distribution if any could be observed in relation to its original test pattern.

RESULTS AND DISCUSSION

Figure 2a shows photograph of an autoradiogram of the test specimen (obtained at step 9 of the procedure) before the application of the bias between two parallel resistors of a test pattern. The radioactivity of this test vehicle measured at step 5 of the procedure was about 15,000 counts/minute, indicating that the activity remaining with the specimen is about 2 x 10^{-2}** micro curies after the rinsing. (Step 4.)

Figure 2a

Autoradiogram of Parallel Resistor Test Pattern (R_1, R_2) Before Biasing (X34)
Radioactive Isotope Used: Na^{22}
Emulsion: Kodak Nuclear Track Plate NTB
Exposed one week at 4°C

---

* Density: Defined as relative concentration of developed dark grains in the emulsion or in the photograph.

** Compared with the 10^{-2} micro curies Na^{22} standard which gives 7500 counts/min. under the same counting conditions.

The relative resistance value did not change appreciably when compared with the control (step 6), the leakage current between the two adjacent parallel resistors were changed from the order of 10^{-12} amperes to the order of 10^{-11} amps, but a similar increase was observed with the control patterns. Thus it was decided that in this case the introduction of the radioactive sodium did not produce any unusual adverse change in electrical characteristics of the test pattern of parallel resistors.

Kodak nuclear track plates Type NTB (emulsion thickness 10μ) were used to obtain the above autoradiogram. The specimen was exposed for one week in a light-tight box at 4°C. The plate was developed as outlined in ref. 8.

In spite of scattered areas of high density regions, the pattern of the resistor unit is clearly discernible in the autoradiogram. Figure 2b, an enlarged view of a portion of Figure 2a, is representative of the general characteristics of the entire pattern. Except for a few spots, the distribution of the density is rather uniform between the two resistors. Figure 3, an autoradiogram of the same area of the test pattern*, depicted in Figure 2b, was produced after the application of 500 volts between the two resistors** for over three days at 220°C.

Figure 2b

Enlarged View of Figure 2a
(X113)

* Exposed for 10 days, otherwise all conditions were the same.

** The terminal A of R_1 (inner resistor) was positively, the terminal C of R_2 (outer resistor) was negatively biased.

Figure 3

Autoradiogram of Test Pattern
(X113)
Biased 500 volts
A(R_1) - Positive
C(R_2) - Negative
Emulsion: Kodak Nuclear Track Plate
NTB Exposure 10 days at 4°C

When Figure 2b and Figure 3 are compared, the distinct change in the density of the area under the positively biased resistor R_1 can be easily seen. Since there was no appreciable change in the measured radioactivity of the test specimen before and after the first exposure to the NTB plate (step 10), the density change seen here should be due entirely to the biasing of the resistors. Since the photographic registration in the emulsion is due to the radiation from the radioactive sodium ion (or atom, whichever it may be) the change of the density between Figure 2b and Figure 3 is due to the re-distribution of the radioactive sodium. Thus the visual evidence is established that the redistribution of the sodium occurs in or on the insulating glass substrate (such as 7059 glass) under the influence of an electrical field. Further comparison of these two figures yields the following facts.

- Figure 2 -

a. The density of the area between two resistors R_1 and R_2 is rather uniform.

- Figure 3 -

a. The area under the inner resistor R_1, which has been biased positively, reduced its density greatly.

b. The area under the outer resistor R_2, which has been negatively biased, did not reduce its density.

c. The width of both the lower density region under R_1 and the higher density region under R_2 increased when compared with the width of the corresponding region in Figure 2b.

d. It appears that the width of the lower density region under R_1 extended both sides of the R_1 width, while the width of the high density region under R_2 extended towards the R_1 direction.

e. The surroundings of both terminals (A, B) of the inner resistor R_1, which were positively biased, are reduced in density, while there appears to be no appreciable change in the surroundings of both terminals of R_2, which were biased negatively. (Terminals B and C are not shown in this figure.)

From these facts it is evident that under the influence of the electrical field, radioactive sodium was repelled further from the higher positive potential region. However, it is difficult to make a precise analysis from these autoradiograms alone of the change in spacial distribution of density, without aligning the autoradiogram with the test pattern from which these autoradiograms are made. It is, however, not always a simple task, though not impossible, to align exactly the autoradiogram and the original test patterns,[9,10] particularly when the size of the original test pattern is very small.

Figure 4 shows another autoradiogram of the same portion of the test pattern, as that shown in Figures 2b and 3 which was made after the second autoradiogram. This autoradiogram was obtained using an emulsion layer stripped from a Kodak AR10 stripping plate, and was applied to the sample as per instructions provided in the stripping plate package. The stripped emulsion layer was exposed to

the specimen for two weeks in a light-tight box at 4°C and was processed, as outlined in ref. 8, with the emulsion layer in permanent contact with the specimen, so that the final autoradiograph is in perfect register with the same specimen. There was no decrease in measured activity (step 16) after the second autoradiogram was taken, so that the test pattern still retains the same distribution and the same amount of radioactive sodium. Therefore, the photograph in Figure 4 may be easily correlated to that of Figure 3. However, Figure 4 is the photograph of the final autoradiogram, which was viewed with the specimen using transmitted light. Thus, this photograph represents the composite of two photographs, one of which is the photograph of the test specimen itself, the other being that of the autoradiogram produced by the distributed radioactive sodium on the surface of the test specimen. The spacial distribution of the radioactive sodium with regard to the resistors, terminals, etc., can easily be studied by examining the distribution of the density in these pictures. Also, by comparing Figure 4 with the photographs previously examined, more detailed information can be obtained concerning the motion of radioactive sodium on a glass substrate. Furthermore, with the aid of densitometer, the quantitative treatment of the subject appears possible. When Figure 4 is compared with Figures 2 and 3 the following observations can be made.

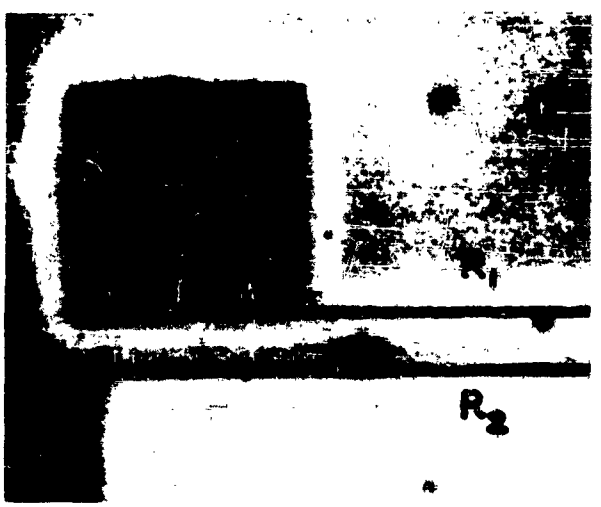


Figure 4

Stripping Film Autoradiograph of the Biased Specimen
(X113)
Emulsion: Kodak Stripping Plate AR-10
Exposure: two weeks at 4°C

a. Areas of depletion of density developed along both sides of the positively biased resistor R_1 and the area surrounding terminals A of R_1.

b. The width of the depleted area on both sides of R_1 and the area surrounding terminals A and B is practically the same.

c. The width of the area of increased density along the side of R_2 facing R_1 can be measured.

d. A very high density area appears in the space between terminal A and resistor R_1. Between this high-density region, and terminal A and resistor R_1, there exists a narrow but distinctly highlighted boundary. Toward the open end of this space, the densely grained ribbon narrows in width somewhat abruptly, forming an acute angle. This angle does not reduce, however, to its vertex, but bleeds off in a curved path around the corner of terminal A, i.e., curved away from resistor R_1. The radius of this curve is such that any point on the bend remains approximately equidistant from resistor R_1 and terminal A. This high-density line (bend) forming the curve decreases in intensity and fades into the depleted area. This indicates that sodium on a glass substrate is repelled strongly by the positive bias.

From these observations it is obvious that sodium on a glass substrate does migrate under the influence of an electrical field, and the direction of the migration is the same as that of the electrical field. It is apparent that sodium does become positively ionized. However, such sodium ions appear to be more strongly influenced by a positive bias. It appears to be possible to obtain by this technique much useful quantitative data, such as depletion rate, activation energy for migration, etc. As can be seen in Figure 4, the grain size is small enough to enable the study of sodium migration in much shorter distances, perhaps as short as one-eighth the separation of R_1 and R_2. Since the actual distance between R_1 and R_2 is 2 mils, with this technique one may be able to study ion migration in one-quarter of a mil. Furthermore, one may be able to study ion migration in much shorter distances, probably in the order of one-tenth of the present practical limits of 1/4 mil, by use of finer-grain emulsion, such as NTE (Kodak) in conjunction with the electron microscope.[11]

The migration of many other isotopes can be studied with this technique. And many isotopes are available now in a carrier-free state, or having very high specific activity. The amounts of total isotopes (stable and active) associated with the minimum amount of radioactive isotope which are required to produce meaningful autoradiograms can be extremely minute. Since the chemical and physical properties of radioactive isotopes and stable isotopes can be considered identical (except isotopes of the elements of low atomic numbers), the behavior of the former represents that of the latter. Thus, the results obtained in this work can be considered as that of ordinary sodium.

Conclusions

Autoradiographic techniques are used to study lateral migration of sodium in insulating glass substrate for microcircuits. NTB plate and AR10 stripping film can be used for this purpose. In this work the visual evidence has been established that sodium does in fact move laterally in insulating glass substrate under the influence of an electrical field in certain ambient conditions. The direction of migration of sodium is the same as that of the electrical field as is to be expected. Use of stripping film technique, which involves processing emulsion in permanent contact with the sample, provides an excellent basal index of measurement. Because of the high resolution obtainable, and the inherent high sensitivity of detection, this technique should prove useful in allied fields as well as with the problem presently under study.

For the systematic study of ion migration phenomena in insulating substrates work is underway using radioactive sodium prediffused material as the substrates.

REFERENCES

1. A. F. McKelvey, G. L. Schnable, M. Sharp, and M. Walker, "Accelerated Aging and Failure Mechanism Analysis of Thin Tantalum Film R-C Networks," to be presented at the Fourth Annual Symposium on Physics of Failure in Electronics, Nov. 16-18, 1965, Chicago, Ill.

2. W. H. Orr and A. J. Masessa, "The Effect of Substrate Ion Migration on Thin Film Tantalum Nitride Resistors," presented at the Electrochemical Society Fall Meeting, Oct. 10-14, 1965, Buffalo, N. Y.

3. S. M. Cox, "Ion Migration in Glass Substrates for Electronic Components," Physics of Chemistry of Glasses, Vol. 5, pp. 161-165, Dec., 1964.

4. D. A. McLean, N. Schwartz, and E. D. Tidd, "Tantalum-Film Technology," Proc. IEEE, Vol. 52, pp. 1450-1462, Dec. 1964.

5. G. Siddall and B. A. Probyn, "Vacuum-Deposited Metal Film Resistors," Brit. J. Appl. Phys., Vol. 12, pp. 668-674, Dec. 1961.

6. R. H. Kingston, "Review of Germanium Surface Phenomena," J. Appl. Phys., Vol. 27, pp. 101-114, Feb. 1956.

7. E. H. Snow, A. S. Grove, B. E. Deal, and C. T. Sah, "Ion Transport Phenomena in Insulating Films," J. Appl. Phys., Vol. 36, pp. 1664-1673, May, 1965.

8. "Kodak Materials for Nuclear Physics and Autoradiography," Kodak Scientific and Technical Data P-64, Eastman Kodak Company, Rochester, N. Y.

9. G. A. Boyd, Autoradiography in Bidogy and Medicine, New York: Academic Press, Inc., 1955, p. 99.

10. J. Novek, "Improvement of a High-Resolution Autoradiographic Apposition Method," International Journal of Applied Radiation and Isotopes, Vol. 15, pp. 485-489, 1964.

11. M. M. Salpeter and L. Bachman, "Autoradiography With the Electron Microscope," Journal of Cell Bidogy, Vol. 22, pp. 469-477, 1964.

SECTION VI

MINUTEMAN II CQAP PROGRAM

OPENING REMARKS

MINUTEMAN II, PHYSICS OF FAILURE PROGRAM

J. F. Wiesner, Capt., USAF

The "Component Quality Assurance Program" (CQAP) was part of contract AF04(694)-247 between Ballistic System Division of the United States Air Force and Autonetics Division of North American Aviation Inc., and was directed towards improving the reliability of the Minuteman II weapon system. The reliability of the Minuteman I weapon system was achieved via life testing components, tight process controls and extremely rigid failure analysis. In order to reach the failure rate goals established for Minuteman II, it was quickly realized that life testing was impractical because of the extremely large sample sizes that would be required to get statistically significant data within a reasonable time. Tight process controls and rigid failure analysis was continued in the Minuteman II reliability program. The solution to the reliability problem then was the "Component Quality Assurance Program" which was carried out by various component manufacturers. By applying various step stresses such as temperature, voltage, heat, vibration, etc., it was possible for the manufacturers to uncover weak links in the device construction or fabrication techniques and to apply corrective actions in the form of design, material or process changes. If the corrective action resulted in the ability of the device to withstand the stress it had previously failed, it was assumed that its inherent reliability was improved. It was also evident that a basic approach was needed to identify the causes and mechanisms underlying the failures that were uncovered by the application of these step stresses. The Physics of Failure Program was just such an approach and complemented the CQAP effort.

Figure 1 illustrates the organization of the CQAP Physics of Failure Program showing how it was controlled by Ballistic Systems Division of the United States Air Force. Rome Air Development Center (RADC) was asked to provide technical direction to the program because of their reliability experience with electronic components and their technical direction would not cause any proprietary problems. One of the major obstacles that had to be overcome in this program was the management problem associated with scientific work and its direction towards a specific goal. The results of this program indicate that this obstacle was successfully overcome and that the efforts were instrumental in improving the reliability of the Minuteman II weapon system.

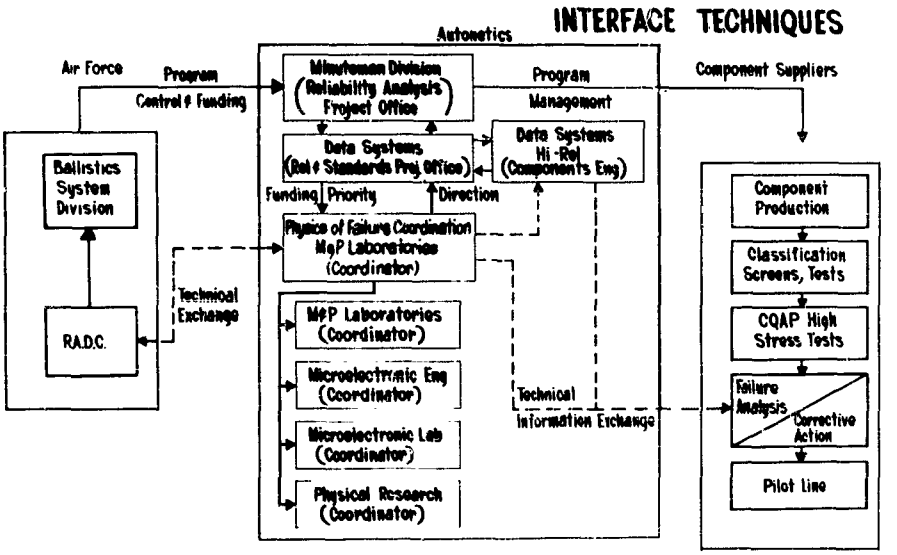


Figure 1. WS 133B Physics of Failure/Component Reliability Program

The devices that were to be investigated by CQAP and the Physics of Failure Program were selected on the basis of their projected impact on Minuteman II reliability. Figure 2 series illustrates the kind of chart used to compare the various devices, their failure modes and their impact on Minuteman II reliability. All numerical references on this chart are not actual values but do serve to demonstrate the method used. Figure 3 serves to illustrate the kind of chart used to determine a suitable grouping of failure modes into suitable Physics of Failure investigatory areas. Tens of thousands of parts from the various CQAP component manufacturers together with small quantities of unstressed Minuteman II parts from system failures were made available to the

Physics of Failure Program. Figure 4 illustrates the phased technical approach used in the program while exploring the areas illustrated in Figure 3. A wide variety of sophisticated instrumental analytical equipment such as the electron beam microprobe, mass spectrograph, gas chromatograph, etc., was used to characterize abnormalities in the electronic devices of interest. All the abnormalities noted did not lend themselves to the complete phased approach illustrated in Figure 4. As a result the methodology illustrated in Figure 5 evolved. In the left hand column it may be seen that an initial Physics of Failure payoff results because corrective action by the manufacturers can occur at this point for mechanisms that are obviously going to cause a failure. The second payoff can be demonstrated by the middle column which represents those abnormalities which can be shown to be associated with a failure mode in a reasonable time such as within the study period associated with Minuteman II. The information in the right-hand column can result in a third payoff by being of use to future programs concerned with new types of devices that may have sensitivity to these abnormalities in failure modes.

FAILURE MODES

	OPEN	INTERCONNECT PREFORM	INTERFACE FAILURE	BREAKDOWN PASSIVATION	CONTAMINATION CRACKED	METALLIZATION SHORTED DIE	BOND OFF	INTERCONNECT WIRE OFF	BOND OR LEAD UNSEALED	RESISTOR WIRE	PLASTIC PACKAGE FAILURE	INVERSION LAYER FORMATION	CURRENT CONCENTRATION	CONTAMINATION	INDUCED FAILURE	PROBE DAMAGE	MECHANICAL FAILURE	BROKEN RESISTOR WIRE	SHORTED RESISTOR TURNS	ΔFR
TRANSISTOR I	5	10	X	X	20	5	0	X	X	40	10	5	5	X	X					2.95
TRANSISTOR II	0	5	X	X	5	5	0	X	X	60	10	5	10	X	X					1.58
TRANSISTOR III	0	5	X	X	5	10	0	X	X	60	5	5	10	X	X					2.51
RESISTOR	X	X	X	X	X	X	5	65	X	X	X	10	5	10	5					0.27
DIODE	X	10	X	X	15	X	0	X	30	25	15	5	0	X	X					0.60
CAPACITOR	X	0	40	25	X	X	10	X	0	X	X	15	10	X	X					0.29
INTEGRATED CIRCUIT I	20	5	X	X	5	15	0	X	X	30	0	15	10	X	X					0.26
INTEGRATED CIRCUIT II	15	10	X	X	5	15	0	X	X	35	0	10	10	X	X					0.55
FR	.26	.62	.12	.07	.92	.60	.04	.18	.10	4.05	.67	.54	.68	.0..	.21					9.00
%	3	7	1	1	10	6	0	2	2	44	7	6	7	3	1					100

ΔFR = ACTUAL FR-GOAL FR

Figure 2. Systems Impact vs Failure Modes for Minuteman CQAP Devices

FAILURE MODES

TECHNICAL AREA	OPEN METALLIZATION	OPEN INTERCONNECT	ALUMINUM DIFFUSION	PASSIVATION DEFECTS	TANTALUM OXIDATION	CRACKED SILICON DIE	WIRE BOND OPEN	WIRE OPEN BOND TO LEAD	INSTABLE RESISTOR WIRE	PLASTIC PACKAGE CRACKS	INTERMETALLIC FORMATION	CHEMICAL CORROSION	CONTAMINATION INDUCED FAILURE	PACKAGE HERMETICITY FAILURE	BROKEN RESISTOR WIRE	SHORTED RESISTOR WIRE
METAL DEPOSITION	X					X				X		X				
PLATING		X					X				X	X				
CAPACITOR MATERIALS			X	X								X				
SILICON MATERIALS		X			X					X	X	X				
JOINING		X			X	X	X					X	X		X	
RESISTOR WIRE METALLURGY								X				X		X	X	
PLASTIC PACKAGING MATERIALS									X			X	X			
SEMICONDUCTOR SURFACE INSTABILITY									X	X		X				
THERMAL CONDUCTANCE		X			X						X	X			X	
CONTAMINATION		X	X	X		X	X	X	X	X		X	X	X	X	
MATERIAL COMPATIBILITY		X			X	X	X					X	X		X	
DEGRADATION PROCESSES		X				X	X	X		X		X	X	X	X	
PACKAGE INTEGRITY									X			X	X			

Figure 3. Failure Modes as Related to Technical Areas

PHASE I CHARACTERIZATION OF ABNORMALITIES

PHASE II POSTULATION OF MECHANISMS

PHASE III VERIFICATION OF MECHANISMS

PHASE IV DETERMINATION OF REACTION KINETICS

PHASE V ASSESSMENT OF INTERACTIONS

PHASE VI VERIFICATION OF INTERACTIONS

Figure 4. Technical Approach

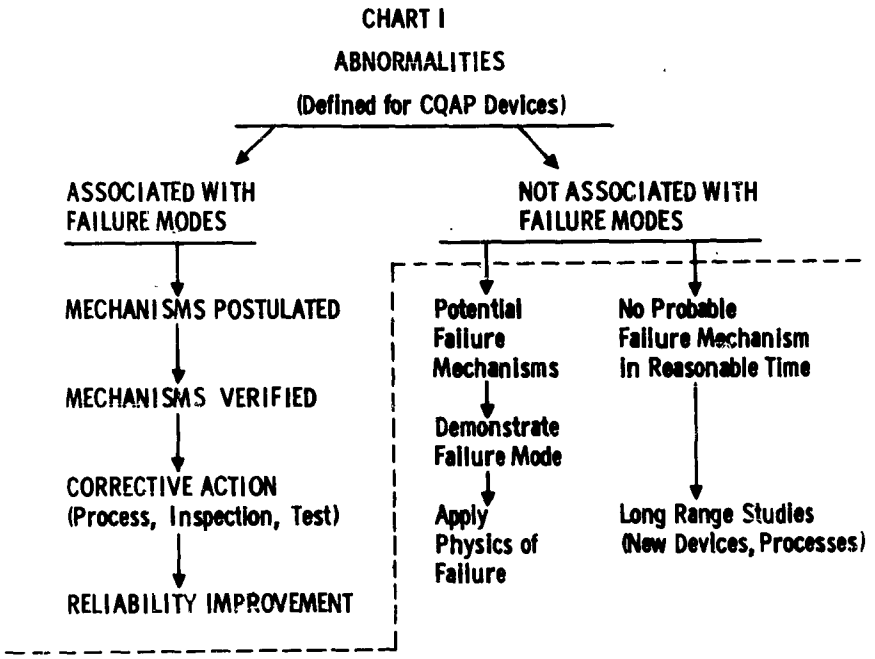


Figure 5. Types of Abnormalities Investigated

With the increase in performance requirements, yield and reliability envisioned for the future there must be an accompanying increase in sophistication in reliability techniques. This can come about only by achieving an understanding of the materials - structures and electronic properties of electronic devices on an atomic and molecular level. This is modern reliability which can become a reality via joint efforts between manufacturers, system builders and the United States Air Force.

The following papers represent highlights of the Physics of Failure Program directed towards improving the reliability of the electronic components used in the Minuteman II weapon system. Although this program was directed at specific components and was designed to respond to the reliability needs of Minuteman II, it did produce information of use to the electronic components industry in general.

FAILURE MECHANISMS ASSOCIATED WITH THERMOCOMPRESSION BONDS IN INTEGRATED CIRCUITS

G. V. Browning, L. E. Colteryahn, and D. G. Cummings

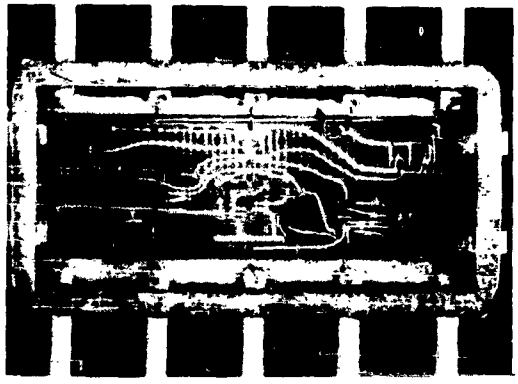
Autonetics, A Division of North American Aviation, Inc., Anaheim, California

I. INTRODUCTION

About a year ago, Autonetics encountered a problem with gold-aluminum thermocompression bonds between gold wire and aluminum metalization on SiO_x-silicon dice in integrated circuit devices (Figure 1). A tendency for bonds to become electrically open was the apparent problem encountered. The open condition was observed to be permanent in some devices and intermittent in others. As part of the Minuteman II Physics of Failure Program, an investigation was undertaken to understand the cause of such failures and then to propose suitable corrective action to increase component reliability.

Complete electrically open bonds were not difficult to recognize since a slight lift on the gold wire or prodding of the bond on opened devices permitted observation of bonds detached from the bonding pad (Figure 2). Intermittent behavior, however, was not recognized until it was observed that some open bonds closed when a static charge accumulated on the case of the device. Following this tenuous lead further, intermittent behavior was observed to occur rather frequently, was established to be dependent upon voltage impressed during electrical testing, was influenced by test temperature, and showed some time dependency. Some bonds could be made to behave intermittently almost at will, thus virtually making them act as a mechanical on-off switch.

In spite of the ability to induce intermittent behavior readily, the susceptibility of good (electrically closed) bonds to become open or intermittent could not be predicted from electrical test data. Slowly increasing the applied voltage to a fully rated value and above, did not yield evidence that showed any correlation with subsequent behavior. Because electrical tests were not discriminatory, a decision was made



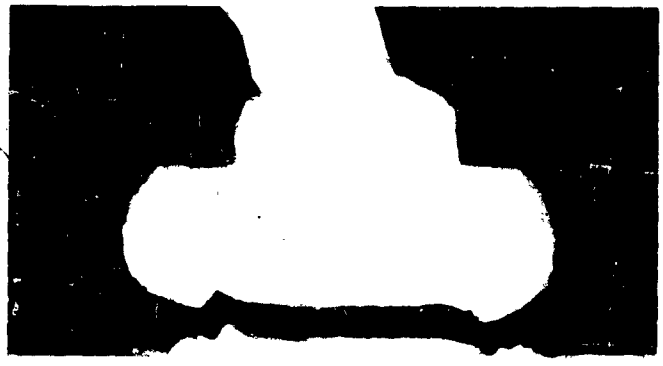


Figure 2

Open Thermocompression Bond — 750X

to conduct a physics of failure type investigation of both good and bad bonds in the same device as well as in different devices to determine, if possible, the cause of failure.

II. MATERIALS AND EXPERIMENTAL PROCEDURES

Integrated circuit electronic devices from the Minuteman II program provided the bulk of specimens examined in this investigation. Some high-power switches from Minuteman II, and a few transistor devices from Minuteman I were also included in the metallurgical examination. The examination of production devices was supplemented by an extensive examination of Au-Al TC bonds made in Autonetics M&P Laboratories so as to simulate processes and tests used in device manufacture.

Standard metallographic preparation procedures were used to polish bonds in cross-section (Figures 3 and 4). Normally employed and special etching techniques were used in order to reveal the microstructure of the intermetallic bond. Photomicrographs were made with Polaroid film, both black and white and in color. Subsequently, other techniques were used to clarify results obtained metallographically. Several Au-Al alloy compositions were made by vacuum fusion to provide "standard" materials needed for phase identification. X-ray diffraction and electron beam microprobe techniques were used to determine as rigorously as possible the composition of the various microconstituents observed in the "standard" vacuum fusion alloys as well as those in the intermetallic band.

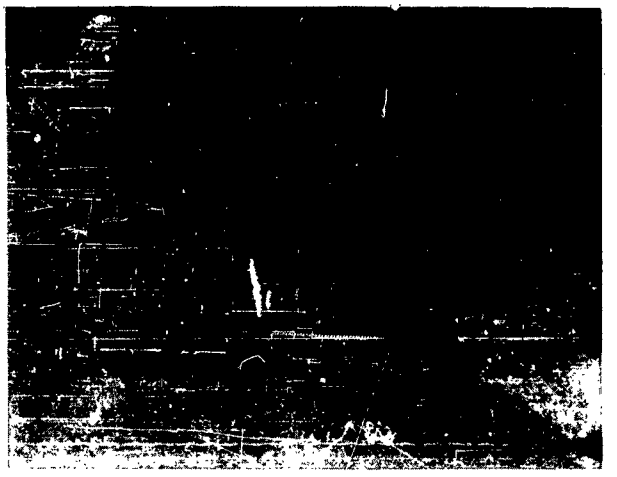


Figure 3

Open Thermocompression Bonds — 200X

Figure 4

Residue From Thermocompression Bond - 1000X

III. RESULTS AND DISCUSSION

A summary of device processing history is provided in Table 1. With the exception of device burn-in procedure, the material and processing variables were exhaustively investigated in laboratory experiments.

The metallurgical work clearly showed quite early in the investigation that TC bond failures were in all probability associated with a band of intermetallic compounds. Until a suitable etching reagent to delineate the various constituents was found during this investigation, the constitution of the bond was imagined to be a single-phased aluminum-rich region (Figure 5). Etching with various aluminum etching reagents, however, did not attack the band. This lack of etching suggested that the intermetallic band was probably composed of gold-rich phases. Furthermore, thorough microscopic examination of open and intermittently open bonds showed that a crack, void, or discontinuity of some kind was visible at the outer edges. There was some microscopic evidence that a discontinuity was entering the intermetallic band near mid-thickness rather than along the interface between the band and the unaffected gold mass.

The observed location of the discontinuity was additional indication that the band was of multiphase structure. A suitable etching reagent was needed to define the new structure. On the hypothesis that the band was composed of gold-rich (differentiating on the basis of atom percent Au) constituents, specifically the intermetallic compounds Au_4Al and Au_5Al_2, several reagents used for gold and gold alloys were tried.

Table 1

TC Bonding Considerations

Supplier	Gold Wire Size (Mils)	Bonding Parameters				Other Variable Factors
		Aluminum Thickness (Å)	Capillary Tool Temp (°C)	Heated Stage Temp (°C)	Bonding Force (gm)	
A	1.0	8,000	Cold	325	45	1. Si Oxide
B	2.0	80,000	Cold	310	150	2. Al Deposition Temp
	5.0	80,000	Cold	310	70	3. Al Thermal Treatment
C	1.0	10,000	260	*	*	4. Total Bonding Time
D	1.0	8,000	Cold	280	100	5. Dwell Time
E	1.0	7,000	140	310	50	6. Bonding Sequence
						7. Hermetic Sealing
						8. Stabilization
						9. Burn-in
						10. A/N Processing
*Information not presently available.						11. System Use

Figure 5

Polished Cross-Section of TC Bonds - Unetched

Based on knowledge at Autonetics and its suppliers, extensive examination of this structure was carried out for the first time (Figures 6 and 7). The condition was observed in all bonds, to more or less the same degree, in a device and subsequently was found to be a general condition in all devices regardless of small differences in manufacturing techniques. Between 500 and 1000 bonds were cross-sectioned, polished, etched, and examined to establish the generality of the observed multiphase condition of the intermetallic band. The universality of the multiphase condition indicated that the real crux of the bond failure problem was the nature of the metallurgy of the gold-aluminum system (Figure 8). Further insight into the problem depended upon identification of the phases observed and a knowledge of the kinetics of the reactions leading to their formation.

Alloys synthesized to produce the various intermetallic compounds represented in the gold-aluminum phase diagram were identified for structure by x-ray diffraction techniques. Electron beam microprobe emission characteristics of the intermetallic compounds were also established with specimens from these "standard" alloys. Microprobe emission patterns obtained by scanning the standard alloys were compared with those obtained by scanning polished cross-sections of TC bands. In this manner, the identity of the various phases constituting the intermetallic band was deduced within the limits of accuracy of resolution of the microprobe. This work confirmed the presence of the previously predicted Au_4Al and Au_5Al_2 intermetallic compounds in the band and also the compound Au_2Al. It is interesting to note that very little, if any, evidence was found for the compounds AuAl, $AuAl_2$

Figure 6

Polished Cross-Section of TC Bonds – Etched

Figure 7

Thermocompression Bond, Etched – 750X

(commonly known as "purple plague" among device manufacturers and users), or unreacted aluminum. This does not imply that AuAl, $AuAl_2$, or unreacted Al are not present but simply that, if present, they are beyond the resolution capability of the electron beam of the microprobe apparatus. Furthermore, no detectable amount of Si was found in any of the bonds scanned with the microprobe.

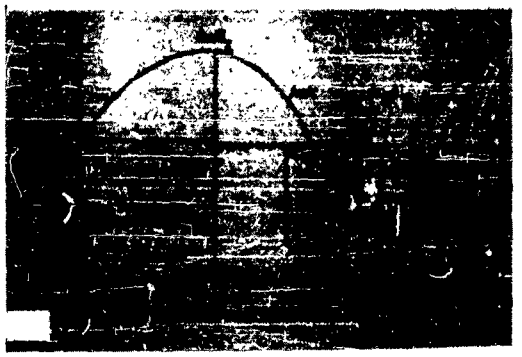


Figure 8

Gold-Aluminum Equilibrium Diagram

Another condition that was apparently unrelated to intermetallic formation and bond failure was observed in many devices: A blackening of the aluminum bonding pad and conductor stripe near the TC bond (Figure 9). This condition could be reproduced in the laboratory by holding a device at an elevated temperature (laboratory experiments were conducted at temperatures of 400 C to 500 C for several hours). The blackening was found to be caused by gold diffusing outward from the TC bond, reacting with aluminum to form intermetallic compounds, the same as observed in the bond, which caused a very uneven or jagged surface condition that randomly dispersed incident light. A unique feature of this migration or diffusion of gold was that the leading edge of the black area showed visible evidence of $AuAl_2$ in polished cross-sections (Figure 10). The significance of gold diffusion, away from the TC bond, lies in the fact that a mechanism is provided that explains the enlargement of cracks and voids around the periphery of the bond (Figure 11). A microcrack is thought to exist around the periphery because during the last few microseconds of the bonding operation, insufficient interdiffusion of gold and aluminum occurs for a metallurgical bond to obtain. The original microcrack becomes enlarged because of a large difference in diffusion rates of the two atom species; thus, it is an example of the Kirkendall effect.

With the information obtained from extensive metallographic examination of polished and etched bond cross-sections, phase identification, and gold diffusion away from the TC bond, a failure mechanism can be postulated. TC bonds are most assuredly the result of a diffusion-controlled reaction between the bonding wire and the bonding pad metalization. TC bonding of gold wire to aluminum pads, of course, is no exception, and the bonding operation is one of gold-aluminum interdiffusion. Gold diffusion continues even after the aluminum beneath the bond is consumed. Evidence of continued gold

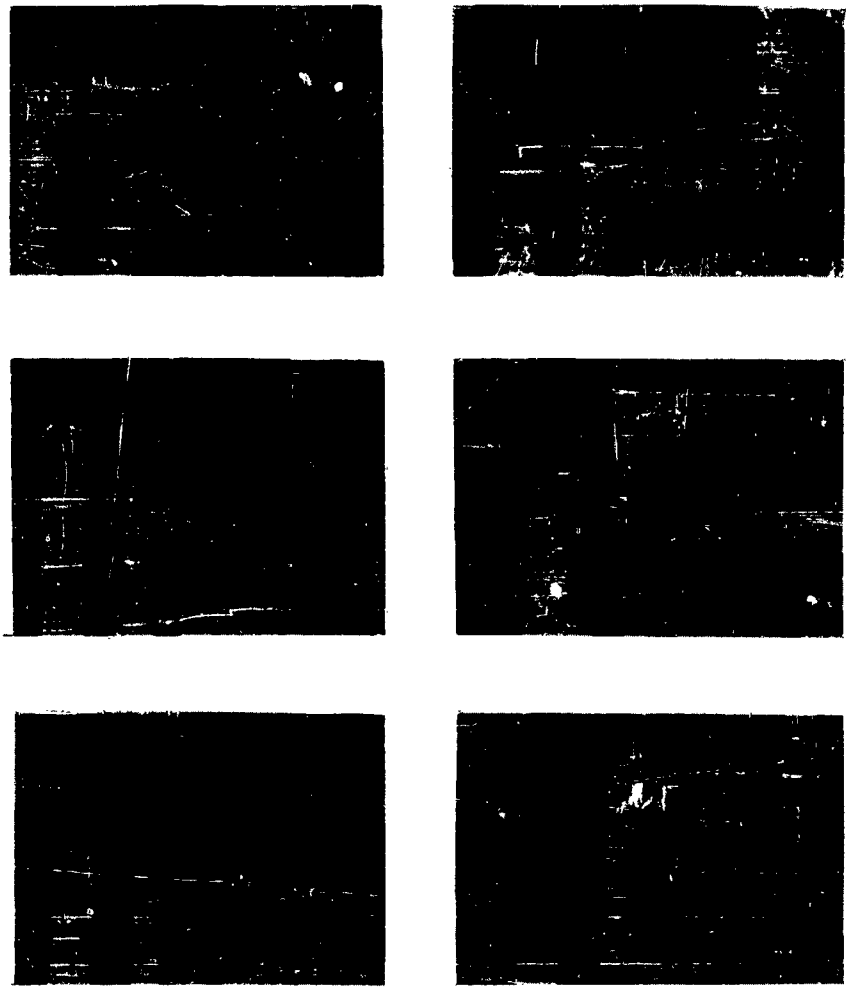


Figure 9

Time-Lapse Photograph Showing Surface Migration

Figure 10

Diffusion of Gold From TC Bond Along Aluminum Conductor Strip at About 400 C for 647 Hours, Unetched - 1500X

diffusion was obtained from cross-sections of TC bonds having been held at the bonding temperature for times from 1 to 30 minutes Figure (12). The intermetallic band was observed to grow until it reached its maximum thickness in 5 to 7 minutes. Further holding caused the onset and growth of a second phase next to the gold ball with no further increase in band thickness; thus, gold is believed to be diffusing into the original phase to cause a transformation to occur (Figure 13). Meanwhile, aluminum atoms are still available around the periphery of the gold ball from the bonding pad and the conducting stripe. This source of Al serves to retain the original phase only around the inside edge of the gold ball. Thus, we can postulate that all of the gold-aluminum phases form and are present during the first stages of the bonding process. Further diffusion of gold atoms into the intermetallic band will transform the aluminum-rich phases progressively to more gold-rich phases. The rates of formation of the various phases and the relative amounts of each at any point in this process of phase growth and transformation will depend upon the diffusion coefficient of gold and aluminum in the phases and the stability of each. Thermodynamic and metallurgical considerations require that a diffusion couple in a two component system consist of a set of single-phase layers occurring in a sequence corresponding to their location in the appropriate phase diagram. The single-phase regions in the gold-aluminum system are aluminum (solid solution of gold in aluminum), $AuAl_2$, AuAl, Au_2Al, Au_5Al_2, Au_4Al, and gold (solid solution of aluminum in gold). As soon as the aluminum beneath the bond is consumed, the processes are gold diffusion controlled

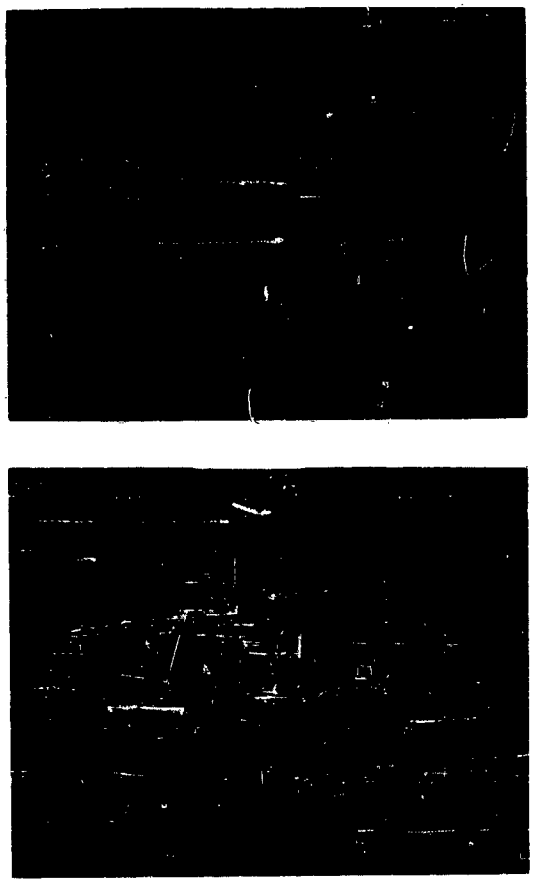


Figure 11

Open TC Bond Due to Annular Crack – 250 C for ~120 Hours

(except for the edge of the bond where a source of aluminum still exists), and the transformation equations are as follows:

$$2Al\ (Au\ s.s.) + Au \rightleftharpoons AuAl_2\ (\Delta H_1)$$

$$AuAl_2 + Au \rightleftharpoons 2AuAl\ (\Delta H_2)$$

$$AuAl + Au \rightleftharpoons Au_2Al\ (\Delta H_3)$$

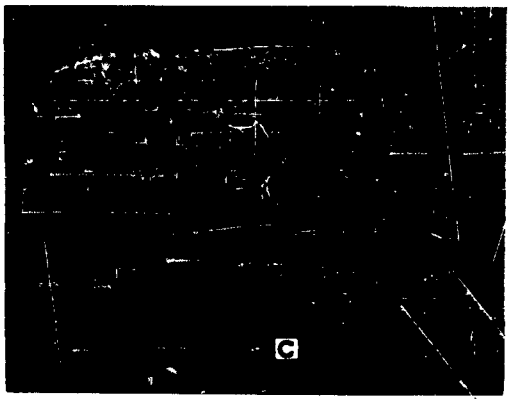


Figure 12

Bond Made in Laboratory After 6 Minutes at 300 C — 1770X

Figure 13

Bond Made in Laboratory After 21 Minutes at 300 C — 1770X

$$2Au_2Al + Au \rightleftharpoons Au_5Al_2 \ (\Delta H_4)$$

$$Au_5Al_2 + 3Au \rightleftharpoons 2Au_4Al \ (\Delta H_5)$$

$$Au_4Al + nAu \rightleftharpoons Au(Al \text{ s. s.}) \ (\Delta H_6)$$

The relative amounts of each phase will depend on the values (and sign) of ΔH, to some extent on the stoichiometry of the transformation, and on the rate of diffusion of gold through the preceding phases. It is known that one reaction - gold and aluminum to form probably $AuAl_2$ - is highly exothermic, but the relative values of the - ΔF (free energy heat of formation) of each of the intermetallics are unknown. The - ΔF's may probably, of course, change in different manners with temperature, and it is this fact that would account for the various combinations and relative amounts of the several intermetallics observed in bonds made at different temperatures and subjected to a wide variety of heat treatments. This transformation is accompanied by appropriate changes in crystallographic structure by rearrangement of atom positions and the formation of phase interfaces. It is not improbable that the crystallographic mismatch between two adjacent phases is so great that the resulting phase interface is quite highly strained.(Figure 14). Furthermore, the various mechanical and physical properties will be different. The most important of these properties, insofar as devices are concerned, are relative density and relative expansion. Both can cause additional stresses to be superposed on the weak interface. The induced stresses can initiate and propagate fracture to complete failure. This is presumed to cause bonds to become completely open by detachment.

SCHEMATIC REPRESENTATION OF STRAIN LINES INDUCED BY VOLUMETRIC CHANGES DUE TO PHASE FORMATION

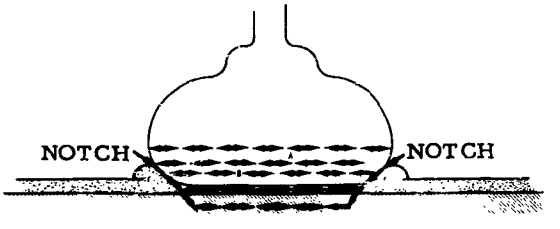


Figure 14.

Strain Lines Induced by Volumetric Changes Due to Phase Formation

The mechanism whereby a bond becomes open or intermittent in electrical behavior but mechanical integrity is retained has already been described – the diffusion of gold from the bond to form or enlarge cracks or voids. When a void extends from the outer periphery of the bond inward and downward to the SiO_x film on the silicon die, electrical isolation from the conductor stripe and the active region will occur (Figure 15). If the void gap is small, thermal fluctuations could cause intermittent contact to be made.

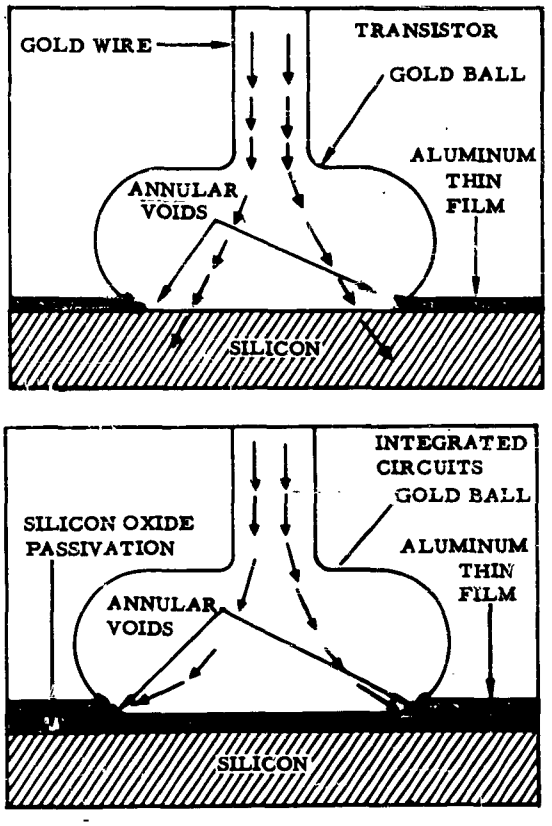


Figure 15.

Typical TC Bond Configurations for Transistors (Top) and ICs

Another facet of this investigation was to explain the difference in failure rate between Au-Al bonds over silicon in discrete devices used for Minuteman I Program and bonds of integrated circuits. Two plausible explanations are offered. One concerns the difference in the size of the wire used and the bond formed. In discrete devices, especially power transistors and switches, up to 5-mil diameter wire was generally used. The diameter of the resulting bond was about 5- to 15-mil diameter or four to five times that of bonds on integrated circuits (1-mil diameter forming a 3- to 4-mil diameter bond). The greater mass of the bonds in discrete devices combined with their larger contact area could cause a greater heat-sink effect and somewhat lower temperature during bonding even though stage and capillary temperatures of the bonder were exactly the same. The second explanation of the superior behavior in discrete devices in reality concerns the design of the interconnections. In discrete components, the bonds are made to aluminum evaporated directly over silicon. Thus, the electrical path follows along a line from silicon die directly into the bond through the intermetallic band and along the wire. In IC's, however, bonding directly over active areas was abandoned in favor of the expanded contact concept; the bond is remote from the active area, and electrical continuity is provided by the evaporated Al stripe. The manner in which gold migrates along the conductor stripe causing an isolating void to form around the periphery of the bond has already been noted. Gold diffusion away from the bond could occur in discrete devices, but this mechanism would not cause electrical opens inasmuch as the bonds make direct ohmic contact with the silicon. The expanded contact is an example of designing a favorable condition for failure into devices because of a lack of awareness of the vital role materials of construction can have.

Examination of TC bonds that have failed in discrete devices has shown: (1) open bonds are generally completely detached; (2) the complete gold-aluminum phase diagram has been observed in quite a few devices, and (3) the same failure mechanism is equally applicable to bond failures in discrete devices as well as those in IC's. Electron microprobe scanning of the surface remnants of detached bonds in several transistors showed that fracture was associated with the gold-rich phases as previously described. This was also found to be true of small tension test specimens prepared from a diffusion couple; fracture was observed to occur in the gold-rich region.

The postulated failure mechanism appears valid for other metal combinations in addition to gold and aluminum. For instance, one type Minuteman II transistor was observed to have failed in the brazement joining the silicon chip to the copper pedestal. To make the composite of materials somewhat more compatible with silicon, a heavy film of molybdenum was evaporated onto the surface of the copper.

A flash of nickel was then applied to the molybdenum surface for adherence and wettability. A gold-silicon eutectic brazing alloy was used to attach the silicon die to the pedestal. The location of the failure was along the interface between the nickel flash and the eutectic brazing alloy. Metallographic examination indicated an almost continuous line of voids interconnected by the fracture. Electron microprobe indicated the presence of a nickel-silicon intermetallic compound, Ni_3Si_2. The depletion of a thin zone in the eutectic braze alloy of Si atoms without replacement by Ni atoms caused the formation of pores and voids because of the Kirkendall effect. This failure is another example of a design deficiency wherein the material selection was made without sufficient evidence that the combination would be metallurgically compatible.

Since it appears that the Au-Al system under such conditions is inherently unreliable, it would be of interest to discuss what the manufacturers have done in response to the findings discussed above; and further, it is important to determine, if possible, how reliable gold-aluminum TC bonds are.

An obvious solution would be to use some substitute for Al, Au, or both. Elimination of aluminum would eliminate all of the advantages of using aluminum, viz, the extremely good adherence to oxide films, its ability to make excellent ohmic contact with silicon, chemical inertness, and its economic advantages. A substitution could be made for the gold wire as has been done by one manufacturer who uses aluminum wire. Not only is there some difficulty in making an aluminum-to-aluminum bond, but in order to circumvent transfer of the problem area from the die to gold plated Kovar output lead-throughs, the protective gold plating is replaced by a nickel-aluminum film. Evaporation of nickel and aluminum is a two-step process and requires masking off the remainder of the package interior so that only the Kovar lead-throughs are coated. A simple method of preventing Au-Al reaction is the evaporation of a diffusion barrier metal between the aluminum film and a gold film. TC bonding is then made gold-to-gold. The potential barrier metals obviously must be selected with care so as not to create combinations just as reactive as the Au-Al combination. Several barrier metals have been introduced by different device manufacturers. One of the best turns out to be molybdenum. Palladium has been suggested but as a lead replacement rather than a barrier material, and it is being evaluated. One bond made with platinum gave indications of intermetallic formation. Metallographic examination of sample devices with diffusion barrier layers indicates that pinholes in the barrier metal and film overlap may be the most serious causes of subsequent failures. This defect can be eliminated by proper manufacturing controls.

Some of the work performed at Autonetics indicates that a short, high-temperature treatment of Au-Al bonds may promote the formation of the solid solution of aluminum in gold. Consideration on the lines of the earlier comments on the driving force of these reactions might be a useful approach. This should be ideal insofar as the metallurgy of the TC bond is concerned. Some risk might result, however, from gold atoms poisoning the junction area to cause later failure. This would only occur, however, if the gold migrated all along the aluminum conductor from the pad to the ohmic contact point since at the pad area, silicon oxide is a barrier between gold and silicon.

Inasmuch as Au-Al TC bond failures have been shown to be time-temperature sensitive, one naturally is concerned about meeting the reliability requirements for operational equipment. Knowing the mechanism of failure and its acceleration by high-temperature treatments, it was believed that data obtained at elevated temperatures could be extrapolated with reasonable confidence and permit projection of a failure rate.

For the purpose of reliability projection, TC bonds subjected to different processing were exposed to various temperatures. The thermal processing of Group I devices (bonds) included a prebake treatment at 425 C in dry nitrogen (-60 F dew point) followed by a stabilization treatment at 200 C for 96 hours. Thermal processing of Group II devices (bonds) included a prebake at 425 C in air followed by a stabilization treatment at 150 C for 120 hours.

The test temperature-time matrix along with the number of bonds exposed at each temperature and the test results are shown in Table 2. A temperature as near as possible to the meximum junction operating temperature anticipated was included to give further validity to the entire investigation.

The method used to measure the integrity of the bonds was to observe diode breakdown characteristics of each bond with a curve tracer as the voltage was increased stepwise to 1.04 volts, maximum, with the current limited to 50 microamperes.

A statistical analysis of the test data for Group 1 devices at 275 C and 200 C showed that the failure rate was indeed exponential. Extrapolation of this data by use of the exponential relationship (Arrhenius) showed the bond reliability of Group I devices would be 0.00097 percent per 1000 hours at a temperature of 50 C (the maximum junction operating temperature anticipated). A comparison of failure rate of Group I and Group II devices exposed to the same temperature shows Group II devices to be about 120 times superior. Using this factor,

Table 2.

Results of Temperature-Time Matrix Study

Bond Group	Number of Bonds	Cell Temperature, C	Test Hours	Bond Hours	Comments
I	5200	275	48	249,600	Test terminated (all failed)
I	1300	200	2000	2,600,000	Test terminated (all failed)
II	228	200	2870	659,408	2 failed
I*	1100	125	2880	3,168,000	0 failed
I*	2900	85	2880	8,352,000	0 failed
II*	1200	85	2880	3,456,000	0 failed

*Tests performed by suppliers

the failure rate of Group II devices at 50 C can be estimated to be approximately 0.000008 percent per 1000 hours (Figure 16). Or in a ten-lead device, the failure rate of bonds would be expected to be 0.00008 percent per 1000 hours or about 1/3 the rate originally estimated for bonds at the very onset of the Minuteman II program.

In summary, gold-aluminum interconnects in integrated circuit devices display a time-temperature dependent mode of failure. It is the result of interdiffusion of gold and aluminum and the tendency to form gold-rich intermetallic compounds at elevated temperatures. There are some indications that air treatment causing oxidation of the pad and the associated circuitry has some retarding effect. Use of other metals, such as molybdenum, to act as diffusion barriers further retards interdiffusion. Temperature variation, static charges, and normal operating voltages can cause intermittent closure of open bonds.

IV. ACKNOWLEDGEMENT

The Autonetics Division of North American Aviation, Inc., wishes to express its appreciation to its suppliers of integrated circuit devices for their contributions in the solution to this problem. The support received from the Ballistic Systems Division and Rome Air Development Center of the United States Air Force for their vital and timely interest requires special acknowledgement.

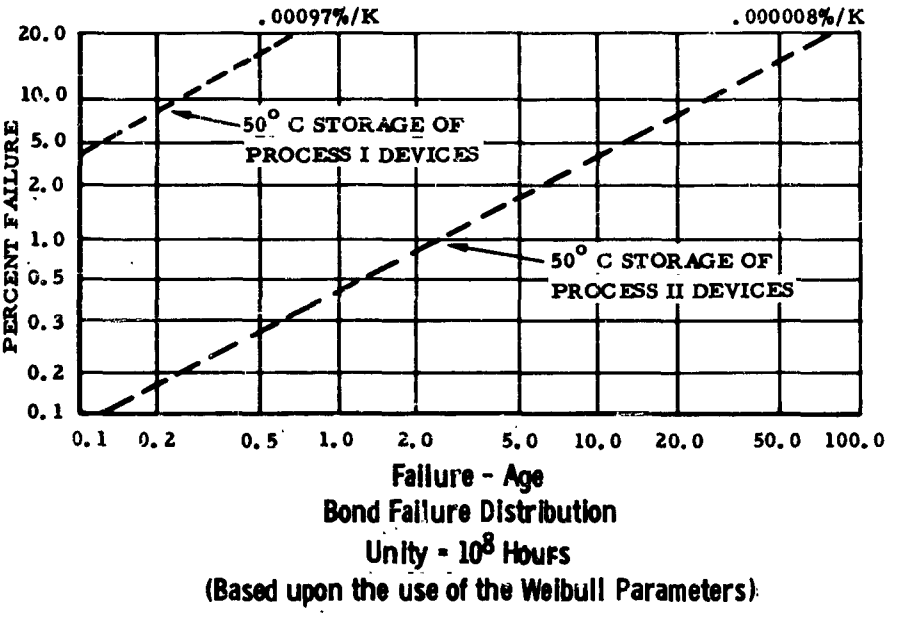


Figure 16.

Bond Failure Distribution

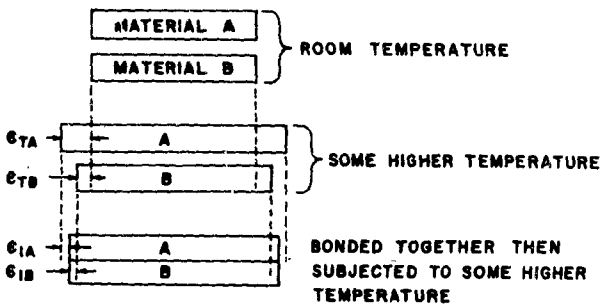
FAILURE MECHANISMS ASSOCIATED WITH THERMALLY INDUCED MECHANICAL STRESS IN MINUTEMAN DEVICES

C. G. Jennings

Autonetics, A Division of North American Aviation, Inc.,
Anaheim, California

Semiconductor devices are subjected to a considerable number of temperature changes during fabrication, screening tests, and storage. Operational mechanical strains in the composite materials from which the semiconductor devices are made are induced as a result of this temperature change due to incompatible rates of thermal expansion in adjacent materials (Ref 1). Two materials of the same length have different lengths at some other temperature due to different rates of thermal expansion as shown in Figure 1. When these two materials are attached to each other in some manner, mechanical strains are set up in each material as they are forced to expand with temperature at the same rate. This rate is different from either one of their unrestrained rates of thermal expansion. The analytical expression for mechanical stress is also given in Figure 1 assuming no bending or plastic strain in either material (Ref 2). Note that the stress is a function of the difference between the two materials' coefficients of expansion, the change in temperature, the respective moduli of elasticity, and the thicknesses. The experimental mechanical stress is determined by measuring the mechanical strain and converting it to stress utilizing Hooke's Law, wherein the stress is directly proportional to the strain and the modulus of elasticity.

All semiconductor devices experience a thermally induced mechanical stress in one degree or another. Generally speaking, devices are made of materials with as close a rate of thermal expansion to each other as possible to thus minimize the amount of thermally induced stress. Power transistors, however, require that the heat generated during operation be conducted away from the die. Materials that will conduct the heat away generally have a much higher coefficient of thermal expansion than the semiconducting materials (especially silicon). A good example of this material incompatibility is shown



$$\epsilon_{TA} - \epsilon_{IA} = \epsilon_{TB} + \epsilon_{IB}$$

ANALYTICAL EXPRESSION FOR STRESS

$$\sigma_A = \frac{E_A(\gamma_B - \gamma_A)\Delta T}{1 + \frac{E_A T_A}{E_B T_B}}$$

Figure 1

Thermally Induced Stress Due to Mismatched Linear Coefficients of Thermal Expansion

in Figure 2 in the cross-section of a power transistor. The large mass of copper used to conduct the heat away has a coefficient of expansion of approximately 16 ppm/C, while the thinner molybdenum pedestal and the silicon die have coefficients of expansion of 4.8 and 2.0 ppm/C, respectively. The calculated stress in the silicon die of this device utilizing the equation in Figure 1 is 23,000 psi for a 100 C temperature change.

The experiment stress in the silicon of this power transistor was determined with resistance strain gages adhesively bonded to the die as shown in Figure 3. The strain causes a change in resistance in the strain gage associated with the stress which unbalances a Wheatstone Bridge. These very small gages (0.015 inch gage length) were bonded in the center of the die as shown, as well as near the edge, to determine where the strain was maximum. The maximum strain occurred at the edge, as predicted in Ref 3. The resistance temperature sensor shown in Figure 3 was used to accurately measure the temperature of the device at the time the strain was measured, as well as to insure that temperature equilibrium had been reached. The principles of operation are the same for the strain and temperature gages. The temperature gage can sense temperature changes of 0.01 F.

The environmental chamber used for temperature cycling and the strain indicator used to measure the strain from the gages is shown

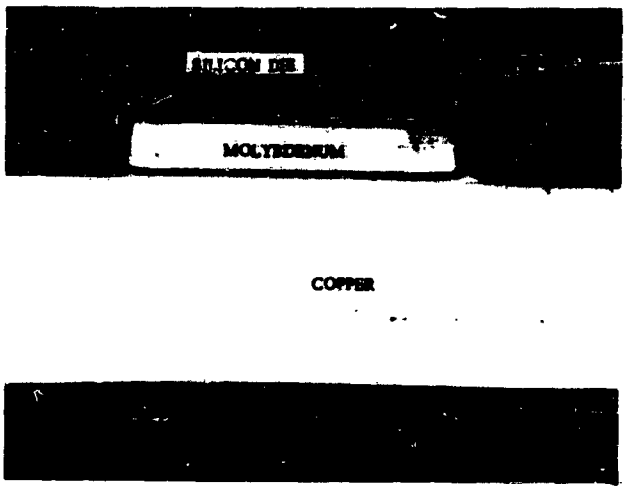


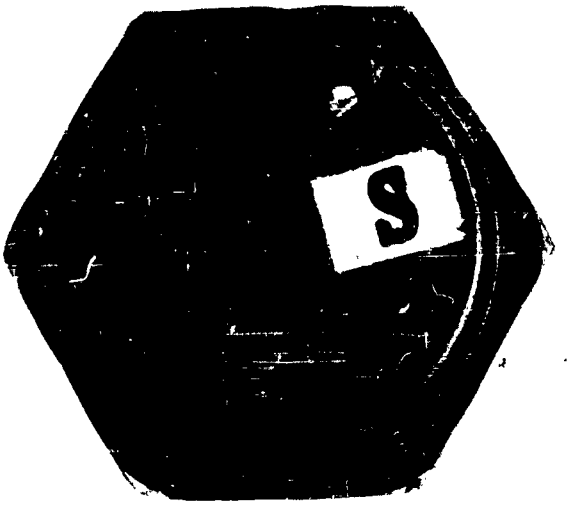
Figure 2

Cross-Section of Power Transistor
Showing Relative Thicknesses of Copper,
Molybdenum, and Silicon

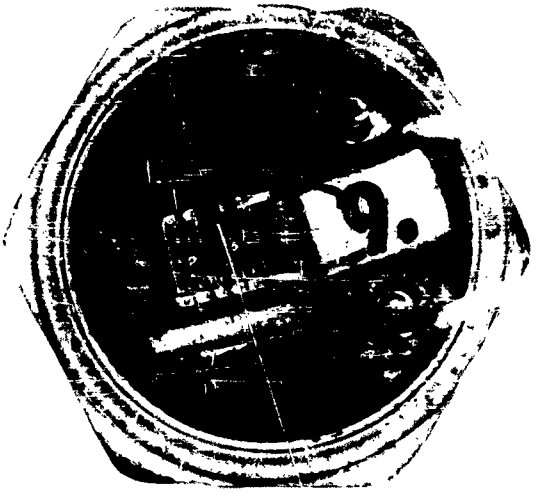
in Figure 4. The leads from the strain gage mounted on the device are brought through the outlet in the chamber wall and connected to the instrument. The temperature control on the chamber is ±3 C at both elevated temperatures (+175 C) and cold temperatures (-75 C).

The maximum thermally induced stress observed in this power transistor during temperature cycling with the resistance strain gages is shown in Figure 5. These stresses were observed near the edge of the silicon. The bar chart shows a tensile stress of 11,000 psi at 150 C. Upon returning to room temperature, a 1000 psi residual compressive stress was present. This was attributed to plastic flow in the brazement at the elevated temperature. The 13,000 psi compressive stress observed at the -55 C temperature was also followed by a tensile residual stress due to plastic flow in the opposite direction. The fact that the experimental stress is considerably lower than that calculated by the analytical method is a result of the plastic flow.

The plastic flow in the die-to-header brazement reduced the stress to a level which would not ordinarily cause failure. Cracked dies and ruptured die-to-header brazements, however, were reported in the failure analysis reports submitted by the supplier of this transistor as part of a combined Component Quality Assurance Program (CQAP). Metallographic examination of the same transistors exposed



a. Resistance Temperature Sensor Mounted on Molybdenum Pedestal



b. Resistance Shown Gage Mounted In Center of Silicon Oil

Figure 3

Sensing Devices Mounted on Power Transistor to Accurately Measure Temperature and Strain

Figure 4

Bemco Environmental Chamber and Baldwin-Lima-Hamilton Strain Indicator Used for Measuring Thermally Induced Mechanical Strains

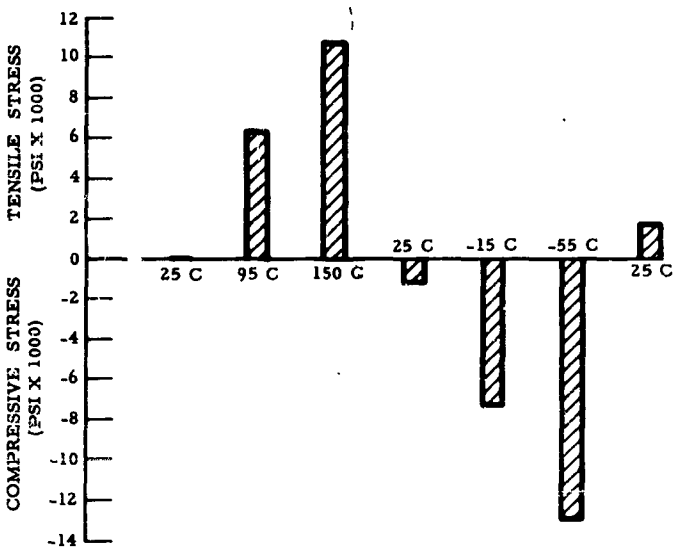


Figure 5

Thermally Induced Mechanical Stress in Silicon Die of Power Transistor When Temperature Cycled

to a selected assortment of thermal conditions indicated that these failure modes were associated with thermally induced mechanical stress. The evidence of plastic flow and abnormalities observed by the metallographic examination led to the postulation of three dependent failure mechanisms.

1. Thin brazements between die and pedestal, and pedestal and header reduces plastic flow capability.

2. Thermal cycling devices as screening technique causing strain aging in brazement.

3. Nickel-silicon intermetallic formation reduces strength and/or plastic flow capability of brazement.

It was postulated that the thermally induced mechanical stress is a fundamental failure mechanism that couples with the three dependent failure mechanisms stated above.

Special power transistors were purchased from the supplier to verify the postulation on the effect of the braze thickness on the amount of stress transmitted to the silicon die. Devices were fabricated which had a relative braze thickness of one-half and twice normal braze thickness. The photomicrographs in Figure 6 show the relative thickness of brazement extremes. Abnormal braze thicknesses in this range were observed in the metallographic examination of production devices. The effect of the braze thickness on the amount of thermally induced stress is shown in Figure 7. The change in the room temperature stress after the elevated temperature cycle is attributed to a relief of the residual die bonding stress and/or a breakdown in the die brazement. Partial ruptures have been observed after the temperature cycling especially in the thinner brazements.

The postulation verification on the effect of temperature cycling on the amount of stress transmitted to the silicon die is shown in Figure 8. The devices receiving the least severe temperature cycling treatments exhibited the lowest stress. Devices that had received 100 temperature cycles from 200 to -65 C and devices cycled 10 times between 270 and -100 C appear as a paradox in the chart. Metallographic examination of the test devices revealed more extensive ruptures in these two groups of test specimens which accounts for the lower stress state. A partially ruptured brazement will not transmit as high a thermally induced mechanical stress as a sound brazement. The stresses shown in Figure 8 are considerably lower than the stress shown in Figure 5. The stresses portrayed in Figure 8 are average stress values in the lower stressed center of the die, while the stress shown in Figure 5 is a maximum stress taken near the edge of the die.

a. Thick AuSi Eutectic Brazement (620X)

b. Thin AuSi Eutectic Brazement (620X)

Figure 6

Photomicrographs of Power Transistor Die-to-Header Brazements Showing Thickness Extremes

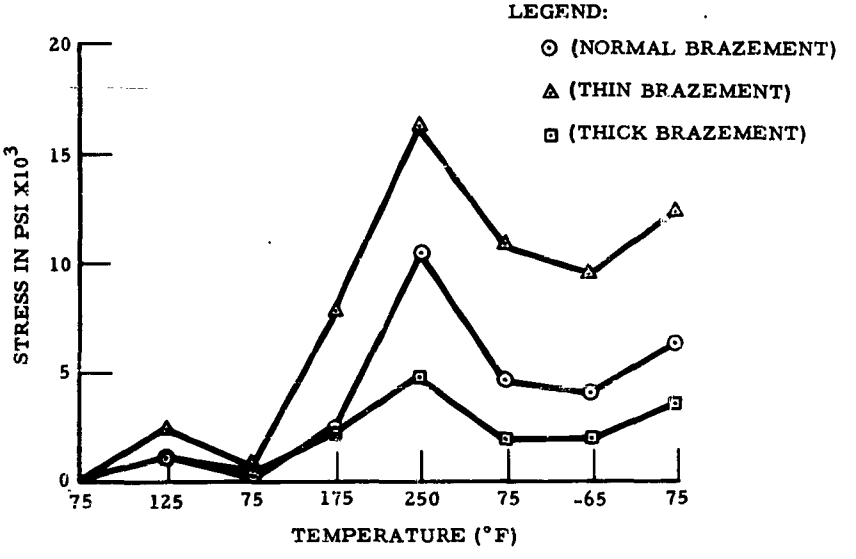


Figure 7

Average Thermally Induced Mechanical Stress in the Silicon Die of Power Transistors with Normal, Thin, and Thick Brazements

The postulation made on the nickel-silicon intermetallic formation weakening the brazement resulted from the metallographic examination of the supplier CQAP test parts. Some of the devices that had been exposed to the more extreme thermal conditions exhibit ruptures in the die-to-pedestal brazement. Figure 9 shows a high magnification of the rupture. The rupture always took place adjacent to a light blue colored compound. Electron microprobe traces, also shown in Figure 9, show the blue phase to be a nickel-silicon alloy. The presence of the nickel-silicon alloy at the rupture point prompted a characterization study. Three nickel-silicon compounds were synthesized as metallurgical standards for the electron probe. The compounds were identified by X-ray diffraction. Microhardness readings indicated all the compounds were very hard and brittle. Microprobe traces utilizing the standards identified the nickel-silicon alloy in devices as an intermetallic compound with the stoichiometric ratio of $Ni_3 Si_2$. Electron microscope replicas of diffusion samples (as well as devices) exhibited a row of very small voids at this interface as shown in Figure 10. This row of voids was attributed to the Kirkendall effect where the nickel diffuses more rapidly through the intermetallic zone than the silicon diffuses within the intermetallic zone.

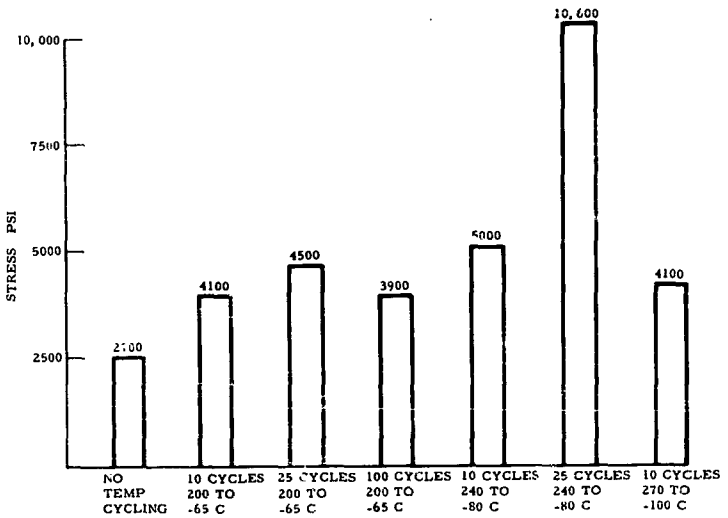


Figure 8

Thermally Induced Mechanical Stress at -65 F in Die of Power Transistor After Selected Thermal History

Power transistors similar to the ones previously discussed were also tested for thermally induced mechanical stress. The header and braze materials were the major difference between the two devices. The amount of thermally induced stress as well as fabrication stresses is given in Figure 11. The fabrication stress induced into the silicon die from cold welding the cap to the package was also measured with the resistance strain gages. An applied or elastic stress was measured during the welding operation. Also, a residual stress remained in the silicon die after the welding operation which then was algebraically added to the other stresses. Other residual stresses are induced into the silicon die which have not been completely determined such as from die bonding, silicon oxide growth, and lead forming. The total stress applied to the die is a superposition of the applied or service stress to the composite residual stresses.

Devices built by the supplier may be made of materials that are very compatible and as a result possess low residual stresses. They will also exhibit relatively low thermally induced mechanical stresses. These same devices may still be subjected to higher mechanical stresses during system manufacturing. This problem arose at Autonetics where the high density of integrated circuits (IC's) used in the Minuteman II computer required that external heat sinks be used to dissipate the heat. The IC's were bonded to the copper heat sink with a good thermal conducting organic adhesive. Mechanical stressing of the IC's was detected when a test was conducted to evaluate

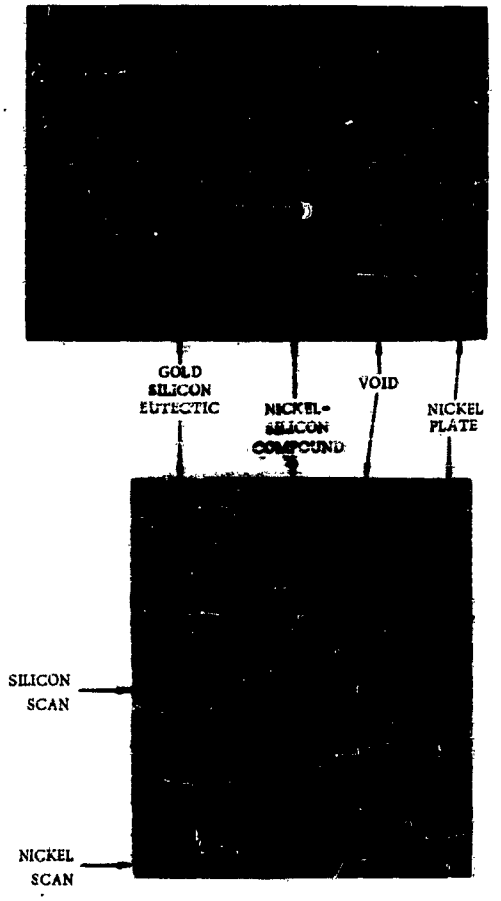


Figure 9

Photomicrograph and Electron Probe Trace on Ruptured Die-to-Header Brazement of Power Transistor (Subjected to Temperature Cycling -85 to 315 C)

a new more fragile IC package design which possessed advantages of fewer electrical shorts and better heat dissipation. The stresses were measured by installing small resistance strain gages to the silicon die which were then monitored during temperature cycling within the environmental range. Figure 12 shows the gage mounted on the top surface of the IC. The stresses were generated because the copper heat sink and multilayer board have much higher coefficients of thermal expansion (16 ppm/C) than the composite IC package (approximately 5 ppm/C) which is adhesively bonded to it.

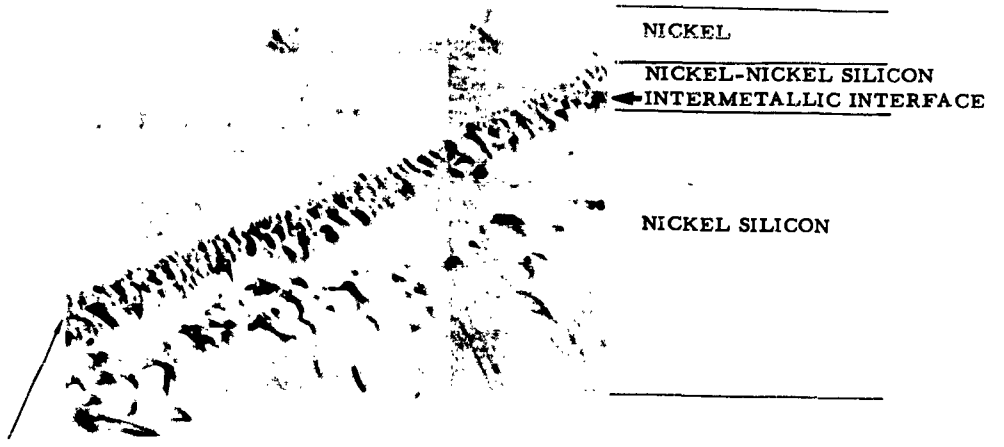


Figure 10

Electron Micrograph of Nickel-Silicon Diffusion Sample Replica

A comparison of the thermally induced stress in the silicon die of the two package designs is shown in Figure 13. The graph shows the new bare base package exhibiting the highest thermally induced stress when bonded to the heat sink with the high thermal conductive organic adhesive. The more rigid alumina pedestal package did not transmit as high a stress. Both packages were shown to transmit low stresses to the die providing they were not bonded to the heat sink, which indicates the materials used within the package were relatively compatible. A degradation of the die-to-header bond is evident by the decrease in stress from cycle 1 to cycle 3 at -65 C when bonded to the heat sink.

The results of an evaluation of other selected adhesives for bonding the IC's to the heat sink is shown in Figure 14. With the exception of devices attached to the heat sink with two flexible solvent activated adhesives, the stresses were always higher in IC's of a new package design (bare base) compared to present package design (alumina pedestal). The stress levels in the two packages bonded with the flexible solvent activated adhesives were relatively low and roughly

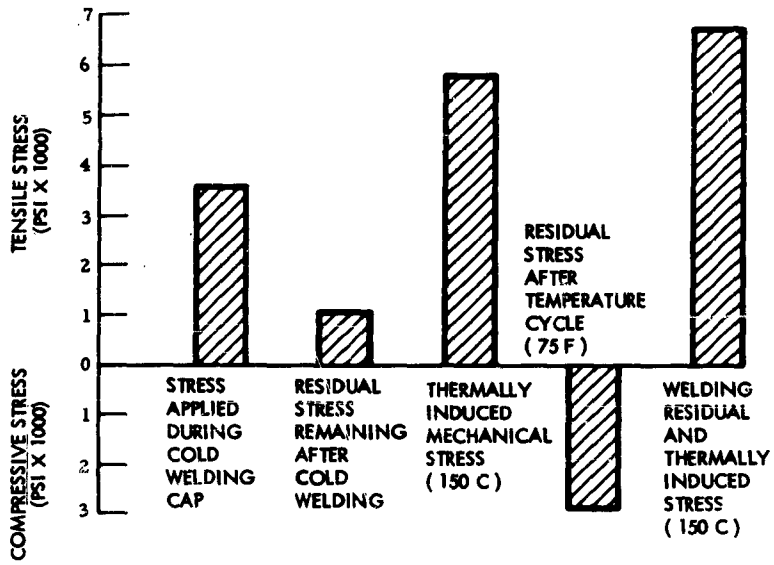


Figure 11

Mechanical Stress Induced into Silicon Die of Power Transistor From Cold Welding Cap and Temperature Cycling

equivalent to each other. The use of the more flexible adhesive makes it possible to utilize the new bare base package with its inherent advantages of fewer electrical shorts and better heat dissipation.

The most common type of failure mode attributed to the thermally induced stress in the IC's bonded to the heat sink is ruptured die-to-header bonds. Figure 15 shows the crack in the pyroceram which attached the die to the alumina pedestal in the package. This crack developed during the temperature cycling test. A similar type failure is shown in Figure 16 which shows the rupture between the pyroceram and the gold plated Kovar. The pyroceram does not adhere as well to the gold plated Kovar as it does to the alumina, and the weakest point of the bond is at this interface.

The effect of thermally induced mechanical stress has been illustrated as being an underlying failure mechanism in a number of semiconductor devices. This stress superimposes on other residual fabrication and service stresses to give a total composite stress. For the normal high reliability device, this stress is not sufficient to cause failure. It does, however, couple with other dependent failure mechanisms to

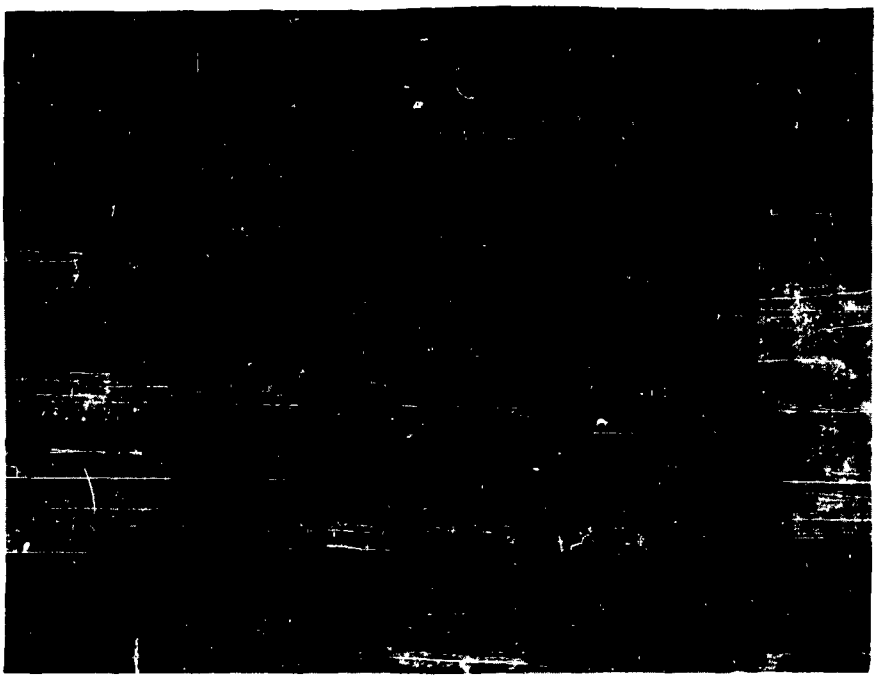


Figure 12

Integrated Circuit with Resistance Strain Gage
Mounted on the Surface (130X)

produce failures of significant importance in high reliability systems such as Minuteman.

References

1. Taylor, T.C., "A Study of Thermally Induced Cracking in Semiconductor devices," Raytheon Technical Report R-35, September 8, 1954

2. Timoshenko, S., Strength of Material, Part II, 2nd Ed., D. Van Nostrand Co., Inc., New York, 1941

3. Taylor, T.C., and Yaun, F. L. "Thermal Stress and Fracture in Shear Constrained Semiconductor Device Structure," IRE Transactions on Electron Devices, May 1962

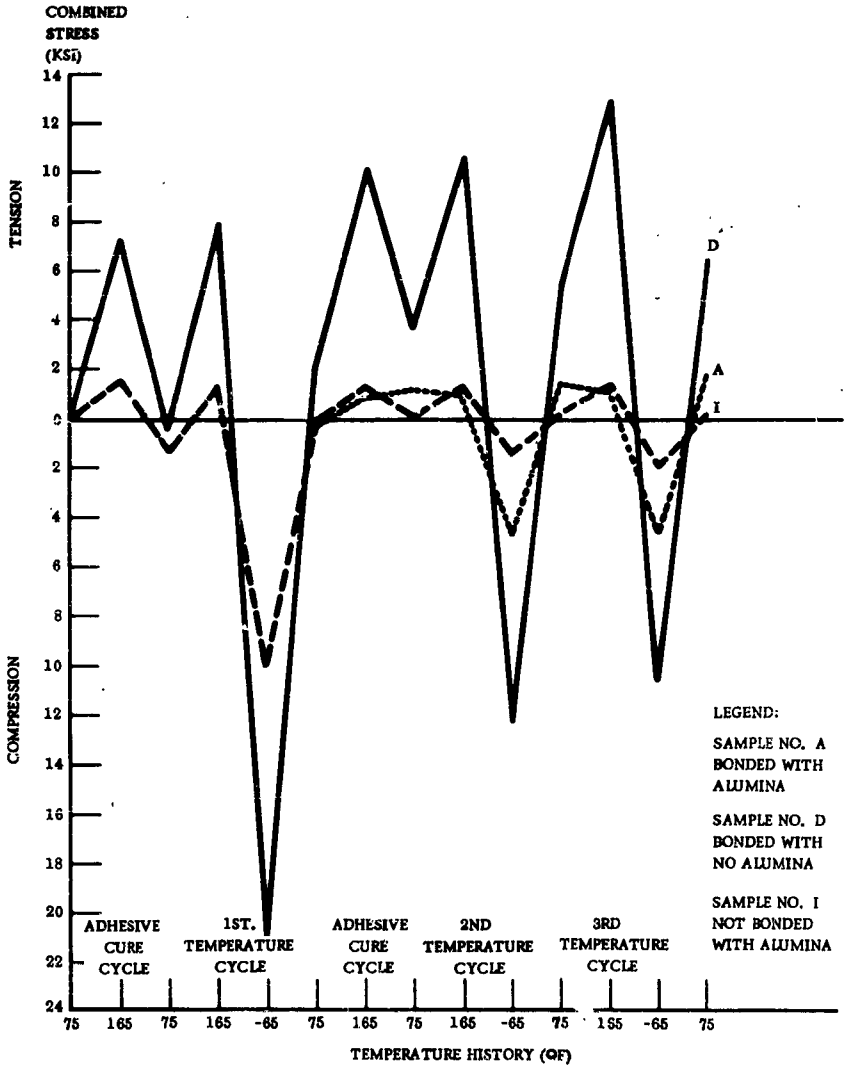


Figure 13

Combined Residual and Thermally Induced Stresses for Successive Temperature Cycles on Integrated Circuits Mounted on Copper Heat Sinks

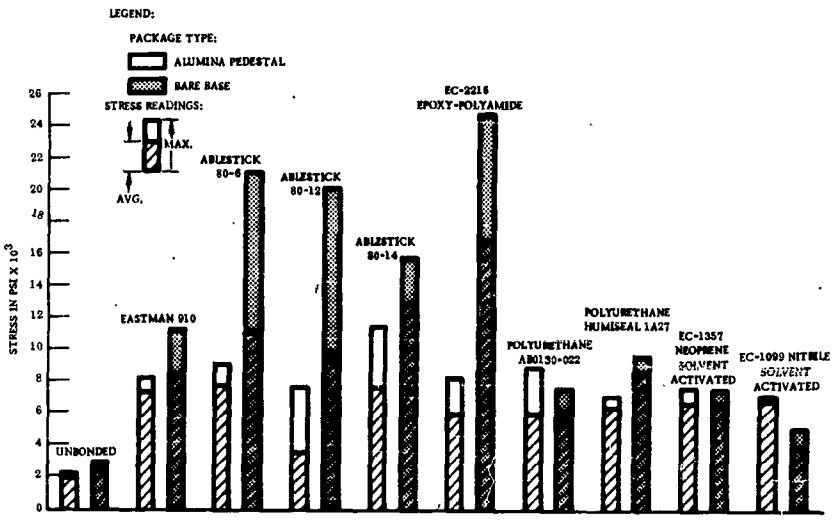
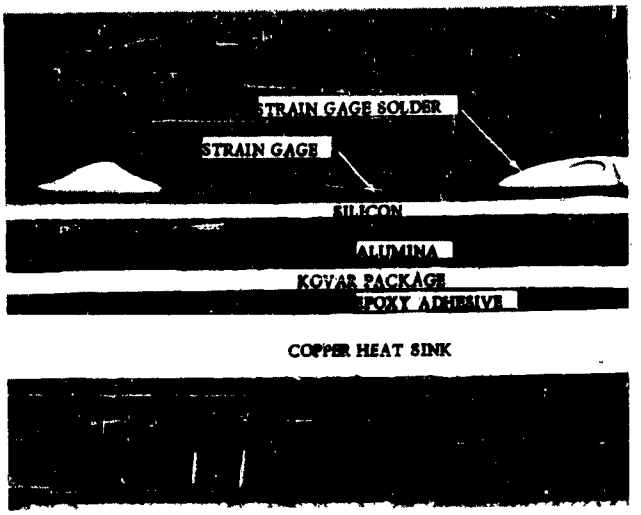
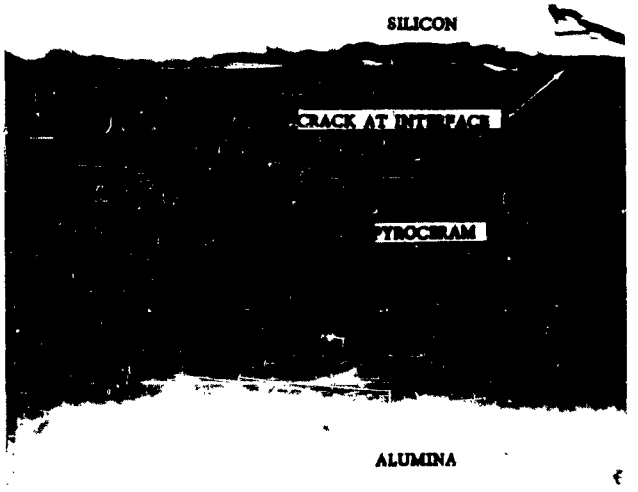


Figure 14

Comparison of Stresses Induced by a -65 F Temperature in the Silicon Die of Two Integrated Circuit Packages Bonded with Selected Adhesives to a Copper Heat Sink



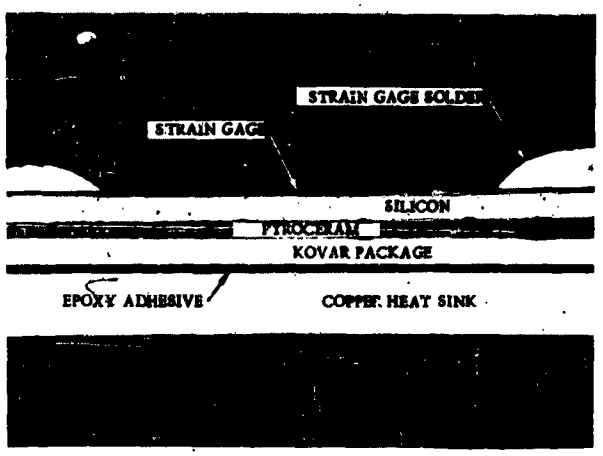
a. Cross-Section Showing Relative Material Thicknesses



b. High Magnification of Crack at Pyroceram Silicon Interface 740X

Figure 15

Photomicrographs of Alumina Pedestal Integrated Circuit Package



a. Cross-Section Showing Relative Thicknesses of Material Thicknesses (28X)

b. High Magnification Showing Void and Poor Adhesion of Pyroceram to Gold Plated Kovar

Figure 16

Photomicrographs of Bare Base
Integrated Circuit Package

PROPERTIES OF PLASTIC MATERIALS AND HOW THEY RELATE TO DEVICE FAILURE MECHANISMS

S. M. Lee, J. J. Licari, and A. Valles

Autonetics, a Division of North American Aviation, Inc.,
Anaheim, California

I. INTRODUCTION

A variety of plastic materials are in use by component manufacturers for protection of semiconductor devices. The inherent economic advantages are the main reason for their increasing usage in diodes and transistors. Plastics offer low material costs, simplicity and versatility of application techniques, and a great diversity of chemical, physical, and mechanical properties from which to choose. Their economic appeal is especially significant where large-scale production and automation are prime factors.

The optimization of organic polymer based materials for packaging semiconductor devices is a sizable task since the materials are required to perform several functions including protection against mechanical, chemical, physical, and electrical stresses. For example, the plastic encapsulant may serve to dilute or eliminate mechanical stresses applied to the leads thus preventing damage to the semiconductor element, act as a heat transfer path for the dissipation of heat, or provide a barrier against contaminants and the penetration of moisture vapor or other gases. Conversely, the plastic must not deleteriously affect the component's performance (e.g., by exerting excessive cure shrinkage stresses or by contributing injurious contaminants which may cause inversion or provide low electrical resistance shunt paths.

A survey conducted at the inception of the Autonetics Physics of Failure Program revealed a notable dearth of published information on the use of plastics over semiconductor junctions. Although it is realized that some work must remain unpublished because of a proprietary nature, it is suspected that, to a large extent, empiricism was the basis for the choice of plastics.

II. TECHNICAL APPROACH

At the beginning of the program diode abnormalities were reported to be excessive reverse leakage currents under stress conditions and poor lead wire to resin pull strength. The excessive reverse leakage currents were thought to be promoted by inversion formation at semiconductor junctions, which was hypothesized as being possibly due to contaminants contained or released by the plastic encapsulants.

The test program initially established was broad and encompassed a multitude of tests for a number of properties, including a comparison of the ionic and ionizable or gaseous contaminants contained or generated by plastics. Data obtained by probing in many directions was intended to be used in postulating failure mechanisms which could be subsequently either verified or disproven by tests on the actual components. Physicochemical property determinations were, for the most part, conducted on "synthetically fabricated specimens" since it was difficult or impossible to directly evaluate the plastic on the diode surface.

Microdiode encapsulants from several vendors were investigated and the chemical types including silicone, phenolic, and epoxy were ascertained by infrared spectroscopy. In addition, fluorinated polymers such as Teflon not ordinarily used for diode encapsulation were comparatively evaluated to elucidate the importance of the relationship of polymer structure and contamination to the optimum device performance.

III. EXPERIMENTAL RESULTS

A. Infrared Identification of Plastics

Infrared spectra were determined on either the pyrolysis products or on cast films of the products. Table 1 presents the results of these tests.

One of the most interesting findings in this phase of the investigation was the detection of ammonia in the AM-1 molding compound. This was of particular concern since other investigators (Ref 1 and 2) have reported that ammonia increases the reverse leakage current in contact with semiconductor junctions.

Table 1

Infrared Identification of Plastics

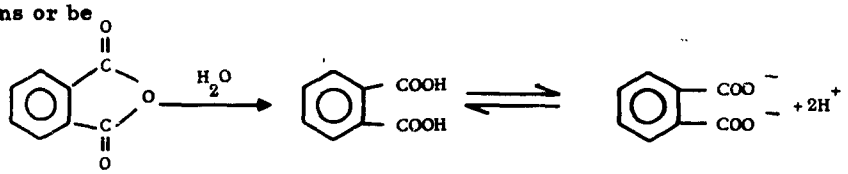
Compound and User	
Vendor A (R-11) Inner Coating AC-1	Primarily alkyd with some indications of silicone.
Vendor A (Durez 22845) Molding Compound AM-1	Phenolic resin, ammonia present.
Vendor A (DC305) Molding Compound AM-2	Primarily methyl silicone with small amounts of phenyl silicone.
Vendor B (SC-12) Inner Coating BC-1	Primarily alkyd with some indications of silicone.
Vendor B (SC-15) Inner Coating BC-2	Methyl-phenyl silicone.
Vendor B (C-26) Intermediate Coating BC-3	Anhydride-epoxy. The anhydride is apparently a mixture of methyl nadic and dodecyl succinic anhydride.
Vendor B (DC304) Molding Compound BM-1	Primarily methyl silicone with small amounts of phenyl silicone.
Vendor C (DC1400) Transistor Coating CC-1	Alkyd resin with small percentage of silicone.

B. Water Extract Electrical Resistivities

These tests were conducted to indicate the amount of water extractable ionic constituents present in the plastic, which may be introduced in the numerous manufacturing steps involved in synthesizing the base polymer from fillers and additives used in compounding the plastic (particularly true with some transfer molding compounds) or from unreacted constituents (such as an organic amine or acid anhydride) arising from nonstoichiometric quantities.

The procedure involved extracting weighed amounts of the ground plastic with deionized water for 8 days at 160 F and measuring the electrical resistivities of the extract solutions at ambient temperature, using an Industrial Instruments, Inc., Conductivity Bridge Model RC16B (Ref 3). Resistivity data (Table 2) showed the phenolic molding compound to have a low value which indicated a high ionic content;

conversely, methyl-phenyl silicone plastics had the highest resistivities. The difference in the resistivity of the cured and uncured anhydride-epoxy originates from the acid component, which can form ions or be



rendered relatively nonconductive through esterification and etherification reactions taking place during epoxy cure.

Table 2

Water Extract Resistivities

Vendor	Material	Average Resistivity (10^3 ohm-cm)
A	Alkyd-silicone Coating (R-11) AC-1 (Uncured)	73
	Alkyd-silicone Coating (R-11) AC-1 (Cured)	65
	Phenolic Molding Compound (Durez) AM-1 (Uncured)	4.5
	Phenolic Molding Compound (Durez) AM-1 (Cured)	4.7
	Silicone Molding Compound (DC305) AM-2 (Uncured)	137
	Silicone Molding Compound (DC305) AM-2 (Cured)	155
B	Alkyd-silicone Coating (SC-12) BC-1 (Uncured)	30
	Alkyd-silicone Coating (SC-12) BC-1 (Cured)	14.7
	Methyl-phenyl-silicone Coating (SC-15) BC-2 (Uncured)	86
	Methyl-phenyl-silicone Coating (SC-15) BC-2 (Cured)	270
	Anhydride-epoxy Coating (C-26) BC-3 (Uncured)	1.8
	Anhydride-epoxy Coating (C-26) BC-3 (Cured)	9.0
	Silicone Molding Compound (DC304) BM-1 (Uncured)	270
	Silicone Molding Compound (DC304) BM-1 (Cured)	290
	Alkyd-silicone Coating (DC1400) CC-1 (Uncured)	9.0
	Alkyd-silicone Coating (DC1400) CC-1 (Cured)	11.5
	R-2 (Cured)	370
	R-2 (Uncured)	11.8
	Epiall 1960 AM-3 (Uncured)	107

C. Chemical Investigation of Phenolic Molding Compound AM-1

Microdiodes encased in phenolic molding compound were marginal in passing performance specification tests. In addition they were shown to contain ammonia and exhibited a high water extract conductivity, so it was decided to further investigate the phenolic molding compound in this diode. Spectrographic analysis of the ash (Table 3), determination of resin content on the cured and uncured compound, and measurement of the pH on the water extract of the phenolic were studied. The resin content determination consisted in ashing a known weight of AM-1 molding compound and calculating the weight loss, which gave 54.87 percent (with an uncured sample) and 52.40 percent (using the cured resin). The lower resin content of the cured compound was attributed to the loss of volatiles during the cure. The pH measurements averaging 8.3 were on the alkaline side and coincided with a weak solution of ammonium hydroxide.

Table 3

Spectrographic Analysis of AM-1 Phenolic Molding Compound Ash

Dry*			Water Extract		
Magnesium	-	Major (10% or more)	Magnesium	-	Trace
Aluminum	-	Major	Calcium	-	Trace
Silicon	-	Major	Silicon	-	Trace
Calcium	-	Major	Aluminum	-	Trace
Sodium	-	Minor (Approximately 10%)			
Boron	-	Minor			
Zinc	-	Minor			
Titanium	-	Minor			
Iron	-	Trace (0.1% or less)			
Nickel	-	Trace			
Manganese	-	Trace			
Chromium	-	Trace			
Copper	-	Trace			
Silver	-	Trace			

*Ash from cured and uncured samples gave same results.

D. Thermogravimetric Analysis

To determine the relative thermal stability of various plastic encapsulants, pyrograms (weight loss versus temperature curves) were obtained using a Stone model thermogravimetric analyzer.

Analyses were conducted with a programmed temperature increase rate of 6 C per minute and pyrograms were obtained for samples of cured and uncured AM-1 phenolic molding compound. The initial weight loss was subtly evident in the cured sample at temperatures above 50 C as is shown on the pyrogram shown in Figure 1. The initiation of the weight loss in the uncured sample occurred sharply at about 120 C, which appears to indicate that the rates of gas producing reactions do not become appreciable until around 120 C. The weight loss in the cured sample is due to release of ammonia (cf. Section H on Mechanism). It is believed that breakdown of the polymer begins around 350 C, with the weight loss rate increasing above this temperature producing gaseous decomposition products of the polymer.

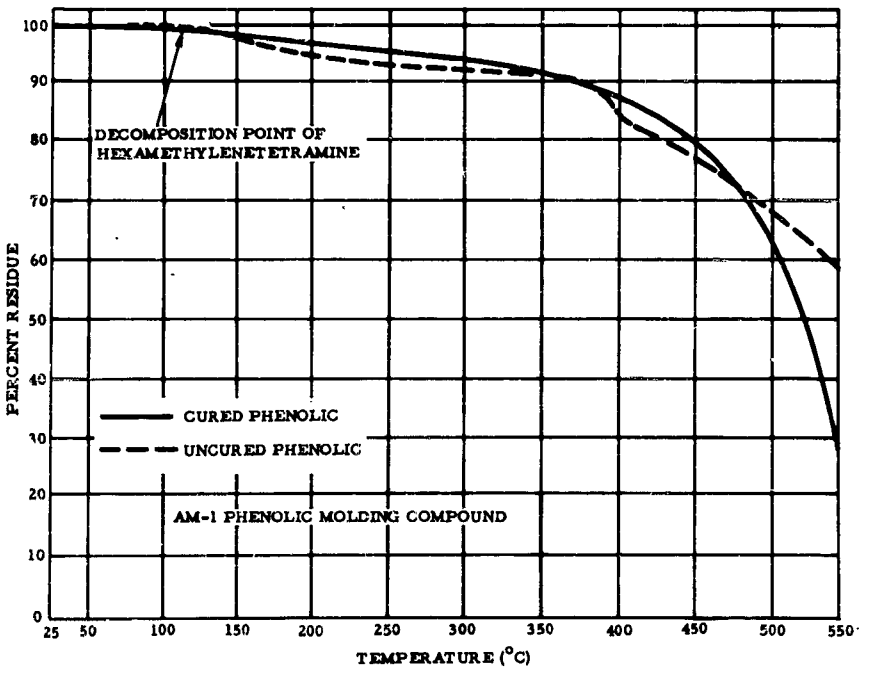


Figure 1

Pyrogram of Phenolic

Table 4 presents instantaneous weight loss rates taken at two temperatures and average weight loss rates between 60 to 350 C. Pyrograms made on both cured and uncured epoxy molding compound indicated no appreciable weight losses below 200 C which is well above the maximum temperature which diodes are likely to experience. Therefore, from the standpoint of deleterious effects due to gassing, epoxy material appears superior to the phenolic under these conditions.

Table 4

AM-1 Phenolic Molding Compound Thermal Weight Loss

State of Molding Compound	Temperature °C	Weight Loss %	Weight Loss Rate μg/C	Weight Loss Rate μg/Min
Cured	120	1.1	3.37	20.22
Uncured	120	0	0	0
Cured	200	3.3	3.37	20.22
Uncured	200	4.3	4.32	25.92
Cured	60* to 350	0 to 6.6	3.48	20.90

*The weight loss rate was approximately constant over this temperature range.

E. Gas Permeability

Diodes encased in the AM-1 molding compound (Figure 2) also contain an inner coating (AC-1), which is directly applied over the passivated diode surface. Gases originating from AM-1 coating would have to permeate the AC-1 coating before contacting the diode surface. Experiments were therefore devised to measure moisture vapor and ammonia gas permeabilities on the AC-1 coating in an effort to obtain pertinent data relative to this.

In the determination of ammonia permeability a test film was sealed over a bottle partly filled with 28-percent chemically pure ammonium hydroxide, which at equilibrium contained approximately 98 percent by volume ammonia gas (Ref 4).

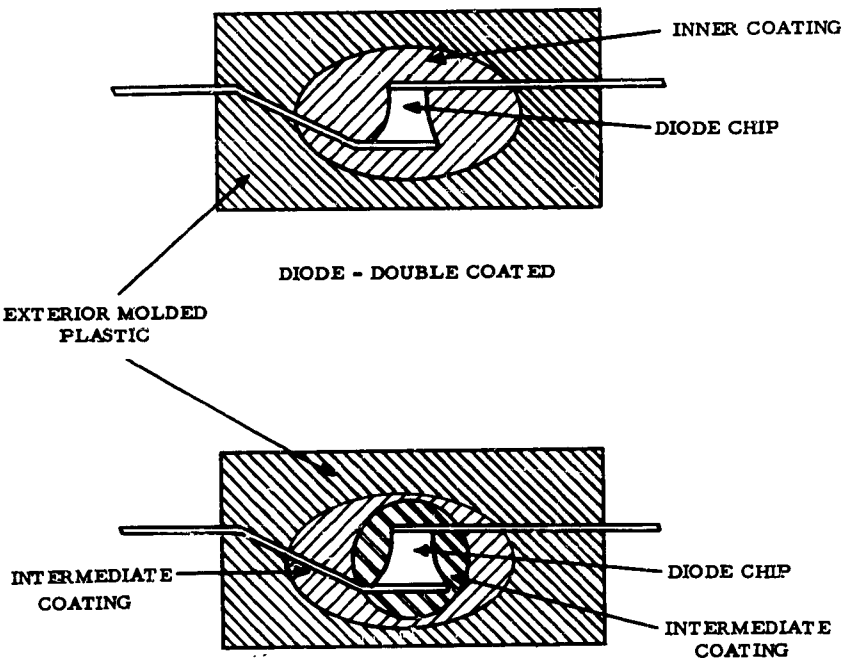


Figure 2

Plastics Used in Diodes

Moisture vapor permeability tests were conducted using the cup method in accordance with Ref 5 which consisted of clamping the test film over a flanged cup containing water and placing in a desiccator. The weight loss from the cup over a period of time was used to calculate the permeability rates.

Table 5 presents the ammonia gas and moisture permeability results with the AC-1 coating. Although the test methods for the ammonia and water vapors are slightly different, the large difference in average rates clearly demonstrated the higher permeability of the AC-1 coating to ammonia gas.

Table 5

Ammonia and Moisture Vapor Permeability of AC-1 Coating

Average Ammonia Permeability in Micrograms-cm/hour-cm^2	Average Moisture Vapor Transmission in Micrograms-cm/hours-cm^2
26.1	1.3

The moisture vapor transmission rates of the AC-1 coating as compared with other materials was investigated (Table 6) and illustrates the greater permeability of this material. The AM-1 molding compound was found to be a great deal more permeable as compared to the other materials which can be attributed to the microporosity in the phenolic resulting from gassing during cure.

Table 6

Moisture Vapor Permeability

Material	Average Moisture Vapor Transmission in Micrograms-cm/hours-cm^2
AC-1 (Silicone alkyd coating)	1.3
AM-1 (Phenolic molding compound)	18.5
BC-1 (Silicone alkyd coating)	1.0
BM-1 (Silicone molding compound)	6.8
CC-1 (Alkyd-silicone coating)	0.9

F. Purified Resins

1. Preparation and Tests. Concurrent with the study of the commercial plastics used on diodes, a project was undertaken to purify plastic constituents and evaluate them as protective systems over uncased diodes. An epoxy resin together with several aromatic amines and acid anhydride curing agents utilized as co-reactants were more rigorously purified. The epoxy resin was purified by molecular distillation, which separates the resin into fractions relatively free of inorganic contaminants with narrower molecular weight ranges. The curing agents were purified by distillation through a Vigreaux distillation column.

Water extract resistivity measurements conducted on the as-received epoxy resin, the molecular still residue, and the light molecular weight fraction were all high and approached those of the deionized water blanks. However, spectrographic analysis on ash from the three samples (Table 7) revealed a considerable difference in metal content. Since the water extract tests did not reveal a significant difference in resistivity among the three materials, it appears that the metal ions are insoluble and bound.

Table 7

Spectrographic Analysis of Epoxy Resin Fractions

As-Received Epon 828	Still Residue	Light Molecular Weight Fraction
Copper - Trace*	Silicon - Major**	Copper - Trace (less than original)
Aluminum - Trace	Magnesium - Trace	
Magnesium - Trace	Aluminum - Trace	
Silicon - Trace	Iron - Trace	
Calcium - Trace	Copper - Trace	
	Chromium - Trace	
	Calcium - Trace	
	Manganese - Trace	

*0.1 percent or less
**10 percent or more

2. Stoichiometry - Water Extract Resistivity Correlation. The effects of a plastics on semiconductor junctions could conceivably originate from one or a combination of several factors including:

1. Extraneous contaminants contained in the plastic, i.e., ionic substances in contact with semiconductor surfaces which could lead to inversion.

2. Structural physical properties innate in the polymer molecule such as polarity and polarizability.

3. Mechanical stress formation due to plastic cure and/or thermal shrinkage which could conceivably produce changes in resistivity in doped semiconductor.

4. The use of nonstoichiometric equivalents of curing agent and resin in the formulation of epoxy plastic containing available organic ions can lead to an increase in conductivity.

Factor 1 could be minimized by using purified reactants. Factors 2 and 3 are fixed because of the preselection of the resin to be used. To minimize factor 4, a method was developed to check the completeness of reaction in plastic "mixes" containing different epoxy-acid anhydride ratios. This technique consisted of measuring water extract resistivity on cured 1-gram ground samples of each sample "mix." Table 8 presents the results, while Figure 3 depicts the curve. From the results, it is evident that the maximum resistivities are exhibited by 60-100 and 50-100 anhydride-epoxy ratios for the technical and purified reactant "mixes," respectively. It can also be seen that the resistivities decrease on extending the digestion (conducted at 165 F) time from 72 hours to 240 hours. Since a portion of the cross-linking reaction is esterification (etherification is also a reaction mechanism) this decrease in resistivity can be attributed to the hydrolysis of the ester cross-linkages and the attendant formation of carboxylic acids. The presence of these ionizable organic acids increases the conductivity.

3. Stoichiometry - Elevated Temperature Electrical Resistivity Correlation. Elevated temperature volume resistivities were conducted on cured plastic samples containing the same resin anhydride proportions as used in some of the water extract resistivity-stoichiometry study. In an effort to determine the causes for excessive room temperature diode reverse leakage currents after the diodes had been subjected to sustained reverse bias voltages at elevated temperatures, the volume resistivities were obtained at three temperatures: 100, 200, and 300 F using the guarded electrode test. This method minimizes the error due to surface conductivity by shunting the current away from the meter.

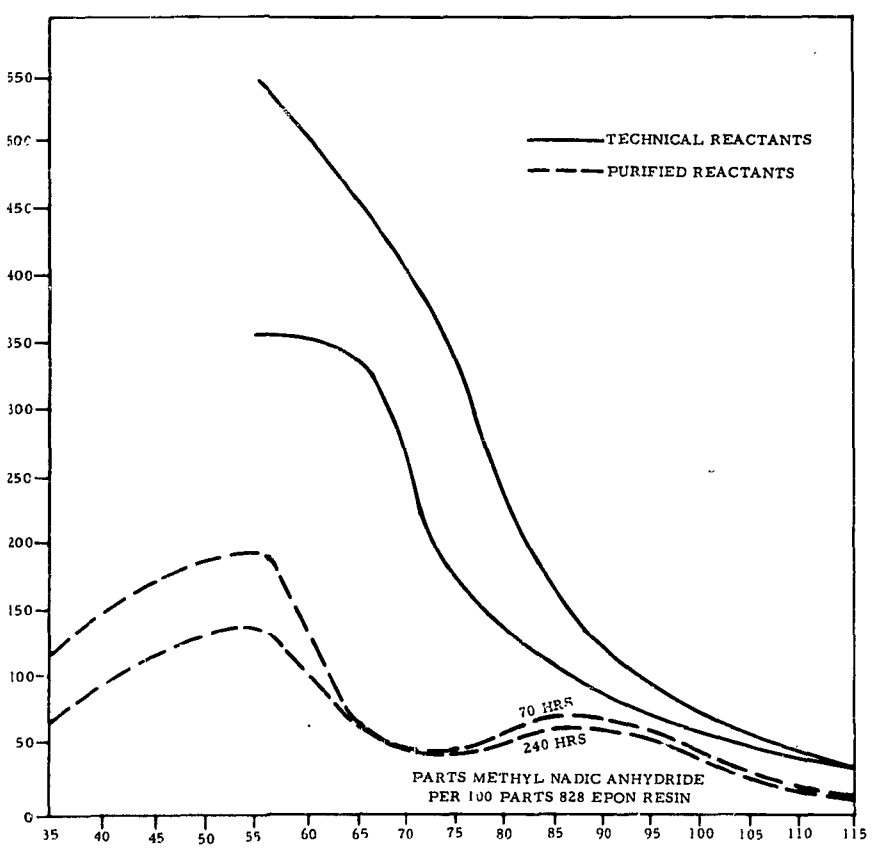


Figure 3

Water Extract - Technical Reactants

Table 8

Water Extract Resistivity-Technical and Purified Reactants

Proportions**			Average Specific Resistivity (Ohm-cm)			
			After 72 Hours		After 240 Hours	
MNA	BDMA	Epon 828	Technical	Purified	Technical	Purified
35	0.5	100	---	103,000	---	66,500
45	0.5	100	---	171,500	---	111,500
55	0.5	100	550,000	190,000	355,000	138,000
65	0.5	100	460,000	73,000	345,000	67,000
75	0.5	100	335,000	48,000	170,000	43,500
85	0.5	100	165,000	71,000	120,000	62,000
95	0.5	100	100,000	64,500	82,000	55,000
105	0.5	100	61,000	31,500	50,000	27,000
115	0.5	100	33,500	21,250	29,000	17,500
Blank	Blank	Blank	430,000	850,000	---	550,000

* The method for purification of the reactants is given in Progress Report No. 8

**Per 100 gms epoxy
MNA = Methyl Nadic Anhydride
BDMA = Benzyldimethylamine
Epon 828 = Shell Chemical Company Epoxy Resin

To explain the high leakage currents after elevated temperature-reverse bias stresses, it was hypothesized that at elevated temperatures, the ions in the plastic made mobile by the high temperature will move under the influence of an electric field. Those ions near the silicon surface will then induce a layer of opposite charge in the semiconductor near the surface which may result in channeling. If the semiconductor is cooled with the reverse bias voltage on, there is a tendency to "freeze" the channeling condition since the ions are less mobile at room temperature, and will then exhibit high leakage currents. Evidence which corroborates this hypothesis was observed when baking the diodes under reverse or forward bias. This tended to remove the high leakage current. Ideally, then, the best plastic to prevent this type of leakage should be one which does not have a high increase in conductivity with temperature.

Figures 4, 5, and 6 depict the current (in nanoamperes) as a function of inverse temperature. The "stoichiometric" ratios are those resin anhydride ratios which were found to exhibit the maximum water extract resistivity, while "nonstoichiometric" mixes are those which gave the lowest water extract resistivity of the ratios tested. As the word implies, the "intermediate" mixes mentioned in these figures are mixes which contained resin-anhydride ratios intermediate between the stoichiometric and nonstoichiometric ratios.

Figure 7 is a comparison of the 300-volt curves for both technical grade and purified reactants. It is evident from these curves that there is an apparent correlation between the water extract resistivities and the elevated temperature resistivities of the plastics. By the same token, there is a correlation between deviations from stoichiometry and the elevated temperature conductivities of plastics. As a comparison, Figure 8 shows current versus inverse temperature relationships for AM-1 phenolic molding compound.

Tables 9 and 10 present the calculated volume resistivities for the different ratio mixes of the purified and technical grade epoxy anhydride mixes, while Table 11 contains the AM-1 phenolic molding compound data.

As can be seen from the current-inverse temperature curves, not all of them are straight lines. No explanation can be given at this time for these anomalies, but it is suspected that changes in the power supply or test fixture might be a source of error.

An effort was made to express these current-inverse temperature curves by means of a mathematical equation. The expression $I = Ae^{-B/T}$ suggested in Ref 6 was chosen as the appropriate equation to represent this data, where I = electrical current, e = natural logarithmic base, T = the absolute temperature in degrees Kelvin, A = a constant, and B = a constant. To solve for A and B, two points were arbitrarily selected on the 200-v curve in Figure 6 to give $A = 24$ and $B = -8.39 \times 10^3$. The calculated values using these constants in the equation coincided with those taken from the curve. Additional elevated temperature resistivity tests conducted on Teflon yielded no discernible currents to 400 F as anticipated.

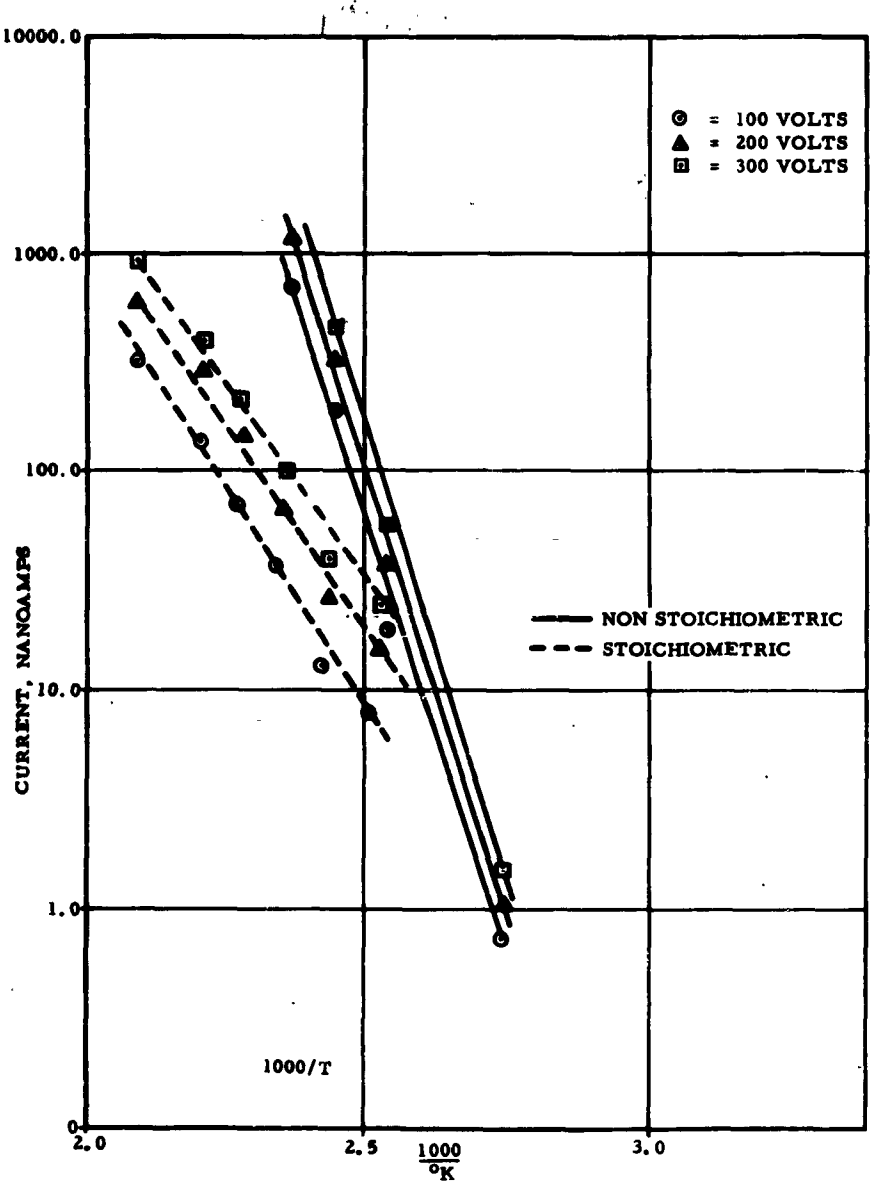


Figure 4

Epoxy-Anhydr.de (Purified), Current as a Function of Temperature

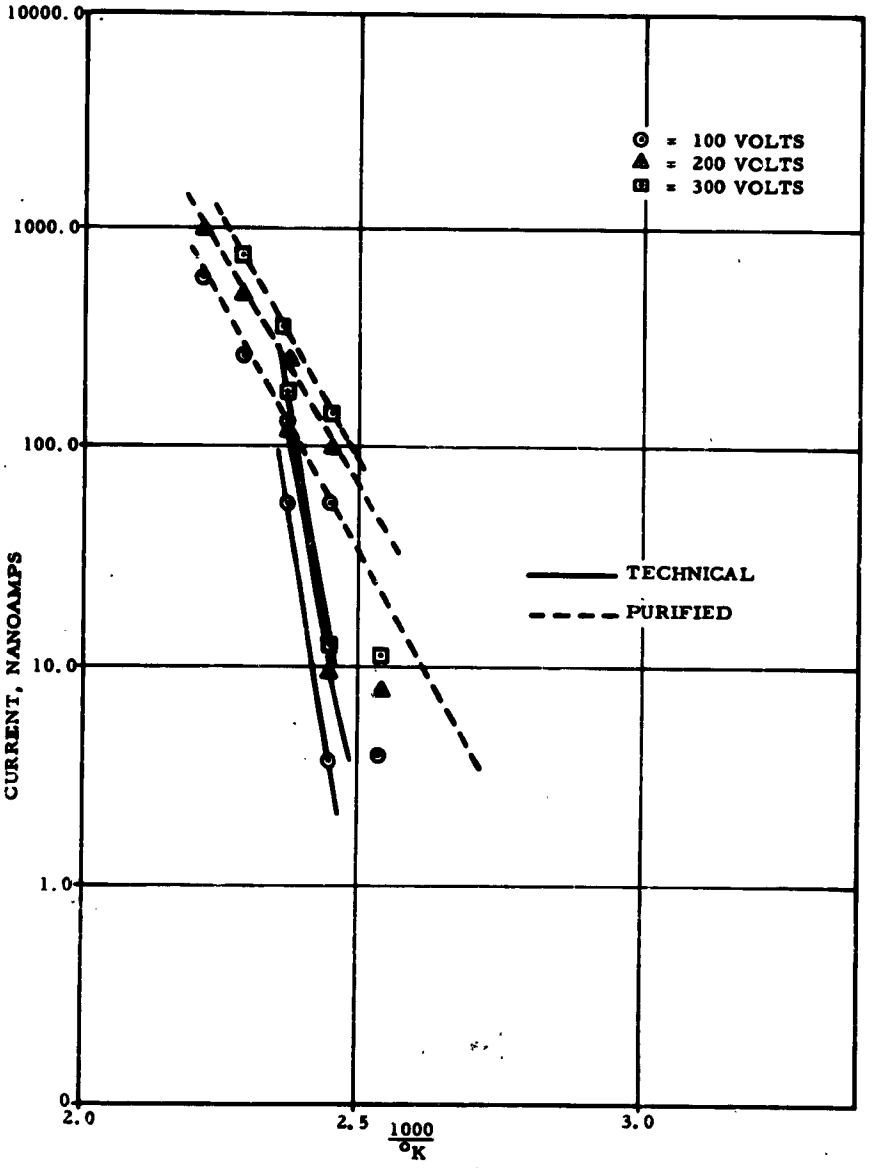


Figure 5

Epoxy-Anhydride, Current as a Function of Temperature

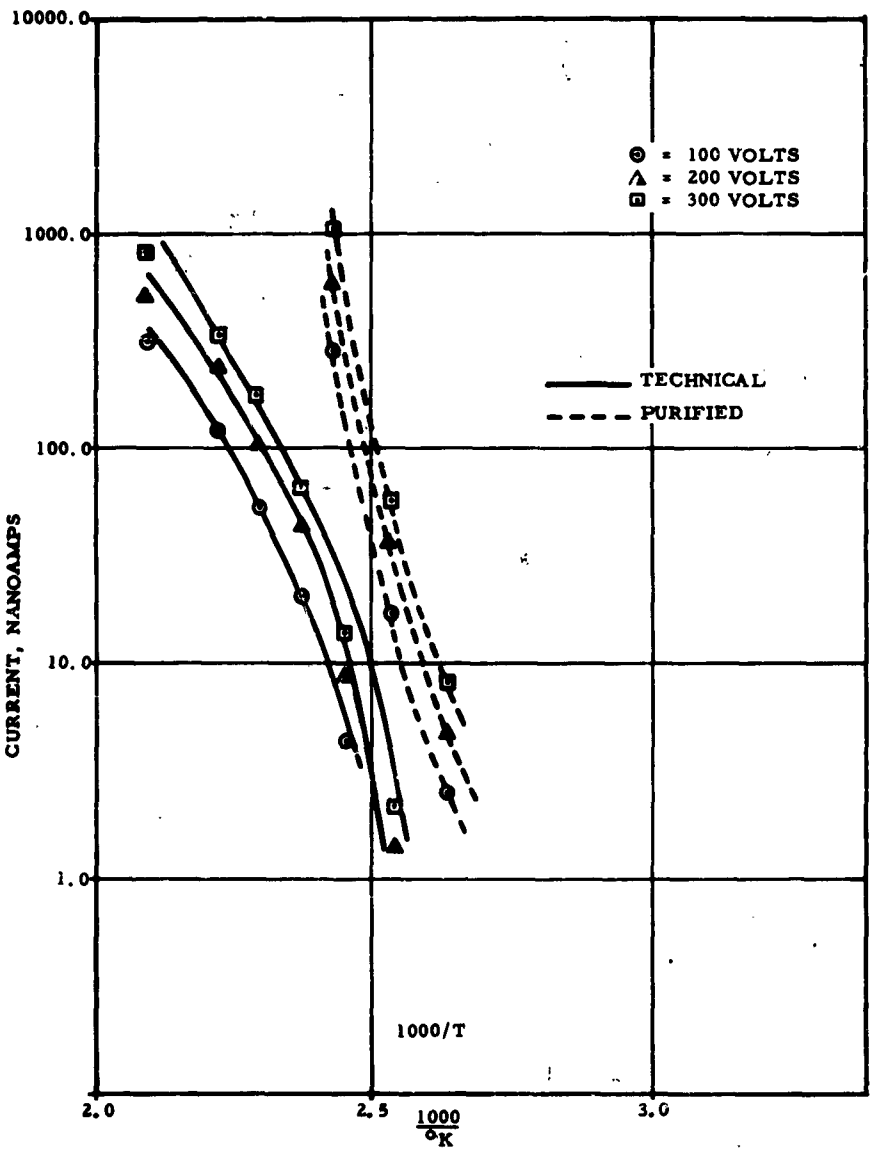


Figure 6

Epoxy-Anhydride (Stoichiometric), Current as a Function of Temperature

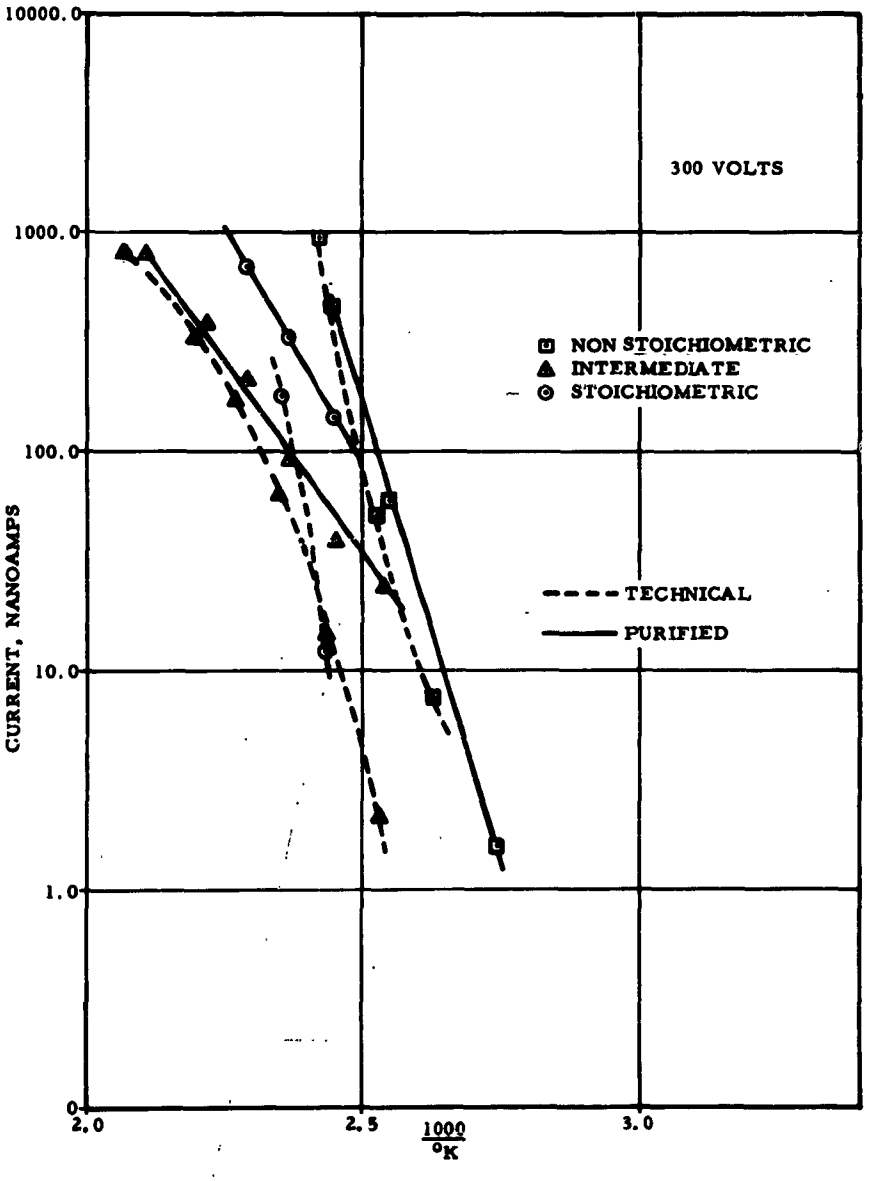


Figure 7

Epoxy-Anhydride, Current as a Function of Temperature

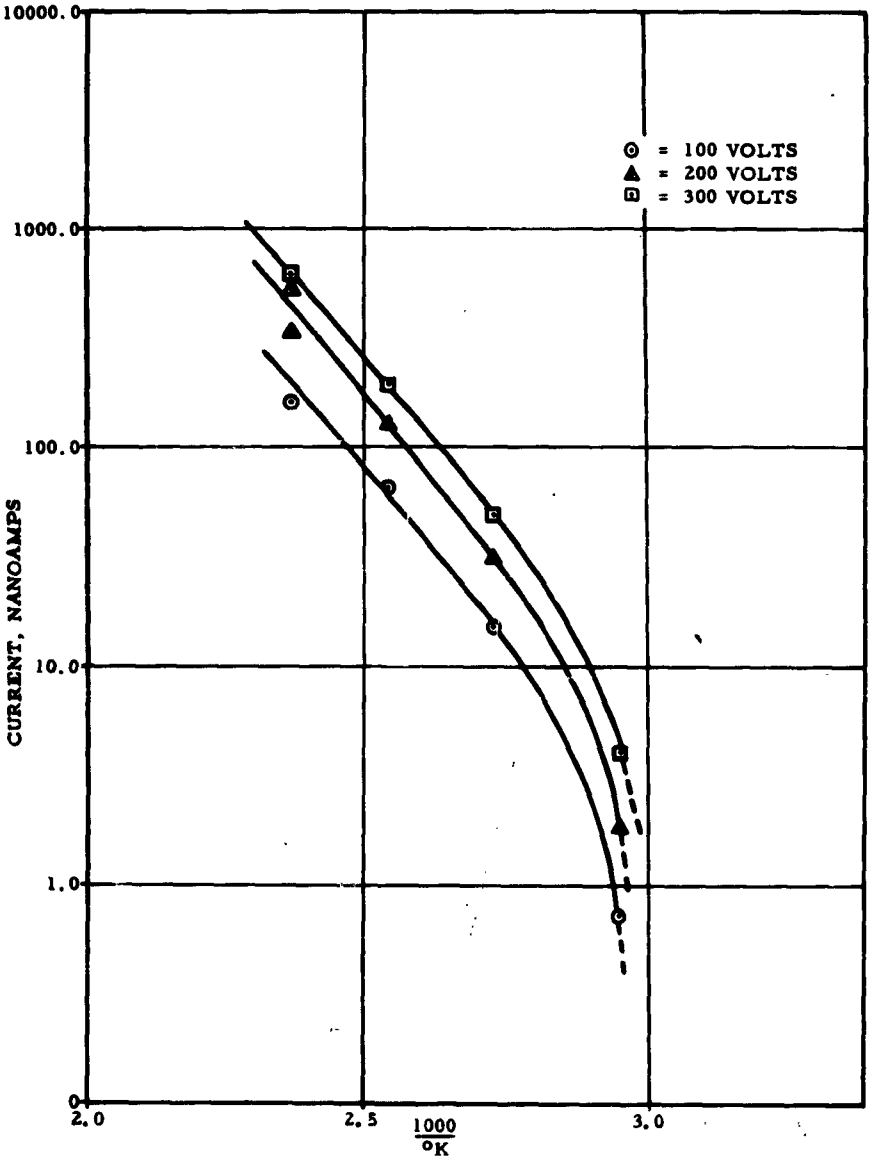


Figure 8

AM-1 Phenolic Molding Compound, Current as a Function of Temperature

Table 9

Volume Resistivities (Technical Reactants) of Epoxy-Anhydride Mixtures

Material (Thickness)	Test Temp °F	E Volts	I (NA)	Volume Resistivity (Ohm-In x 10^{10})	Average
Stoichiometric Ratio	100	100	---	---	
	200	100	---	---	
(0.062 Inch)	300	100	21.0	14.40	
	100	200	---	---	
	200	200	---	---	
	300	200	40.5	14.29	
	100	300	---	---	
	200	300	---	---	
	300	300	60.75	15.00	14.56 @ 300F
Intermediate Ratio	100	100	---	---	
	200	100	.070	4630.0	
(0.058 Inch)	300	100	58.5	5.55	
	100	200	---	---	
	200	200	.135	4700.0	
	300	200	118.5	5.47	3910 @ 200F
	100	300	---	---	
	200	300	.375	2600	
	300	300	197.2	4.93	5.32 @ 300F
Non-Stoichiometric Ratio	100	100	289	1050	
	200	100	.590	514.5	
(0.062 Inch)	300	100	1450	0.21	4883 @ 100F
	100	200	.069	8790	
	200	200	1.25	485.0	
	300	200	3000	0.20	462.2 @ 200F
	100	300	.185	4810	
	200	300	2.35	387.0	
	300	300	7500	0.12	0.1773 @ 300F

Table 10

Volume Resistivities (Purified Reactants) of of Epoxy-Anhydride Mixtures

Material Thickness	Test Temp. °F	E Volts	I (NA)	Volume Resistivity (Ohm-In x 10^{10}	Averages
(Purified)	100	100	.020	12,500	
Stoichiometric	200	100	1.48	169.5	
Ratio	300	100	35.0	7.17	10,583 @ 100F
(0.075 Inch)	100	200	.045	10,800	
	200	200	2.80	208.5	
	300	200	65.00	7.69	176.03 @ 200F
	100	300	.089	8,450	
	200	300	5.00	150.1	
	300	300	97.00	7.76	7.54 @ 300F
(Purified)	100	100	.010	28,000	
Intermediate Ratio	200	100	2.35	119.2	
(0.067 Inch)	300	100	125.00	2.24	
	100	200	.045	12,490	
	200	200	5.85	95.8	
	300	200	265.00	2.12	
	100	300	.038	21,700	
	200	300	6.75	124.5	
	300	300	360.00	2.34	
(Purified)	100	100	.00014	1,800,000	
Non-Stoichiometric	200	100	0.77	376.0	
Ratio	300	100	750.00	0.35	
(0.072 Inch)	100	200	.00010	5,225,000	
	200	200	1.10	475.0	
	300	200	1,250.00	0.42	
	100	300	.00016	4,900,000	
	200	300	1.65	474.0	
	300	300	2,100.00	0.373	

Table 11

Volume Resistivity from Electrical Conductivity Data - AM-1 Phenolic Compound

Material Thickness	Test Temp °F	E Volts	I (NA)	Volume Resistivity (Ohm-In x 10^{10})	Average
(0.070 Inch)	100	100	0.5	587.0	
	200	100	15.0	18.0	
	300	100	190.0	1.4	556.0 @ 100F
	100	200	1.0	540.0	
	200	200	32.0	16.9	
	300	200	420.0	1.3	17.0 @ 200F
	100	300	1.5	540.0	
	200	300	50.0	16.2	
	300	300	620.0	1.3	1.3 @ 300F

G. Component Testing

The latter part of the program was devoted primarily to review of the data acquired and to conducting experiments which could conceivably correlate the data and failure mechanisms. Of particular interest was the ammonia discovered in the AM-1 molding compound and the information acquired by the elevated temperature experiments.

1. Diode Ammonia Exposure. As indicated previously, ammonia was found in the AM-1 molding compound used by a vendor to encase diodes. To evaluate the effect of ammonia gas environments on diodes, bare and passivated or passivated coated with one of several plastic systems were exposed to the different gas environments and 100-vdc reverse bias. Table 12 presents the protective plastic systems and the test conditions. Due to the insufficient number of test diodes available initially, only three or four were used in each combination of protective coatings and gas environment. The dry ammonia and dry nitrogen gases used in these tests were as received from the manufacturer. Calculations based on the moisture maximum content specified for the ammonia revealed that it contained less moisture than air dried over calcium chloride. The moist gases were obtained by bubbling the dry nitrogen and the dry ammonia through water and concentrated ammonium hydroxide, respectively, which were then filtered through glass wool prior to injecting into the test chamber, which consisted of a test tube containing the mounted diode with electrical connections. Reverse bias voltages were applied to the diode under these test conditions after approximately 24 hours of exposure to the gases. The results including average reverse currents as a function of time are shown in Figures 9 and 10 and indicate that higher and more erratic currents are obtained in the presence of ammonia. In comparison

most of the diodes in nitrogen remain fairly constant. In addition, the reverse currents of the diodes exposed to ammonia remained stable for a number of hours and then changed, sometimes drastically, as indicated in curve C. The relatively high currents of the normal case diodes may result from the porosity of the AM-1 molding compound coupled with the previous subjection of the diodes to ammonia gas under heat and pressure during the molding operation.

Table 12

Diode Ammonia Exposure

Designation	Coating System	Gas Atmosphere
A	Uncoated diodes	Dry nitrogen
B	Normal case diodes*	Dry nitrogen
C	AC-1 alkyd-silicone coating	Dry nitrogen
D	CC-1 alkyd-silicone coating	Dry nitrogen
E	Uncoated diodes	Dry ammonia
F	Normal case diodes	Dry ammonia
G	AC-1 alkyd-silicone coating	Dry ammonia
H	CC-1 alkyd-silicone coating	Dry ammonia
I	Uncoated diodes	Moist nitrogen
J	Normal case diodes	Moist nitrogen
K	AC-1 alkyd-silicone coating	Moist nitrogen
L	CC-1 alkyd-silicone coating	Moist nitrogen
M	Uncoated diodes	Moist ammonia
N	Normal case diodes	Moist ammonia
O	AC-1 alkyd-silicone coating	Moist ammonia
P	CC-1 alkyd-silicone coating	Moist ammonia

*The standard completed diodes were coated with the AC-1 coating and then molded with the AM-1 compound.

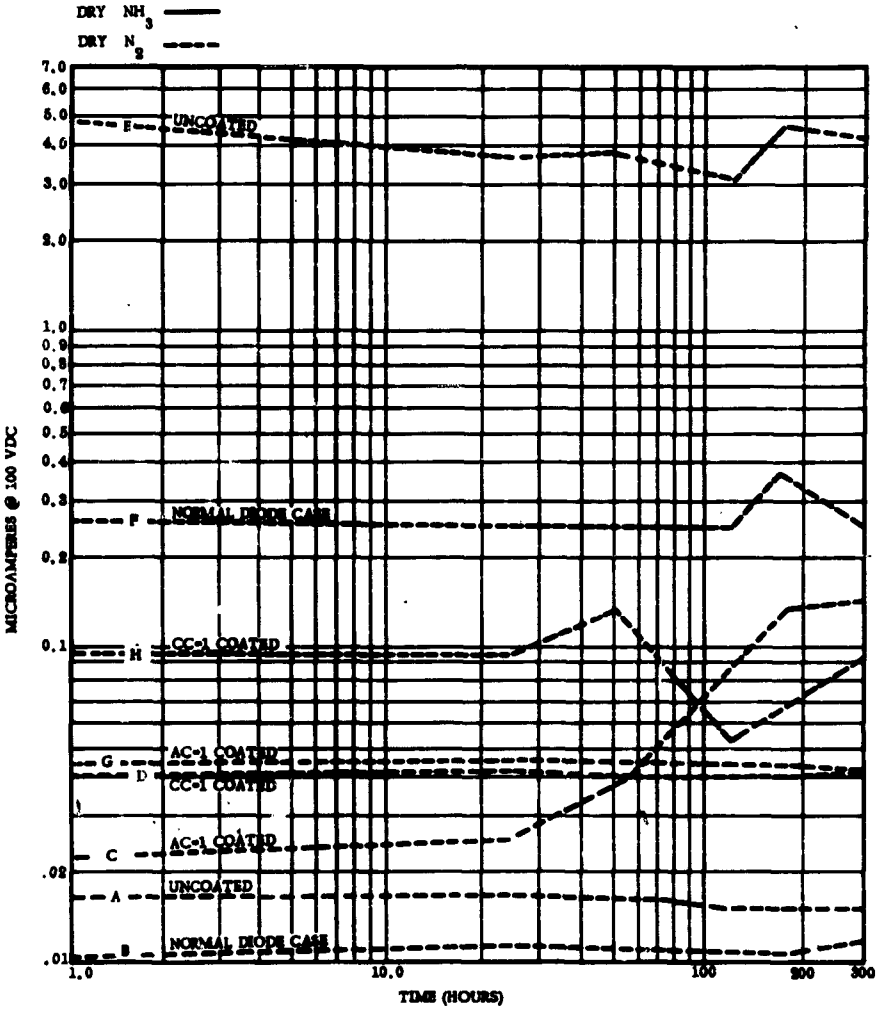


Figure 9

Diode Leakage in Dry Ambients

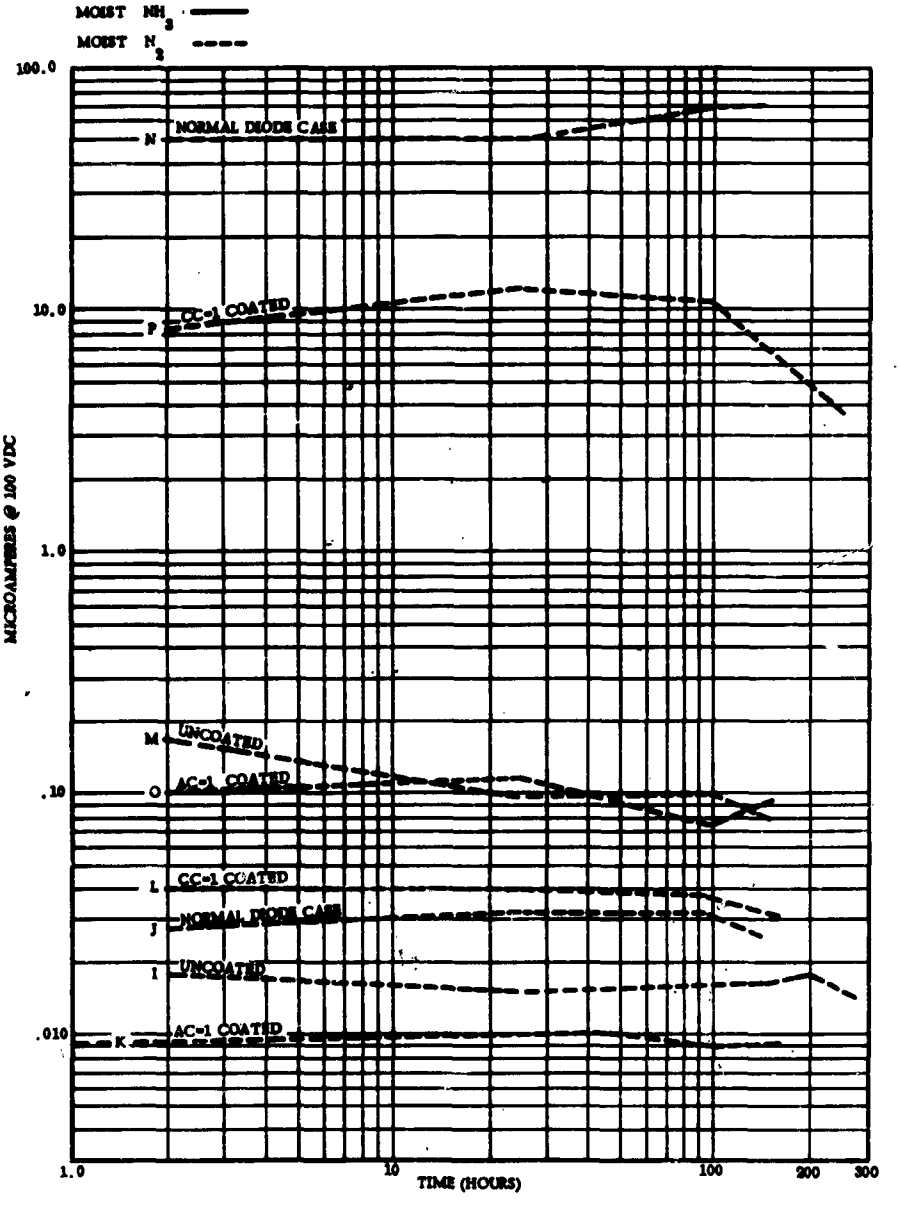


Figure 10

Diode Leakage in Moist Ambients

2. Diode Elevated Temperature - High Humidity Testing. In the second series of component tests, diodes were coated with stoichiometric and nonstoichiometric mixtures of the epoxy-anhydride and Teflon as previously reported and were mounted in test tubes and subjected to reverse bias and high relative humidity or to 100-v reverse bias and elevated temperatures.

1. Elevated Temperature Test

From the results in Table 13, the test data on I_R and V_f, it appears that the diodes protected with Teflon were outstanding in their resistance to change. The reverse currents increased with many of the epoxy coated diodes regardless of the purity or adherence to reactant stoichiometry which may be due to the coating being innately conductive at elevated temperatures. The forward voltages (V_f) did not change drastically on any of the diodes regardless of protective coating.

Table 13

Diode Elevated Temperature (Reverse Bias Tests)

S/N	Diode Surface Condition	Coating System*	Initial Reading I_R (na) @ 50 V	I_R(na) @ 50 V After 20 hrs @ 150°C and 75 V Rev. Bias	I_R(na) @ 50 V After Addition, 20 hours @ 150°C No Rev. Bias	V_f (mv) @ 1ma Initials	V_f(mv) @ 1ma after 20 hrs @ 150°C and 75 V Rev. Bias
739	Pyrolyzed	Epoxy, Stoichiometric Purified	10	548	572	14	573
739	Unpyrolyzed	Epoxy, Stoichiometric Purified	11	18 μa	586	1.0 μa	557
743	Unpyrolyzed	Epoxy, Stoichiometric Purified	14	150 μa	569	60 μa	543
744	Unpyrolyzed	Epoxy, Non-Stoichiometric - Purified	10	220 μa	579	76 μa	543
745	Unpyrolyzed	Epoxy, Non-Stoichiometric - Purified	16	152 μa	572	---	535
746	Pyrolyzed	Epoxy, Non-Stoichiometric - Purified	9	10 μa	575	10 μa	576
754	Unpyrolyzed	Epoxy, Stoichiometric Technical	12	237 μa	572	121 μa	647
759	Unpyrolyzed	Epoxy, Non-Stoichiometric - Technical	12	31 μa	571	14 μ a	541
767	Pyrolyzed	Non-polar Inert Resin	13	13 (na)	575	13	578
768	Unpyrolyzed	Non-polar Inert Resin	17	18	574	16	593
769	Pyrolyzed	Non-polar Inert Resin	9	9	609	9	599
770	Unpyrolyzed	Non-polar Inert Resin	11	9	586	8	591
771	Unpyrolyzed	Non-polar Inert Resin	11	13	574	13	577
772	Pyrolyzed	Non-polar Inert Resin	12	13	572	11	574

2. High Humidity Tests

High humidity tests were conducted on at 160 F and 75-volt back bias under 95-percent relative humidity to determine the effects on I_R and V_f which is presented in Table 14. The Teflon coated diodes showed no change in I_R. However, the diodes coated with purified and stoichiometrically mixed constituents also had very little change in I_R.

Table 14

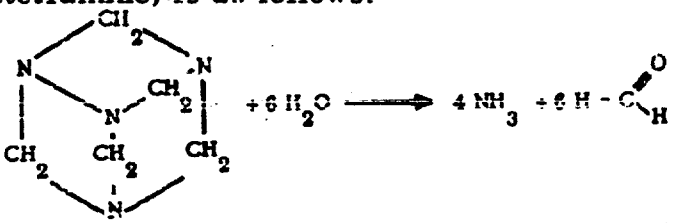
Diode Humidity - Reverse Bias Tests

S/N	Diode Surface Condition	Coating System	Initial Reading I_R (na) @ 50 V	I_R (na) @ 50 V After 96 hours @ 160°; 95° to Relative Humidit 75 V Back Bias
701	Unpyrolyzed	Epoxy, Stoichiometric - Purified	8	11
702	Pyrolyzed	Epoxy, Stoichiometric - Purified	8	7
703	Pyrolyzed	Epoxy, Stoichiometric - Purified	8	9
704	Pyrolyzed	Epoxy, Stoichiometric - Purified	6	5
705	Unpyrolyzed	Epoxy, Stoichiometric - Purified	8	8
706	Unpyrolyzed	Epoxy, Stoichiometric - Purified	11	12
707	Pyrolyzed	Epoxy, Non-Stoichiometric - Purified	10	10
708	Pyrolyzed	Epoxy, Non-Stoichiometric - Purified	8	7
709	Unpyrolyzed	Epoxy, Non-Stoichiometric - Purified	13	8
710	Pyrolyzed	Epoxy, Non-Stoichiometric - Purified	11	69 µa
711	---------	----------	890 µa	00
712	Unpyrolyzed	----------	11	00
713	Pyrolyzed	Epoxy, Stoichiometric - Technical	10	9
714	Pyrolyzed	Epoxy, Stoichiometric - Technical	8	9
715	Unpyrolyzed	Epoxy, Stoichiometric - Technical	9	00
716	Unpyrolyzed	Epoxy, Stoichiometric - Technical	13	650
717	Unpyrolyzed	Epoxy, Stoichiometric - Technical	---	00
719	Pyrolyzed	Epoxy, Non-Stoichiometric - Technical	9	9
720	Pyrolyzed	Epoxy, Non-Stoichiometric - Technical	11	9
721	Pyrolyzed	Epoxy, Non-Stoichiometric - Technical	10	193
722	Unpyrolyzed	Epoxy, Non-Stoichiometric - Technical	16	00
723	Unpyrolyzed	Epoxy, Non-Stoichiometric - Technical	12	199
724	Unpyrolyzed	Epoxy, Non-Stoichiometric - Technical	14	11
725	Pyrolyzed	Uncoated	9	00
726	Pyrolyzed	Uncoated	6	00
731	Pyrolyzed	Teflon	11	8
732	Unpyrolyzed	Teflon	12	12
733	Pyrolyzed	Teflon	8	8
734	Pyrolyzed	Teflon	32	13
735	Unpyrolyzed	Teflon	9	9
736	Unpyrolyzed	Teflon	17	19

H. Mechanism of Ammonia Evolution

Ammonia was found to be present on diodes encapsulated with phenolic molding compound as indicated by pH (Section III. C) and confirmed by infrared spectra (Section III. A). This was found to affect the I_R as shown in the test tube experiment in which a diode was sealed in a moist ammonia environment (Section III. G). Thermogravimetric analysis of the phenolic molding compound indicated that the cured material began to lose weight almost immediately above room temperature whereas the uncured material was stable up to 120 C. This

difference may be explained by an understanding of polymer chemistry involved. In the curing of phenolics, a latent curing agent is used which remains unreactive until the desired decomposition temperature is reached. In the decomposition, formaldehyde is formed which is the active cross-linking agent, but a byproduct, NH_3, is also produced which apparently becomes entrapped in the molded plastic. The mechanistic equation for the decomposition of the curing agent (hexamethylenetetramine) is as follows:



It is apparent from the curves that the latent curing agent is decomposed at about 120 C, whereas in the cured system, it has already undergone decomposition. The release of this gas is manifest in the curves for the cured material. The products formed depend on the ratio of the resin to hexamethylenetetramine with about 50 to 75 percent of the nitrogen added becoming bound to the molecule, while the remainder is released as ammonia.

IV. SUMMARY

The investigation has been primarily concerned with a search for factors pertinent to mechanisms of failure in microdiode values which entailed design and screening evaluation of various experimental and analytical approaches. The data produced by this effort was reviewed for applicability to known or conjectured mechanisms of failure which were then applied to these components to verify postulated hypotheses.

The interesting results of the program include:

1. The discovery of ammonia in a phenolic molding compound used by a diode manufacturer. Further work revealed that ammonia contamination produced increased and erratic reverse currents and was more pronounced in the presence of moisture.

2. The revelation of the presence of ammonia and other deficiencies in this molding compound induced the manufacturer to investigate other compounds and substitute a better one.

3. The elevated temperature resistivities of coatings were a measure of the performance of these materials as protective systems over diodes under elevated temperature test conditions.

4. In some plastic compounds, a correlation was found between elevated temperature resistivities and water extract resistivities.

5. Teflon, an inert coating with a very high elevated temperature resistivity and high water extract resistivity, was demonstrated to have superior performance under elevated temperature reverse bias and high humidity reverse bias test conditions.

6. The resistivity of water extracts of evaporants was found to be an effective means of indicating the amount of ionic and other conductive impurities in plastic materials.

V. REFERENCES

1. *G. A. Katsev, et al, "Effect of the Absorption of Certain Amines on the Semiconductor Surface on the Principal Parameters of Germanium Transistor Triodes"

2. *R. O. Sitvinor and Hsu Tung-liang, "Effect of the Surface State in Silicon p-n Junction on the Reverse Currents and Drift"

3. "Molded Plastic and Plastic Parts Glass Fiber Filled, Diallyl Phthalate Resin," Specification MIL-M-19833A (SHIPS) 16 July 1962

4. John H. Perry, Editor, Chemical Engineers Handbook

5. "Insulating Compound Electrical Embedding," Specification MIL-M-16923D

6. Eric Baer, Ed, Engineering Design for Plastics, Soc. of Plastic Engineers, Inc.

*From a compilation of Translation of Russian papers entitled "Surface Properties of Semiconductors," edited by A. N. Frumkin, Consultant Bureau, 1964

INVESTIGATION OF SURFACE FAILURE MECHANISMS IN SEMICONDUCTOR DEVICES BY ENVELOPE AMBIENT STUDIES

G. V. Brandewie, P. H. Eisenberg, and R. A. Meyer *

Autonetics, a Division of North American Aviation, Inc.,
Anaheim, California

I. INTRODUCTION

Routine failure analysis at Autonetics indicated that a large percentage of devices with surface associated failure modes could be returned to original operating levels by opening the device package and releasing the enclosed gas ambient. Such a recovery indicated that the composition of the ambient encapsulated in a semiconductor device package could have an important effect on the stability of the electrical parameters of the device.

Some of these effects, e. g., I_R in diodes and I_{CBO} in transistors, have been directly related to the extreme surface sensitivity of semiconductor materials (Ref 1 and 2). Surface phenomena have been the subject of many investigations since a large portion of undefined failure mechanisms in solid state devices have been postulated to be attributable to surface effects. Adsorbed gas atoms on the surface have been found to affect surface charge concentrations, thus contributing to inversion layers or conductivity type changes (Ref 3).

The Physics of Failure Program on Minuteman II electronic devices further indicated the need to study gas ambient mechanisms in greater detail. A method was developed to analyze the gas ambient of good and failed devices by use of a CEC 21-103C mass spectrometer with a mass range of 1-1400 and unit resolution at mass 700. These techniques were developed so that the device junction would not be destroyed when the package was opened. Therefore, electrical measurements could be made before, during, and after removal of the ambient and after backfilling with selected ambients. Other techniques were developed utilizing the gas chromatograph for analysis.

*North American Aviation Science Center.

This paper describes several examples in which device package gas ambients were an important factor regarding the failure mechanisms. The survey of Minuteman II transistor and integrated circuit package ambients indicated that the gases present are a result of the manufacturing processes and are not well controlled.

II. ANALYTICAL TECHNIQUES FOR GAS AMBIENT ANALYSIS

The two analytical techniques utilized for this investigation were gas chromatography and mass spectrometry. Each has distinct advantages and disadvantages when applied to the analysis of micro quantities of gas within electronic packages. A mass spectrometer is a molecular sorter based solely on the weight of the molecule. It also generates a fragmentation pattern which is uniquely representative of the molecule. Figure 1 shows a simplified diagram of a 180° single focusing mass spectrometer. Each chemical compound has a unique repeatable fragmentation pattern and sensitivity. Thus, both qualitative and quantitative examination of a mixture is possible. No prior knowledge of the sample composition is necessary since all samples are run under the same conditions. Routine analysis can be made for gases such as hydrogen, helium, oxygen, nitrogen, benzene, and for as many as 25 other compounds in a single run. A view of the CEC 21-103C mass spectrometer used in this program is shown in Figure 2.

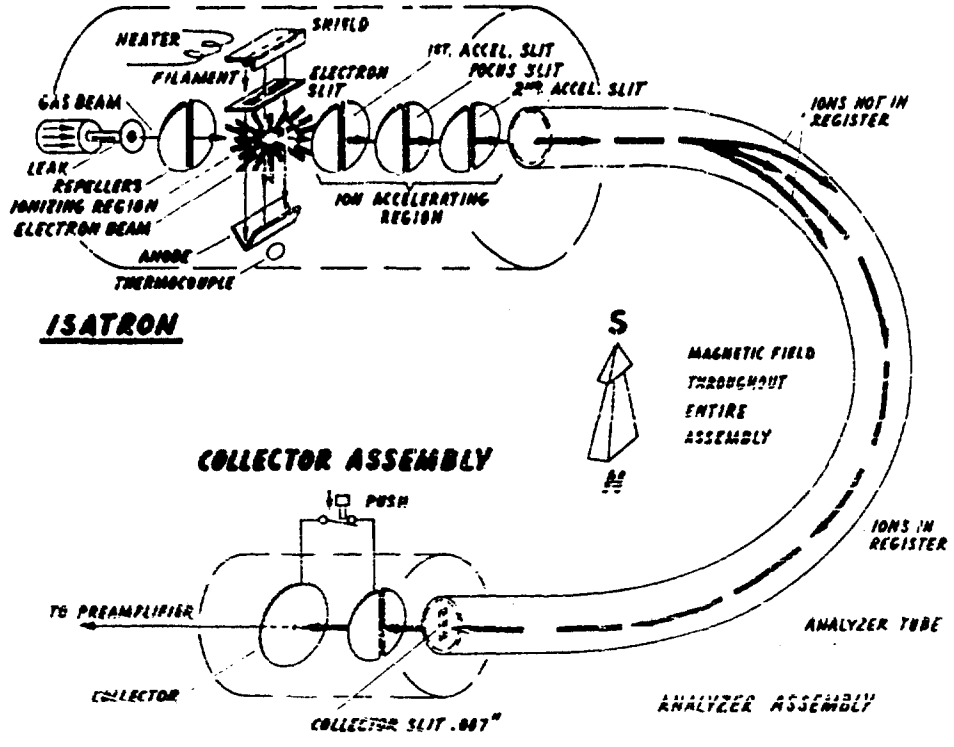


Figure 1

Simplified Diagram of a 180° Single Focusing Mass Spectrometer

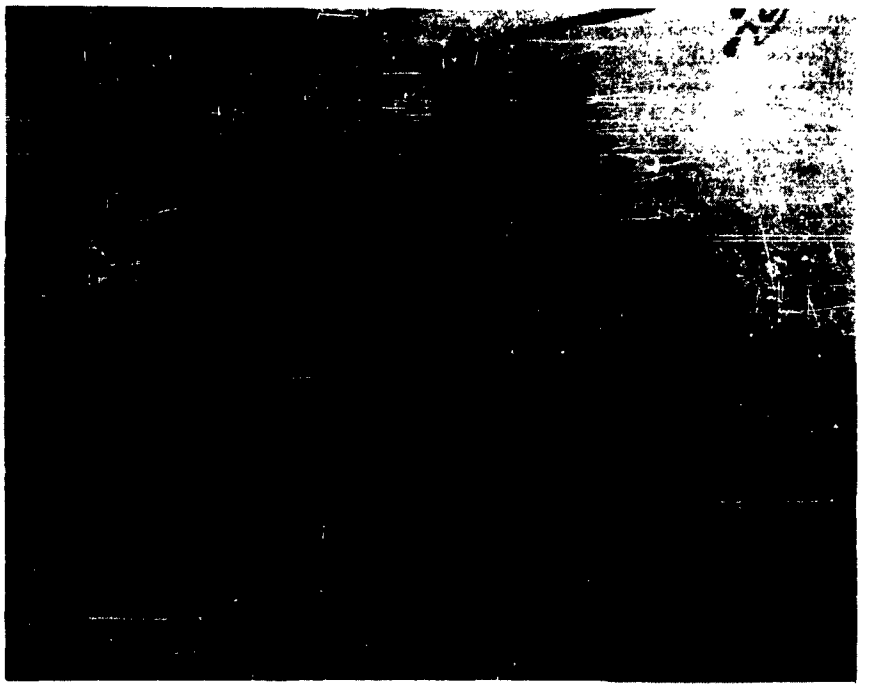


Figure 2

The Mass Spectrometer Installation

The volume of gas in a TO-50 flat package is 5000 micron-cc which is well within the sample size requirements of 30 micron-cc for the mass spectrometer. The gas quantity unit, micron-cc, refers to one cubic centimeter of gas at a pressure of one micron of mercury. This is the result of multiplying gas volume in cubic centimeters by the pressure in microns of mercury. There are 16.4 cc per cubic inch and 760,000 microns per standard 29.9-inch atmosphere.

The apparatus shown in Figure 3 was used with the mass spectrometer for the encapsulated gas analysis studies. The device was mounted on the Teflon pedestal and positioned so that the puncturing point entered the package without damaging the device. The apparatus was connected to the mass spectrometer inlet system, and the system was evacuated; the point was then advanced to puncture the device top and release the encapsulated gas into the mass spectrometer for analysis. Electrical connections were available via a vacuum seal in the base of the apparatus. The device leads were connected to an

external test circuit for continuous monitoring of electronic parameters before and during the removal of the encapsulated gas. Other leads furnished power for a small heater in the mounting pedestal which allowed heating the devices up to 200 C. Conventionally, a 30-second period was allowed for puncture and release of the encapsulated gas.

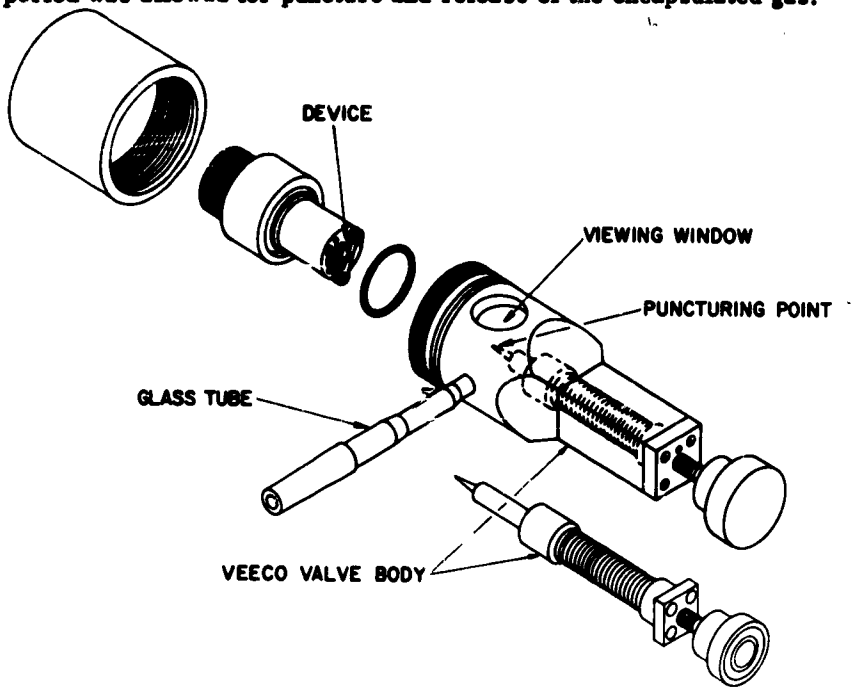


Figure 3

Apparatus for Device Package Ambient Analysis
Without Electrical Destruction

The apparatus shown in Figures 4, 5, and 6 has been designed for the testing of large numbers of parts at the sacrifice of continuous electrical parameter monitoring of the device. Even though the number of devices being tested was increased from one to ten, the evacuation time remained constant.

The single device apparatus, Figure 3, is adaptable to the determination of the effect of ambient gas upon electronic parameters by additional valving for backfilling. After the encapsulation gas had been removed, the system was backfilled with the desired gas while continuously monitoring the electronic parameters through the vacuum seal connections.

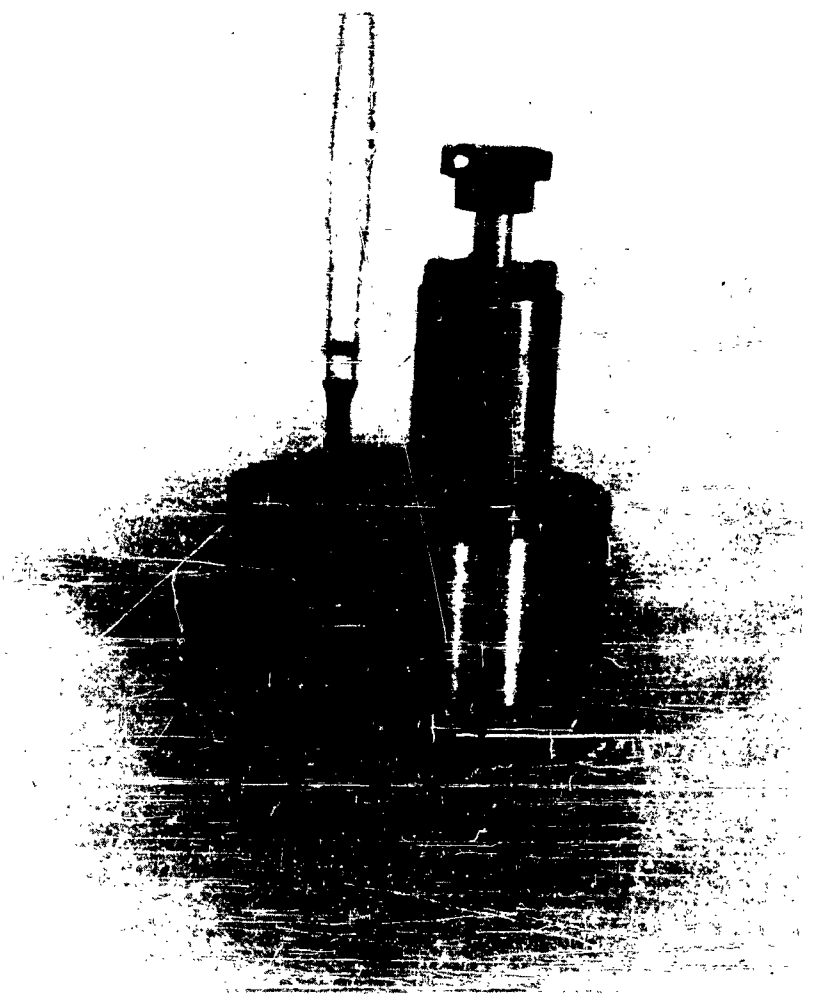


Figure 4

Outer Appearance of the Multipuncturer (With This Device, 10 Packages May be Punctured Consecutively)

Figure 5

Inner Appearance of the Multipuncturer

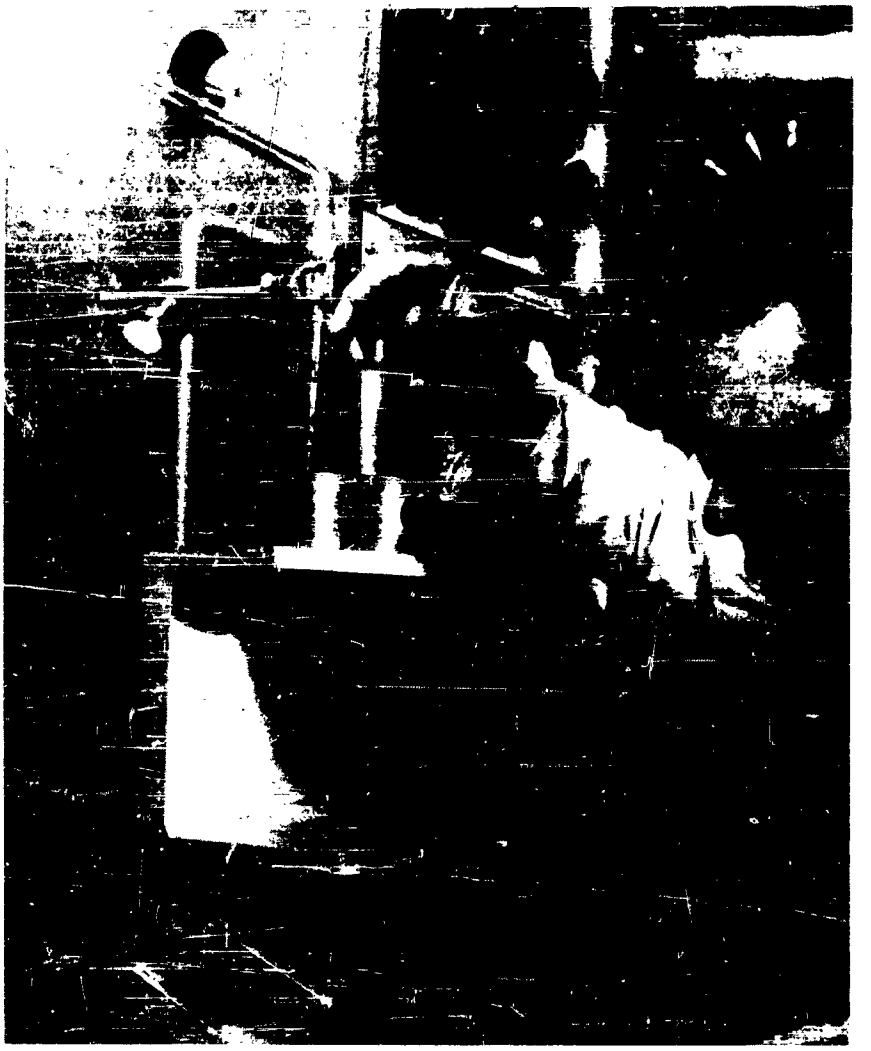


Figure 6

The Multipuncturer, Connected to the Mass Spectrometer, in the Process of Puncturing a Package

The gas chromatograph may be best compared to a distillation column. Figure 7 shows the progress of an analysis in six time cross-sections. In the first or top line, the sample mixture has just been injected into the flowing carrier gas stream. In the second, the sample slug had advanced into the column impelled by the carrier gas, and component A, by virtue of its more rapid transit character, is separating from B. In the third, A and B are well separated, and in the last three, A and B successively pass through the detector causing excursions of the recorder pen. In general, the area under the peak is a measure of the gas quantity. The separation columns must be carefully selected to separate the compounds of interest. Thus, a column suitable for oxygen and nitrogen separation holds carbon dioxide and heavy hydrocarbons indefinitely. A column suitable for hydrocarbons of the C_5 range will not separate hydrogen, air, and methane, etc. It is therefore evident that if there is only a small amount of gas available such as that within a device, prior knowledge of the sample composition is mandatory. It is possible to have components in a package which are not separable in a single run due to column limitations. The sensitivity of a gas chromatograph varies greatly with the thermal conductivity of the compounds to be detected. Typical easily attained values for oxygen and nitrogen are 3 micron-cc or about 10 times the sensitivity of the mass spectrometer. Hydrocarbons may be analyzed by another type of detector and may be detected in quantities as small as 1/1000 of a micron-cc.

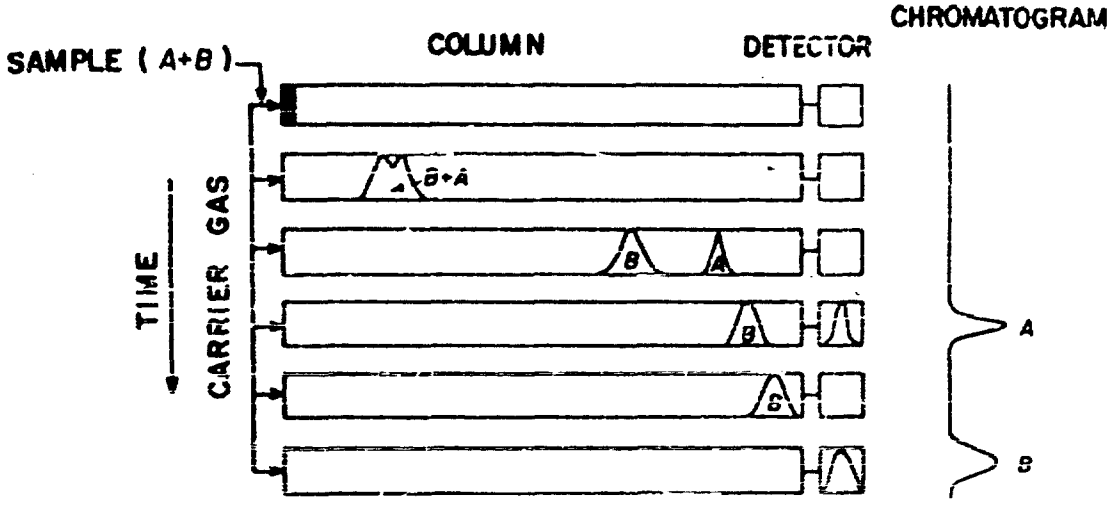


Figure 7

Gas Chromatograph Analysis

Most devices having leaky seals may be identified by their lower pressure or entire lack of gas. During the evacuation of the opening apparatus, the outside of the device is exposed to a very low pressure (10^{-5} Torr). The encapsulated gas will be withdrawn through the leak at a rate proportional to the size of the hole. Assuming complete equilibration with atmospheric pressure, a TO-50 would contain approximately 11,000 micron-cc of gas. An initial leak rate of 2×10^{-7} cc-atm/sec would halve the internal pressure in the 15-hour evacuation period, and a leak rate of 2×10^{-6} would not leave enough gas to be measured under these experimental conditions. Hermeticity testing of devices before gas analysis will, however, screen out leakers.

The determination of water warrants special mention. Surface adsorption of water in the sample handling system complicates the water determination. It is very easy to lose 2×10^{-7} grams of water, which represents 5 gas volume percent, on the walls of the sample system. Therefore, the data reported as gas volume percent water must be viewed as a minimum value. The numbers are comparative, however, and indicate that one device contains more water than another.

III. SEMICONDUCTOR DEVICE GAS AMBIENT EXPERIMENTS

A. Minuteman Transistor Gas Ambients

The gas content in packages has been reported as being an important factor contributing to electrical failure. The objective of this program was to discover what relations existed between the types of gases found enclosed in transistor packages and their I_{CBO}, h_{FE}, and other electrical characteristics.

A summary of the gas contents identified in several types of CQAP Minuteman transistors is shown in Table 1. The percentages represent average values of all the units of each device type analyzed. The data shows the predominant encapsulating ambient to be N_2. However, assuming the manufacturer's attempt to package with pure nitrogen (100 percent), the values range from 87.5 percent to 99.7 percent. An exception is the power transistor (Manufacturer Y) who packages in an atmosphere containing a high percentage of helium as an aid in leak detection. Analyses of device ambients consistently include CO, CO_2, O_2, H_2, and ssorted hydrocarbons.

Table 1

Gas Ambients in Minuteman Transistor Packages

	Transistor		Gas Volume Percent (Ave.)											Package Volume
No.	Type	Manufacturer	H_2	N_2	O_2	CO_2	CO	A	He	C_6H_6	CH_4	H_2O	Misc H-carbons	μ-cc x 10^3
1	Power	X	0.3	91.3	0.4	5.6	2.1	0.05	-	0.25	-	-	-	130
		Y	3.0	9.9	0.15	3.2	2.6	0.05	80.5	0.4	-	-	0.2	210
2	PNP Signal (Dual chip)	X	3.3	87.5	2.4	2.1	1.9	0.8	-	-	0.5	2.5	-	4.0
		Y	0.4	96.6	0.4	0.3	1.6	0.1	-	0.2	0.4	-	-	3.7
3	NPN Signal	X	1.0	98.5	0.1	0.2	0.1	-	-	-	0.1	-	-	6.7
		Y	0.1	98.6	-	0.4	0.5	0.1	-	-	0.3	-	-	5.5
4	NPN Signal (Dual chip)	X	0.1	94.0	0.9	0.3	0.5	0.1	-	-	-	-	4.1	4.0
5	PNP Signal	X	1.6	96.6		0.2	0.3	0.1	-	-	0.1	-	1.1	5.9
6	Power	X	0.08	99.7	0.03	0.1	0.07	-	-	0.04	-	-	-	1100

Measurable quantities of moisture were detected in the package of only one manufacturer. However, as mentioned earlier in describing the equipment, it is believed that moisture from the sample is condensed inside the mass spectrometer. It is estimated that in a typical flat type package, moisture contents below 5 percent by volume may be undetectable due to the adsorption on the walls of the equipment. Tests are currently in process to quantitatively determine this adsorption loss. Samples of various gases will be saturated with moisture at carefully controlled temperatures and the quantity of moisture determined using dew-point apparatus. Samples will then be placed in the mass spectrometer at a given temperature and analyzed using normal procedures. The difference in the moisture content before and after the analysis will be attributed to adsorption on the system walls.

A summary of gas ambients noted in pnp signal dual-chip transistors from one manufacturer as a function of electrical failure indicators is given in Table 2. The results indicate deviations from the planned pure N_2 ambient. Water vapor was found in I_{CBO} and h_{FE} failures but not in electrically good devices. Also, good and I_{CBO} failed devices had less O_2 than the other two types of failed devices.

Table 2

Gas Ambient in PNP Signal Transistors from Manufacturer X as a Function of Electrical Condition

Average Percent of	Good Devices	I_{CBO} Failures	h_{FE} Failures	Beta Ratio Failures	Empty Packages
H_2	3.2	2.4	1.5	1.4	-
CH_4	0.6	0.2	0.6	0.5	-
A	0.7	0.7	0.9	0.8	0.5
N_2	90.0	80.7	91.0	90.3	89.5
O_2	1.8	1.5	3.3	4.2	8.8
CO_2	1.9	2.6	1.2	2.3	0.4
CO	1.8	3.4	1.2	0.5	0.8
H_2O	-	8.5	0.3	-	-
Average Volume (μ-cc)	4,300	4,000	3,400	3,900	4,000

In contrast, experiments at Motorola (Ref 3) have shown that an npn device h_{FE} degradation occurred when stored at 300 C in N_2 and would recover upon storage in O_2 at 300 C, or in an ambient containing O_2. The degradation was related to surface channel formation. It is possible that the mechanism described is not only a function of the ambient present but also dependent on the composition or partial pressure of each gas as described by Nicholson (Ref 4). Obviously, more work is needed to understand gas-surface reactions to facilitate the postulation of general models which will explain the apparent discrepancies among experimenters.

In an attempt to determine the origin of the contaminants in the packages of these transistors, the manufacturer was asked to seal some "empty" packages (without the active semiconductor die and related process steps). The results, also shown in Table 1, indicate the empty packages contained no detectable CH_4, very little H_2, less CO_2 and CO, and more O_2. These differences may be related to the absence of the die.

B. Effects of Temperature + Ambient + Bias on Transistors

A manufacturer X dual pnp signal transistor, determined to have the most stable leakage current among many tested, was selected for gas contamination studies. Gas analysis showed only nitrogen present in the package when it was opened. Table 3 shows the variation of I_{CBO} for the two transistors in this package when the transistors could be inverted under normal stress conditions of -46 volts at 200 C for several hours. In Table 3, it is seen that even though A was biased while B was not biased before and after puncturing the can, both units were stable after a 150 C vacuum bake overnight. After the can was filled with hot benzene and baked overnight, the leakage current of A (biased) increased three orders of magnitude while that of B (unbiased) remained unchanged. After the benzene was pumped out and both transistors were subjected to hot water vapor over 60 hours, the biased unit degraded almost another order of magnitude but the unbiased unit remained unchanged. The leakage current of the biased transistor was restored to its original low value after it was vacuum baked with no bias for 5 hours. Later the condition of A and B were reversed in that unit; A was not biased while unit B was biased. After a hydrogen bake of 150 C overnight, unit A was still unchanged while B degraded almost three orders of magnitude. A 5-hour 300 C bake without bias did not alter A and restored B to its original current value.

Table 3

I_{CBO} for Manufacturer X Dual PNP Signal Transistor Subjected to Various Conditions

(All Measurements in Nanoamps at Room Temperature with 20 Volts Reverse Bias)

Condition	Transistor A	Transistor B
1. Before vacuum bake-out	3.0	2.5
2. 150 C vacuum bake overnight with 120 volts sweep bias on A but no bias on B	3.0	2.5
3. Heated to 150 C, punctured at 150 C	2.5	2.5
4. Filled with hot benzene vapor overnight	2,400	2.5
5. Benzene pumped out at R. T., over weekend in hot water vapor	10,000	2.5
6. Five hour 300 C vacuum bake with no bias on A or B	2.5	2.5
7. Filled with hydrogen followed by 150 C bake overnight with no bias on A but 100 volts sweep bias on B, cooled to R. T. followed by vacuum of 10^{-3} Torr	2.5	30
8. Filled with carbon dioxide followed by 150 C bake overnight with no bias on A but 100 volts sweep bias on B, cooled to R. T. followed by vacuum of 10^{-3} Torr	3.0	650
9 Five hour 300 C bake with no bias on A or B, cooled to R. T. followed by vacuum of 10^{-3} Torr	3.0	3.0

The above results show that these types of transistors can be inverted under certain conditions only when bias voltage is applied, and that they can be restored to their original leakage current values by baking them without bias. These results agree with the surface inversion model of Atalla (Ref 5) in which surface charge separation is

induced by a fringing electric field. It is postulated that the surface charge distribution is dependent on the gas composition present.

C. Minuteman Integrated Circuit Gas Ambients

A survey of the ambients used in integrated circuits indicated considerable differences among manufacturers as indicated in Table 4. Major constituents of nitrogen, hydrogen, and oxygen were found in the three manufacturer's devices analyzed. Device to device variations were high indicating poor process control. For example, mass spectrometric results on integrated circuits from manufacturer Y showed up to 95 percent H_2 in the package ambients while the manufacturer indicated the ambient as pure N_2.

Individual device abnormalities not easily detectable from the average values in the figure are worth mentioning. A high percentage of manufacturer X's integrated circuits contained at least a trace of benzene (C_6H_6) which possibly originates as a curing agent for the silicone coating used to coat the die surface. Two manufacturer Y units showed the presence of Freon 12 (dichloro-difluoro methane) which was attributed to the use of this gas for gross leak testing by this manufacturer.

Manufacturer Y used H_2 as part of a reducing atmosphere during his lidding operations to effect a hermetic package. Thermocompression bonding experiments at Autonetics had indicated a drastic increase in beta (or h_{FE}) when forming gas (15 H_2/85 N_2) was used as a protective blanket during the bonding, usually at 330 C. To evaluate the observed changes, several devices were tested before and after "baking" at 330 C for different lengths of time, while being subjected to various ambients (forming gas, pure hydrogen, and pure nitrogen). The results indicated that the beta increase was due to the hydrogen. Experiments with samples from several vendors consistently indicated an increase in beta at different rates. This phenomenon was associated with H_2-oxide interactions which varied from manufacturer to manufacturer depending on the oxide-silicon structure which is a function of oxide processing variables. It must be pointed out that these experiments were on special test devices from various manufacturers and did not include actual devices from manufacturer Y.

Since so-called "hermetic" packages have a leak rate over a period of time, the package would become depleted in H_2 and result in electrical drift. The effect of a package leak, which would change the internal package ambient, was simulated in manufacturer Y devices by puncturing packages with a small hole. Initial electrical measurements before and immediately after showed no significant changes in the output transistor betas of this integrated circuit type (general purpose amplifier). It is planned to continue to take electrical measurements to establish whether or not changes will occur after longer exposure to

Table 4

Gas Ambients in Minuteman Integrated Circuit Packages

Manufacturer	Gas Volume Percent (Ave.)												Gas Volume μ-cc $\times 10^3$
	H_2	N_2	O_2	CO_2	CO	A	He	C_6H_6	CH_4	H_2O	CCl_2F_2	Other H-carbons	
X	-	69.7	1.8	1.0	-	0.1	2.6	21.0	-	2.8	-	1.0	6.2
Y	60.2	35.5	0.9	0.4	1.1	0.1	-	-	0.1	-	1.7	-	2.6
Z	-	41.4	52.8	2.1	1.4	0.4	1.7	-	0.1	-	-	-	3.4

nonhydrogen containing ambients. Also, sealed and unsealed units will be baked at elevated temperatures in order to accelerate changes.

Further reference to Table 4 indicates on first glance what might appear to be an error. The flat packs used by the three manufacturers have approximately the same internal pressure-volume (pv). However, the data indicates the gas volume of manufacturer A (6200 μ-cc) to be much greater than that of manufacturer Y (2600 μ-cc) and manufacturer Z (3400 μ-cc). Examination of the packaging processes used, however, explains this difference. Manufacturer X uses a room temperature welding operation, manufacturer Y uses a metal brazing operation at approximately 350 C, and manufacturer Z uses a glass sealing operation at above 300 C. After sealing, the gas in the heated parts cools which forms a partial vacuum inside the package and accounts for the lower observed gas pressure-volume.

D. Sources of Ambient Contaminants

The most logical source of contaminants in the enclosed package gas is the gas immediately outside of the package. The hermeticity of packages is a relative factor based on gas leak rates expressed in atm-cc/sec. High reliability semiconductor devices, such as used on the Minuteman Program, have a specification required leak rate of less than 1×10^{-8} atm-cc/sec of helium.

For one hole through a package wall .10 mils thick, the diameter of the hole for 10^{-8} atm-cc/sec is about 0.1 micron assuming viscous flow, molecular streaming, or diffusion. Therefore at this particular leak rate the leakage is as fast by interdiffusion of gases with no total pressure difference as by net mass flow with a pressure difference. At higher leak rates viscous flow becomes relatively more important and at lower leak rates diffusion becomes relatively more important.

A program was established based on the gas leak equations to examine leak rates near this value in terms of the effectiveness of a package to retard in-leakage of contaminants (as most packages are sealed with a partial vacuum). In performing this analysis, it was assumed that the leakage is due to molecular diffusion as contrasted with viscous flow. The equation used to describe the leak rate was:

$$\frac{dp}{dt} = K (P_o - P_i)$$

or

$$\ln \frac{P_o}{P_o - P_i} = kT$$

where:

P_o = outside partial pressure of gas

P_i = inside partial pressure of gas

t = time

$$k = \frac{R \text{ (leak rate) (cc/time unit)}}{\text{internal volume (cc)}}$$

$$R = R^1 \frac{\sqrt{\text{mw leak rate gas}}}{\sqrt{\text{mw gas in question}}}$$

A typical example is illustrated. Consider the data for manufacturer Y devices:

internal volume - 0.014 cm^3

external pressure O_2 - 15.8 cm

leak rate detector gas - helium

initial internal O_2 pressure - 0 cm.

The package internal pressure of O_2 as a function of time was determined for various leak rates between 1×10^{-7} cc-atm/sec of helium and 1×10^{-12} cc-atm/sec of helium.

The results, shown in Figure 8, show that at the commonly specified leak rate of 1×10^{-8} cc-atm/sec, a package from manufacturer Y should have one-half the external partial pressure (7.9 cm) inside the package 30 days after final sealing, assuming the package was stored in air. The data in Figure 8 indicates the rapid decrease in rate of oxygen entering the package as the leak rate decreases. At a leak rate of 1×10^{-10} cc-atm/sec, approximately 3.5 mm of O_2 would be inside the package after 3 years. Assuming the critical amount of oxygen needed to shift device properties outside of the 3-year specification endpoint is 7.9 cm (or 10.5 volume percent), this leak rate would be acceptable as far as oxygen is concerned. Consideration must be made of all possible gases entering the package and also of gases such as hydrogen exiting from the package.

Figure 9 shows the time necessary to reach 7.9 mm O_2 inside the package as a function of leak rate. If we assume that this is the amount of gas necessary to degrade the part outside the 3-year endpoint, we can then determine from this graph the useful life of the part as a function of the relative hermeticity of the package.

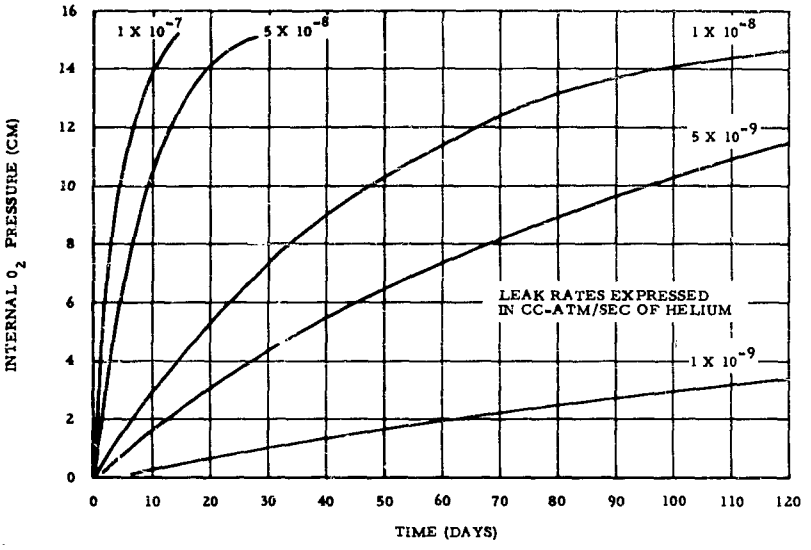


Figure 8.

Amount of O_2 in an Integrated Circuit Package Vs Time After Sealing for Different Package Leak Rates

Other sources of contaminants include outgassing of metal parts inside the package, residual organic solvents and cleaning solutions not thoroughly removed during cleaning, and gaseous products from chemical reactions.

IV. CORROSION OF ALUMINUM INTRACONNECTS IN INTEGRATED CIRCUITS

Integrated circuit failures with electrical opens or high resistance connections were analyzed and found to contain corroded aluminum intraconnects (Figure 10). Electron microprobe analysis revealed, in most cases, the presence of chloride ion at the point of corrosion, Figure 11. In two cases, the presence of fluoride ion (and no chloride) was noted. A program was initiated to study the failure mechanism in more detail and to determine the ability of normal burn-in and screen tests to cull out such corrosion-prone devices.

The proposed reaction was postulated to be as follows (Ref 6, 7, and 8):

$$6HCl + 2Al \rightarrow 2AlCl_3 + 3H_2\uparrow \quad (1)$$

$$AlCl_3 + 3HOH \rightarrow Al(OH)_3 + 3HCl \quad (2)$$

$$2Al(OH)_3 + \text{aging} \rightarrow Al_2O_3 + 3H_2O \quad (3)$$

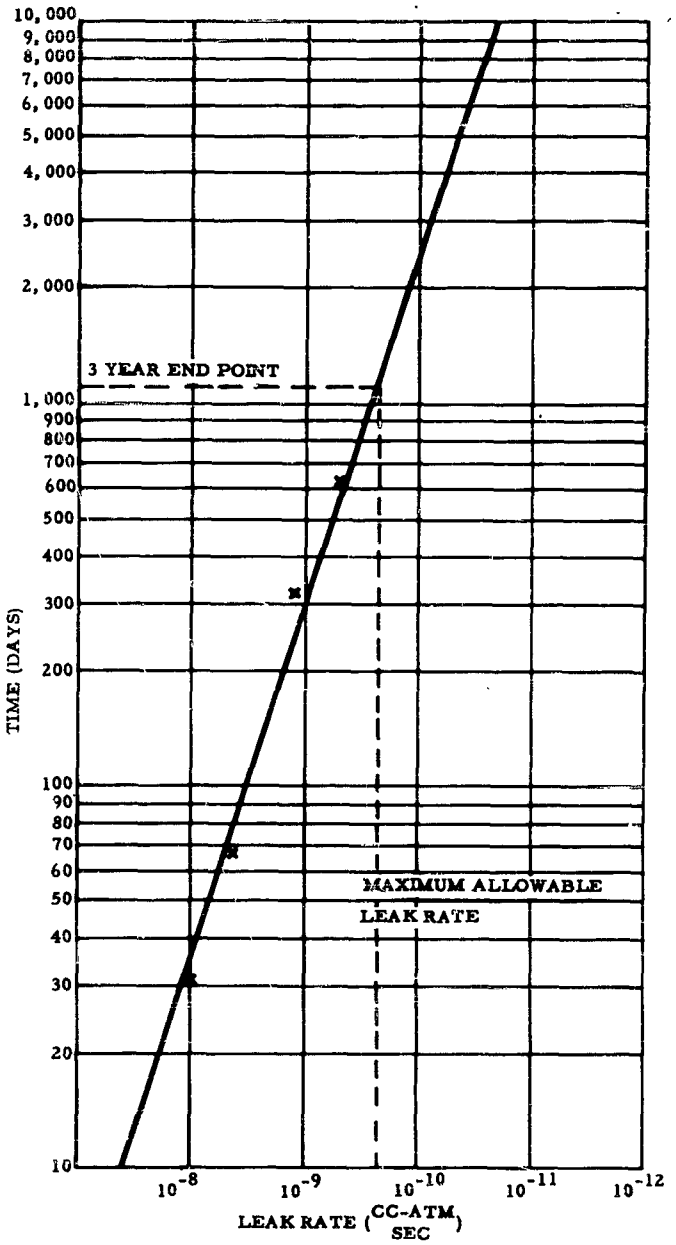


Figure 9.

Time Necessary to Obtain 7.9 cm O_2 Pressure Inside an IC Package Vs Leak Rate

Figure 10.

Photomicrograph of Aluminum Corrosion on IC Surface (104 X)

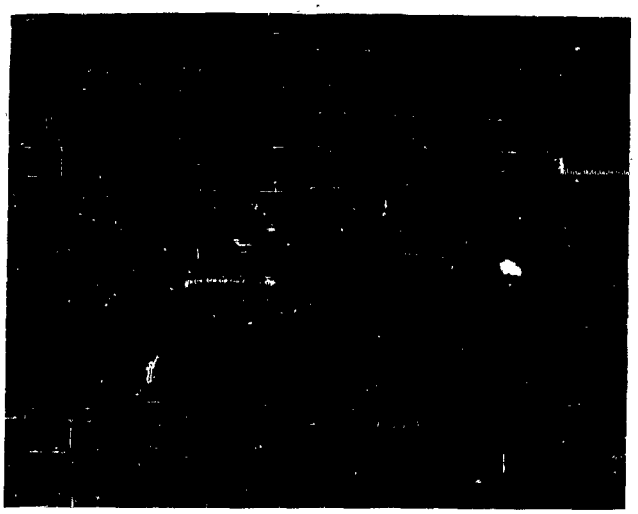


Figure 11.

Microprobe Analysis Showing Chlorides in
Corroded Region (280 X)

Analysis of the reaction equations indicates the cyclic nature of the mechanism. The aluminum initially reacts on the chloride ion (HCl) to form aluminum chloride (Eq 1). The aluminum chloride reacts with available water to form aluminum hydroxide and regenerate the HCl. The reaction can then start over. These reactions indicate that a specific amount of moisture is required in order to allow the reaction to progress. Otherwise, the amount of corrosion would depend on the amount of residual HCl in the package after sealing. That is, if the package was hermetic and perfectly dry, no corrosion would occur after the initial HCl had been used up in the reaction occurring in Eq 1 above.

The possibility of obtaining corroded devices in hermetic packages was evaluated as a partial verification of the failure mechanism. Approximately 50 devices were subjected to hermetic leak tests (both the helium fine leak test and the ethylene glycol bubble test). The devices were carefully opened so as not to allow contaminants to enter the package during the opening operation. The device surfaces were examined at 100 magnification for evidence of corrosion. Corrosion was noted on devices and was verified by electron microprobe analysis. Although the observed corrosion was not severe enough to affect the electrical properties at that time, further reaction could reduce the aluminum cross-sectional area with a resulting reduction in current carrying capacity.

Samples from the same lot, which were also hermetically tested but were not opened, were submitted for mass spectrometric analysis of the package ambient. Analysis showed a wide variation in the package ambients from device to device, but the major constituents of the gas were a H_2-H_2 mixture. In two out of ten devices, Freon gas (CF_2Cl_2) was detected. Microprobe analysis of the aluminum intraconnects of the device with the most Freon (37 volume percent) indicated extensive fluoride corrosion. Attempts to link the Freon with possible chemical corrosion phenomena have not as yet been successful since Freon 12 is reportedly highly stable and should not react with the aluminum under normal conditions.

Since devices with chlorine contamination were electrically good, it was necessary to investigate the possibility of the reacton continuing to cause failure at some later date. Environmental chambers which simulated the device package were built with feed-throughs for electrical operation, Figures 12 and 13. Devices with initial signs of corrosion and devices with no visible corrosion were placed in the chambers, and the chambers were evacuated and backfilled with those gases which were determined by mass spectrographic analysis to be

present. The devices were operated under normal conditions and checked periodically for further reaction. A vacuum sealed glass on top of the chamber allowed examination of the device surface on the metallograph without removal from the test chamber. Gas chromatographic analysis verified the gas composition in the test chambers and showed moisture content below 0.05 percent by volume. No increase i corrosion has been noted after 450 hours of operation.

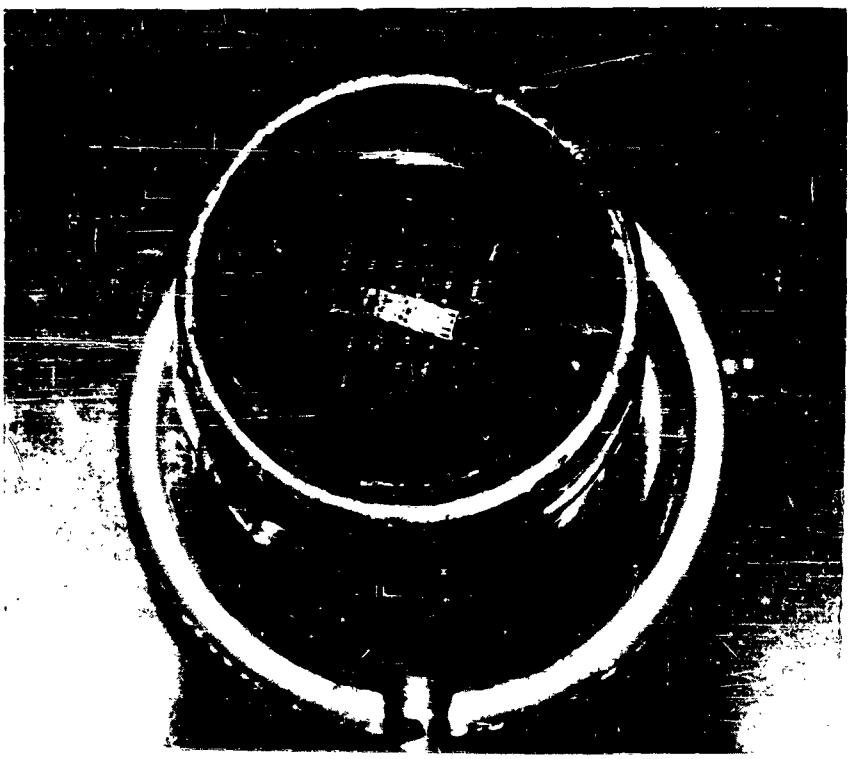


Figure 12.

Hermetic Chamber for Testing IC's in Selected Gases

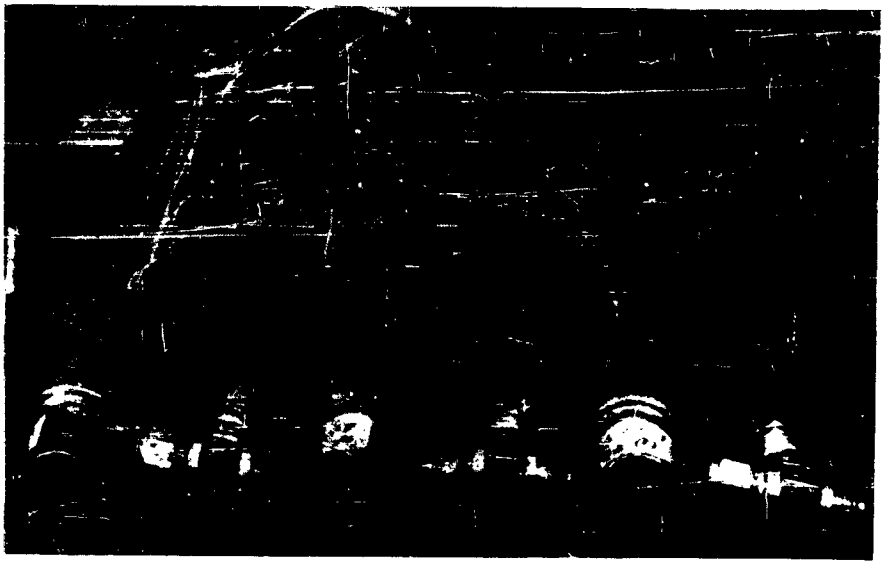


Figure 13.

Integrated Circuits Under Electrical Operation
in H_2-N_2 Atmosphere

A calculation was made to determine the critical amount of moisture needed to maintain the reaction and cause an open circuit. Corrosion was assumed to proceed from one edge of an aluminum stripe in all directions at an equal rate. The total corroded area would then be in the form of a semicircle of area $1/2\ \pi\ W^2$ where W is the width of an aluminum stripe. If we assume the stripe is 3 mil wide and 10,000 Å thick, the total amount of aluminum to be corroded is approximately 0.55×10^{-9} in.^{3} or 9×10^{-9} cm^3. Using an aluminum density of 2.7 gm/cm^3, the amount of aluminum to be corroded is 2.4×10^{-5} mg. Using Eq 1 and 2 above, 4.8×10^{-5} mg or 6.0×10^{-5} milliliters of water are needed to react with all of the $AlCl_3$ reproduced. This corresponds to a volume percent of 1.7 or a vapor pressure of 12.9 mm of H_2O. The corresponding relative humidity is 65 percent.

The thermodynamics of the reaction was studied to determine the possible effect of package hydrogen pressure on the reaction equilibrium. The free energy for the reaction at 300 K was found to be -85 K-cal per mole (Ref 9). Using the equation

$$-\Delta F_{300} = RT \ln K$$

where ΔF = the reaction free energy at 300 K
R = universal gas content
T = temperature in degrees Kelvin
K = equilibrium constant

In this case, K can be related to the pressure of H_2 so that

$$\ln_{10} P_{H_2}^{\ 3} = \frac{85}{2.3\,RT} = 62 \text{ or}$$

$$P_{H_2} = 10^{21} \text{ atm}$$

Obviously, this reaction is not affected by the package hydrogen (less than 1 atmosphere).

V. OUTGASSING OF PLATED METAL LAYERS IN TRANSISTORS

A possible source of voids in transistor die bonds, as well as package ambient contaminants, is the outgassing of plated layers on the header during the bonding operation, thereby forming gas pockets in the braze alloy, Figure 14. To study the contribution of plate outgassing to the formation of voids, samples of plated wire were sealed in evacuated glass vials, Figure 15, and then were heated at elevated temperatures for varying lengths of time. These vials were broken and the gases collected for gas chromatographic analysis. Moisture, hydrogen, and oxygen, as well as hydrocarbons including methane, ethane, and acetylene, were identified among the outgassing products of copper, nickel, and gold plates. A typical chromatogram of the hydrocarbon portion of gases evolved from gold plate at 550 C is shown in Figure 16.

The kinetics of the outgassing of gold plate were also studied. Sample vials containing gold plated wire were sealed at a pressure of 35 microns of mercury and then were heated at temperatures of 350, 450, and 550 C for 10 minutes, 1 hour, and 4 hours at each temperature. The volume of gold plate in each sample was determined, and an outgassing rate in units of milliliters of gas per cubic centimeters of plate per second was computed for each condition of test. These values were plotted versus the midpoint of the time interval in Figure 17 to obtain a family of rate curves. The volume of gas evolved by a given volume of plate was determined for each die bonding temperature and time. For example, the die of a power transistor is bonded at 520 C in an operation requiring a total of about 120 seconds. Using the dimensions of the transistor die, 0.250 x 0.250 inch, it was calculated that

Figure 14.

Gas Pocket in the Gold-Silicon Eutectic Layer of a Transistor

1.03×10^{-4} ml of gas was evolved in 120 seconds. The average volume of the gold-silicon eutectic bonding layer in this transistor is 6.14×10^{-4} cm^3, with a minimum value of 2.05×10^{-4} cm^3. Using these values, it was found that outgassing of the plate resulted in an average of 16.8 percent voids with a possible maximum of 50 percent voids in the bonding layer. Similar computations for a single transistor, bonded at 450 C for 60 seconds, indicated that an average of 5.6 percent with a maximum of 17.5 percent voids resulted from outgassing of the plated layers. It is apparent that outgassing of plated layers can account for voids in die to header brazements.

Figure 15.

Sample Vial for Plate Outgassing Study

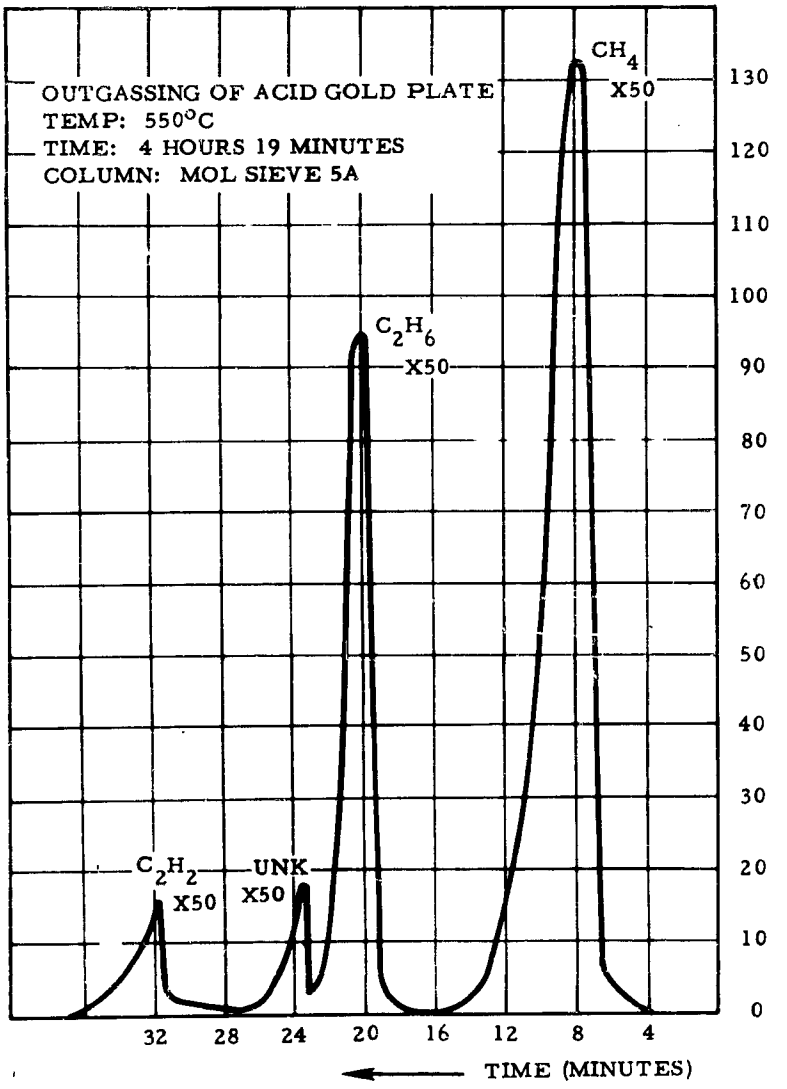


Figure 16.

Chromatogram Showing Hydrocarbon Portion of Outgassing Products from Gold Plate

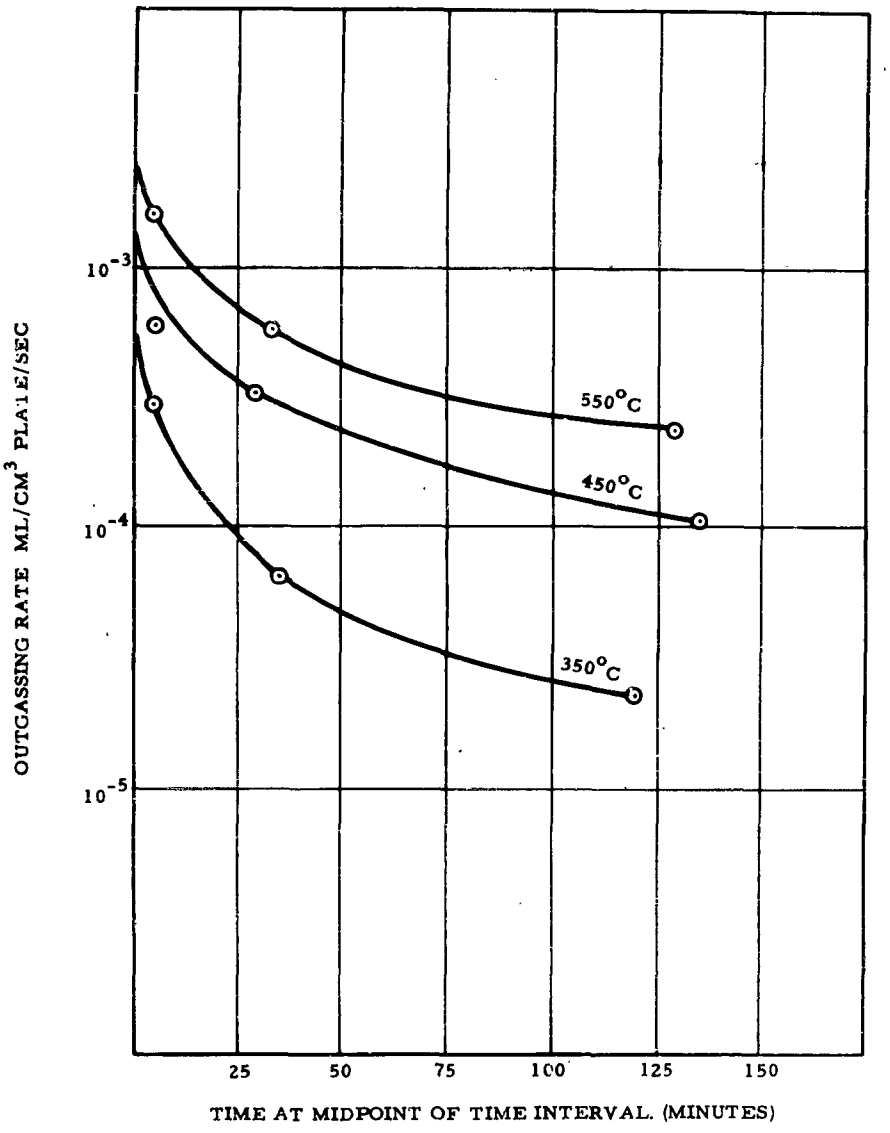


Figure 17.

Outgassing Rates of Gold Plate

VI. REFERENCES

1. R. H. Kingston, Semiconductor Surface Physics, University of Pennsylvania Press, 1957.

2. F. J. Biondi, Transistor Technology, Volume II, D. VanNostrand Company, Inc., 1958.

3. J. D. Kong, Detailed Study of Deleterious Effects on Silicon Transistors, Technical Report No. RADC-TR-65-35, Air Force, 1964.

4. D. Nicholson, Physics of Failure in Electronics, Volume 3, "Silicon Surface Leakage," Air Force, 1965.

5. M. M. Atalla, A. R. Bray, and R. Lindner, Proc. Inst. Elect. Engrs. (London), Pt. Suppl., 1959, Vol. 106, p. 1130.

6. J. Heyrovsky, J. Chem. Soc., 117 (1920).

7. R. Gerstl, Ber. 8 (1), 831 (1875).

8. Private Communication with J. McCloskey, Autonetics.

9. C. J. Smithells, Metals Reference Book, Butterworth & Co. Limited, 1962.

IMPERFECTIONS AND IMPURITIES IN SILICON ASSOCIATED WITH DEVICE SURFACE FAILURE MECHANISMS

by J. E. Forrester, R. E. Harris, J. E. Meinhard, and R. L. Nolder

Autonetics, a Division of North American Aviation, Inc., Anaheim, California

I. INTRODUCTION

Semiconductor devices depend for their operation on the optimum geometrical disposition of the bulk, surfaces, and interfaces of judiciously selected materials. Material selection traditionally has been based mainly on bulk properties on the assumption that surface and interface properties would have only minor effects on the device function or could be controlled by existing metallurgical and crystallographic techniques. As device miniaturization has progressed the active bulk to surface ratio has decreased steadily until surface and interface effects have assumed a dominating role in device function. Surface-related device malfunction problems have increased proportionately, becoming significant with the development of planar technology and acute with the inception of MOS technology. The present investigation addresses itself to the structural characterization of device malfunction originating at or near surfaces and interfaces.

II. EXPERIMENTAL APPROACHES

Two important topics considered in this investigation are inversion phenomena associated with ionic redistribution in dielectric layers, and silicon crystal imperfections situated at or near the silicon-silicon dioxide interface. Inversion effects were examined in relation to junction spreading using electron beam microprobe technique, reverse current leakage, 1/f noise, and the kinetics of ion redistribution in the oxide layers of planar transistors. A tentative chemical bond model was evolved to explain the observed effects. Silicon crystal imperfections were examined by transmission electron microscopy and correlated in part with phosphorus doping profiles obtained by electron beam microprobe analysis. These experiments were conducted on planar power transistors and revealed line dislocations in the region of P - N

junctions and other types of defects in the vicinity of the silicon-silicon dioxide interface.

III. INVERSION

The electron beam microprobe has been utilized in two general modes, both of which employ the impingement of a focused, traversing electron beam on the open device surface. By means of appropriate spectrometers one can identify x-radiations, and therefore the atoms emitting them, from micron-sized regions in the device material or intrusive particles of contamination. Some contaminant particles that have been discovered by this means include lead, sulfur, (probably as sulfate), chlorine (as chloride), and the migration of gold in aluminum contacts.

The second mode of operation pertains more directly to the observation of inversion and consists of monitoring the beam induced current as a function of focal location on the device surface. These currents are governed by the electrical impedances distributed between the beam impingement location on the surface and the ground connection to the device. By this means the surface geometry of planar devices is clearly and reproducibly delineated in terms of image density, as shown in Figure 1. The figure also reveals the spreading of the collector base junction associated with inversion. In this case the inversion was induced by the electron beam itself with an applied collector-base reverse bias of 45 V, and it appears as a white shadow contiguous with the original (dark) junction contour in the right-hand photograph. The associated leakage current increased from 32 μa to 200 μa. Continued beam exposure was found to produce further increases in leakage current until a maximum was reached, after which further treatment reduced leakage current to a point at or near that characteristic of the uninverted condition. These effects probably can be explained by the accumulation of static charges within the device.

Figure 1

Electron Beam Induced Current of Inverting PNP Transistor

For the purposes of this investigation reverse current leakage associated with inversion was specified as that produced under the combined influence of bias and baking at elevated temperature and cured by baking alone. Several examples of the behavior of small signal transistors under various temperature and bias treatments are shown in Figures 2 and 3. Reverse leakage currents (in na) are plotted as ordinates resulting from the sequence of temperature, bias voltage and duration of treatments tabulated horizontally. Transistors A and B (Figure 2) show characteristic inversion behavior; transistor C has a high reverse leakage which is not associated with inversion. Transistor D (Figure 3) shows erratic behavior with indication of a possible inversion component; transistor E is an example of a previously invertible specimen becoming increasingly more resistant to inversion. The last-mentioned phenomenon suggests the presence of electro-chemical processes occurring within the device which remove the ions responsible for inversion. Only devices behaving like transistors A and B were selected for correlative studies.

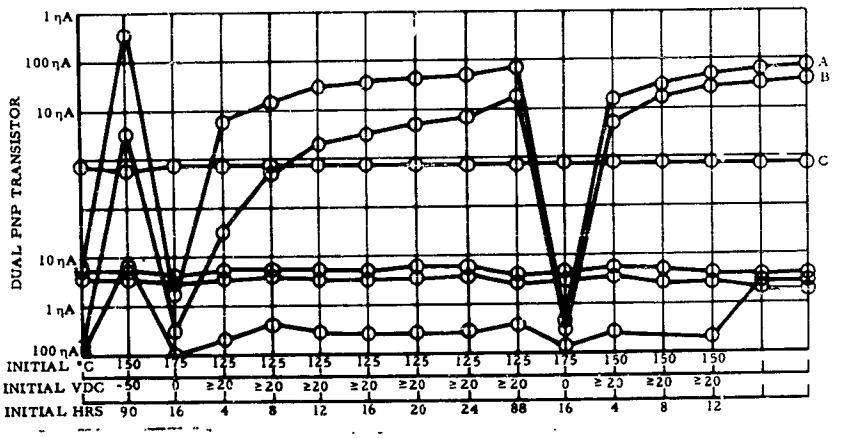


Figure 2

Collector-Base Leakage Current Through Several Cycles
(Sheet 1)

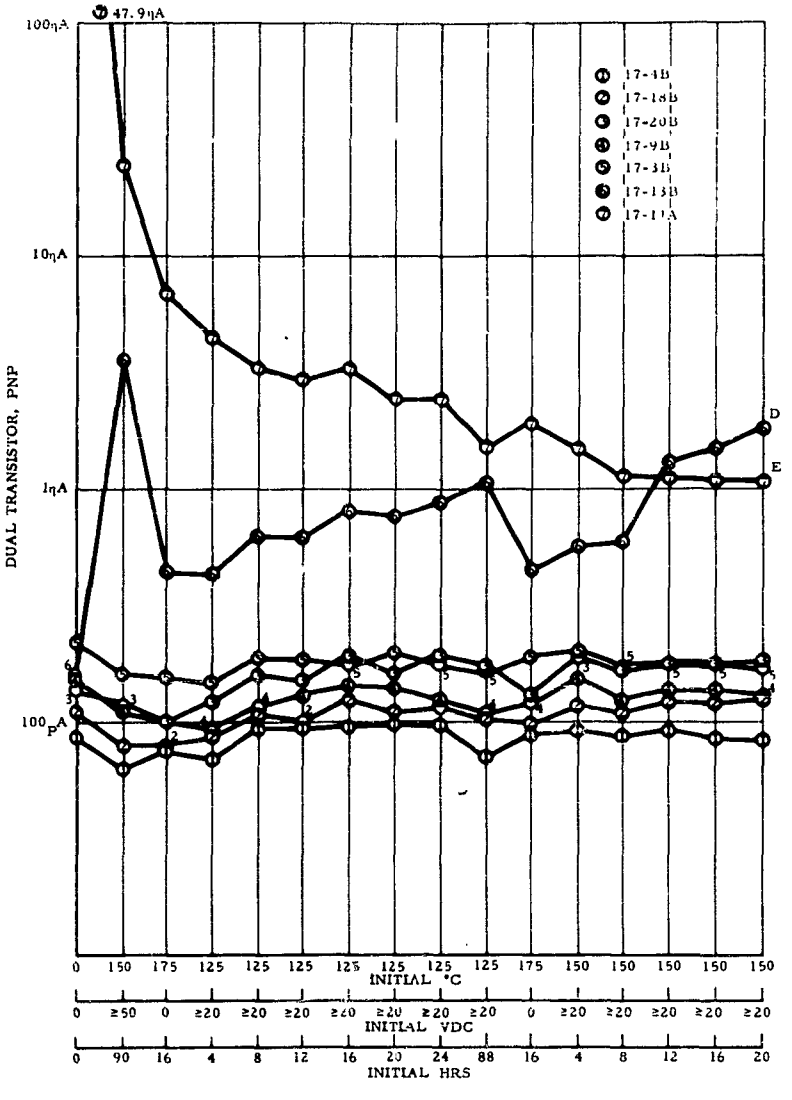


Figure 3

Collector-Base Leakage Current Through Several Inversion Cycles
(Sheet 2)

Correlation of inversion with noise current measurements was undertaken on the assumption that the oxide charge accumulation associated with inversion would induce surface states of varying recombination rates in the silicon. Such a correlation is made in Figure 4, which shows a series of oscilloscope traces of I_{CBO} at a 10 millisecond sweep rate with the corresponding 1/f noise current (I_N) indicated under each trace. The slightly blurred character of these traces near the origin arises from the fact that they were produced in a series of sweep-voltage steps. In the required current range it was necessary to operate the oscilloscope preamplifier at its highest gain factor which introduces a load distortion at ordinary sweep voltages (-2 to 150 V) and obscures the actual magnitude of the leakage current. Limiting the sweep initially (from 2 to 5 V) clearly defines the characteristics near the origin, and leakage is obtained as a definite step in current reading. Sweep voltage is then increased, making any necessary adjustments in vertical deflection, to obtain the remainder of the curve.

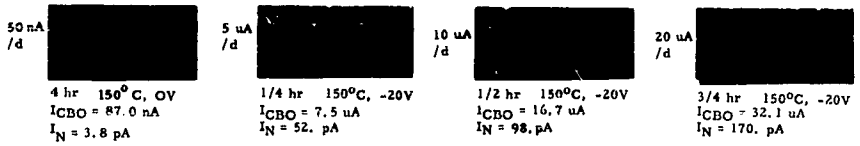


Figure 4

Reverse Characteristic of Transistor

The left hand trace in Figure 4 shows a partially recovered small signal transistor (of the type characterized in Figures 2 and 3) after baking four hours without bias. The 1/f noise, measured with a Quan-Tech Model 311 Transistor Noise Analyzer, is 3.8 pa. Succeeding traces show the effects of temperature-voltage treatments on this device. Data listed under each trace indicate increases in noise current corresponding to the increases in leakage current. Tracking of inversion leakage current by noise current is more clearly illustrated by the plot in Figure 5. However, not all normally-behaved inversions track 1/f noise so closely, indicating that more than one mechanism may be responsible for inversion leakage. The recovering transistor E (Figure 3) did not display a significant noise current, again suggesting the successive removal of ions capable of inducing surface states in the silicon.

The behavior of the leakage current with time, when the devices are subjected to selected stress condition appears to give better reproducibility and more information than any of the other electrical parameters. It is also the most sensitive parameter to changing conditions. Such kinetic investigations are not sufficient by themselves,

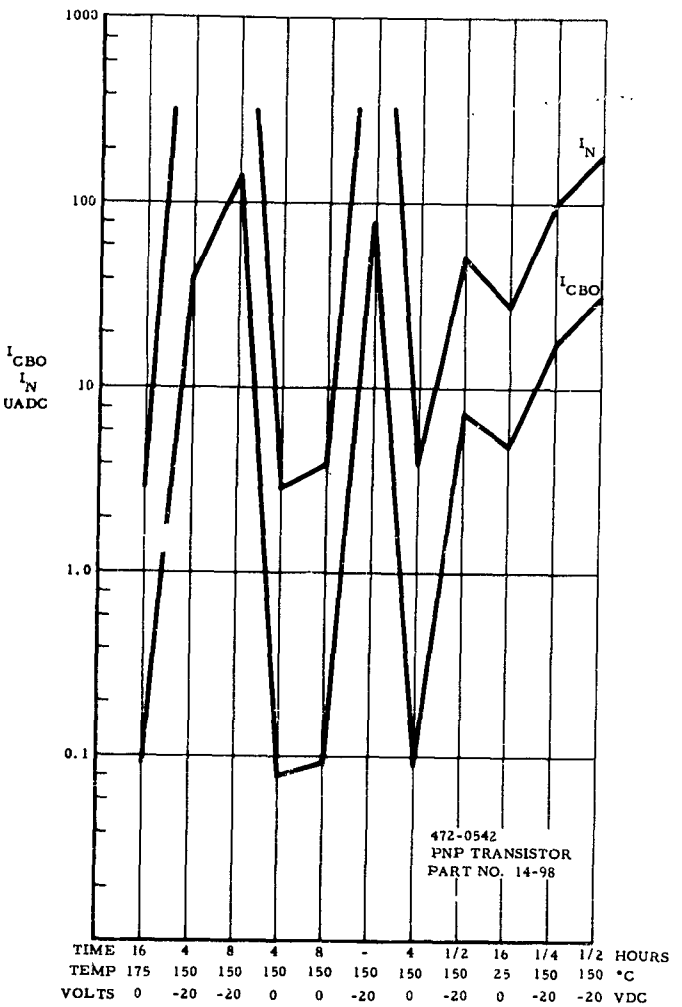


Figure 5

Correlation of Noise and Leakage Current

however, to provide positive identification of the migrating ionic species causing inversion or to reveal the structural features of the oxide layer contributing to the distribution and trapping of such ions.

Figure 6 illustrates a typical plot of I_{CBO} behavior under thermal stress and annealing conditions. Here a PNP small signal transistor was baked at 200 C with -46 volts bias for 30 minutes and then baked at 130 C without bias for the periods indicated in the figure. The leakage currents measured at room temperature show that the longer the bake without bias, the better the unit became. The high I_{CBO} after heating with bias is due to the increased area surface junction.

A model was developed for the study of this behavior assuming that the surface junction shrinks according to a parabolic law (implying a diffusion process) when subsequently heated without bias. The general junction geometry is illustrated in Figure 7. The junction area is given approximately by:

$$A \cong 2\pi r_1 \Delta r, \tag{1}$$

where,

$$\Delta r = r_2 - r_1.$$

Then from the original assumption,

$$\Delta r = \Delta r_{max} - (Dt)^{1/2}, \tag{2}$$

where D is the rate constant of junction shrinkage and is defined by:

$$D = D_o \exp(-\Delta E/_{kT}) \tag{3}$$

where ΔE is the activation energy of the diffusing species.

The activation energy expresses the amount of work required to move particles from one place to another in their environment. In the case of mobile ions on an oxide surface at a fixed temperature with constant electric field, like-charged ions with low activation energy will migrate faster than those with larger activation energy. Also, since $I_{CBO} = k_1 A$, the leakage current can be expressed as:

$$I_{CBO} \cong 2\pi r_1 k_1 \left[\Delta r_{max} - (Dt)^{1/2}\right] \tag{4}$$

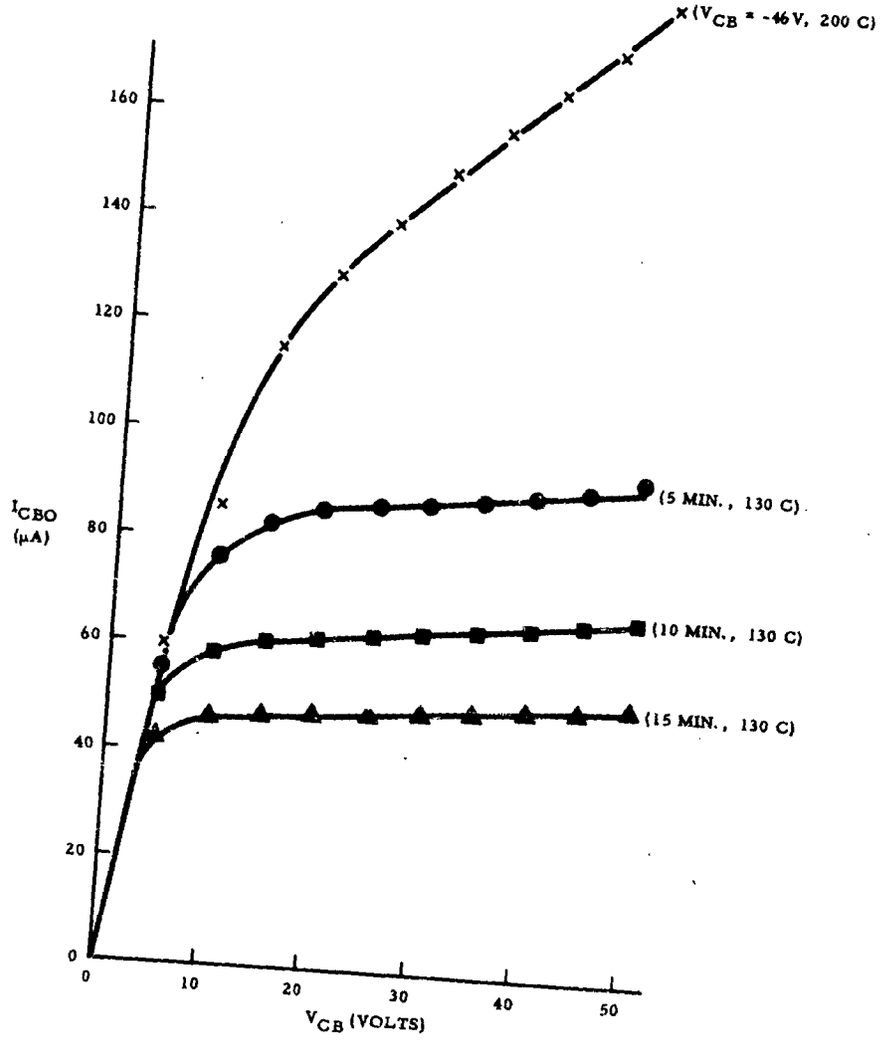
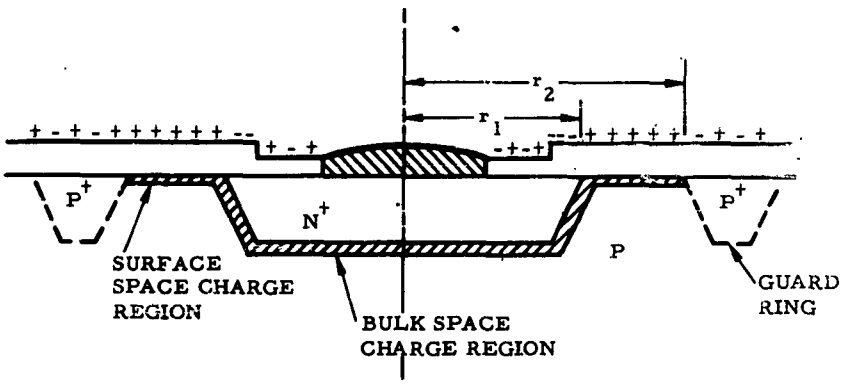


Figure 6

I_{CBO} Versus V_{CB} After Baking at 130 C for PNP Signal Transistor



$$\text{Area} \cong 2\pi r_1 \Delta r = 2\pi r_1 \left[\Delta r_{max} - (Dt)^{\frac{1}{2}}\right], \quad (1)$$

where $\Delta r = r_2 - r_1$ and D is the rate constant of junction shrinkage

$$I_{CBO} = \alpha A \cong 2\pi r_1 \alpha \left[\Delta r_{max} - (Dt)^{\frac{1}{2}}\right]. \quad (2)$$

$$D = \frac{(\Delta I_{CBO})^2}{\beta r_1}, \quad (3)$$

where β = constant and t_1 is the time for the current change

Figure 7

PN Junction Shrinkage Model

Relative I_{CBO} (the I_{CBO}, after heat treatment, divided by the maximum I_{CBO}, which is defined as I'_{CBO}) is now adopted:

$$\text{Relative } I_{CBO} = I_{CBO}/I'_{CBO} \quad (5)$$

where:

$$I'_{CBO} \cong 2\pi k_1 r_1 \Delta r_{max} \quad (6)$$

defining I'_{CBO} in terms of Δr_{max} (the location of the guard ring). The expression for Relative I_{CBO} then can be written as:

$$I_{CBO/I'_{CBO}} = 1 - \frac{(Dt)^{1/2}}{\Delta r_{max}} \tag{7}$$

From the data of Figure 6 (130 C bake) along with similar data for three other temperatures, the Relative I_{CBO} versus $(t)^{1/2}$ is plotted to obtain Figure 8. It can be seen from equation (7) that the slopes of the lines are proportional to the square root of the rate constant D at each of the temperatures. Instead of determining the rate constant D directly from the slope, the time, $(t)_{1/2}$ required for the Relative I_{CBO} to become 0.5 is used to determine the activation energy, ΔE, for the rate constant. The constant is expressed by:

$$D = \frac{\Delta r^2_{max}}{4\,(t)_{1/2}} \tag{8}$$

The logarithm of $(t)_{1/2}$ is then plotted as a function of reciprocal absolute temperature. The experimental values taken from Figure 8 will then produce the plot illustrated in Figure 9. From this graph the activation energy for the rate constant of the junction shrinkage was calculated to be 0.68 electron volts. Lee (Ref 1) reported that the activation energy for the diffusion coefficient of hydrogen producing hydroxyl in natural fused silica was 15.8 kilocalories per mole (0.68 electron volts). Also, Schmidt (Ref 2) reported that the activation energy of the formation of surface charge under vacuum or hydrogen aging is approximately the same as that for the formation and diffusion of hydroxyl in silica (15 kilocalories per mole).

If equation (4) is used for the leakage current, then the change in the leakage current can be expressed as:

$$\Delta I_{CBO} = I_{CBO}\,(t=o) - I_{CBO}\,(t=t_1) \tag{9}$$

Equation (9) gives the value of the rate constant as:

$$D = \frac{\left(\Delta I_{CBO}\right)^2}{\beta t_1} \tag{10}$$

where β is a constant ($\beta = 4\pi^2\, r_1^2\, k_1^2$) and t_1 is the time interval for the current change.

If the logarithm of D in equation (10) is plotted as a function of the reciprocal absolute temperature, then the activation energy can be calculated from the graph. Figures 10 and 11 are such plots for PNP

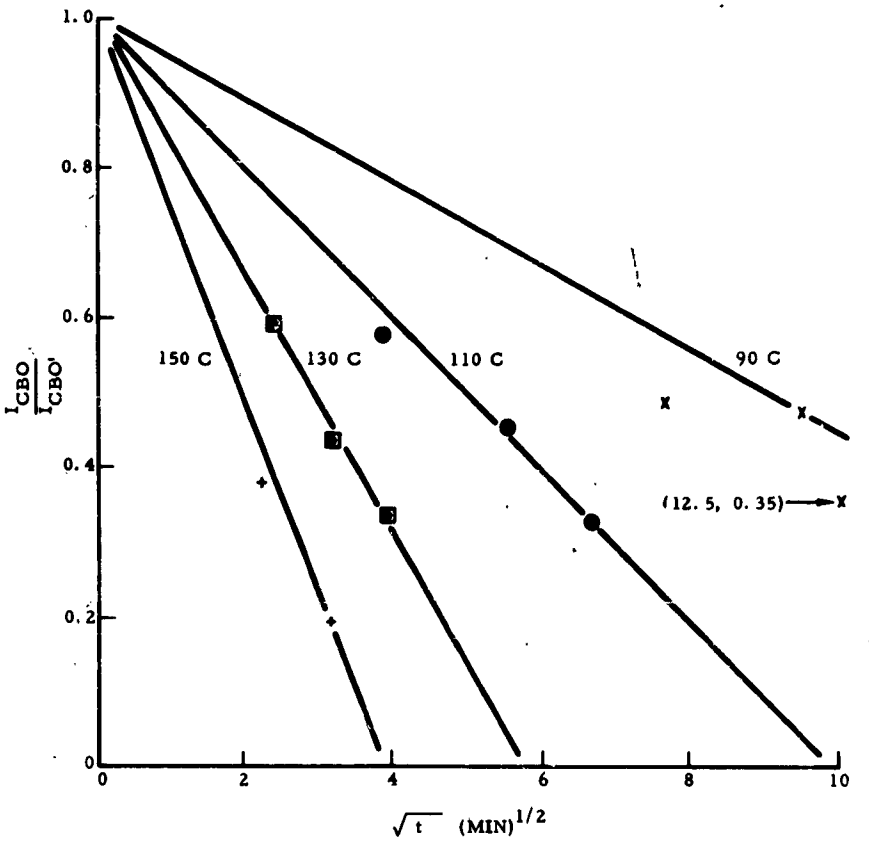


Figure 8

Relative $I_{CBO} = \frac{I_{(CBO)}}{I_{CBO'}}$ Versus Bake Time at Several Temperatures for PNP Signal Transistor

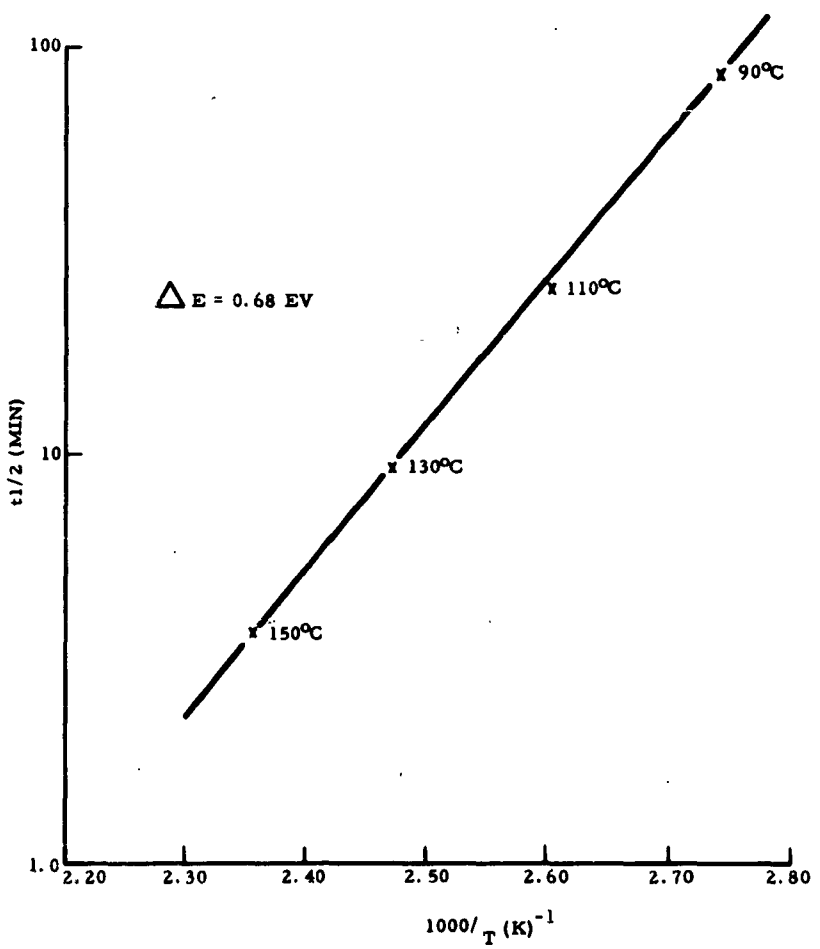


Figure 9

Time to Reach $\frac{I_{CBO}}{I_{CBO'}}$ = 0.50 for Four Different Temperatures for PNP Signal Transistor

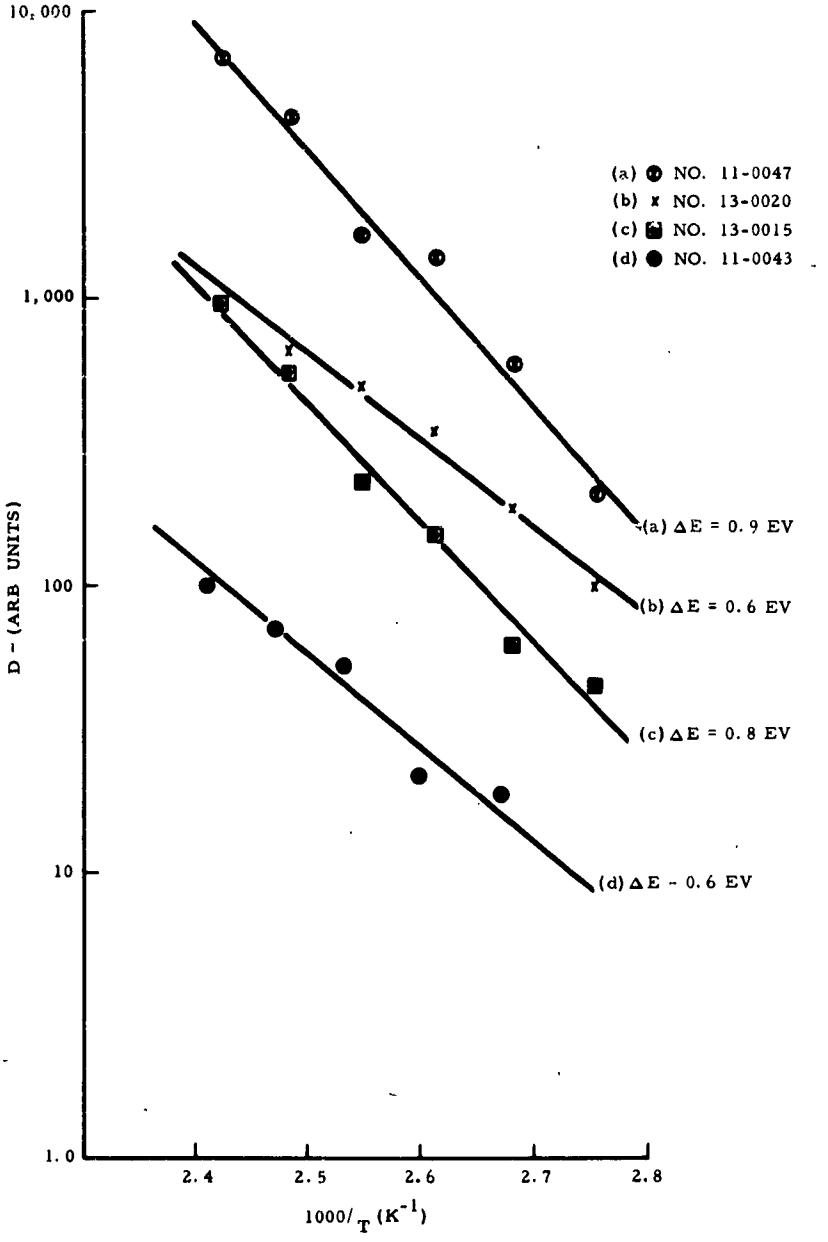


Figure 10

Diffusion Constant Versus Temperature for PNP Signal Transistors

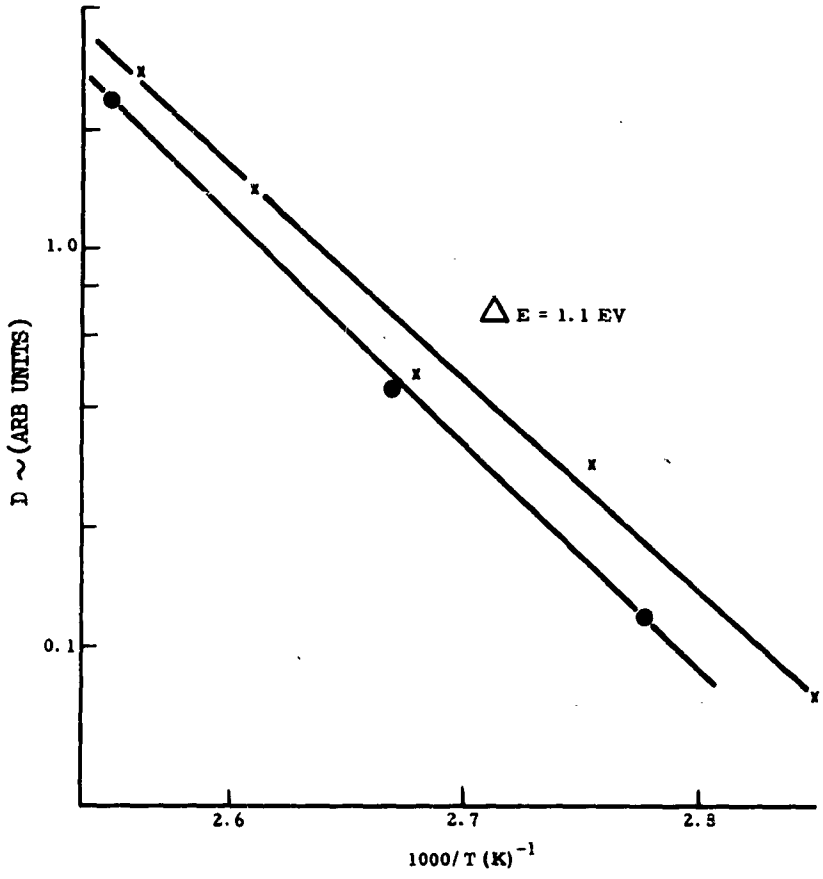


Figure 11

Diffusion Constant Versus Reciprocal Absolute Temperature for the Dual Transistors (A and B)

signal transistors and Figure 12 is for a computer microdiode. In Figure 10 are plots of diffusion constant versus temperature for several PNP signal transistors of the same type giving activation energies from 0.6 to 0.9 electron volts. Figure 11 illustrates typical plots of the diffusion constant versus reciprocal absolute temperature for two additional transistors from a different source for which activation energies of 1.1 electron volts were obtained.

NPN signal transistors also were studied. Figure 13 illustrates the I_{CBO} behavior of an NPN transistor at the indicated bias of -46 V and temperature of 150 C. This unit was stressed for 16 hours, and then baked without bias for time intervals indicated in the figure. The

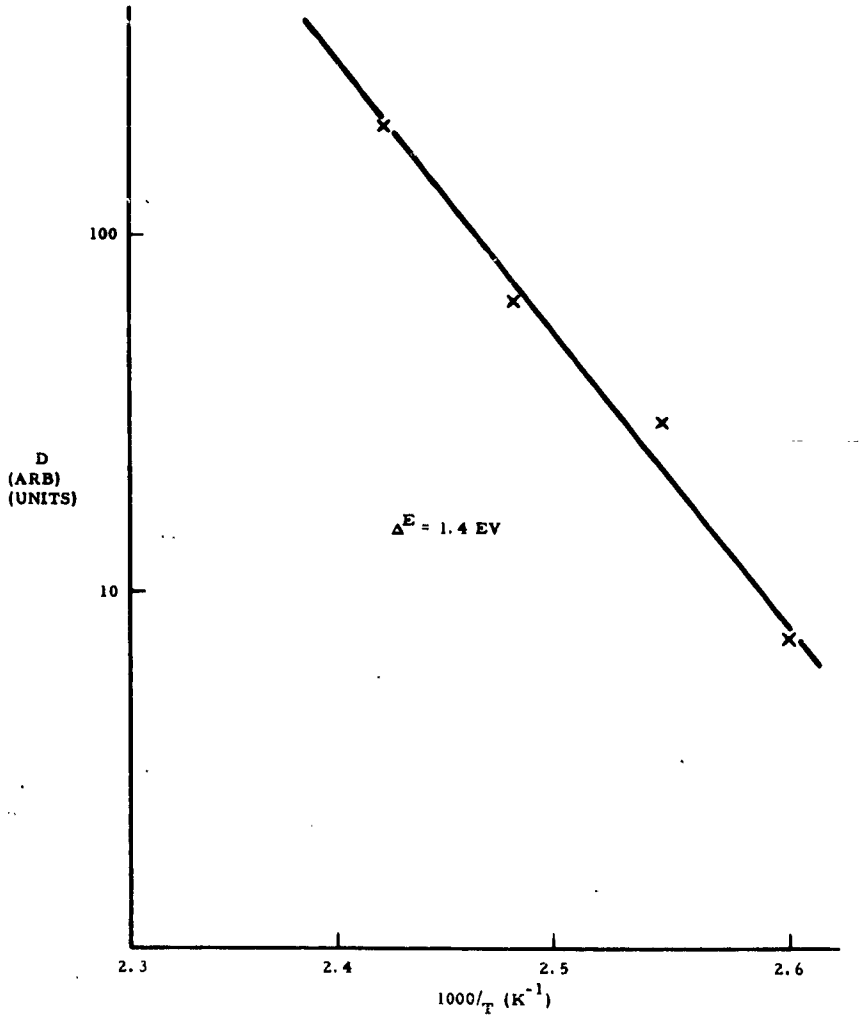


Figure 12

Diffusion Constant Versus Temperature for a Computer Diode

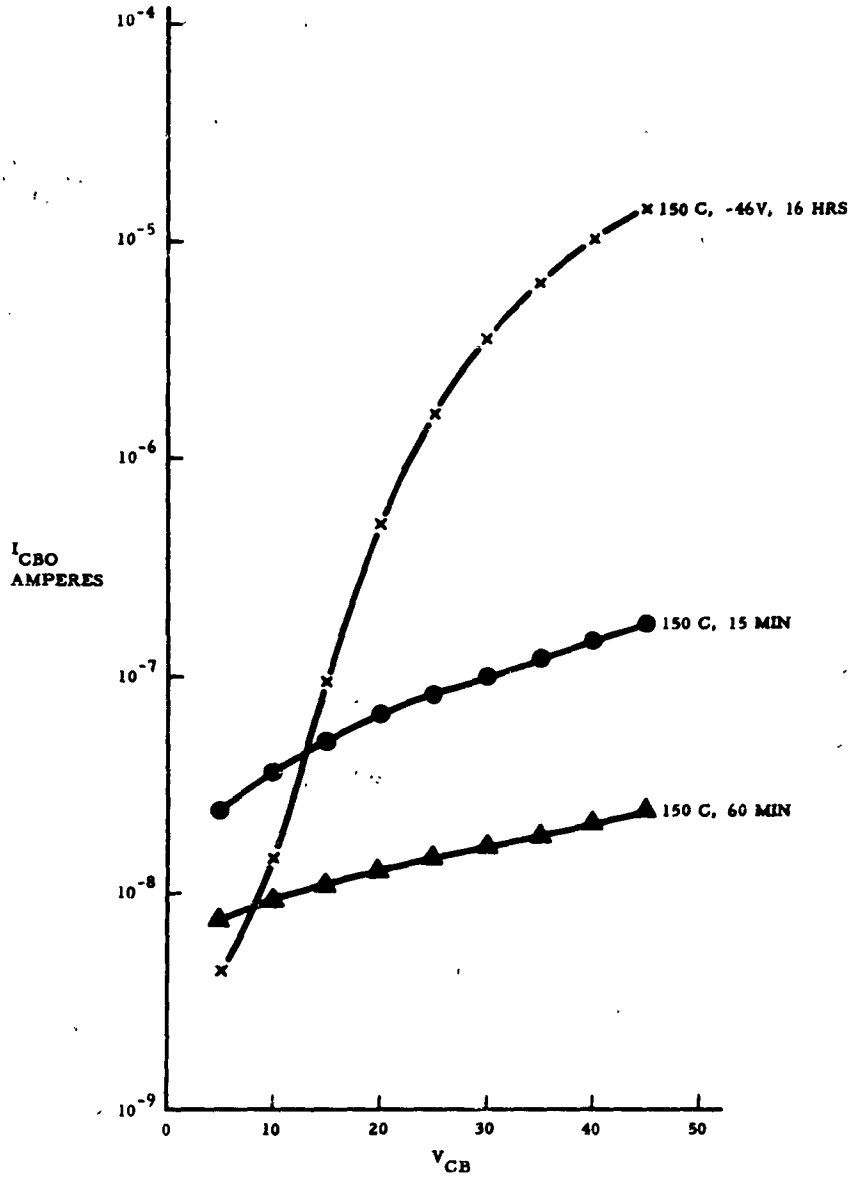


Figure 13

I_{CBO} Versus V_{CB} for NPN Signal Transistor

leakage current behavior of NPN devices was somewhat similar to that of PNP. These devices were more difficult to invert and not as sensitive to changing conditions. Only a few were selected for detailed study.

PNP signal transistors secured at a later date from the same source behaved differently from those previously obtained. After voltage and temperature stress, these devices were characterized by an I_{CBO} versus V_{CB} characteristic that shifts abruptly from a bulk dominated to a surface dominated characteristic at some value of V_{CB} as illustrated in Figure 14. Heating the device without bias shifts the various segments of the curve as indicated by the arrows at A, B, and C in the figure. The A region drops as it does in the other PNP transistors studied.

Using the data obtained, the following model is proposed. The dynamics of this model, like that of the previous one, are characterized by a thermal activation energy. Figure 15 shows a cross-section view of a PN+ planar junction terminating at the oxide surface. The position of the depletion layer is expressed by:

$$X = K_1 V_{CB}^{1/2} \tag{11}$$

It is assumed that under reverse bias at high temperatures there is an accumulation of positive charge on the p-side of the junction.

The position of the leading edge of this charge distribution as a function of annealing time t is given by:

$$X = W_o + K_2 t^{1/2} \tag{12}$$

Where it is assumed that the attrition of charge from the accumulated region is governed by a diffusion mechanism. K_2 is then proportional to the square root of the diffusion constant.

It is further assumed that for

$$K_1 V_{CB}^{1/2} < W_o + K_2 t^{1/2} \tag{13}$$

bulk leakage characteristics are observed, while for

$$K_1 V_{CB}^{1/2} > W_o + K_2 t^{1/2} \tag{14}$$

surface leakage current is dominating. The characteristic voltage V'_{CB} is then given by

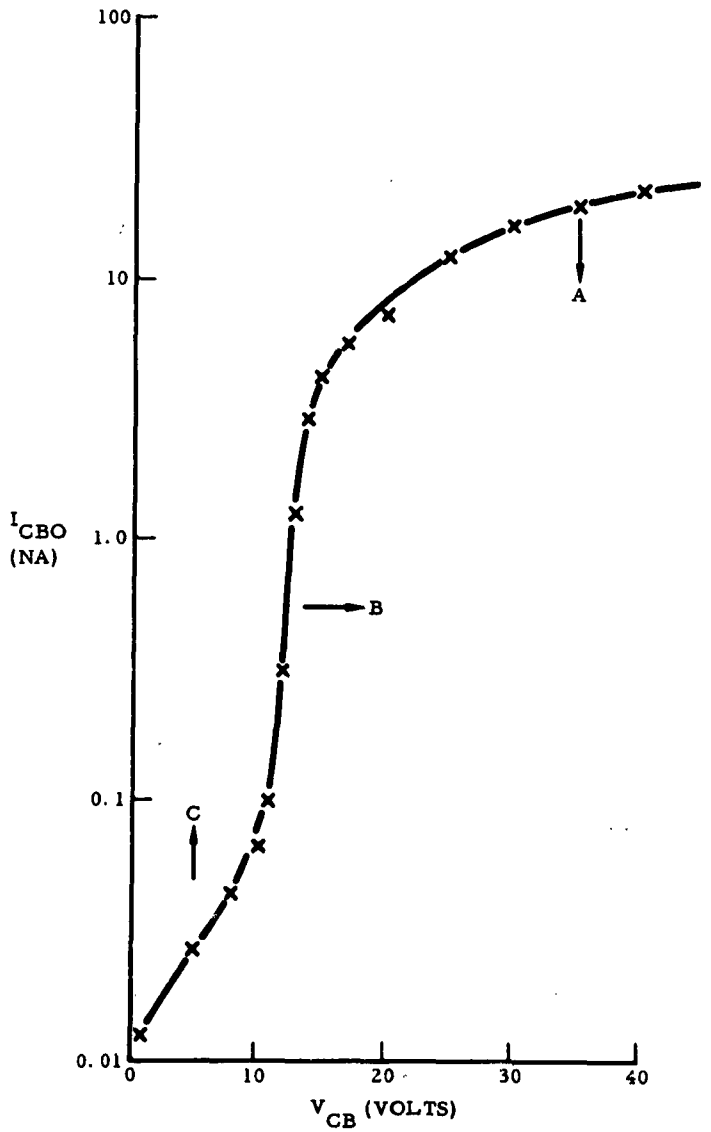


Figure 14

I_{CBO} Versus V_{CB} of PNP Signal Transistor Stressed for 100 Min. at 205 C, -46 V

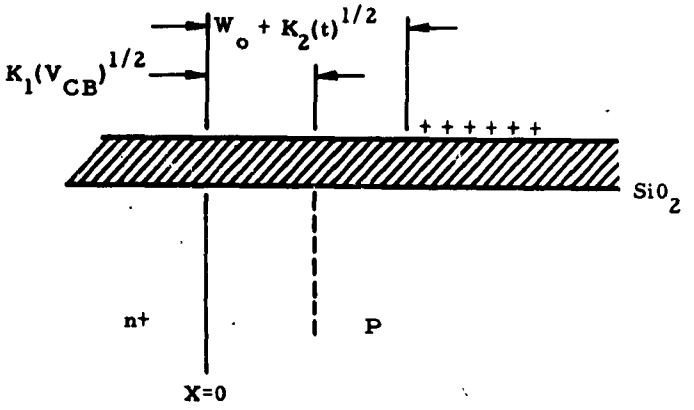


Figure 15

PN Junction Terminating at Oxidizer Surface

$$(V_{CB})^{1/2} = \frac{Wo}{K_1} + \left(\frac{K_2}{K_1}\right) t^{1/2} \qquad (15)$$

The ratio K_2/K_1 which is proportional to the square root of the diffusion constant for the charge redistribution is determined from the slope of the $V'_{CB}{}^{1/2}$ versus $t^{1/2}$ plots for different annealing temperatures. Figure 16 shows plots of $\sqrt{V'}$ (V_{CB} to get one nA) as a function of $\sqrt{t}$ for a PNP signal transistor for the temperatures 100 C and 129 C. Figure 17 shows $(K_2/K_1)^2$ versus reciprocal absolute annealing temperature. An activation energy of about 1.5 electron volts is apparent from this plot implying a thermal activation energy of 1.5 electron volts for the diffusion constant.

Figure 18 shows I_{CBO} versus reciprocal absolute temperature for a PNP signal transistor from a different source that has been stressed with reverse bias at high temperature and then annealed at a lower temperature. Here the temperature coefficient of the I_{CBO} increases and the I_{CBO} level decreases as the device is annealed. At each stress level, the curves can be described by an equation of the form:

$$I_{CBO} = A \exp\left(^{-\Delta E}/_{kT}\right) \qquad (16)$$

As can be seen from the table on Figure 18, almost all of the change in I_{CBO} can be accounted for by the change in activation energy (ΔE). In fact, the experimental limits of error do not exclude the

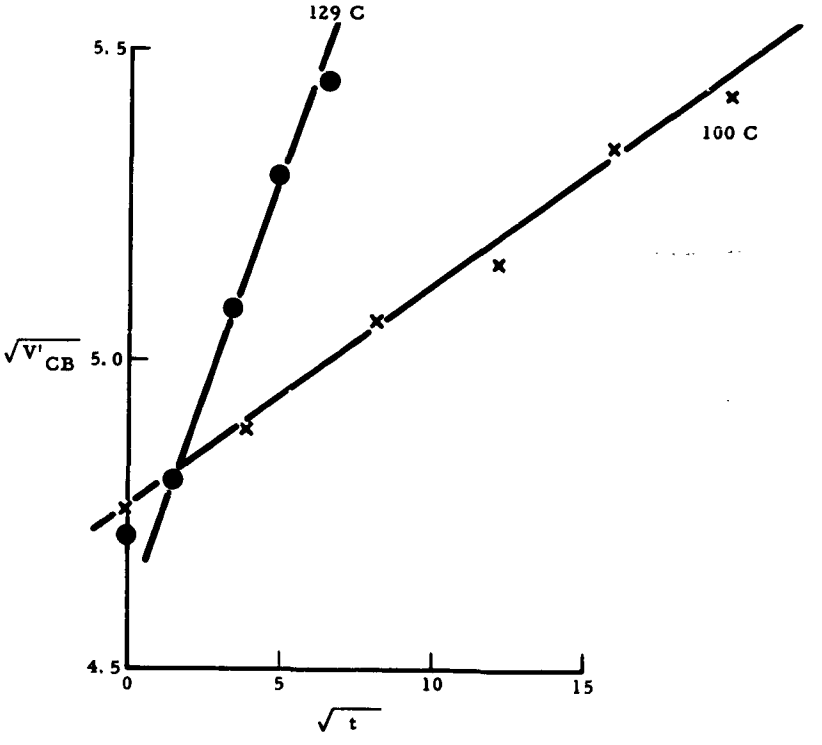


Figure 16

V_{CB} to get 1 na Versus Tine Graphs for PNP Signal Transistor Baked at Two Temperatures

possibility that the pre-exponential factor may be a constant. This result suggests that, for at least some transistors, temperature and bias cycling is accompanied by an irreversible process not accounted for in the previous models. This process, as observed earlier, is characterized by an increasing resistance to inversion. The origin of this effect is difficult to deduce without detailed knowledge of the processes used in manufacturing the device.

In spite of this difficulty it is possible to advance a structural model of the silicon dioxide dielectric layers capable of explaining these effects. This model is based on the assumption that the catalytic effect of steam on oxide growth involves the initial binding of hydroxyl

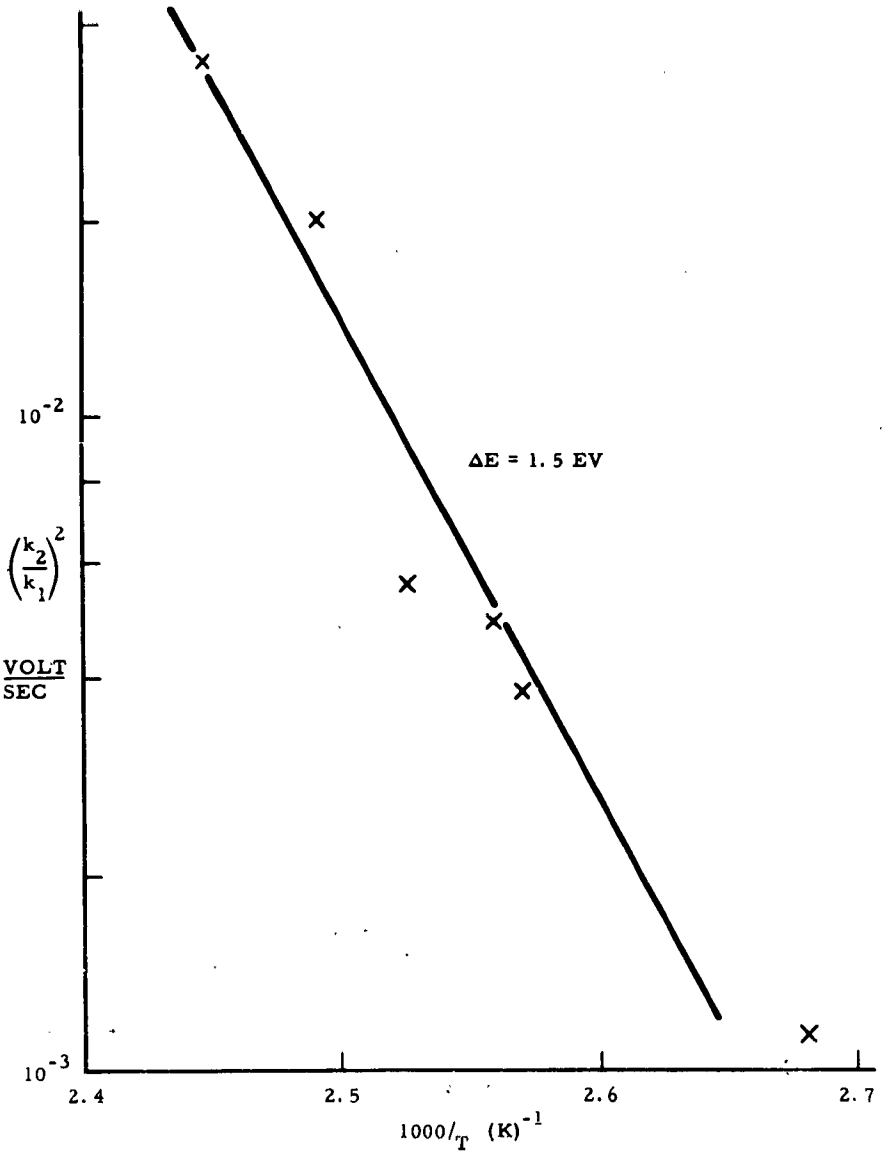
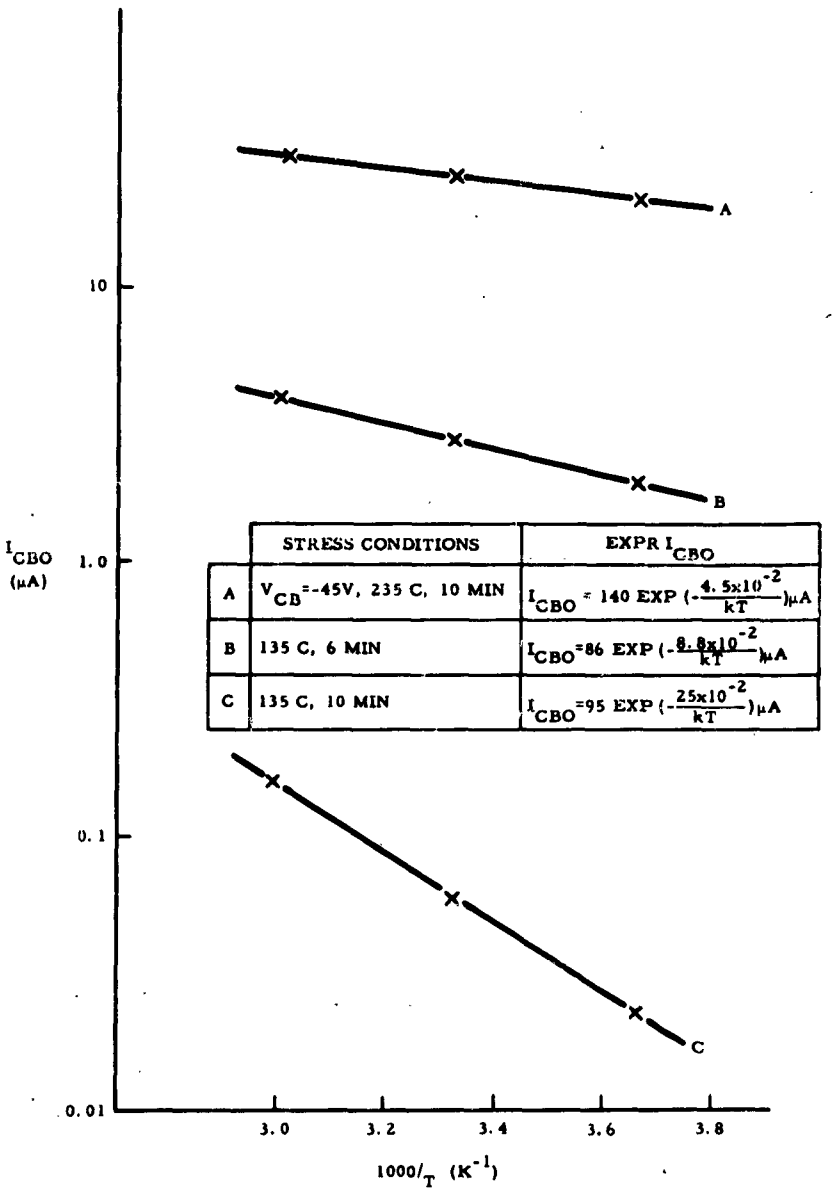


Figure 17

Thermal Relaxation Plot of Transition Voltage of a PNP Transistor



	STRESS CONDITIONS	EXPR I_{CBO}
A	V_{CB}=-45V, 235 C, 10 MIN	$I_{CBO} = 140\ EXP\ (-\frac{4.5\times10^{-2}}{kT})\mu A$
B	135 C, 6 MIN	$I_{CBO}=86\ EXP\ (-\frac{8.8\times10^{-2}}{kT})\mu A$
C	135 C, 10 MIN	$I_{CBO}=95\ EXP\ (-\frac{25\times10^{-2}}{kT})\mu A$

Figure 18

I_{CBO} (-20V) Versus Reciprocal Absolute Temperature of PNP Signal Transistor

groups to silicon atoms involved in producing the dielectric layer. At the oxidation temperatures employed (1100 to 1200 C), however, hydroxyl groups on adjacent silicon atoms would quickly react with each other to produce an oxygen bridge and water, as follows:

$$\geq Si - OH + HO - Si\leq \rightarrow \geq Si - O - Si + H_2O\leq \qquad (17)$$

This reaction can proceed only as long as neighboring hydroxyl groups are available. Isolated hydroxyl groups remaining after the reaction are securely fixed because the silicon oxide matrix is too viscous to allow them sufficiently close approached for interaction.

The residual hydroxyl groups in reality are weakly acidic, consisting of anionic radicals (silicic acid) bound to the matrix, and replaceable hydrogen ions. Other cations, such as sodium, which may be present as impurities, are capable of displacing the hydrogen from these anionic sites. Thus, the process of inversion, according to this model, involves the electrochemical transport of cations by a field induced hopping process from one fixed anionic site to the next. Accumulation of excess positive charge in the vicinity of the negatively biased silicon-silicon dioxide interface probably involves the splitting of silicon-oxygen bonds to form temporary anionic sites:

$$\geq Si - O - Si\leq + Na^+ \underset{\text{no bias}}{\overset{\text{bias}}{\rightleftharpoons}} \geq Si - O^{-}Na + {}^{+}Si\leq \qquad (18)$$

Baking under zero bias reverses the reaction and permits redistribution of the mobile cation.

The observed increase in resistance to inversion by some transistors on repeated inversion-deinversion cycling implies, on the basis of the present model, an irreversible electrochemical transport process probably occasioned by a limited source of mobile cations which are discharged at the negatively biased interface. If hydrogen ion is the mobile species this may result in the discharge of molecular hydrogen into the transistor can atmosphere. The physical details of the process must remain speculative until more delicate experiments can be devised for this investigation.

IV. CRYSTAL DEFECTS IN SILICON

It is well known that the electrical properties of silicon are extremely sensitive to purity and crystal imperfections. Transmission electron microscopy is a useful method for the study of defects which may be troublesome in silicon. Dislocations, stacking faults, small precipitates, and low-angle grain boundaries are among the primary defects observed by this means. Since the effects of such defects on silicon device operation were of primary concern, a typical power transistor was selected as a study vehicle. This device possessed the

advantage of having a relatively large silicon substrate which contributed to the ease of manipulation and specimen preparation. Although only one type of transistor was investigated, the results are considered typical of those to be expected from a power transistor, and the techniques developed are potentially useful for the study of many other types of devices.

The most important observations were considered to be those of dislocation networks associated with the final phosphorus diffusion producing the emitter region of the device. A variety of structures were observed, some previously unreported. Some of these networks may be related to microplasma phenomena. Some evidence of dopant induced inversion layers not observable by other techniques was found. The transistor investigated is depicted in Figures 19, 20, and 21. Figure 19 shows the device package with the cap removed. Figure 20 shows a top view of the pellet pattern. The base and emitter fingers are covered with aluminum and the boundary region between the fingers is covered with silicon oxide. Figure 21 is a one-finger representation of a cross section of the silicon pellet. The collector was (111) antimony-doped epitaxial silicon which was subsequently doped with boron and then phosphorus to form the base and emitter, respectively. A silicone resin protective coating lay over the entire pellet.

In order to prepare the silicon material in a device, it was necessary to remove it completely from its pedestal and remove all other materials attached to it (silicone, gold ribbon, aluminum, and silicon oxide). It was then necessary to thin the silicon to thicknesses suitable for transmission electron microscopy (1000 to 4000^{0}A). This was done by chemically polishing in a 2:15:5, HF, HNO_3, acetic acid solution.

A total of eight devices was prepared for investigation. In every device, transmission electron microscopy observations could be made in each of the silicon regions exposed at the epitaxial surface, i.e., the base region originally covered with SiO_2, the base region originally covered with aluminum, the emitter region originally covered with aluminum, and the base-emitter junction region.

A typical light micrograph showing the surface of a wafer following oxide removal is shown in Figure 22. Both the base and emitter regions which had been covered with aluminum have an etched surface. In the base region, it can be seen that the surface contains triangular etched forms similar to those created by the presence of stacking faults. The emitter regions show no well-formed patterns but have a generally etched character. The regions which had been covered with SiO_2 are smooth in character. An angle lapped pellet is shown in Figure 23. It is seen that the regions which had been covered with aluminum are depressed below the regions which had been covered with SiO_2. In this case, as well as with other specimens, both etched regions have the generally etched character without the well-defined patterns.

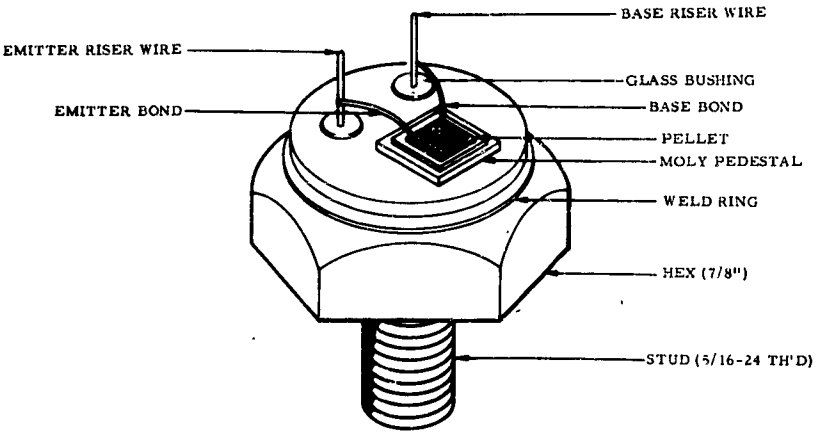


Figure 19

Power Transistor with Cap Removed

A typical electron micrograph of one of the well-defined triangular defects is shown in Figure 24. It is seen that it is a pit in the shape of a shallow inverted truncated pyramid. As a comparison, Figure 25 shows an example of a triangular stacking fault defect. It is seen that this defect does not have a depression in the center region, but etching occurs along the edge of the triangle as shown by the lower left side edge. Stacking fault contrast fringes appear along the other two edges. No contrast fringes were observed in the triangular defects found in the base regions; therefore, it is a reasonable assumption that these defects are caused by the alloying of the aluminum with the silicon. In both the

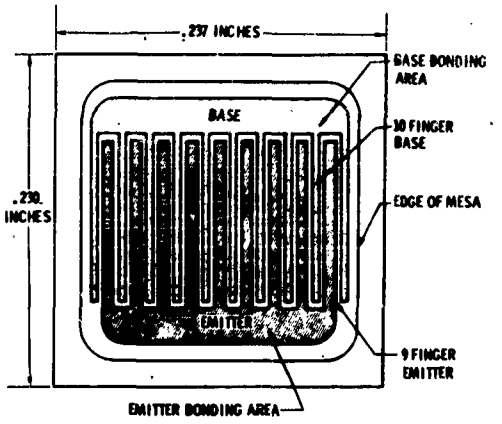


Figure 20

Top View of a Power Transistor

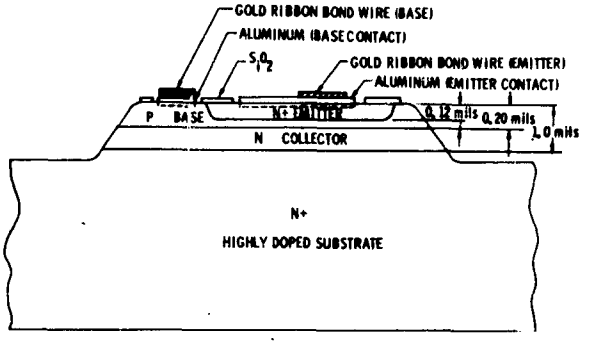


Figure 21

Cross Section of a Power Transistor

base and emitter regions having no well-defined patterns by light microscopy, transmission electron micrographs revealed that there were also truncated pyramids present, as shown by Figures 26 and 27. It is seen that the defects in the emitter region were not as well defined as those in the base region.

Dislocation networks were observed in every specimen. Two types were observed. In the five wafers where the silicon adjacent to the epitaxial surface was examined, dislocations were observed primarily in the base region within about 50 μ of the junction. The

Figure 22

Micrograph of Transistor from which Oxide and Aluminum Have Been Stripped

Figure 23

Angle-Lapped Power Transistor Pellet

Figure 24

Triangular Etch-Pit

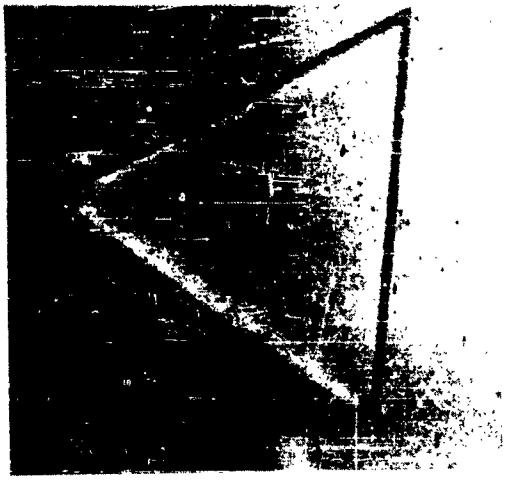


Figure 25

Etched Stacking Fault

Figure 26

Defect Etch-Pits Observed in Transistor Base Region

Figure 27

Defect Etch-Pits Observed in Transistor Emitter Region

dislocation configurations varied between those shown in Figures 28 and 29. In a few areas, no dislocations at all were observed in this region. The presence of these dislocations, particularly those lying in a dense network about 25μ from the junction, was unanticipated and considered significant for reasons described later. A small density of dislocations was observed in the emitter regions. In two devices a regular hexagonal network of dislocations was observed throughout the emitter region, ending abruptly at the base-emitter junction. These observations were made in the silicon adjacent to planes 0.03 and 0.003 to 0.006 mils below the original epitaxial surface. A typical example of this network is shown in Figure 30. Networks displaying both hexagonal and other configurations have been observed in phosphorus diffused zones by other investigators (Ref 3, 4, and 5) and are attributed to dopant concentration gradients. Our observations differ slightly from the previous work by Joshi and Wilhelm (Ref 6) who observed hexagonal networks in the material adjacent to the surface of diffusion. This implies that in the device investigated, a large concentration gradient did not exist at the surface.

Figure 28

Dislocations in Transistor Base Region Near Emitter-Base Junction

In many cases, there appeared to be a superimposed dislocation network which was not as regular as the hexagonal one. An example of this is shown in Figure 31. Both networks appear to lie along the (111) plane parallel to the surface with an occasional dislocation on the more irregular network leaving the (111) and ending abruptly at one of the surfaces. Varying light and dark contrasts were seen at some of the places where the dislocations leave the (111) plane. At the junction, the irregular network does not end abruptly but extends some 4 to 6μ

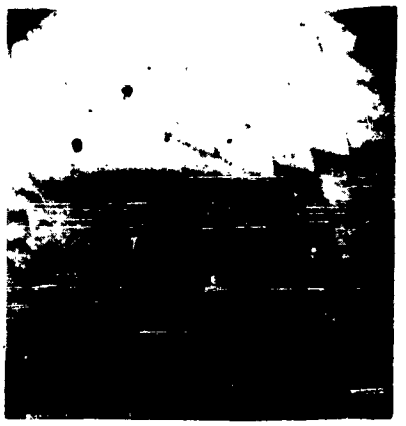


Figure 29

Dislocations in Transistor Base Region Near Emitter-Base Junction

Figure 30

Dislocation Network in Transistor Emitter Region

Figure 31

Superimposed Dislocation Networks Found in Transistor Emitter Region

in the base region. An example of this is shown in Figure 32. It should be noted that this is not nearly as far as the penetration into the base region of the dislocations observed in the silicon adjacent to the surface.

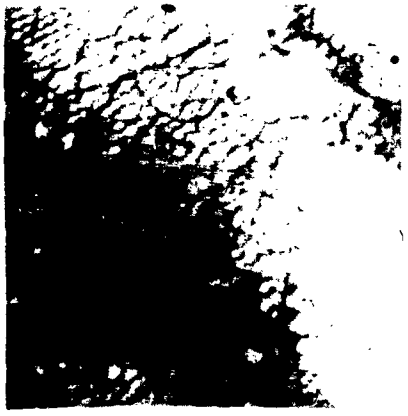


Figure 32

Extension of Dislocation Array Beyond Emitter Base Junction

The microplasma emission at the base-emitter junction of three devices was investigated. Subsequent transmission electron microscopy of two of the emitting devices showed the dislocation structures presented respectively in Figures 30 to 32, and in Figures 28 and 29. These observations are consistent with those described previously for devices which have been etched briefly or not at all. The observed absence of dislocation networks in the third specimen was attributed to more extended etching which completely removed the emitter region. Examples of low voltage microplasma emissions are illustrated in Figure 33 (a) and (b). Very few associations with surface defects were found. It is usually considered that such microplasma emissions at junctions are associated with dislocations. Although the present study does not appear to demonstrate any such correlation, these results must be regarded as preliminary and inconclusive.

The expectation of a relatively high concentration of phosphorus at the silicon-silicon oxide interface within the base region adjacent to the junction led to attempts to measure the phosphorus level by means of electron beam microprobe scans. This was done on one device with the silicon oxide still present, and on another with the oxide removed. The corresponding phosphorus x-radiation profiles are shown in Figure 34 (a) and (b). Following microprobe analysis, the latter device was angle lapped and stained with the results shown previously in Figure 23. No phosphorus was detected within the base region of either device.

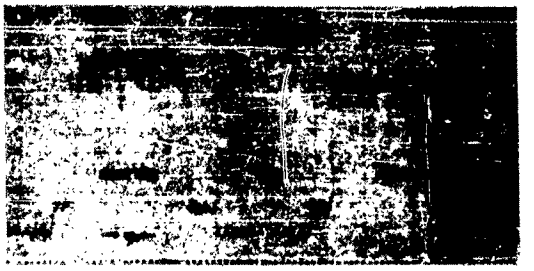
This investigation represents a departure from the bulk of previous related work in demonstrating the feasibility of working directly on completely fabricated silicon devices which evokes deductions more immediately related to device function. A significant feature of this study medium is the perpendicular termination of P-N junctions in the surface under investigation. The present observation that the poorly defined network extends slightly beyond the junction where the hexagonal network ends abruptly may be explained by the possibility that these dislocations are created by concentration gradients which are not parallel to the (111) surface. This is supported by previous observations (Ref 4) of a non-hexagonal network created by a concentration gradient perpendicular to the (001) plane. It is expected that a concentration gradient would extend at least 3 μ beyond the oxide step defining the junction. The dislocation structure, existing as far as 50 μ beyond the junction in the region adjacent to the original epitaxial surface is not as easily explained. The assumption that these dislocations are caused by a concentration gradient would imply that diffusion has occurred at the silicon-silicon oxide interface at a rate up to 16 times as fast as that in the bulk material. Diffusion is known to occur at faster rates at interfaces, grain boundaries, and dislocations than within perfect material. However, in this situation, there is a smaller solubility of phosphorus in silicon oxide than in silicon, and it is expected that this might impair the diffusion rate at the interface. This low solubility is demonstrated in Figure 34 (a) where the relative

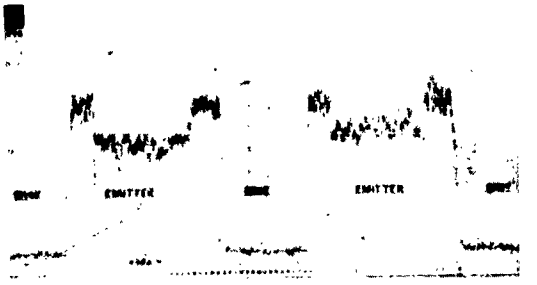
(A)

(B)

Figure 33

Microplasma Emission at Emitter-Base Junction





(B)

Figure 34

Phosphorus Scan by Electron Beam Microprobe (a) Silicon Dioxide Present (b) Silicon Dioxide Removed

concentration of phosphorus in the silicon oxide is much lower than the region not covered by oxide. The lack of detection of phosphorus within the base region by means of the electron microprobe does not preclude the possibility of its presence. An estimate is easily made that a nominal amount of phosphorus could be present in a layer 500 Å thick without being detected by the probe. If such a layer exists, it would be a dopant-created surface inversion layer. If this inversion layer could extend as far as the aluminum deposit over the base region, a completed channel would result. An example where this possibility nearly may have occurred is shown in Figure 35 where a dislocation network approaches the regions where etching of the base region by the aluminum deposit has occurred. This result indicates the utility of transmission electron microscopy in revealing possible structural features that may contribute to channeling. A potential result of such channeling might be a breakdown of the type shown in Figure 36. It can be seen that the breakdown crosses the junction at one small region. If channeling was occurring here, a short circuit could have resulted with the subsequent catastrophic failure.

Figure 35.

Approach of Dislocation Network to Defect Array
Associated with Aluminization of Base Region

It should be further noted that the electron microprobe analysis correlated with angle lapping and staining demonstrates diffusion effects which are expected. The low solubility of phosphorus in the silicon oxide previously mentioned can create what is referred to as the snowplow effect (Ref 6). As the silicon oxide layer is formed, the

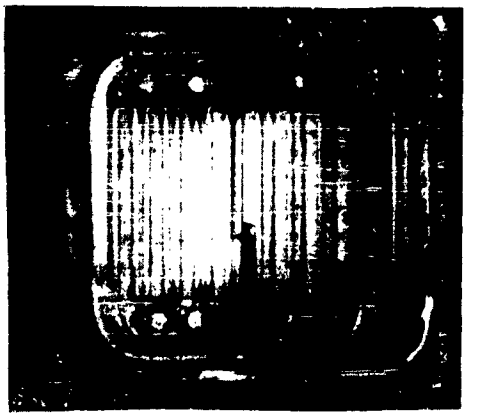


Figure 36

Thermal Breakdown Path in Power Transistor

phosphorus content in silicon at the interface is increased due to the depletion of phosphorus in the newly formed oxide. The larger amount of phosphorus detected by the probe in the silicon which had been covered by the silicon oxide is an indication of this, as shown in Figure 34 (b). The region which has been covered with aluminum also had the same phosphorus-rich layer at the surface since it had, at one time, a similar oxide layer over it which was removed before aluminum alloying.

The demonstration of this phosphorus-rich layer presents a paradox. It is expected that a higher than normal concentration gradient would exist at the surface. On the other hand, electron microscope observations demonstrated no hexagonal network at the surface in contrast to previous observations (Ref 5). At present, this situation cannot be explained.

It is well known that low voltage microplasma breakdown can be associated with the presence of dislocations (Ref 7) and that microplasma does not occur at every dislocation (Ref 8). A recent report (Ref 9) suggests that copper segregation at dislocations can be a cause of localized microplasma. In this instance, it has been shown that many more dislocations are present than the number of localized microplasmas. Definite correlation of microplasma effects with local structural defects obviously will require a continued instrumental investigation.

V. CONCLUSION

It seems evident from these investigations that a more complete understanding of processes occurring at the surfaces and interfaces of planar silicon devices will depend on a continued attack by a multi-disciplined instrumental approach and by the refinement of individual analytical techniques to smaller volume and lower concentration limits. The experimental evaluation of thermal activation energies of diffusing species in silicon oxide, although phenomenologically useful, reveals little concerning the identity of the migrating ions. Processing modifications aimed at controlling their effects must depend initially on their identification. This approach, in conjunction with other experimental investigations, appears to provide the basis of a significant step, however, in the formulation of a conceptual model describing the transport of positive ions in silicon dioxide layers. The proof of such a model will depend on additional experiments designed to reveal the structural detail of these layers.

The high sensitivity of electron microscopy to structural features makes it a potent investigative tool. A problem of interpretation is posed, however, by the frustrating profusion of detail it reveals. Since the technique cannot be employed on operating devices, the effects of structure on function must be drawn from inference and from contingent observations made with other instruments, such as the electron beam microprobe. Such correlations as that considered to exist between microplasma emission and crystal defects require patience and skill to demonstrate. The contribution of dislocation networks to channeling, and to thermal breakdown, requires external corroboration by other techniques. Yet in the latter case, at least, each experimental approach destroys the evidence required by the other, and direct corroboration is not possible. In spite of these problems, electron microscopy will continue to increase in usefulness to any program investigating the structural origins of electronic phenomena occurring at the surfaces and interfaces of electronic devices.

VI. REFERENCES

1. R. W. Lee, Physics and Chemistry of Glasses, 5, No. 2 (April 1964).

2. R. Schmidt, "Electrical Behavior of Oxidized Silicon Surfaces," Electrochemical Society Meeting, Washington, D. C., October 1964

3. R. J. Jaccodine, Appl. Phys. Letters 4, 114 (1964)

4. J. Washburn, G. Thomas, and H. J. Queisser, J. Appl. Phys., 35, 1909 (1964)

5. M. L. Joshi and F. Wilhelm, J. Electrochem. Soc., 112, 185 (1965)

6. R. M. Warner, Jr., and J. N. Fordemwalt (ed.,), Integrated Circuits, p. 305, McGraw-Hill, New York (1965)

7. A. G. Chynoweth and G. L. Pearson, J. Appl. Phys., 29, 1103 (1958)

8. A. Goetzberger, Proceedings of the International Conference on Semiconductor Physics, Prague, 1960, p. 808, Academic Press, New York (1961)

9. J. E. Lawrence, J. Electrochem. Soc., 112, 796 (1965)

DESIGN AND PROCESS CONTRIBUTION TO INHERENT FAILURE MECHANISMS OF MICROMINIATURE ELECTRONIC COMPONENTS FOR MINUTEMAN II

A. J. Borofsky and D. C. Fleming

Autonetics, A Division of North American Aviation, Inc.
Anaheim, California

ABSTRACT

High-stress testing of 200,000 microminiature components used on Minuteman II, and subsequent analyses of failures, have demonstrated deficiencies in design and processing of these microminiature components. These tests and analyses were made during the course of the Component Quality Assurance and Component Evaluation Programs conducted by Autonetics.

A summary of these results is given. It provides criteria for establishment of designs and processes for microminiature transistors, diodes, capacitors, resistors and integrated circuits.

Solutions for demonstrated deficiencies are identified and catagorized in terms of design improvement, process and tool changes, inspection changes, and screens.

The role of component procurement documents and component qualification procedures in the assurance of proper design and processing of high reliability microminiature components is discussed. The effectiveness of these procedures in elimination of design and process inherent failure mechanisms is shown.

INTRODUCTION

At the start of the Minuteman II Program, it was recognized that the required technological advances in electronic components would be achieved through successive cycles of identification and correction of deficiencies in component design and processing.[1] In support of this objective, high-stress testing and subsequent analyses were conducted on 200,000 Minuteman II components. Testing was conducted on signal and power transistors, computer and power diodes, ceramic and both solid and liquid tantalum capacitors, metal film and wirewound resistors, and integrated circuits as part of the Component Quality Assurance Programs and Parts Monitoring Programs. Chemical, metallurgical and physical analysis analyses on the subcomponent level were conducted by Autonetics do provide a more thorough identification of the causes and mechanisms underlying failure. These Programs are a part of contracts between the Autonetics Division of North American Aviation, Inc. and the Ballistics System Division of the United States Air Force. The Component Quality Assurance Programs[2] are a part of Contract AF04(694)-247; AF04(694)-402 and AF04(694)-626.

The postmortem analysis and laboratory analyses and studies of devices which "failed" during these tests have demonstrated deficiencies in design and processing of these components. A summary of the results of these tests and supplier action resulting therefrom has been made. This report identifies significant areas of component deficiencies and the contribution of design materials and processes to them. In addition, the means for minimization of the effects of the contributing factors which have been proven by subsequent evaluations are shown. It must not be inferred that the modes and mechanisms discussed are universally present in all devices analyzed, nor that the corrections shown are necessarily the most effective or valuable for all designs or all suppliers.

The results reported herein, however, provide significant insight as to the problems which were inherent in the designs, materials and processing of the Minuteman II microminiature devices, and their resolution. As such, they provide a criteria for determining the problems which might be anticipated with other microminiature devices through analysis of the design, materials, and processes proposed.

The term "mechanism and failure mode" has been used herein as much as possible to describe the different factors with reference to failure. For instance, "mechanism" provides an understanding for the means by which a failure occurs and sometimes is employed to refer to a very basic understanding in a molecular or atomic sense. "Failure mode" is employed to describe the observed manner in which the failure occurs.

Discussion of the nature of the mechanisms noted herein are outside the scope of this paper. Characterisations are reported to the extent required to achieve effective corrective actions. Changes in design, process, screens and inspections are purposely non-specific in order to protect our suppliers proprietary information. In depth, discussions for several mechanisms will be presented by Autonetics' scientists who conducted "Physics of Failure" investigations (during the final session of this conference). Additional references to the identified mechanisms are cited in this report.

FINDINGS

The failure modes and mechanisms found in Minuteman II devices were based upon experience. Generically, they arise from problems inherent in surfaces, contamination, material compatibility, and the required microgeometry.

Resistors (Tables 1 through 3). Device degradation, which is accelerated by ability of moisture to penetrate polymeric packages, was a common major problem.[3, 4] Design, material, and process improvements have enabled suppliers to provide devices which are consistently capable of withstanding ten-day exposure to moisture resistance test. All suppliers have made at least one change in material. All have learned the necessity for control of bulk material, storage, pot life, and viscosity. Considerable effort was expended to optimise cure time. Additional package improvements resulted from the use of stirring paddles, lead degreasing, and general process cleanliness.

The separation of the lead from the resistance element on talon construction was a common problem. Highly reliable construction has required a change in design in which the resistance element is welded or soldered to an end cap which is part of the lead wire assembly. The ability to weld consistently good joints is difficult and suppliers have learned the necessity for control of temperature, pressure, time and tools.

The resistor manufacturer buys a relatively low output from resistor wire manufacturers. As a result, the ability to obtain improved wire is limited. The demands for a controlled product require that the resistor manufacturer select the wire for high reliability production. Factors used in selection include temperature coefficient and aging characteristics.

Major improvements in deposited film resistors involved improved procedures for handling, spiralling, and contamination reduction. These improvements lowered the incidence of failures resulting from non-uniform film thickness and by film electrolysis.[5]

Table 1
Failure Modes/Mechanisms and Corrections for Film Resistors

Failure Modes/Mechanisms	Major Contributing Factors	Means to Reduce Contributing Factors			
		Design	Process	Screen	Inspection
Electrolysis of metal film	Porous conformal coating and contamination, poor adhesion	Add inorganic coating over the film	Establish control of leak testing maximum shelf life controls		Particle size controlled
Moisture and ionic permeability		Use of high density epoxy with optimum hardener	Analytical balance for mixing proportions S. S. paddle for mixing Engineering control and laboratory techniques for mixing and application Degrease leads and end cap Preheat parts prior to application Continuing rotation during application and mix Base paint-coating application in clean rooms Mechanized Handling		Pressure cooker or humidity test on each batch
Non-continuous/non-uniform resistance film	Non-uniform evaporation Contaminated substrate surface		Uniform rotation during evaporation Clean and smooth substrate surface	Projection type examination	

Table 1 (Cont)

Failure Modes/Mechanisms	Major Contributing Factors	Means to Reduce Contributing Factors			
		Design	Process	Screen	Inspection
Non-continuous/non-uniform resistance film (Cont)	Cracked substrate Damaged during spiral	Rounded corners for cap assembly		15 x short time overload before coating	
Relative thermal expansion of incompatible materials causing intermittent opens	Talon construction Excessive end cap - Substrate tolerances	Press fit end cap Rounded substrate shoulder Tighten tolerance Compatible materials		Pull test	

Table 2
Failure Modes/Mechanisms and Corrections for Precision Wirewound Resistors

Failure Mode/Mechanisms	Major Contributing Factors	Means to Reduce Contributing Factor			
		Design	Process	Screen	Inspection
Metallurgical changes in resistance wire characteristics with time	Inadequate aging Uneven tension		Optimization stress relief temperature		Select based on temperature coefficient
Metallurgical composition and cold working of wire			Uniform layer winding tension control		
Corrosion	Voids in insulation	Use of less permeable coating		100% insulation pinhole test	
Moisture and ionic permeability of polymeric material Incomplete polymerization	Porosity of polymeric coatings		Bulk blending and cold storage		Viscosity control
Intermittent or open connections Mechanical or thermal stress	Compatibility of materials	Change to soft nickel leads	Use slack loop in resistance wire		Sample pull test

Table 3
Failure Modes/Mechanisms and Corrections for Power Wirewound Resistors

Failure Mods/Mecisnisms	Major Contributing Factors	Means to Reduce Contributing Factors			
		Design	Process	Screen	Inspection
Metallurgical changes in resistance wire with time and temperature	Inadequate aging		Added stress relief cycle		Select low temperature coefficient wire
	Thermal conduction of conformal coating	Added thermal conducting insulated filler	Tension controls on winding machine		
	Thermal conduction through cap and leads	Change to SS cap and copper-weld leads			
Moisture and ionic permeability of polymeric material Incomplete polymerization	Porosity of package	Double coat	Optimized cure time Controlled application time to insure optimum coating		
Metallic inclusions from processing metals	Static charge on element		Static eliminator Cleaning cap terminal and core	100% infra-red scan	
	Weld splash		Control weld splash at end cap Harder weld point material		Visual
	Particles from abrading process		Eliminate abrasion calibration for low ohmic values		
Mechanical and thermal stress causing intermittent or open connections	Inadequate time-pressure cycle in welding		Added pressure sensing anode to welding		
	Misfit of cap to core		Change end cap material to SS or 80 Ni-20CY Radius instead of chamfer core		Sample pull test

Ceramic Capacitors (Table 4). The majority of ceramic capacitor failures early in the program were caused by shorts and low insulation resistance. These failures were in part, due to thinness and porosity of the ceramic sheets used as the dielectric. These types of failures were virtually eliminated by improvements in the homogeneity of the ceramic powder, processing improvements, electrode design, and contamination controls.

Shorts and high dissipation type failures were also related to the methods used to connect the terminals to the electrodes. The introduction of part preheating prior to joining, contributed significantly to the reduction of these failures.

Solid Tantalum Capacitors (Table 5). Miniaturization of solid tantalum capacitors resulted in an increase in leakage current and a higher occurrence of shorts. It became apparent that considerable more development was required in the areas of forming and anodizing the tantalum anode.[6, 7] These development programs resulted in improvements in tantalum purity, methods of forming the anode, deposition and pyrolysis processes, and coating methods to achieve good electrical bonds to the case.

Geometry of the anode itself was a major contributor to the relatively high incidences of shorts observed. Not only did the length and diameter require change, but the configuration of the corners also was altered.

Wet Porous Anode Capacitors (Table 6). The basic degradation of these devices delivered early in the Minuteman II Program was leakage of the electrolyte.[8] The loss of electrolytic results in increases in leakage current and dissipation factor and a decrease in capacitance. Further problems result from the corrosive nature of the electrolyte.

Although no true hermetric seal has yet been developed for the small size devices used, improvements in the following have greatly reduced the possibility of electrolyte losses.

1. Optimum plug material
2. Use of sealants and O-rings over the plug
3. Optimum crimp design and crimp tooling
4. High-temperature screens for leakage

Table 4
Failure Modes/Mechanisms and Corrections for Ceramic Capacitors

Failure Mode/Mechanisms	Major Contributing Factors	Methods to Minimize Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Difference of thermal expansion between lead and ceramic causing dissipation factor increase	Inadequate silver coating on ceramic Silver migration into solder Silver surface contamination		Increase coating thickness Refire silvered parts prior to use Add heat sink solder jigs Mechanized handling Use lower temperature solder		
Dielectric Breakdown	Foreign material introduced during mixing		Mix using SS Paddles Improve sealer material		Use one mill load per lot
Metalic chemical reaction causing Dielectric breakdown	Non-homogenous firing; thin sections		Profile kiln temperature Thicker Sheets Improve sheet casting equipment		Tighten criteria No. holes per square in ceramic sheet Continuous pyrometer recording on kiln
	Compatibility dielectric/electrode material	Thinner electrodes and better ceramic	More precise silk screening	2 direction X-ray	
Moisture penetration of ceramic Porousity of ceramic	Thinness and holes in ceramic sheet	Added ceramic cover plate to each face Use ceramic rather than plastic coating Added two overcoats	Improve cleaning methods after lead attachment Deflash case only after complete cure		

Table 5
Failure Mode/Mechanisms and Corrections for Solid Tantalum Capacitors

Failure Mode/Mechanisms	Major Contributing Factors	Methods to Minimise Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Current leakage due to localised surface impurities and excessive burrs Contaminates in Ta material	Impurities in Ta powder		Improve purity by washing		Specify minimum impurities to Ta suppliers
Breakdown of Ta_2O_5 Dielectric Thin areas of anodize Contamination/ Metallic particles	Insufficient buildup of film Impurity in Ta powder	Rounded Anode edges	Increase number of formation cycles Improve purity by washing		Specify minimum material impurities to Ta suppliers
Silver depletion between case and silver carbon coated MnO_2 cathode Diffusion of silver into solder	Breakdown of Ta_2O_5 Dielectric Inadequate carbon coat on MnO_2	Thicker coating	Wash to reduce MnO_2 surface contamination	100% visual for contamination	
Cracking of tantalum-nickel weld joint	Welding temperature-pressure contamination of leads		Optimised cycle *Degreasing step* added Add heat shunt		
Anode shorting to case Misalignment	Insertion Tooling	Reduce Anode Length	New centering tooling	2 direction X-ray	
Leakage caused by moisture absorption in and on insulation	Non-hermetic	New glass Seal design	Flux removal process added Optimization of sealing temperature	Added seal Test hot part immersed in cold H_2O, then tested electrically	Visual

Failure Mode/Mechanisms and Corrections for Wet Porous Anode Tantalum Capacitors

Failure Mode/ Mechanisms	Major Contributing Factors	Methods to Minimize Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Electrolyte loss because of inadequate seal	Permeability of seal	Double crimp tolerances on plug Add elastomer coating on plug Add O ring to larger case sizes between plug and case	Control electrolyte fill operations Mechanize crimp seal operation Thicker case Decontamination of internal plating at seal end		Leakage at high temperature
Corrosion of cathode lead Nickel lead weld penetration through silver case	Method of welding and thickness of case	Thicker case	From resistance to arc welding		
Anode short to case Design and geometry	Anode too long Anode not centered	Shorten anode Add centering porous glass insulator	New anode centering tool	2 direction X-ray	
Failure of weak tantalum riser-lead weld because of heat concentration	Welding method Short riser	Lengthen riser	Use ultrasonic welding		
Shorting by metallic particles in electrolyte Non-adherence of platinum	Flaking of platinum coating on case. Metalic particles from processes		Improve case surface preparation use electro instead of chemical deposition		
Note: Improvements similar to those reported for solid tantalum capacitor were also made in the processes and used to fabricate the anode slug.					

Reduction in failures arising from corrosion of cathode leads, uncentered and elongated anodes, and coating flaking have been achieved. In addition, the improvements in the formation of the anode, similar to those noted for solid tantalum capacitors were required to reduce the failures caused by anode geometry and defective dielectric coatings.

Diodes (Table 7). Surface problems[9-15] arising from inadequate packaging and careless handling and inadequate attention to cleanliness were shown by the high incidence of inversion and channeling failures. Major improvements were required to overcome the multitude of latent problems inherent in the use of polymeric packaging.[3, 4]

Marked reduction in failures have been achieved through the following actions:

1. Improved lead wire material, plating and cleaning
2. Multicoating and composite coating of active elements
3. Selection of polymeric material more suited to the application; better specification, storage, blending, application and curing.
4. Improved process control, inspection and the use of dye penetrant test.

Further stability of surfaces was achieved by changes in the passivating coatings and the elimination of high pressure leak detection methods using ionic solutions.

The degradation of forward voltage characteristics has been related to deficiencies in joining operations. Control of the gold plating on leads and tight controls of die geometry were required to eliminate degradation caused by delamination of plating and cracked dice. Major improvements in alignment and scribing were necessary to eliminate shorts and thermal runaway.

Small Signal Transistors (Table 8). Major areas of failure were shown by failure mechanisms associated with inversion/channeling[9-1] and joining. As with diodes, minimization of surface problems require attention to both surface sensitive processing and packages.

Procedures such as hot gas flushing, ultrasonic washing, carrier cleaning, and dry box processing were found to be essential needs at several stages of processing. However, the continued discovery of surface degradation phenomena required the implementation of 100% reverse bias or operating stresses to assure a product population relatively free from this defect. Major design changes e.g., low

Table 7
Failure Modes/Mechanisms and Corrections for Diodes

Failure Mode/Mechanisms	Major Contributing Factors	Methods to Minimize Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Inversion/channeling Voids in passivating oxide (pinholes)	Handling induced scratches and chips Dust - Dirt Compound cracks		Clean area processing (micro-air hood)		Mask, water and pellet inspections
High temperature sealing		Lower temperature sealing glass		100% reverse leakage screen at voltage near breakdown	
Incomplete surface passivation	Non-uniform application of passivation material	Passivating compound change to provide better adherence to die	Improved curing		100% microscopic examination prior to molding
Polarization of passivation under T and V	Poor adherence of passivation to die Contamination entrapment	Pyrolytic glass passivation			
Die off-center in molding	Hydroscopic nature of passivating polymer which penetrates to exterior	Two coat intermediate polymeric coating Interior coating changed to provide less permeability to moisture	Improved mold design and use		

Table 7 (Cont)

Failure Mode/Mechanisms	Major Contributing Factors	Methods to Minimise Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Inversion/channeling (Cont) Cracks at lead/polymer interface	Excessive handling (deflashing operation) No chemical bond between lead and polymer	Change polymer compound	Sandblast for flash removal	100% Burn-In	
External contamination	Hydrazine flux prior to lead tinning		Neutral flux		
Leak detection testing (bombing)	Ionic bombing solution		Zyglo leak detecting (non-ionic)		
Moisture bridge lead to lead (between polymeric layers)	Shrinking of internal polymeric materials	Application of bonding agent	Pre-shrinking internal polymeric coating		
Forward voltage degradation Voids are delamination of gold plating from Kovar lead	Non uniform lead plating Contamination from Au diffusion		Lead plating control Sandblast both sides of die Controlled scrubbing prior to bond formation		Bond shear strength test Bonder inspecting
Cracked die (Au diffusion into die)	Bonding on die of varying diameter and thickness (due to excess etching)	Tightening of resistivity and diffusion targets	Improved masking Reduced ratio of anode to cathode areas		100% inspection of lead fillets, tilted die, peripheral voids

Table 7 (Cont)

Failure Mode/Mechanisms	Major Contributing Factors	Methods to Minimize Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Forward voltage degradation (Cont)					
Cracked die (Au diffusion into die) (Cont)	Irregularities of Ag contact	Silvered grown silicon contact			
Short-break through or bridge oxide	Misalignment		Improved mold design. Use mechanical loading		
Thermal runaway caused by crack through junction	Scribing through oxide layer Chips carried with pellet into final device		Scribing without penetrating oxide layer		Pellet inspection
Open caused by processing	Broken-internal lead		Improved mold design and use		

Table 8
Failure Modes/Mechanisms and Corrections for Small Signal Transistors

Failure Mode/Mechanism	Major Contributing Factors	Means for Reducing Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Inversion/channeling (Carrier recombination velocity) Contamination in, on and under oxide layer	Moisture, gaseous and ionic contamination resulting from uncontrolled atmospheres, handling and inadequate cleaning (e.g., inadequate cleaning prior to oxide growth	Low resistivity annular ring Field relief electrodes	Oven in dry box line Inert gas in dry box Dry box sealing Hot gas flushing or bake (a) before dicing (b) after die mount (c) after lead attach. (d) package and lid	100% back bias at elevated temperature	
Reaction between coating and surface	Use of varnish for surface protection Contamination of oven atmosphere by vacuum pump Readsorption of contaminants	Protective glass coating	Wash prior to oxide growth (ultrasonic) Improved handling Improved carrier cleaning Purer water, ultrasonic washing		
Moisture bombing	Retention of small quantities in package		Elimination of high pressure water bombing		
Non-hermetic packages	Cracked ceramic loose pins		Alignment of lid and header	100% gross leak test (hot oil) under magnification	100% microscopic inspection of headers
	Contamination or excess plating at weld area			100% fine leak detection	
	Blisters on metallization				Tightened control on plating thickness

Table 8 (Cont)

Table 8 (Cont)

Failure Mode/Mechanism	Major Contributing Factors	Means for Reducing Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Non-hermetic packages (Cont)	Holes in glass (internal)	Redesign to permit glass flow around internal lead	Mount braze after glazing operation	Emitter floating potential	
	Thermal mismatch of materials	Redesign			
	Mishandling, strenuous cleaning	Redesign			
	High temperature brazing	Low temperature brazing (Pd-Cu-Ag)			
	Cut-of-control welding			Stitch welding base plate to ring frame	Improved welding control
External leakage because of ionizable paint	Low temperature paint	High temperature paint			
Current concentration centers					
High current densities at defect site	Wafer edge cracks		High temperature gas phase etching	100% inspection of die after separation	
Stacking faults			Improved epitaxial process	Improved epitaxial process	
"Cold" thermocompression bond					
Uncontrolled process	Inadequate heat at bond site Misplaced bonds	Reduction in wire size	Hot capillary bonding	100% centrifuge	Sample bond shear test

Table 8 (Cont)

Failure Mode/Mechanism	Major Contributing Factors	Means for Reducing Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Intermetallic formation	Uncontrolled bond size, weakened wire		Constant tension feed		
	Chisel bonding		Heater block alignment Process parameter adjustment		
	Chisel bonding Al-Al		Ball bonding Al-Au		
Contact delamination			Micro alloying of contacts		
Internal shorting Lead wire sagging	Uncontrolled wire tension	"Upbonding"	Controlled wire tension		
Deformed wire	Short wire span				
Internal opens Lead wire broken	Uncontrolled tension (metal fatigue)		Controlled wire tension		
Thin wire	Deformed wire	Thicker wire	For 2 step bonding chip and post		
	Deformed post surface				100% inspection
Intermetallic formation	Contamination on post		Ultrasonic cleaning		

Table 8 (Cont)

Failure Mode/Mechanism	Major Contributing Factors	Means for Reducing Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
High collector resistance/ thermal runaway Cracked die, wafer tilting, and voids	Excess pressure during bonding Voids under die	Back metallisation Changes in composition and geometry of preform	Hot capillary bonding	100% thermal impedance screen	
Delamination at intermetallic interfaces	Use of nickel interface Inactivated plating				
Microplasm formation	Improper masking and resolution, dust		Improved photolithographic process, laminar air flow hoods		

resistivity guard rings, field relief electrodes, and protective glass coatings, have been introduced to minimize this problem. Ceramic packages used for the small signal devices must be carefully fabricated and handled.[16] Despite improvement in design, it has been determined that tight controls, microscopic inspections, and careful hermetricity testing is required to achieve the required reliability. Strict control of alignment and of welding for package closure are required.

Prevention of the degradation of bonds to the die has required many changes.[17] Foremost has been the need to adequately preheat all parts prior to joining. Hot capillary ball bonding has proven effective. The formation of brittle aluminum-gold bonds has not been found to be a problem with these devices. However, centrifuge screening and sample bond shear tests are required to remove those devices in which alloy formation into the base silicon, necessary for adhesion, did not occur.

Internal wire connections in these small devices posed a dual problem. If the wire was too loose or deformed, shorting to the die occurred. If the wire was too tight, the resulting tension fatigued the metal and opened the connection. A major package redesign which permitted "upbonding"[a] and the control of wire-feed tension alleviated these problems.

High collector resistance originating from voids in the back contact[18] and delamination of plating occur. The substitution of metalization of the back surface of the die for preforms and the elimination of nickel plating on the mounting stage were effective corrective actions. However, a 100% thermal impedance screen is considered necessary to meet high reliability requirements.

Power Transistors (Table 9). Inversion and channeling[9-15] problems were widespread on power transistors. Minimization of these problems, as with small signal devices, necessitated careful attention to handling and cleaning procedures which affect the surface of the device. Major changes in the design, materials of construction, and tooling for power transistor packages have been made.

Secondary breakdown[19] at low energy levels in the early devices has necessitated optimization of the resistivity and fluckness of the epitaxial layer. A non-destructive secondary breakdown screen is required to assure that devices can tolerate some degree of transient current while in the circuit.

---

[a] Bonding the terminal wire to a connection higher than the plane of the silicon.

Table 9
Failure Modes/Mechanisms and Corrections for Power Transistors

Failure Mode/Mechanism	Major Contributing Factors	Means to Minimize Effects of Major Contributing Factors			
		Design	Process	Screen	Inspection
Inversion/channeling Variation in protective oxide	Inadequate control of oxide process	Thicker protective oxide			
	Contamination from mineral acid removal of photo-resist		Non-mineral acid removal of photo resist		Tighter control for removal of protective oxide
Contamination on surface	Irregular etching		Ultrasonic etch		
	Dirty carriers		Rapid removal of wax employed during lapping High temperature rinse water Cascade rinsing High temperature bake, carrier cleaning, substitution of DI water for alcohol		
	Contamination from diffusion atm.		Ni gettering		
	Process delay		Change and control of diffusion atm.		
	Dust		Improved dust control		
	Protective coating		Elimination of protective coatings		
	Silver migration	Substitution of gold for silver leads			

Table 9 (Cont)

Failure Mode/Mechanism	Major Contributing Factors	Means to Minimize Effects of Major Contributing Factors			
		Design	Process	Screen	Inspection
Inversion/channeling (Cont)					
Improper main seal weld	Lack of parallelism				Tighter acceptance on parallelism
	Excessive plating	Reduce plating thickness Increase weld area			Tighter control on plating thickness
Cracked glass in glass-to-metal seal	Marginal design	Metal-ceramic seal	Tooling redesign		Tightened inspection of glass meniscus
Cracked metal	Lead deformation during crimping		Tooling redesign		
Shallow diffusion	Surface damage. Diffusion damage. Photomasking lfmitations oor interconnect geometry	Deeper diffusion and geometry changes	Two step aluminum evaporation		
Limitations of KMER and acetate masking	Pinholes Poor definition Non-uniform coatings Poor acid resistance		KFTR Glass masking Hg vapor light source Chemical removal of KFTR		
Triple diffused structure	Rear interconnect geometry	Double diffused epitaxial planar	Two step Al.		

Failure Mode/Mechanism	Major Contributing Factors	Means to Minimise Effects of Major Contributing Factors			
		Design	Process	Screen	Inspection
Secondary breakdown/ thermal runaway					
Current concentration centers	Non-optimum epitaxial layer resistivity and thickness; thin base width	Optimise epitaxial layer resistivity and epitaxial and base thicknesses		100% secondary breakdown screen	Tightened control of epitaxial layer thickness
	Thin interconnect metallisation	Optimise metallisation thickness			
Improper emitter diffusion (pipes)	Incomplete removal of photo resist (oxide islands)		Increased development time Increased time and temperature for photoresist removal	100% base - Collector scope test after emitter diffusion	Reduced time for KPR coated wafer storage
Diffusion irregularities	Acid vapor contamination of KPR prior to development		Improved dust control		
Emitter to base shorts	Excessive penetration of metallization		Double processing for oxide growth and removal; double emitter diffusion Reduction in metallization temperature		
Change in thermal resistance (thermal mismatch) Voids between die and mounting pedestal	Poor wetting (Au-alloy braze) Sn-Ag solder	Molybdenum layer between die and package. Au-Si eutectic bonding	Ultrasonic bonding	100% measurement of high current parameters (die stage)	

Table 9 (Cont)

Failure Mode/Mechanism	Major Contributing Factors	Means to Minimize Effects of Major Contributing Factors			
		Design	Process	Screen	Inspection
Change in thermal resistance (thermal mismatch) (Cont)					
Deterioration of braze by ordering	Brittle Ni-Si couples	Elimination of nickel			
Cracked die	Excessive bonding pressure	Add grid so scribbing done on raw silicon rather than SiO_2		100% x-ray of header before assembly	Monitor small signal gain parameters
Residual strain in die	Excessive die probe pressure		Reduced bonding pressure		
	Au-Ni metallization	Aluminum metallization			Improved plating solution control
	Ag leads	Au leads			V_{sat} measurement after joining
High series resistance					
Insufficient deformation of crimp (to outer lead)	Low crimping pressure	Redesigned header			
Micro cracks, uneven breaks, and fissures	Scribing in low atomic density plane		First scrib along cleavage plane		

The identification of emitter pipes,[19] the junction irregularities,[20,21,22] and emitter-box shorts have precipitated marked changes in masking, photo resist, and diffusion. Double processing for oxide growth, oxide removal, and emitter diffusion have been found necessary.

Changes in thermal resistance[23,24] resulted from degradation of bonds between the die and case and die cracks. These problems had their origins in the metallurgical systems selected for joining die to case. Major redesigns have incorporated metallurgical systems and processes which provide non-brittle, relatively void-free interfaces.

Integrated Circuits (Table 10). A severe problem was inherent in the gold to aluminum bonds used in integrated circuits.[25] Removal of this problem required development of composite metallurgical systems for bonding gold to gold.

Shorts between the interconnections and the silicon[26] have occurred because of oxide defects. Double masking and double oxide growth and electrical testing of dielectric strength are helping to control this problem.

The following problems encountered with diodes and transistors were encountered with integrated circuits and resolved by actions similar to those used for semiconductor devices.

1. Inversion/channeling due to lack of process cleanliness and inadequate packaging[9-15]
2. Improper control of internal lead strength
3. Die cracking
4. Handling damage
5. Photomasking related problems
6. Attachment of die to package

It is to be noted that findings similar to these have been reported by Partridge, Hanley, and Hall.[27]

DISCUSSION

The recurrence (on this new generation of microminiature devices) of many problems which had been identified and resolved on earlier devices is not necessarily surprising since well known factors contribute to the situation such as:

1. Inadequate lead time (as defined by contractual end item delivery requirements) for thorough development of new parts.

Table 10
Failure Modes/Mechanisms and Connections for Integrated Circuits

Failure Mode/Mechanism	Major Contributing Factors	Methods to Minimise Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Inversion/channeling Contamination in, on and under oxide layer	Impurities Non-homogenity Adhesion of alum Open at oxide step Purity of oxides Pinhole in oxide	Al. deposition between input diodes - acts as depletion barrier	TEOS layer PbO layer Silane layer Oxide thickness control Al. evap. time and temp. Double KMER Double/multiple masking/ separate masks Improve mask tolerance Optimise stabilisation bake Mechanized diffusion New die carrier Automatic gas cycling		Inspect for oxide tears 400 X inspection Tighten criteria on Al. width Pre-alum. clean inspection
Inadequate die attachment caused by incompatible materials and surfaces	Poor processes and material selection		More uniform gold plating, control Au salts in solution, post Au deposition cleaning Thicken gold plate 100 to 200 micro inches Remove KMER from wafer back prior to bake Reduce die mount. temperature	In process temperature cycle	

Table 10 (Cont)

Failure Mode/Mechanism	Major Contributing Factors	Methods to Minimize Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Contamination (surface)			Stab. bake temperature increased Increase D.I. wash time Automated handling Vacuum pick up Dry scribe water Automatic gas cycling		
Die bond intermetallic formations Material compatibility	Uncontrolled bond size - temperature material purity	Au-Au systems Larger bonding areas Increase lead diameters Rerouted interconnections and pad locations New package gives more room to bond	Lower bonding temperature Lower stab. bake temperatures One bond for contact, second bond for tail pull Automatic tail-pull Control die-bonding temperature Hot capillary bonding New bonding machines	100% centrifuge In process temperature cycle	Visual

Table 10 (Cont)

Failure Mode/Mechanism	Major Contributing Factors	Methods to Minimize Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Silicon surface epi spikes leading to cracking Surface treatment, handling	Crystalization methods Handling methods	Small die size when possible	Improved controls and cleaning procedure at wafer lap to reduce spikes & impurities Tweezers with blocking mechanism Reduce die mount temp. Sheet resistivity limits tightened Vacuum pencils Handling mechanized Improve die boat		
Air and moisture penetration cause leakage Material compatibility	Design and processes	Thicker walls and lids Increase Au thickness Increase oxide thickness of Kovar leads Lid overlap	Move uniform gold plating New welding tips Vacuum bake prior to seal Pre-oxide clean Kovar parts H_2O content less 35 PPM during lidding Preform cleaning (solvent)		Tighter specs on piece parts

Table 10 (Cont)

Failure Mode/Mechanism	Major Contributing Factors	Methods to Minimize Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Air and moisture penetration cause leakage (Cont) Material compatibility (Cont)			Tension on lidding boat springs Dry box cleaning cycle Control Au salts in plating Post Au cleaning Lidding furnace profile optimization Auto Pak handling Provide stitch weld base to header Introduce Mech-Pak Insulated lids		Check profile daily Welder clean up Visual inspection Material inspection
Oxide defects and contamination on surface causing shorts and leakage	Inadequate control of oxide process		Move pump down time to assure H_2O outgassing Added vacuum bake-completed wafer Improved pre PbO cleaning Improved pre lid cleaning Improved Post P_bO cleaning Pre-oxide cleaning of Kovar parts		Visual for pinholes Visual for cleanliness

Table 10 (Cont)

Failure Mode/Mechanism	Major Contributing Factors	Methods to Minimize Effect of Major Contributing Factors			
		Design	Process	Screen	Inspection
Oxide defects and contamination on surface causing shorts and leakage (Cont)			Solvent cleaning preforms Dry box cleaning cycle Improve pre-alum. cleaning using D. I. rinse New die carrier Mechanised diffusion Optimization stabilization bake		

2. Continuous cost pressure on component manufacturers which encourages shortcuts and gambling on newer designs, materials, and processes.

3. Attempts to upgrade commercial parts to high reliability status without analyses and correction of problems inherent in design, process, and the incompatibility of material combinations used.

4. Inadequate utilization of the knowledge of the contribution of design, process, and materials to device failures.

5. Inadequate understanding of the contribution of design, processes, and material to device failure.

The observations and the findings reported herein dictate the regimen to be followed if one is to obtain reliable devices for aerospace systems. This consists of:

1. Delineation of design, materials, and processes of proposed new parts.

2. Selection of suppliers with the most reliable designs and processes using the most compatible materials. The selected supplier must understand and control these items.

3. Knowledgeable estimation of failure modes inherent in design, processes, and materials of selected suppliers.

4. Inclusion of specification of screens to remove high percentage of parts exhibiting problems estimated in (3) above.

5. Continuous analysis and characterization of devices to ascertain weaknesses of design and processes and incompatibility of materials and to determine factors which contribute to weakness and incompatibility.

6. Technical interfacing with suppliers to upgrade design, processes, and materials to minimize effects of contributing factors.

7. Continuous study of failure mechanisms to provide background for improvement in the state-of-the-art.

These activities must be implemented through component procurement specifications and qualification procedures, as have been done on Minuteman II.

The Component Quality Assurance and Component Evaluation Programs have achieved their objective: accelerated improvement of the reliability of a new generation of devices. The reliability objectives for Minuteman II will be met. Their timely attainment could not have been achieved without these programs. The task, however, is not complete. New technology and economic considerations will force continuing change in the design, processes, and material of construction. Continuing evaluation,[b] characterization on the component level, Physics of Failure studies, and upgrading of specifications remain an essential feature of our component reliability effort.[28]

[b] Special note is made of the very excellent paper "Physics of Failure Accelerated Testing" by G. E. Best, G. R. Bretts, and H. M. Lambert (Electro Technology, October, 1965) which cites many excellent references on this subject. Rather than repeat the thorough literature survey performed by the authors, the reader is referred to their paper.

REFERENCES

1. A. J. Borofsky, "Electronic Components for Minuteman II," 3rd Annual National Electronic Packaging Conference (NEPCON) Long Beach, California, 10 June 1965, Publication X5-1068/33 Autonetics, Anaheim, California

2. A. J. Borofsky, "Component Quality Assurance Programs for Microminiature Components for Minuteman II" in Physics of Failure in Electronics, Volume 3, Rome Air Development Center, Research and Technology Division, Air Force Systems Command, Griffis Air Force Base, N. Y., 1965

3. J. J. Licari, S. M. Lee and A. G. Valles, "Properties of Plastic Materials and How They Relate to Device Failure Mechanisms," presented at the Fourth Physics of Failure in Electronics Conference, Chicago, Illinois, November 16-18, 1965

4. H. A. Perry et al, "The Corrosivity of Plastics," Society of Plastic Engrs, 19th Annual Conference, 1957

5. M. E. Goldberg et al, "Study of Comprehensive Failure Theory," IIT Research Institute Final Report AF30(602)3054-5519 - 551902, RADC TDR-64-309 Sept 1964

6. L. F. Howard and A. W. H. Smith, "Dielectric Breakdown on Solid Electrolyte Tantalum Capacitor," IEEE Transactions on Component Parts, June 1964, Vol CP-11, No. 2 p 187-193

7. E. J. Fresia and J. M. Eckfeldt, "Failure Modes and Mechanisms in Solid Tantalum Capacitors," Physics of Failure in Electronics, Vol 2, RADC Series in Reliability, 1964, pp 483-497

8. J. Burnham, "A New Approach to the Attainment of the Highest Reliability in Tantalum Capacitors," 1964 WESCON Conference, Los Angeles, California, 1964

9. D. Nickolson, "Silicon Surface Leakage," Physics of Failure in Electronics, Edited by M. E. Goldberg and J. J. Vacarro, Volume 3, p 81

10. P. H. Eisenberg, G. V. Brandewie and R. A. Meyer, "Investigations of Surface Failure Mechanisms in Semiconductor Devices by Envelope Ambient Studies," presented at the Fourth Physics of Failure in Electronics Conference, Chicago, Illinois, November 16-18, 1965

11. J. E. Meinhard, R. L. Nolder, R. E. Harris and J. E. Forrester, "Imperfections and Impurities in Silicon Associated with Device Surface Failure Mechanisms," presented at the Fourth Physics of Failure in Electronics Conference, Chicago, Illinois, November 16-18, 1965

12. E. D. Metz, "Silicon Transistor Failure Mechanisms Caused by Surface Charge Separation," Physics of Failures in Electronics, Vol 2, Edited by M. F. Goldberg and J. Vaccaro, p 1963

13. J. Hall, S. S. Baird and R. P. Williams, "Results of High Temperature and Impressed Voltage on Transistor Devices" presented at the Electrochemical Society Meeting, May 6-10, 1962

14. G. L. Schnable, E. S. Schlegel and R. S. Keen, "Failure Mechanisms in Reverse - Biased Oxide Passivated Silicon Diodes," Physics of Failure in Electronics, Volume 3, p 108

15. J. Olmstead et al, "Hydrogen-Induced Surface Space Change Regions in Oxide Protected Silicon," IEEE Transactions of Electron Devices, Vol ED-12 3/65 pp 104-107

16. D. C. Griffin, "Hermetric Packaging for Semiconductor Devices," Semiconductor Products and Solid-State Technology Vol 8, No. 5 May 1965, pp 42-45

17. A. J. Avila, "Metal Bonding in Semiconductor Manufacturing and Survey," "Semiconductor Products and Solid-State Technology," Special Issue, November 1964, Vol 7, Number 11, pp 22-26

18. J. D. Guttenplan and F. H. Stuckenberg, "Failure Mechanisms Associated with Die-to-Header Interfaces of Planar Transistors," presented at the Fourth Physics of Failure in Electronics Conference, Chicago, Illinois, November 16-18, 1965

19. Schafft and J. C. French, "Second Breakdown in Transistors" IRE Transactions of Electron Devices, March 1962, p 129

20. H. J. Queisser, "Dislocations and Semiconductor Device Failure," Physics of Failure in Electronics, Spartan Books, Baltimore, Maryland, 1963, Clearer-Hume Press, London, pp 146-155

21. J. J. Wortmann and R. A. Evans, "Young's Modulus, Shear Modulus, and Poisson's Ratio in Silicon and Germanium," Journal of Applied Physics, Vol 36, No. 1, January 1965, pp 153-156

22. R. C. Miller and A. Savage, "Diffusion of Aluminum in Single Crystal Silicon," Journal of Applied Physics, Vol 27, December 1956, pp 1430-1432

23. C. G. Jennings, "Failure Mechanisms Associated with Thermally Induced Mechanical Stress in Minuteman Devices," presented at the Fourth Physics of Failure in Electronics Conference, Chicago, Illinois, November 16-18, 1965

24. Aeronautical Systems Division, Div/Materials and Processes Physics Lab, Wright Patterson AFB

"Theory of the Thermal Conductivity of Metals, Alloys and Semiconductors," Report Nr. ASD-TOR-62-74
Final Report August 1962

25. G. V. Browning, D. G. Cummings and L. E. Colteryahn, "Failure Mechanisms Associated with Thermocompression Bonds," presented at the Fourth Physics of Failures in Electronics Conference, Chicago, Illinois, November 16-18, 1965

26. J. J. Wortman and R. M. Burger, "A Fundamental Failure Mechanism in Thin Film Metal Dielectric Structures Observable as a Generated Voltage," Physics of Failures in Electronics, Volume 2, Edited by M. F. Goldberg and J. Vaccaro, p 173

27. J. Partridge, L. D. Hanley and E. C. Hall, "Progress Report on Attainable Reliability of Integrated Circuits for Systems Applications," Report No. E-1679, MIT Instrumentation Laboratory, Cambridge, Mass., November 1964

28. WS-133B Components Final Report, October 1964 to October 1965 Autonetics, In preparation.

SECTION VII
PAPERS NOT PRESENTED
AT SYMPOSIUM

SELECTIVE CHROMATE CONVERSION OF INTEGRATED CIRCUIT INTERCONNECTING ALUMINIZATION

D. A. Abdo

The Boeing Company
Aero-Space Division
Seattle, Washington

ABSTRACT

The extremely small size of semiconductor integrated circuits introduces microanalytical problems not normally encountered. E. A. Corl [1] at the 1964 Physics of Failure Symposium described an anion reaction for failure analysis of microcircuits. A similar non-destructive chemical process, chromate conversion of the interconnecting aluminization, has been studied. The process will react selectively producing striking contrasts in the color of the aluminization. A number of factors both in the solution and the integrated circuit itself were found to influence the reaction rate. By the proper application and technique, chromate conversion can be used to locate sites of surface defects and attendant electrical failure. Variations and contrasts in the characteristic brown color will show whether an area is P or N, the relative doping level, areas of high Al_2O_3 concentration, the presence of pinholes in the oxide under the aluminum, breaks in the aluminum, and contaminants on the surface. The advantages of the process are simplicity, economy, sharp variations in color, and it is non-destructive. Compositions of the various solutions and the pertinent reaction parameters are described. A study of a number of factors within the integrated circuit affecting the reaction was also made. A proposed reaction mechanism is discussed. The process appears to have great utility as an analytical tool. It may also have an application as a process step for the protection of aluminum against oxidation and mechanical damage.

---

1. E. A. Corl, "Anion Reaction for Failure Analysis of Microcircuit Components," Physics of Failure in Electronics, RADC Series in Reliability, Vol. 3, pp 342 - 353, 1965

INTRODUCTION

The chromate conversion of aluminum has been utilized as a commercial process for protection against oxidation for over 30 years. About a dozen proprietary compounds meeting the standards of MIL-C-5541 are used in sizeable quantities for bulk processing of the various aluminum alloys.

The present interest in this process is the application to integrated circuit interconnecting aluminization. By use of selective reaction with the metallization, it has been possible to show areas of surface defects and opens in the interconnections.

CHROMATE CONVERSION PROCESS OF ALUMINUM

Since most chromate conversion processes are proprietary, detailed reactive mechanisms are not widely known. However, it is generally believed[2] that hexavalent chromium (such as Cr_2O_7 = and CrO_4 =) is reduced to the trivalent form (probably in the well known form of Cr_2O_3) which precipitates as a complex chromium gel on the metal surface when the pH at the metal solution interface has been increased suitably by the local consumption of acid. X-ray studies have shown that the films are amorphous. However, it has been found that they normally contain both hexavalent and trivalent chromium compounds. The color and other characteristics of the precipitated coating are governed by the pH of the solution, the nature of the activator radicals and the ratio of their concentrations to that of the chromate compound or compounds used. While the formation of a conversion coating requires some attack on the metal surface, the thickness of metal removed normally is small, varying from about 0.0001 to 0.00001 inches (or about 25,000 to 2500 angstroms). Electrical resistivity of chromate films is relatively low. Ranges are broad since resistance varies with film thickness; however, an average figure is 900 - 2500 microhms for a square inch of projected area.

THE ANALYTICAL TECHNIQUE

Since the aluminization of most integrated circuits is nominally 8,000 angstroms, normal proprietary compounds usually result in excessive attack and dissolution of the metal. Therefore, two approaches were attempted in order to find solutions which would selectively react without excessive attack. These were the modification of a proprietary solution and the formulation of laboratory mixtures.

To insure more uniform results, a pretreatment cleaning process was found necessary. The usual procedure is to first immerse the entire device in a non-etch alkaline

2. C. W. Ostrander, "Chromate Conversion Coatings," Electroplating Engineering Handbook, Reinhold Publishing Co., A. K. Graham, Editor, pp. 416 - 435, 1962

cleaner or "degreaser". Several of these are available commercially. The solution should be diluted down about 2 to 4 fold, from its normal concentration. Immersion time is about 15 minutes at 70 - 80°C and at a pH of 10.5. The sample is then rinsed in distilled water for about 15 seconds and immersed in an aluminum deoxidizer solution at 65 - 70°C for 10 to 20 seconds. This solution contains 27 mils of 85% H_3PO_4 and 20 grams of CrO_3 per liter of distilled water and has a pH of about 0.8. The sample is once again rinsed with distilled water, dried rapidly and is ready for chromating. It is convenient to have the device mounted on a microscope slide during the above steps because the chromating itself should be observed under a metallurgical microscope having an approximate magnification range from 50 to 800X.

A proprietary compound covered by MIL-C-5541 was modified by using it at one-half its normal concentration and adding from 1 to 3 grams of a chromate salt such as $K_2Cr_2O_7$ or $Na_2Cr_2O_7$ per 100 mls. of solution to give a pH of approximately 1.5. The solution may be used at 25°C and 1 drop usually covers the entire device.

Typical reaction times are 15 seconds to 4 minutes depending upon the individual circuit and the depth of color desired. In most cases this solution would color aluminum connected to N areas a dark brown, and P areas a tan color.

This same solution may be diluted 10 fold and result in an N reaction only. In this case it was found that only the aluminum covering N-doped areas would color tan or brown.

The same procedure is used in both cases. The reaction may be interrupted at any time by rinsing with distilled water and drying rapidly. In some cases better results were obtained by thus stopping the reaction and repeating the chromating in 15 - 45 second treatments about 4 or 5 times.

Three solutions were formulated that would yield a P reaction, a N reaction or a N and P reaction.

The P-type solution has the composition:

10 grams CrO_3
15 grams $K_2Cr_2O_7$
0.15 grams NaF

Two grams of this mixture is dissolved in 100 mls. of distilled water.

The N-type solution has the following composition:

10 grams CrO_3
5 grams $BaCl_2$
0.01 grams NaF

One gram of this mixture is dissolved in 100 mls. of distilled water.

The N and P type solution has the following composition:

10 grams CrO_3
15 grams $Na_2Cr_2O_7$
5 grams $K_2Cr_2O_7$
1 gram $BaCl_2$
0.15 grams NaF

Three grams of this mixture is dissolved in 100 mls. of distilled water. This has given a predominantly P-solution, that is, aluminum connected to P-doped areas will color darkest. By diluting this solution 5 to 10 fold a predominantly N-solution is obtained. All solutions are used at 25°C and a nominal pH of 1.5.

These solutions were found to work quite well on integrated circuits from one manufacturer. However, the results were not entirely consistent when they were attempted on devices from other vendors.

THE PROCESS AS A FAILURE ANALYSIS TOOL

Since probing for a failure site can often result in mutilating the aluminization, the chromating process shows considerable promise because defects in or under the aluminization will usually result in a deviation from the expected coloring. The series of photomicrographs which appear below are examples of how failure sites and surface defects may be detected by employing the chromating process. Since this is a color reaction having varying shades from tan to red-brown, some contrast is lost in black and white reproduction. In most cases, however, the contrast is sufficient to illustrate the point at hand.

Figure 1 is an unreacted integrated circuit. The bright, silver-appearing geometry is the aluminum which interconnects the various components.

Figure 2 shows an N reaction. In this case the emitter and collector windows and the N portion of the diodes colored brown. Close observation shows what appears to be a pipe in the emitter window where no reaction took place. Figure 3 shows this defect at higher magnification.

Figure 4 is a P reaction with aluminum connected to a base diffusion of a transistor. The break in the aluminum prevented the reaction from proceeding any further.

Figure 5 is an example of an N and P reaction. Here N doped areas resulted in a darker brown color, such as the common emitter aluminization and the N+ collector windows. Aluminum connected to P areas colored tan, an indication of

the relative doping levels. Note the absence of the gold bond resulted in no reaction at all, an indication that dissimilar metals accelerate the reaction.

Figure 1

Integrated Circuit Before Chromating

One can show the presence of relatively high aluminum oxide concentration by eliminating the aluminum deoxidizer cleaning step. Since the chromating will proceed faster where the oxide is thinnest, areas of thickest oxide formation will not color, as shown in Figure 6.

The process has been found useful in locating pinholes in the silicon oxide passivation under the metallization. In Figure 7 the deviation from the normal reaction resulting in the light appearing "bulls-eye" is the site of the pinhole.

Suspected contaminants between the aluminum and the silicon oxide can be demonstrated also. In Figure 8 the unreacted light areas are an indication of surface contamination.

Figure 9 shows a number of deviations from the normally expected N reaction. In some cases aluminum connected

N to P to P colored darkest. The N doped collectors did not react at all. There are several shadings of tan to brown on P aluminum which were not expected to occur. In this case the failure site which was previously predicted as an open by electrical measurement was manifested by a dissolution of a portion of the aluminum at the collector oxide step.

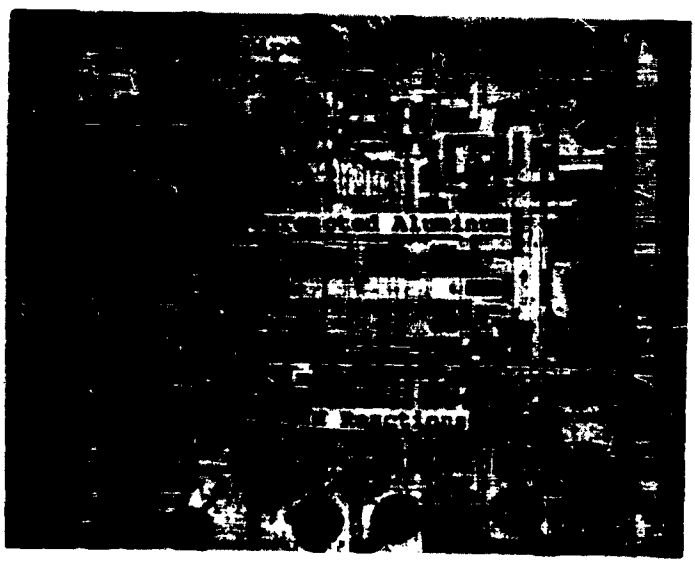


Figure 2

Reaction of Aluminum Over N-Doped Areas

Figure 10 is a near uniform chromate conversion of all metallization with the exclusion of some of the aluminum covering the windows. This solution had a very low concentration of metallic cations which may account for the lack of coloring over the windows.

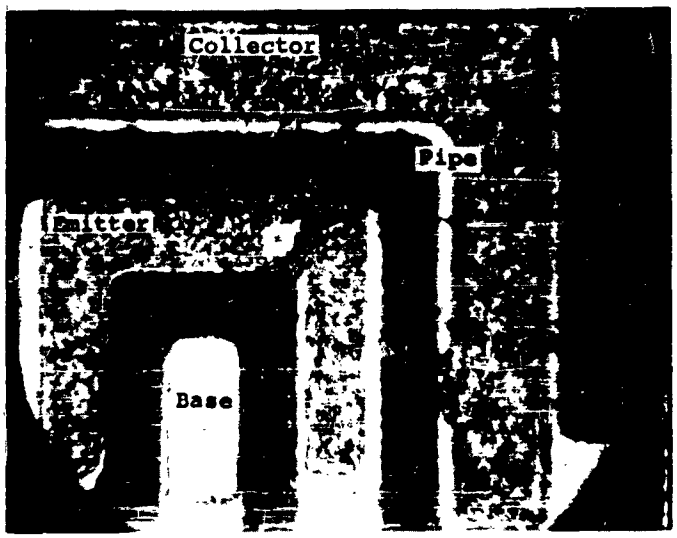


Figure 3

Lack of Expected N-Reaction In Emitter Window Indicates Pipe Down To P-Base Region

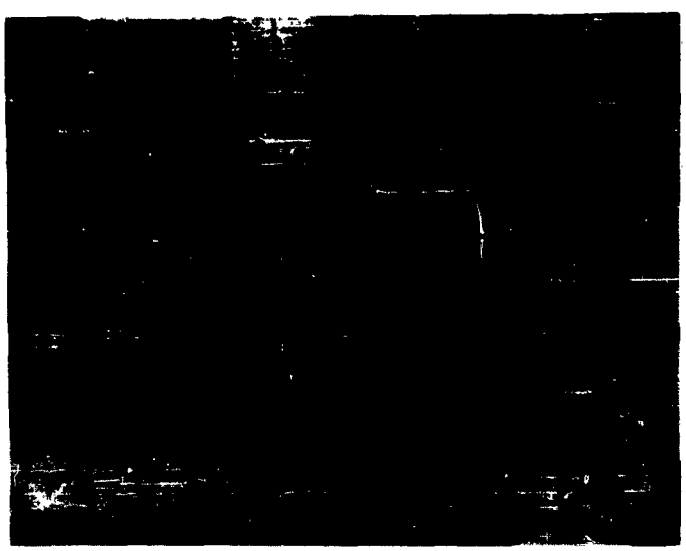


Figure 4

Breaks in the Aluminum are Shown by an Interruption in the Reaction

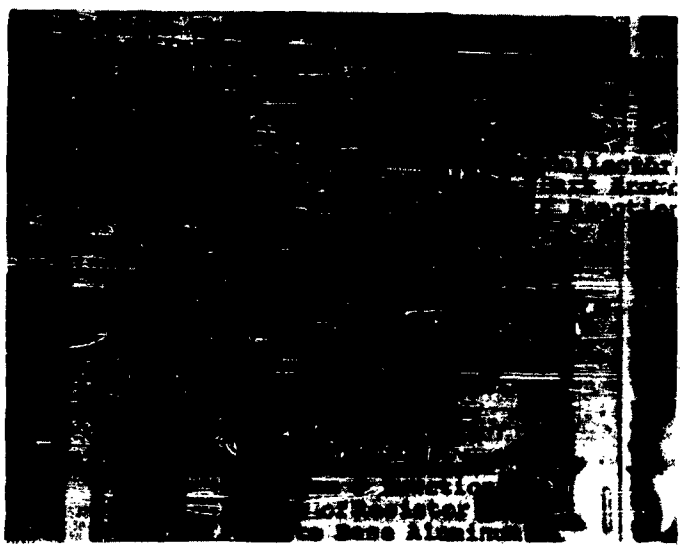


Figure 5

N and P Reaction
N reaction predominated resulting in a darker color.

Figure 6

Sites of high aluminum oxide concentration result in no reaction. Spotty dark areas are predominantly P reactions.

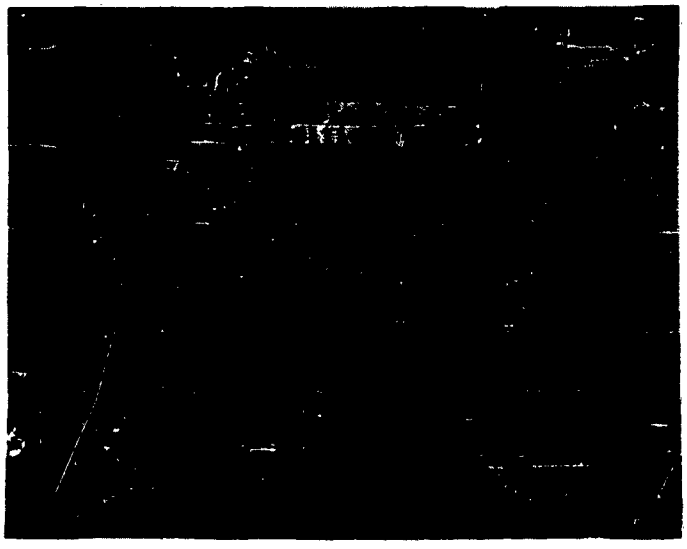


Figure 7

Deviation from Normal Reaction Locates Site of Pinhole in Oxide

Figure 8

Unreacted Light Area Indicates Surface Contamination Between Silicon Oxide and Aluminum

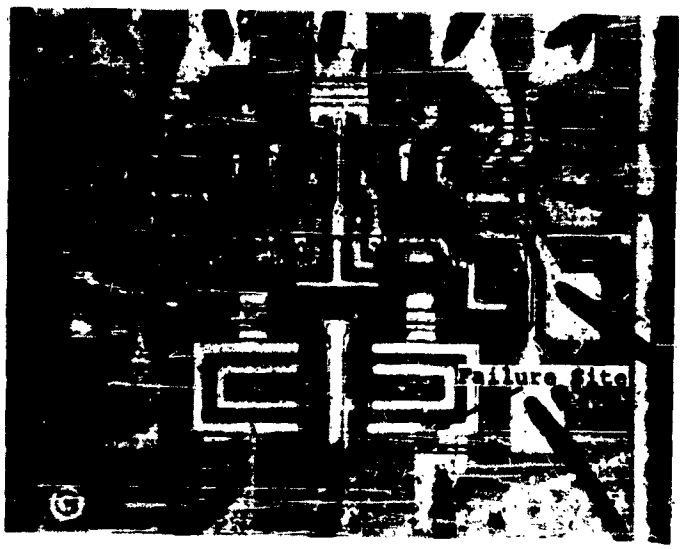


Figure 9

Deviation from Normal N Reaction and Failure Site at Collector Oxide Step

Figure 10

Near Uniform Chromate Reaction on All Aluminization

PROPOSED REACTION MECHANISMS

Since there were deviations in the expected reactions from manufacturer to manufacturer a clear cut theory for the reactions is not possible without further experimental data. Integrated circuit processing parameters have a definite influence on how the chromate solutions will react. Detailed processing data is considered proprietary, so a certain amount of speculation is necessary.

The method of doping and the later processing steps are considered to be those most likely to influence the reaction. Silicon oxide passivation can be grown in several different manners. The impurity content within the oxide will vary. Cleaning techniques and purity of solvents will have a definite bearing on the quality of the oxide. The manner of deposition of the aluminization, and the process of alloying and sintering appear to have strong influences. Also, the presence or absence of dissimilar metals such as a gold bond was shown earlier to affect the reaction (Figure 5).

Disregarding variables in the chromate solutions for the moment, two principal factors within the device are considered to influence the direction of the reaction. The first of these is the difference in potential of the doped areas connected to the various segments of the interconnecting aluminum as well as the potential difference between gold and aluminum. Since this is a reaction of an electrolyte, the larger the potential difference between connected sites the more rapid the reaction and the darker the color.

The other important factor is the impurity level within the aluminum. To insure good ohmic contacts the aluminum is alloyed into the highly doped emitter windows and what is usually the N+ doped collector windows. This will result in dopants diffusing into the aluminum. The sintering process of the aluminum to the silicon oxide might possibly result in some transfer of impurities. From a number of reactions observed it appears that a higher impurity content within the aluminum results in a faster reaction and a darker coloring. The impurity content appears to have a catalytic effect.

Which of these effects will predominate largely depends on the individual manufacturing technique. To some extent variations in manufacture can be compensated by modifying the chromating solutions.

As mentioned earlier, there are a number of variables in the chromating solutions. Concentrations, ratios, solubilities, activators, and pH are among the more important factors. However, by working with solutions known to react in a predictable manner on an integrated circuit from one manufacturer it may be possible to modify the more important variables in the solutions to resolve a reaction common to all manufacturers.

POSSIBILITIES AS A PROCESS METHOD

Since it appears possible to uniformly chromate all of the aluminization, this may offer an application as a process tool. Chromate conversion coatings are used primarily to inhibit oxidation of aluminum. The aluminization of integrated circuits is so thin that oxidation might result in a hot spot and an eventual open. A chromate conversion coating could reduce the likelihood of such an occurrence. There is also the possibility of offering some protection against mechanical abrasion during fabrication.

SUMMARY AND CONCLUSIONS

There are several advantages to the chromate conversion process. The selective color reaction is helpful in locating failure sites and surface defects without the need for time-consuming and sometimes mutilating test probing. It is very economical since only a drop or two is required for a device. Although a thin layer of aluminum is dissolved during the conversion process, the chromate coating itself is a good conductor, so the process is essentially non-destructive. This technique or a similar color reaction technique will become more in demand with increased miniaturization and the corresponding increased difficulty of locating failures.

The shortcoming of the technique is that a selective reaction for integrated circuits from one manufacturer may not work in the same selective manner for circuits from another manufacturer. In spite of this shortcoming the reaction usually reveals a good deal of useful information. It is hoped that further experimentation will resolve the variations due to manufacturing processes.

HOT SPOT MESOPLASMA FORMATION IN SILICON PLANAR TRANSISTORS

E. B. Hakim

United States Army Electronics Command
Fort Monmouth, New Jersey

ABSTRACT

A model is presented to explain the cause of mesoplasmas when secondary breakdown occurs in silicon planar transistors. An analysis of both aluminum and gold metalized devices is presented. To substantiate the model, data are presented from experiments to analyze the second breakdown sites. Three areas of investigation are suggested which may prove effective in reducing the destructive effects of hot spots and mesoplasmas in transistors.

I. INTRODUCTION

The problem of hot spots and secondary breakdown in transistors, a catastrophic failure mode, has plagued the electronics industry for years.[1,2] As a transistor reaches some critical value of current and voltage, it will go into a negative resistance region, to a much lower voltage at higher current and then destroy itself exhibiting either a degraded diode characteristic or a collector-emitter short circuit.

Alloy devices which have been destroyed in this manner can be dissected and areas found where the alloy has penetrated across the junction causing a short circuit.[3] Recently the author and two colleagues have found that hot spots in certain types of alloy transistors result in material being forced out of these hot spot sites in the form of balls, rather than completely alloying into the device.[4] This paper deals with the hot spot phenomenon in silicon planar devices which are the most commonly used devices today and discusses the formation of mesoplasmas, their effect on devices, and means to retard their development.

II. DISCUSSION

Hot Spots and Mesoplasmas

Most silicon planar transistors today use aluminum metalization on the silicon to form contact areas. A device which has been destroyed as a result of second breakdown usually exhibits severe damage to the aluminum and the entire surface of the device. Figure 1 shows four typical examples. This type of destruction is a direct result of the absence of a current limiting device in the circuit. If, however, the hot spot formation and second breakdown are controlled, interesting results can be observed.

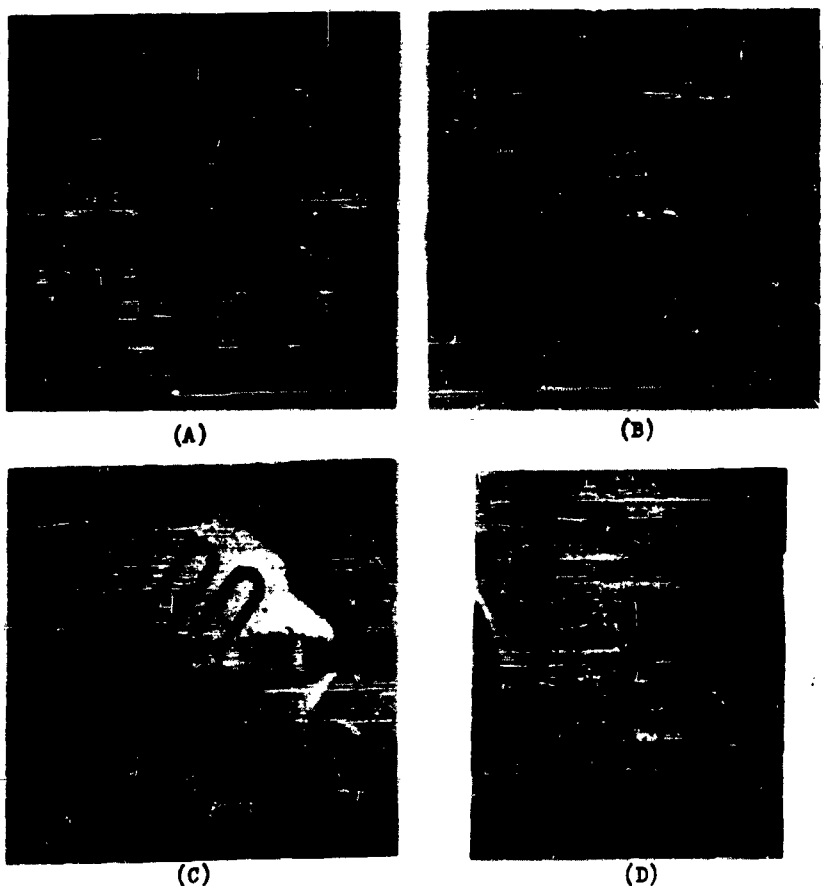


FIG. 1
Secondary Breakdown Damage to Surface of Transistor. (A) & (C) Show Recrystallized Base & Emitter Areas. (B) & (D) Show Damaged Areas at the Surface and Melted Leads.

The term "mesoplasma" was first used by A. C. English and H. M. Power.[5] Mesoplasma was the name given to the light emission which is a strong red spot much larger in dimension than a microplasma, and is common in a reversed bias diode in avalanche. Mesoplasmas were later reported in silicon transistors and related to second breakdown in transistors.[6] A typical microplasma emission is shown in Figure 2, while Figure 3 shows the same device in second breakdown, with a mesoplasma emission. Figure 3 shows the mesoplasma emission in the metalized area. It has also been seen in the junction. An anomaly has been observed where the emission is in the form of an arc rather than a spot. Figure 4A is an example of this arc-over phenomenon and Figure 4B is the V-I trace. The negative resistance region shown in Figure 4B is only observed when the arc-over occurs. It is believed that this oddity was caused by sputtering of aluminum when the hot spot was initially formed. The sputtered aluminum produced a low resistance path across the oxide which was over the junction. At some critical voltage, the arc was produced causing the voltage drop observed on the curve tracer.

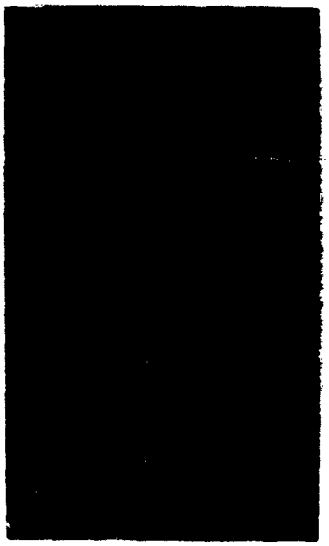
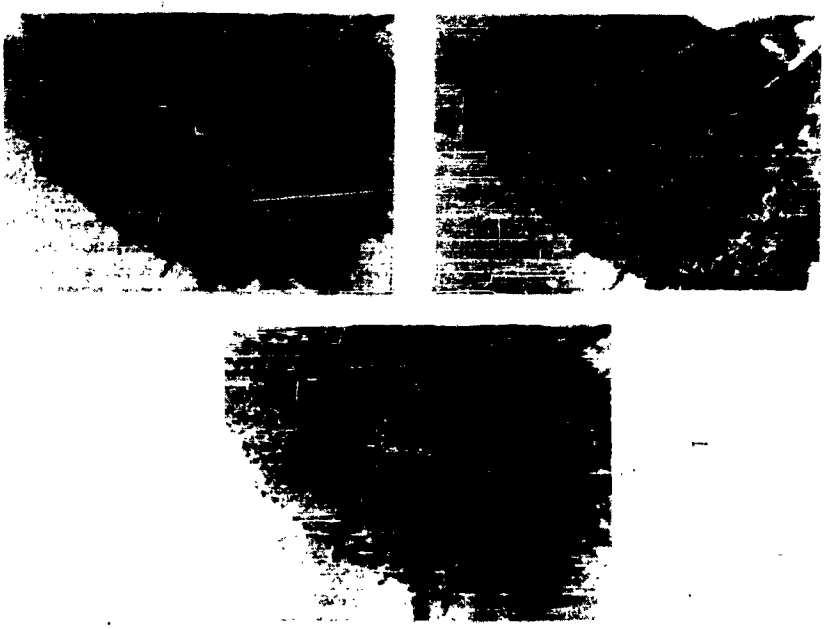


Fig. 2

A Typical Microplasma Emission

Fig. 3

Secondary Breakdown
Mesoplasma Emission



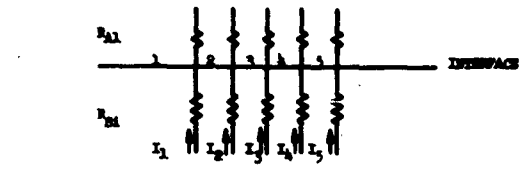
(A)

(B)

Fig. 4
(A) Three Photographs of the Arc-Over in the Same Device. (B) The V-I Trace Illustrating the Lower Voltage Second Breakdown Curve.

Mesoplasma Formation

A possible equivalent circuit for the metalization on the surface of a silicon transistor, and the silicon immediately beneath it, might be as shown in Figure 5. As shown, it is assumed that there is an infinite number of resistors consisting of aluminum and bulk silicon. If, as a result of current crowding or some other effect, R_{Si3} is less than the other resistors, then the current distribution will not be uniform across the junction. As the current flow through the device is increased more current will flow through R_{Si3} until a temperature is reached which will melt the aluminum above this hot spot. As the aluminum melts, it will diffuse into the silicon, decreasing R_{Si3} even further. If the current is suddenly increased, the surge in current will heat the area enough to cause the aluminum to alloy completely across the silicon causing a short circuit.



where; R_{Al} = aluminum metalization resistance

R_{Si} = bulk silicon resistance

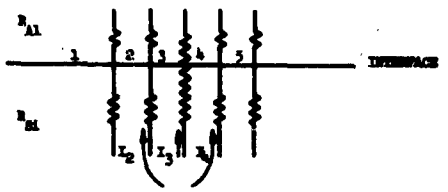
$I_1 = I_2 = I_3 = I_4 = I_5$

$R_{Al} \ll R_{Si}$

Fig. 5

Equivalent Circuit for Metalization and Silicon Surface.

If, on the other hand, the current is increased gradually, the resistor R_{Si3} will go through a minimum and then begin increasing due to heating effects; that is, at the same time the aluminum diffuses into the silicon and the aluminum and/or the silicon is oxidizing. Because the resistance at the surface is increasing due to the oxide, current flow will increase through R_{Si4}. As the resistance R_3 increases further, more current will pass through R_{Si2} and R_{Si4}. This is shown in Figure 6. As before, the melting temperature of aluminum will be reached, causing diffusion of aluminum and decreasing R_{Si2} and R_{Si4} further. At some point the resistance of the surface above R_{Si2} and R_{Si4} will increase, and the process begins again. This effect is shown in Figure 7. The previous remarks have been limited to aluminum metalization. Investigations have also been carried out using gold as the interconnect. The analysis used for the current flow is the same as before; however, the process by which the surface resistance changes is different. Instead of alloying and oxidizing simultaneously as in the aluminum case, the gold melts and vaporizes or possibly aggregation occurs, leaving behind high resistance silicon.



where: $R_3 > R_2$ or R_4

$I_3 < I_2$ or I_4

Fig. 6

Circuit For the Metalization and Silicon Surface After a Hot Spot Formation.

(A) (B)

(C)

Fig. 7

The High Resistance Region Growth Caused by Relocation of the Hot Spot

Initial experiments utilizing infrared techniques tend to add further verification to the theory of the changing current distribution in the second breakdown mode as presented above. If the surface of a transistor is observed with an infrared scanner, it can be determined that a hot spot (the area of highest current density) will vary with the change in resistence at the surface. That is, as the damaged area grows, the hot spot detected by the scanner will follow the outer edge of this area. Preliminary measurements have also indicated that the temperature of the surface is indeed near the melting temperature of the metalization. For a device with aluminum interconnects (melting point of aluminum is 660°C) the surface has been found to be approximately 600° - 700°C. With gold as the interconnect (melting point of 1063°C) the surface temperature was determined to be in excess of 800°C. It is further believed that the actual hot spot temperature is the melting temperature of silicon which is 1420°C. However, no direct measurements have been made to verify this statement. This same conclusion was reached in the article by A. C. English and H. M. Power.[5]

III. CONCLUSIONS AND RECOMMENDATIONS

From the facts and conjectures so far presented, it is possible to draw some conclusions and hypothesize further.

1. A hot spot can be formed in bulk silicon which can reach 1420°C.

2. This temperature is reflected in the rise of surface temperature above the hot spot area, causing melting, diffusion, vaporizing, and oxidation at the surface.

3. The location of the hot spot can move due to a change in the resistivity at the surface.

4. The surface is composed of aluminum-silicon oxide for the aluminum metalized device and silicon-silicon oxide for the gold device.

5. The hot spot will continue to relocate itself until complete alloying exists and a short circuit results.

6. Arc-over between electrodes can result causing the characteristic negative resistance region of secondary breakdown.

There are three approaches which are being investigated at the present time to diminish the effect once a hot spot has been generated. Although the eutectic temperature of gold-silicon is lower than aluminum silicon, gold appears to be more resistant to alloying and parameter degradation than aluminum. If this is true, then higher melting materials should even be better. Molybdenum, which has a melting temperature of 2620°C, is being investigated as a contact metal. Other possibilities are chromium, nickel, platinum, titanium, and rhodium. All of these metals have melting temperatures above that of silicon. If the technological problems associated with the application of these metals are overcome, then it may be possible to go into second breakdown with its negative resistence characteristic, but with no permanent transistor degradation, because little or no alloying has taken place.

A series of tests were performed to determine the composition of the material in the areas where mesoplasmas were observed. Identically manufactured units, except one had aluminum on the surface and the other had gold, were put into second breakdown with controlled mesoplasma emission. The surfaces of the devices were then studied with the aid of an electron microprobe. Figures 8 and 9 show the X-ray scans of both devices. Figure A is an optical photograph of the device; Figure B is the sample current scan; Figure C is the scan for the metalization; Figure D is the scan for silicon. In the case of gold (Figure 8), comparison between the optical photograph and the two X-ray scans shows the complete absence of gold in the hot spot region and only silicon remaining. For aluminum (Figure 9), it is seen that only some aluminum is missing and some silicon is evident. This analysis led to the theory of vaporization of gold and alloying/oxidizing of aluminum.

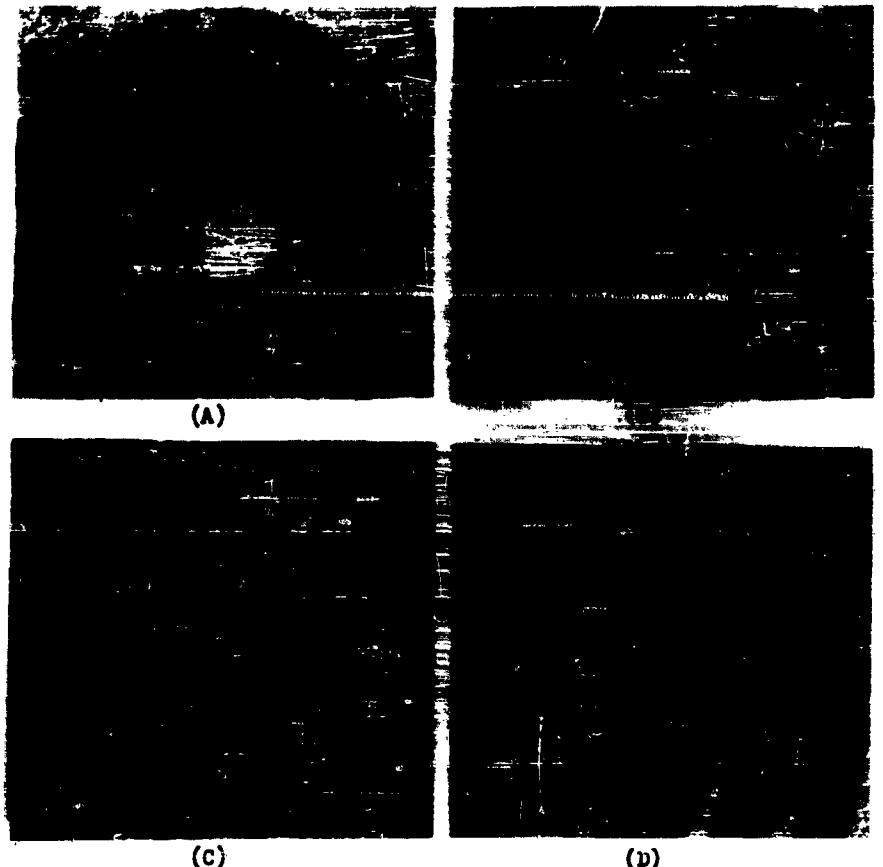


Fig. 8
(A) Optical Photograph of Gold Metalized Device With Second Breakdown Area; (B) Sample Current Scan; (C) Gold X-Ray Scan; (D)Silicon X-Ray Scan

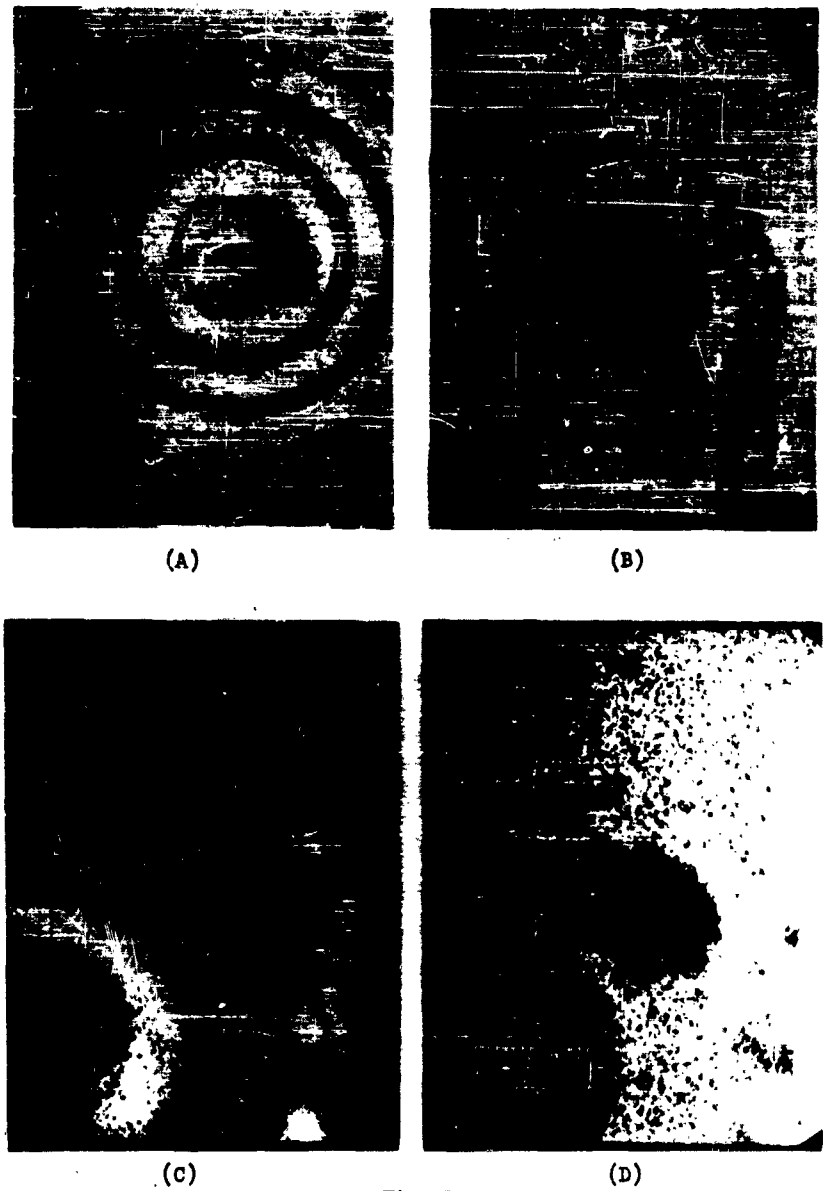


Fig. 9

(A) Optical Photograph of Gold Metalized Device With Second Breakdown Area; (B) Sample Current Scan; (C) Gold X-Ray Scan; (D) Silicon X-Ray Scan.

The second approach under investigation is to have an inert material in intimate contact with the top surface of the device and on top of this is placed a massive heat sink as shown in Figure 10. This scheme may reduce the surface temperature enough so that higher current densities are required to reach the critical alloying temperature. The coating between the surface of the device and the heat sink would serve two possible purposes; it would act as an electrical insulator and as a seal which could prevent rapid oxidation of the hot spot surface. Two coatings being attempted are epoxies and silicone resins.

The last approach is illustrated in Figure 11. Here the possibility of using the metalization as a fuse is being explored. The device shown has the emitter contact area in the form of thin-narrow fingers. If now the device is subjected to a transient, a hot spot will form under one finger. The metal above the hot spot will be displaced causing an open circuit in that finger. When the device returns to its operating point it will be minus one contact finger but the device characteristics should not be drastically altered.

IV. ACKNOWLEDGMENTS

The author wishes to take this opportunity to thank Mr. T. Redgate for his helpful comments and Mr. J. R. Shappirio and Mr. W. F. Nye for the X-ray photographs and their assistance in analyzing them.

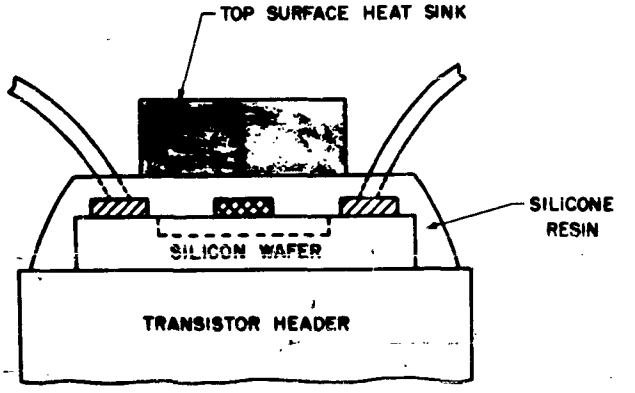


Fig. 10

Transistor Top Surface Heat Sink

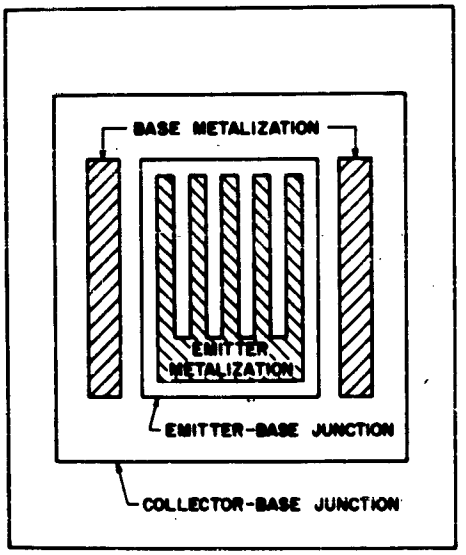


Fig. 11
Finger Emitter Contact Area

V. REFERENCES

1. Taue, J. and Abraham, A., "Thermal Breakdown in Silicon p-n Junctions," Phys. Review, Vol. 108, November 1957, pp. 936-937.

2. Thornton, C. G. and Simmons, C. D., "A New High Current Mode of Transistor Operation," IRE Trans. on Electron Devices, January 1958.

3. Ford, G. M., "Collector to Emitter Breakdown Related to Thermal Runaway in Homogeneous Base Germanium Power Transistors," Solid/State/Design, June 1963.

4. Hakim, E. B., McSherry, L.K., and Reich, B. "Solder Ball Formation in Silicon Alloy Transistors," IEEE Transactions on Electron Devices, June 1965.

5. English, A. C. and Power, H. M., "Mesoplasma Breakdown in Silicon Junctions," Proceedings of the IEEE, March 1965.

6. Hakim, E. B. and Reich, B., "Light Emission from Silicon Transistor Junctions," Proceedings of the IEEE, August 1963.

FAILURE MECHANISMS ASSOCIATED WITH DIE-TO-HEADER BONDS OF PLANAR TRANSISTORS

J. D. Guttenplan and F. H. Stuckenberg

Autonetics, a Division of North American Aviation, Inc.,
Anaheim, California

I. INTRODUCTION

Thermally induced physical changes occurring in the die-to-header bonds of discrete planar transistors were investigated as possible mechanisms which could cause the gradual degradation of transistor electrical characteristics. These physical changes were observed or identified in a physics of failure investigation of many hundreds of transistors which had been subjected to various levels of accelerated stress testing. The conclusions drawn from this investigation have resulted in corrective actions being implemented by the various transistor manufacturers, thereby accomplishing a major objective of the program -- to improve the reliability of discrete devices used in the Minuteman II missile guidance and control system.

The scope of the work reported here was limited to an investigation of failure mechanisms associated with transistor die-to-header bonds. It became apparent early in the work that the die-to-header interfaces of these transistors were complex metallic systems composed of many layers of different materials as a direct result of the fabrication processes. Many different procedures are utilized by the manufacturers to bond silicon dice to metalized headers. Undoubtedly the most popular method is to scrub a silicon die onto a heated, gold-plated ceramic header with an intermediate gold-silicon or gold-germanium preform to form a eutectic bonding layer. The ceramic headers are metalized with layers of screened on molymanganese and plated nickel, both diffused into the ceramic. These are followed in some instances with a plated copper strike to promote adhesion, a barrier layer of plated nickel, and the bonding layer of plated gold. In certain transistors, an intermediate layer of copper-gold-nickel or copper-silver braze

alloy used to fill the through-hole connections in the ceramic header, is found in the interface. Other devices contain an intermediate molybdenum pedestal plated with nickel and gold. In still other transistors, the silicon die is brazed to a copper and nickel plated header utilizing a silver-tin brazing alloy.

These complex metallic systems were further complicated by certain temperature and time dependent physical changes which occur in the interfaces. The following will be discussed in detail together with the possible mechanism by which each may result in eventual degradation of device electrical characteristics:

1. The solid-state diffusion or migration of intermediate materials into the gold bonding layer on the header. It is postulated that this can result in voids in the die bond due to nonwetting of the bonding layer by the preform or brazing alloy.

2. Outgassing of materials in the interface during the bonding operation resulting in void formation.

3. Intermetallic formation giving rise to vacancies due to differences in interdiffusion rates (Kirkendall Effect) and subsequent cracking along these vacancies due to stresses caused by thermal mismatch.

4. Thermal reordering of gold-silicon eutectic alloys causing changes in electrical resistance.

5. Solid-state phase transformation in the copper-gold braze material used to fill header through-holes resulting in electrical resistance changes and saturation voltage drift.

II. EXPERIMENTAL RESULTS

A. Solid-State Diffusion in the Bond Interface

A type of defect that was prevalent in a majority of the transistors examined on this program was voids in the die-to-header bonds. Examples of this are shown in Figures 1 through 3. Figure 1 is a cross-sectional view through the bond of a signal (micromesa type) transistor. Figure 2 is a similar area in a dual switch transistor, and Figure 3 shows voids found in the die bond of a large power transistor. Thermal analyses which considered the geometry and thermal properties of the materials composing the device were performed on models of various types of planar transistors. These analyses indicated that voids in the die bond had little effect on device performance until the voids exceeded 50 percent of the bond interface. However,

Figure 1

Voids in Die-to-Header Bond of a Signal (Micromesa Type) Transistor (350X)

Figure 2

Voids in Die-to-Header Bond of a Dual Switch Transistor (200X)

the results of infrared scans and radiographic examination of devices indicated that even small voids did result in hot spots and localized thermal gradients with corresponding temperature increases up to 15 F in the die. Larger percentages of voids in the bond interface resulted in thermal runaway and device failure. In thermal runaway, temperature rise increases the intrinsic carrier concentration, thereby

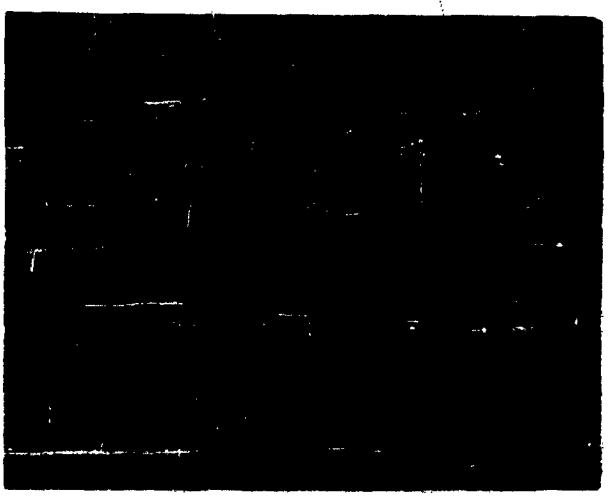


Figure 3

Voids in Die-to-Header Bond of a Power Transistor (140X)

increasing the current. This in turn again increases the temperature. Hence, both current and temperature increase without limit to failure. The failure rate goes up with high incidence of voids while the power load necessary to cause failure decreases.

X-ray analyses of the materials in the die-bond interfaces indicated that in many instances the solid-state diffusion or migration of metals was associated with void formation. For example, the area shown in Figure 3 was probed with the electron microprobe. Analysis of the x-rays, shown in Figure 4, indicated that copper had diffused nonuniformly around the void area. Similarly, the void area shown in Figure 5 was probed and a diffused copper-nickel layer was found directly beneath the gold-silicon eutectic layer. The die-to-header interface in Figure 6 was found to be composed of the following layers of materials, as determined by x-ray analysis, beginning at the ceramic header and moving up towards the silicon die:

1. Molymanganese metalized alumina
2. Nickel
3. Gold-copper-nickel alloy containing some diffused germanium
4. Gold-silicon eutectic with considerable segregation of silicon and containing diffused iron

Figure 4

Electron Microprobe Scan of Area Shown in Figure 3
(Bright Areas Represent Copper X-Rays); 130X

Figure 5

Voids in Die-to-Header Bond of a Power Transistor (300X)

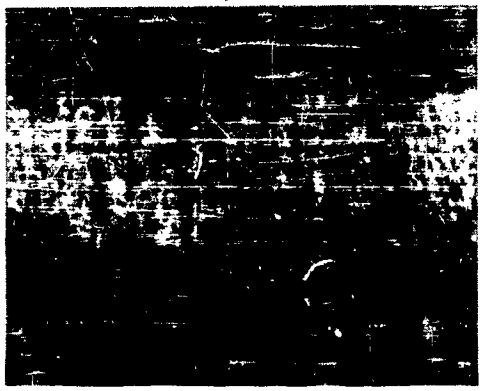


Figure 6

Die-to-Header Interface of a Signal Transistor (1600X)

5. Gold-germanium eutectic containing diffused copper and iron and some silicon

It was postulated from the foregoing results that the diffusion of materials in the interface could cause areas of nonwetting on the bonding surface by the preform or braze alloy, thereby resulting in voids during the bonding operation. To test this premise, simulated copper headers plated with consecutive layers, 200 μin. thick, of copper, nickel and gold were subjected to varying degrees of heat treatment to cause diffusion. This was carried out at elevated temperatures up to 550 C both in vacuum and in air. Two sizes of silicon dice, representing a large die found in the power transistors and a smaller die used in the signal transistors, were then bonded to these simulated headers. A gold-2 percent silicon preform was used to form the eutectic bond. The bonding was done in production die-bonding facilities. Appearance of the dice after bonding is shown in Figure 7. The conditions for pretreatment of the headers and areal percentages of voids estimated from radiographs of each bond are summarized in Table 1.

An analysis of the results of this experiment indicated the following:

1. Diffusion of copper and/or nickel into the gold bonding layer at 550 C in the absence of oxygen caused an approximate 100 percent increase in the overall average of voids. Increased time of diffusion did not significantly increase the percentages of voids, indicating that wettability of the solid solutions formed does not vary significantly over a wide range of concentrations.

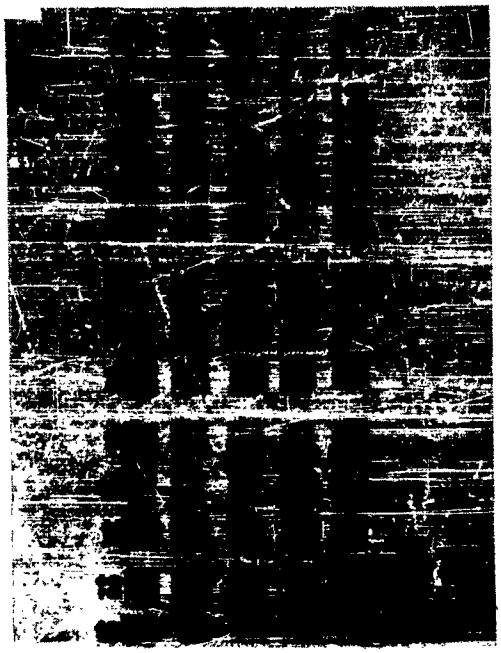


Figure 7

Silicon Dice Bonded to Diffused and Oxidized Headers

2. Oxidation of the gold bonding surface also caused an approximate 100 percent increase in the average percentage of voids. The effect appeared to be independent of temperature (in the range 350 to 550 C) and time (5 min to 120 min).

3. Combined effects of diffusion and oxidation were not additive.

It was concluded that both types of treatment lowered the wettability of the bonding layer, resulting in void formation during the bonding operation.

To determine the kinetics of diffusion in the bond interface as a function of time and temperature, in the actual range of die-bonding temperatures, diffusion couples were prepared and then analyzed with the electron microprobe. The system of consecutive layers of copper, nickel, and gold plates (each 200 μin. thick) was again selected as a system representative of the metalized header. Couples plated in this manner were heated at 300, 400, 475, and 525 C for periods of time up to 4 weeks. Coupons were removed at designated times for analysis.

Table 1

Die Bonding to Diffused and Oxidized Headers

Treatment	Percentage Voids		
	Overall Average	Average Large Dice	Average Small Dice
1. As plated (controls)	14.7	19.0	9.4
2. 30 min at 550 C (in vacuum)	28.5	28.0	29.0
3. 80 min at 550 C (in vacuum)	29.0	29.0	28.0
4. 420 min at 550 C (in vacuum)	29.5	31.0	27.0
5. 71 hr at 550 C (in vacuum)	31.0	35.0	23.0
6. 5 min at 350 C (in air)	27.0	31.0	23.0
7. 30 min at 350 C (in air)	28.5	34.0	23.0
8. 120 min at 350 C (in air)	29.0	34.0	24.0
9. 5 min at 550 C (in air)	27.5	30.0	25.0
10. 30 min at 550 C (in air)	28.5	25.0	32.0
11. 120 min at 550 C (in air)	31.5	25.0	38.0
12. 30 min at 550 C (in vacuum) + 5 min at 550 C (in air)	27.5	28.0	27.0
13. 135 min at 550 C (in vacuum) + 30 min at 350 C (in air)	29.5	32.0	27.0

An electron beam scanning method gave semiquantitative concentration profiles of the three elements of interest -- copper, nickel, and gold. A typical scan of a control coupon before diffusion is shown in Figure 8. For each scan, the ordinate or height of the scan indicates the concentration of the element. The abscissa of the scan indicates the distance across the interface of the three layers of plate, a total distance of approximately 15 microns.

Analysis of the scans from the 300 C diffusion study shows that copper diffused through the nickel layer and was present to a slight extent in the gold layer after 506 hr, (Figure 9). The secondary peak of copper in the gold layer represents approximately 6 percent copper. Some diffusion of nickel into gold and gold into nickel is also

Figure 8

Diffusion Study, Controls at 0 Hr (3200X)

Figure 9

300 C Diffusion Study After 506 Hr (3200X)

apparent. At 400 C, a slight diffusion of copper and nickel into the gold layer was discernible at 4 hr. At 24 hr, the amount of diffusion was equivalent to that found at 300 C after 506 hr. After 672 hr, copper was present in the gold layer to an extent of approximately 30 percent by weight and apparently was uniformly diffused throughout the layer (Figure 10). The nickel layer was still intact; however, nickel and gold were present in the copper layer. At 475 and 525 C, discoloration of the outer gold surface was noted in 30 min, indicating significant diffusion of copper and/or nickel into the gold layer. This was verified by the microprobe scans, Figure 11. Increasing amounts of inter-diffusion were observed at each temperature with increasing time. After 168 hours at 525 C, the microprobe scans show a central solid solution of gold, copper, and some nickel with peaks of nickel at each surface, Figure 12. The foregoing microprobe scans were used to calculate the diffusion coefficients and activation energies for the

Figure 10

400 C Diffusion Study After 672 Hr (3200X)

Figure 11

525 C Diffusion Study, 30 Min (3200X)

interdiffusion of gold-nickel and copper-nickel. However, no straightforward solution of the diffusion equation was available to compute the coefficients for the interdiffusion of copper and gold because of the intervening nickel layer. Therefore, additional copper-gold couples were run at 250, 350, 425, and 500 C for specified lengths of time to determine the constants for this system. Microprobe scans for a control and a copper-gold couple heated for 21 hr at 500 C are shown in Figure 13.

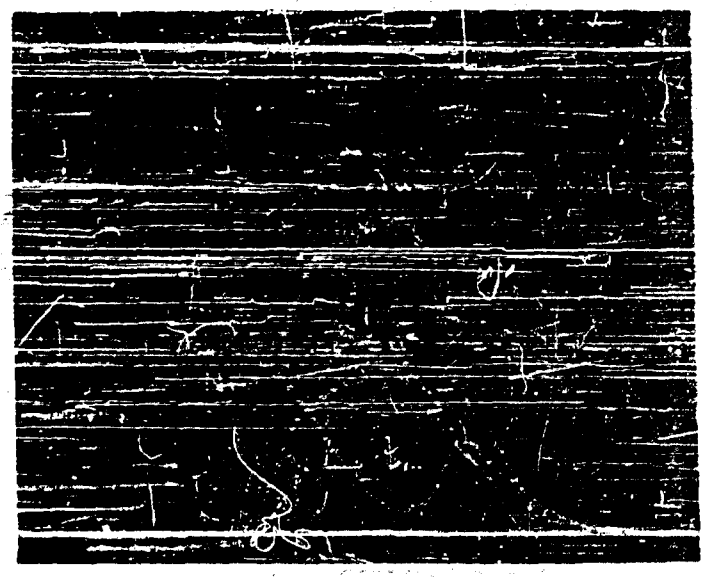
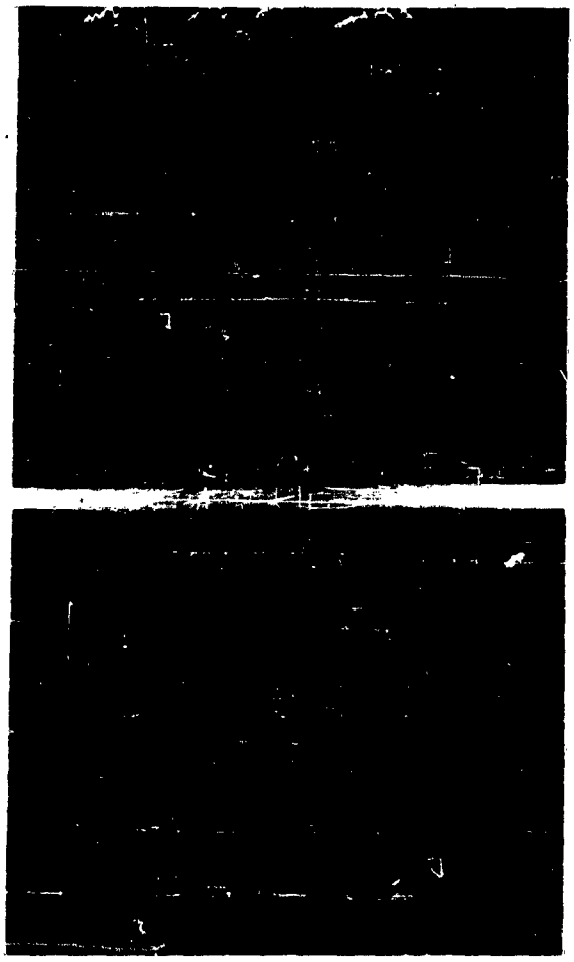


Figure 12

525 C Diffusion Study, 168 Hr (3200X)

An integrated form of Fick's Second Law, assuming a diffusant source of constant concentration, was used as the mathematical model for the solid-state diffusion process (Ref. 1 through 5). An abbreviated derivation is shown in Figure 14. At time, t, the concentration of metal, N(x), diffusing through a semi-infinite solid (bounded at distance x = o) for a distance equal to x is given by Eq (5). This assumes no depletion of the initial surface concentration, N_o, for $x < o$. D is the diffusion coefficient and erfc and erf are the error function complement and error function, respectively. Using data obtained from electron microprobe scans, diffusion coefficients were calculated for each temperature. Arrhenius plots of logarithm D vs the reciprocal of the absolute temperature, as shown in Figure 15, were used to derive the activation energies for the diffusion processes. A summary of the diffusion data is presented in Table 2, together with values obtained from the literature for comparison. Although the literature values are based on somewhat higher temperature ranges and different initial concentrations of the diffusing species, the agreement is apparent.

Using the values from Table 2, the amount of underlying metal that will diffuse to the bonding surface could be determined for each die bonding temperature as a function of the time required to bring the device up to temperature and to perform the bonding operation. For example, it was calculated, using Eq (5), that at a die bonding temperature of 525 C (as used in the fabrication of a large power transistor), approximately 22.5 percent of copper will diffuse into a layer of gold 100 μin. thick in 2 min, with 0.4 percent copper reaching the outer



a.

b.

Figure 13

Copper-Gold Diffusion Study
a. Controls at 0 hr (4.87μ /horizontal division)
b. 500 C for 21 hr (9.74 μ/horizontal division)

surface. Under the same conditions, about 5.5 percent of nickel would diffuse into the gold layer, penetrating about 50 μ in. Lesser, but still significant, amounts of diffusion will occur at the lower temperatures and in the shorter periods of time required for die bonding of the smaller signal and dual switch transistors.

FICK'S FIRST LAW: $\vec{F} = - D \vec{\nabla} N$ (1)

FICK'S SECOND LAW: $\frac{\partial N}{\partial t} = \frac{\partial}{\partial x} \left(D \frac{\partial N}{\partial x} \right)$ (2)

$$\frac{\partial N}{\partial t} = D \frac{\partial^2 N}{\partial x^2} \quad (3)$$

ASSUME DIFFUSION FROM CONSTANT SURFACE CONCENTRATION N_o INTO A SEMI-INFINITE SOLID BOUNDED AT DISTANCE X = O. NO DEPLETION OF N_o for X<O.

$$N(x) = N_o \operatorname{erfc} \left(\frac{x}{2\sqrt{Dt}} \right) \quad (4)$$

$$\frac{N(x)}{N_o} = 1 - \operatorname{erf} \left(\frac{x}{2\sqrt{Dt}} \right) \quad (5)$$

Figure 14

Derivation of Diffusion Equation

B. Outgassing of Materials

Many examples of spherical type voids that could not be attributed to nonwetting of either the plated header or die were found in the eutectic layers of these transistors. These voids resembled gas pockets in many respects as shown in Figure 16. To study the contribution of material outgassing to the formation of voids, samples of plated wire were sealed in evacuated glass vials (Figure 17), and then were heated at elevated temperatures for varying lengths of time. These vials were broken and the gases were collected for gas chromatographic analysis. Moisture, hydrogen, and oxygen, as well as hydrocarbons including methane, ethane, and acetylene, were identified among the outgassing products of copper, nickel, and gold plates. A typical chromatogram of the hydrocarbon portion (approximately 10 percent of the total) of gases evolved from gold plate at 550 C is shown in Figure 18. The remainder of the gases was determined to be moisture.

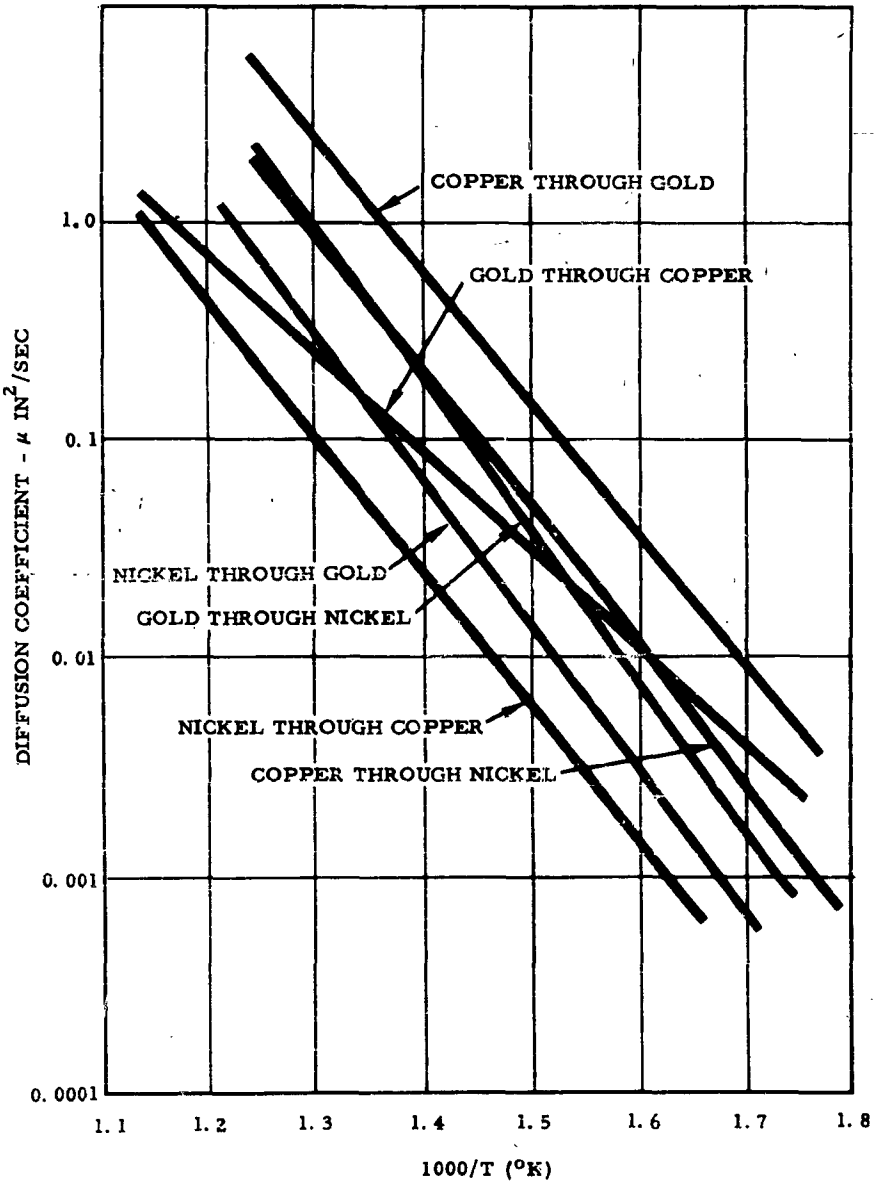


Figure 15

Arrhenius Plots for the Interdiffusion of Copper, Nickel, and Gold

Table 2

Summary of Diffusion Data

Solvent Metal	Diffusing Metal	Concentration of Diffusing Metal	Temperature Interval C	ΔE_{ACT} K cal/mole	D_o cm^2/sec	Reference
Nickel	Copper	Pure	300-525	29.2	1.14×10^{-3}	
Copper	Nickel	Pure	400-525	28.15	6.91×10^{-5}	
*Copper	Nickel	7.5-11.8 atom %	550-950	29.8	6.5×10^{-5}	(6)
Nickel	Gold	Pure	300-525	31.0	3.45×10^{-3}	
Gold	Nickel	Pure	300-525	30.5	1.0×10^{-3}	
*Gold	Nickel	15 atom %	800-1003	31.2	1.74×10^{-3}	(7)
Gold	Copper	Pure	350-500	27.8	1.22×10^{-3}	
*Gold	Copper	Pure	301-616	27.4	1.06×10^{-3}	(8)
*Gold	Copper	25.6 atom %	443-740	27.4	5.8×10^{-4}	(8)
Copper	Gold	Pure	350-500	22.06	3.23×10^{-6}	
*Copper	Gold	2.4-3.5 atom %	400-970	22.5	6.8×10^{-6}	(6)

*Literature value

Figure 16

Gas Pocket in Gold-Silicon Eutectic Layer of Transistor (400X)

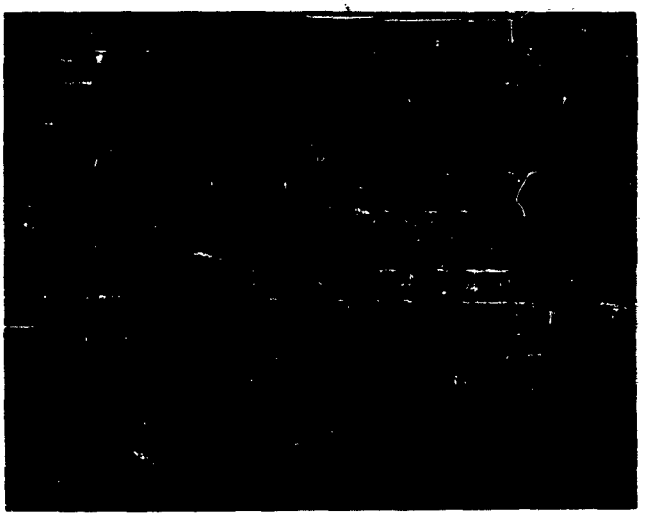


Figure 17

Sample Vial for Plate Outgassing Study

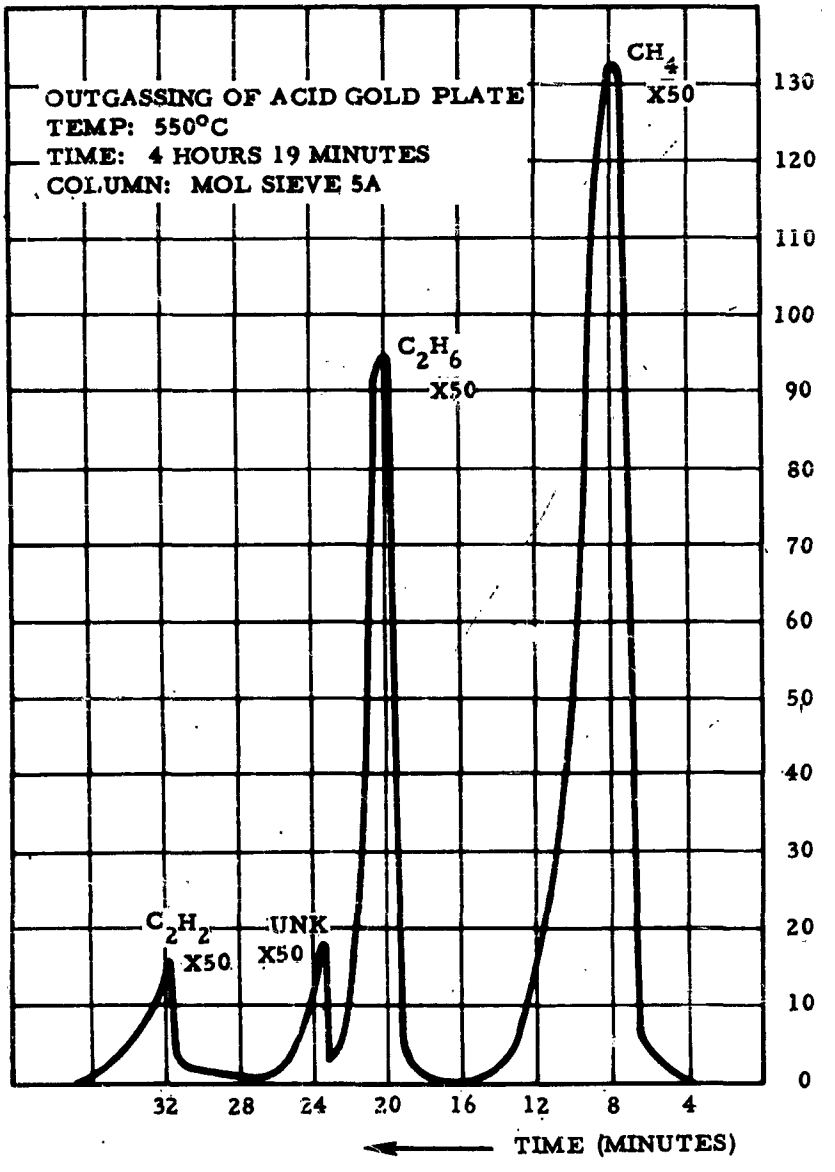


Figure 18

Chromatogram Showing Hydrocarbon Portion of Gases Evolved from Gold Plate

The kinetics of the outgassing of gold plate were studied at three temperatures in the transistor die-bonding range. Evacuated vials containing gold plated wire were heated at temperatures of 350, 450, and 550 C for periods of 10 min, 1 hr, and 4 hr at each temperature. The volume of gas evolved from each sample vial was determined from the pressure on breakage. An outgassing rate in units of milliliters of gas per cu cm of plate per sec was then computed for each condition of test. These values were plotted versus the midpoint of the time interval to obtain the family of rate curves shown in Figure 19.

Using these rate curves, the volume of gas evolved by a given volume of plate was determined for each die bonding temperature and time. For example, the die of one large power transistor is bonded at 520 C in an operation requiring a total of about 120 sec. Using the outgassing rate for 520 C at 60 sec and the dimensions of the die, it was calculated that 1.03×10^{-4} milliliters of gas were evolved in 120 sec. Comparing this with the average and minimum volumes of the gold-silicon eutectic bonding layer for this transistor, it was found that outgassing of the platings on the header could result in an average of 16.8 percent voids with a possible maximum of 50 percent voids in the bonding layer. Similar computations for a smaller signal transistor bonded at 450 C for 60 sec, indicated that an average of 5.6 percent with a maximum of 17.5 percent voids in the eutectic layer could result from outgassing. Comparing these values with the values found in Table 1, it is apparent that outgassing of materials on the header can account for a large percentage of the voids in the controls. It is also evident that the effects of diffusion and oxidation would be additive to the above values since the headers were effectively outgassed by the respective treatments shown in Table 1 prior to die bonding.

C. Intermetallic Formation

An unusually large number of cracks and/or voids, some extending the entire length of the die, were found in the die bonds of a certain make of signal transistor. An example is shown in Figure 20. It should be noted that these transistors had not failed accelerated stress testing prior to this examination. Electron microprobe analyses, Figure 21, indicated that separation occurred between a nickel-rich layer and a nickel-silicon intermetallic layer.

A similar situation was found in the die bond of one make of power transistor. This is discussed in greater detail in an accompanying paper. The cracking was attributed to stresses induced by differences in thermal expansion of the intermetallic and nickel-rich layers and appears to proceed along voids or vacancies formed by differential diffusion rates of nickel and silicon (Kirkendall Effect).

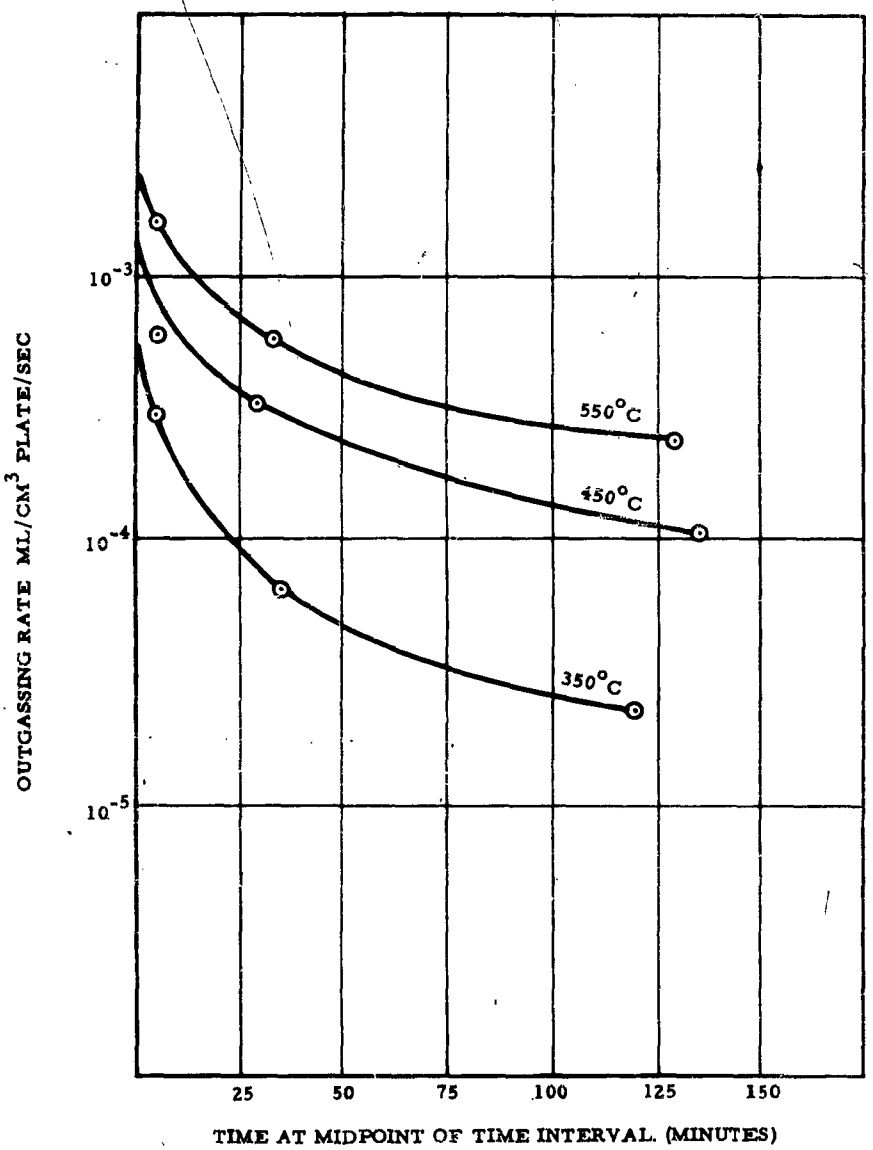


Figure 19

Outgassing Rates of Gold Plate

(a) Cross-Section Showing Horizontal Crack Under Entire Length of Die (95X)

(b) Closeup of Horizontal Crack (1600X)

Figure 20

Crack in Die-to-Header Bond of a Signal Transistor

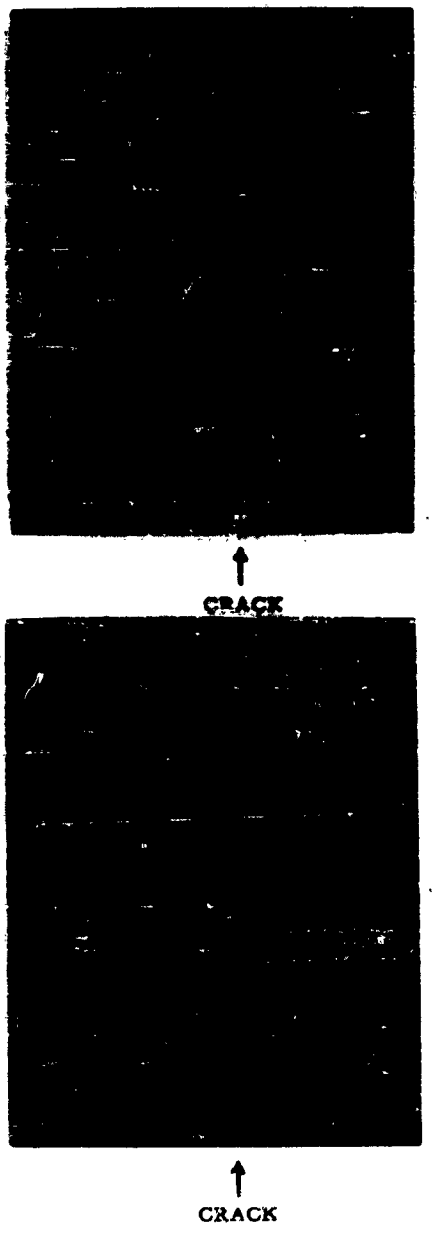


Figure 21

Electron Beam Microscans Across Crack in Signal Transistor Die Bond (1600X)

To investigate the kinetics of nickel-silicon intermetallic formation, silicon specimens were plated with thick layers of nickel and then heated at four temperatures for varying lengths of time up to 135 hr. Experimental temperatures included 250, 350, 450, and 550 C. These specimens were sectioned and analyzed by electron microprobe. No diffusion was apparent at 250 C after 135 hr. At 350 C, a narrow zone of diffusion was noted at 110 hr. This zone contained approximately 4 percent silicon. At 450 C after 135 hr, a diffusion zone containing 10 percent silicon was apparent. At 550 C after only 8.25 hr, at least two and possibly three zones of intermetallics were visible, as in Figure 22. It was calculated that the zone nearest to the silicon contained both Ni_2Si and Ni_3Si_2.

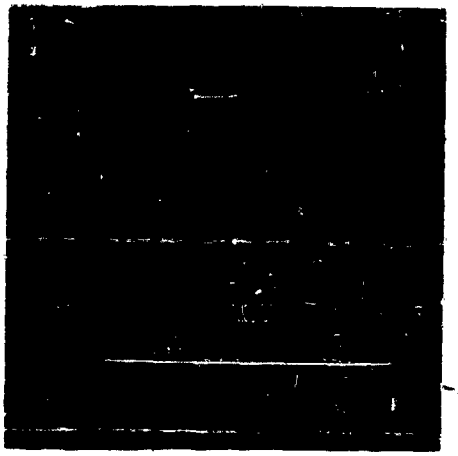
It is apparent from these results that nickel-silicon intermetallics can form at temperatures as low as 350 C over long periods of time. At higher temperatures, for example 520 C (the upper end of the die-bonding range), it is possible that nickel-silicon intermetallic formation may occur in the short period of time required for die bonding. However, it is probable that both the long periods of time at high temperature testing or operation and the shorter periods of time at die-bonding temperatures contribute to this condition.

D. Thermal Reordering of Gold-Silicon Eutectic

Large differences in metallurgical structure of the gold-silicon eutectic layer were observed. For example, in the die bond shown in Figure 23, the structure consisted of acicular precipitates of silicon, in the form of needles or platelets, in the eutectic melt. At the other extreme, a spherodized structure, Figure 24, consisting of small particles of silicon in the melt, was found. In a majority of the devices, the eutectic structure fell in between these extremes.

To determine the cause of the structural differences, gold-silicon eutectic alloys were prepared and then heat treated at elevated temperatures for various periods of time. The eutectic alloys were prepared by vacuum depositing and then plating gold to a thickness of about 1 mil on silicon chips. These chips were heated to 525 C and quickly cooled. Immediately after alloying, the eutectic in all tests had a structure similar to that shown in Figure 23. Heat treatment of the eutectic alloys at temperatures in the range of 150 to 350 C apparently caused a thermal reordering process. The silicon needles agglomerated and, depending on the time involved, the structure approached the spherodized appearance shown in Figure 24. The time required for complete stabilization of the structure varied from about 55 hr at 350 C to more than 200 hr at 150 C.

Accompanying these structural changes was a change in electrical resistance of the eutectic alloy. Resistance measurements were made at room temperature, from one side of the silicon chip to the other.



(a) BSE Areal Scan of Interface (1600X)

(b) Line Scans of Ni and Si X-Rays Across Interface (1600X)

Figure 22

Microscans of Nickel-Silicon Diffusion Couple, 8.25 Hr. at 550 C

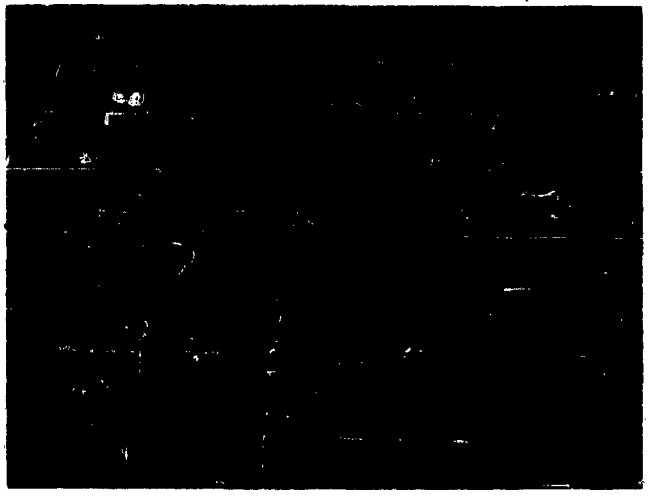


Figure 23

Acicular Precipitates of Silicon in Gold-Silicon Eutectic Melt (1700X)

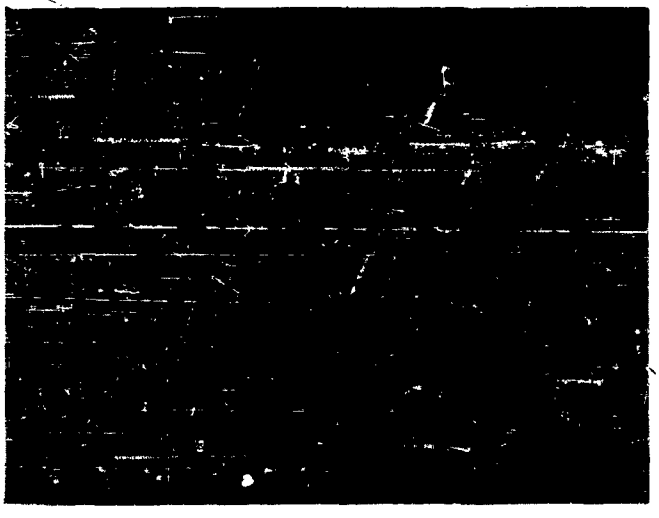


Figure 24

Spherodized Structure of Silicon in Eutectic Melt (2500X)

Figure 25, a plot of electrical resistance vs time at 200 C, indicated a general increase in electrical resistance of the eutectic alloy during the thermal reordering process until the structure became stabilized. Similar trends were observed at the other temperatures studied.

It is apparent from these results that in transistors, where the gold-silicon structure has not been stabilized, slow changes may occur in the eutectic alloy at device operating temperatures. These changes can affect the electrical resistance of the back contact and cause electrical parameter drift.

E. Phase Transformation in Copper-Gold Braze Alloy

In a common type of signal transistor (micromesa package), through-holes in the ceramic header are filled with braze alloy to provide electrical contact from the external leads to the base, collector, and emitter pads. When the through holes are filled the molten braze alloy is allowed to overflow the hole and wet the internal pads to ensure good contact, Figure 26. This hole-filling step is done prior to die-bonding. Thus, a layer of braze alloy becomes an integral part of the die-to-header bond.

Problems in saturation voltage drift encountered by one manufacturer prompted an investigation of this braze alloy. Electron microprobe analyses indicated that the braze was essentially a copper-gold alloy with a small percentage of nickel. In the devices examined, the gold concentration varied from 30 to 50 percent, nickel from 0 to 3 percent, with the remainder copper.

The concentration of gold in the braze alloy is a critical factor. As apparent in the copper-gold phase diagram, Figure 27, a solid-state phase transformation can occur in copper-gold alloys containing over 45 weight percent gold (Ref. 9). This phase transformation will occur at temperatures encountered in the high-stress testing to which these transistors had been subjected. Accompanying the phase change is an increase in electrical resistance which will vary from 16 to 54 percent in the temperature range encountered, Figure 28 (Ref. 10). It was calculated that these changes in electrical resistance could readily cause drift in the saturation voltage.

The kinetics of the phase transformation have not been well defined. However, it is apparent that the problem can be eliminated by reducing the gold concentration in the braze alloy to 40 weight percent or less.

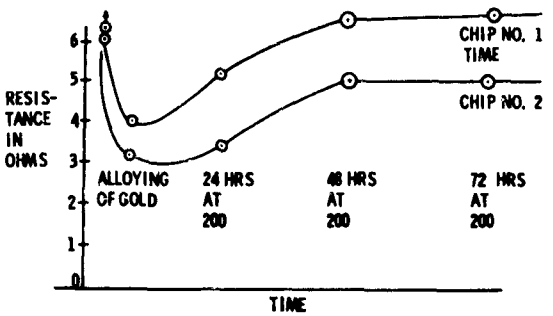


Figure 25

Change of Electrical Resistance of Gold-Silicon Eutectic Alloy During Heat Treatment

Figure 26

Cross Section Through Signal Transistor (Micromesa Package) Showing Overflow of Braze Alloy on Pad (140X)

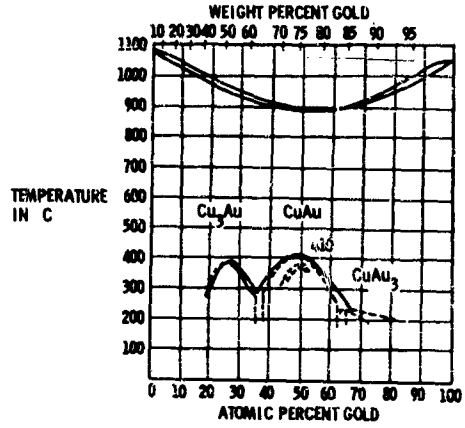


Figure 27

Copper-Gold Phase Diagram

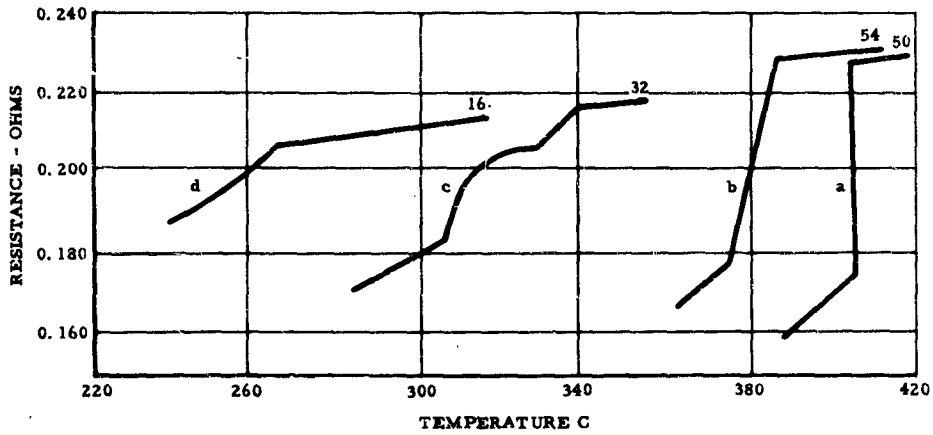


Figure 28

Resistance Changes Accompanying Phase Transformation

III. SUMMARY

An investigation of failure mechanisms associated with die-to-header bonds in planar transistors was centered on five thermally induced physical processes that occur in these bonds. The five processes include solid-state diffusion of intermediate materials into the bonding layer and outgassing of materials on the header, both resulting in void formation; nickel-silicon intermetallic formation leading to cracking; thermal reordering of gold-silicon eutectic alloys accompanied by electrical resistance changes; and a phase transformation in copper-gold braze alloys, also resulting in changes in electrical resistance. The kinetics of these processes were examined with relation to actual device materials and geometries to determine their effects on transistor electrical characteristics.

The results indicate that these processes are probably not major life-controlling mechanisms. However, any one or a combination of these processes may result in parameter drift and incipient failure during extended device operation. Certain immediate corrective actions are available to prevent these deleterious affects. Thermal stabilization or outgassing steps in the device fabrication and close control of materials are suggested. Other corrective actions will require an investigation of more efficient barrier layers on the header, substitution of materials to improve wettability, or perhaps a complete redesign of certain devices.

IV. REFERENCES

1. A. Fick, Ann. Phys., 94, 59 (1855).

2. F. M. Smits, "Formation of Junction Structures by Solid-State Diffusion", Lecture, Intensive Course in Semiconductor Theory and Technology at the University of Michigan, June 1962.

3. P. G. Shewmon, Diffusion in Solids, McGraw-Hill Book Co., Inc., New York, 1963.

4. W. Jost, Diffusion in Solids, Liquids, Gases, Academic Press Inc., New York, 1960.

5. C. J. Smithells, Metals Reference Book, Vol. II, Butterworth and Co., Ltd, Wash, D. C. 1962.

6. C. Matano, Japan Journal of Physics, 9, p. 41 (1934).

7. O. Kubaschewski, and H. Ebert, Z. Elektrochem, 50, 138, 1944.

8. W. Jost, Z. Phys Chem, B16, 123 19[illegible].

9. M. Hanse, Constitution of Binary Alloys, M[illegible]Graw-Hill Book Co., Inc., New York, pp. 198-203, 1958.

10. F. C. Nix and W. Shockley, Reviews of M[illegible]rn Physics, 10, 1938.