MIL-HDBK-175
1 MAY 1968

MILITARY STANDARDIZATION HANDBOOK

MICROELECTRONIC DEVICE DATA HANDBOOK

FSC-5962

DEPARTMENT OF DEFENSE
WASHINGTON D.C. 20360

MIL-HDBK-175
MICROELECTRONIC DEVICE DATA HANDBOOK

1. This handbook was developed by the Department of Defense and the National Aeronautics and Space Administration-1968 in accordance with established procedure.

2. This publication was approved on for printing and inclusion in the military standardization handbook series.

3. This document provides general information on microelectronic devices and their applications. It will provide valuable information and guidance to personnel concerned with the design, development, and production of equipment and systems employing microcircuits.

4. Every effort has been made to reflect the latest information on technology, reliability, testing and design considerations. It is the intent to review this handbook periodically to insure its completeness and currency. Users of this document are encouraged to report any errors discovered and any recommendations for changes or inclusions to the Commanding General, U. S. Army Electronics Command, Fort Monmouth, New Jersey Attention: AMSEL-TD-S.

FOREWORD

This document provides guidance for the selection and application of microelectronic devices in Military systems. Emphasis is placed upon considerations affecting reliability of systems employing such devices.

This handbook comprises six sections of user-oriented technical discussion, ranging from design, manufacture, use of the devices in subsystems, and specifications to reliability and failure physics.

Some of the material used is copyrighted; permission for its use is gratefully acknowledged, and acknowledgement is also made to all publications and microelectronic-device manufacturers, who responded generously to all requests for information from ARINC Corporation working under contract to NASA.

APPLICABLE DOCUMENTS

The following documents of the latest issue in effect form a part of this Handbook to the extent described herein:

SPECIFICATIONS

MILITARY

MIL-S-19500	Semiconductor Devices, General Specifications For
MIL-M-55565	Microcircuits, Packaging of

STANDARDS

MILITARY

MIL-STD-105	Sampling Procedures And Tables for Inspection By Attributes
MIL-STD-202	Test Methods For Electronic And Electrical Component Parts
MIL-STD-750	Test Methods For Semiconductor Devices
MIL-STD-883	Test Methods And Procedures For Microelectronics
MIL-STD-1313	Microelectronic Terms And Definitions

HANDBOOK

MIL-HDBK-108	Sampling Procedures And Tables For Life And Reliability Testing

CONTENTS

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

INTRODUCTION

O N E

This handbook is intended as a quick-reference document for use by design engineers, technicians, parts specialists, and by contractors. The test is addressed to readers with little or no experience in microelectronics. It is intended to provide general guidance for employing the technology. Solutions to the specific problems of equipment design must be considered in the context of cost, schedule, environment, and the other constraints of a particular application and are therefore beyond the scope of this handbook. The handbook does provide general information that will be of substantial assistance in the solution of specific problems.

This handbook is composed of six chapters including this introductory chapter; the remaining chapters are titled as follows:

Two:	Basic Processes and Design Considerations
Three:	System Design Considerations
Four:	Testing
Five:	Specifications and Procurement
Six:	Reliability and Physics of Failure

A flat pack requires a surface area of at least 125 x 250 mils, excluding leads, while the dual-in-line package (DIP) typically requires 250 x 700 mils of surface area, excluding leads. Leads require additional surface area. When simple circuits are packaged separately, the surface-area and volume efficiency is poor.

Space efficiency can be improved when multiple circuits are contained in a single package. Identical circuits (typically up to four) in a single package are frequently used. Circuit density is limited by the number of package leads on standard packages.

The hybrid microcircuit is one technique currently being pursued to increase circuity density. The ultimate approach is that of LSI -- in which perhaps as many as a thousand circuits could be packaged in a space no larger than 3 x 3 inches.

The surface-area and volume requirements of film and multichip circuits are usually greater than those of the monolithic integrated circuit. The multichips are typically put into a modified TO-5 package with a surface-area requirement of about 10^5 square mils. Film circuits typically require a minimum of 250 x 250 mils of surface area, excluding leads.

The weight of the smallest integrated-circuit package in general use is approximately 0.1 gram. Again, the assembled equipment averages a much higher weight -- an estimated 1 gram per unit -- than the simple sum of the weights of the integrated-circuit packages. As with volume, the weight of multichip and thin-film microcircuits is greater than that of the monolithic integrated circuit. Nevertheless, microelectronic circuits weigh 1/5 to 1/50 as much as conventional circuits when connected into equipments.

The estimates given above are quite rough and are valid only for comparisons at the circuit level. A given equipment's volume or weight may be changed drastically or only slightly with microelectronics conversion, depending mainly on the weight and volume of the parts that are not converted. Further, there is a strong trend to increase the functional complexity of the integrated circuit; increased complexity can often be achieved without materially increasing overall package weight or volume.

It is often mistakenly believed that an integrated circuit consumes much less power than an equivalent conventional circuit. While microelectronic equipments are often designed to consume considerably less power than conventional versions, this reduction cannot be attributed to an inherent power efficiency in the individual microelectronic circuits. Any power-speed combination available today in a microelectronic circuit can also be achieved with conventional components. The sharp reductions in power that often accompany conversion of an

equipment to microelectronic form are the result of changes in design. There is also a reduction in lead lengths and mechanical joints in the integrated circuit, with a corresponding reduction in power loss. This power difference is negligible in most cases at present, but as the power levels continue to decrease, the difference will become significant. If nanowatt circuits become available, the microelectronic circuit will indeed be significantly more efficient than conventional circuitry.

1.2 PERFORMANCE CHARACTERISTICS

The integrated-circuit technology has concentrated on digital circuitry because such circuitry does not require passive elements with tight tolerances and broad ranges of values and because there is usually a high degree of repetition of the same circuit in digital equipment. All digital functions are currently available in integrated-circuit form.

More than 2000 items are available off-the-shelf although many of these circuits overlap in function. This is not to say, however, that such circuits are interchangeable; most of the major manufacturers have developed their own compatible circuit family to provide most of the required digital functions. Mixing items from different families is generally not feasible; thus, in working from the standard inventory, a designer must initially choose the logic scheme and characteristics that best suit his needs and then design his system around a single family of circuits.

Special requirements, not covered by the stock circuits, can be met by custom-connecting (metalizing) the elements of standard matrix wafers, which are maintained in inventory for this purpose by some manufacturers. To fill even more exacting needs, fully custom circuits can be made by all manufacturers.

Although the monolithic structures are especially appropriate for digital circuitry, some multichip and hybrid circuits are used -- generally for the extremes of the operating frequency range.

Because simple monolithic circuits are satisfactory for digital applications, they received almost the full attention of manufacturers for a number of years. While an unsatisfied demand for digital circuits continued, other applications were neglected. Recently, however, the rush to establish positions in the pioneering digital market has leveled off somewhat, allowing effort to be directed into other areas -- notably linear circuits and complex arrays such as LSI and hybrid microcircuits.

Significant progress has been made in the application of microelectronics to analog functions. Circuits for performing linear functions are becoming

readily available. The most common linear circuit available is the operational amplifier. Other linear circuits include the following:

Audio amplifiers

Oscillators

Preamplifiers

If-limiting amplifiers

R-f amplifiers

Video amplifiers

Wide-band amplifier-discriminators

The offerings in the linear area, however, are by no means as complete as those in the digital area. The proportion is out of balance with the potential usage of the two types -- partly, of course, because of the later start made on linear functions. However, one of the factors that made the digital system so attractive to manufacturers of integrated circuits was its unitized design based on the repetitious use of a few basic circuit forms. An equivalent advantageous design situation does not exist to any such degree with linear devices, so that there is a constraint against the establishment of a stock inventory and a tendency to rely more on the custom approach.

1.3 COST

The price of the integrated circuit has been steadily decreasing for some time and now is competitive with that of conventional circuitry -- in the sense that the integrated circuit can be procured for no more than is required to procure, assemble, and test a group of conventional components that will perform the same electrical function. Much of the cost of an integrated circuit is incurred in the packaging of the crystal. This packaging cost remains essentially constant for a considerable range of circuit complexity. The cost per circuit function is expected to drop well below that of conventional circuitry in the next few years as the capability for increasing complexity (with constant yield) improves. Prices for the packaged device itself can be expected to decrease, but the added cost of testing will also have some effect on the price paid by the consumer.

Savings resulting from the widespread use of microelectronics begin at the device-procurement stage but will be far more significant after the equipments are placed in operation. Both maintenance and logistics costs can be drastically reduced even without standardizing circuit-performance characteristics. These reduced costs for maintenance and logistics result directly from increased reliability.

Microcircuits are tested extensively, including screening (see Section 4.3.3) and failure analysis (see Chapter 6), to improve equipment reliability -- since a reduction in equipment failures will decrease the number of maintenance actions required and thus reduce maintenance costs. If all other costs are equal, the life-cycle cost is also reduced.

Dertinger* presents several examples of the effect of MTBF on support cost. The examples illustrate that considerable savings can be realized by modest increases in MTBF. It is important to determine the cost per failure realistically to quantify the results of comparative evaluations based on MTBF's. Cost-per-failure figures are difficult to derive. In addition, many items of cost associated with the maintenance action may initially be sunken costs; for example, the savings associated with the spares pipeline may be initially zero because it may be necessary to maintain the supply organization at its current state. The size of such an organization is a function of what was required in the past and does not reflect current needs. The inertia alone will partially offset potential cost savings.

Nevertheless, the life-cycle costs can be reduced by improving device reliability. Since device prices are decreasing, it would seem advantageous to invest some if not all of the savings in more comprehensive testing and screening programs that would provide more reliable devices for installation in operational equipment. A small investment at the device level will provide a significant savings at the operational level.

1.4 MICROELECTRONIC DEVELOPMENTS

A common reaction to one's first encounter with an integrated circuit (properly magnified) is that "the whole production process must be completely automated". This is not the case, however; a major portion of integrated-circuit manufacturing is performed by highly skilled personnel.

The manufacturing steps are broken into two major groupings -- wafer processing and device assembly. The first group relies heavily on automation and the other on human labor.

The major role of people in the wafer-processing steps is in the operation of automatic equipments, loading and unloading these equipments, and transporting groups of wafers from place to place. Little increase in automation is expected in this area. However, improved automatic process control and testing by computers can be expected.

*E. F. Dertinger, "Status of Reliability Requirements in Government Contracts" 11th Nat. Symposium on Reliability and QC, Miami Beach, Florida.

Manufacturing steps that involve device assembly offer promise for automation (device testing is already highly automated in most plants). The chip-mounting and lead-attachment steps are particularly inviting for automation since they are almost completely dependent on human skill, with the accompanying cost and reliability penalties. Further, errors at these points are very difficult to detect. (There is no practical technique available for the nondestructive testing of minute connections and joints.) The flip-chip technique, in which the chip is turned face downward and the metalized contact areas are joined to similar areas preformed on a substrate, or some variation of it may provide the solution to the mounting and lead-attachment problems. Such techniques also introduce difficulties such as bond-inspection restrictions and sharply reduced thermal-transfer capabilities.

In addition to manufacturing steps, major areas of concern in the automation of integrated-circuit production are the production tooling (particularly the diffusion and intraconnection masks) and circuit design.

The electrical design of current integrated circuits is much more constrained by the topological layout of the circuit elements than that of conventional circuits. The coupling problems are more severe (because of the p-n junction method of isolation); typically a single level of intraconnections is utilized. As the circuits become more complex, the number of variables that must be considered becomes enormous. Considerable effort has therefore been expended on development of computer techniques to assist the designer in handling these variables.

The trend in microelectronics packaging is toward higher circuit density. Complex arrays are available to a limited degree; research and development is being performed on extremely complex devices that have the required intraconnections on the silicon wafer. Such an approach is called Large-Scale Integration (LSI) and has as its objective the intraconnection of hundreds of undiced circuits packaged in a container comparable in size to a silver dollar.

Considerable effort is being directed toward increasing the upper frequency limit associated with microelectronics. Microwave integrated circuits have been demonstrated as practical; microwave IC's are currently being fabricated in a number of laboratories around the country.

1.4.1 Large-Scale Integration

There is no clear definition for LSI within the industry. It is agreed, however, that it involves a high degree of complexity on a single silicon substrate and a component density requiring at least two levels of metalization. The LSI concept implies the fabrication of a complete function block or subsystem on a single substrate.

The LSI concept uses monolithic IC techniques at the system and subsystem levels. With LSI, IC's on the same substrate are intraconnected by metalization techniques. The circuits are arranged in a matrix and interconnected to achieve a desired signal processing.

The matrix elements may be identical IC's or groups of components that can be intraconnected to form multiple gates or various types of flip-flops. The first level of metalization provides the intraconnections for the components that make up the matrix element. It is also possible to place different circuits on the same substrate; for example, flip-flops may be interspersed with gates by combining the masks for each circuit during the step-and-repeat masking preparation. When the circuits are different, the circuit requiring the greater area is usually designed with a surface area that is an even integral number larger than the area of the other circuit.

The LSI applications have many advantages; there is no doubt that the concept will be applied more as the technology matures. Some of these potential advantages are the following:

- Fewer part types
- Improved performance
- Lower equivalent-device cost
- Improved reliability
- Smaller equipment size, weight, and volume
- Lower power requirements
- Simplified maintenance

Associated with these advantages are disadvantages whose relative importance is only a function of the current state of the art. As the technology improves, these constraints will be relaxed. Some of the important disadvantages associated with today's LSI technology are the following:

- Application to only repetitive circuits
- Sensitivity to cell yield
- Packaging problems
- Complicated test procedures
- Specifications
- Power density
- Increased turn-around time
- Mask complexity
- Coordination between circuit, package, and system designers

Both advantages and disadvantages are identical to those that are identified when integrated circuits and conventional circuits are compared.

The LSI device is fabricated in the same manner as silicon monolithic integrated circuits up to the point at which the wafer is diced. The wafer may contain hundreds or thousands of circuits depending upon circuit complexity, component density, or various design considerations. Before separating the dice

(scribing), each circuit is electrically tested to determine its performance characteristics. Those not meeting the performance requirements are marked. If the wafer is to be used for IC's, the marked dice are thrown away after scribing.

When the wafer is to be used for LSI, the wafer is not scribed and thus the "bad" circuits remain. They are, however, not intraconnected. The LSI device may not require the use of the complete wafer but only a matrix of circuits in an array. Ideally it is desirable to select an array of circuits from the wafer that does not include any bad circuits. If this is possible, a single set of metalization masks can be used for all subsequent identical arrays. When bad circuits are interspersed with good ones within the array, metalization must be done on a custom basis for each array (see Section 3.4).

The matrix of circuits, once metalized according to signal and bias requirements, is mounted in a package; the inputs and outputs are brought out from the array through the package leads. The result is a complex function in a single package instead of a single circuit (or a few identical independent circuits) in a single package. Design considerations for LSI are discussed in Chapter 3.

LSI will probably find initial application in digital assemblies with a high incidence of repetitive circuits, such as registers. Ultimately LSI may advance to a micro-computer (or "computer on a slice"). Any application of LSI has the advantage of reducing the back-panel wiring at the expense of increasing device metalizations, because a large amount of circuit interconnecting is done on the silicon substrate.

The proximity of the circuits reduces interconnection parasities and thus increases frequency response. While substrates will be large, component density will be high, making more efficient use of the wafer as compared with present IC efficiency.

The LSI device has the potential for being considerably less expensive than IC's in a design in which each is technically feasible. The LSI's initial limited volume and special one-time cost will keep its price high, but its cost advantage can be so great that for highly repetitive digital applications its use will be mandatory.

LSI has many disadvantages; some simply require maturation of the technology to be eliminated, while others will persist. New packages are needed with a variety of pin requirements; packages with as many as 160 leads are being developed concurrently with the devices.

LSI devices will require custom design for some time to come. While some applications are similar enough to use the same off-the-shelf LSI, these will be the exception unless system designers adopt a set of procedures for designing at a level higher than the circuit. Thus the re-use factor of any LSI device will be poor.

Testing the LSI device will be complex and costly if complete testing at each pin is required. It may be necessary to perform only selective tests or to test by substitution. Regardless of how testing is performed, LSI's present formidable testing and screening problems as compared with IC's.

Thermal properties of the device package and methods of heat transfer must be carefully analyzed in LSI applications. The high circuit density will require packages with excellent thermal properties. While thermal management is always a serious consideration, it is even more important for LSI.

LSI is a natural extension of the monolithic IC technology. It promises reduced size, weight, and volume, better reliability, lower cost, better performance, and lower power requirements. These advantages are the same as those predicted for microcircuits when they were compared with discrete circuits only a few years ago. Although the gains are similar, so are the problems -- which in many cases have been complicated in proportion to the complexity of the devices.

1.4.2 The Computer as a Microelectronic-Design Tool

Computers are currently being used to aid in the design and analysis of microelectronic circuits. They can be expected to play an ever-increasing role in future microelectronic developments, aside from their obvious role as a major consumer of these circuits. Their use as an engineering tool is not simply desirable but could become mandatory for design and analysis of the complex devices that are currently being developed.

Computer analysis of electronic circuits has became common in recent years. This is due, in large part, to the development of general-purpose circuit-analysis programs, such as the electronic circuit analysis program (ECAP)* and the network analysis program (NET-1).** These programs can be used by engineers who do not understand machine programming and may be applied to a range of problems dealing with a-c, d-c, and transient analysis. Graphic† man-machine interfacing may be used with the ECAP program.

Both ECAP and NET-1 have been developed to simulate any circuit that can be represented by lumped parameters. However, in practice there are a number of restrictions.†† One restriction is the requirement that only circuit elements for which the programs can provide models be included; these include fixed

* Available from IBM.

**NET-1 Network Analysis Program, Los Alamos Scientific Laboratory of the University of California.

† G.R. Hogsett and D.A. Nisewarger "An Application Experiment with On-Line Graphics-Aided ECAP," ISSSC (1967), p.72.

††D. Christiansen, "Computer-Aided Design: Part 6, "Electronics, (February 6, 1967), p. 74.

resistors, capacitors, inductors, mutual-inductive couplings, voltage and current sources, diodes, transistors, nonlinear elements, etc. One program requires that transistors be identified only by type numbers.

The complexity of the circuit being analyzed is restricted by the memory capacity of the computer; for example, the NET-1 is restricted to 300 each resistors, capacitors and inductors, 63 fixed-voltage sources, 63 time-dependent sources, and 200 nodes -- for a memory capacity of 32,000 words. ECAP is restricted to 50 nodes for the 7094 and 20 nodes for the 1620. Aside from these restrictions there are limitations in handling nonlinear circuits and in achieving accurate device models.

Circuit analysis, such as worst-case design, and a-c, d-c, and transient analysis, are applicable across the board. Computers also have more specialized applications in microelectronics, such as component layout to optimize surface-area requirements, design of diffusion and metalization masks (especially the complex metalization patterns required for LSI), partitioning for optimum grouping of circuits into LSI devices, and real-time process control of the fabrication sequence in device manufacturing.

To illustrate the possible application of the computer in microelectronic design and fabrication, a general discussion of various stages in design and processing of an electronic subassembly to be fabricated as a complex IC array will be considered. It is assumed that the subassembly has been specified at least to the extent of the input-output requirements.

The first step is determining the number and type of circuits that are required. Once the circuit requirements have been specified, preliminary design can be performed by use of a graphics-interfaced general-purpose computer program. D-c, a-c, and transient analysis can be performed and changes made as required to meet the circuit specifications. Performance sensitivity to parameter drift and element tolerances can be analyzed with computer-aided worst-case design studies.

Assuming that the subassembly is reasonably complex, it will be necessary to partition the subassembly into complex arrays that are technologically realizable. This involves determining the type and number of circuits on each array that will optimize some objective function, subject to a set of constraints.

The objective function may represent a figure of merit indicating the effective integration of each possible grouping of circuits.* The constraints represent the real-world restrictions imposed upon the fabrication of such arrays. For complex arrays, computer analysis is mandatory if meaningful comparisons of all possible groupings are to be made.

* A.R. Habayeb, C.E. Holland, Jr., and T. McDuffie,"System Partitioning for Microelectronics," Proceedings of Symposium on Microelectronics Applications, Garden City, N.Y. (September 1967).

Once the number and circuit type of the required arrays have been determined, computer-aided component and circuit layout may be performed to optimize surface-area requirements. Ultimately, the information generated by the computer may be used to cut the required fabrication masks automatically on a digitally controlled coordinograph.

The arrays are now ready to be fabricated. Computers can be used on a real-time basis to control the various process steps such as diffusion and etching. Continual testing and comparison of the results with a set of standards provides the necessary data for the computer to make decisions, such as whether to increase doping concentrations or alter temperatures or etching times. Errors that cannot be corrected will require the computer to make a decision on whether or not to start the processing over. Such tests and decisions may be made by the computer up to the time the array is ready for shipment.

After the wafer has been processed, each circuit is tested and the data are stored in the computer for subsequent use in determining a feasible metalization routing for the array. This is an iterative process that can easily be performed by the computer, assuming that a solution actually exists.

After the array has been packaged, a computer-controlled automatic tester may be used. Test data accumulated for each device (including results of various screening tests) may be stored to be easily accessible for future reference. (This might include examination of the data in relation to the results of subsequent failure analysis performed on the array.)

This example is intended to illustrate possible uses for computers in microcircuit design and fabrication. The illustration has not exhausted all possible applications. Computers are used in varying degrees in some of the applications mentioned in the example, while other applications are only in the conceptual stage.

1.4.3 Microwave Integrated Circuits

Integrated microwave devices are made with both silicon monolithic and hybrid fabrication techniques. By "integrated" the microwave engineer means a group of components or circuits that are fabricated in a single package. "Microwave IC" represents a grouping of microelectronic components or circuits in a single package. Such groupings might include monolithic elements, thin-film elements, beam-lead elements, or silicon monolithic circuits. Thus microwave IC's use both monolithic and hybrid techniques applied either together or separately.

Hybrid fabrication is usually performed on an aluminum-ceramic substrate whose dimensions are of the same order of magnitude as that of thin-film circuits. Quartz or other suitable materials may be used as the substrate. The basic requirements are that it be a good insulator (for component isolation),

have acceptable thermal properties, have the capability for achieving a micro-smooth surface (because thin-film elements are fabricated on the surface), and have an acceptable dielectric constant.

Microstrip transmission lines are used to transmit the signals. The substrate acts as the dielectric between the ground plane and the top conductor. The characteristic impedance (Z_o) of an unshielded line* is a function of the dielectric constant of the substrate (ϵ_r), the width of the top conductor, and the separation between the top conductor and the ground plane. Most applications require a Z_o of 50 ohms or greater. To achieve these values within the constraints of substrate thickness (typically 20 mils) and top-conductor width (about 2 mils minimum), a dielectric constant whose value lies in the range of 9 to 35 appears to be best.

The active elements of a hybrid device are typically separate, unpackaged semiconductor devices. Passive components such as resistors, capacitors, and inductors are formed by thin-film techniques. Ferrite devices are used where no reciprocal functions are required. When low-frequency circuits -- such as control circuits -- are required, a monolithic silicon circuit may be used and mounted directly on the alumina substrate.

Planar techniques that are used to manufacture low-frequency monolithic silicon IC's may also be used to fabricate microwave IC's.** The substrate for such circuits is high-resistivity (1500 ohms per cm) p-type silicon. For this value of resistivity, silicon is a reasonably good insulator and provides the required isolation between components.

Active components are fabricated in low-resistivity n-type pockets epitaxially grown on the high-resistivity silicon substrate. Inductors and capacitors are fabricated on the substrate by thin-film techniques. The transmission lines are microstrip; typical conductor dimensions are about 6 mils. The silicon substrate is 10 mils thick. The number of circuits that can be fabricated on a single wafer is dependent upon the surface-area requirements of the various components. Ertel has reported getting 41 transmit-receive (TR) switches on a one-inch silicon wafer.

A variety of microwave IC's has been fabricated -- included are an X-band balanced mixer, a 500-MHz amplifier, a 9-GHz receiver, oscillators, circulators, and TR switches. These circuits are typically restricted to operation of about one watt of continuous power. Ertel has reported that one circuit can sustain peak powers of 50 watts.

---

* Harold Sobol, "Extending IC Technology to Microwave Equipment," Electronics, (March 20, 1967), p. 112.

** Alfred Ertel, "Monolithic IC Techniques Produce First All-Silicon X-band Switch", Electronics, (January 23, 1967), p. 76.

Most of the microwave IC's that have been fabricated are hybrid circuits. These are simpler to build than monolithic circuits and are currently much less costly to build. The cost is very dependent upon yield, and if high yields can be obtained, monolithic circuits could be the least expensive.

1.4.4 Hybrid Microcircuits

A hybrid microcircuit is one that is fabricated by combining two or more circuit types (i.e., film circuits and semiconductor circuits) or a combination of one or more circuit types and discrete elements. The primary advantage of hybrid microcircuits is design flexibility. Components may be selected a priori by testing. Hybrid microcircuits find wide application in specialized applications, such as low-volume circuits, high-frequency circuits, and interim LSI (sometimes referred to as Medium-Scale Integration).

Several elements and circuits are available for hybrid applications. These include components that are descrete and that are electrically and mechanically compatible with monolithic IC's. They may be used to perform functions supplementary to monolithic IC's, and can be handled, tested, and assembled with with essentially the same technology and tools.

Such devices may be packaged individually or in groups, or they may be fabricated as beam-leads devices. Their small size allows their use in both thin- and thick-film circuits. Such devices are easily tested.

Elements are available in a variety of packages (including flip chip) that may be soldered, welded, or ultrasonically welded to an appropriate substrate. Size, cost, and performance are dependent upon the manner in which the elements are packaged. In addition to elements, complete circuits are available in the form of uncased chips (unencapsulated IC dice) and as flip chips. These chips are usually identical to those sold as part of the manufacturer's regular production line. It is necessary that they be handled with extreme care and be properly packaged and connected by the user if a high-quality final assembly is to be obtained.

An interesting application of the hybrid technology is the use of flip-chip circuits in complex arrays. The flip chip has small conducting lumps instead of bonding pads on the die for making electrical connections. The electrical connections are made by ultrasonically bonding or by soldering these lumps to a set of bonding pads on the substrate. Thus the die is inverted when bonded to the substrate. Such an approach introduces alignment problems during bonding and also may sharply reduce the heat-transfer capability of the device, because the die is attached to the substrate only through the bonding lumps.

The advantage of hybrid microcircuits is increased packaging density and reduced cost per circuit, because the cost of individually packaging each die is eliminated and the labor requirements for bonding are reduced. Reliability may be improved because the lead wires and thus the lead bonding are eliminated. All of the lumps are bonded at the same time; this should improve device reliability.

As an interim approach to LSI, flip-chip circuits or uncased dice can be mounted on a substrate (usually alumina). The dice can be mounted in a matrix array and the interconnections printed on the substrate. The array is sealed as a module in a ceramic package to protect the chips from mechanical and environmental stresses. Appropriate pins are brought out of the package to interface the dice and make external connections.

BIBLIOGRAPHY: CHAPTER 1
INTRODUCTION

Aleksander, I. "Integrated Circuits -- New Thoughts on Computer Circuit Design," Microelectronics and Reliability, 11(1964), 81-91.

Bohman, J. "Large-Scale IC Arrays: The Next Step Forward," EDN, (1966), 96.

Boehm, W.R. "Microelectronics -- Its Impact on Naval Ships," Naval Engineer Journal, 111(1964), 365-376.

Boysel, B."Memory on A Chip: A Step Toward Large-Scale Integration," Electronics, (1967), 93-97.

Brooks, F.P., Jr. "The Future of Computer Architecture," Proceedings of LFIPS, (1965), 93-97.

Brauer, J.B. Microelectronics for Command and Control Systems: Today and Tomorrow. Rome, N.Y., 1966.

Brauer, J.B. The Numbers Game -- Who Wins?, Proc. IEEE/ASQC.

Ertel, Alfred."Monolithic IC Techniques Produce First All-Silicon K-Band Switch," Electronics, (1967), 76-91.

EEE. "Tom Hyltin of TI and his Microwave IC's,"(Jan. 1967), 74-79.

Evans, R.J., A.G. Revesz, and J.T. Wallmark. Investigations of Fundamental Limitations Determining the Ultimate Size of Micro-Structures. RCA Report. 1964.

Flynn, M.J. "Complex Integrated Circuit Arrays: The Promise and the Problems," Electronics, (1966), 111-16.

Foley, W. "The Chips are Finally Down -- Monlithic Circuits Now," Product Engineering, (1966), 69-73.

Henle, R.A., and L.O. Hill. "Integrated Computer Circuits: Past, Present, and Future". Proceedings of the IEEE, (1966), 1849.

Hobbs, L.E. "The Impact of Hardware in the 1970's" Datamation, XII(1966),36-44.

Kaplan, I.E. The Projected Effects of Electronic Developments on Navy Personnel Requirements During the Period 1965 - 1980; A Critique of the Original Report and A Discussion of the Effects of Late Developments in Microelectronics on the Navy's Human Element. No. 65-20. 1965.

Lathrop, J.W. "The Status of Monlithic and Thin Film Circuits" Electronic Industries, (1965), 38-42.

Lyau, C.F.G. A Selected Bibliography on Microelectronic Integrated Circuits. Autonetics. 1965.

Madland, G.R., and J. Jolly. "The Importance of Integrated Circuits to Engineers," Electronic Industries, (1966), 40-44.

Mays, H.C. "Computer-Aided Design for Large-Scale Integration," International Solid State Circuits Conference, (1967).

NASA. Microelectronics in Space Research. SP-5031. 1965.

Naval Ordnance Laboratory. Microelectronics Requirements At The Naval Ordnance Laboratory. NOLTR 65-229. 1966.

NAVSHIPS. Integrated Circuits Handbook: Miniature Parts and Integrated Circuit Devices for Electronic Equipment. 1965.

Parks, M.S. The Story of Microelectronics: First, Second and Future Generations. North American Aviation. 1966.

Sack, E.A., R.C. Lyman, and G.Y. Chang. "Evolution of the Concept of a Computer on a Slice" Proceedings of IEEE, (1964), 1713-20.

Seelbach, W.C. Computer Sub-Systems on a Chip of Silicon. Motorola Internal Report.

Sterrett, D.S. "Introductory Remarks for Bureau of Ships Microelectronics Program," Proceedings 3rd Conference Navy Microelectronics Program, (1965), 81-82.

Spitalny, A., W.S. Mann, and G. Hartstein. "Computer Aided Design of Integrated Circuits," Proceedings of 1965 SHARE Conference, (1965), 1-32.

Sprague, J.L. "Integrated Circuits," International Science and Technology, (1964), 43.

U.S. Naval Applied Science Laboratory. Handbook for Systems Application of Redundancy. 1966.

Vadaz, L. "The 'Micromatrix' Approach to MOS Complex Arrays," Wescon (1966), 1-5.

Wolf, H., and K. Greenough. "The Trend Toward Monolithic Subsystems," Second International Symposium on Microelectronics, (1966).

BASIC PROCESSES AND DESIGN CONSIDERATIONS

T W O

The design of microelectronic circuits is similar to the design of discrete-component circuits. Most of the early microcircuit designs were accomplished by a component-for-component substitution in an existing discrete-circuit design. The only engineering design differences were those associated with the manner in which the circuits were fabricated.

The various fabrication methods (i.e., monolithic silicon and thick and thin films) introduce design constraints. These result from the small dimensions of components, parasitics introduced by certain component-isolation techniques, component-tolerance limitations, and various voltage and current limitations.

The monolithic fabrication technology limits performance because all of the components are fabricated in situ and different components or parts of different components are fabricated at the same time. Thus no advantage can be taken of optimum processes for constructing various components. When discrete circuits are fabricated, each component can be tested and optimum choices can be made.

Thin- and thick-film microcircuits provide more design freedom and approach more closely the flexibility associated with discrete circuits. Components can be adjusted to achieve desired close tolerances when required. In addition,

discrete transistors, which may be tested before attachment in the circuit, are required in these circuits because there is currently no adequate method for fabricating active film components. This has been a serious constraint and has limited the usefulness and acceptance of such microcircuits.

This chapter is intended to provide insight into the monolithic fabrication techniques so that the designer can understand the nature of the design constraints imposed by the fabrication method. The methods and techniques discussed are of only general interest to the circuit designer because the problems associated with this phase of microcircuit design are largely in the domain of the microcircuit manufacturer. This discussion, however, forms a sound base for the remaining topics, which include circuit specification and applications.

The salient characteristics of both linear and digital circuits are presented and various applications are discussed. A design example of a modified DCTL NOR gate is explained and evaluated, and design and fabrication concepts for MOS microcircuits are presented. The chapter is concluded with a brief discussion of film components, including a comparison of various characteristics of monolithic, MOS, and film components.

2.1 BIPOLAR TRANSISTOR DEVICES

2.1.1 Preparation of Semiconductor Materials

The rapid advances in semiconductor-device technology over the past fifteen years would not have been possible except for the impressive progress made in the purification and preparation of semiconductor materials. Near-perfect crystals of virtually absolute purity are needed if consistent, high-quality devices are to be made. The impurity level must be controlled to less than one impurity atom per million -- at times, to less than one atom per billion. This level is not detectable chemically or spectroscopically. Its measurement depends entirely on electrical conductivity. The expression for conductivity is

$$\sigma = q(\mu_p p + \mu_n n) \qquad (2\text{-}1)$$

where

μ_p = mobility of holes

μ_n = mobility of electrons

p = number of holes per cubic centimeter

n = number of electrons per cubic centimeter

q = electron charge

For an intrinsic semiconductor n or p can be assumed equal to the donor or acceptor impurity density, respectively. When n is very much greater than p, the semiconductor is n type (electrons are the majority carriers); also, generally, n is much greater than n_1 (the intrinsic free-electron density), so that n_1 can be ignored when the concentration of donor impurity is being determined.

2.1.2 Measuring Conductivity

The heated-probe check is a convenient method for determining the polarity of the charge carrier -- that is, to determine whether the semiconductor wafer is n-type or p-type. A galvanometer is connected to the semiconductor wafer by means of two probes that are conveniently placed on the wafer. One of the probes is at room temperature, and the other probe is attached to a heated soldering iron, as shown in Figure 2-1. With the hot probe connected to the negative side of the galvanometer, the needle of the galvanometer registers a negative reading for the n-type material and a positive reading for the p-type material.

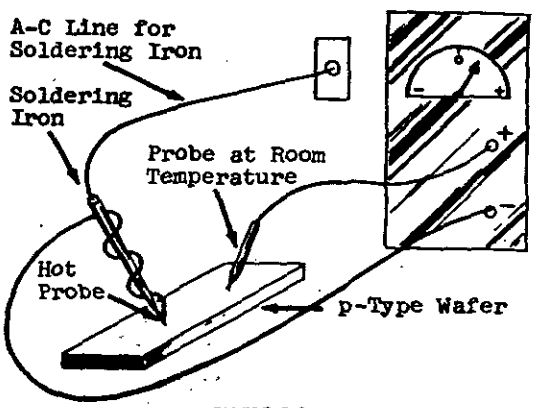


FIGURE 2-1
THE HEATED-PROBE CHECK

The heated-probe check can be set up with a microammeter instead of a galvanometer. When current is indicated on the microammeter with the hot probe connected to the negative terminal of the meter, the wafer is p-type. When the hot probe is connected to the positive side of the microammeter and current is indicated on the meter, the sample is n-type. These indications occur because the majority carriers diffuse in the direction of the negative temperature gradient. The majority carriers are electrons in the n-type material and holes in the p-type material. In an n-type semiconductor, the majority-carrier electrons diffuse away from the hot probe; thus they cause it to become positive with respect to the colder probe.

A commonly used technique for measuring resistivity is the 4-point probe technique (ASTM Test Method T43-64T), shown in Figure 2-2. Four equally spaced probes are aligned on the semiconductor wafer. In the figure, the resistivity is given by

$$\rho = \frac{V}{I} 2\pi SF\left(\frac{W}{S}\right) \qquad (2\text{-}2)$$

where $F\left(\frac{W}{S}\right)$ is a geometrical correction factor that corrects for alteration of current-flow lines through the sample for special boundary conditions such as a finite sample thickness, W.

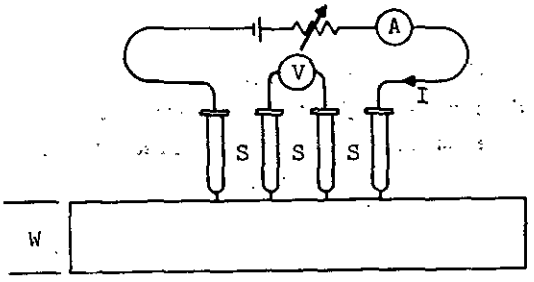


FIGURE 2-2
ARRANGEMENT FOR FOUR-POINT PROBE MEASUREMENT

In addition, the Hall effect can also be used to measure conductivity type and carrier concentration. When a sample piece of semiconductor is placed in a magnetic field and a current is made to flow perpendicular to the field, a voltage is produced across the sample along an axis perpendicular to both the magnetic field and the direction of current flow, as shown in Figure 2-3. The magnitude and direction of this voltage can be used to determine the concentration and sign of the current carriers in the semiconductor sample. For a sample of width, W, the Hall voltage is given by

$$V = 10^{-8} \frac{RIH}{W} \qquad (2\text{-}3)$$

where V is in volts, I in amperes, H in Gauss, and W in centimeters, and R is the Hall constant given by

$$R \approx \frac{3}{8} \qquad (2\text{-}4)$$

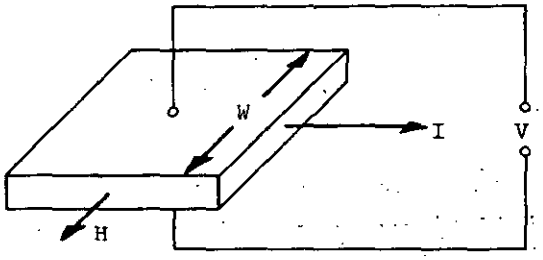


FIGURE 2-3
HALL EFFECT

The expression for sheet resistivity is developed as follows:

The resistance of the conductor illustrated in Figure 2-4 is

$$R = \rho \frac{\ell}{tw} \text{ ohms} \qquad (2\text{-}5)$$

where

- ρ = resistivity of the diffused conductor
- ℓ = length of the diffused conductor
- t = thickness of the diffused conductor
- w = width of the diffused conductor

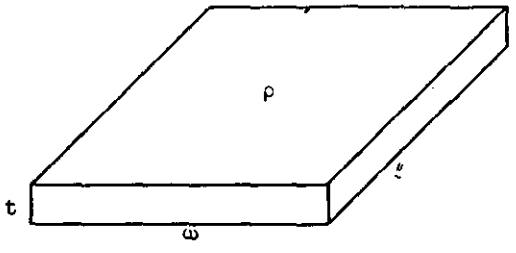


FIGURE 2-4
PARAMETERS FOR DETERMINING SHEET RESISTIVITY

The sheet resistivity is defined as

$$\rho_s = \rho/t \tag{2-6}$$

This is valid when the depth of the conductor is small (a few microns) and it is relatively constant, as is the case for most diffused resistors. Substituting ρ_s for ρ/t in the above equation for resistance gives

$$R = \rho_s \frac{\ell}{W} \tag{2-7}$$

or

$$R = \rho_s K \tag{2-8}$$

where

$$K = \ell/W$$

The term K is referred to as the length-to-width ratio and is important in specifying resistor values. If ρ_s is known, the resistance of any resistor is determined simply by specifying the ℓ/W ratio.

As an example, consider a system that requires a 5-kΩ resistor with sheet resistivity of 50 ohms/sq.

Since R and ρ_s are given, it is only necessary to solve for K:

$$K = \ell/W = R/\rho_s = \frac{5K}{50} = \frac{500\emptyset}{5\emptyset} = 100 \tag{2-9}$$

The resistance is determined by $\ell/W = 100$. There are an infinite number of solutions, and the actual choice of dimensions will depend on other factors such as required surface area for power dissipation, or distributed capacitance.

2.1.3 Alloying

In an npn transistor, three regions alternate in the conductivity type. In an alloy transistor the starting material is p-type, and an n-conductivity layer has to be realized on either side of the wafer, as shown in Figure 2-5.

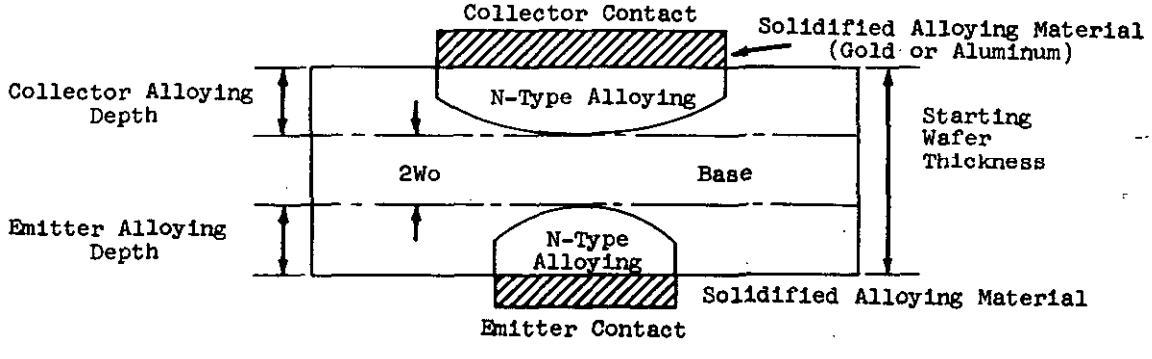


FIGURE 2-5
ALLOY TRANSISTOR

These regions are formed by placing spheres or discs of doping material on opposite sides of a prepared wafer, and heating to allow the dope to melt and dissolve part-way into the wafer. As the molten alloy is carefully cooled, dissolved germanium recrystallizes at the liquid-solid interface of the wafer, which acts as a single-crystal seed. The temperatures at which these processes are carried out vary from 700° to 900°C depending on the alloying material used.

Attempts to reduce the dimensions of alloy transistors for high-frequency applications led to the introduction of electrochemical etching and plating techniques, which made possible the development of interface-barrier transistors. In this device, very close emitter-collector spacings are possible. It is necessary to start with a thin wafer of semiconductor and then etch it still further by subjecting it to two coaxial jets of etching solution. When the thickness of the central web has been reduced to about 0.2 mil, metal contacts are electroplated into the same points, with the same jets being used as electrolyte vehicles.

2.1.4 Diffusion

By diffusing donor and acceptor impurities into semiconductor materials, it is possible to fabricate p-n junction devices that have superior electrical characteristics. The diffusion process has many distinct advantages over other fabrication processes, particularly if silicon is the material used. It is the basic process and in most cases the only process in which superior transistor characteristics can be realized. Junction depths and impurity concentration of the layers can be controlled more precisely than in alloyed structures.

After the introduction of these diffusion techniques into semiconductor fabrication, it became possible to achieve base widths, with high accuracy, of less than 1 micron, which improves considerably the transport factor (and therefore α) and the transport time (and therefore f_α). Other means of improving the high-frequency response would be to reduce the collector-base junction area, as has been done with mesa transistors. A typical mesa transistor is shown in Figure 2-7.

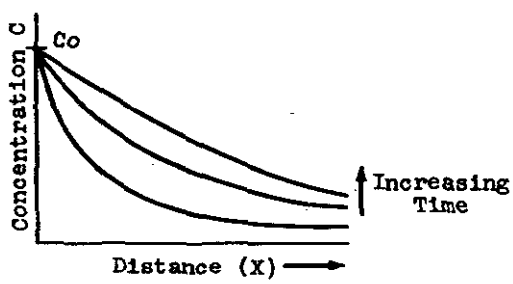


FIGURE 2-6
IMPURITY CONCENTRATION VS DISTANCE

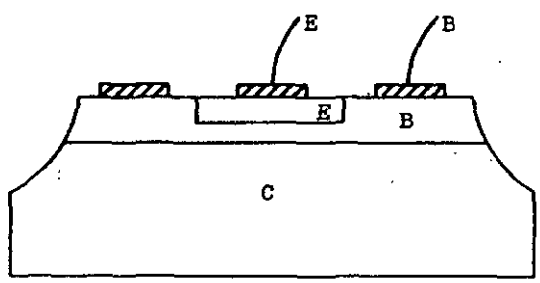


FIGURE 2-7
MESA TRANSISTOR

The defect of such a structure is apparent. The collector-base junction is exposed to the surrounding atmosphere, and surface effects at this point are severe. To minimize surface effects on junctions, the planar diffused structure was introduced by Fairchild. A typical planar structure is shown in Figure 2-8.

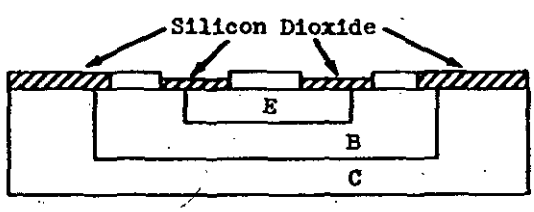


FIGURE 2-8
PLANAR STRUCTURE

The planar process permits the passivation of the surface by an oxide layer at an early fabrication stage. The silicon-oxide coating is grown on the surface before any junctions are diffused. This greatly improves the parameters that are particularly sensitive to surface conditions. These parameters are the reverse leakage currents, breakdown voltages, noise figure, and low current β. Planar structures also made it possible to solve many problems in the realization of integrated structures, in which several elements are built simultaneously on the same piece of semiconductor.

To show how the planar process is used to fabricate devices for microcircuits, the steps taken in building a planar-diffused transistor are described here. This process is common to all bulk-type microcircuits.

The steps used in the fabrication of an npn planar-diffused transistor are shown in Figure 2-9. The starting material is n-type silicon (A). The entire surface of the n-type silicon wafer receives a thermally grown oxide layer. A window is chemically opened through the oxide layer (B). A p-type diffiusion is then performed to construct the base region of the transistor (C). The wafer with the window is placed in a diffusion furnace, with an impurity such as boron, for a few hours at about 1300°C. The junction formed in (C) is formed under the original oxide layer and never sees the ambient environment of the diffusion furnace. A new oxide layer is formed during the diffusion.

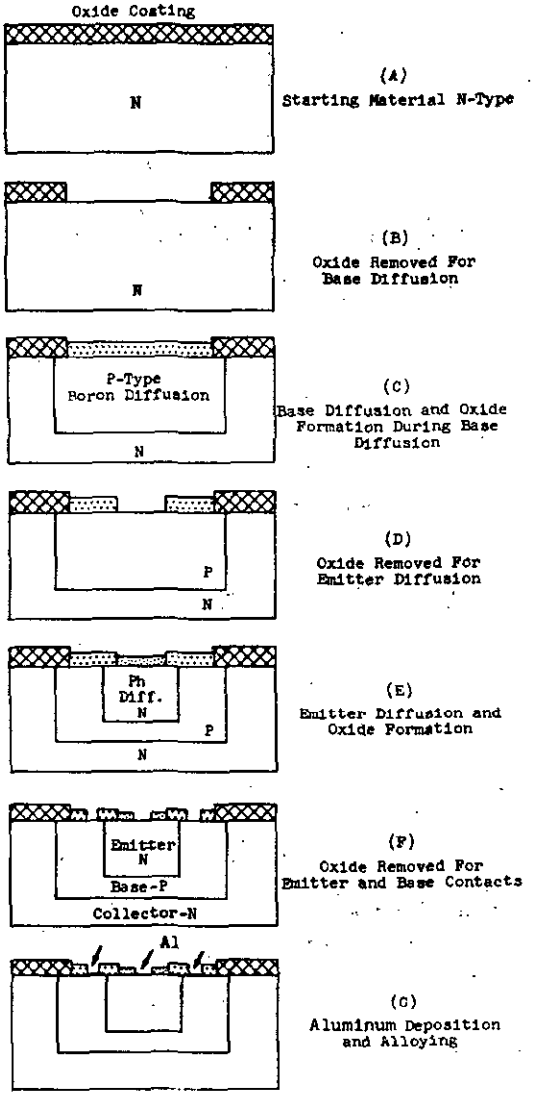


FIGURE 2-9

FABRICATION STEPS

A window is now opened through the new oxide layer (D). This window is much smaller than the previous window used for the base diffusion. The n-type emitter diffusion is carried out in the same manner as the base diffusion except that now an n-type dopant such as phosphorous is used (E). During the diffusion, the oxide layer is regrown across the window.

Through precision masking techniques very small windows are opened in both the base and emitter regions (F). These windows are about 0.001 inch in width for many transistors. Aluminum is then deposited into these windows and alloyed to form the ohmic contacts to the transistor (G). The lead wires may now be attached. The collector connection is frequently made by soldering the substrate directly to the header. It should be noted that the p-n junctions (throughout the fabrication process and on the finished device) have always been beneath the oxide layer. The oxide layer protects or passivates the p-n junctions from outside effects

that can cause degradation of the device.

Thus, in addition to furnishing the masking during diffusion, the planar oxide layer or coating is also a means of protecting or passivating the p-n junction. Planar devices are used in bulk-type semiconductor microcircuits instead of mesa devices. As was shown in Figure 2-7, the p-n junctions on mesa-type devices are exposed.

2.1.5 Oxide Masking

The opening of windows in the silicon dioxide layer forms a stencil-like structure through which diffusion and alloying can be carried out. This stencil-like structure or oxide mask is made by the following process (usually referred to as the planar process):

(1) The clean silicon wafer is placed in a furnace, where the oxide layer is grown.

(2) The oxide layer is coated with a photosensitive acid-resistant film such as Kodak KPR.

(3) The KPR film is stabilized with infrared light for about thirty minutes.

(4) A positive (the areas where diffusions are desired are black areas on a transparent film) of the configuration desired is placed over the wafer with the KPR film.

(5) Ultraviolet light is used to illuminate the KPR film through the positive. Areas of the KPR film exposed to the ultraviolet light are hardened and are made impervious to acid etches; areas not exposed are removed by developing.

(6) The KPR is developed so that it is removed where not exposed. The result is an acid-resistant film stencil.

(7) The wafer is then placed in hydrofluoric acid, which removes the silicon dioxide where there is no KPR film.

(8) A window is thus opened through the oxide, forming a mask for diffusion and alloying.

2.1.6 Precision Evaporation

The extreme thinness of the diffused layer created obstacles, which required a search for new methods to make alloyed electrical contacts to the layer without puncturing or otherwise destroying it. Another problem required a solution just as urgently as that of the thin base layer. For high-frequency performance these alloyed regions have to be as close together as possible -- without electrically shorting. This problem was solved by a vacuum-evaporation technique. The evaporation

technique allows the controlled deposition of thin layers of materials. It is also a method for depositing highly reactive materials, such as aluminum, that are difficult to work with in air. Further, it allows greater cleanliness in technique and hence gives a more intimate contact between the layer and the surface upon which it has been cooled. In addition, the fact that the evaporation beams travel in straight lines permits the use of precisely dimensioned masks to create similarly precise evaporated patterns. An alloying operation follows the evaporation of the gold antimony or aluminum contact.

2.1.7 Epitaxial Techniques

Epitaxy is a means of growing a very thin (3-25μ), uniformerly doped monocrystalline region on a relatively low-resistivity semiconductor substrate. The epitaxial process involves the decomposition of $SiCl_4$ with hydrogen at approximately 1200°C; the silicon that is freed is deposited on the basic silicon substrate. The deposition is perfectly oriented; the result is a layer of monocrystalline silicon with a uniform doping level at all depths. Subsequent diffusions are more predictable because of this uniform doping. The combination of planar diffusion techniques with epitaxial structures gives improved transistor characteristics. A typical planar-epitaxial transistor is shown in Figure 2-10.

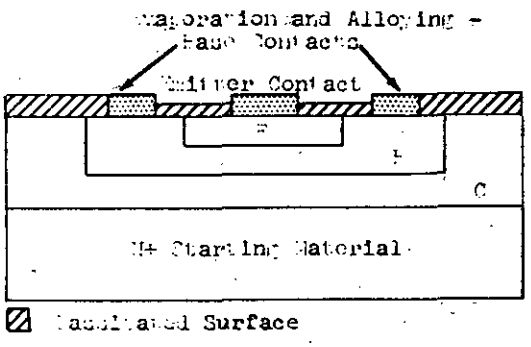


FIGURE 2-10
PLANAR-EPITAXIAL TRANSISTOR

As Figure 2-10 implies, there are several advantages of incorporating an epitaxial layer on the collector substrate. The epitaxial layer has comparatively higher resistivity than the starting material, so that base-collector junction capacitance is low. The series-collector bulk resistance remains low since the epitaxial layer is thin and the substrate is heavily doped. Thus the saturation resistance is low, which permits higher current flow for a given dissipation, and shorter collector-saturation time. Also, the collector-base breakdown is high because of the high resistivity of the epitaxial layer.

In one process, a mixture of hydrogen and silicon tetrachloride flows through a quartz tube into the reaction zone of the furnace. The flow path may be vertical or horizontal. Consider the horizontal case, shown in Figure 2-11. Energy is delivered to the reaction zone by external r-f coils that couple to a graphite "susceptor". The wafers lie on top of a quartz sleeve, which encases the susceptor, and they are thus protected from possible contamination by impurities from the graphite.

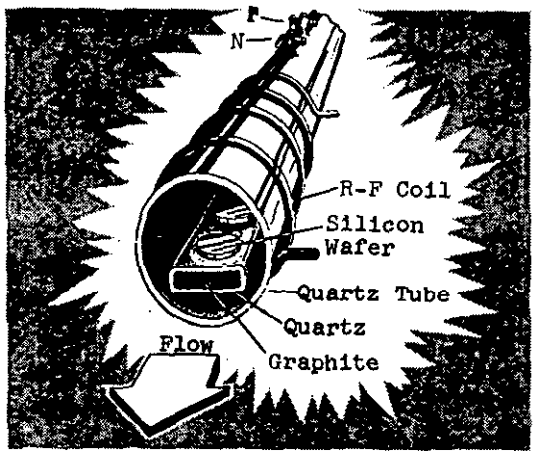


FIGURE 2-11

REACTION ZONE OF EPITAXIAL FURNACE

In the overall reaction that proceeds in the vicinity of the heated silicon wafers, hydrogen reduces silicon tetrachloride to form pure silicon and hydrogen chloride gas. Silicon evolved in the reaction is deposited on the silicon wafer and has a strong tendency to continue the crystallographic pattern embodied in the wafer. This continuation of structure is "epitaxy". While this picture of the process is imperfect with respect to detailed mechanisms, it is reasonable to assume that to maintain the crystal pattern a silicon atom on the growing surface must be given sufficient time to migrate to the proper crystallographic site or to leave the surface before it is trapped by other atoms in a wrong position. This qualitative picture is supported by the observation that disordered growth occurs if the growth rate is pushed above a certain value at a given temperature. Raising the temperature raises the tolerable growth rate. Since the rate of the deposition process can be controlled, and since it can be stopped at any time by shutting off the reactants, the thickness of the epitaxial layer can be controlled to within a few tenths of a micron.

It is also possible to adjust impurity doping precisely. Again, a number of methods can be used. An impurity compound can be mixed in a small quantity with the silicon tetrachloride; this mixture can then be reduced in the reaction zone, and the impurity atoms will be included in the resultant crystal. Alternatively, separate vessels can be provided for the main silicon tetrachloride supply, for a p-doped supply, and for an n-doped supply, giving separate n and p ports, as suggested in Figure 2-11. As in the main supply, a carrier gas can be sent through the doped vessels. Also, diffusion can be employed as the delivery mechanism. In the latter, it is possible to control vapor pressure by adjusting temperature. Still another approach employs gaseous impurity compounds diluted in a carrier gas.

2.1.8 Isolation

The isolation between components on a monolithic substrate is usually accomplished by reverse-biased p-n junctions. Associated with the p-n junction are capacitance and normal leakage currents. Therefore, complete isolation is not accomplished, and the effect is less-than-optimum performance.

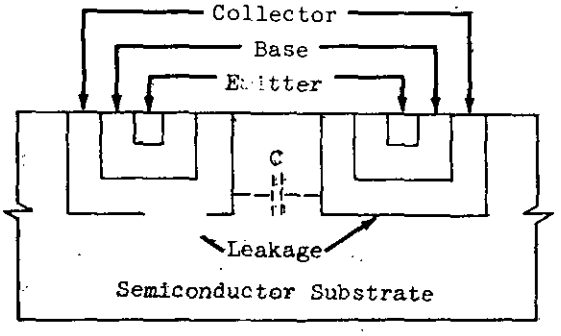


FIGURE 2-12
CAPACITANCE AND LEAKAGE IN P-N JUNCTION ISOLATION

As indicated in Figure 2-12, there is capacitive coupling between the collectors of the two transistors because of the capacitance associated with the junctions. At normal doping levels, the value is about 0.1 pf/sq mil at 0.5 volts of reverse bias. In addition, the leakage current is a function of temperature and may be appreciable at high temperatures. It would be best to isolate the devices from each other by a dielectric and obtain isolation more closely related to discrete components. A number of techniques have been developed to accomplish this. Some manufacturers are also developing methods for growing single-crystal silicon on an insulating substrate.

A number of monolithic isolation techniques* are currently in use. One of the earliest** to be developed is illustrated in Figure 2-13. The starting material is an n-type silicon substrate with a heavily doped epitaxial layer, as shown in (A).

Grooves are etched in the substrate according to some predetermined layout. These grooves are several mils deep. An oxide is either grown or vapor-deposited over the entire surface of the substrate. This oxide is SiO_2 and will perform the isolation function (B). Polycrystalline silicon is epitaxially deposited over the oxide (C). The single-crystal substrate is lapped to a point where single-crystal islands are isolated from one another by the oxide and imbedded in polycrystalline silicon (D). The sole purpose of the polycrystalline silicon is to hold the single-crystal islands together. These islands can now be doped and the devices fabricated.

One difficulty with this technique is the degree of control required in the lapping operation. The thickness of the n region is important in device fabrication, and the necessary lapping tolerance is in the order of 1 to 2 microns.

A number of techniques have been proposed for overcoming this difficulty†. One is illustrated in Figure 2-14. In this method a highly doped, uniform single-crystal substrate, as illustrated in (A), is processed in exactly the same

*J. W. Lathrop, "The Status of Monolithic and Thin Film Circuits", Electronic Industries, June 1965, p. 38.

**D. McWilliams, C. Fa, G. Larchian, and O. Maxwell, "A New Dielectric Isolation Technique for High-Quality Silicon Integrated Circuits," Journal Electrochemical Society, Vol. 111, No. 7. p. 153C, 1964.

†N. Schwartz, "Reactive Sputtering", Tenth National Vacuum Symposium, American Vacuum Society, (New York, 1963).

COURTESY OF ELECTRONIC INDUSTRIES

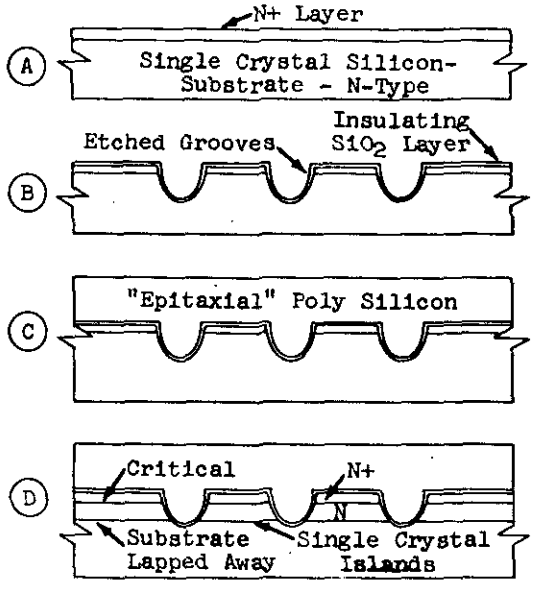


FIGURE 2-13

DIELECTRIC ISOLATION (1)

COURTESY OF ELECTRONIC INDUSTRIES

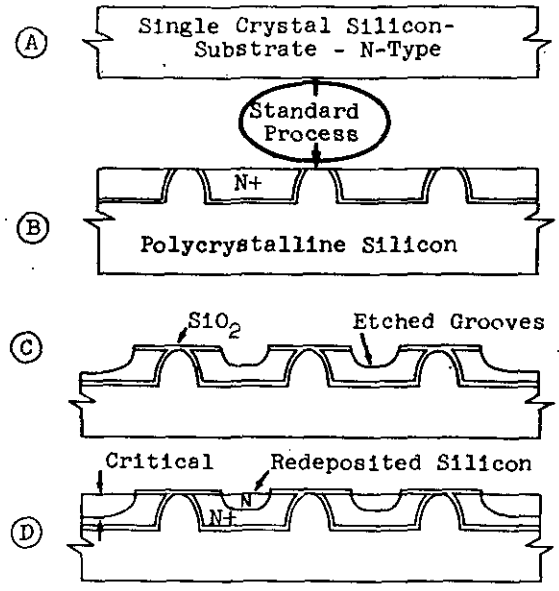


FIGURE 2-14

DIELECTRIC ISOLATION (3)

way as in Figure 2-13 to yield isolated islands of low-resistivity material (B). By use of a SiO_2 film for selective masking, regions within these islands can be vapor-etched (C) and lower resistivity silicon redeposited, as in (D). The critical distance is now controlled by the etch and deposit processes of epitaxial deposition, which are superior to mechanical lapping.

Another method* is shown in Figure 2-15. The critical distance is controlled during the initial step of epitaxially depositing an n-film on an n+ substrate,

COURTESY OF ELECTRONIC INDUSTRIES

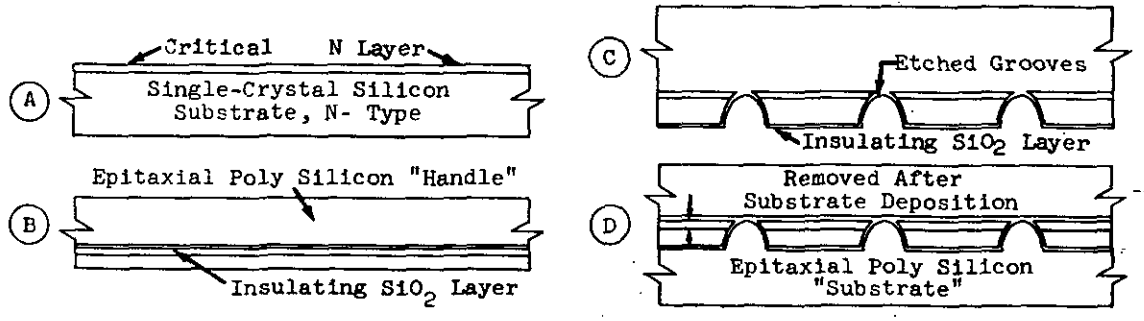


FIGURE 2-15

DIELECTRIC ISOLATION (2)

* G.L. Schnable and A.F. McKelvey, "A Technique for Preparing Oxide-Isolated Silicon Wafers for Microcircuits", Electron Devices Meeting, Washington, D.C., (October 1964).

as shown in (A). Next, an insulating SiO_2 layer is deposited over the n-layer, and polycrystalline material over this. The only purpose of this polycrystalline layer is to serve as a handle during operations and allow the substrate to be thinned to a few mils, as shown in (B), without breakage. Grooves are then etched in this thinned single-crystal material, creating mesas over which an oxide can be grown as in (C). Finally, an epitaxial polycrystalline substrate is deposited over thin oxide and the handle material removed to leave isolated regions with a common flat surface, as shown in (D).

The processes of Figures 2-14 and 2-15 minimize the lapping control for dielectric isolation, but at the expense of additional processing. The p-type, as well as n-type, material can be processed, and both types of islands can be formed on the same slice.

The result of these methods is isolated islands of semiconductor material into which impurities can be diffused to form active and passive components.

A circuit fabrication procedure* combining oxide isolation with other techniques -- which include localized epitaxial growth, localized etching and backfilling, and localized gold doping -- is described below. The process steps are illustrated in Figure 2-16.

(1) An epitaxial substrate wafer is prepared by conventional processing techniques with an n+ layer on one surface. The starting resistivity is chosen to be that required for one or more of the finished circuit elements.

(2) Depressions are etched into the back of this wafer in locations corresponding to regions where the required conductivity type and impurity density will be different from those of the region selected in (1).

(3) Each depression is filled with epitaxial silicon that has the impurity density and type required for additional devices.

(4) Grooves are etched into the silicon around each of the desired regions from the back side of the wafer.

(5) In a two-step process, metal and oxide layers are deposited.

(6) Polycrystalline silicon is then grown over the entire back side of the wafer to a thickness approximating that of the original wafer.

(7) The top surface of the wafer is then lapped or etched down to expose each of the desired isolated regions.

(8) The wafer is then processed conventionally to form each of the desired devices.

*C.G. Thornton, "The Application of New Metallurigcal Technique to Silicon Integrated Circuits," National Electronics Conference, (Chicago, 1964).

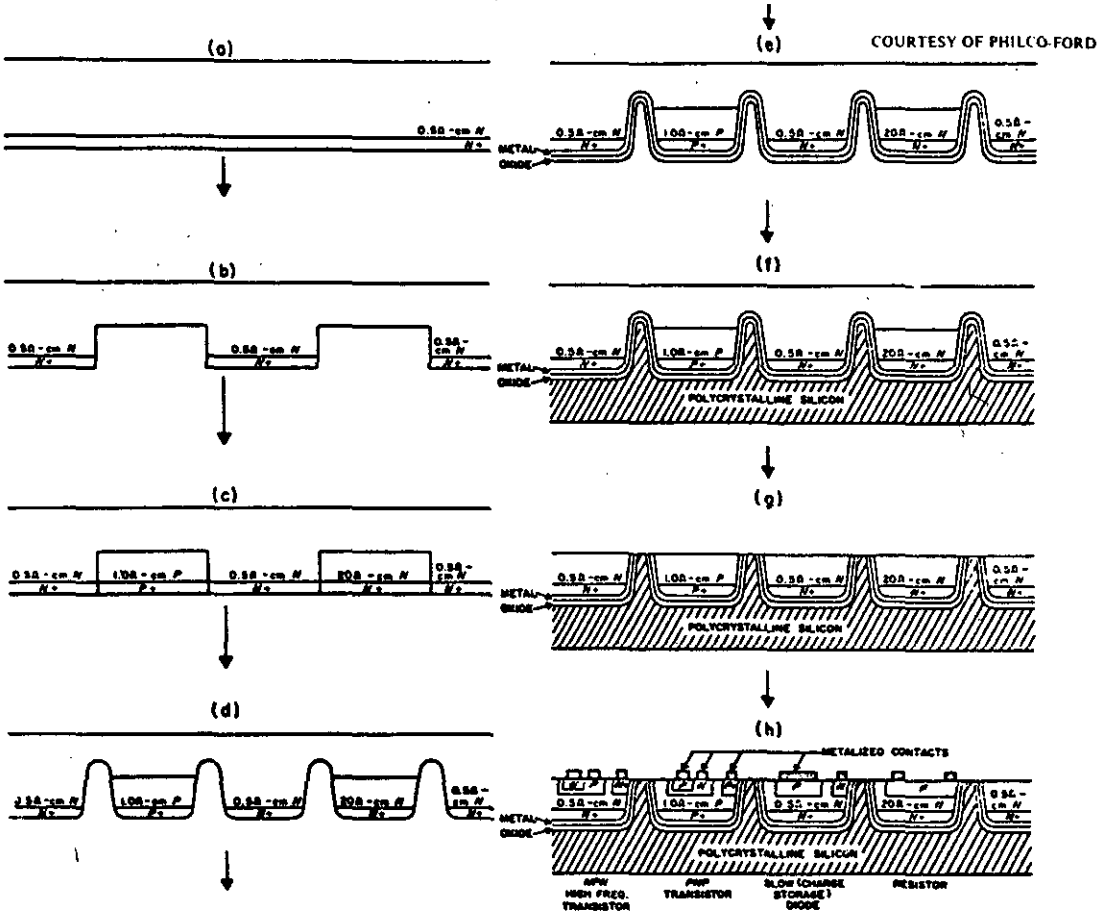


FIGURE 2-16

FLOW CHART FOR OPTIMIZED MICROCIRCUIT PROCESS

Before final metalization, gold is deposited through openings in the oxide in those regions where it is desired to reduce lifetime. A short heat treatment (approximately 1000°C for 15 minutes) is sufficient to distribute the gold throughout the desired regions.

The capacitance associated with oxide isolation is in the range of 0.02 pf per square mil per micron of oxide thickness. Thus a 5-micron-thick oxide reduces the capacitance coupling by a factor of 25 below that of p-n junctions that have a slight reverse bias.

2.1.9 Specifying the Circuit

The design engineer can follow one of two approaches in microelectronic circuit design. In one approach he will keep current with and be experienced in the state of the art of monolithic-circuit fabrication. He will then be in a position to give the circuit manufacturer detailed specifications on all components required in his circuit. The other approach is to provide the manufacturer with complete "black box" terminal specifications to which the final circuit must be manufactured and guaranteed.

Basic to both of these approaches is an understanding of the limits or boundary conditions that put constraints on the circuit design. One such constraint involves the fabrication technology itself. What types of devices can be fabricated on the same substrate? Table 2-1 is a monolithic compatibility chart that illustrates the possible combinations of devices on a single chip. It should be understood that the more complicated the circuit the more costly it will be to fabricate.

TABLE 2-1

MONOLITHIC COMPATIBILITY CHART

Devices	Signal		MOS-FET	Junction FET	Power Transistor	Matched Pairs*	Signal Diodes	Complementary
	NPN	PNP						
Signal NPN	-	Y	Y	Y	N	Y	Y	Y
Signal PNP	Y	-	Y	Y	N	Y	Y	Y
MOS-FET	Y	Y	-	Y	N	D	Y	D
Junction FET	Y	Y	Y	-	N	D	Y	N
Power Transistor	N	N	N	N	-	Y	Y	N
Signal Diodes	Y	Y	Y	Y	Y	Y	-	Y

Y - Yes
N - No
D - Difficult

*Useful matching is achievable for most devices; close matching of gain is difficult with any of the devices.

Microcircuits are fabricated in a variety of ways. Monolithic, hybrid, thin film, and thick film are some of the approaches. Generally, only monolithic silicon integrated circuits and some variation or hybrid of the monolithic form are considered here. The sheet resistance of monolithic circuits is limited to that associated with the base and emitter diffusions. This is necessary from a fabrication standpoint. The passive components are formed at the same time the transistor base or emitter is fabricated.

The hybrid offers a wider variation in sheet resistance than is possible with diffusion only. Thin-film resistors are available for use with monolithic circuits in a wide range of sheet resistances. Diffused resistors above 300-400 ohms per square are not recommended for close tolerances, because of the change in value with the temperature associated with lead-bonding operations. Ranges of resistor values are given in Table 2-6 at the end of this chapter.

Capacitors are either silicon-oxide (nonpolar with relatively high Q at high frequencies) for bypass and high-frequence tuning or junction type (low Q due to effective series resistance of the top contact) for bypass and voltage tuning. Typical characteristics are given in Table 2-7 at the end of this chapter.

The yield and, therefore, the cost of a circuit is more dependent on the area the circuit requires than on the number of components. In this respect it is wise to keep the circuit area to a minimum consistent with other requirements, such as component value and power requirements.

Area efficiency is a key factor with regard to capacitance. Large capacitors (500 to 10,000 pf) become progressively more expensive because of the area requirement. Values above 10,000 pf are not practical.

Though state-of-the-art monolithic transistors impose certain constraints, geometry nevertheless can be varied, and many combinations of transistor characteristics are possible. Typical characteristics of monolithic transistors are given in Table 2-2.

TABLE 2-2**

MONOLITHIC TRANSISTOR CHARACTERISTICS

Characteristics	Emitter Dimensions In Mils			
	1x1-1/2	3x3	2x10	2x20
BV_{CBO}	← 35 Volts →			
BV_{CEO}	← 15 Volts →			
BV_{EBO}	← 7 Volts →			
$h_{FE/IC}$	40/100μA	40/1mA	40/2.5mA	25/2.5mA
$h_{FE/IC}$	80/5mA	60/10mA	50/50mA	50/50mA
f_T (MHz)	550	220	--	--
Collector-Base C (pf) at 6V	2	7	10	16.3
Emitter-Base C (pf) at 6V	2	5	10	16
Collector-Substrate C (pf) at 6V	4	10.5	10	17
R_{CS}(ohms)	70	120	35	27

Another constraint is imposed upon the designer by the fact that the inductance function as such is not available in the monolithic technology. If inductance is required, it is necessary to synthesize or design around it.*

2.1.10 Linear Amplifier Checklist**

The nine-point checklist given below is a handy guide for developing a conventional set of performance specifications for a final integrated circuit.

(1) Establish circuit specifications (specify black-box functions, establish environmental conditions† and restrictions imposed on the design). A black-box specification is illustrated in Table 2-3. It provides the guide for fabricating the circuit. The more complete and realistic the specification, the simpler it is to begin fabrication. For the design engineer who is not going to fabricate the circuit, this specification is all that is required providing he has not violated the constraints imposed by the technology.

* V. Uzunoglu, "Six Possible Routes to Noninductive Tuned Circuitry", Electronics, (15 November 1965), p. 114.

** D.C. Bailey, "Converting Amplifiers to Integrated Circuit Format". EEE, (February 1964), p. 70.

† D. C. Bailey, "Black-Boxing Your Linear Integrated Circuit", Electronic Design, (12 June 1964), p. 74.

COURTESY OF ELECTRONIC DESIGN

TABLE 2-3

BLACK-BOX CHECKLIST

Environmental
- Operating-temperature range
- Storage-temperature range

Gain Specifications
- Frequency
- Source impedance
- Load impedance
- Power supply tolerance
- Open loop or gain setting
- Voltage gain, V_o/V_{in}
- Transconductance, $\Delta I_o/\Delta V_{in}$
- Current Gain, I_o/I_n
- Transresistance, $\Delta V_{in}/\Delta I_o$
- Gain stability (gain vs temperature)
- Gain linearity (gain vs signal level)

Bandwidth
- Source impedance
- Load impedance
- Power supply tolerance
- 3, 6, or 0 db points
- Open or closed loop (if closed loop, what gain)
- Frequency response
- Gain-bandwidth product
- f_t
- Cutoff frequencies
- Gain margin
- Phase margin
- Slope of gain vs. frequency characteristics

Stability
- Temperature range
- Input-output conditions (capacitances and resistances)
- Expected life
- Maximum equivalent input drift (Voltage and/or current)
- D-c stability
- A-c stability
- Phase margin with maximum feedback
- Maximum output capacitance

Noise
- Frequency and bandwidth
- Source Resistance
- Noise figure (or equivalent input noise voltage)

Maximum Output (dc)
- Power supply voltage
- Load impedance
- Minimum linear output voltage
- Minimum linear output current
- Maximum output impedance

Input (dc)
- Minimum input impedance
- Differential mode
- Common mode
- Maximum common-mode voltage
- Minimum common-mode rejection ratio (specify frequency and common-mode voltage swing)

Dynamic Range (ac)
- Power supply voltage
- Load impedance
- Source impedance
- Maximum input before clipping
- Minimum unclipped output
- Dynamic range of input signal
- Minimum power output

Power Supply
- Output voltages
- Tolerances
- Ripple and noise
- Impedance vs. frequency
- Output power

Package
- Form factor (TO-5, flatpack)
- Environmental requirements
- Salt spray
- Leakage
- Linear acceleration
- Shock

COURTESY OF EEE CIRCUIT DESIGN

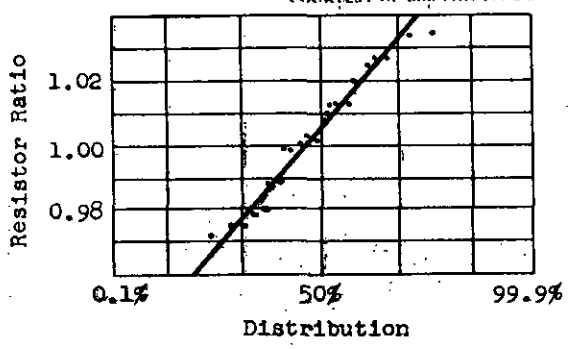


FIGURE 2-17
DISTRIBUTION OF RESISTOR RATIOS ON A MONOLITHIC STRUCTURE (SHEET RESISTANCE: 200 Ω/sq.)

COURTESY OF EEE CIRCUIT DESIGN

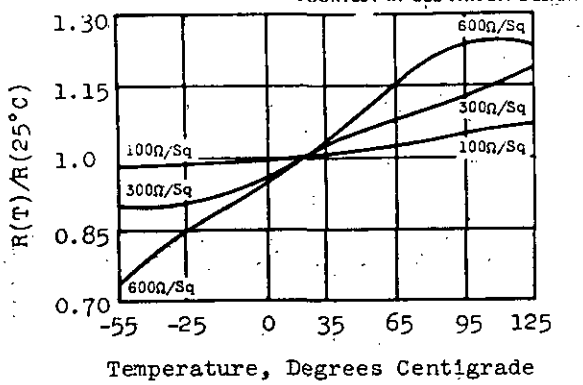


FIGURE 2-18
PERCENTAGE CHANGE OF RESISTANCE WITH TEMPERATURE FOR DIFFUSED RESISTORS WITH THREE VALUES OF SHEET RESISTANCE

COURTESY OF EEE CIRCUIT DESIGN

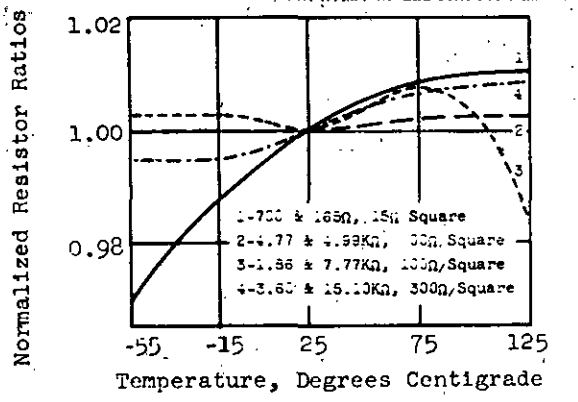


FIGURE 2-19
TRACKING ERROR OF DIFFUSED-RESISTOR RATIO FOR THREE DIFFERENT SHEET RESISTANCES (3 SEPARATE TAPPED CHIPS)

(2) Synthesize circuit configurations within the boundaries fixed by integrated-circuit technology. The active devices that can be used are found in Table 2-1. If inductance is required, a method must be determined to realize this function. Will an all-monolithic form be used or will hybrid techniques be required to give optimum performance?

(3) Establish an analysis method that utilizes measurable quantities in the final integrated circuit.

(4) Perform d-c analysis of the circuit. After the circuit has been synthesized, it must be analyzed from the standpoint of d-c stability. Much of this can be done by statistical methods, since the data on integrated-circuit elements are generally available in that form. The component values and their tolerances must be specified.

Figures 2-17, 2-18, and 2-19 are useful for d-c analysis. Resistor tolerances are normally wide and expensive to control or adjust to close values. However, resistance ratios are easily controlled and, if proper design is used, are frequently more important than the absolute values.

Integrated circuits have distributed parameters and parasitics associated with them. If possible, these characteristics should be employed as useful circuit elements.

(5) Breadboard the circuit, using discrete components. The circuit performance is compared with the

black-box specifications for performance. If the required performance levels are not achieved, the circuit can be modified or a different circuit approach can be synthesized. In some cases it may be necessary to repeat these steps several times to achieve the desired results.

(6) Draw a mask layout and compute the parasitics. These include resistance and capacitance interconnection parasitics, isolation parasitics, and substrate resistance. The calculated parasitic elements are inserted into the original breadboard in the form of discrete-equivalent lumped resistors and capacitors. Again circuit performance is measured and compared with blackbox specifications.

(7) Breadboard the circuit with integrated components. Compare the results with the black-box specifications. At this point it should be possible to determine if the specifications can be met with this particular design, although modifications may still be necessary before optimum results are achieved.

(8) Submit mask drawing for mask preparation. Include pin connections, lands, and spacing.

(9) Fabricate sample circuits and determine if they meet the required black-box specifications. Now the design can be optimized for the black-box specifications. Sometimes it becomes necessary to go through several process runs to achieve the desired distribution of electrical properties that best meet the original specifications.

2.1.11 Applications

Circuit analysis plays a major role in the design approach outlined above. The applications to follow use two techniques that have been found satisfactory. For d-c amplifiers a technique developed by Middlebrook* is used, and for high-frequency circuits a method proposed by Linvill** (which characterizes the integrated circuit as a black box with admittance parameters) is used.

Optimum analysis techniques have yet to be developed. The small-signal "nonlinear model" is more accurate at high frequencies because the two-pole alpha is included. Below a megacycle, capacitive parasitics can usually be neglected since small-geometry transistors will be used and since the passive parasitics present a negligible impedance at these frequencies. Resistive parasitics such

* R.D. Middlebrook, Differential Amplifiers, (New York, 1963).

** J.G. Linvill and J.F. Gibbons, Transistors and Active Circuits, (New York, 1961).

as the series collector resistance cannot be neglected. At low frequencies, series collector resistance can be lumped into the normal T-equivalent or small-signal nonlinear model.

2.1.11.1 D-C Amplifiers

An ideal d-c amplifier would have zero output with zero input; provide constant gain regardless of time, temperature, and input-level variations; provide an output that is a magnified replica of the input; and have an infinite bandwidth. To approach the ideal the following are needed: low noise figure, low equivalent-input offset voltage and drift, good linearity and dynamic range, gain stability with variations in temperature, and wide bandwidth.

Little information is available on integrated-circuit-type transistors, but indications are that noise figures are slightly greater than those of discrete types made by the diffusion process. With temperature variations, the noise figure becomes masked by internal d-c drift of the transistor.

Silicon diffused transistors exhibit a temperature coefficient of base-to-emitter voltage of about 2.3 mV/°C. In a single-transistor (unilateral) input stage, this coefficient is multiplied by the entire amplifier gain; thus to use such an amplifier for d-c amplification, temperature would have to be held constant, or it would be necessary to calibrate the amplifier with temperature. A better circuit configuration is the differential amplifier (Figure 2-20). Through canceling of the drift of one transistor with an equal and opposite drift of another, d-c drift can be reduced by several orders of magnitude.

COURTESY OF EEE CIRCUIT DESIGN

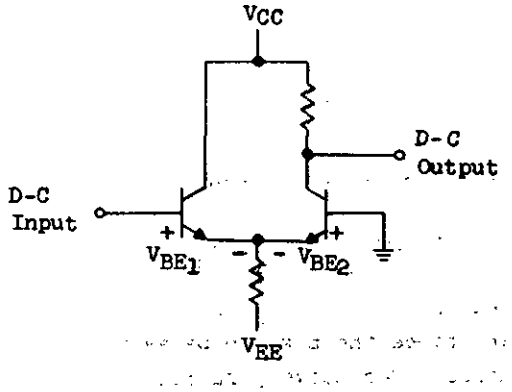


FIGURE 2-20
DIFFERENTIAL AMPLIFIER

A differential amplifier used as an input stage for a d-c amplifier helps reduce d-c drift. For a unilateral d-c amplifier, the performance of the discrete version and that of the integrated version are both subject to the d-c drift problem.

2.1.11.2 A-C Amplifiers

A well-known limitation of present integrated-circuit technology is its inability to produce inductors or large capacitors. Thus direct coupling should be used where possible. The following ground rules are recommended:

(1) Use a differential input stage.

(2) Design for as much d-c gain as can be tolerated with the specified amount of drift in the first stage.

(3) If gain is insufficient, add a second d-c amplifier stage, using an external coupling capacitor.

(4) Use overall feedback for gain stability.

(5) Obtain high impedance (FET's will provide this if it cannot be achieved otherwise) to permit the use of smaller coupling capacitors.

Figure 2-21 illustrates the application of these principles. The differential amplifier permits a gain of about 100 with sufficiently low drift to maintain a reasonable dynamic range. With the FET stage, a large external resistor is needed, but the small capacitor required can be integrated.

High a-c gain and good d-c stability are obtained in the circuit of Figure 2-22. Here a large amount of d-c feedback provides d-c stability, and a small amount of a-c feedback results in high a-c gain. The amount of a-c feedback is controlled by external capacitor C_1 and resistor R_1 (if R_1 is zero, a-c gain is maximized).

COURTESY OF EEE CIRCUIT DESIGN

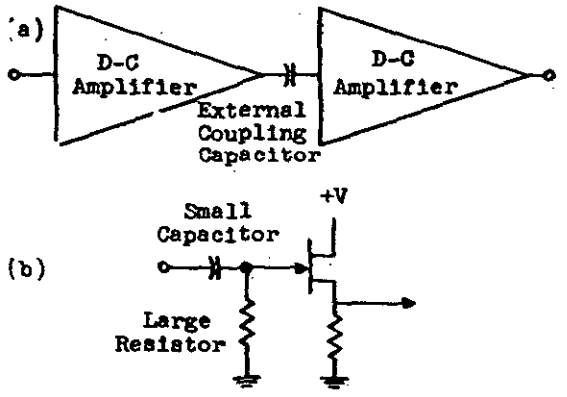


FIGURE 2-21
A-C AMPLIFIER TECHNIQUES

COURTESY OF EEE CIRCUIT DESIGN

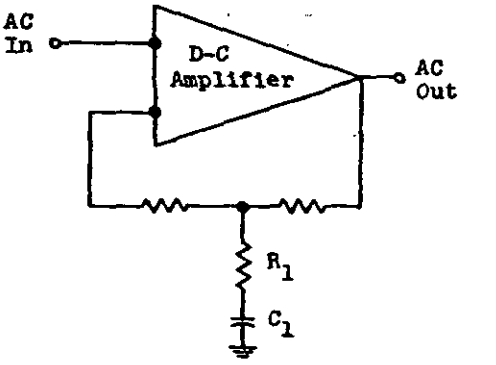


FIGURE 2-22
HIGH A-C GAIN, D-C STABLE CIRCUIT

2.1.11.3 Signal-Processing Circuits

A major goal in the servo-control field is the elimination of transformers. The integrated choppers, demodulators, and quadrature rejection circuits illustrated in this section should be examined with that in mind.

The two major factors in the design of choppers are offset current and voltage, and drift. (Offset causes an output in the absence of an input, and drift produces a change in output with time and temperature.) The inverter transistor connection reduces both offset and drift.

Figure 2-23 shows how even lower offsets can be achieved. Here the drift is determined by tracking of the transistor characteristics with temperature and time, and by the difference in junction temperature between the two units. Cancellation of offset is achieved by equal and opposite offset of another device. The integrated version has an obvious advantage over the discrete version since the transistors will be matched and operating with small differences in temperature.

COURTESY OF EEE CIRCUIT DESIGN

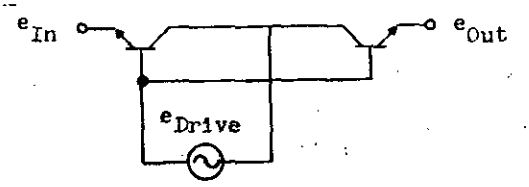


FIGURE 2-23
BASIC CHOPPER CIRCUIT

Offsets and drifts of standard integrated circuit pairs (Figure 2-24) in the inverted chopper connection average 50 μV offset and 1 μV/°C drift. Figure 2-24(a) shows two transistors diffused into a single die; and Figure 2-24(b) shows two emitters diffused into a single collector-base junction area, which provide even tighter control of drift and offset. Double-emitter devices have been reported to have offsets under 20 μV and drifts less than 0.2 μV/°C.

COURTESY OF EEE CIRCUIT DESIGN

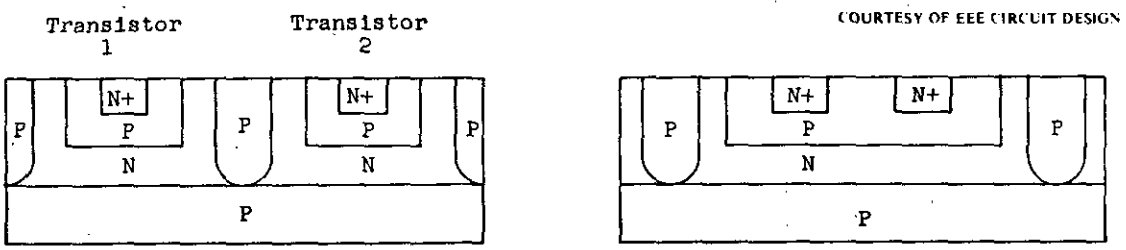


FIGURE 2-24
STANDARD CHOPPER CONFIGURATIONS

2.1.11.4 Demodulators

Although demodulators vary widely in configuration, only diode-quad types are discussed here.

In discrete diode-quad bilateral switches, the usual procedure is to select diodes in quads or pairs to achieve low offsets. Offsets from 10 mV down to 1 mV (with successively poorer yields) are typical. In integrated circuits, two diodes

can be diffused side by side into the same silicon substrate. An n-on-p diffused pair can be connected to a p-on-n pair to form a bridge (Figure 2-25). The p-on-n

COURTESY OF EEE CIRCUIT DESIGN

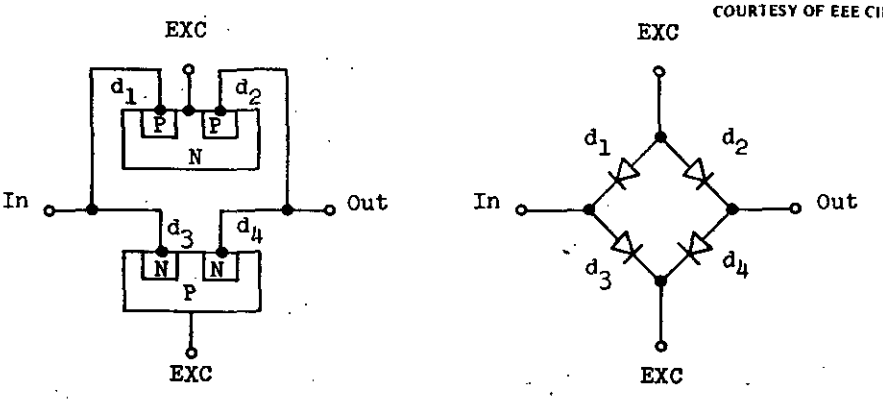


FIGURE 2-25

DIODE-QUAD SWITCH CONSTRUCTION

pair should match and track, as should the n-on-p pair. Figure 2-26 shows a diode quad in a demodulator. Quadrature rejection is accomplished by controlling the conduction angle. The d-c output is the average value of the in-phase sine wave. At that time the quadrature signal is passing through zero. The output will be zero for quadrature signals and nearly equal to the peak magnitude for in-phase (o°) signals. A paralleled R and C is used to achieve a peak-conducting circuit.

COURTESY OF EEE CIRCUIT DESIGN

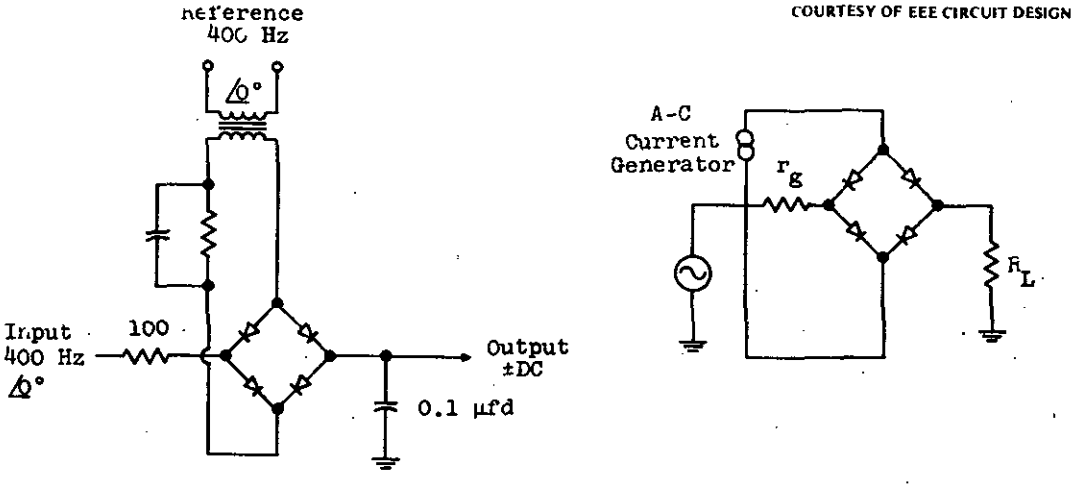


FIGURE 2-26

DIODE-QUAD DEMODULATOR AND EQUIVALENT CIRCUIT

A transformerless version of the demodulator is shown in Figure 2-27. In this version, the isolation property of the transformer is approached by use of the high output-impedance of the common-base connection. When Q_1 and Q_2 are cut off, the impedance is essentially infinite. When Q_1 and Q_2 conduct, the drive

nodes of the bridge are connected to the current generators with the output impedance of Q_1 and Q_2 across them (it is above 500 KΩ at low frequencies).

COURTESY OF EEE CIRCUIT DESIGN

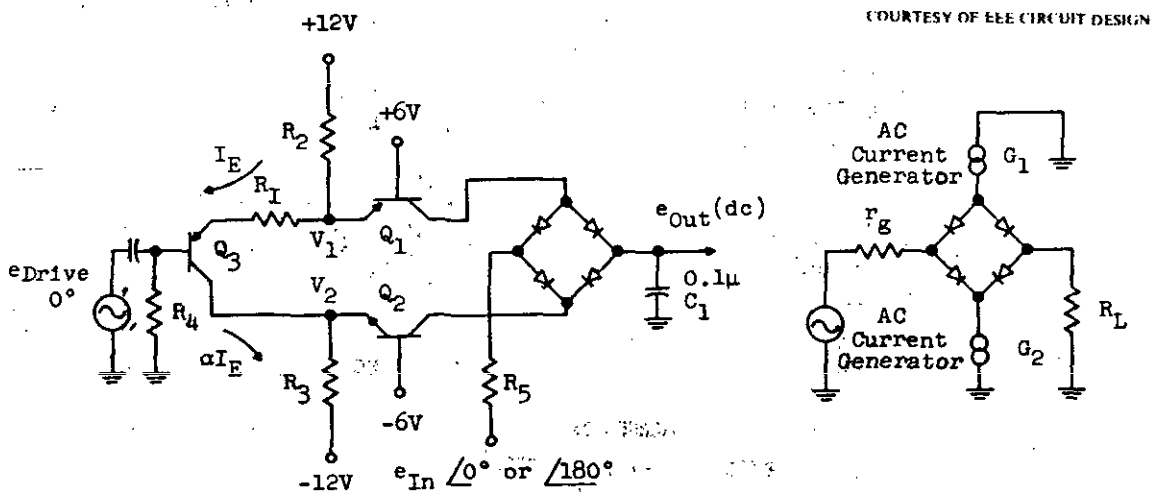


FIGURE 2-27

TRANSFORMERLESS VERSION OF DEMODULATOR SHOWN IN FIGURE 2-26

The demodulator functions of the transformer and transformerless circuits are essentially the same. In the transformerless version, the bridge conducts for 30 to 50 degrees about the peaks of the cycle, causing C_1 to charge up to the average value of the peaks of the input signal through R_5. When the drive circuit cuts off, C_1 holds the charge until the next half cycle. The output is a bipolar d-c voltage, corresponding to the zero and 180-degree input signals. Another transformerless demodulator is shown in Figure 2-28.

COURTESY OF EEE CIRCUIT DESIGN

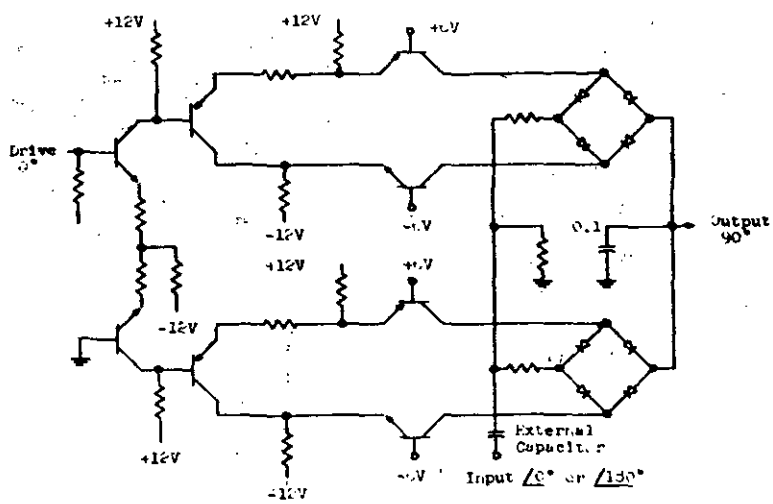


FIGURE 2-28

TRANSFORMERLESS QUADRATURE REJECTION DEMODULATOR

2.1.11.5 Communication Circuits

R-f and i-f amplifiers, mixers, oscillators, and multipliers with good performances over a wide range of frequencies can be designed and fabricated in integrated-circuit form. Monolithic integrated circuits can often match the performance of discrete versions up to about 10 MHz. Above this, parasitics begin to play a significant role.

For monolithic r-f circuits, it is recommended that impedances be kept low or that resistive isolation techniques be used. Vhf and uhf circuits are not yet practical in monolithic form. At these frequencies it is often desirable to separate the individual circuit elements on a high-frequency dielectric (e.g., ceramic), as in hybrid circuits.

COURTESY OF EEE CIRCUIT DESIGN

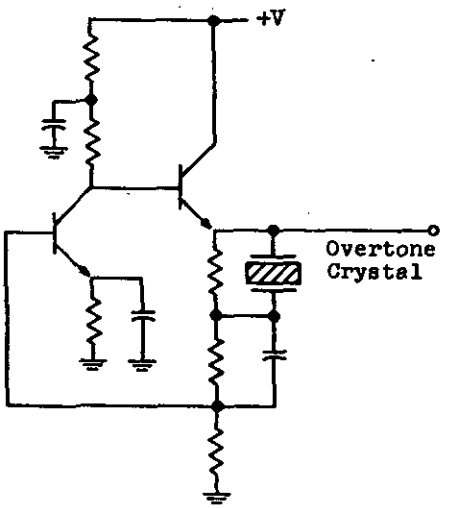


FIGURE 2-29
OVERTONE CRYSTAL OSCILLATOR

Again, it is desirable to eliminate chokes and transformers. Figure 2-29 shows an overtone crystal oscillator without inductors. It can provide a frequency stability better than 0.005% from 0°C to 50°C. Naturally, the crystal is not included in this monolithic integrated circuit.

2.1.11.6 Digital Circuits

While design of digital circuits is similar to that of analog circuits, there are differences in the particulars of the design. Some of the more important digital characteristic are as follows:

(1) fan-in, (2) fan-out, (3) noise immunity (4) propagation delay, and (5) best logic type for a particular application. Design factors that must be considered before determining the desired diffusion profiles and mask layout include the following: (1) reliability, (2) fan-out, (3) temperature range, (4) switching speed, (5) noise immunity, (6) power dissipation, and (7) production yield. Often a design that is best for one parameter acts to degrade another; for example, high switching speeds are not compatible with large fan-out and low power.

In preparing the specifications for the circuit it is necessary to determine the following:

(1) The type of transistor required, which includes setting the diffusion profiles and determining the geometrical layout

(2) The center values of the resistor

(3) The tolerance on transistor and resistor parameters

To determine these values it is necessary to state the basic black-box objectives of the digital circuit.

The following example* will serve to illustrate some of the important design considerations for digital circuits. The electrical equivalent circuit for this example is illustrated in Figure 2-30. It is a three-input, modified direct-coupled transistor-logic (DCTL NOR gate). Logic 0 is approximately +0.2V, and logic 1, +0.8V. A pulse at inputs 1, 3, or 5, or any combination of them, will drive one or more of the transistors into saturation and produce a 0 at output pin 7.

COURTESY OF ELECTRONIC DESIGN

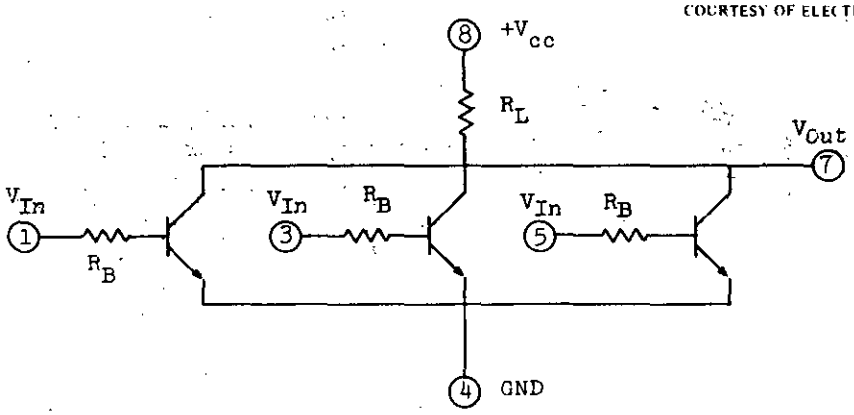


FIGURE 2-30
MODIFIED DCTL NOR GATE

External Black-Box Parameters		Internal "Device" Parameters	
Fan-in	≈ 3	Load Resistance R_L	≈ 600 ohms
Fan-out	≥ 5	Base Resistance R_B	≈ 400 ohms
Power Dissipation	≤ 15mW	Saturation Resistance	≈ 40 ohms
Operating Voltage	≈ 3 V	Current Gain h_{FE}	> 20
Noise Immunity	≥ 0.3V	h_{FE} Peak at I_c	≈ 5mA
Maximum Supply Voltage	≤ 10V	Frequency Cutoff f_T	≈ 500 MHz
Propagation Delay	≤ 25 nsec	Saturation Voltage V_{BE}	≈ 0.7V
Operating Temperature Range	-55 to +125C	Saturation Voltage V_{CE}	< 0.25V
		Base Sheet Resistance P_S	≈ 100 ohm/sq

*L. B. Valdes, "Case History: Integrating a NOR Circuit", Electronic Design, 2 March 1964.

The noise-immunity objective requires that when a transistor is locked in saturation, a 0.3-V signal superimposed on the output voltage will not turn on the transistor in the next circuit system (assuming that one NOR gate drives another). Ideally, the integrated circuit should be designed with V_{CE} sat as low as possible.

A minimum value of load resistance, R_L, can be obtained from the power-dissipation requirements; 15 mW at 3 V gives a maximum current consumption of 5 mA.

Thus

$$R_L \text{ min} \approx \frac{3.0 - V_{ce\ sat}}{0.005} \approx 530 \text{ ohms} \qquad (2\text{-}10)$$

If diffusion control limits the load resistance to ± 15%, the center value of R_L is set at 620 ohms. An upper value is approximately 720 ohms.

This calculation also indicates that in a typical case a transistor in saturation is at I_c = 5 mA. For maximum fan-out it would be desirable to have the transistor in saturation with the base current I_B as small as possible. Thus a transistor with a large value of h_{FE} is wanted. The curve of h_{FE} vs. I_c should also peak at I_c = 5 mA. From switching-speed considerations (maximum propagation delay of 25 nsec) it is also known that a fast transistor will be needed, one with an f_T of at least 500 MH_z.

Choosing the proper value of R_B is somewhat more complicated. If R_B is small, the transistors will draw too much base current from the previous stage. This not only limits the fan-out of each gate, but may lead to "current-hogging", which can be illustrated with the aid of Figure 2-31. Current available from Gate A is limited to a maximum value given as:

$$I_{avail} = \frac{V_{cc} - V_{in}}{R_L} \qquad (2\text{-}11)$$

This gate drives five others, which have slightly different input characteristics. Gate B1 has the lowest $V_{BE\ sat}$ of the group, and B5 has the highest. Thus it may happen that Gates B1 through B4 draw so much I_B that virtually all the available current is directed into these four gates and there is not enough left to turn on Gate B5; the first four gates have "hogged" the available current.

Most silicon transistors are in saturation at I_C/I_B = 10 or $I_B \approx 0.5$ mA. At this point, $V_{BE\ sat} \approx 0.7$ v, and it is evident that too much series voltage drop cannot be added across R_B. If this series voltage drop is arbitrarily

COURTESY OF ELECTRONIC DESIGN

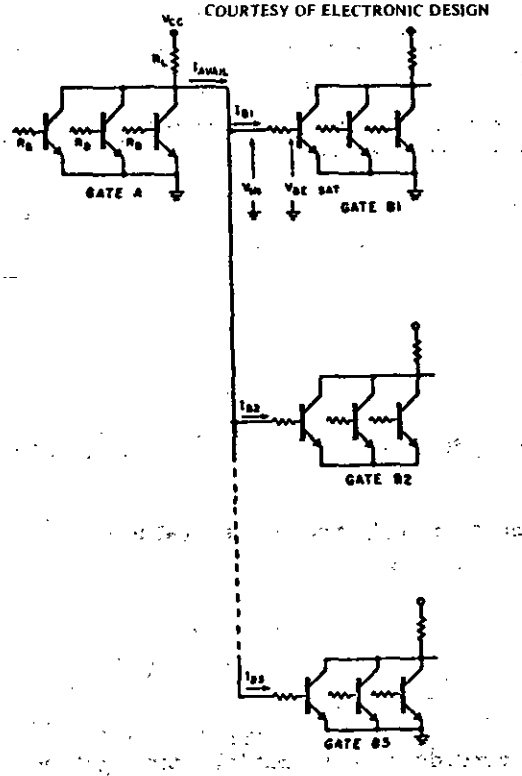


FIGURE 2-31
REPRESENTATION OF A NOR GATE DRIVING FIVE LOADS (GATES B1 TO B5)

chosen as 0.2 V, the result is R_B = 400 ohms. Figure 2-32 illustrates what happens if two transistors are paralleled, one with V_{BE} sat = 0.7 V and the other with V_{BE} sat = 0.8 V. The input current is only 0.25 mA to the second transistor, but it remains in saturation as long as h_{FE} is greater than 20.

The actual selection of R_B = 400 ohms was made after a worst-case analysis of the circuit. It is obviously not sufficient to consider the various requirements independently. It is necessary to examine what happens when all component parts of the integrated circuit are at their worst possible values, and to consider the variation of these parameters with temperature. In addition, the transient characteristics of the device must be examined, together with all the parasitic terms that are an integral part of every integrated circuit.

The desired electrical characteristics can be satisfied with a transistor having the geometrical layout of Figure 2-33. This unit has a base width of 0.8 μ and enough gold doping to produce a storage time of 20 nsec. It has been designed to minimize transverse voltage drop in the collector region and obtain a saturation resistance of less than 40 ohms.

COURTESY OF ELECTRONIC DESIGN

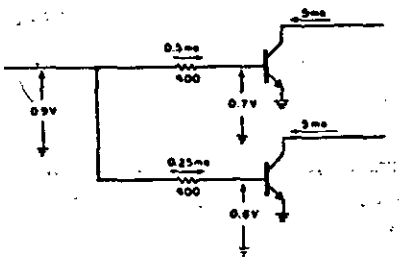


FIGURE 2-32
EFFECT OF DIFFERENT $V_{BE_{sat}}$ VALUES ON NOR-GATE CIRCUITS DRIVEN IN PARALLEL

COURTESY OF ELECTRONIC DESIGN

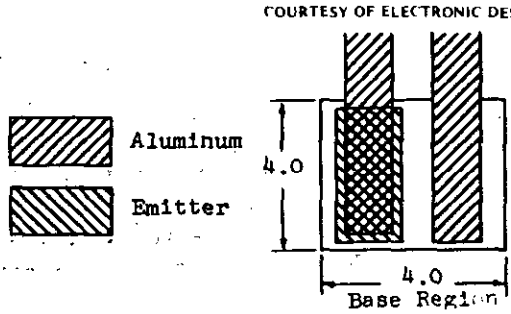


FIGURE 2-33
TOP VIEW OF EMITTER AND BASE REGION OF TRANSISTOR USED IN AN INTEGRATED CIRCUIT

A heavily doped n+ region has been diffused around three-fourths of the base region, reducing the possibility of channeling..

The load resistor R_L and input base resistors R_B are produced by use of the same p-type diffusions as for the base of the transistor. As explained earlier, their values are adjusted by controlling the ratio of length ℓ to width w of the resistor pattern and using the equation

$$R = \rho_s \frac{\ell}{w} \qquad (2\text{-}12)$$

where the sheet resistivity, ρ_s, is about 100 ohms/square.

The fact that the resistors are at a more positive potential in the circuit than the transistors requires that they be placed in separately isolated n-type regions. Figure 2-34 is a cross-section of the circuit, showing one transistor and one resistor.

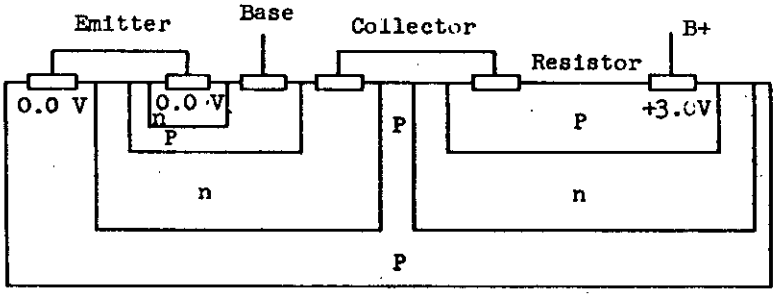


FIGURE 2-34
CROSS-SECTIONAL VIEW OF AN INTEGRATED CIRCUIT SHOWING ONLY ONE TRANSISTOR AND THE LOAD RESISTOR

The measured characteristics of two devices constructed from the foregoing information are given in Table 2-4. Two saturation voltage tests were performed: one with 0.825 V applied to the input, and a second with 1.5 V. The output voltage, V_o (for the light saturation condition), illustrates the ability of the transistor to turn on at a prescribed signal level.

The hard-saturation voltage condition provides a measure of the ability of the circuit to handle a large input signal. It is also used to ascertain that there are no parasitic elements limiting the operating of the device. For example, there may be a parasitic clamping diode between the input

terminal and the substrate, which would tend to lift the output voltage as the input voltage is increased.

The output current is measured with 0.55 V applied to all three transistors. This is a worst-case condition. It should be noted that if one of the transistors turns on too soon (at $V_{BE} = 0.55$), the current through R_L will be diverted through such a transistor rather than go through the output terminal and into the next circuit. Early turn-on of a transistor is also characterized by the value of I_{CEX}.

The fan-out of the NOR gate is theoretically the ratio of the minimum output current divided by the maximum input base current -- in this case, 2.9/0.37 = 8.

COURTESY OF ELECTRONIC DESIGN

TABLE 2-4

MEASURED PARAMETER VALUES FOR NOR CIRCUIT

Base Input Current I_B (0.825 V applied to input)	370	μA max
Output Voltage V_o (0.825 V applied to input)	300	mV max
Saturation Voltage V_B (1.5 V applied to input)	250	mV max
Output Current I_o (output at 0.825 V)	2.9	mA min
Threshold Current I_{CEX} (input at 0.55 V, output at 1 V)	200	μA max

COURTESY OF ELECTRONIC DESIGN

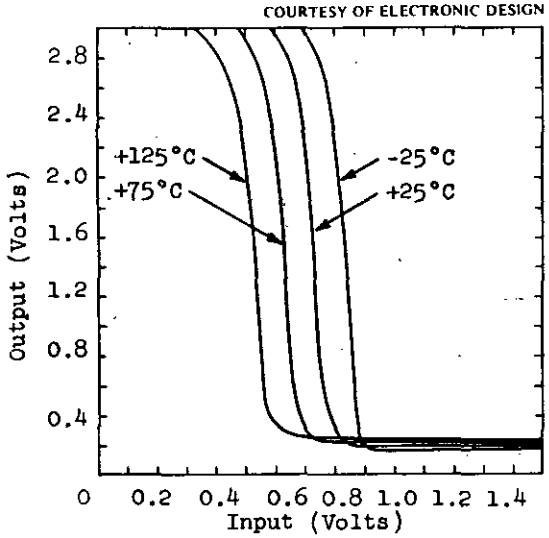


FIGURE 2-35
OUTPUT VOLTAGE VS INPUT VOLTAGE

Figure 2-35 shows the output voltage vs input voltage of a typical gate. These curves can be used to determine the noise immunity of the gate. Take the highest temperature (+125C curve), or worst-case, condition. The gate turns on at approximately $V_{BE} = 0.55$ V. If $V_{CE_{sat}} < 0.25V$, the required 0.3-V noise immunity is achieved at maximum operating temperature.

The temperature variations of some of the other significant gate parameters are illustrated in Figures 2-36 through 2-38. The propagation delay is defined in the test circuit illustrated in Figure 2-39; it is the average of the turn-off time t_1 and the turn-on time t_2. Both of these are measured at the 50% point in the input waveform A and the output waveform B.

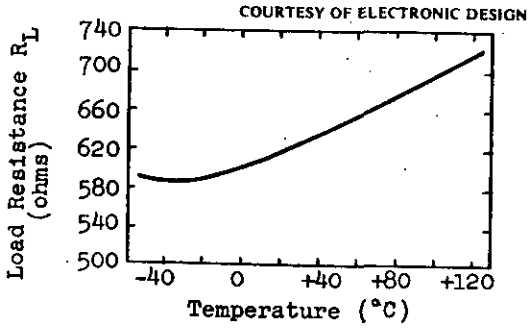


FIGURE 2-36
EFFECT OF TEMPERATURE
ON THE LOAD RESISTANCE R_l

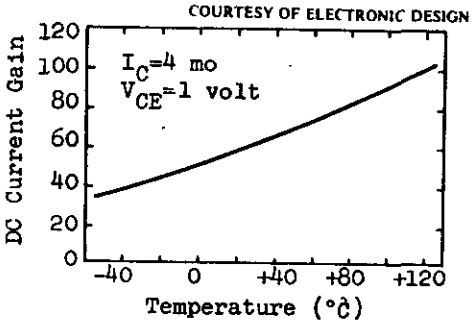


FIGURE 2-37
EFFECT OF TEMPERATURE ON
THE CURRENT GAIN OF AN
INTEGRATED-CIRCUIT TRANSISTOR

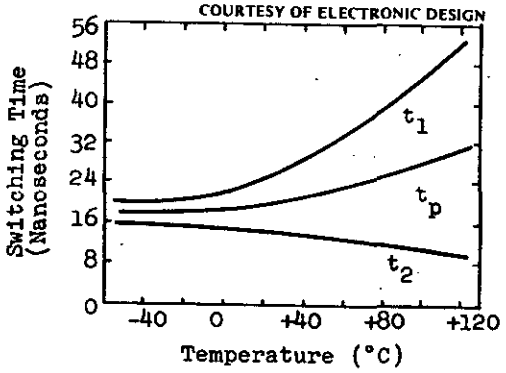


FIGURE 2-38
SWITCHING TIMES t_1 AND t_2 AND
PROPAGATION DELAY t_p

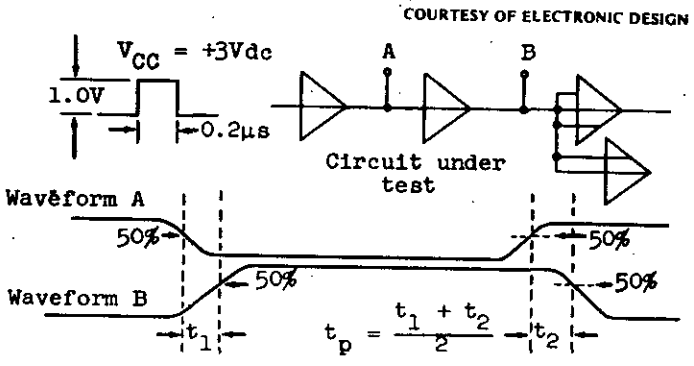


FIGURE 2-39
CIRCUIT USED FOR THE MEASUREMENT OF
PROPAGATION DELAY t_p

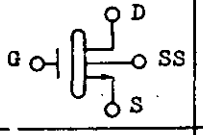
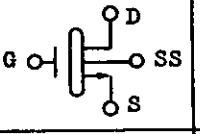
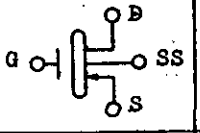
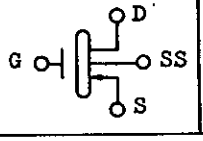
2.2 MOS FIELD-EFFECT DEVICES

The metal-oxide-semiconductor (MOS) field-effect transistor (FET) uses an electric field applied through an insulated gate electrode to modulate the conductance of a channel layer in the semiconductor under the gate electrode. The channel consists of a lightly doped region contained between two highly doped regions called the "source" and "drain". The conductuance of the channel is dependent upon the intensity of the field and hence the voltage applied to the gate electrode. Most MOS transistors now available are fabricated on a silicon substrate. The insulating oxide is silicon-dioxide. (Some results of using silicon-nitride as the insulator have been published.) The MOS has high input resistance ($>10^{15}$ ohms), and at high frequencies the input impedance becomes highly capacitive. Unlike the junction transistor, which is a current-controlled device, the MOS is voltage- or charge-controlled.

There are two basic types of MOS devices: enhancement and depletion. Each of these types can be fabricated so that either electrons or holes are the majority carriers. Thus four distinct types of MOS transistors exist. These are listed in Table 2-7, with associated bias-voltage requirements.

TABLE 2-5

CHARACTERISTICS OF FOUR BASIC TYPES OF MOS DEVICES

MOS Type	Symbol	Gate-to-Source Voltage to Cutoff	Incremental Gate-to-Source Voltage to Increase Conduction	Normal Drain-to-Source Voltage	Normal Substrate-to-Source Voltage
N-Channel Depletion		Negative	Positive	Positive	Zero or Negative
N-Channel Enhancement		Zero	Positive	Positive	Zero or Negative
P-Channel Depletion		Positive	Negative	Negative	Zero or Positive
P-Channel Enhancement		Zero	Negative	Negative	Zero or Positive

G - Gate S - Source D - Drain SS - Substrate

2.2.1 N-Channel Enhancement MOS

The majority carriers are electrons in the n-channel enhancement transistor. A positive voltage (with respect to the source) applied at the gate induces a positive potential in the oxide at the oxide-semiconductor surface. This in turn attracts electrons to the semiconductor surface, and these carriers form the transistor's induced channel. An enhancement-type transistor structure is shown in Figure 2-40.

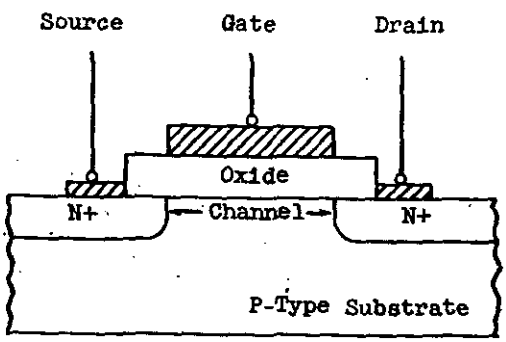


FIGURE 2-40

N-CHANNEL ENHANCEMENT MOS

At zero gate voltage there is only a very small current flow ($\approx 10^{-9}$ amps) between source and drain. Since there is no gate voltage, there are no free carriers in the space between the two N+ regions. The current that does flow is that of the reverse-biased junction formed by the P-type substrate and the N+ drain region.

When a positive voltage is applied to the gate electrode, minority carriers (electrons) are attracted from the P-region to the surface to form the channel. The surface has an excess of holes (empty valence states) since it is originally P-type, and as the electrons are drawn to the surface of the P-region underneath the oxide, these empty states are filled. Therefore, a gate voltage is reached at which there are just enough electrons to fill the empty valence states. At this value of gate voltage the channel is intrinsic. For further increases in gate voltage the channel becomes N-type. When this happens, the surface is referred to as "inverted." Figure 2-41 is an illustration of the drain current as a function of drain voltage for an enhancement field-effect transistor. Because of the nature of surface states, surfaces change easily from P-type to N-type, which makes it quite difficult to fabricate the enhancement FET; they have a tendency to become depletion-type devices. The surface resistivity must be kept low to prevent the channel from forming when there is no gate bias.

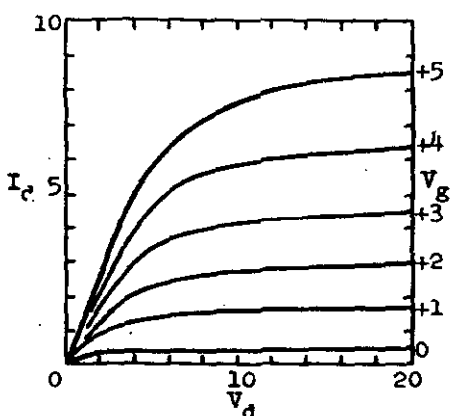


FIGURE 2-41

DRAIN CURRENT VS DRAIN VOLTAGE
Vertical: 1mA per division
Horizontal: 2V per division

2.2.2 P-Channel Enhancement MOS

The P-channel enhancement MOS utilizes holes as the majority carriers. The substrate is N-type, and the induced channel

is P-type. The drain and source regions are P+, and the drain-to-source voltage is negative. The channel is induced by application of a negative gate-to-source bias. The drain current as a function of drain and gate voltage is similar to that shown in Figure 2-41 except that all the voltages are reversed.

The metalized gate electrode shown in Figure 2-40 covers the entire channel region. It is characteristic of the enhancement-type transistor that the gate electrode overlaps both source and drain N+ regions. This is necessary to prevent a high resistance from appearing in series with the source and drain, since no carriers would be attracted into that part of the channel not included in the gate electric field.

This geometry, however, results in a large capacitance from gate to drain and gate to source. The exact value depends on the degree of overlap and the thickness of the oxide. The thicker the oxide the smaller the capacitance value; yet, for sensitivity, the oxide must be thin enough to provide high gate fields for reasonably small gate voltages. To meet these opposite demands, the oxide is usually made thin over the channel and thick over the N+ regions (P+ regions for P-channel devices), as shown in Figure 2-42. Also illustrated is the effect of gate voltage on carriers for both P- and N-types.

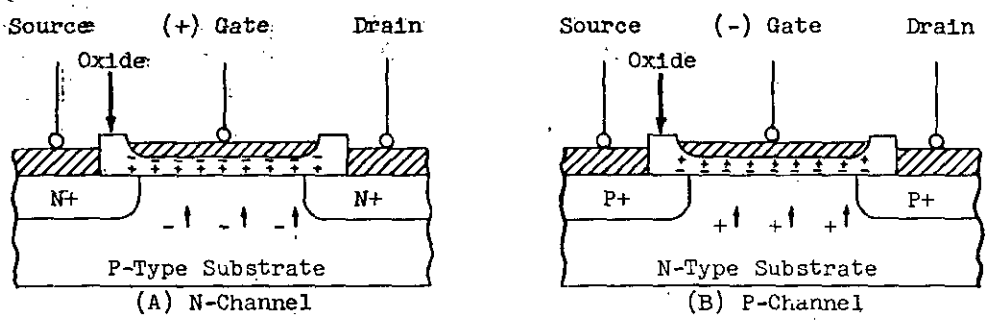


FIGURE 2-42

OXIDE THICKNESS AND CHARGE ON ENHANCEMENT MOS TRANSISTOR

2.2.3 N-Channel Depletion MOS

The depletion-type MOS is illustrated in Figure 2-43. The channel for this device is formed at the time when the oxide is fabricated on the surface. The degree of doping for the channel depends on the saturation current desired at zero gate voltage. The current that flows with zero gate voltage is higher in the depletion-type device since the channel exists at all voltages. The drain

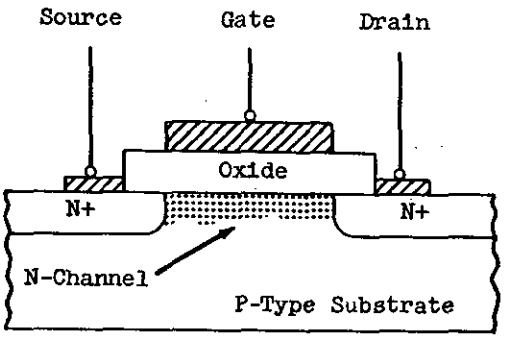


FIGURE 2-43

DEPLETION-TYPE MOS

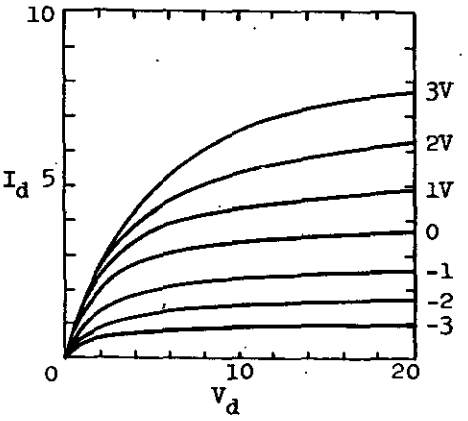


FIGURE 2-44

DRAIN CURRENT VS DRAIN VOLTAGE

Vertical: 2mA per division

Horizontal: 5V per division

current as a function of drain voltage is illustrated in Figure 2-44 for an N-channel depletion MOS transistor.

This device operates for negative as well as positive bias on the gate. It is unique in its applications and can be considered as operating either in the depletion mode or in the enhancement mode.

Operation in the depletion mode requires a negative gate-to-source bias. As the gate bias is made more negative, the channel is depleted of electrons because of the electric field induced in the oxide. The field at the oxide-semiconductor interface is positive and repels the electrons from the surface. The result is a decrease in current as a result of the decreasing conductance of the channel.

2.2.4 P-Channel Depletion MOS

The geometry for the P-channel is the same as that of the N-channel depletion MOS. The majority carriers, however, are holes. This device has all the characteristics of the N-channel except that all voltages are reversed.

The channel is not as highly doped as the drain and source regions. Because of the relatively high free-carrier concentration in the channel, it is not necessary that the gate electrode overlap the drain. This is referred to as "gate off-set". A small series resistance will be induced in the channel at the drain when the transistor is operated in the enhancement mode. The only effect this has on device operation is to increase the drain voltage at which current saturation occurs. The gate electrode does overlap the source. If it did not, the series resistance associated with the source would introduce degenerative feedback, which would have deleterious effects on device gain. The gate off-set significantly reduces the gate-to-drain feedback capacitance.

The two basic MOS structures are illustrated in Figure 2-45. These basic structures were first proposed by D. Hahng and M. M. Atalla.*

COURTESY OF IEEE TRANS ON ELECTRONIC DEVICES

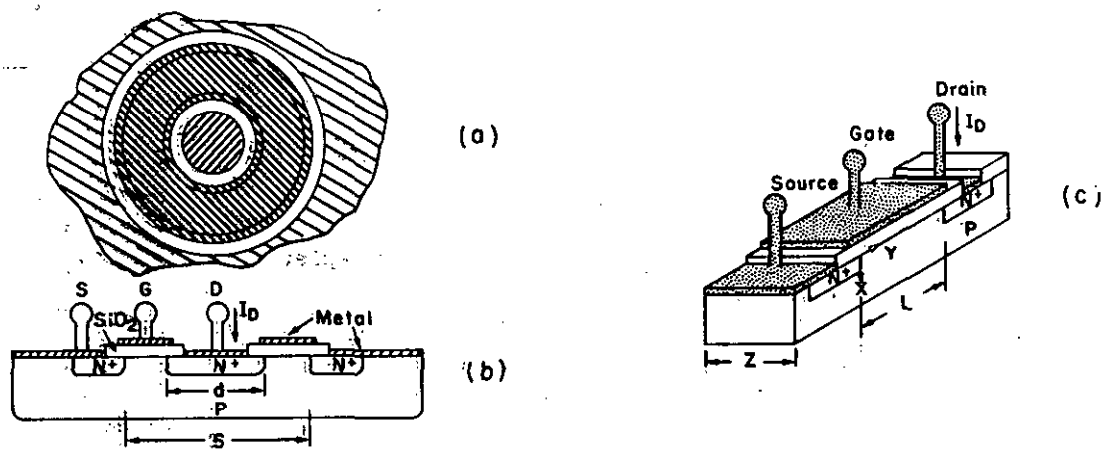


FIGURE 2-45

GEOMETRY OF N-CHANNEL MOS TRANSISTORS (a) TOP VIEW (b) CROSS-SECTIONAL VIEW (c) THE LINEAR STRUCTURE

The N+ regions of the structure shown in Figure 2-45 are obtained by high-temperature diffusion of phosphorus impurity into the P-type silicon substrate by the use of oxide masking techniques.** The N+ region used as the source can be internally connected to the P-type substrate when the source electrode is fabricated. If it is not, the device is a four-terminal device, with the substrate acting as the fourth terminal. The other N+ region is the drain and is usually employed as the output terminal. The source acts as the common terminal for the input and output.

*U.S. Patent No. 3102230, "Electrical Field Controlled Semiconductor Devices", filed 31 May 1960 and issued on 27 August 1963 to D. Hahng; and U. S. Patent No. 3056888, "Semiconductor Triode", filed 17 August 1960 and issued on 2 October 1962 to M. M. Atalla. Some of the materials covered in these two patents were presented at the IRE-AIEE Solid State Device Research Cong., Pittsburgh, Pa., June 1960 by these authors in "Silicon-Silicon Dioxide Field Induced Surface Devices".

**For a comprehensive review of the silicon planar-technology developments see G. E. Moore, "Semiconductor Integrated Circuits," Chapter 5 of Principles of Microelectronic Engineering, E. Keonjian, Ed., (New York, 1962).

The input lead, called the gate, is evaporated over the oxide. In this particular structure the oxide is silicon dioxide. However, work is currently being done with silicon nitride as the insulator. The SiO_2 layer is a determining factor of device characteristics and stability.

The oxide insulator can be grown thermally at high temperature or anodically at room temperature. Oxide thickness of the order of 1000 Å is generally used. Thick oxide reduces the transconductance, gain, and speed of the device for a given d-c operating point; and very thin oxide makes the reproducibility of the device difficult.* The metal gate electrode and the metal electrodes that form the ohmic contacts to the source and drain regions are made by evaporation and photoresist techniques. Metals such as aluminum, silver, and gold are used.

The linear structure illustrated in Figure 2-45c is well suited for integrated-circuit work. It is easy to interconnect the devices, and there is a natural isolation between devices on the same substrate.

2.2.5 Drain Current

The equation for drain current is derived by Sah* as

$$I_D = \frac{\bar{\mu}_n C_o}{L^2} \left[(V_G - V_T) V_D - \frac{V_D^2}{2} \right] \qquad (2\text{-}13)$$

where

I_D = drain current

V_G = gate-to-source voltage

V_D = drain-to-source voltage

V_T = the turn-on voltage if it is positive or turn-off voltage if it is negative

C_o = the total capacitance of the oxide layer under the gate

L = the channel length as indicated in Figure 2-45c

$\bar{\mu}_n$ = average surface mobility of electrons in the channel

Equation 2-13 is based on a simplified model of the channel and is valid only for gradual channels; it ceases to be a good approximation in the region of the channel where it is pinched off or nearly pinched off.

* C.T. Sah, "Characteristics of the Metal-Oxide-Semiconductor Transistors", IEEE Transactions on Electron Devices, July 1964, pp. 324-345.

If V_T in equation 2-26 is positive and V_G is zero, the current will be negative. This cannot be the case in practice. What is indicated here is that V_G must be larger than V_T in order to have current flow in the channel; thus we are dealing with an enhancement transistor. This is easily seen in Figure 2-46, which illustrates the transfer characteristic of an enhancement MOS transistor.

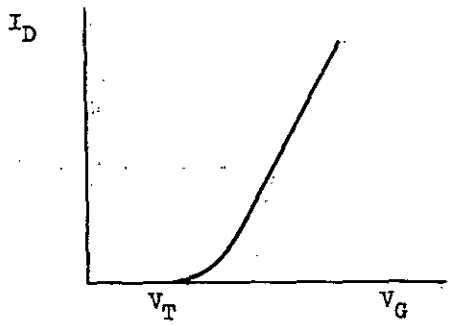


FIGURE 2-46
TRANSFER CHARACTERISTIC OF ENHANCEMENT MOS TRANSISTOR

From Figure 2-46 it can be seen that current does not flow until some value of gate voltage (V_T) is applied between gate and source. This is the voltage required to invert the channel discussed in Section 2.2.1. Current will not flow in the channel until the gate voltage exceeds the transistor turn-on voltage. The determining factor in the magnitude of V_T is the oxide thickness. For thin oxides, V_T is lower than for thicker oxides.

If V_T is negative, Equation 1 indicates a current at zero gate voltage. This is the case for the depletion-type transistors. The transfer characteristic for an N-channel depletion transistor is illustrated in Figure 2-47.

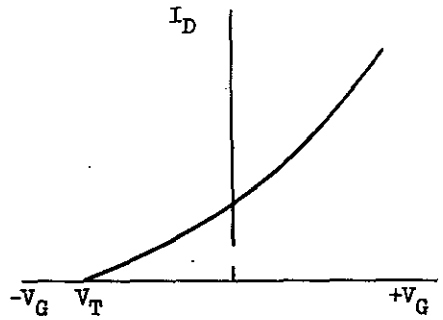


FIGURE 2-47
TRANSFER CHARACTERISTIC FOR DEPLETION MOS TRANSISTOR

Note that for $V_G = 0$ in Figure 2-47, there is substantial drain current. In this case V_T is the voltage required to turn off the drain current.

2.2.6 Pinch-off

Equation 2-26 indicates a linear current-voltage relationship for the MOS. This is true only in the ohmic region of the V_D - I_D characteristic. As the drain voltage is increased beyond some value (assuming constant V_G), drain-current saturation takes place. At this point, the channel is pinched off. This is the pentode region of the V_D - I_D characteristic.

The gradual pinch-off of the channel of an N-type depletion MOS is illustrated

in Figure 2-48. Assume that the gate electrode is shorted to the source ($V_G = 0$). For zero applied drain voltage the channel is unconstricted, as is shown in Figure 2-48a. As the drain voltage is increased from zero in a positive direction, current begins to flow in the channel. This is point (b) on the curve of Figure 2-48e.

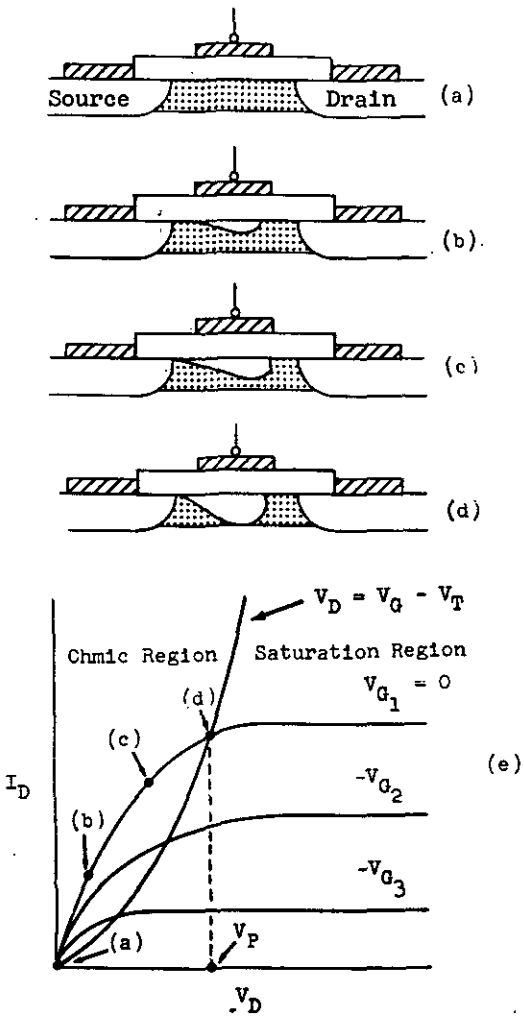


FIGURE 2-48

PINCH-OFF FOR DEPLETION MOS TRANSISTOR

The applied voltage is dropped across the channel, and the effect is to reverse-bias the gate electrode relative to the channel. Since the channel resistance is distributed, the voltage is distributed along the channel and the most negative area is that closest to the drain. The result is an induced depletion region in the channel, with the greatest depletion occurring closest to the drain, as is indicated in Figure 2-48b.

As the drain voltage is increased to correspond to point (c) on the curve of Figure 2-48c, the depletion region is further enlarged. At point (d) the drain voltage equals the pinch-off voltage, and drain-current saturation takes place. At this point it is assumed that the depletion region extends across the channel. There is only a slight increase in drain current for further increases in drain voltage.

From Figure 2-48e it can be seen that the drain current can be divided into two segments: that which appears in the ohmic region and that which appears in the saturation region. These regions are divided by the locus of points determined by

$$V_D = V_G - V_T$$

For values of $V_G - V_T < V_D$, Equation 2-13 is valid. However, in the saturation region, where $V_G - V_T \geq V_D$, this equation is no longer valid; the gradual-channel approximation fails. This condition corresponds approximately to the condition of maximum drain current from Equation 2-13 and can be obtained from it as follows:

If we set $\left(\frac{\delta I_D}{\delta V_D}\right)_{V_G} = 0$ and use Equation 2-13, (2-14)

the pinch-off condition is

$$V_D = V_{DS} = V_G - V_T \tag{2-15}$$

where V_{DS} is the drain voltage at saturation.

The drain current at the pinch-off condition is given by

$$I_{DS} = \frac{\mu_n C_0}{2L^2}\left(V_G - V_T\right)^2 = \frac{\overline{\mu}_n C_0}{2L^2} V_{DS}^2 \tag{2-16}$$

If the drain voltage is further increased beyond the pinch-off voltage given by Equation 2-15, the pinch-off region lengthens into the channel. Most of the additional voltage applied to the drain beyond the pinch-off voltage appears across the length of the pinch-off region and causes little increase in the drain current.

In the saturation region the channel is said to be pinched off. Yet a constant current is being conducted from source to drain. In this region space-charge-dominated currents* are generated in the drain area. This type of current flow differs from ohmic current flow since the channel drift field now controls the distribution of mobile charge as well as the charge velocity. It is analogous in this sense to space-charge current flow in conventional vacuum tubes, where the plate-to-cathode field gradient determines the space charge.

2.2.7 Determining V_T

A rapid experimental determination of the threshold voltage V_T can be obtained from two terminal characteristics. The gate electrode is connected directly to the drain for the enhancement-mode N-channel device, which has $V_T > 0$ and no built-in channel. For the depletion-mode N-channel device a battery of $V_{GG} > |V_T|$ (positive side tied to the drain) is used that has a sufficiently high voltage to pinch off the built-in channel.

*G.T. Wright, Proceedings of the IEEE, Vol. 51 (1963), p. 1642.

In these connections, the device is in the saturation region since $V_D = V_G + V_{GG} > V_{DS} = V_G - V_T$. Thus the two-terminal drain current can be obtained from Equation 2-16 by the use of Equation 2-15:

$$I_{DS} = (\bar{\mu}_n C_o / 2L^2) \cdot (V_D - V_{GG} - V_T)^2, \qquad (2\text{-}17)$$

which shows that the onset of the drain current corresponds to a drain voltage of $(V_T + V_{GG})$. For devices of the enhancement type, the turn-off voltage V_T can be determined readily from a display of this characteristic without the use of a gate battery since $V_T > 0$. For the depletion type, $V_T < 0$; and a gate battery of $V_{GG} > |V_T|$ must be used to determine the turn-on voltage V_T. This slight additional complication comes from the fact that an N-channel MOS transistor, which has a P-type substrate, cannot be biased with a large negative drain voltage, since then the drain junction would be forward-biased and the large forward-drain junction current would mask off the pinch-off point.

2.2.8 Determining the g_d and g_m

The low-frequency values of g_d, the drain conductance below saturation, can be determined from Equation 2-18.

$$g_d = \left(\frac{\delta I_D}{\delta V_D}\right)_{V_G} = \frac{\bar{\mu}_n C_o}{L^2} (V_G - V_T - V_D), \qquad (2\text{-}18)$$

Thus g_d decreases linearly with drain voltage and becomes zero at saturation, where $V_D = V_{DS} = V_G - V_T$.

The transconductance (g_m) is

$$g_m = \left(\frac{I_D}{V_G}\right) V_D = \frac{\bar{\mu}_n C_o}{L^2} V_D \qquad (2\text{-}19)$$

and its maximum, which occurs at saturation, is given by

$$g_{ms} = \frac{\bar{\mu}_n C_o}{L^2} V_{DS} = \frac{\bar{\mu}_n C_o}{L^2} (V_G - V_T) \qquad (2\text{-}20)$$

Since $(V_G - V_T) = V_{DS}$ in the saturation region, one can use Equation 2-16, substitute into Equation 2-19, and rearrange such that

$$g_{ms} = \sqrt{\frac{2 I_{DS} \bar{\mu}_n C_o}{L^2}} \qquad (2\text{-}21)$$

The value of transconductance in Equation 2-21 is also approximately the value of transductance when the device is operated beyond the saturation voltage.

2.2.9 Frequency Limitations of an MOS Device

The upper frequency limit of an MOS device depends on τ, the transit time of a carrier from source to drain. The transit time is a function of device channel length (L), the surface mobility (μ), and the applied voltage from drain to source. The equation for transit time is

$$\tau = \frac{L^2}{\mu V_{DS}} \qquad (2\text{-}22)$$

The intrinsic gain-bandwidth product for the MOS is

$$GBW = \frac{g_m}{2\pi C_c} \qquad (2\text{-}23)$$

where g_m is the device transconductance and C_c is the active gate-to-channel capacitance (in saturation $C_c = 2/3\ C_o$). The gain-bandwidth product is directly related to the carrier transit time as

$$GBW = \frac{1}{2\pi\tau} \qquad (2\text{-}24)$$

In actual practice the upper frequency limit may be substantially lower than $1/2\pi\tau$. For a 10-micron channel, the upper frequency limit will be several hundred megacycles per second.

2.2.10 Equivalent Circuit of the MOS

An equivalent circuit of the MOS is illustrated in Figure 2-49.* The equivalent circuit is composed of six resistors, five capacitors, a constant-current source, and two diodes.

COURTESY OF ELECTRONICS, McGRAW-HILL © 1964

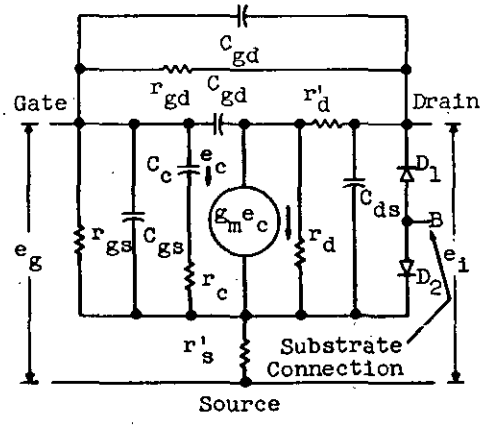


FIGURE 2-49
EQUIVALENT CIRCUIT FOR MOS FET OPERATING IN THE PINCH-OFF REGION

Resistance r_{gs} represents the leakage from gate to source, and r_{gd} is the leakage from gate to drain. These values are very high, typically about 10^{15} ohms.

The series network formed by C_c and r_c is a lumped approximation of the distributed network of the active channel. The capacitance C_c is the sum of the small capacitors distributed over the active channel area; it is expressed as

$$C_c = \frac{\partial Q_c}{\partial V_g} \qquad (2\text{-}25)$$

where Q_c is the total channel charge. C_c is usually a function of the applied

*D. M. Griswold, "Understanding and Using the MOSFET", Electronics, (December 14, 1964), p. 66.

gate and drain voltage. In saturation, this simplifies to

$$C_c = \frac{2}{3} A_c C_{ox} \tag{2-26}$$

where A_c is the gate area overlying the active channel and C_{ox} is the oxide capacitance per unit area. Typically, oxide thickness T_{ox} = 1000 angstroms, so that $C_{ox} = 10^{-8}$ Farads/Cm^2.

The capacitance C_c charges and discharges through the channel resistance r_c The channel resistance, in turn, is composed of innumerable series and parallel resistors between the source, or drain, and the points in the channel where the individual channel-to-gate capacitances take effect. The high-frequency performance of the MOS is strongly dependent on the r_c C_c time constant.

When the MOS is operated in the saturation region, it appears as a constant current source. For this reason the active portion of the circuit is depicted as a constant current source with a value of $g_m e_c$. The low frequency value of g_m can be obtained from the V_D - I_D characteristic as

$$g_m = \left.\frac{\Delta I_D}{\Delta E_g}\right|_{V_D} \tag{2-27}$$

Resistance r_d in parallel with the current generator, $g_m e_c$, represents the dynamic output resistance of the transistor. It can be determined by the slope of the output characteristics as

$$r_d = \left.\frac{\Delta V_D}{\Delta I_D}\right|_{V_{sG}} \tag{2-28}$$

In the pinch-off region, r_d is several orders of magnitude larger than any parasitic resistances. The parasitic resistances appear in series with the source and at the drain.

Resistances r_d' and r_s' represent those portions of the source-to-drain channel which are not controlled by the transistor's gate. These parasitic resistances are mostly caused by the metal-to-semiconductor contact of the source and drain regions. However, when the gate electrode is offset from the drain (depletion-mode transistor), that portion of the channel resistance not under the gate electrode is included in r_d'.

The only effect of r_d' is to shift the value of the external drain voltage required to obtain drain-current saturation. However, r_s' appears as a common element to the input and output circuit and therefore induces degenerative feedback. It lowers the external terminal transconductance, g_m, which is expressed as

$$g_m = \frac{g_{mo}}{1 + r_s' g_{mo}} \tag{2-29}$$

where g_{mo} is the internal transconductance. To keep r'_s as small as possible, the gate electrode always overlaps the N+ source contact.

Capacitances C_{gd}, C_{ds}, and C_{gs} are the physical case and interlead capacitances between gate and drain, gate and source, and drain and source. Capacitances C_{gd} and C_{gs} also include any capacitances that are not dependent on voltage, such as that contributed by the physical overlap of the insulated gate over the source or drain. C_{gd} is reduced significantly when gate off-set is used for depletion-mode devices. Capacitance C_{ds} includes the capacitances of D_1 and D_2.

Capacitance C'_{gd} represents the intrinsic gate-to-drain capacitance, which decreases as the channel voltage approaches the pinch-off region. This capacitor is quite significant since it determines the degree to which drain-current saturation is achieved.

Diode D_1 represents the junction formed between the N+ drain region and the semiconductor substrate; D_2 is the junction formed between the N+ source region and the substrate. D_1 and D_2 are back-to-back diodes in parallel with the channel. At high frequencies the diodes contribute an equivalent series RC network that affects the output admittance.

2.2.11 Amplifier Circuits

There are three basic single-stage circuit configurations in which the MOS is used. Each has its advantages in a particular application.

2.2.11.1 Common Source

The common-source configuration is illustrated in Figure 2-50a. The signal is applied between gate and source. This circuit has a high input impedance and a medium-to-high output impedance. The circuit provides a voltage gain greater than unity, which is given by

$$A = - \frac{g_m r_d R_L}{r_d + R_L} \qquad (2\text{-}30)$$

where g_m is the transconductance, r_d is the drain resistance, R_L is the effective load resistance, and the minus sign indicates the phase reversal from input to output. If an unbypassed resistance is introduced between the source and ground as shown in Figure 2-50a, degenerative feedback is induced in the circuit. The common-source voltage gain A' with an unbypassed source source resistor is

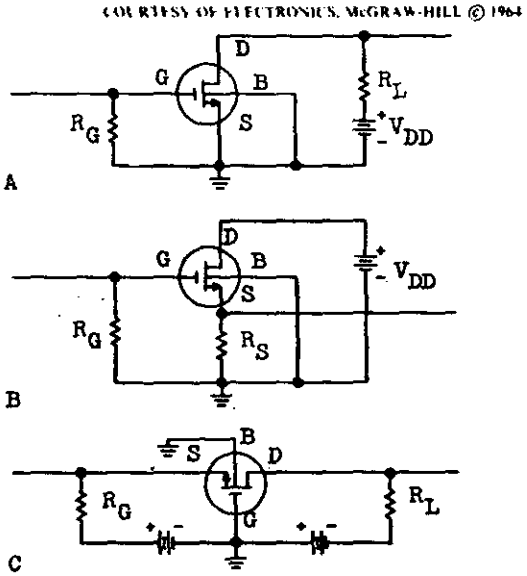
$$A' = \frac{g_m \, r_d \, R_L}{r_d + (g_m \, r_d + 1) \, R_s + R_L} \qquad (2\text{-}31)$$

where R_s is the unbypassed source resistance. The output impedance, Z_o, is increased by the unbypassed resistor:

$$Z_o = r_d + (g_m \, r_d + 1) \, R_s \qquad (2\text{-}32)$$

2.2.11.2 Common-drain Circuit

The common drain circuit, illustrated in Figure 2-50b, is sometimes referred to as a "source follower." The input impedance is very high, but the output impedance is low. There is no phase reversal between input and output. The distortion is low, and the voltage gain is less than unity.



B=Bulk Gate
D=Drain
G=Control Gate
S=Source
R_G=Gate Resistor
R_S=Source Resistor

FIGURE 2-50

THE THREE BASIC AMPLIFIER CONFIGURATIONS FOR A FET:
(A) COMMON-SOURCE OPERATION
(B) SOURCE-FOLLOWER OPERATION
(C) COMMON-GATE-OPERATION

This circuit finds application where an impedance transformation from high to low is required and where low input capacitance is desirable. The input is injected between gate and drain, and the output is taken between source and drain. This circuit provides 100% degenerative feedback. Its gain is given by

$$A' = \frac{R_S}{\left(\frac{\mu + 1}{\mu}\right) R_S + \frac{1}{g_m}} \qquad (2\text{-}33)$$

Since the amplification factor, μ, is usually much greater than unity, this reduces to

$$A' \approx \frac{1}{1 + \frac{+1}{g_m R_S}} \qquad (2\text{-}34)$$

The input resistance is R_G when R_G is connected to ground. If R_G is returned to the source contact of the MOS, the input resistance R_i is given by

$$R_i = \frac{R_G}{(1 - A')} \qquad (2\text{-}35)$$

If the load is resistive, the input capacitance of the source follower is reduced by the negative feedback:

$$C_i = C_{gd} + (1 - A')C_{gs} \qquad (2\text{-}36)$$

The output resistance is given by

$$R_o = \frac{r_d R_s}{(g_m r_d + 1) R_s + r_d} \qquad (2\text{-}37)$$

Since $\mu = g_m r_d$ and is usually much greater than unity,

$$R_o \approx \frac{1}{g_m + G_s} \qquad (2\text{-}38)$$

where

$$G_s = \frac{1}{R_s}$$

2.2.11.3 Common-Gate Circuit

The common-gate circuit (Figure 2-50a) is used to transform from a low to a high impedance. The input impedance of this configuration has approximately the same value as the output impedance of the source-follower circuit. The gain is given by:

$$A = \frac{(g_m r_d + 1) R_L}{(g_m r_d + 1) R_G + r_d + R_L} \qquad (2\text{-}39)$$

COURTESY OF ELECTRONICS, McGRAW-HILL © 1964

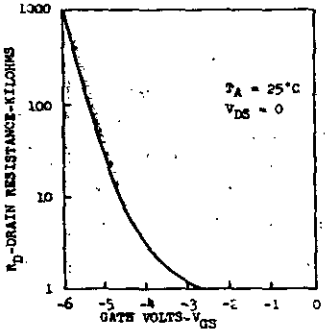


FIGURE 2-51

DRAIN RESISTANCE CHARACTERISTICS

2.2.12 Variable Resistor

The MOS can be used as a variable resistor or voltage-controlled attenuator. A variation in gate voltage causes the drain-to-source resistance to vary. Within a certain range of gate voltage this resistance change is linear. Figure 2-51 is an illustration of the variation of drain resistance with gate voltage. For high negative voltage the characteristic is linear and the range is from 5kΩ to 1MΩ.

2.2.13 Switching Times

For switching applications, such as the inverter circuit shown in Figure 2-52, the switching times consist of both the time constant of charging and discharging the channel and the time constant associated with load resistance, R_L, and capacitance, C'_d. This capacitance, C'_d, consists of the drain-junction capacitance and the stray capacitance of the device header and package and associated circuit wiring. During the turn-on transient, the time constant for charging the channel dominates over $R_L C'_d$, while during the turn-off transient, $R_L C'_d$ dominates.

COURTESY OF IEEE TRANS ON ELECTRONIC DEVICES

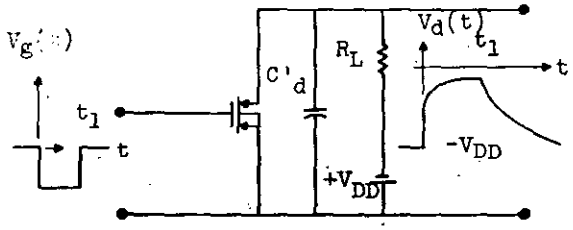


FIGURE 2-52

SWITCHING TRANSIENTS IN A P-CHANNEL MOS TRANSISTOR INVERTER CIRCUIT

Consider the turn-on transient as illustrated in Figure 2-52 for a P-channel device without a built-in channel. The device is initially in the off state, with a negative drain voltage, V_{DD}, and zero gate voltage. At t = 0, a negative gate voltage is applied which is sufficient to induce a hole channel and cause holes to flow down the drain from the source and discharge the capacitance C'_d, which is initially charged up to a charge of $-V_{DD}$. If the time constant, $R_L C'_d$, is large compared with the time constant of the channel, the discharge time of C'_d is essentially the time constant of the channel since during this time the

*C. T. Sah, op. cit.

build-up of charge on C'_d by the load current flowing through R_L from the V_{DD} battery is negligible. The channel time constant is an intrinsic property of a given device and depends on the device geometry and material properties.

This time constant is the transit time of the electrons across an N-type channel for the case of constant electron mobility and is the same as Equation 2-22:

$$\tau = \frac{L^2}{\mu V_{DS}} \qquad (2\text{-}40)$$

From this equation it is seen that the channel time constant decreases slightly as the drain voltage is increased beyond V_{DS}. If the value of $V_{DS} = \sqrt{\frac{2L\omega I_{DS}}{\mu C_c}}$ (where C_c is the total effective gate capacitance, which equals 2/3 C_o in saturation) is substituted into Equation 2-53, the following expression for the channel time constant is obtained:

$$\tau = \frac{L^2 C_c}{2\mu I_{DS}} = \frac{C_c V_{DS}}{I_{DS}} \qquad (2\text{-}41)$$

A numerical example will illustrate the contributions from various sources on the switching time of the turn-on transient. For a P-channel MOS transistor switch, the following numerical values are typical: channel length, L, 10 microns; channel width, Z, 250 microns; bulk resistivity, 10 ohm-cm, corresponding to a donor impurity concentration of 4×10^{14} atoms/cm^3; average surface mobility of holes, 100 cm^2/v-sec; oxide thickness, X_o, 1000Å. The total gate capacitance (C_o) is given as

$$C_o = \frac{K_o \epsilon_o Z L}{X_o}$$

where

K_o = dielectric constant of the oxide ($\approx$4.0)

ϵ_o = permissitivity of free space (8.85×10^{-14} farad/cm)

$$C_o = \frac{[4.0\ (8.85)\ 10^{-14}\ (250)\ 10^{-4}\ (10)\ 10^{-4}]}{1000 \times 10^{-8}}$$

$$= 0.885 \times 10^{-12} \text{ farads}$$

$$= 0.885 \text{ pf}$$

In the saturation region the effective gate-to-channel capacitance is

$$C_o = 2/3\ C_o$$

Therefore, C_c = 0.590 pf.

Assuming that the device is switched on by a gate voltage to a steady-state drain current in the saturation region of $I_D = I_{DS}$ = 10 mA, then the saturation drain voltage can be calculated from Equation 2-16 as

$$V_{DS} = \sqrt{\frac{2L^2 I_{DS}}{\mu_p C_o}} = \sqrt{\frac{210^{-6}(10^{-2})}{(10^2)(0.885)10^{-12}}}$$

$$V_{DS} = 15$$

and

$$\tau = \frac{C_c V_{DS}}{I_{DS}} = \frac{0.590 \times 10^{-12} \times 15}{10^{-2}}$$

$$= 885 \times 10^{-12} \text{sec.}$$

This is the value of the electron transit time or the charging and discharging time of the channel for the given conditions. The gain-bandwidth product for such a device, is

$$GBW = \frac{1}{2\pi\tau}$$

$$= \frac{1}{(6.28)(885)10^{-12}}$$

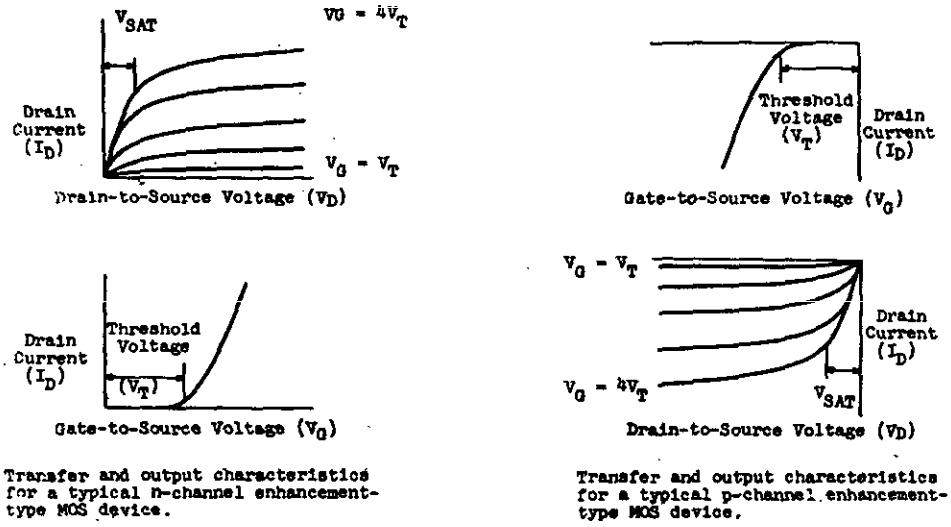
$$GBW = 180 \text{ MHz}$$

For a comparison of the channel time constant with $R_L C'_d$, assume R_L = 1000 ohms. The capacitance associated with the drain-to-substrate is about 0.1 pf in a typical device. The drain lead capacitance through the header is about 0.5 pf for a TO-5 package. Other stray capacitance associated with the circuit wiring may be as high as 3 pf. Assuming 3 pf, it can be determined that $R_L C'_d = 3000 \times 10^{-12}$ sec, which is considerably greater than the channel time constant of 885×10^{-12} sec. The turn-on time constant is determined by the transit time.

However, the relative importance of the circuit and channel time constants is reversed during the turn-off transient. The drain-voltage transient decreases toward $-V_{DD}$ (see Figure 2-52) more slowly since the capacitance C'_d must be charged up to $-V_{DD}$ through R_L. Thus the turn-off time constant is approximately $R_L C'_d = 3000 \times 10^{-12}$ sec. If the circuit time constant is comparable to the transit time, the total switch-off time constant would then be given approximately by the sum of these two time constants. The total switching time (turn-on plus turn-off) is limited by the circuit rather than the device.

2.2.14 Switching Applications

The enhancement-type transistor is well suited for digital circuit applications because direct-coupled signal inversion is possible with no need for level shifting between stages. Characteristics of the devices that are important in this application are illustrated in Figure 2-53, which displays the output and transfer characteristics of both N-type and P-type enhancement-mode MOS transistors. Of particular importance is the fact that the saturation voltage ($V_{D\ Sat}$) is less than the threshold voltage (V_T); this characteristic of MOS devices enables the circuit designer to construct extremely simple direct-coupled logic circuits.



Transfer and output characteristics for a typical n-channel enhancement-type MOS device.

Transfer and output characteristics for a typical p-channel enhancement-type MOS device.

FIGURE 2-53

MOS DIGITAL CIRCUIT*

2.2.15 MOS Digital Circuits*

The potential advantages of MOS devices in integrated digital arrays can be realized only if proper consideration is given to circuit choice. Although many circuit configurations are possible, only one or two take advantage of the unique properties of MOS devices to achieve a high figure of merit of functional complexity per unit at high yield. Before logic and storage circuits are discussed, four possible MOS inverter circuits will be considered. Each will be evaluated for application in an integrated circuit. not as a circuit constructed from discrete components.

To simplify the discussion and to use only positive voltages, it is assumed that all the examples shown for single-channel MOS circuits use N-channel devices. If all voltages are reversed, the same results can be achieved with P-channel devices. In fact, P-channel arrays, because of their somewhat simpler processing, are more common than N-channel arrays at present.

Figure 2-54 shows an inverter that uses an MOS inverting transistor and a resistor. In integrated form, this circuit is a very poor choice. For small MOS structures, a high value of resistance is required (greater than 10,000 ohms). If the resistor is formed by the source and drain diffusion, it occupies too large an area. If a separate diffusion is used, or if the resistor is formed as a thin film over the oxide, the processing is increased in complexity. In either case, the figure of merit for the circuit is reduced.

*R. D. Lohman, "Applications of MOS FET's in Microelectronics", SCP and Solid State Technology, (March 1966).

COURTESY OF SCP AND SOLID STATE TECHNOLOGY

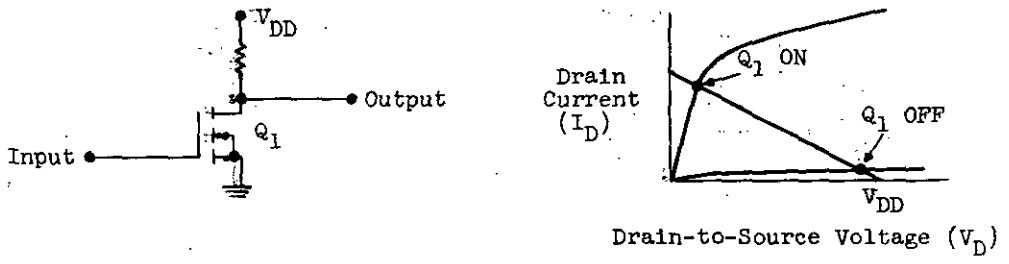


FIGURE 2-54

INVERTER USING MOS TRANSISTOR AND A RESISTOR WITH TYPICAL LOAD LINE

Figure 2-55 shows an inverter that uses a depletion-type MOS transistor as a load for an enhancement-type MOS inverter. (A depletion-type MOS transistor is one that conducts appreciable current with zero bias between gate and source.)

COURTESY OF SCP AND SOLID STATE TECHNOLOGY

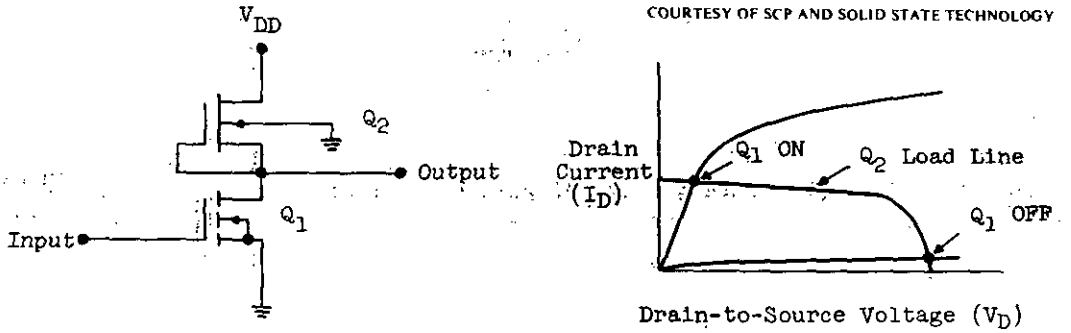


FIGURE 2-55

INVERTER USING DEPLETION-TYPE MOS TRANSISTOR AS LOAD FOR ENHANCEMENT-TYPE MOS INVERTER, AND TYPICAL LOAD LINE

This circuit has a number of advantages. First, its area is potentially very small because it consists only of MOS devices. Second, the shape of the load line that results is capable of improved speed as compared with an ohmic resistor. From a processing standpoint, however, it suffers from the disadvantage that two types of MOS devices, with different threshold voltages, are required. The resulting processing complexities tend to lower its figure of merit drastically.

Figure 2-56 shows an inverter that uses an enhancement-type MOS transistor connected as a source follower to serve as a load for the inverting MOS transistor. Although this circuit is a poor choice from a discrete-component point of view, it possesses a very high figure of merit considered in the light of digital integrated arrays. First, it uses the absolute minimum area because, to operate at

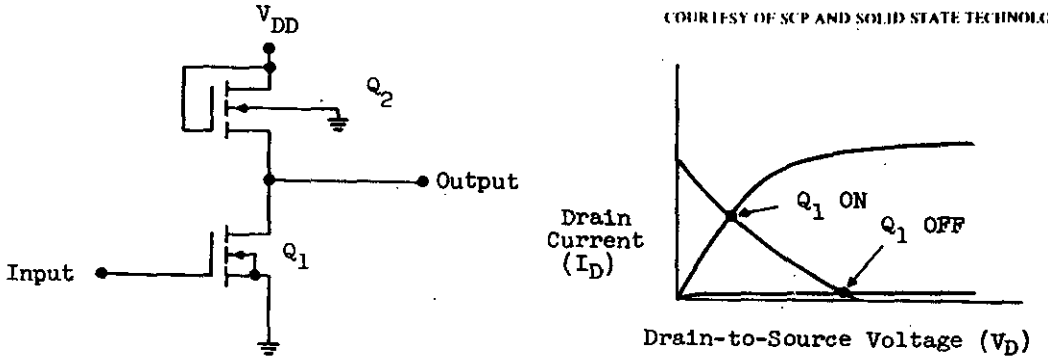


FIGURE 2-56

INVERTER USING ENHANCEMENT-TYPE MOS TRANSISTOR CONNECTED AS SOURCE FOLLOWER AS LOAD FOR INVERTING MOS TRANSISTOR, AND TYPICAL LOAD LINE

all, the load MOS device must be smaller than the inverting MOS device. Second, the processing required to fabricate the circuit is also minimum because the load MOS device is made by the same process as that used for the inverter MOS device. The operation of the circuit depends on the control of the ratio of the transconductance of the two devices. Although the absolute value of the transconductance depends on a number of factors, including the oxide thickness under the gate, the ratio is determined by the geometries of the two devices. When accurate masks are used, therefore, the ratio can be held constant even with variations in processing.

The final inverter circuit to be considered is shown in Figure 2-57. It consists of a pair of complementary MOS devices connected in series, with the

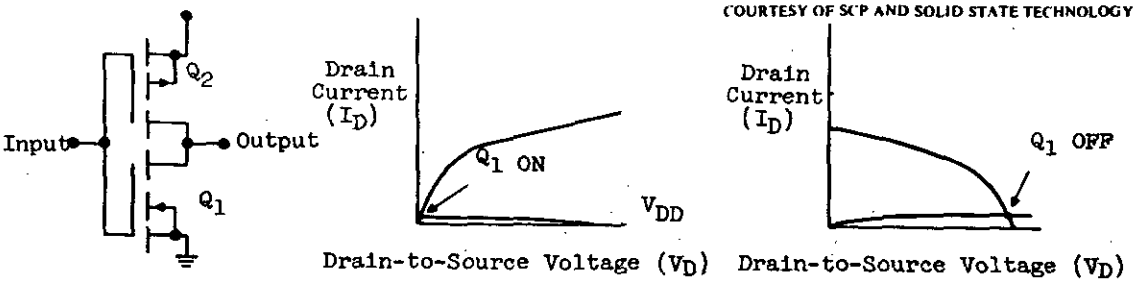


FIGURE 2-57

INVERTER AND LOAD LINES (USING COMPLEMENTARY MOS DEVICES)

gates connected to each other and driven by the input signal. When the input signal is zero, the P-channel device is on and the N-channel device is off. The reverse is true when the input signal is positive. Each device, when on, is required to supply a direct current equal to only the leakage current of the other device. During a transistion of the input signal, however, capacitive loads

are charged through the low output impedance of one or the other of the two devices. Thus, although it requires very low standby power, the circuit is inherently fast. The load lines in Figure 2-57 indicate the very low standby power. The figure of merit for the circuit is not as high as for the source-follower circuit shown in Figure 2-56, because slightly more area is required and the processing is considerably more involved. Despite these disadvantages, arrays of complementary MOS circuits are beginning to find applications where either very low standby power or high speed is important enough to compensate for the reduced figure of merit.

The total power dissipated, neglecting the standby power, in a complementary MOS circuit during switching is given by

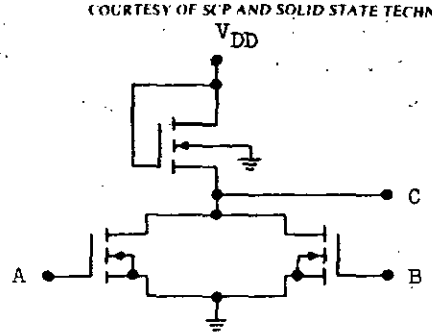
$$P \cong C_L V_{DD}^2$$

where C_L is the total output capacitance and V_{DD} is the supply voltage.

2.2.16 Single-Channel Arrays

The basic circuit shown in Figure 2-57 can be developed into a wide variety of digital circuit arrays. Parallel connection of the inverter transistor forms a NOR gate, as shown in Figure 2-58. Figure 2-59 shows how a NAND gate can be

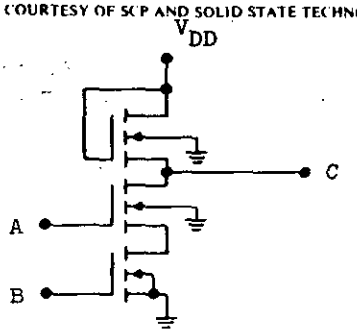
COURTESY OF SCP AND SOLID STATE TECHNOLOGY



A	B	C
H	H	L
H	L	L
L	H	L
L	L	H

FIGURE 2-58
NOR GATE USING PARALLEL CONNECTION OF MOS TRANSISTORS

COURTESY OF SCP AND SOLID STATE TECHNOLOGY



A	B	C
H	H	L
H	L	H
L	H	H
L	L	H

FIGURE 2-59
NAND GATE USING SERIES CONNECTION OF MOS TRANSISTORS

formed by series connection of the inverting transistors. When the NAND gate of Figure 2-59 is used in conjunction with the NOR gate of Figure 2-58, the resulting configuration offers a high degree of logic flexibility. However, the series connection suffers from the disadvantage that the inverter transistors must be twice as large as their parallel counterparts to maintain the same control of logic levels.

2.2.17 MOS Noise

In the MOS transistor, three mechanisms give rise to gate noise and thus input noise in an amplifier.* The main source of noise at high frequencies** is thermal noise due to random fluctuations in the free-carrier concentration in the channel. Noise figures comparable to low-noise vacuum tubes are obtained at frequencies above 50 MHz.

The low-frequency noise spectrum,† which may extend up to tens of megacycles per second in some devices, is controlled by the fluctuations in the number of electrons occupying surface traps; it resembles an f^{-n} distribution. The value of n is generally between 1 and 2, and the spectrum extends down to very low frequencies. Shot noise is produced by fluctuations of the individual charges as they drift and diffuse toward the surface. Leakage noise is associated with the small flow of current through the oxide.

2.3 FILM MICROCIRCUITS

2.3.1 Introduction

Silicon monolithic integrated circuits are by far the most widely used microelectronic approach. However, film microelectronics are also begining to have limited use. Film circuits are usually used where specialized applications such as high power, high frequency, and low component tolerances that cannot be met with silicon monolithic circuits are required.

Both thin-film circuits and thick-film circuits are available. Each has advantages and disadvantages when compared with either each other or with monolithic circuits. The major disadvantage is the lack of a suitable film-type active device. Thus both film circuits involve the deposition of passive film components on flat substrates and then the attachment of appropriate active devices.

2.3.2 Thin-Film Components

Thin-films are deposited in a vacuum chamber at a pressure of 10^{-5} torr. The material being deposited is electrically heated in a vessel in the chamber until it melts and vaporizes. The vaporized material condenses on a glass or ceramic substrate typically one inch square. The deposition pattern may be

---

* A.G. Jordan and N.A. Jordan, "Theory of Noise in Metal Oxide Semiconductor Devices", IEEE Transactions on Electron Devices, March 1965, p. 148.

** A. Van der Ziel, "Thermal Noise In Field-Effect Transistors", Proceedings of IRE, Vol. 50 (1962), p. 1808.

† A.G. Jordan and N.A. Jordan, "Noise in MOS Devices", Solid-State Devices Research Conference, (Boulder, Colorado, 1964); and

C.T. Sah, "Theory and Experiments on the 1/f Surface Noise of MOS Insulated-Gate Field-Effect Transistors", Solid-State Devices Research Conference, (Boulder, Colorado, 1964).

controlled by a metal mask. The mask is usually 0.002-inch-thick stainless steel and is placed over the substrate during vapor deposition. An alternative to this approach is to deposit over the whole substrate and then etch out the desired pattern using photolithographic techniques and selective etchants.

A number of materials have been used for thin-film resistors, including nichrome, SiO-Cr cermets, tantalum, and titanium, with nichrome being the most widely used. Substrates of finely polished glass, glazed ceramic, and oxidized silicon have been used. Nichrome resistors can be deposited with a sheet resistance of between 1 and 300 ohms/square, with a temperature coefficient of resistance of ± 50 ppm/deg C. Resistors can be deposited to tolerances of ± 1 percent. Tantalum is also widely used for thin-film resistors. The advantage of tantalum resistors is that they can be adjusted to close tolerances by electrolytic anodizing of the tantalum. Another advantage of tantalum is that the anodic oxide has a high dielectric constant and is suitable for the fabrication of capacitors in the range of 2.5 pf/sq mil. Thus both resistors and capacitors can be fabricated from a single metal. However, this approach has the serious disadvantage that the tantalum metal reacts with the tantalum oxide at temperatures above 200°C and cannot withstand the temperatures encountered in packaging integrated circuits. A number of metal compounds have been investigated for use in thin-film devices, including oxides and nitrides of such metals as chromium, titanium, and tantalum. In addition, an evaporated cermet resistor consisting of chromium and SiO appears to be promising. A sheet resistance of 1000 ohms/square and a tolerance of 5 percent can be achieved. However, few data on stability and reliability are available.

The process for producing thin-film capacitors involves three deposition steps: the deposition of a bottom electrode, the deposition of a dielectric material, and the deposition of the top electrode. The electrodes and dielectric are all deposited in succession on the substrate. The dielectric materials commonly used are silicon monoxide and silicon dioxide.

Silicon monoxide has a higher dielectric constant than silicon dioxide. However, evaporated silicon monoxide is subject to pinholes. The resulting loss in yield is usually not acceptable; in addition, the dielectric constant of silicon monoxide is strongly dependent on deposition conditions, making good reproducibility difficult to maintain. Silicon monoxide cannot be etched and therefore must be deposited through masks. This introduces the problems of achieving mask alignment and maintaining it during substrate heating and fabrication.

Silicon dioxide can be deposited by the simple process of reactive sputtering of silicon in a mixture of argon and oxygen at a pressure of a few microns. The use of silicon dioxide gives a high yield at a capacitance of 0.25 pf/sq mil

since the sputtered silicon dioxide is relatively pinhole-free. In addition, sputtered silicon dioxide is stable and has the properties of bulk-fused silica. The dielectric constant and the capacitance are thus reproducible. An equally important advantage of using silicon dioxide is the fact that it can be chemically etched with hydrofluoric acid, which is compatible with integrated-circuit processing. This permits the fabrication of the thin-film capacitors and resistors by photo-masking techniques -- techniques that all integrated-circuit manufacturers are currently using.

2.3.3 Thick-Film Components

Thick films are produced by screening or photo-etching formations of conducting and insulating materials on ceramic substrates. Noble metals are favored for conducting materials, and ceramics are normally used for the insulating portions.

Hundreds of different cermet formulations can be used to obtain a wide range of component parameters. The material used for a 10-ohm/sq resistor is quite different from that used for a 100-kΩ/sq resistor.

The formulations are fired at temperatures above 600°C to form an alloy that is permanently bonded to the insulating substrate. The characteristics of the final material can be controlled to a limited extent by the firing temperature/time profile.

The cermet is composed of metal combined with a glass frit. Inks with resistivity of 500 ohms/sq, 3kΩ/sq, 8kΩ/sq, and 20kΩ/sq are used routinely. Inks in the range of 50, 100, and 200kΩ/sq are available for higher resistance values. Resistance values can be specified to a tolerance of ±10 percent. Closer tolerances can be obtained by adjusting each resistor after fabrication.

Because of resolution limits of the screen and spreading of the edges of the resistor during firing, the minimum dimensions for a screened pattern are about 0.030 in.; but 0.050 in. is a more desirable minimum width for design purposes.

For a finer pattern resolution, a technique has been developed for etching cermet resistors that permits an order-of-magnitude reduction in the minimum line width. The process consists of the following steps: A cermet ink is applied uniformly over the entire area of a ceramic substrate; the substrate is coated with a photoresist, such as KPR; the substrate is exposed through a photomask and the photoresist is developed, leaving a protective coat of photoresist material over those areas that will comprise the resistor pattern.

Capacitors are formed by a sequence of screenings and firings; they consist of a high-temperature conductor (capacitor bottom plate and intraconnections), a dielectric, and a top-plate conductor. For R-C networks a resistor material is

also screened and fired. The final step is screening and firing of a glass encapsulant. The temperature profiles are dependent upon the materials used.

Thick-film capacitior sheet capacitance is a function of the material used and typically varies from 5,000 - 100,000 pf/in^2. The temperature and voltage characteristics are dependent upon the materials used. These capacitors have high breakdown voltages and good frequency response into the hundreds of MHz.

2.4 Comparison of Microelectronic Component Characteristics

Characteristics of various types of microelectonic resistors and capacitors are listed in Tables 2-6, and 2-7, respectively. The tables can be used to compare the differences in component types or to determine design constraints.

TABLE 2-8

CHARACTERISTICS OF MONOLITHIC RESISTIVE ELEMENTS

Parameter	Resistor Type				
	Diffused Base	Diffused Emitter	Thin-Film Nichrome	MOS	Thick-Film
Range of Values, ohms	100-30K	10-1K	20-50K	300-10^6	100-10^6
Sheet Resistivity, ohms/square	200	2.5	40-400	N/A	100-200k
Tolerance, Absolute	±10%	±15%	±5%	±20%	±10%
Tolerance, Relative	±3%	±3%	±1%	N/A	±3%
Power Dissipation, mW/square mil	3	3	2	3	0.5
Voltage Breakdown volts	30	5	100	50	100
Parasitic Capacitance, pf/mil^2	0.2-0.5	0.2-0.05	0.02	0.4	0.02
Temperature Effects, ppm/°C	500-2000	100-500	100	3000	±200
Frequency effects	Flat to 30 MHz	Flat to 30 MHz	Flat to 200 MHz	N/A	Flat to 100 MHz

TABLE 2-7

CHARACTERISTICS OF MONOLITHIC CAPACITORS

Capacitor Type	Maximum Capacitance, pf	Characteristic Capacitance pf/sq mil	Breakdown Voltage, Volts	Voltage Dependence	Tolerance Design	Temperature Effects, PPM/°C
Emitter-Base Junction	1000	0.9	5	K/ V	±20%	+60 to 600
Base-Collector Junction	500	0.2	20	K/ V	±20%	+60 to 600
MOS	250	0.1 - 0.8	50	--	±20%	±200
Thin-Film S_iO_2	200	0.2 - 0.4	30	--	±10%	±200
Thin-Film Ta_2O_5	5000	2.5	20	--	±10%	+400
Thick-Film	10,000	0.005 - 0.100	75	--	±10%	±300

BIBLIOGRAPHY: CHAPTER 2

BASIC PROCESSES AND DESIGN CONSIDERATIONS

Aarons, M. W. "Integrated Circuit Device and Fabrication Technology," Semiconductor Products, 111(1965), 42-45.

Abraham, D., and H. B. Goldberg. Microelectric Engineering; Volume 1,Fabrication Technology. No date.

Alberts, R. D. Integrated Silicon Device Technology; Volume IX, Epitaxy. Research Triangle Institute, 1965.

_____. Integrated Silicon Device Technology; Volume X, Chemical/ Metallurgical Properties of Silicon. Research Triangle Institute, 1965.

Allen, J. D. Microelectronic Analog-To-Digital Converter and Sync Generator. Jet Propulsion Laboratory. Technology Report 32-569.

Bailey, D. "Black Boxing Your Linear Integrated Circuit," Electronic Design, (1964), pp. 74-77.

Bell, W. V., J. E. Kernan, and P. H. Holub. Computer-Aided Analysis of a Silicon Monolithic Integrated Current Switch Gate. Army Electronics Command, ECOM-2683. 1966.

Berry, B. M. Integrated Silicon Device Technology; Volume XII, Techniques. Research Triangle Institute, 1966.

Bogert, H. Z., Jr. "Metal Oxide Silicon Integrated Circuits," SCP and Solid State Technology, (1966), pp. 30-35.

_____. "Metal Oxide Silicon Integrated Circuits," Semiconductor Products,III (1966), 30-35.

Burger, R. M. Integrated Silicon Device Technology; Volume VIII, Diodes. Research Triangle Institute, 1966.

Carroll, John M. Microelectronic Circuits and Applications. New York, 1965.

Christiansen, D. "EEE Specifying Guide: MOS Integrated Circuits," Circuit Design Engineering, (1966), pp. 60-70.

Christiansen, D. "Integrated Circuits in Action; Part I, The Great Design Dilemma," Electronics, (1966), pp. 68-73.

Coppen, P. J. "FET Complementary Integrated Circuits; Aerospace Material," Electronics, XXXII(1964), 55-58.

EDN, "A Review of Microelectronic Circuit and System Design," 1964.

Eimbinder, J. "Linear Integrated Circuits," EEE Circuit Design Engineering, (1966), pp. 76-112.

Fairchild Semiconductor Corporation. Integrated Circuit Design Considerations.

Farine, D. E., and D. Trotter. "MOS Integrated Circuits Save Space and Money," Electronics, (1965), pp. 84-95.

Foglesong, R. "Integrated Circuits Design and Application," SCP and Solid State Technology, (1964), pp. 32-42.

Fowler, W. L. "Look at Linear Integrated Circuits," Electronic Industries, IX(1965), 64-69.

Gibbons, J. F. Semiconductor Electronics. New York.

Government Reports. Technology for PNP Integrated Circuits. AD-629126. 1966.

Griswold, D. M. "Understanding and Using the MOS FET," Electronics, (1964), pp. 66-70.

Haehtel, G. D. Semiconductor Integrated Oscillators. ERL Report 65-32. 1965

Hartley, R. L. "Integrated Circuit Compatible R. F. Amplifier," NEREM, 1963.

Hauser, J. R. Integrated Silicon Device Technology; Volume XI, Bipolar Transistors. Research Triangle Institute, 1965.

Heiman, F. P., and S. R. Hofstein. "Metal-Oxide-Semiconductor Field-Effect Transistors," Electronics, (1964), pp. 50-61.

Heiman, F. P., and H. S. Miller. "Temperature Dependence of n-Type MOS Transistors," IEEE Transactions on Electronic Devices, pp. 142-148.

Hofstein, S. R., and F. P. Meiman. "The Silicon Insulated-Gate Field Effect Transistor," Proceedings of the IEEE, pp. 1190-1202.

Holland, C. E., Jr., and L. W. Sumney. "Using an Optical Comparator as a High Resolution Reduction Camera," IEEE Trans, PMP-1(1965) 35-40.

Hulme, J. R. "Design Considerations for Solid State Integrated Circuits," Solid State Design, (1963), pp. 33-38.

IEEE Spectrum. "Compilation of Microelectronic Design and Fabrication Articles," (1964), pp. 62-101.

Ilgenfritz, Robert. "Thick Film Hybrid Microelectronic Circuit Technology," Solid State Technology (1966), pp. 35-42.

Ingraham, R. C. "Photolithographic Masks for Integrated and Thin Film Circuitry," Semiconductor Prod., (1965), pp. 33-42.

Ivesdal, T. A. "Basics of Integrated Circuit Components," Electronic Design, (1964), p. 50.

Jordan, A. G., and N. A. Jordan. "Theory of Noise in Metal Oxide Semiconductor Devices," IEEE Transactions on Electronic Devices, pp. 148-156.

Kane, J. R., and D. L. Wollesen. Field-Effect Transistors in Theory and Practice. Motorola Application Note AN-211.

Keonjian, E. Microelectronics Theory, Design, and Fabrication. New York, 1963.

Kilby, J. S. Device Fabrication. 1966 International Solid State Circuits Conference.

Kile, C., Jr. "Semiconductor Integrated Circuits," SCP and Solid State Technology, (1967), pp. 30-38.

King, D., and L. Stern. "Designing Monolithic Integrated Circuits," Semiconductor Products, (1965).

Larocque, A. P., and N. O. Koralkoff. Technology for PNP Integrated Circuits. Army Electronics Command. ECOM-2665. 1966.

Latham, D. C., F. A. Lindholm, and D. J. Hamilton. "Low-Frequency Operation of Four-Terminal Field-Effect Transistors," IEEE Transactions on Electron Devices, (1964), pp. 300-305.

Levin, S. N. Principles of Solid State Microelectronics. New York, 1963.

Lin, H. C. "Design Considerations for Linear Functional Blocks," Solid State Design, (1964), p. 44.

Lohman, R. D. "Some Applications of Metal-Oxide Semiconductors to Switching Circuits," SCP and Solid State Technology, (1964), pp. 31-34.

_____. "Applications of MOS FET's in Microelectronics," SCP and Solid State Technology, (1966), pp. 23-29.

Luettgenau, G. G., and S. H. Barnes. "Designing with Low-Noise MOS FET's A Little Different But No Harder," Electronics, (1966), pp. 53-58.

Mammano, R. A. "System Design Using IC's," EDN, (1966), pp. 26-37.

Martin, T. B. "Circuit Applications of the Field Effect Transistor," Semiconductor Products, (1962), pp. 30-38.

Matzen, W. T., and R. A. Meadows. "Low-Frequency Integrated Circuits Achieved with Thermal Transfer," Electronics, (1964), pp. 54-57.

Matzen, W. T. Advanced Fabrication Process for Functional Electronic Blocks. 1965.

Maxwell, D. A. "The Minimization of Parasitics in Integrated Circuits by Dielectric Isolation," Wescon, 1964.

Meindl, J. D., R. Gilson, and W. Kiss. "How to Design Micropower Transistor Amplifiers; Part I," Electronics, (1964), pp. 73-82.

Miller, A. Semiconductor Films on Insulators Devices and Circuits. St. Louis, 1966.

Murphy, B. T. "Cost-Size Optima of Monolithic Integrated Circuits," Proceedings of the IEEE, (1964), pp. 1537-1545.

Nester, H. H., T. E. Salzer. "Thick Film Production Techniques," Electronic Packaging and Production, (1967).

Norman, R. H. "Digital Applications of Integral Electronics," Solid State Design, (1964), p. 24.

Olmstead, J. A. "Using MOS Transistors in Integrated Circuits," Electronic Design, (1964).

O'Reilly, T. J. "The Transient Response of Insulated Gate Field Effect Transistors," Solid State Electronics, pp. 947-956.

Peek, John R. "Design of Thin Film Resistor and Capacitor Circuits," Solid State Technology, (1967), pp. 27-35.

Phillips, A. B. Transistor Engineering and Introduction to Integrated Semiconductor Circuits. New York, 1962.

_____. "Designing Digital Monolithic Integrated Circuits," Monitor, (1964), p. 18.

Potterf, G. R., and D. W. Liddell. Preferred Specifications for Microelectronic Digital Circuits. NEL Technical Memorandum 858. 1965.

RCA Corporation of America. RCA Linear Integrated Circuit Fundamentals; Design, Application, Electronic Components, and Devices. 1966.

_____. Development of Integrated Circuits Utilizing Complementary Transistors. Interum Development Report No. 1.

Research Triangle Institute. Integrated Silicon Device Technology; Volume II, Capacitance. 1963.

Root, C. D., and L. Vadasz. "Design Calculations for MOS Field Effect Transistors," IEEE Transactions on Electronic Devices, pp. 294-299.

Sah, C. T. "Characteristics of the Metal-Oxide Semiconductor Transistor," IEEE Transactions on Electronic Devices (1964), pp. 324-345.

Schlegel, E. S. Metal-Oxide Silicon Technology; A Literature Review and Bibliography. Philco Report. 1965.

Sherwin, J. S. "Take the Fog Out of Field-Effect Design," Electronic Design, (1966), pp. 38-44.

Sherwin, J. S. "Gain Insight Into FET Amplifiers," Electronics Design, (1966), pp. 40-45.

Shive, J. N. The Properties, Physics, and Design of Semiconductor Devices. 1959.

Sikina, T., and J. Tsai. First Quarterly Report; MOS Integrated Circuits. Westinghouse Report. 1966.

Stone, H. A., and R. M. Warner. "The Field-Effect Tetrode." Proceedings of the IRE, pp. 1170-1961.

Thornton, C. G. "New Trends in Microelectronics Fabrication Technology: 1965-1966," Semiconductor Products, 111(1966), 42-49.

Tiger, B. and E. Donald. "The Stress Survival Matrix Test As A Tool for Evaluating Integrated Circuit Effectiveness," Analysis of Reliability and Maintainability, (1965), pp. 28-30.

U. S. Army Electronics Command. Technology, Design and Application of Integrated Circuits. Report 2747. Fort Monmouth, N. J., 1966.

Uzunoglu, V. "Feedback: Perils and Potentials in Designing Integrated Circuits," Electronics, 11(1965), 67-71.

_____. "Using Feedback Pairs in Integrated Circuits Part I -- The Series to Shunt Pair," Electronic Design, (1965), pp. 28-30.

Valdes, L. B. The Physical Theory of Transistors. New York, 1961.

_____. "Case History: Integrating A NOR Circuit," Electronic Design, (1964), p. 38.

Warner, R. M., Jr., and J. N. Fordenwalt, eds. Integrated Circuits;Design, Principles,and Fabrication.

Widlor, R. J. "Some Circuit Design Techniques for Linear Integrated Circuits," IEEE Transactions on Circuit Design, IV(1965), 586-590.

Wilson, R. W. "Thin Film Passive Elements for Monolithic Integrated Circuits," Solid State Technology, (1967), pp. 21-26.

SYSTEM DESIGN

THREE

3.1 INTRODUCTION

When microelectronic components are connected to perform an electrical function, their characteristics are modified. Usually the effect of intra-connection and interaction is deleterious to component performance. However, different fabrication approaches (e.g., IC, Hybrid, or LSI) affect different constraints or the same constraints in different degrees; it is thus possible to obtain better performance (both technical and economic) by selecting the approach whose trade-offs are most advantageous for a given application.

The following characteristics that affect system design are considered in this chapter:

- Packaging
- Interconnection
- Intraconnection
- Operating frequency
- Thermal
- Logic choice
- Redundancy
- System partitioning
- Maintainability and logistics

Although these topics are important, they by no means exhaust design considerations. Each system design is unique; thus some characteristics are emphasized more than others. These design topics, however, are believed to be of general importance in any design.

IC packages have evolved from the transistor type (modified to include more leads), which are still used, to the widely used flat pack, and to the dual in-line package (DIP). Generally, these packages are limited to 14 leads, but some configurations are available with many more. Packages with up to 160 leads are being developed for LSI devices. Such techniques as the flip-chip Hybrid Microcircuit, require package configurations that are not currently available.

Interconnections (connections made external to the device) are critical because they affect system performance, cost, and reliability. When a large number of IC's are mounted on a single board, it is necessary to have multiple layers of connections within the board. These multilayer boards are expensive and difficult to fabricate with high reliability.

The number of interconnections can be reduced if LSI techniques are employed. Such a reduction in interconnections would improve frequency response and might improve reliability.

The number of interconnections required with an LSI device is primarily dependent upon the manner in which the system or subsystem is partitioned. Thus partitioning is a major consideration for LSI design.

Interconnections are made external to the package; intraconnections, somesometimes referred to as metalizations, are made on the silicon substrate. Once they are made, they cannot be altered.

The relationship between interconnections and intraconnections is that in complex arrays intraconnections are substituted for interconnections. This substitution is advantageous if metalizations are less expensive and more reliable than the interconnections for which they are substituted, as is the case when a single level of metalization is required. However, LSI devices require two or more layers of metalization. In addition, it is necessary to intraconnect only the "good" circuits on a substrate. Thus, intraconnection techniques can become very complex.

Two intraconnection techniques have been proposed for LSI: fixed wiring and discretionary wiring. The fixed-wiring technique requires all circuits on some portion of the wafer to be good since the intraconnection scheme is never varied. Discretionary wiring provides much more flexibility since it makes it possible to cope with "bad" circuits in an array. With the discretionary approach, a different intraconnection pattern is generated for each LSI device. This method is much more costly than the fixed-wiring approach, but the effective yield may be sharply increased, thus making this method more practical than the fixed-wiring approach.

The intraconnection pattern can be fixed when the hybrid microcircuit approach is used since each device can be tested before it is bonded to the substrate.

One of the most important contributions of complex arrays is that they extend the operating-frequency range of monolithic circuits. Gate delays of less than a nanosecond are possible only with complex arrays. The number of leads, the conductor lengths, and the number of interconnections all introduce parasitics that limit operating frequency. Arrays provide close circuit proximity, reduce transmission-line lengths, and reduce the number of external connections.

Thermal effects and power dissipation are obviously important in monolithic-circuit design. High-power circuits, greater than 1 watt, are usually not compatible with monolithic technology. Power dissipation for IC's is typically about 10 mW per gate. These power levels can easily be accommodated with present-day heat-transfer techniques.

The question of which IC logic type to use is of major importance to the designer. He must make a comprehensive evaluation of the logic forms that are available and correctly choose the type that is best for his application. Trade-offs must be made between speed, power, and cost for each application.

Microelectronic devices, particularly IC's, have characteristics that are highly useful in redundancy. The extremely small size and weight (and frequently small power consumption) of these devices allow the addition of redundancy to a system without severe physical penalties. Also, the cost of integrated circuits does not increase linearly with increased complexity; a single package containing dual gates does not cost as much as two single gates in separate packages. Further as the complexity within a single package increases, reliability decreases slowly. In other words, dual gates in the same package are only slightly less reliable than a single gate and much more reliable than two separate gates. It is generally acknowledged that with constant electrical performance characteristics, all other characteristics of integrated circuits are superior to discrete parts arranged to perform the same function. There are, of course, severe restrictions on the electronic functions that can be performed by integrated circuits.

Once the equipment is designed and produced, it will be necessary to maintain it and provide spares. The repair level must be determined and a decision must be made whether to repair a module or discard it. In addition, the method of supplying spares must be chosen. These decisions must be made while the equipment is in the conception or design phase, not after production.

3.2 PACKAGING

3.2.1 General Requirements of the IC Package

One of the major considerations in applying integrated circuits to a particular system is selecting the proper package. In this regard, certain factors merit special attention by both the manufacturer and the user.

The device* manufacturer has primary responsibility for the reliability of the IC package, whether he fabricates the package or purchases it. The device manufacturer, then, must consider such characteristics as mechanical integrity, hermetic seal, lead stresses, etc.

The user of the device, of course, is concerned with package reliability; but he also investigates such characteristics as space efficiency and design flexibility.

General requirements of an IC package fall under the following headings:

- Mechanical integrity
- Electrical requirements
- Hermetic seal
- Thermal transfer
- Space efficiency
- Design flexibility
- Special requirements

3.2.1.1 Mechanical Integrity

During fabrication and operation, the IC package must withstand exposure to certain levels of shock, acceleration, vibration, soldering or welding heat, temperature cycling, thermal shock, moisture, and lead stress. Various tests are designed to ensure that the package will withstand the stresses encountered in device assembly (installing the chip in the package and performing the final seal), system fabrication (mounting and attaching the package to the mother board), and system operation.

3.2.1.2 Electrical Requirements

Electrical requirements of the IC package are as follows:

(1) The number of leads must be sufficient to accommodate a fairly complex circuit chip. Unused leads can be used for package tie-down or can be clipped off.

(2) The package must not have excessive distributed capacitance and inductance.

(3) In many applications, the package must be capable of providing electrical shielding.

*Device is defined as the integrated circuit plus its package.

3.2.1.3 Hermetic Seal

The internal environment of the IC package must remain constant; variations can cause failure. Although an exact correlation between leakage rate and device life is extremely difficult to establish, packages with a leakage rate of less than 10^{-8} cubic centimeters per second have proven to be reliable from a hermetic standpoint.

3.2.1.4 Thermal Transfer

The operating temperature of an integrated circuit is a major factor affecting its reliability. Of the three methods of heat transfer (Figure 3-1), conduction can generally transfer the most heat; thus the IC package should be constructed of materials that have high thermal conductivity.

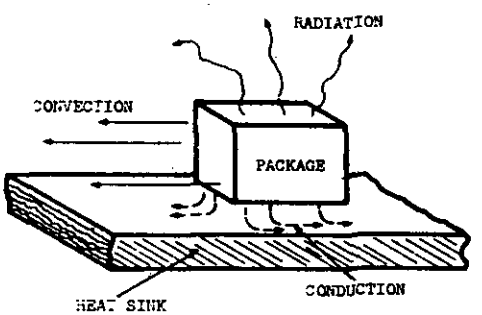


FIGURE 3-1

HEAT-TRANSFER MODES OF IC PACKAGE

3.2.1.5 Package Space Efficiency

The package should conform closely to the shape of its contents. Standardized packages are often unable to meet this criterion, but the poor fit can often be offset by an increase in circuit functions (i.e., more or larger chips).

3.2.1.6 Design Flexibility

The device user will require that the IC package have certain features that facilitate handling and assembly. These features, collectively called flexibility, include the following:

(1) The package must be easily handled during device manufacturing, testing, shipping, and fabrication into a particular equipment. Handling difficulties encountered during any of these phases will result in an increase in equipment costs and often a decrease in reliability.

(2) The package must be amenable to various mounting techniques (planar, stacked, on edge, etc.), to standard interconnection methods (flow-soldering, dip soldering, welding, or thermocompression bonding), and to various means of heat dissipation.

(3) The package should have high-density potential. Integrated circuits play a major role in microminiature systems, where high-density packaging is mandatory.

3.2.1.7 Special Requirements

The package should be resistant to corrosion or other physical change because of external reactive agents -- electrical fields, moisture, etc. In some cases, the package may also have to be immune to light radiation because of the type of circuit enclosed.

For electrical shielding, the package may have to be made of metal, with one of its leads electrically grounded to the case.

3.2.2 Available Package Types

3.2.2.1 Modified TO-5 Package

Modified TO-5 is a standard TO-5 transistor package that has been modified for IC's by increasing its number of external leads. Although the modified package has a new JEDEC TO number, the name "modified" is more popular. For clarity here, the modified TO-5 package will be defined as having more than four leads and with the mechanical outline dimensions of the TO-5 in Figure 3-5.

Modified TO-5 assembly methods are extensions of the techniques used in the production of standard transistor packages. Since the modified TO-5 package typically has 10 leads, compared with the 3 or 4 leads of the standard transistor package, its pin-circle diameter is slightly larger (0.230 inch vs. 0.200 inch). The modified TO-5 package is shown in Figure 3-2. Kovar, a nickel-iron alloy, is used for the leads and eyelet, and 7052 glass for the preform. The can may be German silver, Kovar, or nickel; but Kovar and 7052 glass are frequently used because of their similar coefficients of thermal expansion (Figure 3-3) and because they provide a matched seal.

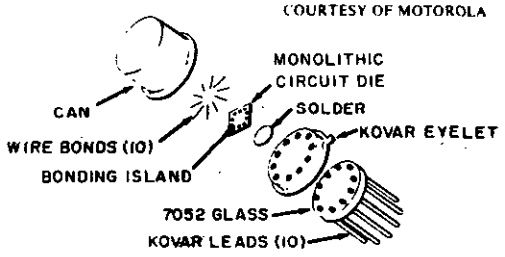


FIGURE 3-2
EXPLODED VIEW OF TO-5 PACKAGE

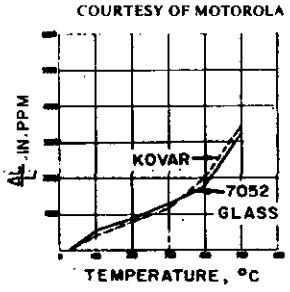


FIGURE 3-3
THERMAL EXPANSION CURVES

The Kovar leads are enclosed in the glass preform and assembled to the eyelet. The glass, leads, and eyelet are sealed by fusion in an oven at 1000°C. After this sealing process, the entire assembly -- called the header -- is cleaned; and the leads are clipped off to obtain the desired post height. The header is then plated with 0.0001 inch of gold cyanide. The monolithic die is bonded to the header by (1) inserting a gold-silicon preform between the die and the header, (2) placing a weight on top of the assembly, and (3) heating the entire header to

approximately 395°C. The gold-silicon eutectic is reached as a result of the heat and pressure; the die, preform, and header are fused together. The circuit is then wirebonded to the posts and baked in an inert atmosphere at 200°C to 300°C for approximately 30 minutes.

In the case of hybrid microcircuit using a ceramic disc, the pattern on the disc corresponding to its location is metalized with molymanganese. When the substrate is then gold-plated, the gold adheres only to the metalized area. The disc is fixed to the header by brazing, the dice are bonded to the metalized pads of the ceramic, and the dice are interconnected and wire-bonded to the posts.

After the circuit is checked electrically on the header, the can is welded to the flange of the eyelet. The welding is done in a controlled atmosphere to ensure stability of the circuit under operating conditions.

Figure 3-4 is a cutaway view of the modified TO-5 package, with heat-flow paths indicated. Most of the heat is conducted from the chip to the header assembly. The header assembly then conducts the heat radially to the lip and sides of the package, where it is transferred to the atmosphere by convection and radiation.

COURTESY OF MOTOROLA

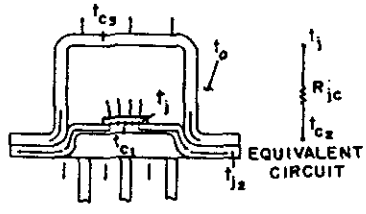


FIGURE 3-4

THERMAL PATHS IN TO-5 PACKAGE

The heat-dissipating capability of the modified TO-5 package can be improved by either of two methods. A metal container (with or without fins) can be fitted over the top of the package to increase its surface area, or forced-air cooling may be used.

The most significant thermal parameters in IC packages are the junction-to-case and junction-to-ambient thermal resistances (θ_{JC} and θ_{JA}, respectively). Values of these parameters vary among manufacturers but are typically 80°C per watt for θ_{JC} and 180°C per watt for θ_{JA} for the modified TO-5 package. The θ_{JC} of a hybrid microcircuit mounted in the modified TO-5 package is approximately 10 percent higher than that of a standard circuit because of the ceramic disc between the chips and the header.

Modified TO-5 Space Efficiency. The space efficiency of an IC device utilizing the modified TO-5 package is generally poor, since the circuit usually occupies only a small fraction of the package height (Figure 3-4). One means of reducing the void space is to apply hybrid techniques. Each silicon circuit is mounted on a ceramic wafer having a diameter approximately that of the modified TO-5 chip-mounting area. The ceramic discs are then stacked in layers to a height sufficient to fill the package (0.260 max). The individual monolithic circuits are interconnected with fine gold wires and terminated at the external leads. This method of increasing space efficiency requires elaborate fabrication equipment and is quite expensive.

The space efficiency problems associated with the modified TO-5 have led to the development of a low-profile variation of this package. This design, known as the TO-80, is illustrated in Figure 3-5, with the modified TO-5 shown for comparison. In spite of its greater inherent space efficiency, the TO-80 has not approached the level of usage of the modified TO-5, because of the TO-80's higher cost and lower power-dissipation capability.

Also detrimental to the space efficiency of the modified TO-5 is the ineffic ent use of lateral area. The IC die is always square or rectangular, since it is impractical to perform die separation in a manner that yields circular dice. Cornered dice do not conform closely to the circular mounting area on the modified TO-5 header (Figure 3-6).

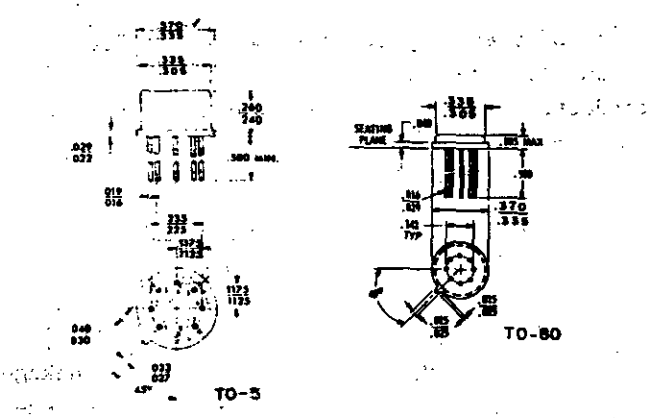


FIGURE 3-5

MODIFIED TO-5 AND TO-80 DESIGNS

COURTESY OF GENERAL ELECTRIC

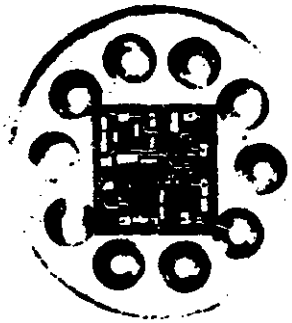


FIGURE 3-6

IC DIE IN MODIFIED TO-5 PACKAGE

Modified TO-5 Design Flexibility. The IC package should have features that facilitate handling and assembly. The modified TO-5 has become a standard package; thus most semiconductor and electronic-system manufacturers have the equipment for handling it efficiently. Also, the fixtures used in manufacturing testing, shipping, and installing the package are available from a variety of vendors at reasonable prices.

Another design-flexibility requirement is that the package be amenable to various mounting techniques. The modified TO-5 package can be either plugged into or embedded in a board (Figure 3-7). The former method is preferred. In either method, the interconnection pattern required for the package leads creates an area that must be avoided by other printed conductors on both sides of the board. The surface (plug-in) method yields insufficient clearance between the islands to route additional circuitry. With the embedding method, there is sufficient area between the islands (because of the increased diameter of the interconnection pattern) for additional conductors; however, the 0.325-inch hole required

to accept the package poses a restriction. Thus the interconnection-pattern problem and the fact that there are only two mounting methods restricts the flexibility of the modified TO-5 package.

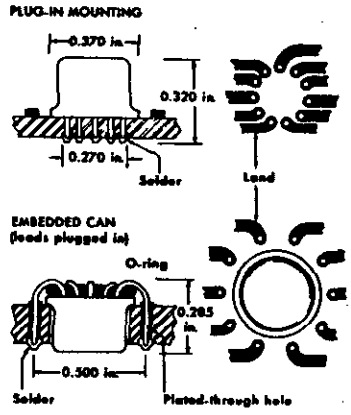


FIGURE 3-7

TWO METHODS OF MOUNTING TO-5 PACKAGE ON BOARD

Generally, the modified TO-5 package is secured to the mounting area by a solder-dip or reflow process. Welding techniques such as parallel gap and axial are seldom used for the modified TO-5 package because of the cylindrical geometry of the individual leads and the overall package.

The modified TO-5 package has a low-density potential relative to the flat package. The effective area of the modified TO-5 base is 0.11 square inch, and its volume is 0.029 cubic inch. As a result, its theoretical density potential (with surface mounting) is approximately 35 circuits per cubic inch, or 60,500 per cubic foot. However, these theoretical figures cannot be achieved due to the wasted space created by the cylindrical form factor, the need for interconnection space, and, frequently, the need for heat dissipating devices. A maximum of 8 surface-mounted modified TO-5 packages are possible per square inch, but it's unlikely that such density could be achieved in any practical arrangement, even with multilayer boards. A more realistic figure is 5 packages per square inch.

3.2.2.2 Flat Pack

Many types of IC flat packages are being produced, in various sizes and materials. These packages are available in square, rectangular, oval, and circular configurations with 10 to 60 external leads. The package can be made of metal, ceramic, epoxy, glass, or combinations thereof. Only the ceramic flat pack will be discussed, since it is representative of all flat packs with respect to general package requirements.

Construction of the ceramic flat pack is illustrated in Figure 3-8. The Kovar leads, which are attached within a frame, are inserted into the castellations of the ceramic mounting base. A 7052 glass is fused to the ceramic base and Kovar leads, creating a hermetic seal. (The castellations in the mounting base provide additional strength to the body-to-lead seal.)

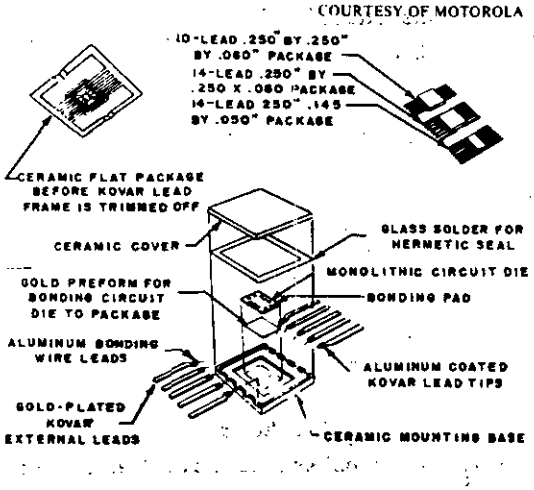


FIGURE 3-8

EXPLODED VIEW OF CERAMIC FLAT PACK

After the leads are sealed to the mounting base, a rectangular area on the inside bottom of the base is treated with molymanganese. This metalization provides a surface suitable for bonding the IC chip to the base. The lead frame is then cut away from the secured leads, and these leads and the metalized area in the bottom of package are gold-plated. The die is then attached by gold-silicon eutectic bonding.

The die bonding is followed by bonding gold or aluminum wires between the bonding islands on the IC die and the inner portions of the package leads.

Following the wire bonding, a glass-solder preformed frame is placed on top of the mounting base. One surface of the ceramic cover is coated with Pyroceram* glass, and the cover is placed on top of the mounting base. The entire assembly is placed in an oven at 450°C; this causes the glass solder and Pyroceram to fuse and seal the cover to the mounting base.

Flat-Pack New Fabrication Technique. Considerable effort is being devoted to eliminating the fine wires connecting the circuit to the Kovar leads. The omission of these wires would materially reduce the cost of integrated circuits by eliminating the bonding labor and would enhance reliability by precluding a potential cause of circuit failure.

A promising fabrication technique is the face-down (flip-chip) mounting method, in which conductive patterns are evaporated inside the package before the die is attached. These patterns connect the external leads to bonding pads on the interior surface of the package. The pads are bonded to appropriate pedestals on the die. The location of these pedestals corresponds to the bonding pads on the interior surface of the package. Alternatively, the pedestals may be located on the surface of the package and the pads on the die. A number of bonding methods have been developed. The most prominent among them is ultrasonic bonding, although

*Trademark of the Corning Glass Company.

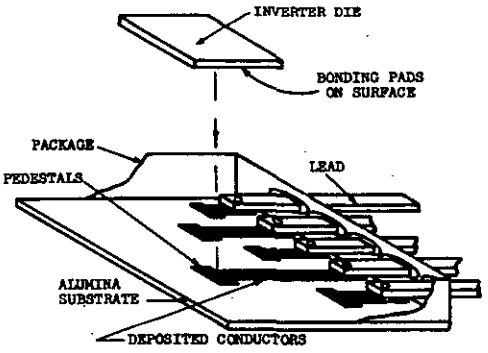


FIGURE 3-9

FLIP-CHIP TECHNIQUE

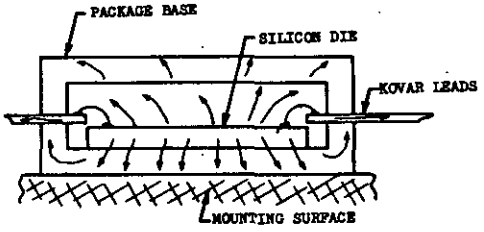


FIGURE 3-10

CUTAWAY VIEW OF FLAT PACK, SHOWING HEAT-FLOW PATHS

thermal-compression bonding and adhesive bonding (such as solder reflow) are also used. This technique is illustrated in Figure 3-9.

Flat-Pack Thermal Characteristics. Figure 3-10 is a cutaway view of the flat pack, with the heat flow indicated. Heat generated by the silicon die is both radiated into the interior and conducted into the mounting base. The heat quickly reaches the package outer surface because of the close contact of the die with the package and the high thermal conductivity of the package material (alumina). The flat pack has a lower θ_{JC} than the modified TO-5 package, with a typical value of 40°C per watt (vs. 80°C per watt for the TO-5). However, its θ_{JA} is higher than that of the modified TO-5 package (typically 250 vs. 180°C per watt) because of its relatively small surface area.

The low θ_{JC} of the flat pack allows efficient removal of heat from the pack by conduction to the heat sink through a thin strip of metal placed under the pack. For this reason, heat transfer by conduction is commonly used today in high-density packaging.

Flat-Pack Space Efficiency. The internal volume of the flat pack can be used efficiently for integrated circuits. The rectangular flat pack conforms closely to the rectangular or square silicon IC die (Figure 3-11). The external leads are brought into the pack through the walls; this leaves the base area unobstructed and available for large dice. The thin package helps to minimize the internal unused space. The silicon circuit often consumes 80 percent of the available mounting space inside the flat pack.

Flat-Pack Design Flexibility. Holding the flat pack secure for fabrication, testing, shipping, and installation constitutes a major problem. The package is very small (1/8 × 1/4 × 1/16 inch), and its leads are quite delicate (measuring 0.004 × 0.012 inch). Fixtures used to retain the flat package are discussed in detail in Chapter 4.

COURTESY OF ELECTRONIC INDUSTRIES

FIGURE 3-11

SPACE EFFICIENCY OF FLAT PACK

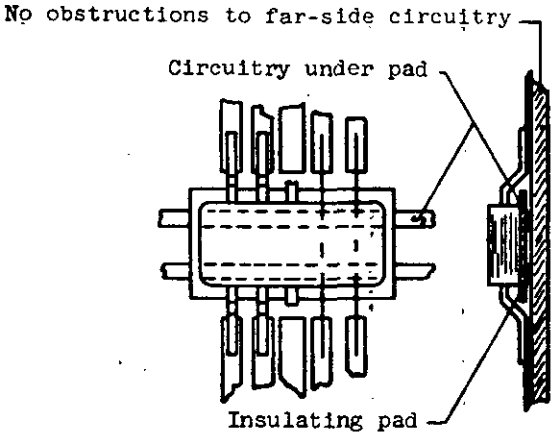


FIGURE 3-12

INTERCONNECTION PATTERN FOR FLAT PACKAGE

Design flexibility is greatly enhanced if alternative means can be employed for mounting the package to a surface. The flat pack can be mounted in several ways: surface, stacked, or on-edge, with the first of these generally preferred.

One interconnection pattern for the flat pack is shown in Figure 3-12. The package leads are attached directly to the printed conductor, and additional printed circuitry is routed underneath the package. An insulating substrate is placed between the printed conductors and the package during assembly, and an epoxy is applied to secure the package in place. Naturally, the insulating pad causes an increase in thermal resistance (by about 5°C per watt) between the package and the mounting area.

Another feature of this type of mounting is the absence of obstructions to printed circuitry on the opposite side of the mounting board. In many designs, flat packs are mounted on both sides of the board, directly opposite each other. The package leads are attached to the printed conductors by soldering or welding.

The flat pack has high-density potential because of its small size, unique interconnection pattern, and adaptability to micro-miniature heat-dissipating techniques. Excluding leads, the area of a 1/8 x 1/4 x 1/16 flat pack is 0.031 square inches; its volume is 0.00195 cubic inches.

The theoretical density potential of the flat-pack is 512 per cubic inch. As with the modified TO 5

package, this density potential cannot be achieved in a practical arrangement. With 1/8-inch lead lengths, a maximum of 10 circuits can be placed on 1 square inch with surface mounting.

3.2.2.3 Dual In-Line (DIP) Package

The dual in-line package was designed primarily to overcome the difficulty associated with handling packages and inserting them into mounting boards. The dual in-line package is easily hand- or machine-inserted and requires no spreaders, spacers, insulators, or lead forming. It is also possible to field-service the device with standard hand tools and soldering irons. This package is finding wide use in commercial applications as a plastic package, as well as in a number of military systems as a ceramic package. The dual in-line package is shown in Figure 3-13.

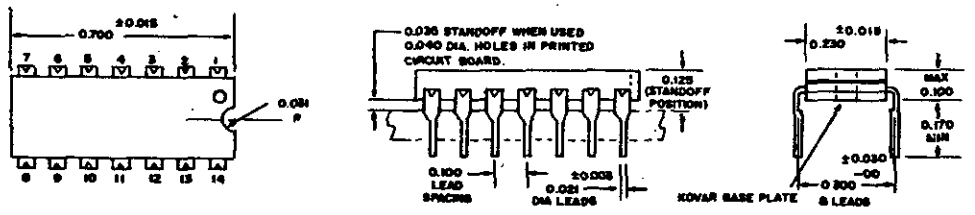
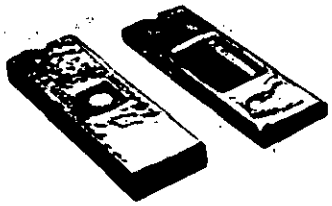


FIGURE 3-13

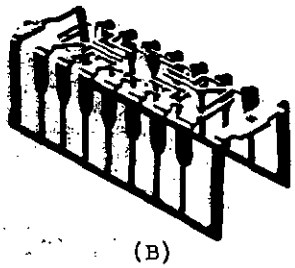
OUTLINE DRAWING OF THE DUAL IN-LINE PACKAGE

The progressive stages in the assembly of a ceramic DIP are illustrated in Figure 3-14. The integrated-circuit die is sandwiched between the two ceramic elements (A). The element on the left is the bottom half of the sandwich and will hold the integrated circuit die. The ceramic section on the right is the top of the sandwich; the large well is intended to protect the IC die from mechanical stress during sealing operations. Each of the ceramic elements is coated with low-melting-temperature glass for subsequent joining and sealing. The lead frame is kovar and is shown stamped and bent into its final shape (B). The excess material is intended to preserve pin alignment. The holes at each end are for the keying jig used in the final sealing operation. The lower half of the ceramic package is forced into the lead frame (C). The die is mounted in the well and leads are attached. The top ceramic element is bonded to the bottom element (D) and the excess material is removed from the package (E).

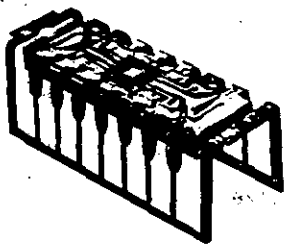
The ceramic DIP is processed singularly while the plastic DIP is processed in quantities greater than one. After complete processing of the packages they are sawed apart. The plastic package also uses a kovar lead frame but the leads are not bent until the package is completed. Molded plastic is used to encapsulate the IC die and unlike the ceramic package there is no void between cover and die.



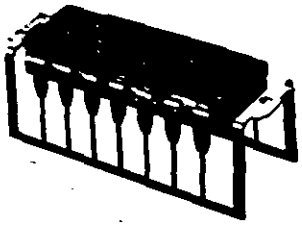
(A)



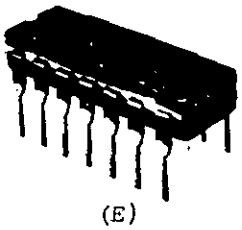
(B)



(C)



(D)



(E)

FIGURE 3-14

DIP ASSEMBLY

DIP Thermal Characteristics. Ceramic is a good heat conductor; plastic is a poor heat conductor. However, heat transfer in both packages is primarily accomplished with the Kovar lead frame; and the thermal impedance from chip to ambient is about the same. If the tab on the Kovar lead frame (or other appropriate means) is used to heat-sink the device, values for θ_{JA} of 50°C/W are possible. In most applications heat sinking is not used, and the value of θ_{JA} will range typically from 150°C/W to 200°C/W.

Plastic packages have wide industrial applications where the specified temperature range is 0 to +70°C. Military applications require a -55 to +125°C temperature range and ceramic is usually required. (Ceramic can also be leak tested, while molded plastic cannot.)

DIP Space Efficiency. The DIP is the least space-efficient IC package in wide use. It is approximately 0.30 inch wide and 0.70 inch long, and its height above the mounting board is about 0.25 inch. Only about 3 packages can be accommodated per square inch; this is about half the packaging density of the modified TO-5.

Although the package is not as small as other package types, it is easy to handle and can be precisely located. Each lead may be individually welded or soldered to a printed circuit board.

Figure 3-15 shows DIP's mounted on a printed circuit board.

COURTESY OF AUGAT

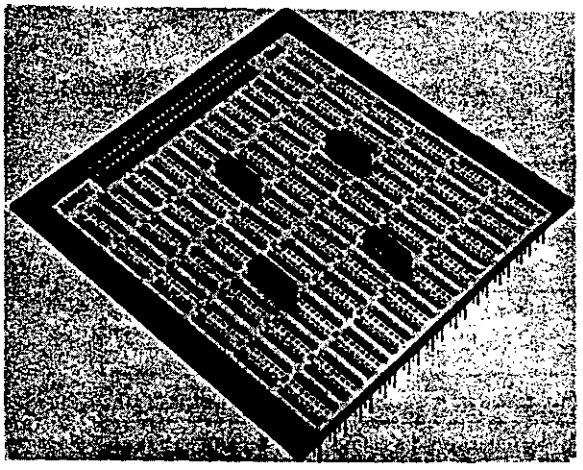


FIGURE 3-15

MOUNTED DUAL-IN-LINE PACKAGES

COURTESY OF AUGAT

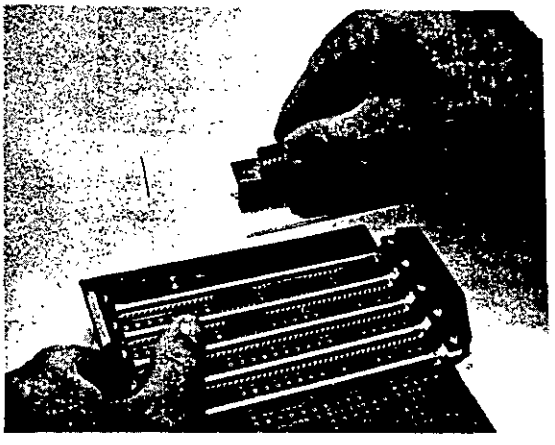


FIGURE 3-16

MOUNTING DIP'S WITH THE GUIDE MATRIX AND ALIGNMENT TOOL

DIP Design

Flexibility - The size of the DIP facilitates handling, and no special lead adapters are required. The exposed leads can be probed from both the top and the bottom of the mounting board for troubleshooting. The rugged construction of the leads allows probing without damage to the leads. The wide lead spacing minimizes the hazards of shorting.

The DIP is available with 14 and 16 leads, but the 14-lead version is the most widely used. At least one manufacturer believes that DIP packages can be constructed with as many as 50 leads.

A guide matrix and aligning fixture (Figure 3-16) can be used for hand insertion of the dual-in-line packages. The guide matrix is positioned over the hole pattern in a PC board. The package is placed in the aligning fixture and then located over the appropriate hole pattern and inserted into the board with slight pressure. This process is repeated until all packages are mounted -- at which time the guide matrix slides off the board without dislocating the packages.

3.2.3 Comparison of the IC Packages

The DIP, flat pack, and modified TO-5 are compared in Table 3-1.

3.2.3.1 Standardization

Many types of package are available for use with integrated circuits, but standardization is difficult because of the rapid advance of the technology. The modified TO-5 and the 1/8 by 1/4-inch flat pack are now used in larger quantities than any of the other package types (plastic DIPS are frequently used in industrial applications). Their popularity is probably the result of the

large quantity of test data available on them. The current trend in commercial systems favors the modified TO-5 or the DIP.

TABLE 3-1

COMPARISON OF IC PACKAGE TYPES

Elements of Merit	Modified TO-5	Flat Pack	DIP
Hermeticity	Excellent	Good	Good (Ceramic)
Number of Leads	12 Maximum	14 usually	14 (usually)
Thermal Transfer (Typical)	θ_{JC} = 80°C/W θ_{JA} = 180°C/W	θ_{JC} = 40°C/W θ_{JA} = 250°C/W	θ_{JC} = 100°C/W θ_{JA} = 150°C/W
Handling Convenience	Good	Fair to poor	Excellent
Surface Mounting Capability (packages per sq. in.)	5	10 (1/4 x 1/8)	3
RFI Shielding	Good	Poor	Poor

The Electronic Industries Association(EIA) has registered eight flat-pack designs (10 and 14 leads) in four basic configurations (Table 3-2). Thus it appears that some package standardization is possible, though new types will continue to enter the market for some time to come.

TABLE 3-2

EIA-REGISTERED FLAT PACKAGES (0.035-INCH THICKNESS)

Designation	Package Dimensions (inches)	Number of Leads	Designation	Package Dimensions (inches)	Number of Leads
TO-84	0.250 x 0.125	14	TO-88	0.330-0.350 x 0.240-0.260	14
TO-85	0.250 x 0.160-0.185	14	TO-89	0.250 x 0.125	10
TO-86	0.250 x 0.250	14	TO-90	0.250 x 0.160-0.185	10
TO-87	0.375 x 0.250	14	TO-91	0.250 x 0.250	10

3.3 INTERCONNECTION TECHNIQUES

New generations of electronic systems will require increasingly dense and complex interconnections, and the trend toward higher operating frequencies will necessitate extremely careful design, even of single conductors. Conductor width, spacing, and dielectric material are electrically critical and in many cases directly related to signal propagation, noise figure, pulse shape, and other

system parameters. Therefore, it is mandatory that designers of systems and circuits become thoroughly familiar with new interconnection techniques. This section is a discussion of the three most used techniques: conventional printed-wiring boards (single- and double-sided), welded-wire planes, and multilayer printed wiring boards.

3.3.1 Conventional Printed-Wiring Boards

The conventional printed-wiring board consists of a glass-epoxy insulating base on which a copper interconnection pattern has been etched (Figure 3-17). The photolithographic etching process, which leaves a copper pattern approximately 0.003 inch thick, is described in MIL-STD-275A.

3.3.2 Welded-Wire Planes and Assemblies

Welded-wire planes are two layers of wire or metal ribbon (usually nickel) positioned at right angles to each other and separated by a thin plastic film. The layers are welded together through prepunched holes in the plastic film. Where external connections are required, the metal ribbons are welded to connect pins. An assembly is then encapsulated in a hard epoxy. The completed unit is shown in Figure 3-18.

COURTESY OF MOTOROLA

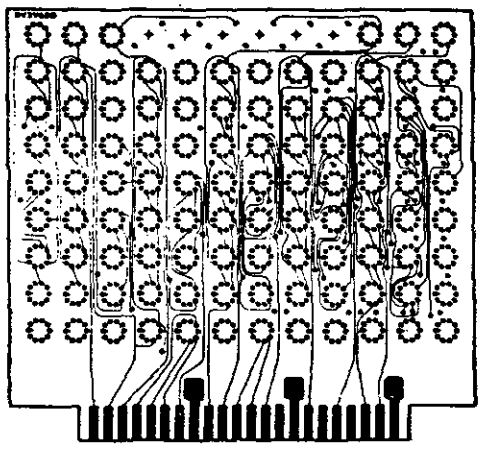


FIGURE 3-17
PRINTED WIRING BOARD

COURTESY OF ELECTRONIC DESIGN NEWS

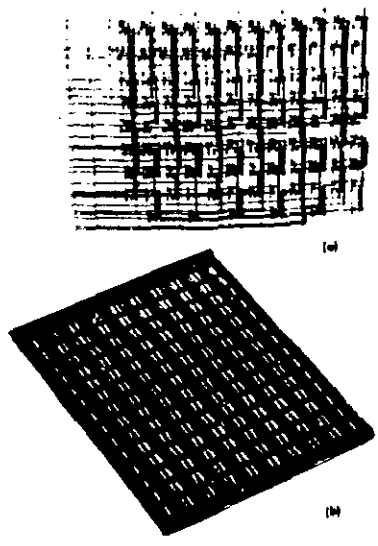


FIGURE 3-18
WELDED WIRE ASSEMBLY

Welded wiring permits 400 or more connections per square inch. Welded-wire networks must be encapsulated, since they would otherwise offer little or no support to attached components; therefore, they weigh more than an equivalent multilayer printed-wiring card. One plane is approximately 0.007 inch thick.

3.3.3 Multilayer Printed-Wiring Boards

Multilayer printed wiring is emerging as the solution to interconnection problems associated with high-density packaging. Consequently, the balance of this discussion will be concerned with this particular technique. Multilayer boards are used (1) to save weight and space in interconnecting circuit modules, (2) to eliminate costly, complicated wiring harnesses, (3) to provide shielding of a large number of conductors, (4) to provide uniformity in conductor impedance in high-speed switching systems, and (5) to provide greater wiring density on boards.

Figure 3-19 illustrates the various individual boards that are mated to form the multilayer unit. Although all multilayer boards are constructed in a similar manner, there are various methods for interconnecting the circuitry from layer to layer. Three proven methods are discussed.

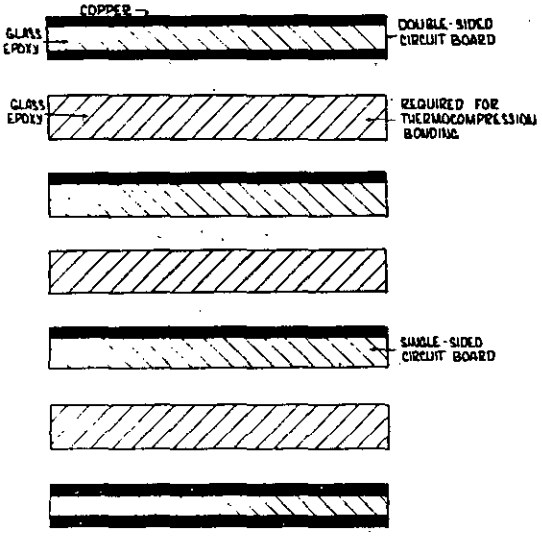


FIGURE 3-19

MULTILAYER BOARD CONSTRUCTION

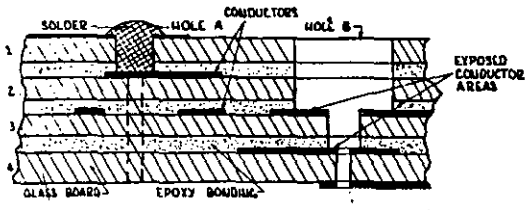


FIGURE 3-20

CLEARANCE-HOLE INTERCONNECTION TECHNIQUE

3.3.3.1 Clearance-Hole Method

In the clearance-hole process, a hole is drilled in the copper island (terminating end) of the appropriate conductor on the top layer, providing access to a conductor on the second layer (Figure 3-20, hole A). The clearance hole is filled with solder, and the desired connection is completed. Usually the hole is extended through the entire assembly at the connection site. This small hole is necessary for the solder-flow process normally used with this interconnection method.

To interconnect conductors located several layers below the top, a stepped-down-hole process is employed (Figure 3-20, hole B). Before assembly, a clearance hole is provided down to the first layer to be interconnected (in Figure 3-20, through layers 1 and 2 to layer 3). The first layer to be interconnected (3) is predrilled with a smaller hole than that of layers 1 and 2; succeeding layers to be connected have progressively smaller clearance holes. After assembly,

a portion of the island areas is exposed on the conductors that are to be interconnected. The stepped-down hole is then filled with solder, completing the connection. Obviously, the larger the number of interconnections required at one point, the larger must be the diameter of the clearance hole on the top layer. Large clearance holes on the top layer result in less space for components and therefore reduce packaging density.

3.3.3.2 Plated-Through-Hole Method

The plated-through-hole method of interconnecting conductors is illustrated in Figure 3-21. The first step is to assemble, temporarily, all the layers into their final configurations. After holes corresponding to required connections are drilled through the entire assembly, the unit is disassembled. The internal walls of those holes to be interconnected are plated with metal (0.001 inch thick), which, in effect, transfers the conductor on the board surface into the hole itself (the process is identical to that used for standard printed-wiring boards). The boards are then reassembled and permanently fixed into their final configuration; all the holes are then plated through with metal.

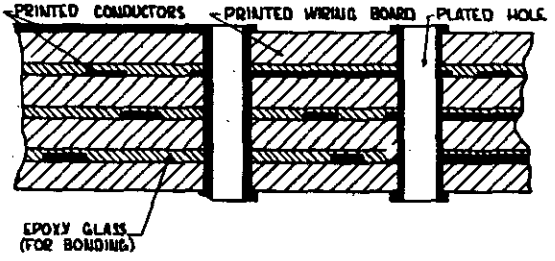


FIGURE 3-21

PLATED-THROUGH TECHNIQUE

3.3.3.3 Layer-Build-Up Method

In the layer-build-up method, conductors and insulation layers are alternately deposited on a backing material (Figure 3-22). This method yields all-copper interconnections between layers and therefore is more reliable than the other two techniques.

3.3.3.4 Comparison of Methods

Each interconnection technique has its advantages and limitations, as follows:

(1) Cost. For small quantities the clearance-hole method is the cheapest since it requires a minimum of equipment and engineering capability. However, the cost of applying this method goes up drastically for larger quantities -- even though fabrication is simple, the operations are quite time-consuming. The plated-through-hole method is usually moderate in cost for both small and large quantities. Built-up layers are generally more expensive than the other two types, regardless of the quantities involved. However, various interconnection designs can have an effect on the multilayer-board price. For example, in the built-up layers, interlayer connections are made without bringing the connections to the surface (as is necessary in the other two methods), allowing greater flexibility in wiring layout. Thus fewer layers are needed for the complete board, and a price lower than that for the plated-through-hole method may result.

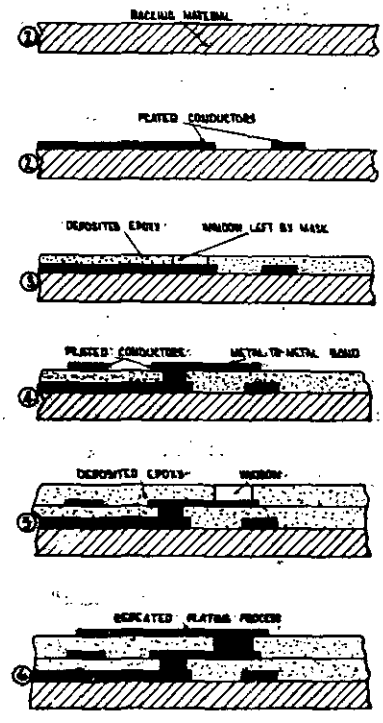


FIGURE 3-22

LAYER BUILD-UP TECHNIQUE

(2) Density. The build-up method yields the highest packaging density, since its internal interconnections do not require corresponding holes in other layers. The clearance-hole technique offers the lowest packaging density of the three methods because of the oversized holes required.

(3) Reliability. As previously mentioned, the built-up layers form the most reliable interconnections, because of their solid-copper makeup. The clearance-hole method is the next most reliable; the plated-through technique the least reliable. In the latter case, uniform metallization has proven difficult to achieve.

(4) Maintainability. In general, interconnections made by the clearance-hole method are the least difficult to rework. Plated-through connections are second; built-up layers are third. In the built-up layers, it is usually difficult -- if not impossible -- to rework any of the internal connections.

Table 3-3 summarizes the advantages and limitations of the three interconnection techniques.

TABLE 3-3

COMPARISON OF MULTILAYER-BOARD INTERCONNECTION TECHNIQUES (RELATIVE RATING, WHERE 1 IS HIGHEST)

Technique	Cost*	Density	Reliability	Ease of Repair
Clearance Hole	3	3	3	1
Plated-Through Hole	2	2	3	2
Built-Up Layers	1	1	1	3

*Small quantities.

3.3.4 Package Tie-Down Techniques

For connecting modules or components to multilayer boards, the conventional method is to solder their leads or pins into the holes of the boards. Other means of component attachment have been tried successfully. These include resistance welding of component leads to tabs, pins, or eyelets protruding over the surface of the boards; and parallel-gap welding to tabs made of nickel, Kovar, special foils, or plain copper on the board surface (Figure 3-23). The common hand, dip, and wave soldering processes have been supplemented with techniques such as resistance microsoldering, controlled-heat-zone soldering, and hot-air soldering.

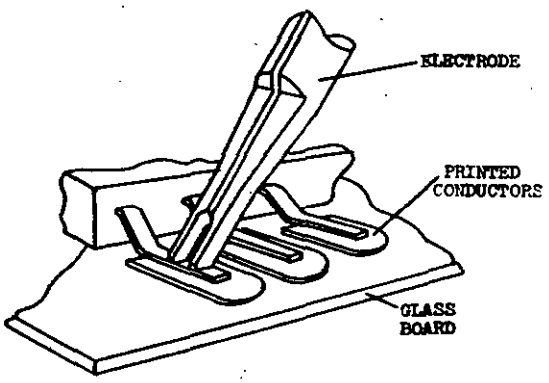


FIGURE 3-23

PARALLEL GAP WELDING

3.3.5 Multilayer-Board Fabrication

A six-bit parallel adder, containing integrated circuits mounted in the modified TO-5 package, will be used to illustrate the multilayer assembly process. A diagram of the interconnections required for each IC is shown in Figure 3-24. Figure 3-25 illustrates the probe-side and component-side signal layers of the multilayer board. Sandwiched between these two signal planes are four voltage planes that distribute voltages to the appropriate pins. Examination of the pin layouts on the individual circuits will show that bias-voltage inputs are well standardized for the various pins; for example, V_{CC} is distributed through pin 3 on the gates, flip-flops, etc. This standardization of pin breakouts lends itself to the multilayer technique shown in Figure 3-26. The smaller openings in these patterns are holes that allow the TO-5 leads to go through unshorted. The dark areas on the boards are copper. Note that on each of the voltage-distribution boards one particular pin is shorted to the copper area.

Figure 3-27A shows the bias-voltage plane that distributes the V_{BB} voltage to the logic unit. This distribution is accomplished by connecting pin 1 (output) of the bias-drive unit to pin 1 (bias input) of all logic units. The V_{BB} plane is subdivided into seven smaller planes, with one bias driver supplying the necessary voltage for each plane.

Figure 3-27B shows the termination-voltage plane, which distributes the V_{TT} voltage generated by the line drivers to the terminating resistor of long signal lines through pin 5. The V_{TT} plane is subdivided into five smaller planes, with one line driver supplying the voltage for each plane.

COURTESY OF MOTOROLA

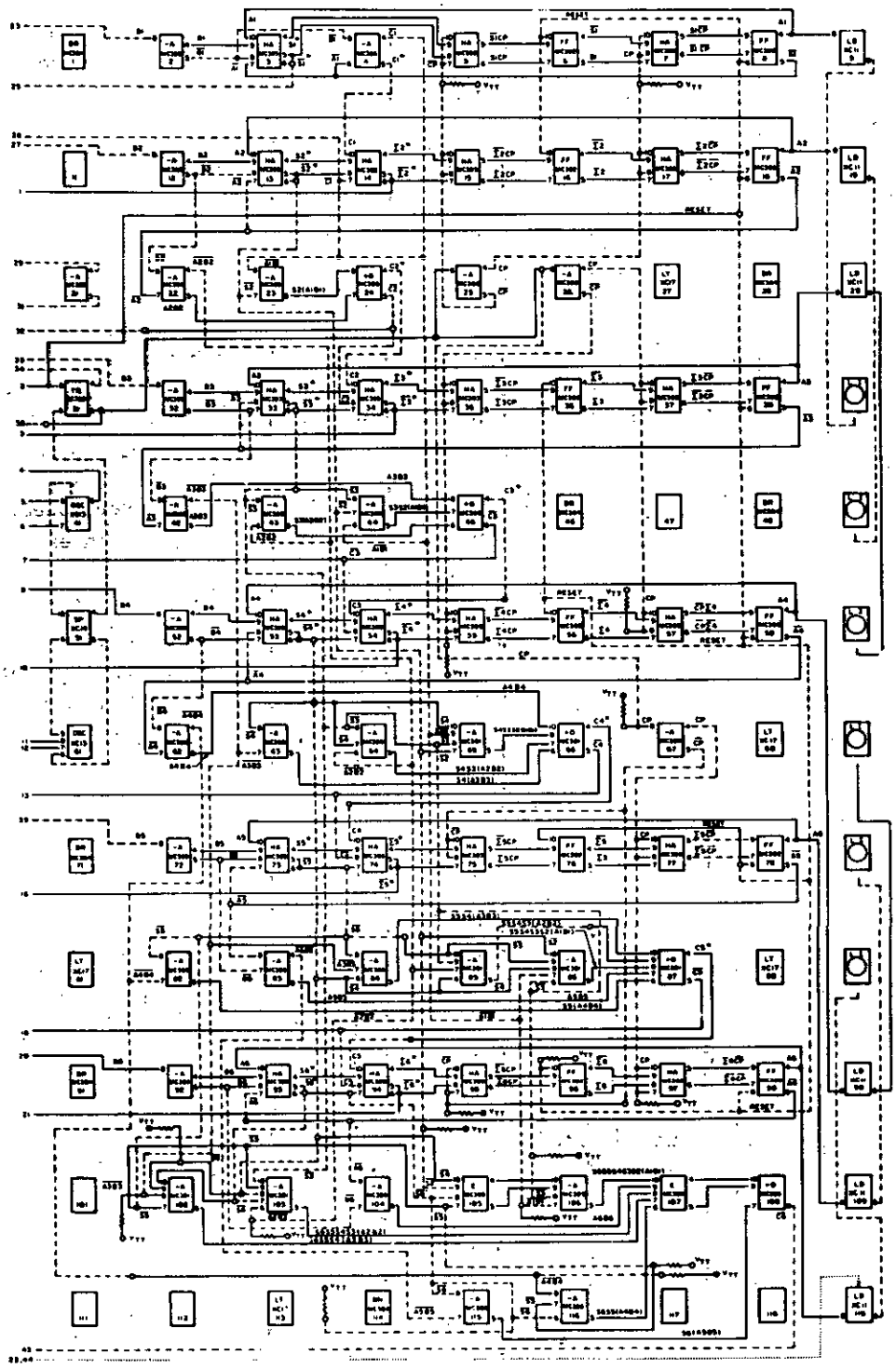
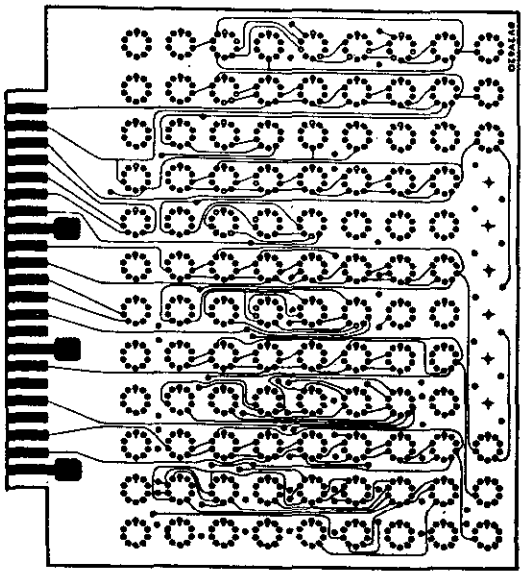


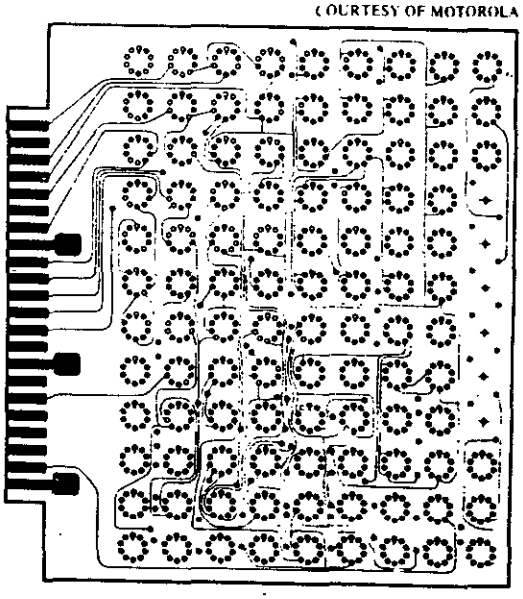
FIGURE 3-24

INTERCONNECTIONS FOR 6-BIT PARALLEL ADDER

COURTESY OF MOTOROLA



Probe-Side Signal Layer

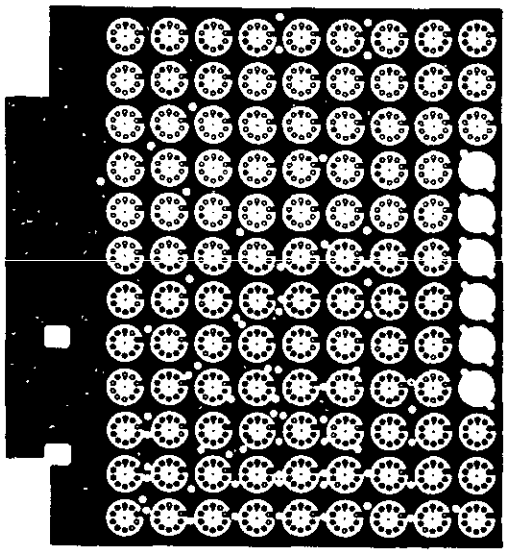


Component-Side Signal Layer

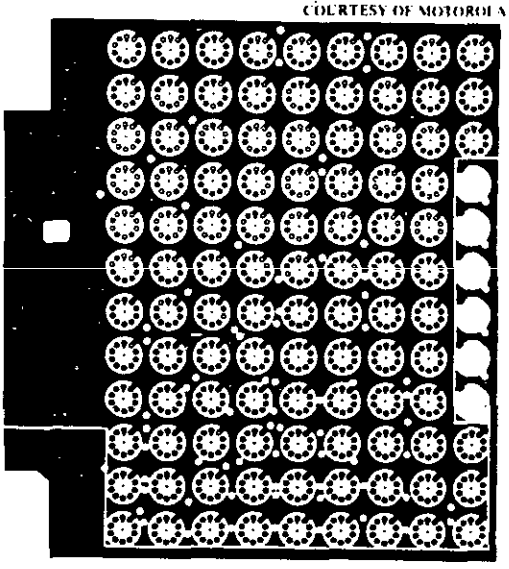
FIGURE 3-25

PROBE-SIDE AND COMPONENT-SIDE SIGNAL LAYERS

COURTESY OF MOTOROLA



V_{CC} VOLTAGE DISTRIBUTION LAYER. PLANE NO. 4

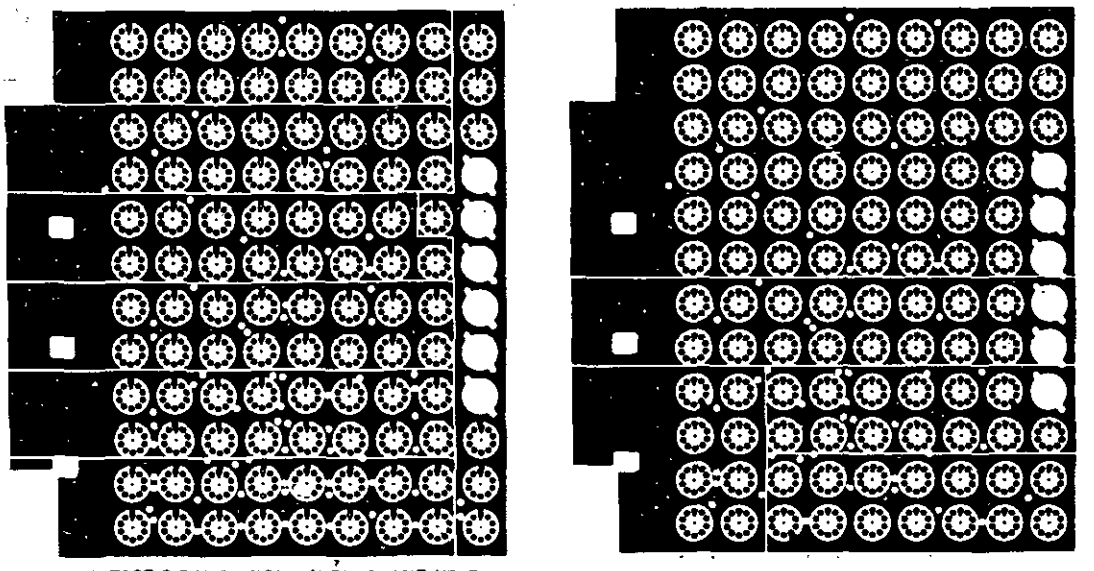


V_{EE} VOLTAGE DISTRIBUTION LAYER. PLANE NO. 2

FIGURE 3-26

VOLTAGE DISTRIBUTION LAYERS

COURTESY OF MOTOROLA



V_{BB} VOLTAGE DISTRIBUTION LAYER, PLANE NO. 3
(a)

V_{TT} VOLTAGE DISTRIBUTION LAYER, PLANE NO. 5
(b)

FIGURE 3-27

VOLTAGE DISTRIBUTION LAYERS

COURTESY OF MOTOROLA

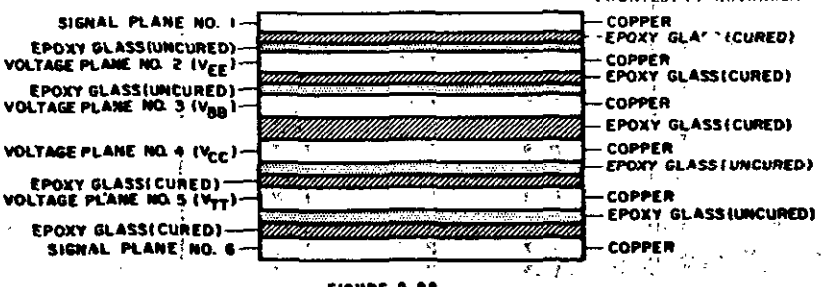


FIGURE 3-28

CROSS SECTION OF BONDED MULTILAYER BOARD

Figure 3-28 is a cross section of the multilayer board, with the various layers identified according to function and material. The bonding of these boards requires accurate registration and a thermocompression technique. The completed multilayer board is shown in Figure 3-29.

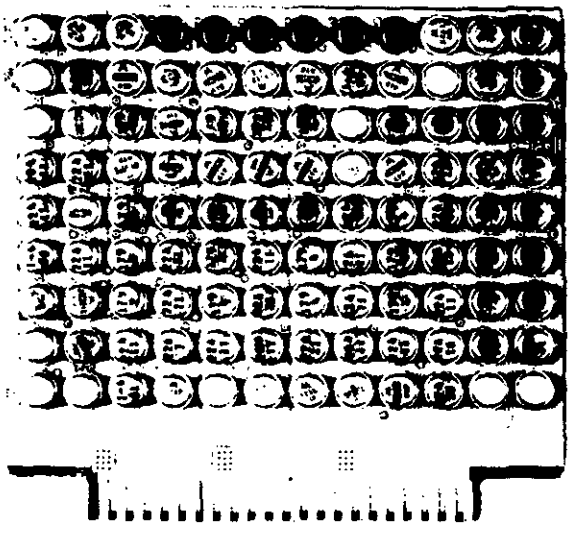


FIGURE 3-29

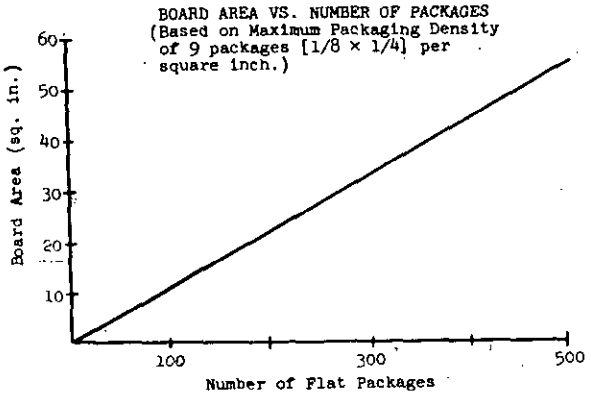
SIX-BIT PARALLEL ADDER

3.3.6 Density

The maximum density of modified TO-5 packages is about five per square inch on two-sided boards and about 8 per square inch on multilayers. The corresponding densities for the smaller flat packs are roughly 6 and 10 packs per square inch. An increase in the number of components on a multilayer board -- with the same component density maintained -- requires an increase in the number of layers as well as in surface area. The additional layers are needed to provide the capacity for the added interconnections. For example, to connect 25 flat packs, three to four layers are usually sufficient (1- by 3-inch board); to interconnect 50 to 60 flat packs, between five and six layers might be required (2- by 4-inch board); to interconnect 450 flat packs, 12 layers might be required (6- by 9-inch board). However, an increase from 5 to 10 layers will increase the cost of multilayer boards (made by the plated-through-hole method) by only 30 to 40 percent. This explains the tendency toward the production of larger boards. Figure 3-30 shows the required number of layers and required board areas as a function of the number of flat packages.

3.3.7 Heat Distribution

Equalizing the distribution and removal of heat in systems utilizing integrated circuits is a major design consideration. A typical heat-sink arrangement for multilayer boards is shown in Figure 3-31. The heat-sink layer is usually about 0.006 inch of copper, added (quite inexpensively) during the normal plating stage of manufacturing. A method for removing heat from closely assembled modified TO-5 packages is shown in Figure 3-32.



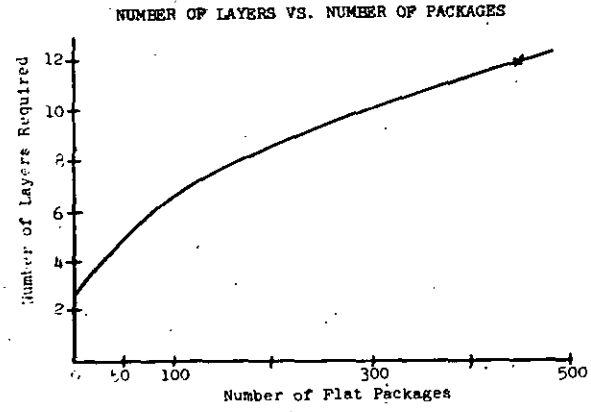


FIGURE 3-30

REQUIRED BOARD AREA AND NUMBER OF LAYERS AS FUNCTIONS OF NUMBER OF FLAT PACKAGES

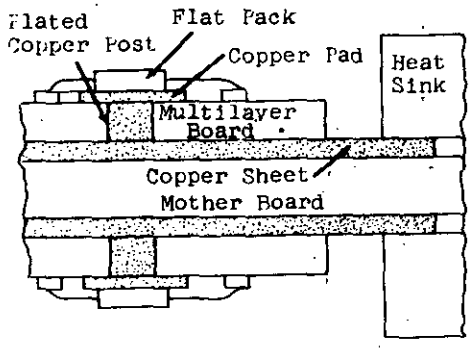


FIGURE 3-31

TYPICAL HEAT-SINK ARRANGEMENT FOR MULTILAYER BOARDS

3.3.8 Electrical Characteristics of Multilayer Boards

The current-carrying capacity of conductors is usually more than adequate for milliwatt circuitry interconnected by multilayer boards. The ohmic resistance is *not* a highly important factor, except in cases where conductive paths are very long. For an approximate calculation of conductor resistance, the following values can be used:

Foil Thickness	Ohms Per Foot Length Per Mil Width of Conductor
1-ounce copper	6
2-ounce copper (standard)	3
3-ounce copper	2

Thus a 2-ounce copper conductor 0.01 inch wide has an approximate resistance of 0.3 ohm per foot; the same conductor 0.020 inch wide has a resistance of 0.15 ohm per foot.

COURTESY OF MOTOROLA

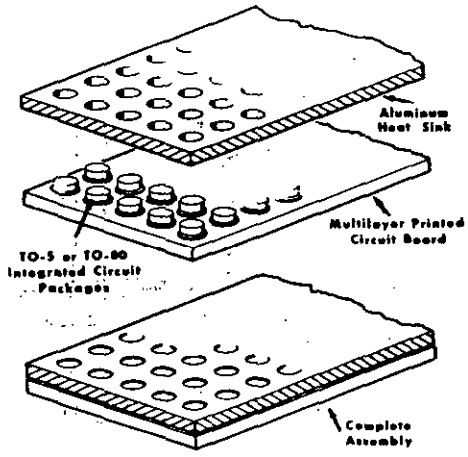


FIGURE 3-32

METHOD FOR REMOVING HEAT FROM CLOSELY ASSEMBLED TO-5 PACKAGES

A more important electrical aspect of multilayer boards is the distributive capacitance between conductors. This capacitance is quite small between two conductors in the same layer, but is sufficiently large between conductors located directly opposite each other on adjacent layers that it must be accounted for in high-frequency designs. The following general equation can be used to obtain an approximate value of distributive capacitance, C:

$$C = \frac{Ke_oA}{d}$$

where K = dielectric constant; e_o = permittivity of free space (3.6×10^{-3} pf/inch); A = conductor area; and d = distance between conductors.

This equation gives values approximating, but lower than, the values obtained by actual measurement. Therefore, the calculated values should not be used in a worst-case analysis. A typical application of the equation is shown below; the data are taken from Table 3-4.

Conductor width = 0.030 inch

Conductor length = 10 inches

Distance between conductors = 0.076 inch

Dielectric constant (K) = 5.4

$$C = \frac{Ke_oA}{d} = \frac{(5.4)\ (3.6 \times 10^{-3}\text{pf/inch})\ (3 \times 10^{-2}\ \text{inch})\ (10\ \text{inches})}{7.6 \times 10^{-2}\ \text{inch}} = 0.77\ \text{pf.}$$

TABLE 3-4
TYPICAL DIMENSIONS (IN INCHES) OF MULTILAYER BOARD MATERIALS

Thickness of copper conductors	0.003 to 0.005
Width of copper conductors	0.020 to 0.030
Thickness of copper heat sink	0.006 to 0.008
Thickness of glass epoxy board (for conductor support)	0.016 to 0.062
Thickness of glass epoxy board (for bonding purposes)	0.060 to 0.100
(Dielectric constant for glass epoxy: 4.8 to 5.4)	

The capacitive coupling between conductors can be minimized by (1) increasing the insulation thickness between layers, (2) narrowing the conductor width and increasing the distance between conductors, and (3) routing circuitry on one layer at right angles to the circuitry on the other layer. In the case of circuits located over the shield or ground plane, the entire length of a conductor is capacitively coupled to that plane, for which allowance must be made in the design.

The calculation of characteristic impedance for microstrip lines in multilayer boards for high-frequency application is a complex procedure and will not be covered here.

3.3.9 Ground Planes and Shielding

Solid copper planes, with clearance holes etched out in areas where no contact is desired, can be placed anywhere within the multilayer board. These planes serve to minimize interference, or cross-talk, between various critical circuits and to shield the circuits within the board from external interference.

3.3.10 Board Design Considerations

Of interest to system designers is the relationship between design requirements and cost of multilayer boards. The overall board size and number of layers do not affect price significantly; the density of terminal points or holes and the corresponding density of interconnections are the main variables controlling cost. The approximate dividing line between the cheaper and costlier versions of multilayer boards is the hole-center separation distance of 0.100 inch. Multilayer boards with hole centers spaced 0.100 inch or more apart have room for at least one conductor between the terminals. This makes possible the placement of 0.020-inch conductors on 0.020-inch spacing, and allows registration and hole-location tolerances of at least ±0.005 inch, all of which are well within the state of the art of printed-circuit manufacturing. The production of such boards with good yields does not present serious difficulty. For boards with hole spaces of less than 0.100 inches, all parameters and tolerances must be decreased proportionally. When the design requires 0.010-inch conductors and spacing, and tolerances of ±0.001 inch on layer thicknesses or ±0.003 inch on hole location and layer registration, the price increases rapidly.

To reduce manufacturing costs, a common practice is to process many small boards on a large panel and separate them in the last steps of the manufacturing cycle. If the tolerances are wide, large panels with small individual circuits can be processed without difficulty. Manufacturing costs increase when close terminal-area spacing and tight tolerances are demanded. Panel size must be reduced to ensure accuracy of registration and hole location, so that only a few multiple circuits can be processed simultaneously on the panel. The difference in cost between wide- and narrow-tolerance circuits can be 80 percent or more.

3.3.11 Device Mounting Considerations

The following are some general rules for mounting components to multilayer boards:

(1) If some discrete components are included in the design, these heavy components should be located near the secured edge of the board to minimize shock and vibration effects.

(2) In high-density packaging all components will normally require heat sinks.

(3) The IC's with the highest dissipation should be mounted on their heat sinks at points nearest the connection to the main heat sink, since such points are the coolest.

(4) Whenever possible, the IC's with the greatest number of interconnections should be located close to each other. This will reduce the complexity of the interconnecting circuitry.

3.3.12 Delivery Time

From the date that artwork for a multilayer board is approved, the minimum delivery time is about four weeks. To produce and check out the artwork may take three or four weeks, giving a minimum delivery cycle of seven weeks. The user frequently prepares the artwork himself.

As with other equipments, there is no assurance that the multilayer board will operate as specified until the first unit is assembled and tested. Hand-wired breadboards do not exactly reproduce the conditions of the completed multilayer boards, since the parameters of conductors in multilayer boards are quite different from those of wires on breadboards. However, after the prototype-development stage has been completed, system production can be achieved within four to five weeks. Thereafter it is only a matter of scheduling to keep the flow of multilayer boards at the required level.

3.3.13 LSI Considerations

A major advantage of LSI is the reduction in interconnections. The total number of connections required for a system is reasonably constant regardless of how the system is fabricated. The type of connection, however, and its frequency of occurrence can vary drastically from one approach to another. LSI substitutes intraconnections for interconnections, i.e., metalization within the package for soldered or welded connections external to the package.

The reduction in external leads afforded by the use of LSI devices instead of equivalent numbers of IC's is difficult to predict because of the effect system partitioning has on external-pin requirements. It can only be stated that the pin requirement for an LSI device will be considerably lower.

Earle* has shown that the external-pin requirement for relatively simple circuits depends on where the circuits are partitioned. The minimum pin requirement appears at a function level or a multiple of it. This is illustrated in Figure 3-33.

*J. Earle, "Digital Computers," Electronic Design, 7 December 1964, p.39.

Figure 3-34 illustrates the same type of comparison as that of Figure 3-33 except on a higher function level. These data* are based on experience with a number of digital systems and on an assumed 1600-IC circuit complexity. The curve of Figure 3-33 has characteristics similar to that of Figure 3-32. It is expected that the minimum point represents some functional level.

Note that the pin requirements are different for the two types of logic having the same circuit complexity. Control circuits will typically have a greater requirement for pins than data-processing circuits, because the control circuits usually require intermodule connections.

To show the effect of pin reduction by the use of LSI, the following hypothetical case is presented. The data and partitioning are based on an actual subsystem** fabricated with IC's mounted on nine double-sided printed-circuit boards.

For purposes of comparison, it is assumed that the number of gates required for each function can be fabricated on a single wafer and intraconnected as required.

Table 3-5 compares the number of packages, pin requirements, and pin-to-gate ratio for LSI and IC. Four different types of IC's are required for the subsystem; a gate, inverter, buffer, and flip-flop. The total number of IC's required is 1550, as compared with seven different types of LSI devices and a total requirement of nine devices.

TABLE 3-5 **

COMPARATIVE LSI AND IC PACKAGE, PIN, AND GATE REQUIREMENTS

Functional	Packages Required		Estimated Equivalent Gates	Pin Requirement		Pins Per Gate	
	LSI	IC		LSI	IC	LSI	IC
Input/Output	1	110	275	78	1,540	0.28	5.6
Format/Control	1	160	400	78	2,240	0.20	5.6
Register A	1	215	926	67	3,010	0.07	3.2
B	1	215	926	67	3,010	0.07	3.2
C	1	215	926	67	3,010	0.07	3.2
Process/ A Unit	1	175	437	68	2,450	0.15	5.6
B Unit	1	210	525	127	2,940	0.24	5.6
C Unit	1	170	425	107	2,350	0.25	5.6
D Unit	1	80	200	67	1,120	0.33	5.6

*Adapted from data by J. J. Staller, Guidelines for Implementation of System Requirements into Microelectronic Mechanical Design, Sylvania Electronic Systems Division, 1964.

**R. Rice, "The Packaging Revolution, Part V: Simpler Designs for Complex Systems" Electronics, 7 February 1966, p.109.

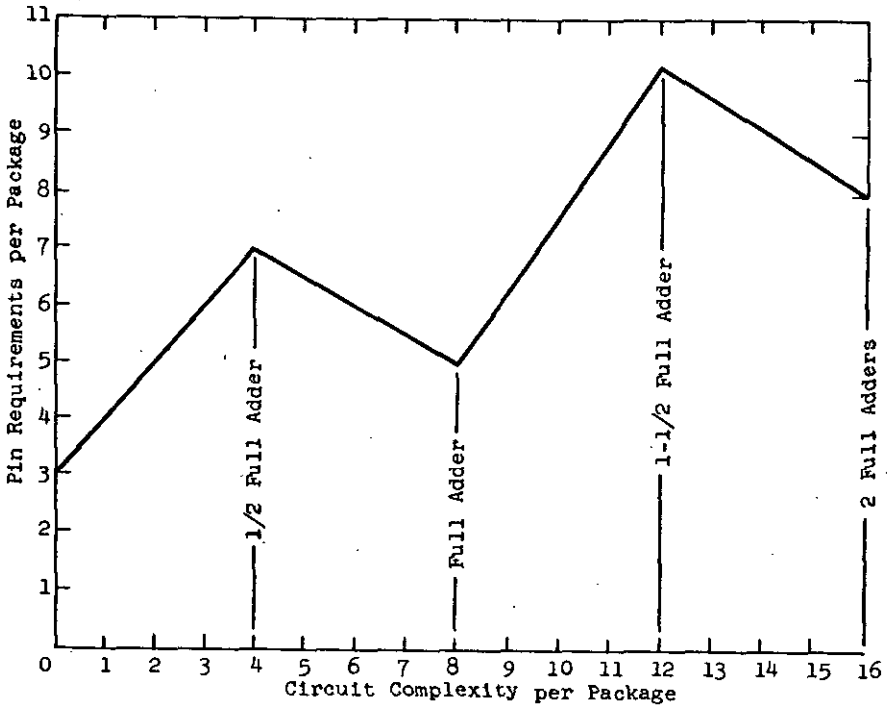


FIGURE 3-33

RELATIONSHIP OF PIN REQUIREMENTS TO CIRCUIT COMPLEXITY

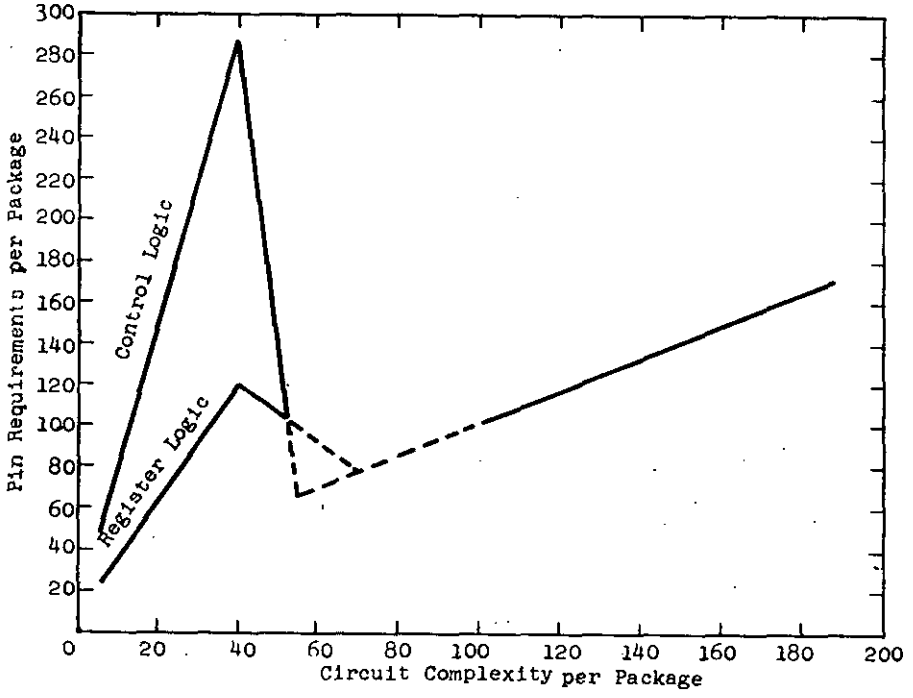


FIGURE 3-34

RELATIONSHIP OF PIN REQUIREMENTS TO CIRCUIT COMPLEXITY

The three registers are identical. This reduces the number of different types of LSI devices that are required in the system. If they were different, two additional types would be required. For logistics purposes, the designer should try to keep the number of different types of parts to a minimum.

It is expected that most LSI devices will, in general, not be used often. In fact, the more complex the device, the less use it will have. Figure 3-35, a plot of the relative usage as a function of the number of gate functions per LSI, indicates that usability drops rapidly as complexity increases. A device containing more than 30 gates has a use factor of less than 0.6. The conclusion to draw is that LSI devices will not have high use factors; thus the designer should go to the highest complexity possible within the state of the art, because he will gain the most advantage from high complexity. The use factor is not an important consideration for gate complexities above 30 and can be disregarded for complexities above 80 gates per wafer.

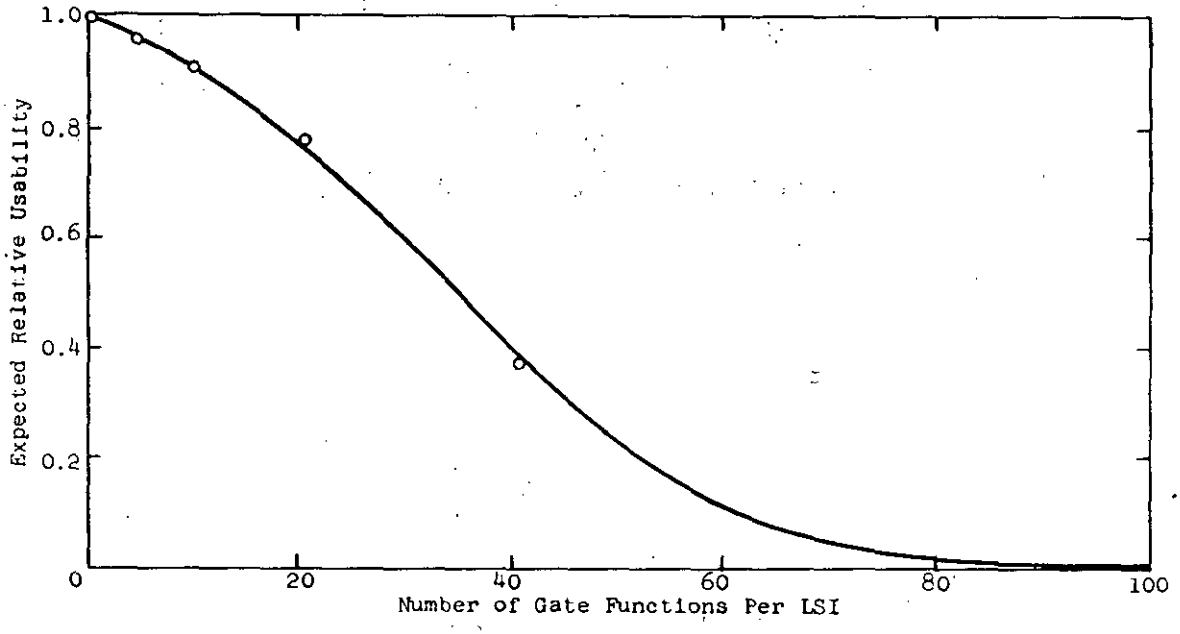


FIGURE 3-35

COMPARISON OF LSI COMPLEXITY VS USABILITY

3.4 INTRACONNECTIONS

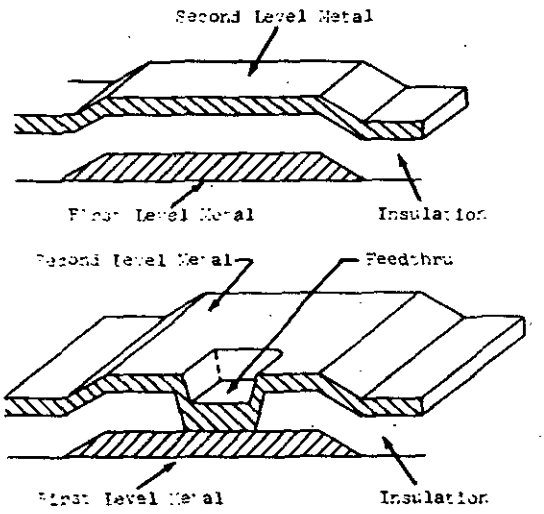
Intraconnections are metalizations which connect various components on the die. They may also be considered as metalizations within a package connecting various flip-chip dice. The difficulties associated with interconnections are also associated with intraconnections, primarily in multilevel metalizations required for high-density packaging.

LSI devices will require a multilevel intraconnection system. The problems associated with intraconnecting the gates on the wafer are toplogical and electrical. Crossovers for bias and signal-voltage metalizations are required; the resistance of the metalizations and contact resistance at the corssover-oxide steps must be low. A multilevel crossover system is composed of the following:

The first level of metalization intraconnects the components that form each circuit. These circuits form a matrix array on the wafer.

A layer of insulation separates the first layer of metalization from the next layer. Holes are etched through the insultation to the first level of metalization.

A second level of metalization is deposited on the insulating layer and forms the second level of intraconnections. Connection is made between the two levels via the holes in the insulation layer. Additional layers of metalization may be fabricated by repeating the insulation and metalization steps. A typical two-layer metalization system is shown below:



TWO-LEVEL LEAD CROSSOVER

Aluminum metalization used in current IC's is one-mil wide and 8000Å thick. This gives a sheet resistance of 0.03 ohms per square. LSI devices will probably require thicker metalization to compensate for thinner line widths and expected contact resistances of between 0.1 and 0.3 ohms per crossover. The intraconnection of a large number of gates on a wafer requires:

A method of intraconnecting only good gates

Multiple levels of intraconnections

There are two basic approaches used for intraconnecting the good gates: fixed wiring and discretionary wiring*.

*For a detailed discussion of discretionary wiring, see J.W. Lathrop, LSI Integration Through Discretionary Interconnection, TI Report, 1966, and H. Freitag, "Design Automation for LSI", Wescon Paper 10-2, 1965.

The fixed-wiring approach uses a single set of metalization masks for identical LSI devices. Each circuit element in the wafer is tested before metalization. The bad circuits are automatically marked, and the wafer is examined to determine which portions of the wafer meet the topological requirement for the mask set. These are then separated from the wafer and metalized. The good portions of the wafer that cannot be used for the LSI device are diced and packaged as individual IC's. This method provides low metalization costs but is highly dependent upon the wafer yield. Many device manufacturers believe this is the best approach to the metalization problem. Those advocating this approach are referred to as the "100 percent club" because of their dependence on high wafer yield.

The discretionary-wiring approach is not as sensitive to wafer yield but is highly dependent upon costly computing equipment. In this process, the circuit elements on the wafer are tested, and the results are stored in a computer. The computer is programmed to provide an intraconnection pattern for a particular array. A mask is made by an automatic-exposure process under control of the computer.

This system uses a flying-spot scanner whose image is projected onto a high-resolution CRT. The CRT image is used to expose a high-resolution plate, which will be used as the mask. The scanning routine starts with the shortest paths and proceeds to the longest. It may be necessary to perform the routine more than once.

The result is that each wafer has its own custom intraconnection pattern of good circuits. The bad circuits remain on the wafer but are not intraconnected.

Either of these methods can be modified by using the "master-slice"* approach at the matrix-element level. The matrix elements are intraconnected by the first metalization, which can vary within certain constraints to provide different circuit arrangements within each matrix element.

Typical two-level metalizations are composed of Al-SiO_2-Al. A serious problem affecting this technique is the incidence of pinholes that reduce the device yield. Pinholes are caused by perturbation or discontinuities in the metal or oxide. The result is a short or near-short through the oxide, causing the device to fail.

Another important phenomenon, which manifests itself by frequent short (or near-short) circuits, is the inherent mechanical weakness of the A_1-SiO_2-Al system. During testing, it is necessary to prove the wafer; a light touch of the steel point will often cause a short. This situation is not peculiar to LSI but may be complicated by the multiple layers of metalization.

*T.I. Tradename for a standard substrate of active and passive components that are custom intraconnected to give the desired circuit function.

A complex array may contain as many as 5000 crossovers. The failure of a single crossover may cause the entire array to fail. The cost to process a wafer to this point is high; thus the frequency of failure must be extremely low if reasonable prices for LSI devices are to be obtained.

3.5 OPERATING FREQUENCY

The IC operating frequency stated by the manufacturer on the specification sheet is usually a maximum device frequency under ideal conditions. Only when the effects of temperature, variations in supply voltage, actual physical layout, and capacitive loading from other devices are determined is it possible to determine the actual operating frequency.

When buffer amplifiers and driver units are required, additional speed may be lost because some of these units are slow. For logic circuits an estimate of the actual maximum operating frequency is about half the ideal repetition rate specified by the manufacturer.

Aside from the frequency constraints associated with the design of the IC circuit, the most important single factor affecting frequency is capacitive loading. This effect can be reduced by understanding where the difficulties occur and packaging the system to minimize these parasitics.

There is a capacitance associated with the input of the device. This is nominal and usually amounts to between 3 and 5 pF. The total capacitance associated with IC inputs will depend upon the fan-out. For a high fan-out, it could be appreciable. There is also a capacitance associated with printed-circuit wiring; it is a function of conductor width, conductor length, and separation between conductors. The capacitance associated with a double-sided printed-circuit board will range from 0.1 to 1 pF per inch of conductor length.

When multiple-layer printed-circuit boards are used, the capacitance load increases because of the decrease in separation between conductors. Typical* values are 0.3 pF per inch between adjacent conductors on the same plane, 3 pF per inch for parallel lines on adjacent layers, and 7.5 pF for lines on layers next to a power plane.

Capacitance is also associated with the printed-circuit board connector. Typical values are from 1 to 3 pF. If these capacitances can be reduced or eliminated, higher operating frequencies can be obtained.

*S. A. Hays, "The Packaging Revolution, Part IV; Bigger, Better Multilayer Boards", Electronics, 29 November 1965.

LSI promises complex circuit arrays with high component density and high speed. Arrays with as many as 1000 gates on a single chip with logic-block delays of only a nanosecond or so per gate are predicted. These short delay times are possible because of the close proximity of the circuits and the reduction of capacitance associated with wiring, printed-circuit boards, packages, and connectors.

As the switching speed increases, serious consideration must be given to the introduction of false signals caused by reflections from unterminated transmission lines. The close spacing of circuits, with their resultant short transmission lines, will minimize this problem on the wafer. However, transmission lines between LSI packages will probably require resistive terminations to reduce signal reflections.

3.6 THERMAL CONSIDERATIONS

Power dissipation is specified by device manufacturers, but different manufacturers specify under different operating conditions. Usually the maximum power dissipation at a 25°C ambient temperature is specified. This is typical the static (d-c) power dissipation. It is also necessary to consider the effects of the a-c component of power dissipation on the device.

Operating characteristics* that affect the power dissipation are ambient temperature, capacitive loading, operating frequency, and bias voltage. In all cases, the power dissipation of the device increases for increases in these parameters.

A temperature change from -55°C to +25°C will typically increase the dynamic power dissipation by 400 percent; from +25°C to +125°C the dynamic power-dissipation increase is 200 percent. The average power dissipation increases as a function of capacitance loading and is different for different frequencies of operation. An increase in capacitance from 15 pF to 100 pF will cause an increase of about 48 percent in average power dissipation in the frequency range of 10 to 25 MHz. The power dissipation may increase as much as 35 percent for a 10-percent increase in bias voltage.

The amount of energy required to energize a load is fixed and independent of the circuit bandwidth speed. However, the power dissipated in a circuit will vary according to the frequency of operation. The relationship between power and frequency is illustrated in Figure 3-36**. The minimum power and the maximum speed of the circuit are generally determined by component or device properties. Figure 3-36 shows that power dissipation increases as frequency increases.

*S. Klein, "A Guideline for Selecting Integrated Circuits", Computer Design, September 1966.

**J. J. Suran, "Circuit Considerations Related to Microelectronics," Proceedings of the IRE, February 1961.

A specific illustration of this principle is shown in Figure 3-37. There the PRF of a flip-flop is plotted against power dissipation. For a change from 1 to 25 MHz at 25°C, a typical increase in power dissipation is about 300 percent.

COURTESY OF IRE PROCEEDINGS

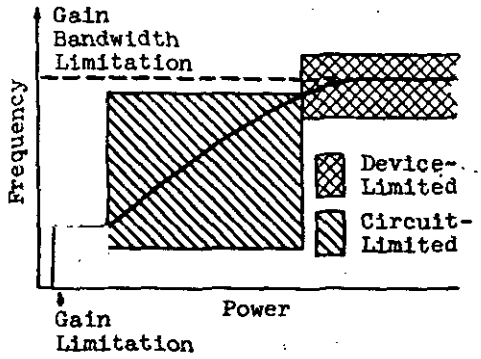


FIGURE 3-36

SPEED VS POWER DISSIPATION

COURTESY OF IRE PROCEEDINGS

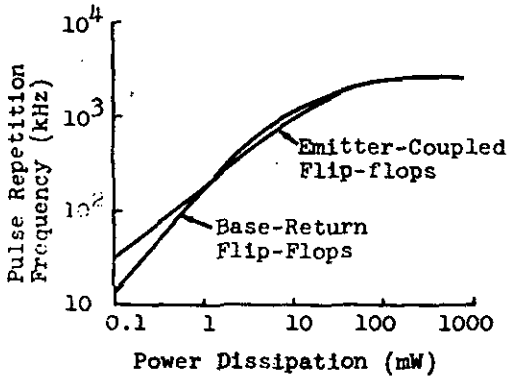


FIGURE 3-37

FLIP-FLOP PRF VS POWER DISSIPATION

The ability of an IC to operate at a given power dissipation is a function of the power density on the substrate. This is not a problem for simple IC's but becomes important as complexity increases while substrate size remains the same.

3.6.1 LSI Thermal Considerations

High component densities in LSI devices will require effective heat-transfer mechanisms. Digital IC's operate at a few mW power dissipation per gate function and present no serious thermal problem. For a complex array with perhaps as many as 1000 gates, the power dissipation of the array would be tens of watts. The average power density of a 1000-gate LSI with an average power dissipation of 10 mW per gate on a 0.3 by 0.3-inch substrate would be in excess of 100 watts per square inch. This is average power density; component power densities will be much higher. To achieve integrated-circuit gate delays in the nanosecond region, much higher component power densities from 5 to 20 kW/inch^2 may be required. (This is based upon the estimate that resistor dimensions will approach 0.25-by-0.25 mil.)

The relationship between power requirements and delay is illustrated in Figure 3-38*, a plot of rise time

*M. J. Flynn, "Complex Integrated Circuit Arrays: The Promise and the Problems", Electronics, 1966.

versus power for an idealized circuit operating under three different conditions. The rise time is a function of both power and loading conditions. The curves of Figure 3-38 show, left to right, the time-power relationship when (1) all driven circuits are on the same chip and a termination resistance is not required, (2) circuits not on the same chip are driven and no termination resistance is required, and (3) there are external circuits plus a transmission-line termination.

The requirement of higher power for higher speed and its detrimental effects on failure rate will require effective thermal management for the LSI device. The pedestal to which the chip is bonded must have a sufficiently high thermal conductivity to distribute the heat flux across the chip and then transfer the heat to the package without an excessive thermal gradient. The heat must then be transferred from the package to a heat exchanger. To keep thermal gradients within acceptance limits it may be necessary to mount the chip on a metal pedestal instead of an insulator. The package may be required to have fins; the use of forced-air cooling may also be required to keep junction temperature within acceptable limits.

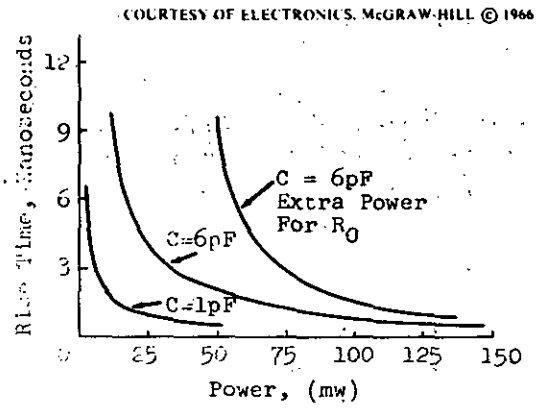


FIGURE 3-38

RISE TIME VS POWER

3.7 LOGIC SELECTION

3.7.1 Introduction

Integrated circuits are available in a variety of logic types. Most digital processing c n be performed by using IC's. The problem is usually not which IC to use, but rather which type of logic is best for a particular application. The user must be able to specify critical system-design parameters and have a good understanding of the functions (and their limitations) that each logic type can perform. With these data he can choose the best logic type for a given application.

The most prevalent logic functions are the NOR function and the NAND function. These functions are performed with an OR gate followed by an inverter and an AND gate followed by an inverter, respectively. The NOR function is performed by logic types such as DCTL and RTL for positive logic. (Positive logic means that the voltage level assigned to the "one" state is more positive than that assigned to the "zero" state. Negative logic is the opposite condition, i.e., the "one"-state voltage level is less positive than the "zero"-state voltage level.)

Logic types such as DTL and TTL perform the NAND function for positive logic. Often the manufacturer will specify the logic function for both positive and negative logic on an IC data sheet. A gate that performs the NAND function for positive logic will perform the NOR function for negative logic. The AND gate for positive logic performs the OR function for negative logic.

Some important factors that determine the selection of the logic type are the following:

- Speed. This is limited by the switching characteristics of the circuit and associated fabrication parasitics.
- Noise Margin. The magnitude of the smallest extraneous input signal that causes an error in the following chain of logic circuits. The value is usually referenced to the worst-case input-voltage level.
- Noise Immunity. The minimum noise margin divided by the maximum logic swing.
- Fan-out. The number of like loads the logic element is required to drive.
- Power Dissipation. The average d-c power required to operate the device.
- Power-Speed Product. The product of the device propagation delay (t_p) and its power dissipation.
- Logic capability. This includes fan-in, inversion, and availability of complemented outputs.

Other considerations are threshold levels and packaging. Also important is whether a particular circuit family includes triggered flip-flops, drivers, and other needed circuits, as well as logic gates.

The main function of any logic circuit is to provide a logic output, to one or a number of loads, in response to a logic input. An ideal logic-coupling network should provide high isolation between input and output, introduce as little signal delay as possible, keep the power dissipation as low as possible, and keep the number of components to a minimum.

3.7.2 Direct-Coupled-Transistor Logic (DCTL)

DCTL was one of the earliest logic types. Its simplicity lends itself well to integrated-circuit fabrication. However, it has serious disadvantages; these, coupled with the vast improvement in IC fabrication, have reduced it to

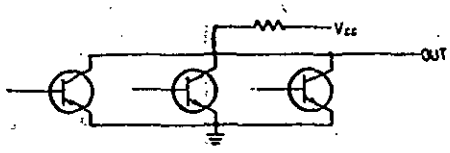


FIGURE 3-39

DCTL GATE

virtual obsolescence. Its major disadvantages are high power dissipation, relatively small logic swing (0.6 volt), and only fair noise immunity. Figure 3-39 illustrates a DCTL NAND/NOR gate.

DCTL suffers from so-called current hogging. This refers to unequal distribution of base current flowing into many bases in parallel from a common collector. Current hogging limits the fan-out and speed performance. In Figure 3-40 the gate element ideally should supply equal base current to all loads. However, small variations in the base-to-emitter threshold characteristics result in a wide variation in current supplied to the load.

This current waste limits the loading capability of DCTL gates and limits their application, except where the base-to-emitter characteristics of the load transistors can be closely matched.

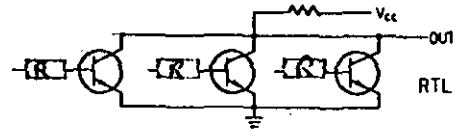


FIGURE 3-40

RTL GATE

3.7.3 Resistor-Transistor Logic (RTL)

To eliminate current hogging, DCTL was modified by placing resistors or resistor-capacitor networks in the base leads. The resulting circuits were RTL and resistor-capacitor transistor logic (RCTL) circuits, respectively. Figure 3-40 is an RTL gate.

The addition of a base resistor increases the input resistance and assures proper operation when more than one load is being driven. The addition of R to the circuit forces some of the input voltage to appear across it, as well as the base-emitter junction. As the value of R is increased, the percentage of input voltage across R increases; the base currents are more evenly divided. This increases the fan-out capability but reduces the speed somewhat since the addition of R to the circuit causes longer storage and turn-off times.

Typically the number of loads that an RTL output can drive is five, but special circuits are available that can drive many more. However, fan-out is limited because logic swing and noise immunity are adversely affected for high values of fan-out.

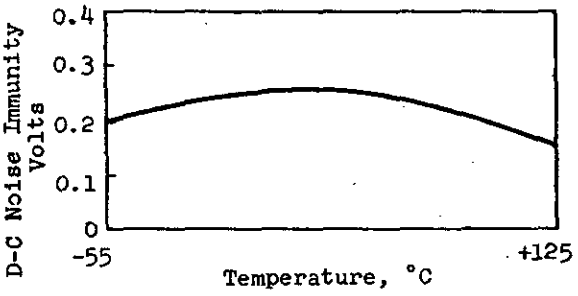


FIGURE 3-41
TYPICAL INPUT DC NOISE THRESHOLD
VS.
TEMPERATURE FOR AN RTL CIRCUIT

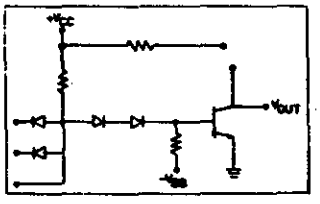


FIGURE 3-42
BASIC DTL GATE CIRCUIT

RTL noise immunity is affected by temperature as shown in Figure 3-41. The logic swing is typically about 1 volt and is dependent on fan-out.

3.7.4 Diode-Transistor Logic (DTL)

DTL logic circuits are widely used and readily available as IC's. A typical DTL gate is illustrated in Figure 3-42. The diodes provide an additional input threshold, thus increasing the impedance over that for RTL. In addition, DTL has a larger voltage swing and better noise margin than RTL. The circuit shown in Figure 3-42 requires two power supplies; this is something of a disadvantage. However, some DTL circuits, such as the 930-Series, are operated with a single power supply. DTL has additional advantages such as lower power dissipation, larger drive capability, higher speed of operation, and better noise immunity compared with RTL.

A modification of DTL is the Variable Threshold Logic (VTL). VTL operation is intended for use in circuits requiring very high noise immunity, such as relay and solenoid circuits. It is not intended to be used in computer logic applications because it has a poor power-speed product. It also requires two power supplies.

3.7.5 Current-Mode Logic (CML)

CML (also referred to as emitter-coupled logic, ECL) was designed especially to take advantage of monolithic fabrication techniques. CML uses the technique of current steering, that is, switching well defined currents with small controlling voltages. Unlike other forms of logic discussed thus far, it is a nonsaturating logic and is therefore exceptionally fast. Propagation delays of 1 to 2 nsec are possible. It also features high fan-in and fan-out and high noise immunity.

A basic ECL gate is illustrated in Figure 3-43. Typically the ECL circuit is designed with a differential-amplifier input and emitter-follower output to restore d-c levels and provide a large output-drive capability. Two power supplies

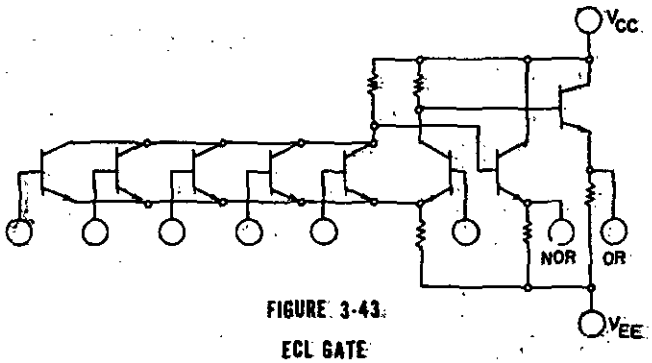


FIGURE 3-43
ECL GATE

are required for this circuit. However, some CML circuits can be operated from a single power supply. CML circuits are not compatible with saturating logic types.

The ECL structure contains some interesting and important advantages. The circuit is reasonably independent of resistance tolerance if resistor ratios are maintained. This is an important factor when one is considering diffused resistors. Another unusual feature of this structure is the symmetry in its d-c input-output transfer characteristic, which results in approximately equal noise margins for both positive and negative noise. These margins are found to be substantially independent of temperature over a wide range.

3.7.6 Complementary Transistor Logic (CTL)

The CTL logic type is intended for high-speed digital applications similar to those of the CML. The CTL circuits use PNP emitter followers as the input AND diodes. Small currents at these inputs control the circuit. CTL exhibits high input impedance and low output impedance.

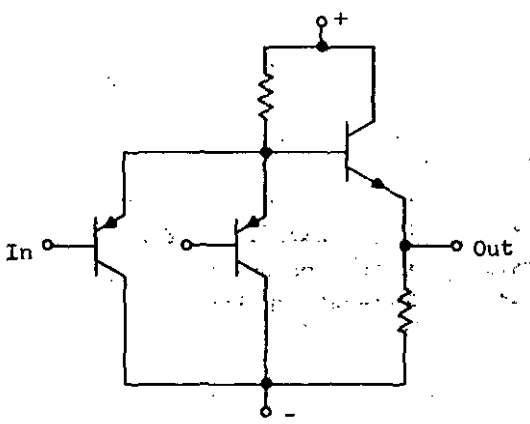


FIGURE 3-44
CTL GATE

The emitter follower introduces a stability problem in both CTL and ECL. However, the problem is reported to be more severe in CTL, because there one emitter follower drives another emitter follower. In addition, the gain through the CTL AND/OR gate is less than unity; this limits the number of gates that may be cascaded. A CTL gate is illustrated in Figure 3-44.

3.7.7 Transistor-Transistor Logic (TTL)

TTL is a modification of DTL, in which faster responding transistors are used in place of input diodes.

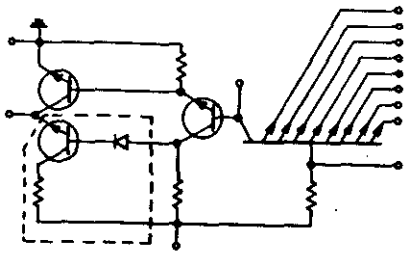


FIGURE 3-45

TTL GATE

A TTL gate is shown in Figure 3-45. This form of logic has a high speed -- typically in the order of 3 nanoseconds. High speed can be maintained for high capacitive loads, because this logic form typically has both a pull-up and pull-down transistor in the output.

The threshold diode in Figure 3-46 can be in the base as shown or in the emitter leg of the transistor enclosed by the dashed lines. When the diode is located in the base circuit, it reduces the speed of the circuit because of substrate parasitic capacitance introduced at the collector node of the previous transistor. Locating the diode in the emitter removes this capacitance and gives an improved speed characteristic.

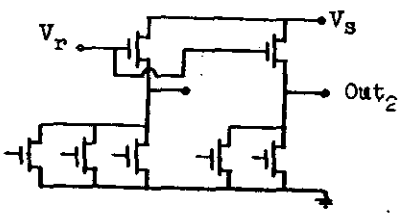


FIGURE 3-46

MOS NOR GATE

TTL circuits are becoming more widely used and can be expected to replace DTL gates for many applications because of their higher speed. However, it is possible that high leakage of the multiple-emitter transistor may load a circuit excessively.

3.7.8 Comparison of Logic Types

Parameters related to the various logic types discussed in the previous sections are tabulated in Table 3-6. These values are representative; the exact value of each parameter will vary widely depending upon the exact device being evaluated. A true performance evaluation will require user testing. Although some insight can be gained from manufacturers' data sheets, it should be remembered that these data are usually specified under ideal conditions and are usually optimistic.

3.8 REDUNDANCY TECHNIQUES *

3.8.1 Introduction

Microelectronic devices, particularly integrated circuits, exhibit characteristics that are highly useful in redundancy. The extremely small size and weight (and frequently small power consumption) of these devices allow

*Material in this section is abstracted from Handbook for Systems Application of Redundancy, NASL (August 1966).

TABLE 3-6

COMPARISON OF LOGIC TYPES

Logic Type	Device Delay, nsec	Average Power Dissipation, mW	Noise Margin (25°C), mV	Fan-out
DCTL	15	20	100	3
RTL	20	10	250	5
DTL	30	8	600	5
CML	5	50	300	10
CTL	2	50	300	5
TTL	10	15	750	20
MOS	500	10	3,000	5

the addition of redundant devices to a system without severe physical penalties. Also, the cost of integrated circuits does not increase linearly with increased complexity; a single package containing dual gates does not cost as much as two single gates in separate packages. Further, as the complexity within a single package increases, the reliability changes at a much slower rate. In other words, dual gates in the same package are only slightly less reliable than a single gate and much more reliable than two separate gates. It is generally acknowledged that, with constant electrical performance characteristics, all other characteristics of integrated circuits are superior to discrete parts arranged to perform the same function. There are, of course, restrictions on the electronic functions that can be performed by integrated circuits.

Redundancy is a method for maintaining a system function at a specified performance level for a specified time by including in the system more than the minimum number of parts, paths, or subfunctions to overcome performance and time limitations caused by internal and external influences.

Failures caused by accumulating and catastrophic processes are time-dependent and are produced through functionally related physical processes, imperfections, and impurities. To cope with these failures (i.e., to protect against a reduced probability of accomplishing a particular mission), disciplines such as physics of failure, derating, worst-case design, feedback, tolerance analysis, improvement of fabrication techniques, and redundancy can be used. Although it would be advantageous to apply all the disciplines, time and monetary restrictions usually require that trade-offs be made among them.

Physics-of-failure and fabrication-improvement efforts attack the component's imperfections and deficiencies caused by age. The application of derating, worst-case design, feedback, and redundancy attack the problem at the system level through system-design techniques.

Failures can occur in either the hardware domain or the time domain, the two domains requiring different types of engineering solutions. For example, repetition (time domain) will reduce the effect of transient noise but will not reduce the effect of hardware failures. Failures in both domains, however, can be prevented by hardware redundancy.

System failures result from a variety of causes:

(1) External physical environmental factors (heat, humidity)

(2) External-signal environmental factors (radiation, transients)

(3) Internal-signal environmental factors (noise)

(4) Faulty design or fabrication

It is, therefore, not surprising that no single type of extra structure (type of redundancy) can cope with all system failures. Failures caused by external disturbances like noise require system solutions and cannot be solved by physics of failure or by component perfection alone.

It is necessary to evaluate failure-correction methods in light of the expected increase in the probability of maintaining a specified function and the cost of implementation.

In comparing the reliability of a redundant system with that of a series system, a number of factors should be considered:

(1) Reliability of detection and switching devices

(2) Independent failure of redundant parts, paths, or subfunctions

(3) Reliability of additional connectors

(4) Availability of spares, maintenance manpower, and diagnostic equipment

(5) Probability distribution of time to failure of redundant configurations

(6) Incorporation of extra structures in a system, which can take such forms as the following:

  (a) Increase in material costs

  (b) Increase in weight, volume, power consumption, and cooling facilities

  (c) Extra time to perform functions

  (d) More complex and expensive design

Therefore, in assessing the worth of using a redundant structure, it is essential to answer the following questions:

(1) Is the problem one of time or of hardware? If it is hardware [x% of a group of parts are working properly, while (1-x)% of the group are not working properly], no amount of repetition will correct it. However, if the problem is in time (100% of a group of parts are working properly x% of the time), repetition may correct it without any additional structures.

(2) What are the causes of the problem? Failures can be caused by external and internal environmental factors, faulty design, or faulty fabrication. To cope with the problem, the cause must be known.

(3) What type of failure correction is necessary? Failure correction can increase reliability by preventing or postponing the occurrence of a failure or, when a failure occurs, circumventing it or reducing the time to detect, diagnose, and correct it.

(4) Does the redundancy lend itself to ease of diagnostic repair and maintenance? It is possible to have a redundant structure in an operating system without deriving any benefit because diagnostic repair and maintenance are not readily performed. If all redundant paths were in a failed state and the user did not know it, the reliability of the system would be the same as if the redundant structure were not included in the system.

(5) What is the effect on pertinent reliability parameters? Depending on the system configuration, there are a number of reliability parameters, such as mean time to failure and probability of success for a specified time and availability, whose effects should be investigated for the applicable redundant structures.

(6) What is the cost of employing the redundant structure? The additional cost for using a redundant structure -- extra material, extra time to perform functions, extra power and cooling, more complex and expensive design, and added weight and volume -- should be used as a basis for evaluating alternate redundant structures that may be applicable in a specific situation.

(7) What potential gains are provided to the user by a particular redundant structure? Whether or not the cost of the increased reliability is worth the increased cost can be determined only by the user in the framework of the specific system application. The selection of the appropriate reliability parameter to use in making comparisons can be a problem in itself.

3.8.2 Mathematical Methods

3.8.2.1 General Probability Approach

For a comprehensive understanding of the mathematics involved in the compilation of reliability measures for redundant structures it is necessary to have a fundamental understanding of the theory of probability.

Definition of Probability. There are many definitions of the word "probability," and it is a popular pastime of mathematicians to debate the relative merits of each. No attempt will be made here to participate in the debate, or even to reconcile the different interpretations, but it is necessary to mention some of them.

The probability that an event will occur can be thought of as the proportion of the time one would expect it to occur in a very large number of trials. If there were no sampling variations, one would expect 120 throws of a single unbiased die to yield 20 ones, 20 twos, and so on up to 20 sixes. The word "unbiased" means that each face would appear as often as each other face. Since there are six faces, the probability of each appearing is one sixth.

The foregoing is a rough statement of the relative-frequency definition of probability. This definition says that if an event occurs in s out of a total of n trials, and does not occur in the remaining n - s trials, the probability of its occurrence is the limit of the ratio of s to n as n goes to infinity. Some mathematicians argue that this definition is not logically consistent, but they all agree that in a proper experiment, s/n is a fairly good approximation of the probability if n is large.

Another definition is based on a count of the number of possible results of a trial. The dice example applies here. Since there are six possible outcomes of a throw of a single die, the probability of any one number's occurring is one sixth, assuming no bias. In general, if an event can occur in m ways and can fail in n ways, the probability of its occurrence is $m/(m + n)$, provided the ways are exhaustive, equally likely, and mutually exclusive.

The concept "equally likely" is basic; it can play the role of the undefined element. Indeed, it is quite difficult to define the term "equally likely", but it can be described roughly as the lack of any bias favoring one outcome over another in a trial.

Definitions of Related Terms. Some of the terms that are used in the definitions of probability and some other terms that will be needed later are defined as follows:

(1) Random Outcome. The value of an empirical observation that cannot be predicted (lack of deterministic regularity) but that has statistical regularity in that the value has a relative frequency of occurrence

in a series of independent observations of the phenomenon (the result of tossing a die, the time-to-failure of a device, etc.).

(2) Exhaustive. As used in the definitions of probability, the term "exhaustive" means that all possible ways for an event to happen are included. The reasons for this restriction in the definition should be obvious.

(3) Trial. A trial is an attempt under a certain set of rules to produce an event A where the outcome of the event A is uncertain. Thus each repeated throw of dice at craps is an attempt to make one's point. One usually speaks of a random trial. The term "random" implies "without bias".

(4) Independent Trials. If the outcome of one trial does not influence the outcome of a subsequent trial, the two trials are said to be independent. Each throw of dice in a crap game meets this criterion. However, the drawing of cards from a deck without replacement does not meet it, since the number of ways an event (drawing a specified card, for example) can happen changes with each drawing.

(5) Event. A set of outcomes. The event has occurred if one of the outcomes of the set is observed on a trial. (If the event is an even number coming up on a die, it occurs if the number 2, 4, or 6 is observed.)

(6) Independent Events. Sets of outcomes based on independent trials.

(7) Mutually Exclusive Events. Two or more events that cannot occur simultaneously (odd and even numbers on one toss of a die).

(8) Independent Trials. Trials of which the outcome of one has no effect on the outcome of others that follow.

3.8.2.2 Conditional and Unconditional Probabilities

The conditional probability of an event is encountered when information about the occurrence of some other event is available. If one is informed that a certain event has occurred, does this tell him anything about the probability of the occurrence of another event? Knowing that B has occurred, what is the probability of A's occurring? If the events are dependent, the knowledge that one has occurred does modify the probability of the other, and this probability is conditional. If no information is available as to the occurrence of an event on a previous trial, the probability is unconditional.

The conditional probabilities of short and open failures are sometimes used to represent element-failure probabilities.

3.8.3 Forms of Redundancy

3.8.3.1 General

A simple description of an idealized requirement for components or functions in a redundant configuration will identify the conditions necessary for the most desirable characteristics.

In a redundant configuration, an idealized device will detect a failure (within itself), make a decision, and complete a switching operation. This over-simplification will identify design requirements for components in a redundant configuration. The deficiencies of a component relative to the idealized device will indicate the level of additional circuitry required for ideal operation. It is emphasized that in a redundant structure a particular component failure mode may have more or less effect in one circuit than it does in another, depending on how the circuit fails (e.g., open or short) in addition to the circuit's effect on performance. The component must be used in such a way that its failure mode does not impede the circuit function.

The simplest example is a component that parallels another, or circuit that parallels another, where the failure is an open. It is apparent that the failure mode does not load or change the circuit function (performance is not considered), and the simple expression for component reliability of parallel structures holds. The component has detected a failure (within itself), made a decision, and acted as a switch.

In another example, two transistors are used in tandem, one in saturation and the other operating in the linear region. A collector-to-base or collector-to-emitter short in the operating transistor can turn off the transistor and cause it to act as if it is saturated, and the saturated unit will become operable. Again, the component detected its failure, made a decision, and switched.

If the element in the first example also had a short mode, then the analysis would be more difficult. There are mathematical expressions for this case; but if the undesirability of a short is indicated, then the short must be detected, decision made, and switching effected. An additional component, such as a fuse, would be necessary in this case to accomplish the detecting, decision-making, and switching.

It can be seen that the nature of the component's failure mode, and the circuit it is designed into, determines what information and self-action are necessary when a component in a redundant circuit fails, and what (if any) additional circuits must be used for adequate circuit functioning. This can be used as a basis for determining how to classify a redundant structure; decision, if the decision capability is present; and, if switching is present, the additional component that must be added to complete the switching action. This leads to a broad classification of the forms of redundancy as non-decision, decision without switching, and decision with switching.

These classifications are used in preference to "passive", "active", and mixtures of the two, which do not provide clear separation or distinction between forms. This classification scheme describes the components that have been added to the circuit, in addition to the components necessary to insure functional integrity. In all cases of multiple hardware, there must be inherent detection, decision, and switching. It is also assumed that detection must precede a decision; thus a separate category of detection is not needed.

It is generally true that in the ideal case, where the failure mode is known, system success can be modeled. In other cases only approximations can be given; and the results will be misleading unless proper precautions or better modeling are used. This is apparent in the case of the Moore-Shannon analysis for relays, or other four-pole components, in which the control and signal paths are separate. This concept is used for series-parallel quad circuits that do not have components with a four-pole characteristic. For bimodal failures of components with common signal-contact paths, separate mathematical forms (not Moore-Shannon) are required to describe the success paths properly.

The redundancy forms are discussed below, and illustrations of typical circuits are given. Where possible, reliability block diagrams are provided. The mathematical model for each circuit is given, along with requirements for its applicability. In all of the time-dependent models, it is assumed that all components are operable at time zero. In general, the redundancy models yield increasing-failure-rate (IFR) functions for similar elements. In many instances the mathematical model is difficult to obtain or, if obtainable, difficult to solve. More and more computer evaluation is resorted to in programs such as the IBM Electronic Circuit Analysis Program (ECAP), ARINC Research Circuit Analysis Program, and Los Alamos Scientific Laboratory's NET-1 program.

The reliability of the form can be given in terms of the probability gain or loss (nuisance factors) as measured by the ratio of system to component success, e.g., whether this ratio is greater or less than unity.

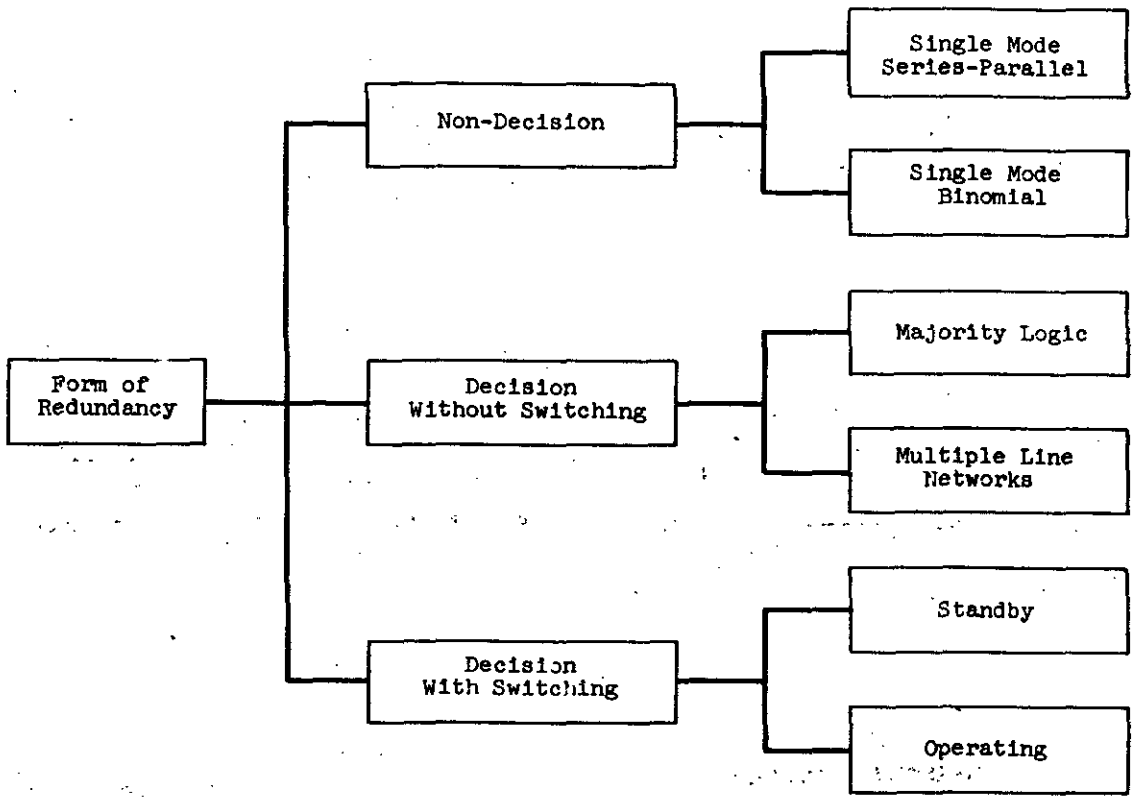


FIGURE 3-47

REDUNDANCY TREE STRUCTURE

3.8.3.2 Tree Structure

As outlined in Section 3.8.3.1, there are three forms of redundancy (Figure 3-47).

(1) Non-Decision Redundancy

(2) Decision Redundancy Without Switching

(3) Decision Redundancy With Switching

Non-decision redundancy forms encompass those structures that do not require external components to perform the function of detection, decision, and switching when an element or path in the structure fails. Examples are Single Mode/Series-Parallel and Single Mode/Binomial.

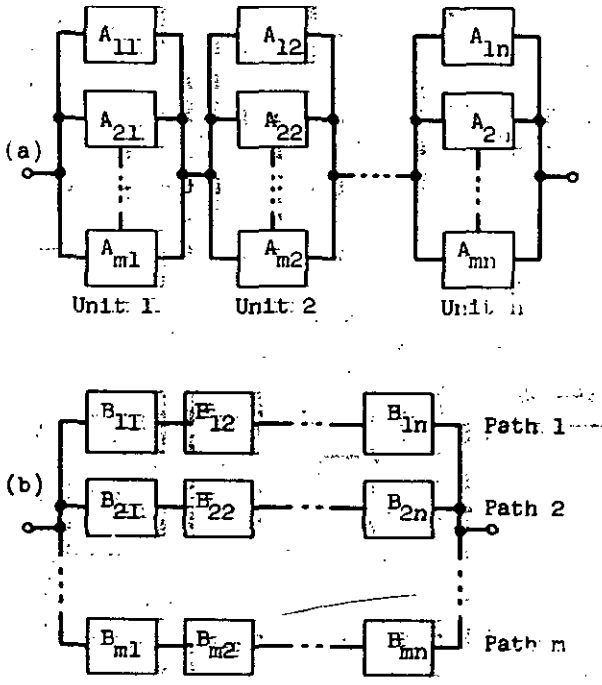


FIGURE 3-48A
SERIES PARALLEL CIRCUITS

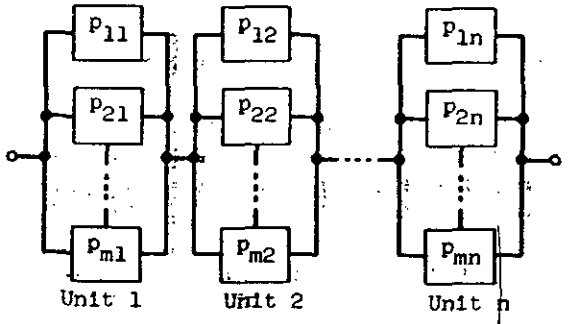


FIGURE 3-48B
GENERAL REPRESENTATION OF A SERIES OF MULTIPLE PARALLEL ELEMENTS

Forms of decision redundancy without switching are structures that require an external element to detect and make a decision when an element or path in the structure fails; these do not need an external element to perform the switching function. Examples are Majority-Logic and Multiple-Line Networks.

Forms of decision redundancy with switching include those structures in which external elements are required to detect, make a decision, and switch another element or path as a replacement for a failed element or path. Examples are Standby and Operating.

3.8.3.3 Nondecision Redundancy

Single-Mode/Series-Parallel

Definition of Form

A series of n units in series with m parallel elements in each unit where only one mode of failure can occur.

Typical Circuits

In the circuits of Figure 3-48A, A_{ij} elements are only subject to open-type failures while B_{ij} elements are only subject to short-type failures.

Reliability Block Diagrams

Both of the circuits in Figure 3-49A would have the block diagrams of Figure 3-48B.

Mathematical Models:

Discrete requirements.

For proper use, each of the elements should be independent of each other, with success probability of P_{ij} for the i^{th} element in the j^{th} unit so that

$$R = \prod_{j=1}^{n} (1 - q_{1j}\, q_{2j} \cdots q_{mj})$$

where q_{ij} is the failure probability of the i^{th} element in the j^{th} unit.

Physical Properties

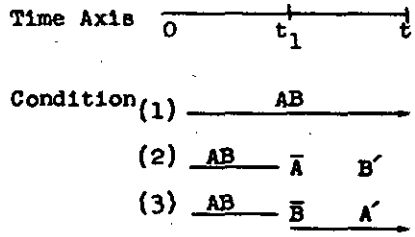
Time - This form of redundancy is usually applied in situations that call for continuous operation in time.

Performance - Frequency response with time could be a factor that must be considered. The circuit can be used without consideration of frequency response if only continuity of path is needed.

Reliability Obtainable

The gain for a specified mission can be measured by the ratio of the reliabilities of the form to the nonredundant structure. If the load is divided among the alternate paths in the structure, the failure of a path may increase the load that the remaining paths will have to carry.

For a relatively simple example of two elements in parallel, a direct derivation of the reliability function can be performed by considering all possible ways for the system to be successful, as given in the following diagram.



Success = Conditions (1), (2), or (3)

(The bar above a letter ($\bar{A}$) represents failure of that element. The prime (A') represents operation of that element under full load; absence of a prime represents operation under half load.) To show the derivation, let

f (t) represent the failure-time density under partial or half load (i.e., when both elements are operating)

let

g(t) represent the element failure-time density under full load (i.e., when one element has failed)

let

$t_1 < t$ represent some fixed point in time

For identical elements:

$$f_a(t) = f_b(t) = f(t)$$
$$g_a(t) = g_b(t) = g(t)$$

The system is operating satisfactorily at time t if either A or B or both are operating successfully. Under the assumption that the elements are independent if both are operating, the probability that both will operate until time t is

$$\left[\int_t^\infty f(t)\,dt\right]^2 = \left[R_f(t)\right]^2$$

The density for one element failing at time t_1 and the other surviving to t_1 under L/2 and from t_1 to t under L is

$$f(t_1)\int_{t_1}^\infty f(t)dt \int_{t-t_1}^\infty g(t)dt = f(t_1)\,R_f(t_1)\,R_g(t-t_1)$$

Since t_1 can range from 0 to t, this density is integrated over that range, and the resulting probability is doubled because the event can occur in either of two ways. Hence:

$$R(t) = \left[R_f(t)\right]^2 + 2\int_0^t f(t_1)\,R_f(t_1)\,R_g(t-t_1)\,dt_1$$

For the case of the two different elements with half-load densities, $f_a(t)$ and $f_b(t)$, and full-load densities, $g_a(t)$ and $g_b(t)$,

$$R(t) = R_{fa}(t)\,R_{fb}(t) + \int_0^t f_a(t_1)\,R_{fb}(t_1)\,R_{gb}(t-t_1)\,dt_1$$
$$R_{fa}(t_1) + \int_0^t f_b(t_1)\,R_{fa}(t_1)\,R_{ga}(t-t_1)\,dt_1$$

If the failure times of the elements are exponentially distributed and each has a mean life of θ under load L/2, and θ' equals θ/k under load L(k 0), solution of the above equation gives

$$R(t) = \frac{2\theta'}{2\theta' - \theta}\,e^{-t/\theta'} - \frac{\theta}{2\theta' - \theta}\,e^{-2t/\theta}$$

when $k \neq 2$. When $k = 2$,

$$R(t) = \left(\frac{2t + \theta}{\theta}\right) e^{-2t/\theta}$$

The system mean life can be shown to be equal to

$$\theta_s = \frac{\theta}{k} + \frac{\theta}{2}$$

Note that when $k = 1$, the system is one in which load-sharing is not present or increased load does not affect the element failure probability. This is exactly the assumption made for previous discussions of active-parallel redundancy, and therefore θ_s for $k = 1$ is equal to $\frac{3}{2}\theta = \frac{3}{2\lambda}$. If there were only one element, it would be operating under full load; therefore, the system mean life would be $\theta' = \theta/k$.

Hence, addition of a load-sharing element increases the system mean life by $\theta/2$. This gain in mean life is equivalent to that gained when the elements are independent, but the overall system reliability is usually less because θ' is usually less than θ $(k>1)$. Therefore,

$$\theta' + \frac{\theta}{2} < \theta + \frac{\theta}{2} = \frac{3}{2}\theta$$

The use of a single improved element when this dependent model is assumed is of interest. To illustrate the effects of using improved single elements or redundant standard elements, the following configurations will be considered:

(1) Single standard element; $\theta = 50$

(2) Single improved element; $\theta = 100$

(3) Dependent model, standard elements; θ (half load) = 100, θ' (full load) = 50

The mean lives and reliability functions of these three configurations are

(1) $\theta_A = 50$, $R_A(t) = e^{-t/50}$

(2) $\theta_B = 100$, $R_B(t) = e^{-t/100}$

(3) $\theta_C = 100$, $R_C(t) = e^{-t/50}\ (1 + t/50)$

The reliability functions are shown in Figure 3-50. Note that although systems B and C have the same mean life, the redundant system has greater reliability in early life. After approximately 125 hours, the improved single-element system is superior. If such factors as effectiveness, cost, weight, and complexity are approximately equivalent for systems 2 and 3, the choice would relate to the mission time.

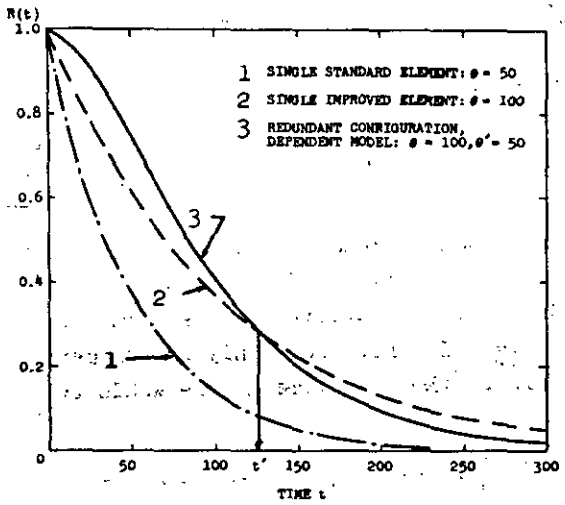


FIGURE 3-49

RELIABILITY FUNCTIONS FOR REDUNDANT CONFIGURATION (DEPENDENT MODEL) AND NONREDUNDANT CONFIGURATIONS

Nuisance Factors - Figure 3-49 shows a crossover point (3 over 2) where the redundant configuration becomes inferior in reliability to the nonredundant configuration. In general, this will be true of all redundant versus nonredundant configurations.

3.8.3.4 Single-Mode Binomial

Definition of Form

A redundant structure of m parallel paths where at least k (k<m) paths must be satisfactory for system success.

Typical Circuits

Same as in Figure 3-48A.

Reliability Block Diagrams

For the case in which k = 2 and m = 3, the block diagram would be as in Figure 3-50.

Mathematical Models

Discrete Requirements. If independent element failures are assumed, the reliability for n units each requiring k out of m successful paths can be obtained by the state probabilities representing failure which are found from the following expression:

$$[(A_{11} + \overline{A}_{11})\,(A_{21} + \overline{A}_{21})\,\ldots\,(A_{m1} + \overline{A}_{m1})]\,[(A_{12} + \overline{A}_{12})\,(A_{22} + \overline{A}_{22})\,\ldots\,(A_{m2} + \overline{A}_{m2})$$
$$\ldots\,[(A_{1n} + \overline{A}_{1n})\,(A_{2n} + \overline{A}_{2n})\,\ldots\,(A_{mn} + \overline{A}_{mn})]$$

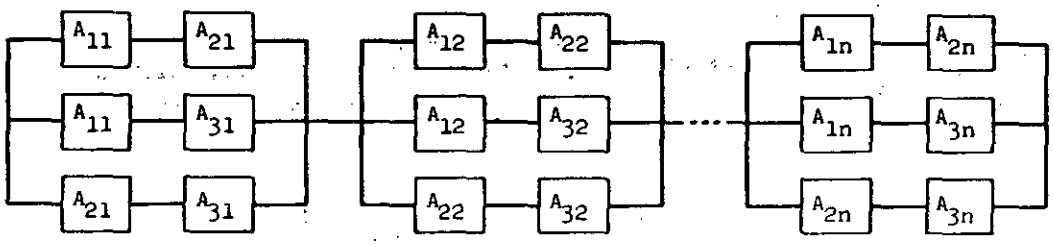


FIGURE 3-50

SINGLE-MODE BINOMIAL REDUNDANCY

The preceding expression can be written as

$$\prod_{j=1}^{n} (A_{ij} + \bar{A}_{ij})^m;$$

$$i = 1, \ldots, m$$

which is simply the product of n binomial expansions.

If p is the constant probability of a success for all elements, the reliability of a system requiring at least k successful paths out of a possible m in each unit would then be

$$R = [P_r (X \geq k_{jm})]^n = \left[\sum_{x=k}^{m} \frac{m!}{x!(m-x)!} p^x (1 - p)^{m-x} \right]^n$$

In the above equation p can also represent a time-dependent failure density, so that the following can be substituted for p:

$$p = \int_t^\infty f(t)\, dt = \text{element reliability density function}$$

3.8.3.5 Decision Without Switching Redundancy

Majority Logic

Definitions of Form

Majority logic is a form of decision redundancy wherein the correct output is assumed to be the one found in a majority of the channels.

This concept of majority logic was propounded by von Neumann and has since been enlarged upon by many authors. Von Neumann's original concept required extremely high redundancy to achieve high reliabilities, but various modifications since proposed have afforded high reliability with a rather low degree of redundancy.

Typical Circuits

See Figures 3-51 and 3-52.

Reliability Block Diagrams

See Figure 3-53.

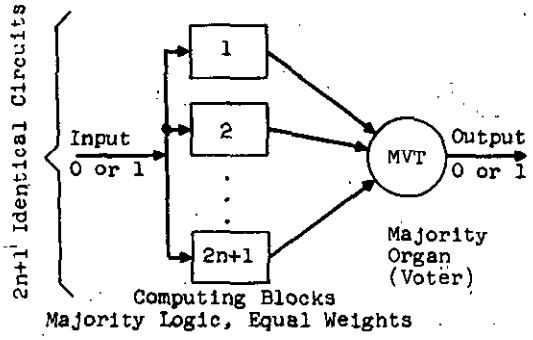


FIGURE 3-51

TYPICAL CIRCUITS

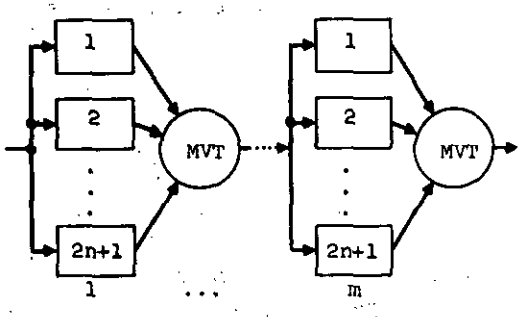


FIGURE 3-52

REDUNDANT DIGITAL SYSTEM USING M MAJORITY VOTE TAKERS AS RESTORING ORGANS

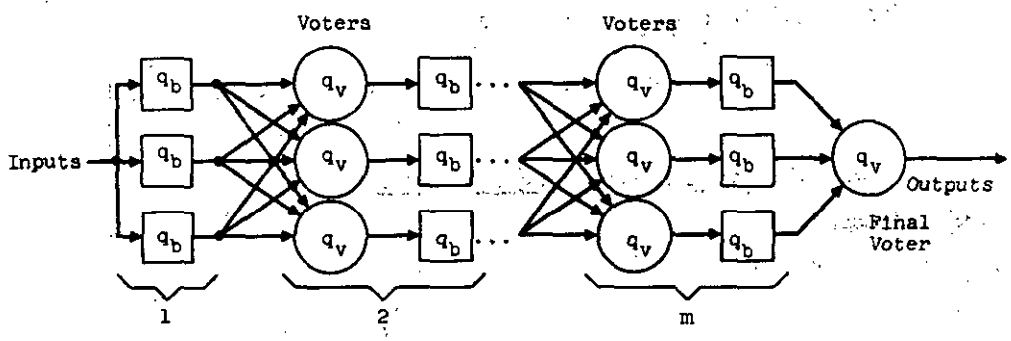


FIGURE 3-53

THREEFOLD MAJORITY LOGIC

Mathematical Models

Discrete Requirements. Let p denote the probability that a circuit is operating properly and let q = (1 - p) be the probability that the circuit has failed; then the probability of failure for the majority group is as follows:

$$q_n = \sum_{i=n+1}^{2n+1} \binom{2n+1}{i} p^{2n+1-i} q^i$$

where

$$\binom{2n+1}{i} = \frac{(2n+1)!}{i!\,(2n+1-i)!} = \binom{2n+1}{2n+1-i}$$

under the assumptions that the vote taker is infallible and that circuit failures are independent.

Similarly, the probability of success (no failure) for the majority group is

$$p_n = \sum_{i=0}^{n} \binom{2n+1}{i} q^i \, p^{2n+1-i} = 1 - q_n$$

Now if the vote taker is not infallible, and q_v is the fault probability of the voter, the failure probability of the redundant block is

$$1 - R_2 = 1 - (1 - q_v)\left[1 - \sum_{i=n+1}^{2n+1} \binom{2n+1}{i} q^i (1 - q)^{2n+1-i} \right]$$

$$\approx q_v + \binom{2n+1}{n+1} q^{n+1}$$

The lower degrees of redundancy give the following approximate failure probabilities:

Degree of Redundancy	Approximate Failure Probability
3	$1 - R = q_v + 3q^2 - 2q^3$
5	$1 - R = q_v + 10q^3 - 15q^4 + 6q^5$
7	$1 - R = q_v + 35q^4 - 84q^5 + \ldots$
9	$1 - R = q_v + 126q^5 - 420q^6 + \ldots$

It is apparent that higher degrees of redundancy are futile, since the voter fault probability q_v soon becomes limiting. In fact, even for threefold redundancy, q_v becomes the major contributor if q is reasonably small in the non-redundant case.

If majority logic is applied to each block, and every voter is triplicated except the last one, the resultant failure probability for the general case, using a 2n+1 fold majority logic and m blocks, is as follows:

$$1 - R = (1 - q_{tv}) \left[1 - \sum_{i=n+1}^{2n+1} \binom{2n+1}{i} q_b^i \, p_b^{2n+1-i} \right]$$

$$\left[1 - \sum_{i=n+1}^{2n+1} \binom{2n+1}{i} (q_b + q_v - q_b \, q_v)^i \, (1 - q_b - q_v + q_b \, q_v)^{2n+1-i} \right]^{m-1}$$

If it is assumed that all the failure probabilities are reasonably small, this becomes

$$1 - R \approx q_{fv} + \binom{2n+1}{n+1} (q/m)^{n+1} + (m-1) \binom{2n+1}{n+1} (q_v + q/m)^{n+1}$$

where q is the nonredundant probability of failure for the system.

In particular, for threefold majority logic, that is, n = 1, the probability is

$$1 - R \approx q_{fv} + 3 (q/m)^2 + 3(m-1) (q_v + q/m)^2$$

If the vote taker is considered to be ideal* rather than infallible, where p_v equals the probability that a vote taker is working properly, and λ_v equals failure rate of a vote taker, and it is assumed that the number of vote-taker failures in a given length of time obeys the Poisson distribution, then

$$p_v(t) = e^{-\lambda_v t}$$

and the probability that m vote takers are working properly is given by

$$R = [p_v(t)]^m = e^{-\lambda_v m t}$$

It is assumed that if a vote taker is not functioning correctly, its output will be the complement of the correct output.

If the failure rate of the vote takers is too large to be neglected, redundant vote takers can be used. In this case, the failure rate of an individual circuit can be considered to include the circuit and the vote taker feeding that circuit. The overall system then becomes equivalent to a system using nonredundant ideal vote takers. If the probability of survival for an individual circuit is given by

$$p = p_v p_o = (e^{-\lambda_v t}) (e^{-\lambda_o t}) = e^{-(\lambda_v + \lambda_o) t},$$

then

$$R = \left[\sum_{i=0}^{n} \binom{2n+1}{i} q^i p^{2n+1-i} \right]^m$$

which is equivalent to the probability of success for m majority groups.

---

*The vote taker is considered ideal if

$$\lambda_v m T_o \ll 1$$

where T_o is the time of interest.

It can be shown* that the maximum reliability is achieved with nonperfect vote takers if $\lambda_o/\lambda_v = 1/n$

where

λ_o = failure rate of the circuit

λ_v = failure rate of the vote taker

and

n = (2n + 1) identical circuits or elements

It is generally necessary to carry system output on a single line, in which case the redundancy scheme proposed by Moore and Shannon could be used to improve the reliability of system output, thus eliminating the final voter from the analytic expression.**

Microelectronics Requirements. The mathematical models presented above may not be applicable to microelectronic circuits. For example, if a 2n+1-fold majority logic were to be placed on a single chip, consideration would have to be given to whether the failure distribution on the chip is continuous, discrete, or a combination of the two. Also, if all the m blocks are on one chip, it may not be appropriate to divide q by m in the above models. To develop an appropriate model, it would be necessary to know how the system is partitioned among the chips and what the probabilities of the various modes of failure are.

If the system contained redundant majority vote takers, problems would arise in placing the paths between the circuit and the vote takers on the chip. In most instances it would be impossible to eliminate crossovers, which would mean additional complexity in constructing the circuit because multiple layers required by the crossovers on the chip are currently difficult or impossible to achieve. A combination of monolithics and multilayer circuit boards may be a feasible solution.

Physical Properties

Time. This form of redundancy can be applied in situations that call for either intermittent or continuous operation in time.

Performance. This form of redundancy is usually associated with binary inputs and output.

*Knox-Seith, J. K., A Redundancy Technique for Improving the Reliability of Digital Systems, Technical Report No. 4816-1, Stanford Electronics Laboratories.

**Technical Report No. 4816-2.

Reliability Obtainable

Gain. Assuming ideal vote takers, the digital system will be most reliable if majority logic is applied at as low a level as possible, i.e., when the system is divided into as many digital subsystems, each followed by a majority vote taker, as possible.

On the other hand, it is clear that the MTTF* for the system will always be less than the MTTF for the individual circuit. In the limit as $n \to \infty$ the system MTTF could be 0.69 times the MTTF for the individual circuit.

It is seen that the use of redundancy and majority logic gives the greatest improvement in reliability in the case of large systems, i.e., in systems for which it is possible to achieve large values of m.

Nuisance Factors. It is emphasized that the full improvement in reliability is realized only if all circuits are working properly at time $t = 0$, that is, at the time when the mission is about to start.

There is another basic problem with this approach to the application of majority logic. Unless the nonredundant fault probability, q, is small, very high degrees of redundancy are required to reduce 1-R. Indeed, it can be shown that for $q > 0.5$, any degree of majority-logic redundancy will actually degrade reliability. Also, if nonredundant vote takers of limited reliability are used anywhere in a redundant system, they will for some period of time constitute the most likely source of system failure.

Multiple-Line Networks**

Definition of Form

Multiple-line redundancy has been studied extensively by Westinghouse and found to be one of the most efficient types of circuit redundancy. Multiple-line redundancy is applied by replacing the single circuit of a nonredundant network by nonidentical circuits operating in parallel, where m is called the order of the redundancy.

The reliability improvement expected with the use of redundant circuits depends on the ability of the network to experience circuit failures without degradation of the network operation. The use of restorers within the network provides this characteristic. The restorer consists of m-restoring circuits which, when operating correctly, have the ability to derive the correct output if k of its m inputs are correct

*Mean time to failure.

**P. A. Jensen, "The Reliability of Redundant Multiple-Line Networks," IEEE Transactions on Reliability, March 1964; and M. K. Cosgrove et. al., "The Synthesis of Redundant Multiple-Line Networks," AD 602749, 1 May 1964.

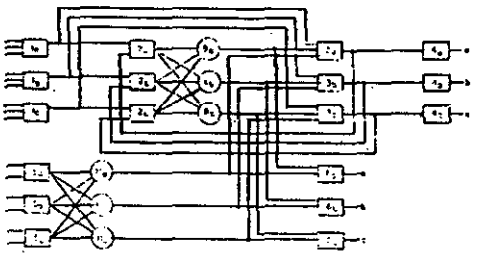


FIGURE 3-54

MULTIPLE-LINE REDUNDANT NETWORK WITH RESTORERS AFTER FUNCTIONS 2 AND 5

Typical Circuits

See Figure 3-54.

Reliability Block Diagram

See Figure 3-54.

Mathematical Models

Discrete Requirements .

(1) The reliabilities of the circuits in the network are independent.

(2) Only an approximation to the exact reliability will be given, and it is based on techniques developed by Proschan and Esary.*

(3) The approximation given is good if the reliabilities of the circuits in the network are close to 1.

(4) The approximation is based on the concepts of minimal cuts, discussed previously, and coherent systems. A system is coherent if it meets the following four conditions:

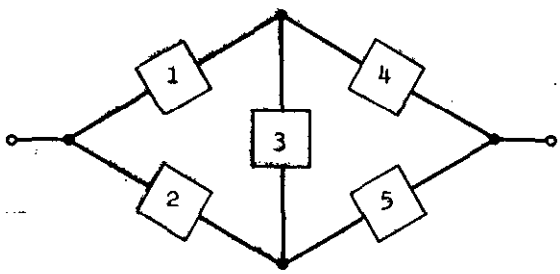
(a) If a group of circuits in the system is failed, causing the system to fail, the occurrence of any additional failure or failures will not return the system to a successful condition.

(b) If a group of circuits in the system is successful and the system is successful, the system will not fail if some of the failed components are returned to the successful condition.

(c) When all the circuits in the system are successful, the system will be successful.

(d) When all the circuits in the system fail, the system fails.

*J.D. Esary and F. Proschan, "The Reliability of Coherent Systems," in Redundancy Techniques for Computing Systems, R. H. Wilcox and K.C. Mann, Eds., Spartan Books, Washington, D.C., pp. 47-61, 1962: and J.D. Esary and F. Proschan, "Coherent Structures of Non-Identical Components," Technometrics, Volume 5, (May 1963) pp. 191-209.



The system shown at left is an example of a coherent system.

The lower bound to system reliability is the probability that none of the system's minimal cuts fail; for the example it is given by the following equation, where

R_s = system reliability

$R(C_i)$ = probability of success for the c^{th} circuit

$Q(C_i) = 1 - R(C_i)$:

$$R_s = \left\{1 - \left[1-R(C_1)\right]\left[1-R(C_2)\right]\right\}\left\{1 - \left[1-R(C_4)\right]\left[1-R(C_5)\right]\right\}$$
$$\times \left\{1 - \left[1-R(C_1)\right]\left[1-R(C_3)\right]\left[1-R(C_5)\right]\right\}\left\{1 - \left[1-R(C_2)\right]\left[1-R(C_3)\right]\left[1-R(C_4)\right]\right\}$$
$$= \left[1-Q(C_1)\,Q(C_2)\right]\left[1-Q(C_4)\,Q(C_5)\right]\left[1-Q(C_2)\,Q(C_3)\,Q(C_4)\right]$$

This is an approximation because the failure of minimal cuts is assumed to occur independently, which is generally not true, since one component may appear in several minimal cuts.

If the j^{th} minimal cut is denoted by the set S_j, and the members of the j^{th} minimal set are given by $i \epsilon S_j$, then $\prod_{i \epsilon S} \left[1-R(C_1)\right]$ is the general relationship for the probability of failure for the j^{th} minimal cut. The lower bound to the system reliability is given by

$$R_s = \prod_j \left\{1 - \prod_{i \epsilon S_j} \left[1-R(C_1)\right]\right\}$$

Thus the determination of the lower bound on reliability requires that the minimal cuts of the network be identified. In a multiple-line network with restorers, a cut is any group of circuits whose failure causes the outputs of at least one restored function to have m-k+1 or more failed lines. This would constitute a network failure.

The minimal cuts of a multiple-line redundant network have three characteristics that are sufficient to establish their identity:

(1) All the members of the minimal cut are circuits in a restored function or restorers that are the input sources of that restored function.

(2) The failure of each member of the minimal cut will cause one output line of the restored function to be in error, and each member will be in a different position.

(3) The failure of a minimal cut will cause exactly $m-k+1$ output lines of the restored function to be in error; hence a minimal cut will have $m-k+1$ members.

If one lists for each of the restored functions in the network all the sets of circuits that fulfill these characteristics, he will have all of the minimal cuts of the network, which will enable him to find the lower bound for the network reliability.

Physical Properties

Time. Multiple-line redundancy is intended primarily for digital circuits operating on binary information.

Performance. Multiple-line redundancy should result in improved reliability of the system unless the individual circuit reliabilities are very low. Low circuit reliabilities would cause the restorers to choose the wrong value if k of the m circuits have failed.

Reliability Obtainable

Gain. The improvement in system reliability should be comparable to the improvement in the reliability of a circuit when a particular element is made redundant. The improvement will not be of the same magnitude, because of the addition of restorers in the multiple-line network.

Nuisance Factors. The lower limit approximation given for the multiple-line network is not good if the circuit reliabilities are not close to 1. If the order of the redundancy exceeds 3, the determination of the input sources becomes quite difficult. Boolean matrices can be used for determining the input sources of a function.

3.8.3.6 Decision With Switching Redundancy

Standby Redundancy

Definition of Form

Circuitry in which a component or unit is standing by idly and begins to operate only when the preceding unit fails is said to be using standby or sequential redundancy. A standby system usually requires failure-sensing and/or switching networks or devices to put the next unit into operation.

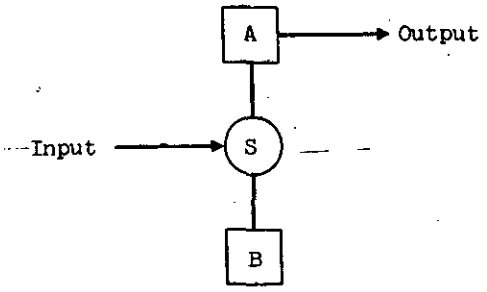


FIGURE 3-55

STANDBY REDUNDANCY

Typical Circuits

Figure 3-55 shows two redundant elements where A is operating and B is waiting until A fails, and S is the sensing and switching mechanism.

Reliability Block Diagram

See Figure 3-55.

Mathematical Models*

Discrete Requirements.

Consider a system connected by a sensing and switching mechanism (S) to another system (B). If A fails, then S senses the failure and switches B into operation.

The device, ASB, operates in the following mutually exclusive ways:

(1) S is operating properly. It monitors A, and if A fails, it turns B on, and the device operates until B fails.

(2) S fails by not being able to sense and/or switch, and when it fails, A is operative and the device fails when A fails.

(3) S fails and in failing it switches to B. A is still operating when S fails, but the device fails when B fails.

(4) A is operating and S fails. The signal path through S becomes open or shorted and the entire device fails at the time S fails.

The unreliability Q(t) will be derived where $Q(t) = 1 - R(t)$, let $\phi_1(t)$ be the failure density function for A, $\alpha(s)$ for S, and $\phi_2(t)$ for B_1 where $\phi_1(t) = \phi_2(t)$ is not required.

p_1 = probability that S fails and the switch stays on A

p_2 = probability that S fails and the switch goes to B

p_3 = probability that S fails in such a way that the signal path is shorted or open, $p_1 + p_2 + p_3 = 1$

For case (1),

$$Q_1(t) = \int_{t_2=0}^{t} \int_{t_1=0}^{t-t_2} \left[1 - \int_{s=0}^{t_1} \alpha(s)ds\right] \phi_1(t_1)\, \phi_2(t_2)\, dt_1\, dt_2$$

*Aroian, L. A., "The Reliability of Items in Sequence with Sensing and Switching," in Redundancy Techniques for Computing Systems, edited by R. H. Wilcox and W. C. Mann, Spartan Books, Washington, D. C., 1962, pp. 318-327.

where $1 - \int_{s=0}^{t} \alpha(s)ds$ is the probability that the device S is operating correctly at t, when A fails.

For case (2),

$$Q_2(t) = p_1 \int_{s=0}^{t} \alpha(s)\left[\int_{t_1=s}^{t} \phi_1(t_1)\, dt_1 \right] ds$$

For case (3),

$$Q_3(t) = p_2 \int_{t_2=o}^{t} \phi_2(t_2)\, dt_2 \int_{s=o}^{t-t_2} \left[1 - \int_{t_1=o}^{s} \phi_1(t_1)\, dt_1 \right] \alpha(s)\, ds$$

For case (4),

$$Q_4(t) = p_3 \int_{s=o}^{t} \left[1 - \int_{t_1=o}^{s} \phi_1(t_1)\, dt_1 \right] \alpha(s)\, ds$$

Hence for the entire device, $Q(t) = Q_1(t) + Q_2(t) + Q_3(t) + Q_4(t)$ and $R(t) = 1 - Q(t)$.

Error Protection

Standby redundancy protects against hardware failures.

Physical Properties

Time. Standby redundancy is applied in situations that require continuous operation in time.

Performance. For the special case of the exponential failure law where γ = failure rate of the switching mechanism, and μ = the failure rates of the two systems (A and B), standby redundancy is better than 2 systems in parallel if $\mu > \gamma$. If $\mu = \gamma$, the two types of redundancy are equal; and if $\mu < \gamma$, parallel redundancy is superior.

Reliability Obtainable

Gain. The gain for a specified mission can be measured in terms of the ratio of the reliability of the structure with standby redundancy to the reliability of alternate structures.

Nuisance Factors. Whenever the ratio calculated for gain measurement is less than unity, standby redundancy has no utility. Also, the failure rate of the switching mechanism can overcome any gain using standby redundancy.

Operating Redundancy*

Definition of Form

In operating switching redundancy, independent identical units operate simultaneously with a common input. A failure detector is associated with each unit, and a switch is connected to the outputs. All units are operating initially, and the output of one unit is used until that unit fails. The switch then steps to the next operating unit and remains there until that unit fails.

Typical Circuits

Figure 3-56 shows a typical switching circuit; C represents the redundant components and D the individual detectors.

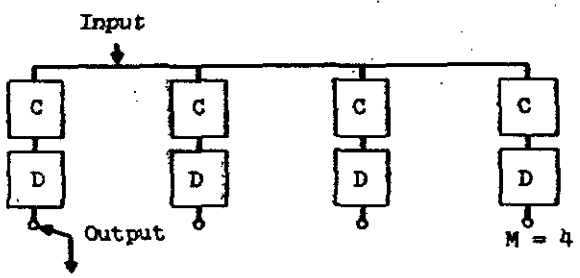


FIGURE 3-56

SYSTEM CONSISTING OF M REDUNDANT CHAINS

Reliability Block Diagrams

See Figure 3-56.

Mathematical Models

Discrete Requirements. The following assumptions will be made to compute system reliability:

(1) There are m chains ordered 1, ..., m; all m operate from the initial time until each fails.

(2) The stepping switch is connected so that its inputs are the outputs of the m chains, with the output of the switch being the output of the system. The switch operates sequentially, starting with chain 1. The switch will also indicate when all m chains have failed.

(3) A failure-detecting device operates in conjunction with each chain and performs the following functions:

(a) If failure occurs in the chain to which the switch is connected, a signal is immediately sent to the switch, causing it to step.

(b) If a failure occurs in a chain to which the switch is not connected, a signal is stored; and if the switch steps to that chain, it is signaled to step once more.

*B. J. Flehinger, "Reliability Improvement through Redundancy at Various System Levels," IBM Journal, April 1958, pp. 148-158.

(4) No significant time is consumed by the failure-detecting and switching operations.

(5) The reliability of a chain is the product of its components:

$$R_c = \prod_{i=1}^{n} R_i$$

The reliability of the system depends on the reliabilities of the chains, the failure detectors, and the switches. For the detectors and switches, there are two modes of behavior with which reliabilities are associated. The first class of reliability (D_a and S_a on Figure 3-57) is a probability that the device operates when failure occurs. This function can be performed only once for each chain, and the probability is defined for a single operation that takes place in negligible time. The second class of reliability (D_b and S_b on Figure 3-55) is a probability that the device does not spontaneously operate during a period of time in which no failure occurs. This type of probability, like the reliability of a chain, is defined for the length of time required for the machine to complete the assigned task. Thus the following probabilities are defined:

R_c is the reliability of the chain, i.e., the probability that it performs its functions adequately for the duration of the assigned task.

D_a is the conditional probability that when a failure occurs in a chain, the failure is detected and a signal is sent to the switch under conditions a or b. (A factor of D_a is the probability that the switch control is connected to the error detector for the chain at which the switch is positioned.)

D_b is the conditional probability that when no failure occurs in a chain for the duration of the task, no signal is transmitted to the switch when it is positioned at that chain.

S_a is the conditional probability that when the switch receives a failure signal, the connection at which it stands is broken and a good connection is made to the next chain.

S_b is the conditional probability that if the switch does not receive a failure signal for the duration of the task, it does not step at any time during the run. (It is assumed that if it does step, it makes contact on the next chain.)

S_c is the conditional probability that if a good connection is made every time the switch steps, a good connection exists between some chain (or the device indicating system failure) and the system output at all times during the run. (It is assumed that switching occurs in zero time, so that switching can be thought of as making contact over the entire time interval except for a finite number of points in time.)

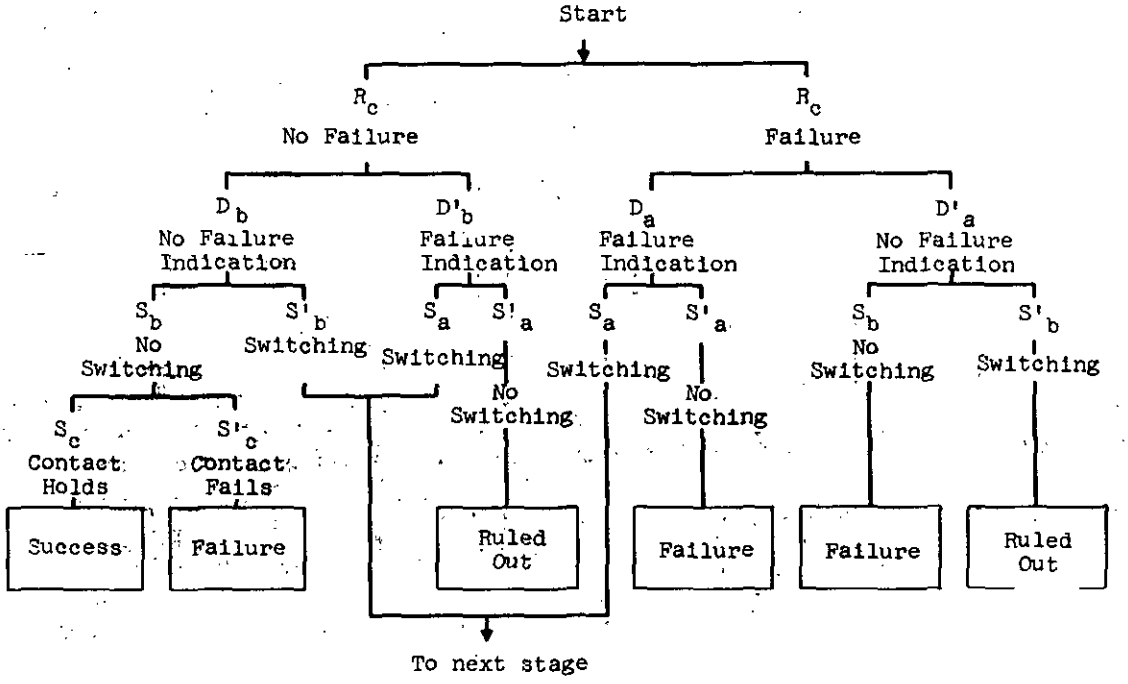


FIGURE 3-57

FAILURE DIAGRAM OF A CHAIN

The reliability of the system of m redundant chains is defined as the probability that it performs the assigned task successfully. This occurs if, for the duration of the task, the switch constantly (except for the points in time required for switching) makes a good connection to a chain that is functioning adequately. This can take place in m mutually exclusive ways, corresponding to the final connection to the m switch contact.

The possible modes of behavior of a chain are diagrammed in Figure 3-57. Successful operation through a given chain requires that the chain function adequately (R_c), that the failure detector not signal an error (D_b), that the switch not step spontaneously while connected to this chain (S_b)*, and that the switch contact remain good (S_c). This has probability $R_1 = R_c D_b S_b S_c$.

A stepping of the switch can occur in three ways:

(1) The chain fails ($F_c = 1 - R_c$), the detector signals failure (D_a), and the switch steps (S_a).

*An approximation is made by the use of one value of S_b for the probability of no spontaneous stepping of the switch from any position. A precise analysis would use S_b as defined above only for the first chain, with successively larger values for this probability for chains 2, ... m. Thus the final reliability is somewhat lower than the correct result. However, since the probability of spontaneous switching in all practical applications will be exceedingly small, the more precise analysis does not seem to be warranted.

(2) The chain does not fail (R_c), but the detector erroneously signals failure ($D'_b = 1 - D_b$), and the switch steps (S_a).

(3) The chain does not fail (R_c), the detector does not signal failure (D_b), but the switch steps spontaneously ($S'_b = 1 - S_b$).

Thus, the probability of one stepping of the switch is

$$\alpha = (1 - R_c)\, D_a S_a + R_c\, (1 - D_b)\, S_a + R_c D_b\, (1 - S_b)$$

There are several modes of behavior of one chain that lead immediately to system failure without any failure indication, because of a bad switch contact (S'_c), to failure of the switch to respond to an error signal (S'_a), or to failure of the detector to indicate failure (D'_a). In addition, there are modes of behavior in which the detector and switch both make errors that cancel each other. These second-order effects will be arbitrarily ruled out.

Now the probability of successful operation with the final connection to the i^{th} switch contact is equal to the probability of i-1 steppings of the switch times the probability of successful operation through one chain, or $\delta^{(i-1)}R_1$.

Then the reliability of the system is the sum of these probabilities for the m switch contacts:

$$R = \sum_{i=1}^{m} \delta^{(i-1)} R_1 = R_1 \frac{1-\delta^m}{1-\delta}$$

or

$$R = R_a D_b S_b S_c \frac{1 - [(1 - R_c)\, D_a S_a + R_c (1 - D_b)\, S_c + R_c D_b (1 - S_b)]^m}{1 - [(1 - R_c)\, D_a S_a + R_c\, (1 - D_b)\, S_a + R_c D_b\, (1 - S_b)]}$$

where

$$R_c = \prod_{i=1}^{n} R_i$$

Since $D_a \leq 1$ and $S_a \leq 1$,

$$R \leq R_c D_b S_b S_c \frac{1 - (1 - R_c D_b S_b)^m}{1 - (1 - R_c D_b S_b)}$$

so that

$$R \leq S_c$$

Furthermore, since $S_c \leq 1$, $D_b \leq 1$, and $S_b \leq 1$,

$$R \leq 1 - (1 - R_c)^m$$

Thus the well-known expression corresponding to perfect failure detection and switching represents an upper limit for the reliability obtainable.

In the present application the initial machine, with no redundancy, is considered to have a reliability R_o. It is assumed that it is possible to break the machine up into p chains of equal reliability, $R_o^{1/p}$. It is further assumed that the failure detector for the complete machine consists of p units, each associated with a chain, such that indications of failure originating from any of these units are equally probable. Then, if D_a and D_b are probabilities associated with the failure detector for one complete machine, the corresponding probabilities for the unit associated with a chain will be $D_a^{1/p}$ and $D_b^{1/p}$. If each chain is made m times redundant, the reliability of the final system is

$$R_s = R_o^{1/p} D_b^{1/p} S_b S_c \times \left[\frac{1 - [(1 - R_o^{1/p}) D_a^{1/p} S_a + R_o^{1/p}(1 - D_b^{1/p}) S_a + R_o^{1/p} D_b^{1/p} (1 - S_b)]^m}{1 - [(1 - R_o^{1/p}) D_a^{1/p} S_a + R_o^{1/p} (1 - D_b^{1/p}) S_a - R_o^{1/p} D_b^{1/p} (1 - S_b)]}\right]^p$$

For perfect failure detection and switching, this becomes

$$R_s = [1 - (1 - R_o^{1/p})^m]^p$$

Physical Properties

Time. Switching operating redundancy is used in continuous time applications primarily, but it can be used in intermittent situations if the failure-detecting device is capable of signaling the switching mechanism at the proper time.

Performance. The performance in many instances will be limited by the reliability of the failure-detecting and switching assemblies.

Reliability Obtainable

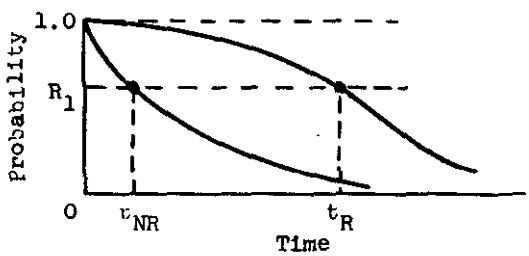
Gain. Flehinger provides tables and charts so that, given an estimate of the initial unreliability, F_o, of a nonredundant system, and the tolerable unreliability, F_s, to be permitted in the final system, one can judge from the appropriate curve the degree of redundnancy, m, and the number of chains, p, that will meet the specifications, assuming that redundancy is required.

Nuisance Factors. For initially unreliable systems and a moderate degree of redundancy, high-reliability can be achieved only by applying the redundancy to relatively small units. Imperfect switching limits the reliability attainable in all cases such that the unreliability is not a steadily decreasing function of p but has a definite minimum beyond which it increases.

3.8.4 Measures of Redundancy Gain

Several measures of redundancy gain have been used in the past. Among these were the following:

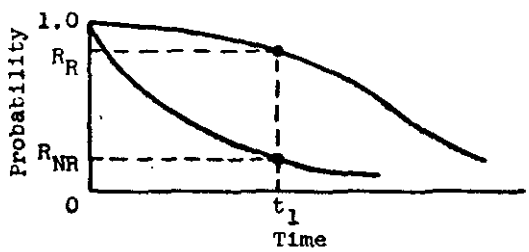
(1) The ratio of the time to failure of the redundant structure to the time to failure of the nonredundant structure at a fixed reliability (R) or a fixed unreliability (Q) (Figure 3-58).



Measure:
t_R/t_{NR} at R_1 or
t_R/t_{NR} at Q_1 where
$Q_1 = 1-R_1$

FIGURE 3-58
TIME-GAIN MEASURE

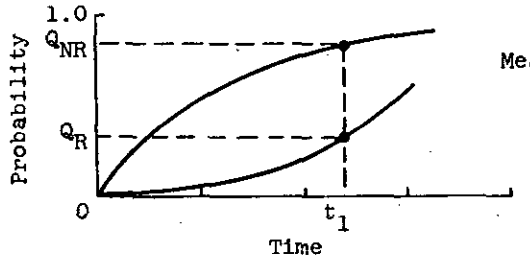
(2) The ratio of the reliability of the redundant structure to the reliability of the nonredundant structure at a fixed time (Figure 3-59).



Measure:
R_R/R_{NR} at
time (t_1)

FIGURE 3-59
PROBABILITY-GAIN MEASURE

(3) The ratio of the unreliability of the nonredundant structure to the unreliability of the redundant structure at a fixed time (Figure 3-60).



Measure:

$Q_{NR}/Q_R = (1-R_{NR})/(1-R_R)$ at time (t_1)

FIGURE 3-60

UNRELIABILITY-GAIN MEASURE

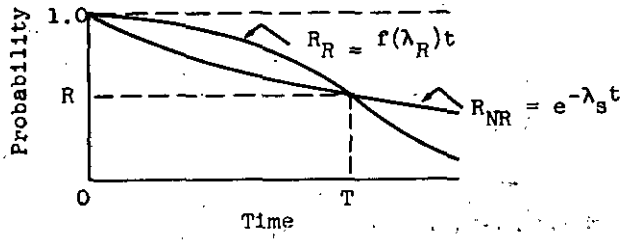
(4) The ratio of mean time between failures (MTBF) of the redundant structure to the MTBF of the nonredundant structure or, conversely, the ratio of the failure rates (λ) of the nonredundant structure to the redundant structure.

Measure:

$(MTBF)_R/(MTBF)_{NR}$ at time (t_1) or λ_{NR}/λ_R at time (t_1)

A review of the meaning and physical interpretation of the four measures reveals a shortcoming: Unless the exponential distribution is assumed, each measure fails to compare the effect of obtaining a specific reliability at a specific mission time. This failure does not give the proper interpretation of the effect or purpose of redundancy.

The proposed measure of gain is the ratio of the component failure rate of the redundant-structure components, at a specified R and T, to a theoretical component failure rate of a nonredundant structure that gives the same R and T (Figure 3-61).



Measure:

λ_R/λ_S at $t_R = t_{NR} = T$ when $R_R = R_{NR} = R$

FIGURE 3-61

COMPONENT-FAILURE-RATE GAIN MEASURE

This gain measure satisfies the purpose of redundancy, i.e., the use of components of relatively high failure rate to perform a task that would normally require extremely reliable components. Moreover, the effect of redundancy is to permit the use of real components where it would be impossible to use a nonredundant structure because of failure-rate limitations.

3.9 MAINTAINABILITY AND LOGISTICS

3.9.1 Introduction

The purpose of this section is to provide guidelines for achieving maximum maintainability and supportability when microelectronic devices are employed in a system/equipment design. Guidelines are provided in two major areas: maintenance concepts and maintenance techniques. Subjects discussed under maintenance concepts include selection of replacement level, discard or repair, spares concepts, location of repair facilities, and skill requirements. In the maintenance-technique area, diagnostic procedures and repair procedures are presented. The use of a Manning and Support Model* and a Discard-at-Failure Maintenance (DAFM) versus Repair-at-Failure Maintenance (RAFM) Model** is referenced where applicable.

3.9.2 Maintenance Concepts

The initial maintainability/logistics task in a system development program is to establish a maintenance concept. This concept is the basis for succeeding development/design decisions that affect maintenance and logistic support. The elements of the maintenance concept include the level of equipment assembly replaced at each maintenance echelon; whether the replaced assembly is repaired or discarded; the location, types, and quantities of spare assemblies/parts; the location, number, and types of repair facilities; and the number and types of maintenance personnel at each facility. Although the concept must be established early in a development program, it should be made sufficiently flexible to allow changes, if necessary, as the program progresses. In many programs, a number of the maintenance elements are fixed before inception, and the remaining elements and a design concept must be selected to enhance the use of the predetermined elements.

Selection of each element of the maintenance concept is presented individually in the following paragraphs. However, the selection of a maintenance method for any one element affects the selection of a method for the other elements; the interactions must be considered. It is usually necessary to make preliminary selections for each element and then change each element in turn to make it compatible with the others, until the most effective overall concept is achieved.

---

* Queuing Tables for Determining System Manning and Related Support Requirements, December 1964, DDC Document AD458206.

** Validation of Discard-at-Failure Maintenance Mathematical Model, February 1966, DDC Document AD479903.

3.9.2.1 Selection of Replacement Level

The possible levels of replacement can be generalized to the following categories:

- Component Part
- Module (subassembly)
- Assembly (black box)
- Equipment

A component part is an item that cannot be physically separated without destruction of its function. A module is a group of parts mounted on a common structure that can be replaced as a unit. An assembly is a group of modules, or modules and parts, that can be replaced as a unit and that performs a unique function for the system. An equipment is a group of lower-level items that fulfills an essential operational requirement.

It is possible to restore a malfunctioning system or equipment to operational capability by eliminating the cause of failure through replacement at any one of the above levels. The selection of the replacement level is contingent on the system-availability requirements, the level of skill of available personnel, the type of spares provisioned, the capability of the maintenance organization, and the available resources. The number and types of maintenance echelons available are also important considerations in selection of the replacement level. Maintenance tasks can be divided among the echelons to increase system availability and reduce support costs.

Replacement-Level Effect on Maintenance

In general, the higher the level of replacement for operating equipment, the greater the availability achieved and the lower the maintenance skill and test-equipment requirements. If repair is effected at the equipment level, downtime can be reduced significantly through automatic fault detection and substitution. This method can be used at lower levels also but with the penalty of increased equipment complexity and decreased equipment reliability. At the opposite end of the scale, if repair is effected at the part level, equipment downtime is the greatest (lowest availability) and requirements for maintenance skills and test equipment are maximum. However, higher levels of replacement generally require greater expenditures for spares procurement, handling, and storage.

The greatest contributor to active downtime is the time required to isolate the fault to a replaceable unit. As the complexity of the replaceable unit is reduced, the number of units to select from becomes greater. It has been found that for a given maintenance concept and design philosophy, the time to locate

the cause of failure varies approximately as the log to the base two of N, where N is the number of replaceable units in the equipment. A desired level of availability can be obtained by making trade-offs between the level of replacement and the degree of diagnostic capability provided (test points, built-in test equipment, automatic fault location, etc.). An additional consideration in the selection of the replacement level is whether discard-at-failure maintenance (DAFM) or repair-at-failure maintenance (RAFM) is to be used. This consideration is discussed in detail in Section 3.11.2.2.

When more than one echelon of maintenance is provided, high equipment availability can be obtained through high-level replacement, while spares requirements can be minimized through subsequent repair of the replaced unit. Through repair of the replaced units for a number of equipments at a centralized location, the total requirements for maintenance skills and test equipment can be reduced. The Manning and Support Model can be used to determine optimum replacement levels and division of maintenance tasks. The procedure used is to enumerate all of the practical alternative replacement levels and task divisions that will provide the desired availability level and then select the one that requires the smallest amount of resources (spares, maintenance personnel, test equipment, and facilities). In most situations, many of the elements of the maintenance concept are predetermined, and thus the practical alternatives and the number of comparisons to be made are limited.

Recommended Monolithic-Circuit Replacement Level

The inherent advantages of monolithic circuits can be realized through replacement at the module level or higher level. The increased reliability and reduced circuit cost should make it feasible to replace modules containing from 5 to 300 IC's depending on the overall maintenance concept. The replaced module could be either repaired or discarded depending on the economics of the particular situation. The chief advantage of this policy is the elimination of testing and replacement at the individual IC level. Replacement of individual IC's would require special skills, test equipment, and tools, and would be detrimental to equipment availability because of the additional time necessary to locate and replace a single IC.

In the packaging of LSI devices, the minimum module would contain at least one device if the device were highly complex, and two or more for less complex devices. The design of both LSI and IC modules should be such that they perform unique functions, and the number of different types should be minimized. This design concept will enhance the capability for diagnosing equipment malfunctions and reduce logistic-support requirements. The concept also will facilitate the use of DAFM and logistic self-support, which are discussed in detail in a later chapter.

3.9.2.2 Discard Versus Repair

At any level of maintenance there are two alternatives: (1) to repair a failed item and return it to the spares inventory, or (2) to discard the failed item and procure a replacement. The failed unit can be repaired locally or sent to another echelon. The general advantages of DAFM are: (1) reduced skills requirements, (2) fewer tools, (3) less complex and fewer types of test equipment, and (4) simpler administrative procedures. The disadvantages are higher costs of spares and increased storage facilities. The decision for DAFM versus RAFM is based on the lowest total support cost.

In general, DAFM is considered the best approach for monolithic-circuit equipment. The lower circuit cost and higher reliability of IC's allow the discarding of more complex modules. Other advantages to be realized from the general use of DAFM are the elimination of depot maintenance, reduced training requirements for maintenance personnel, and simplified support systems. As the level of DAFM is raised, the maintenance requirements will be reduced until the need for a maintenance organization is completely eliminated. This will be accomplished through self-checking equipment, with replacement and discard of assemblies or whole equipments by the operator personnel.

3.9.2.3 Logistic-Support Concepts

There are three general supply concepts: (1) sparing at one level (parts) with periodic resupply, (2) sparing at more than one level with periodic resupply, and (3) self-support (initial provisioning of spares for the planned equipment life time). Each of these concepts can have variations in cycle periods (except self-support), level of confidence, types of spares, and storage locations. The self-support concept is basically the same as periodic resupply, with the cycle period equal to the planned equipment life. The selection of a support concept is based on achieving the desired level of equipment availability at the lowest cost. Controlling factors are other elements of the overall maintenance concept and aspects of the support system that have been predetermined.

Self-Support

The self-support concept involves a number of techniques by which sufficient spares are provided concurrent with the delivery of the equipment to maintain the equipment in operational status throughout its useful life. It can be accomplished by providing spare replaceable units separately, by mounting spare parts on each module of the equipment, or by providing sufficient redundancy to preclude equipment failure. With present circuit reliabilities and costs, the latter two methods will probably not be feasible in most applications. However, with certain types of equipment [highly repetitive part (IC) types and high reliability], it may be feasible to provide spares as part of each installed module.

The number of separate spares required initially to ensure (to a desired level of confidence) no equipment downtime because of lack of spares has been calculated for various unit reliabilities and equipment complexities.* The calculations are based on an assumption of a constant failure rate for the parts under consideration; the probability, then, of having r or fewer failures in time t is given by

$$P(r) = \sum_{N=0}^{r} \frac{(\lambda t)^N}{N!} e^{-\lambda t}$$

where

N = Number of failures

λ = Failure rate of parts being considered

t = Operating time

Table 3-7 illustrates the results of for an equipment life of 2×10^4 hours (approximately ten years). The captions used in the table are defined as follows:

Equipment Life -- Required equipment operating life (hours)

Part MTBF, Hours -- Mean time between failures (hours) of each part (IC, MFC, or LSI)

Probability Level -- Probability that the equipment will not have any failures for which a spare part is not available during the required equipment life

Total Parts per Equipment -- The total number of parts (IC or LSI) used in the equipment

Number of Circuit Types -- Number of different types of circuits (modules) used in the equipment

Two numbers appear for each probability level listed in the table. The upper number is the number of spare circuits required for maximum use of a single circuit type. For example, if the equipment contains a total of 100 circuits of 10 different types, the upper number would be the spare circuits required when one circuit type is used 91 times and each of the other 9 types is used once. The lower number is the number of spares needed for equal use of all circuit types. While these two situations may not necessarily correspond to the part-type combinations for absolute maximum and minimum spare-parts requirements, they do represent extremes that will include most combinations of part types.

*Investigation of Factors Affecting Early Exploitation of Integrated Solid Circuitry, ASD-TDR-7-998-2, Contract AF 33(657)-8785, 12 October 1962 -- 1 January 1963, ARINC Research Publication 234-2-342.

TABLE 3-7
SPARE PARTS REQUIREMENTS FOR LOGISTIC SELF-SUPPORT
(Equipment Life = 2 x 10^4 Hours)

MFM MTBF Hours	Probability Levels	Total Parts per Equipment: 100			Total Parts per Equipment: 500					Total Parts per Equipment: 1000					Total Parts per Equipment: 5000					Total Parts per Equipment:10,000				
		Number of Part Types			Number of Part Types					Number of Part Types					Number of Part Types					Number of Part Types				
		1	10	100	1	10	100	250	500	1	10	100	250	500	1	10	100	250	500	1	10	100	250	500
10^3	.50	20	40	200	100	150	400	750	1000	200	270	600	1000	1500	1000	1150	1800	2750	3500	2000	2210	3100	4250	5500
		-	28	-	-	109	279	551	-	-	210	380	653	1110	-	1014	1183	1461	1935	-	2017	2185	2467	2951
	.75	23	50	200	107	160	500	750	1500	209	290	700	1250	2000	1021	1190	1900	2750	4000	2030	2270	3200	4500	6000
		-	32	-	-	119	287	569	-	-	224	392	684	1609	-	1044	1209	1540	2434	-	2060	2223	2579	3450
	.90	26	60	300	113	180	500	1000	1500	218	300	800	1250	2000	1041	1230	2000	3000	4500	2057	2330	3200	4500	6500
		-	42	-	-	130	389	818	-	-	235	495	933	1612	-	1059	1318	1789	2441	-	2076	2336	2828	3460
	.95	27	60	300	117	190	600	1000	1500	223	310	800	1250	2000	1052	1260	2000	3250	4500	2074	2370	3300	4750	6500
		-	44	-	-	134	393	818	-	-	241	500	933	1618	-	1072	1331	1789	2458	-	2095	2353	2828	3485
	.99	30	80	300	124	200	600	1250	2000	233	320	900	1500	2500	1074	1300	2100	3500	5000	2104	2430	3400	5000	7000
		-	45	-	-	149	401	1067	-	-	259	513	1182	2120	-	1100	1360	2038	2968	-	2131	2394	3077	4000
10^4	.50	2	10	100	10	30	100	250	500	20	40	200	500	500	100	150	400	750	1000	200	70	600	1000	1500
		-	2	-	-	10	107	254	-	-	21	117	264	509	-	102	197	344	590	-	203	297	445	691
	.75	3	10	100	12	30	200	250	500	23	50	200	500	1000	107	160	500	750	1500	209	290	700	1200	2000
		-	4	-	-	14	109	256	-	-	25	120	267	512	-	113	204	352	598	-	218	301	455	702
	.90	4	20	100	14	40	200	500	500	26	60	300	500	1000	113	180	500	1000	1500	218	300	800	1200	2000
		-	13	-	-	23	111	258	-	-	35	123	270	517	-	122	211	360	613	-	227	317	467	724
	.95	5	30	100	15	40	200	500	1000	27	60	300	500	1000	117	190	600	1000	1500	223	310	800	1200	2000
		-	13	-	-	24	113	261	-	-	36	125	275	1016	-	126	216	372	1112	-	233	324	484	1223
	.99	6	30	200	18	50	300	500	1000	30	80	300	750	1000	124	200	600	1250	2000	233	320	900	1500	2500
		-	15	-	-	27	213	510	-	-	39	226	524	1016	-	133	319	621	1112	-	243	429	733	1223
10^5	.50	0	0	0	1	10	100	250	500	2	10	100	250	500	10	30	100	250	500	20	40	200	500	500
		-	0	-	-	1	1	1	-	-	2	2	3	500	-	10	11	12	508	-	20	21	23	518
	.75	0	0	0	2	10	100	250	500	3	10	100	250	500	12	30	200	250	500	23	50	200	500	1000
		-	0	-	-	2	2	250	-	-	3	4	252	501	-	12	14	261	510	-	23	26	272	521
	.90	1	10	100	2	10	100	250	500	4	20	100	250	500	14	40	200	500	500	26	60	300	500	1000
		-	1	-	-	2	101	250	-	-	4	103	252	501	-	14	113	263	512	-	26	125	274	524
	.95	1	10	100	3	10	100	250	500	5	20	100	250	500	15	40	200	500	1000	27	60	300	500	1000
		-	1	-	-	3	101	251	-	-	5	103	253	502	-	16	114	264	513	-	28	126	276	525
	.99	2	10	100	4	20	100	250	500	6	30	200	250	500	18	50	300	500	1000	30	80	300	750	1000
		-	11	-	-	13	102	252	-	-	15	105	254	503	-	27	117	266	516	-	39	129	279	528
10^6	.50	0	0	0	0	0	0	0	0	1	10	100	250	500	4	20	100	250	500	9	20	100	250	500
		-	0	-	-	0	0	0	-	-	1	1	1	1	-	4	4	5	6	-	9	9	10	11
	.75	0	0	0	1	10	100	250	500	1	10	100	250	500	6	20	100	250	500	11	30	200	250	500
		-	0	-	-	1	1	1	-	-	1	2	2	500	-	6	6	8	505	-	11	12	14	510
	.90	0	0	0	1	10	100	250	500	2	10	100	250	500	7	30	100	250	500	13	40	200	500	500
		-	0	-	-	1	2	250	-	-	2	3	251	500	-	7	9	257	506	-	13	18	263	511
	.95	1	10	100	2	10	100	250	500	3	10	100	250	500	8	30	200	250	500	14	40	200	500	1000
		-	1	-	-	2	101	250	-	-	3	102	251	501	-	8	108	257	507	-	14	115	263	513
	.99	1	10	100	3	10	100	250	500	4	20	100	250	500	10	40	200	500	1000	17	50	300	500	1000
		-	2	-	-	3	101	251	-	-	5	102	252	502	-	12	109	259	508	-	18	115	265	515

The use of the table is illustrated below, with the following conditions given:

- Required equipment life: 2×10^4 hours
- Part MTBF: 10^7 hours
- Probability level: 0.99
- Total parts per equipment: 5,000
- Number of circuit types: 250

The first step is to locate the required equipment life, which appears directly under the title on the table. Then the part MTBF is located in the first column. The probability level is located next in the second column. The "total parts per equipment" is located in one of the next six major column headings, and under this the column for the number of circuit types is located. Thus the two spares numbers for the conditions assumed are listed as 500 and 266. These numbers correspond to extreme combinations of circuit-type usage, and most usage will fall between these two extremes.

No more than 15 percent of the total equipment parts would be required as spares for either extreme combination of circuit types. Two other points are of interest: (1) there is little difference in the spare-parts requirements for probabilities of equipment survival of 0.90, 0.95, and 0.99; and (2) the spares requirement usually increases at a slower rate than the number of circuits. Tables for equipment lives of less than 2×10^4 hours are given in ASD-TDR-7-998-2.

Self-support may be accomplished also by providing at least one spare for each part type in a module. These spares are mounted in the same manner as the active circuits and can be used to replace failed IC's by soldering jumper wires. For LSI, spare circuit elements can be provided that also can replace the operating devices through external jumpers. Redundancy can be used as a self-support method through the provision of two or more IC's or LSI's for each device required by the equipment. By proper selection of redundancy configuration and with a sufficient number of alternate paths, an equipment can be designed that has a very low probability of failure during its useful life. In applying redundancy to IC's, extreme caution is necessary to ensure that certain failure modes do not cause failure of the alternate paths or associated circuits.

Periodic Resupply

Ebel and Lang* have developed a rather simple procedure for determining spare-parts requirements. The mathematical basis is the same as that for self-support. From this equation a set of curves was developed to show the number of

*G. Ebel and A. Lang, "Reliability Approach to the Spare Parts Problem", Proceedings, Ninth National Symposium on Reliability and Quality Control, (San Francisco: January, 1963).

spares to stock for a given period to back up a given number of items that have a known failure rate. These curves, illustrated in Figure 3-62, are used as follows:

(1) Determine the failure rate (λ) of each part or assembly.

(2) Determine the number of times the item is to be replaced, i.e., the number of times the item appears in a system, multiplied by the number of systems.

(3) Determine the combined failure rate for this item for the total number of times it is used in a system and the total number of systems involved.

(4) Determine the total operating time for the item; e.g., a part operating eight hours per day for one year has a total operating time of four months.

(5) Enter the left side of the chart at the appropriate cumulative failure-rate level (point a). Draw a horizontal line to the curve representing the desired operating time.

(6) At the intersection of the horizontal line and the curve (point b), drop a vertical line. The number of spares required is read at the intersection of the vertical line and the abscissa (point c).

Assume that an equipment uses two subassemblies, each of which has an expected hourly failure rate of 500×10^{-6}. The equipment is to be used 8 hours per day. How many spares are required for a period of three calendar years? The cumulative hourly failure rate is 1000×10^{-6}. This value is indicated as (a) on Figure 3-62, and a horizontal line is drawn from that point. The total operating time is 12 months. The intersection of the horizontal line and the operating-time curve for 12 months is indicated in Figure 3-62 as (b). Draw a vertical line from this point to the abscissa. The intersection with the abscissa is indicated as (c); the number of spares required is 14.

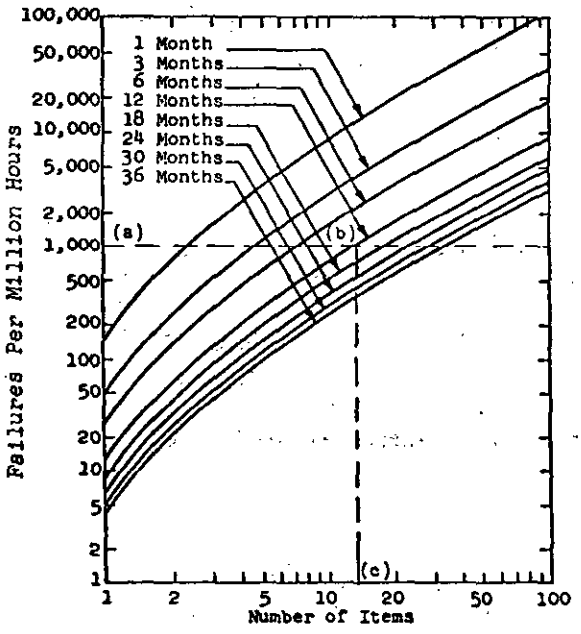


FIGURE 3-62

MINIMUM SPARE REQUIREMENTS FOR VARIOUS OPERATING PERIODS (90-PERCENT CONFIDENCE LEVEL)

The procedure can be used for parts, modules, or higher-level assemblies. Where multiple supplylevels are required (parts at a depot and modules at the equipment installation, etc.), the procedure can be used independently at each level to fulfill the requirements at that level.

The initial spare requirements for each level (maintenance echelon) can be determined through the use of the Manning and Support Model described in Section 3.10.2.1.

Recommended Support Concept

The initial provisioning of spares sufficient for the planned equipment life is economically and physically practical at this time. As indicated in the previous section, only ten to fifteen percent of the total parts (IC's) in the equipment need be provisioned initially to provide a high confidence of sufficient lifetime spares. The small size of an IC will allow easy storage of quantities in this range. As the reliability of IC's increases, this method will become more and more attractive.

The advantages to this approach are the reduction of equipment downtime due to the unavailability of a spare, reduction in paperwork by the maintenance organization, and lower overall support cost for the equipment. On a larger scale, exclusive use of this concept would lead to the elimination of many supply organizations and warehousing facilities. An additional benefit is that production facilities for items peculiar to an equipment do not have to be maintained. However, in this approach great care must be exercised in determining the support requirements so that spares are not depleted early in the equipment's life.

3.9.2.4 Location of Repair Facilities

Maintenance is performed at essentially three locations: on the equipment, in the vicinity of the equipment, or at a depot. The vicinity of the equipment includes the base to which the equipment is assigned or, in case of a fleet, another ship or a tender. A depot is a centralized repair facility capable of complete equipment overhaul; it may be operated by a civilian contractor, equipment manufacturer, or the equipment owner.

Under various maintenance concepts, any combination of these facilities may be used. In some cases, two or all three activities are combined into one facility. The requirements for these facilities are determined by the maintenance skills, test equipment, and tools required by the equipment. Where special skills or expensive equipment are required for maintenance, these tasks are best relegated to a centralized facility for maximum use of resources. However, equipment design should be such that these requirements are minimized or eliminated. With the DAFM concept, it may be possible to eliminate all but on-equipment maintenance.

It is recommended for monolithic-circuit equipment that the single facility or "self-sufficiency" concept be utilized. Under this concept, equipment is repaired by replacing a module, and the failed module in turn is either repaired

locally by the same organization or is discarded. Failed modules that are beyond the repair capability of the local maintenance organization are returned to the manufacturer for repair. This concept, along with DAFM and self-support, will allow the elimination of depot facilities and support pipelines and reduce maintenance-skill requirements. Additional advantages are maximum equipment availability and significantly reduced support costs. In cases where this approach cannot be used, the Manning and Support Model should be used to determine the optimum allocation of tasks and resources to the various facilities.

BIBLIOGRAPHY: CHAPTER 3
PACKAGING AND INTERCONNECTIONS

Air Force Avionics Laboratory. Compatible Techniques for Integrated Circuits. Motorola Report, 1965.

Balderston, J. M. Application of Integrated Circuits to Telemetry Systems. NASA TN-D-1905. 1964.

Basham, G. R., J. Y. Payton, and B. C. Sigal. "Integrated Circuit Applications In the LC-909 Computer," Solid State Design, VII (1965), 36-42.

Battelle Memorial Institute. A Study of the Effect of Space Radiation on Silicon Integrated Micro Circuits. NAS 5-3985.

Bilous, O., I. Feinbery, and J. L. Langdon. "Design of Monolithic Circuit Chips," IBM Journal, (1966).

Binder, D., and B. M. Kuehne, Surface Effects of Radiation on Microelectronic Devices. USNRDL-TRC-19. 1965.

Bogert, H. Large-Scale Integrated Circuit Array. Philco Report. 1966.

_____. Large-Scale Integrated Circuit Array. Quarterly Report. 1966.

Botos, B. "Nanosecond Pulse Handling Techniques in I/C Interconnections," Motorola Integrated Circuit, AN-270, pp. 1-7.

Bowman, W. C., and R. S. Coldwell. Transient Radiation Effects on Microelectronics. Final Report D2-9069. 1965.

Brand, D. E., "Flat Pack or TO-5?" EDN, (1966).

Brauer, J. B., and D. F . Barber. "Are Reliable Microconnections Really Attainable?" Electronic Packaging and Production, (1966), pp. 91-96.

Burger, R. M., and R. P. Donovan. "A Broader Choice of Components for Silicon Integrated Circuits," Electronics, IX (1965), 48-59.

Burks, A. W., A. H. Goldstine, and J. Von Neuman. Preliminary Discussion of the Logical Design of an Electronic Computing Instrument.

Buse, E. G. The Use of Solid Circuits in Satellite Instrumentation. NASA TN D-1758. 1964.

Christiansen, D. "Integrated Circuits in Action; Part 2, Trends and Tradeoffs," Electronics, (1966), pp. 128-135.

Computer Control Company. Interim Development Report for Evaluation of Flip-Chip Integrated Circuit Interconnections. NObsr-95052. 1966.

Cooper, D. F. "An Integrated Circuit Design for a High Speed Commercial Computer," Solid State Design, X(1965), 21-25.

Cragon, H. G. Large Scale Integrated Circuit Array. Texas Instruments, 1966.

Cunningham, J.A. "Expanded Contacts and Interconnections to Monolithic Silicon Integrated Circuits," Solid State Electronics, IX(1965), 735-745.

Davidson, U., and J. Soloman. Integrated Circuits for Field Data Computers. Motorola First Quarterly Report, 1965.

EIA Committee G-42, "Maintainability Index of Package Concepts Using Integrated Circuits," Electronic Industries Association, (1966), pp. 1-20.

Eimbinder, J. "Linear Integrated Circuits," EEE-Circuit Design Engineering, pp. 76-112.

_____. "Microwave IC's Hybrid or Monolithic," Electronics, (1966), p. 135.

Ertel, A. "Monolithic IC Techniques Produce First All-Silicon X-Band Switch," Electronics, (1967), pp.76-81.

Farmer, F. E., A. M. Hague, and H. Groot. Modules (Integrated Circuits) For Communications. ECOM-0045-F.

Freitag, H. "Design Automation For Large-Scale Integration," Wescon, 1966.

Garter, H. G. Integrated Circuit Packaging Techniques. Westinghouse Engineering, 1965.

Garth, E. C. "Ultrahigh-Speed IC's Require Shorter, Faster Interconnections," Electronics, (1966), pp. 103-110.

Garth, E. C., and I. Catt. Examining Ultra High-Speed I/C System Interconnections. Motorola Report AN-256.

Garrett, L. S. The MECL Line of Digital Integrated Circuits. Motorola Report AN-244.

_____, Milliwatt MRTL Family of Integrated Circuits. Motorola Report AN-236.

Grimmer, R.C. "Interconnecting High-Speed Integrated Circuits," Computer Design, (1967), pp. 36-45.

Goering, I. G. "The Application of Integrated Circuits in the Minuteman Missile System," Solid State Design, VII (1965), 28-35.

Goldberg, M. J., and M. W. Aarons, "Computer Methods for Integrated Circuit Design," Electro-Technology, (1965), 77-81.

Grannemann, W. W. Laboratory Testing and Theoretical Studies in Transient Radiation Effects on Microelectronics. Air Force Report. 1966.

Gregory, J.,and R. McReynolds. "The Solomon Computer," IEEE Transactions on Electronic Computers, EC-12(1963), 774-881.

Hallman, L. B. "The Impact of Integrated Electronics on Aerospace Avionic Subsystems," IEEE Transactions on Aerospace and Electronic Systems, II (1965), 152-157.

Hague, A. M. Modules (Integrated Circuits) for Communications. First Quarterly Report DA 28-043-AMC-00454(E). 1966.

Hamburger, T., and S. Gourse. Microelectronic Packaging Concepts. Final Report. 1965.

Hay, S. T. "The Packaging Revolution; Part IV, Bigger, Better Multilayer Boards," Electronics, (1965), pp. 90-95.

Herzog, G. B., A. H. Medwin, and P. Gardner. Large-Scale Integrated Circuit Arrays. RCA Report. 1966.

Hugle, William B. "Integrated Logic Circuits: A Comparative Evaluation",Computer Design, (1967), pp. 36-47.

Hurley, R. B. Transistor Logic Circuits. 1961.

Hyltin, T. "Tom Hyltin of TI and His Microwave IC's," EEE Circuit Design Engineering, pp. 74-79.

Johnson, C. W. "Assembly and Packaging of IC System," Sperry Engineering Review, Volume 19 (1966), 43-49.

Keelan, R. Choosing Very High-Speed Integrated Circuits (ECL and T2L). Motorola Report AN-265.

_____. Choosing DTL Integrated Logic Circuits. Motorola Report AN-263.

_____. Comparison of Seven Digital Integrated Circuit Logic Lines. Motorola Report AN-255.

Keelan, R. Choosing RTL Integrated Logic Circuits. Motorola Report AN-252.

_____ Using The Motorola MDTL Line of Integrated Circuits. Motorola Report AN-235.

_____. Variable Threshold Logic Family of Integrated Circuits. Motorola Report AN-200.

Keister, F. Z., R. D. Engguist, and J. H. Holley. "Interconnection Techniques for Microcircuits," IEEE Transactions on Component Parts, pp. 33-44.

Keonjian, E., and F. Danner. "Micro-Avionics: Important Considerations for Aerospace System Design," Solid State Design,1(1965).

Kinoshita, G., and C. T. Kleiner. "Radiation Induced Regeneration Through the P-N Junction Isolation in Monolithic I/C's," IEEE Transaction Nuclear Science, V(1965), 83-90.

Klein, S. F. "A Guideline for Selecting IC's," Computer Design, (1966), 52-57.

_____. "A Guideline for Selecting IC's in High Speed Digital Systems; Part 2," Computer Design (1966), pp. 52-57.

LaChapelle, T. J., and F. L. Morritz. "Thin Film Silicon-On-Sapphire Diode Matrix Memory," IEEE Wescon Technical Paper, 1965.

LaRue, J. S. Compatible Techniques for Integrated Circuits. Motorola Report. 1965.

Lathrop, J. W. Discretionary Wiring Approach to Large Scale Integration. Wescon Report. 1966.

_____. "Large Scale Integration Through Discretionary Interconnection," Texas Instruments, 1966.

_____. Large Scale Integrated Circuit Array. Texas Instruments Interim Technical Reports 1, 2, 3.

_____. Large Scale Integration Through Discretionary Interconnection. TI SC0574. 1966.

Lincoln, J., and L. M. Spandorfer. "Integrated Electronic Arrays," Sperry Engineering Review, Vol. 19 (1966), 17-24.

Logue, J. C. System Utilization. University of Pennsylvania, 1966.

Lohman, R. D., and S. T. Bosara. "Integrated Circuits for Use in the RCA Spectra 70 Series Computer," Wescon Technical Papers, 1965.

Long, D. M., H. L. Flescher, and S. E. Harrison. Radiation Effects On Insulated Gate Field Effect (MOS) Integrated Circuits. Martin, 1965.

Lowell, A. C., T. Mitsutomi, and S. A. White. A New Dimension In Microelectronic Systems. Wescon Report, 1966.

Lyman, R. C., and L. R. Collins. "How to Optimize Microelectronic Packaging," Electronic Design, (1964).

MacDonald, N. C., and T. E. Everhart. An Investigation of Electronic Beam Irradiation Effects on Metal-Oxide-Semiconductor Transistors. ERL, 1966.

Marshall, D. E., Jr. "Large Scale Integration Problems and Possibilities," Computer Design, (1967), pp. 18-21.

Martin, H. B. "Thin Film Packaging and Interconnections: The Development of A Packaging Philosophy," IEEE Transactions on Product Engineering and Production, pp. 47-50.

Martin, J. W. "What is Practical with Complex Integrated Circuits?" Electronics Engineer, (1966), pp. 118-126.

Martin, W. O., and T. Steltzer. "Physical Design of an Integrated Circuit, Airborne Digital Computer," Wescon Technical Papers, 1965.

McCormick, John E. "Multilayer Reliability", NEP Conference. Long Beach California, 1967.

McIver, G. W., and J. C. Bvie., Advanced Techniques for Integrated Circuit Distributed Array. Air Force Avionics Laboratory, 1966.

Meindle, J. D., and P. H. Hudson. "Low Power Linear Circuits," IEEE Journal of Solid State Circuits, (1966), pp. 100-111.

Microwave Associates Incorporated. Research on Integrated Microwave Circuits. 1965.

Mitchell, E. P., and D. Pope. "Which IC Logic Form?" Electrical Design News, 4 September 1966.

Murphy, B. T. "High Speed Integrated Circuits with Load-Compensated Diode-Transistor Logic," Electronics 11(1963), pp. 68-74.

National Aeronautics and Space Administration. An Examination of the Applicability of Micro-Electronic Circuits to the Telemetry & Command Subsystems of Several Applications of Spacecraft. NASA Report CR-233.

National Aeronautics and Space Administration. Survey/Study of the Interconnection Problem in Microelectronics. NASA CR-414. 1966.

Nevala, R. D., and E. R. DeAltey. "Digital Integrated Circuits and Their Limiting Factors," Electronics Industry, IX(1965), 70-74.

Perkins, C. W., R. W. Marshall, and A. M. Liebschultz. Radiation Effects on (Monolithic) Microelectronic Circuits. Report 01313F. Fort Monmouth, J. J., 1966.

Nixen, D. "Materials for Packaging Microelectronic Devices," Electronics Industry, (1965), pp. 66-69.

Phelps, M. B., Jr. "Choice of Logic Forms for Integrated Circuits," EDN, (1964).

Phipps, C. H. "Trends in Semiconductor Circuit Packaging," IEEE Wescon Technical Packaging," IEEE Wescon Technical Paper, 1965.

Plemenos, F. A. "The Packaging Revolution; Part VI, Converting to Microelectronics," Electronics, (1966), pp. 103-120.

Potter, G. B., and J. Mendelson. "Integrated Scratch Pads Sire New Generation of Computers," Electronics, (1966).

Raisanen, W. R. "Current Mode Logic and Its Application to Large Scale Integration" NEREM Record, 1966.

Rajchman, J. A. Memories In Present and Future Generations of Computers. IEEE Report, 1965.

Rhoades, W. T. "Logic IC's Don't Live Long,"Electronics, March 6, 1967.

Rice, R. "The Packaging Revolution; Part V, Simpler Designs for Complex Systems," Electronics, III(1966), 109-116.

Rigling, W. S. "Designing and Making Multilayer Printed Circuits," Electro-Technology, (1966), pp. 54-57.

Robertson, J. J., and R. A. Hirscheld. "Application of Micro-Electronics to IF Amplifiers," IEEE Trans. Vehicular Comm., 1(1965), 162-172.

Rome Air Development Center. Final Report; Study of Transient Radiation Effects on Microelectronics.

Sanford, Robert. "Understanding IC Logic", Electronics, March 6, 1967.

Sarine, C. D. "Microelectronics in the System IC Applications," Sperry Engineering Review, Vol. 19 (1966), 11-16.

Schulz, R. B., and R. L. Clapsaddle. "Environmental Effects of RFI on Micro-electronics," Electro-Technology, 1-5.

Schulz, R. B., and R. L. Clapsaddle. "Environmental Effects of RFI on Microelectronics," Electro-Technology, IV(1966), 46-50.

Shafer, O. B. "Reducing Wiring Lengths," Electro-Technology, (1962), 92-167.

Shannon, M., and D. L. Cannon. The Microelectronics Fabrication of VHF Systems. 1965.

Signetics Corporation. Final Development Report; Economical Flat Package for Integrated Circuits. 1965.

Slemmons, J. W., and H. R. Howell. "Interconnection Methods for Microcircuits," Electronic Designs, (1964), pp. 78-83.

Smith, M. G. Systems Considerations For Large-Scale Integration. IBM Watson Research Center Report, 1966.

Smith, M. G. "Systems Considerations for Large Scale Integrations," Wescon, 1966.

Spandorfer, L. M., and J. E. Schwartz. "Microelectronic Logic Circuits," Solid State Design, (1964), p. 50.

Staller, J. J., and G. Sideris. "The Packaging Revolution; Parts II and III, Design and Manufacturing Overlap," Electronics, (1965), pp. 75-89.

Staller, J. J. "The Packaging Revolution; Part I, Form and Function Interact," Electronics, 21(1965), 72-87.

Staller, J. J.,and H. A. Ullman. Guidelines for Implementation of System Requirements into Microelectronic Mechanical Designs. Sylvania Electronic Systems Division, Needham Heights, Mass., 1966.

Strahan, V. H. Radiation Effects on Thin-Film Microelectronics. Autonetics Report CR-2017/3111, 1965.

Strull, G., M. N. Giulizno, R. C. Gallagher. "A Systems Approach to a Compatible Family of Linear Integrated Circuits," IEEE Journal of Solid State Circuits, (1966), 118-122.

Suran, J. J., and R. A. Marolf. "Integrated Circuits and Integrated Systems," IEEE Proceedings, 12(1964), 1661-1668.

Triebwasser, S. Programable Interconnection Techniques. ISSCC Report, 1966.

Uhler, A. "Microwave Applications of Integrated-Circuit Techniques," IEEE Proceedings, 12(1964), 1617-1623.

Walker, G. P. "Semiconductor Packaging for High Component Density Applications," pp. 125-128.

Walker, R. M. "System Design and Packaging Using Integrated Circuit," EEE Circuit Design Engineering, pp. 79-83.

Wang, P. Development of Epitaxial Technology for Microelectronics and Large Area Device Application. Sylvania Report, 1966.

Welty, J. M. "Microelectronic Packaging as Applied to MEMA," Proceedings of National Electronics Conference, 1966.

Widly, G. J. Field Experience with a Microelectronic Multimode Radar. AGARD & NATO Report, 1966.

Wrieden, Eugene G. Criteria for Discard-at-Failure Maintenance. IBM Report. Owego, N. Y., 1963.

TESTING

F O U R

4.1 INTRODUCTION

Quality-assurance tests of the type performed on transistors and diodes have frequently been applied to integrated circuits in the past. However, MIL-STD-883, Test Methods and Procedures for Microelectronics, is now available for environmental mechanical and electrical tests of microcircuits. Figure 4-1 shows the sequence of events and the organizations involved in a typical test program. The extent to which the tests are employed depends on the criticality of the device application. Devices intended for noncritical use (industrial equipment, radio, television) usually receive minimum testing before being installed. The manufacturer might, for example, perform only a partial electrical test, and the customer's incoming inspection might consist of only a part count and a visual examination for damage.

However, a device earmarked for a high-reliability system is generally subjected to extensive testing. The manufacturer is likely to perform mechanical, electrical, and hermetic-seal tests, along with a reliability study to verify the failure rate and screening tests to detect potential failures. The customer's incoming inspection includes both visual and electrical checks. In many cases the customer subjects a sample of the devices to tests identical to those performed by the manufacturer. The manufacturer's and customer's data are then compared to verify device reliability.

As indicated in Figure 4-1, defective devices may be sent to a failure-analysis group for detailed analysis. Data obtained from such analyses are usually communicated internally to the components group for use in specifying future devices and evaluating specifications. These data may also be transmitted to the device

manufacturer. The function and procedures of the failure-analysis group are discussed in detail in Chapter 6. Figure 4-2 indicates some of the problems associated with an integrated-circuit test program. For incoming inspection, new test fixtures and procedures are needed for handling the IC packages. Additionally, since IC's cannot always be tested on a go/no-go basis, many tests are quite complex; thus test personnel must be retrained.

Section 4.2 deals with the tests that may be performed by the device manufacturer, the user, or both. Factors such as purpose of test, effectiveness of the test to detect various failure modes or mechanisms, and relative cost are described for most of the tests performed on IC devices. Tests that may also be used in a screening program are tabulated and compared in Section 4.3. Section 4.3 also contains a discussion of screening approaches. Sections 4.4 through 4.6 cover tests, equipments, and fixtures required for the evaluation of integrated circuits.

4.2 QUALITY-ASSURANCE TESTS

4.2.1 Introduction

Quality-assurance testing of integrated circuits is becoming the most important aspect of IC procurement, principally because it is difficult to verify the failure rates of IC devices. The failure rate is a quantitative indication of the probability of survival; this value is difficult to obtain for highly reliable devices because of the large number of test hours required to verify a low failure rate. Such rates are thus usually not available. If they are, their validity is questioned by the user unless he has verified the tests.

To overcome this uncertainty, the user who requires highly reliable devices has turned increasingly to intensive quality-assurance testing. Typically, quality is assured by a combination of screening (100-percent testing) and acceptance (sampling) tests -- including visual inspection and physical, environmental, electrical, and life tests. The type and extent of the test performed depend on the reliability requirements of the application.

Because the application determines the assurance-test procedure, it is not possible to define all the tests as either screening tests or acceptance tests. At one extreme, it is possible to include all the tests described in this section, except those that are destructive, in a screening procedure. At the other extreme a screening procedure or sampling test might include only visual inspection and electrical testing.

Incoming inspection and sampling tests should identify devices that are actually defective or determine the probability that more than a specified number are defective in a specific lot. Screening should serve an additional and much more important function -- the elimination of incipient failures, those defects that are manifested when the equipment becomes operational.

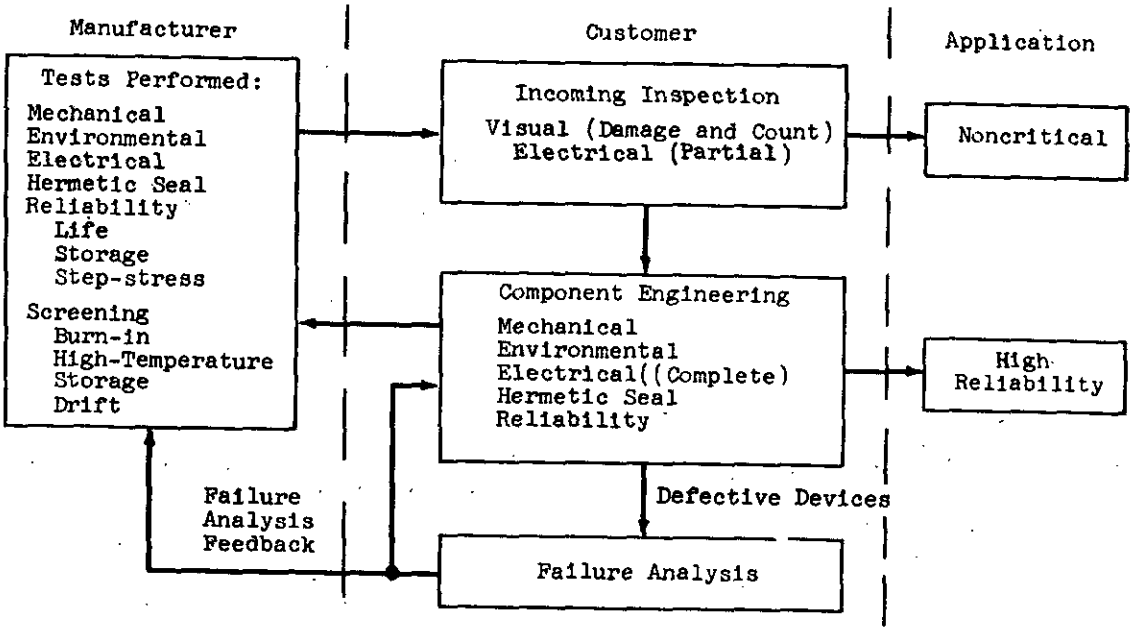


FIGURE 4-1

SEQUENCE OF EVENTS IN TYPICAL TEST PROGRAM

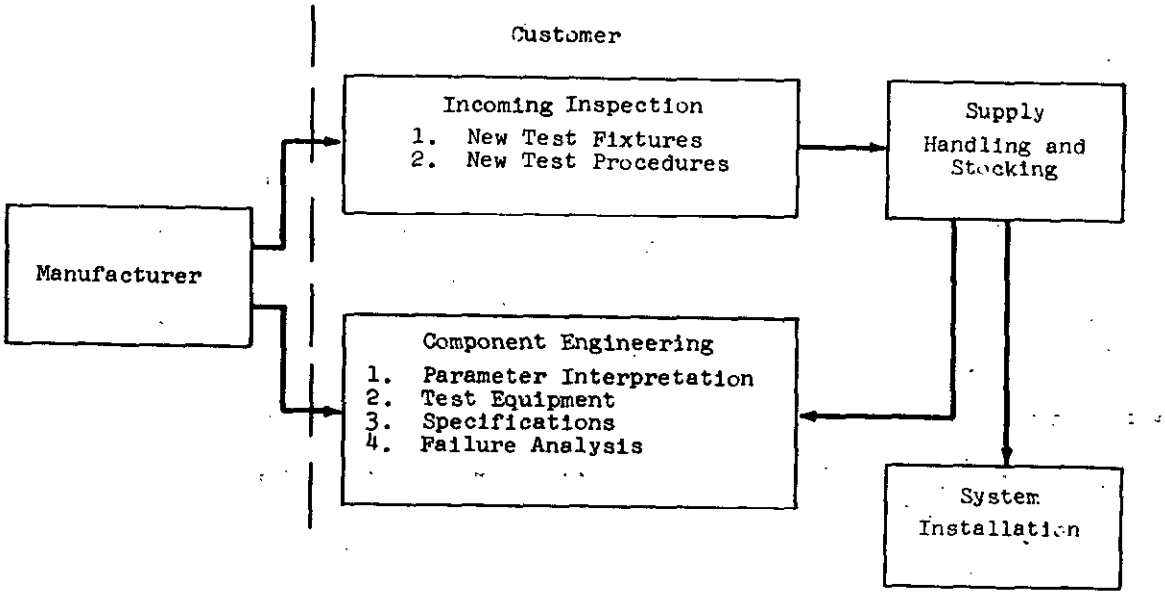


FIGURE 4-2

PROBLEMS ENCOUNTERED IN TEST PROGRAMS

While it may be possible to standardize certain tests, it is not possible to develop a universal testing procedure. Some applications require higher reliability than others under varying stress levels. It is the responsibility of the user to develop the most effective test procedure that satisfies the requirements for a specific application. From this viewpoint, the most effective test procedure is the one that meets the application requirements in the shortest time at the lowest cost.

Sections 4.2.2 through 4.2.5 are descriptions of the tests commonly performed on IC's.

4.2.2 Wafer Testing

For a highly reliable device, it is necessary to start with high-quality silicon material. The material must then be processed with excellent workmanship, control, and inspection.

Wafer tests are performed by the device manufacturer, principally for process control. Some of the tests are performed repeatedly during wafer processing, including dicing.

4.2.2.1 Wafer Inspection

The wafer is inspected at various stages of its processing to detect faults and measure critical physical characteristics. Common faults include the following:

Diffusion faults
Wafer crystal imperfections
Oxide faults
Physical damage
Photolithographic defects
Pinholes

Critical physical characteristics include oxide thickness, resistivity, and resistivity uniformity.

Microscopic visual inspection can detect oxide faults, physical damage, and photolithographic defects. Oxide thickness is measured by interferometer techniques. Electrical probing is used to detect pinholes, diffusion faults, and wafer crystal imperfections.

A four-point probe can be used to determine resistivity and resisitivity uniformity. Electrical probing can be performed on the wafer components or on a test evaluation group (TEG) specifically fabricated on the wafer for evaluation purposes. Resistivity measurements are performed on the wafer. Required tests are conducted before and after each of the appropriate processing steps.

4.2.2.2 Epitaxial Inspection

Defects in expitaxial silicon layers may be on the surface or they may be within the epitaxial layer. Typical of the types of defects observed are pinholes, stacking faults, and pyramids. These may be caused by contamination or substrate crystal defects. Tests are performed after expitaxial growth; and the instruments used include high-power microscopes for detecting etched defects and topological defects, phase-contrast microscopes for detecting stacking faults, and four-point probes for measuring resistivity.

4.2.2.3 High-Voltage Isolation Test

The high-voltage isolation test is used to detect incomplete isolation and diffusion faults. An electrical probe can be used to measure the breakdown voltage at any process step desired.

4.2.2.4 D-C Electrical Tests

Component parameters are measured after various diffusion steps; the complete circuit is measured after metalization. The major reason for measuring component values, primarily those of transistors, is that it is often necessary to rediffuse the wafer to optimize parameters after the components have been fabricated. Often, to increase the wafer yield, the emitter will be rediffused to optimize the device characteristics for the specification to which the devices are being built. Although all previous diffusions are affected at the time the emitter is rediffused, the emitter area contains the highest impurity concentration. The tests performed on transistors are usually measurements of the transistor β and V_{CEO}; they are made with a microscopic electrical probe.

The device manufacturer performs tests during processing and before die mounting for reasons of economics. The cost of a die is nominal until it is mounted on the header. The cost of the finished device is reduced by eliminating at this stage devices that will not pass the final inspection.

4.2.2.5 A-C Electrical Tests

A-c electrical testing is usually not performed on the wafer. If it is, an electrical probe is used. Parasitics can be a problem for fast pulses and high frequencies. The purpose of such tests is essentially to eliminate devices that will not pass final dynamic testing, at which point the cost of throwing away the device is high. There may be some reliability advantage inherent in a-c wafer testing, but it is not agreed upon within the industry.

4.2.3 Dice Visual Inspection

After the wafer is fabricated, it is diced; each die is visually inspected to detect oxide defects, metalization defects, cracks or chipping of the die, and misalignment defects. This is usually performed at 7× to 20× magnification.

4.2.4 Pre Cap Tests

4.2.4.1 Wire Pull Test

The elements of the wire pull test are as follows:

Failure Mechanisms. Defective wire bonds due to poor workmanship or machine control.

Deleterious Effect. Destructive (performed on a sample basis).

Position of Test in Testing Sequence. This test is performed by the bonding-machine operator at the beginning of the work shift.

Relative Cost. The cost of performing this test is very low although all devices tested must, of course, be discarded.

Effectiveness. This test is simple to perform, and from the standpoint of cost and time versus reliability, it is an excellent indication of wire-bond integrity.

Example of Detailed Stress Levels. Gold wires bonded at both ends should exhibit strength in excess of 5 grams; aluminum wires bonded at both ends should exhibit strength in excess of 2 grams. These values are read on the gauge and are not the actual force on the bond, which is about 8 times greater for a 100-mil wire length.

Discussion. An example of such a test for poor post and chip bonds is the use of a tension gauge to pull bonded wires. The gauge may be either automatic or manual. All the bonds on three devices from each bonding machine being operated should be pulled at the start of each shift to ensure that bonding is under control. The tension reading should be greater than some predetermined value.

4.2.4.2 Infrared Test

The elements of the infrared test are as follows:

Failure Mechanisms. Excessive localized heating due to voids between the substrate and header or poor thermal design.

Deleterious Effect. None.

Position of Test in Testing Sequence. Currently limited to use in design evaluation.

Relative Cost. Expensive.

Effectiveness. This type of test is a good engineering design tool, but its slowness makes it inefficient as a process inspection tool.

Discussion. The use of infrared scanning is currently limited to the design phase. Present equipment will not permit 100-percent in-process testing. The industry is not sure that such screening is even required if the circuit has been well designed.

4.2.4.3 Visual Inspection

The elements of visual inspection tests are as follows:

Failure Mechanisms

Lead dress	Die Bond
Particles	Wire Bond
Corrosion	Contamination
Metalization defects	Oxide Defects
Chip mounting and chipped or cracked die	

MIL-STD. 750A, Method 2071.

Deleterious Effect. None.

Position of Test in Testing Sequence. Immediately prior to encapsulation.

Relative Cost. Low to moderate.

Effectiveness. When this test is performed by a qualified inspector using appropriate microscopes, it is highly productive; it is essential in high-reliability programs.

Discussion. The equipment required to perform this inspection is any magnifying apparatus of 50x (or greater) with a collimated light source. It should include a manipulator that provides circuit orientation and handling facilities for visual ease in inspecting.

Lead dress presents a serious reliability problem, primarily because intermittent shorts due to poor lead-dress quality often pass many or all of the required device tests. The use of X-ray screens to illuminate poor lead dress is discussed in Section 4.2.5.13. However, X-ray is ineffective when aluminum leads are used. Defects and poor process control or workmanship should be looked for in the following categories:

Process Uniformity. A topological representation should be provided for inspection, and all devices should conform to this topology. All devices of the same part number in each lot should be identical in appearance. There should be no inconsistencies in topology orientation, bond patterns and placement, etc., that cause a heterogeneous appearance.

Cleanliness. The active device should be free from any chemical residues (including lacquer, varnish, or jelly) or discoloration resulting from device production and handling. Both the device and package should also be free of foreign particles greater than 0.5 mil (0.0005 inch) across the widest dimension. This includes extraneous encapsulated material, weld splatters, and excessive build-up or flaking of gold preform.

Bonds. Metalized bonding pads should be provided for each terminal where a bond is to be made on the substrate. Any bond contact area made on the metalized intraconnection is cause for rejection of the device.

The entire bond, as defined by the bonding-tool impression, should be within the periphery of the bonding pad, and there should be at least 2 mils (0.002 inch) of interconnection material around the periphery of the bonding-tool impression.

There should be no evidence of loose, misplaced, or open bonds, or bonds that have been partially stripped from the pad. There should be no evidence of multiple attempts to bond on any single bonding pad or lead. Such evidence is cause for rejection of the device.

Internal Bonding Wires. Internal bonding wires should be rejected for the following:

Wires of sufficient length (slack) to cause shorting to another lead, die edge or surface, or to the package sides, bottom, or top.

Nicks, cuts, crimps, or scoring that cut into or deform the wire by more than 25 percent of the original diameter.

When viewed from above, leads that cross one another or cross any metalization not electrically connected to the lead (complex circuits may require modification of this condition).

Lead material greater in length than 2 mils (0.002 inches) that is fixed on only one end (pigtails).

Extra wires, i.e., wires other than the ones connecting specified areas on the chip to the external leads. Only one wire should be connected to a specified area of the chip or to an external lead, except where the design of the integrated circuit calls for the use of additional wires and has been previously approved.

Any internal bonding wire that is missing from its intended location.

Inadequate clearance. No wire should be within 3 mils (0.003 inches) of another wire, ball bond, or the case.

Substrate Defects. Any substrate that exhibits cracking, fracture, pitting, chipping, or other signs of physical damage in the active circuit, metalization, or bond areas greater than 1 mil (0.001 inch) should be rejected.

Diffusion Masking. Devices with diffusion irregularities and masking defects should be rejected if oxide is not visible between active junctions or isolation junctions.

Metallic Interconnection. The metallic interconnections should be deposited in accordance with the layout design. No two interconnections should be closer than 0.5 mil (0.0005 inch) or 50 percent of the distance detailed in the layout design, whichever is smaller.

Metalization Voids. Voids in the metallization should not reduce the width of any lead, pad, or fillet to less than 50 percent of design width or below 0.4 mils (0.0004 inches) whichever is smaller.

Tool Marks and Scratches. The following defects are cause for rejection:

Scratches or tool marks that reduce the width of any metalization to less than 50 percent of design width or expose silicon dioxide along the scratch.

Any smear of metalization that extends contiguously more than one lead width from the design lead path, or that reduces the spacing between adjacent leads to less than 50 percent of the design lead spacing.

Metalization Defects (Others). Any one of the following metalization defects should also be considered cause for rejection:

Excessive peeling or bubbles

Corrosion (chemical reaction)

Metalization with less than two-thirds coverage of electrically active contact areas

Oxide Damage. Oxide voids exposing an electrically active junction area or exposing a silicon surface to an electrically active lead that is not, by design, already in contact with the same surface should be cause for rejection.

Periphery. The periphery of the circuit dice should be well defined and encompass the entire active circuit. No active area of the circuit, bonds, bonding pads, or junctions (including under or side diffusion) should be closer than 0.1 mil (0.0001 inch) to the dice periphery.

Package and External Leads. Any one of the following physical defects in the package and leads should be cause for rejection:

Physical damage to case or leads

Cracks or voids in the glass-to-metal or ceramic-to-metal seal, greater than one lead thickness

Bar (chip) tilted or loose in the header

External terminals (leads) shorted to the case

Contamination or conducting particles on external leads

Other mechanical faults

The above examples of inspection criteria are generally used but are not necessarily approved or recommended by NASA/DOD.

4.2.5 Device Tests

4.2.5.1 High-Temperature Storage (Non-operating)

The elements of high-temperature-storage tests are as follows:

Failure Mechanisms

Electrical Parameter Drift

Corrosion

Bulk Silicon Defects

Metalization Defects

MIL STD. 750A, Method 1031.3.

Deleterious Effect. None.

Position of Test in Testing Sequence. Usually performed by the device manufacturer after encapsulation and before electrical tests. Electrical tests may be performed during the bake. If performed by the user, it is performed before physical and thermal environmental tests.

Relative Cost. Very inexpensive.

Effectiveness. The enhancement of device stability makes this test desirable. It will not diminish reliability and may improve it provided the temperature limit is selected with a knowledge of the possible effects on the metalization.

Example of Detailed Stress Levels. Typically, this test requires storage in a prescribed high-temperature ambient (usually 150°C) for up to 1000 hours.

Discussion. This test is sometimes referred to as "stabilization bake". It is intended to accelerate parameter drift with time and temperature. Electrical testing may be performed on a screening or sampling basis at various intervals during the test. Failures are detected during these periodic electrical tests. When electrical tests are not performed during the bake, the defective devices are detected during regular electrical tests.

Careful consideration should be given to the upper temperature limit. Temperatures above 175°C may create long-term failure mechanisms such as plague.

4.2.5.2 Post-Cap Visual Inspection

The elements of post-cap-visual inspection are as follows:

Failure Mechanisms

Lead contamination

Improper sealing

Cracked package

Poor lead plating

Incorrect marking

MIL STD. 750A, Method 2071.

Deleterious Effect. None.

Position of Test in Testing Sequence. After high-temperature storage and as part of incoming inspection.

Relative Cost. Inexpensive

Effectiveness. While this test, as well as other operator-dependent tests, is subject to operator error, it is a necessary test. This inspection can eliminate defects that will usually not be detected by any other test.

Discussion. The device should be inspected by a trained operator using a 20× (or greater) microscope.

4.2.5.3 Temperature Cycling

The elements of temperature cycling tests are as follows:

Failure Mechanisms

Package and seal defects

Metalization and lead defects

Thermal mismatch between various materials used in the IC

Cracked substrate

MIL STD. 202C, Method 107.

Deleterious Effect. None.

Position of Test in Testing Sequence. This test is performed after high temperature storage. To derive the maximum benefit from the test it should be performed before a hermetic seal test. The hermetic seal test will detect package defects induced by the stresses associated with this test.

Relative Cost. Very inexpensive (see discussion).

Effectiveness. Good (see discussion)

Example of Detailed Stress Levels. The IC device should be subjected to ten temperature cycles from -65°C ±3°C to +200°C ±3°C with a minimum of 15 minutes at each temperature extreme and a maximum of two minutes at room temperature between extremes.

Discussion. This is an essential test. It is inexpensive and effective, and comes close to duplicating the thermal stresses of space (a rotating satellite is subjected to such a stress).

Temperature cycling may be one of the few tests that are effective for all-aluminum systems. More cycles are usually required for Al-Al systems than for Au-Al systems, and this results in a slightly higher cost. There is some evidence that the stress levels (above a particular threshold) are not as important in inducing failures as the number of cycles.

Defects induced during this test are detected during subsequent failure-detection tests such as X-ray, leak tests, and electrical tests.

4.2.5.4 Thermal Shock

The elements of thermal-shock testing are as follows:

Failure Mechanisms

Package and seal defects

Bonding and metalization weaknesses

Thermal mismatch

Cracked substrate

MIL STD. 750A, Method 1056.1

Deleterious Effect. See Discussion.

Position of Test in Testing Sequence. This test is performed after other temperature tests and before mechanical tests such as acceleration and shock. Also see Section 4.2.5.3.

Relative Cost. Inexpensive.

Effectiveness. Probably not significantly better than temperature cycling.

Example of Detailed Stress Levels. The IC is immersed in a liquid bath at +125°C for not less than five minutes, transferred immediately to a liquid bath at -65°C, and left immersed for not less than five minutes. This is repeated for a total of five cycles. The volume of the liquid in each bath should be large enough so that the temperature of the bath does not change by more than 10°C when the samples are immersed. The samples must stay in the bath until the liquid has returned to its specified temperature.

Discussion. Defects induced during this test are detected during one of the subsequent failure-detection tests such as X-ray, leak, or electrical tests.

Although this test is similar to temperature cycling, there is a significant difference that may affect its usefulness: The severe strain generated by the extreme temperature gradient is not indicative of any actual operating conditions.

While this test is inexpensive and may be effective for sampling, it is not recommended for screening. One reason is that it may severely penalize a package fabricated with glass or ceramic; another is the possibility that it will weaken the surviving devices. Temperature cycling assesses essentially the same failure modes and mechanisms as thermal shock and is preferred for screening.

4.2.5.5 Constant Acceleration

The elements of constant acceleration tests are as follows:

Failure Mechanisms

Loosening of bonds or die

Lead-to-case shorts

Cracked substrates

MIL STD. 750A, Method 2006.

Deleterious Effect. None.

Position of Test in Testing Sequence. This test is performed after thermal tests. This sequence is desirable since mismatches in thermal-expansion coefficients will weaken marginal-device structures, causing failures under mechanical stress.

Relative Cost. Moderate.

Effectiveness. At a stress level of 20 to 30 thousand G it is effective for gold lead wires. At this stress level, its effectiveness for aluminum leads is questionable and higher levels should be used.

Example of Detailed Stress Levels. The IC device is held by its case with suitable protection for the leads. The device is subjected to a centrifugal acceleration of 20,000 G's in at least two orientations. The run-up and run-down times should not be less than 20 seconds each.

Discussion. Centrifuge yields the highest available acceleration and therefore permits the greatest probability of screening out certain incipient failures. Acceleration in the Y_1 axis stresses weak areas since it tends to lift marginal bonds off the bonding pads and separate a poorly adhering die from the header. Acceleration in the Y_2 axis tends to compress bonding wires against their attachment points and stress unsupported substrates. Further, Y_2 accelerative forces tend to produce a lateral force that will slide the bond in a direction tangential to the pad.

The tensional force induced in a lead depends on many parameters, expecially the density of the material. The tension in grams induced in the leads for 2-mil gold leads at various G levels is shown in Figure 4-3. Similar information for 1-mil aluminum leads is illustrated in Figure 4-4. In each case the values are determined for lead lengths of 100 mils. If

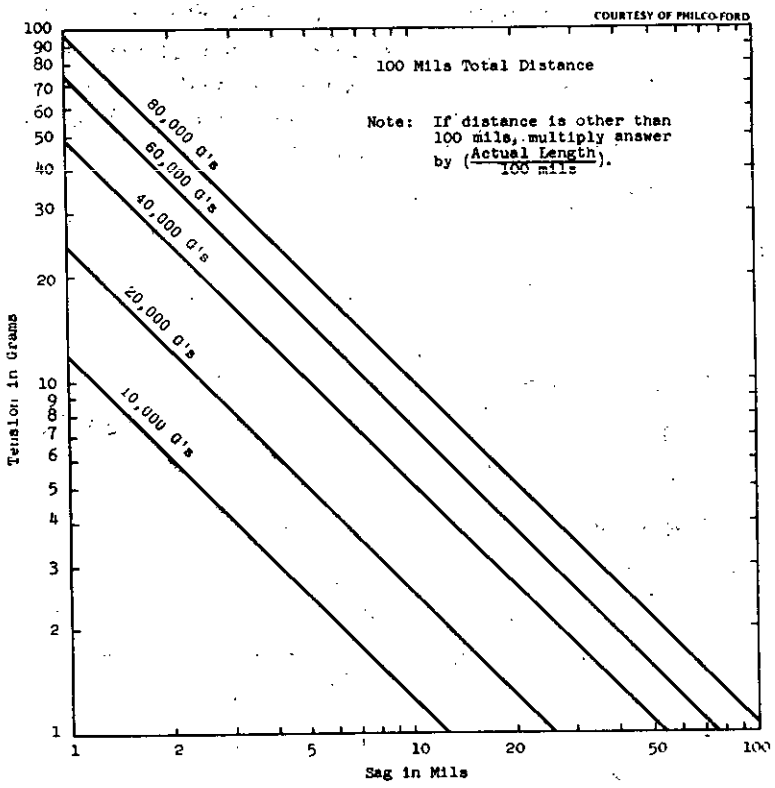


FIGURE 4-3
TENSIONAL FORCES PRODUCED ON 2-MIL GOLD WHISKER WIRES BY A CENTRIFUGE

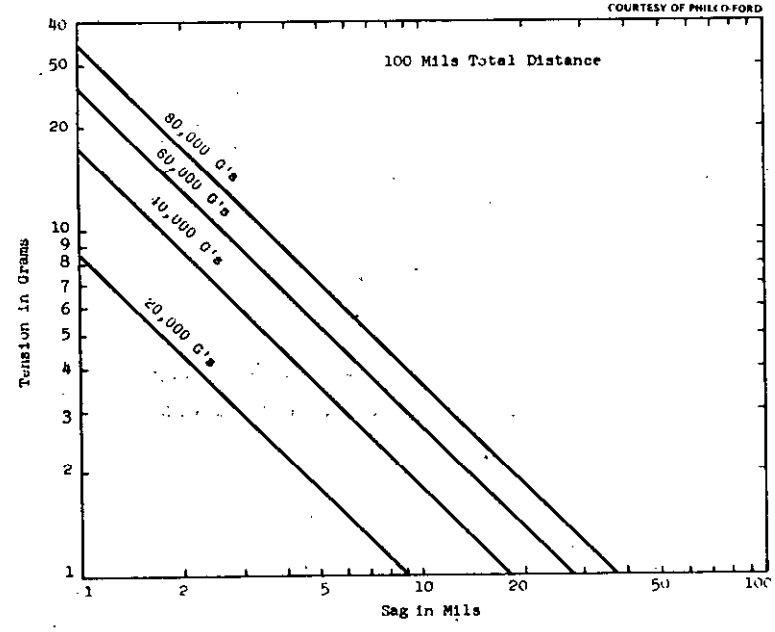


FIGURE 4-4
TENSIONAL FORCES PRODUCED ON 1-MIL ALUMINUM WHISKER WIRES BY CENTRIFUGE

the G force and sag are held constant, the tension for 20,000 G's on Au leads is quite different from that for Al leads. For 20,000 G's and a sag of 1, the tensional forces in grams are 12 for gold and 0.1 for aluminum.

A 20,000-G stress is standard in the industry. It is not known why this value is popular although it is known that a good device using Au leads can withstand this level, and many defective bonds have been found at this level.

Additional questions arise about the 20,000-G level now that aluminum leads are being used extensively in IC's. From the example above, it is obvious that aluminum is not stressed nearly as much as gold at 20,000 G's.

When conducted in the proper fixtures and at the proper level, this test is ineffective for checking internal contacts (especially bonds) at a moderate cost. The experience of one manufacturer indicates that a high "G" pneupacter* shock test is just as effective and is more economical. However, this shock test may result in permanent damage to either the case or the device.

4.2.5.6 Shock (Unmonitored)

The elements of shock (unmonitored) tests are as follows:

Failure Mechanism

Lead Dress	Wire Bonds
Package Defects	Cracked substrate
Die Bonds	

MIL STD. 750A, Method 2016.2

Deleterious Effect. Possible damage to case or circuit.

Position of Test in Testing Sequence. This test is performed after the constant acceleration test or in place of it.

Relative Cost. Moderate.

Effectiveness. Because of the lower stress levels free-fall shock tests are less effective than constant acceleration. The pneupactor shock test induces higher stress levels and is more effective than free fall, and it may be as effective as constant acceleration.

Example of Detailed Stress Levels. The drop-shock test typically provides a stress level of 1500 to 3000 G's for 0.5 millisecond, 25 blows in each of 4 directions: X_1, Y_1, Y_2, Z_1 minimum -- total of 100 blows.

* The pneupacter is an air-actuated accelerator that fires the IC under test (projectile) down a barrel so that it will impact on a target. An accelerometer is used to determine the impact forces.

Discussion. Drop-shock testing is basically a median test, lying between centrifuge and vibration. A shock test possesses both a force-frequency spectrum and a high acceleration. However, its acceleration cannot be achieved on a level comparable to that of a centrifuge, nor is the frequency spectrum as comprehensive as that of vibration testing. Thus, drop-shock testing is merely a compromise containing weaker versions of the salient features of both centrifuge and vibration tests.

Shock can to some degree replace aspects of both centrifuge and vibration tests. However, for a high-reliability program both vibration and centrifuge tests should be used.

The pneupactor shock test is currently being evaluated for shock testing. G levels as high as 50,000 are possible with the air-actuated pneupactor. This test, if feasible, may rival the constant-acceleration test in effectiveness. It is relatively inexpensive.

If package damage can be eliminated and the test conditions do not weaken the device, there may be an economic advantage in using the pneufacter shock test as a screen test in place of centrifuge screen testing. In any event, it is unlikely that both shock and accelerator screens are both required since they induce similar stresses.

Defects induced in this test are detected during subsequent failure-detection tests such as X-ray, electrical, or leak tests.

4.2.5.7 Shock (Monitored)

The elements of shock (monitored) tests are as follows:

Failure Mechanism

Lead Dress	Wire Bonds
Package Defects	Cracked substrate
Die Bonds	Intermittant lead opens
Conducting Particles	

MIL STD. 750A, Method 2016.2.

Deleterious Effect. See Section 4.2.5.6.

Position of Test in Testing Sequence. See Section 4.2.5.6.

Relative Cost. Expensive.

Effectiveness. Poor.

Example of Detailed Stress Levels. See Section 4.2.5.6.

Discussion. The purpose of this test is essentially the same as that discussed in Section 4.2.5.6. However, since it also includes the monitoring of an electrical parameter of the test device to detect intermittent conditions, the scope of the test is extended into particle detection.

Its capability to detect conducting particles is small, and the added expense and difficulty of performing it make it a poor test.

4.2.5.8 Vibration Fatigue

Failure Mechanism

Lead Dress	Wire Bond
Package Defects	Cracked Substrate
Die Bond	Hardening of materials

MIL STD. 750A, Method 2046.1.

Deleterious Effect. Destructive (use on sample basis only).

Position of Test in Testing Sequence. This test is usually performed after shock testing.

Relative Cost. Expensive.

Effectiveness. This test is ineffective for mechanical defects typically found in integrated circuits.

Example of Detailed Stress Levels. The IC device and its leads are rigidly fastened to the vibration platform so that there is a one-to-one transference of motion. The device is then subjected to a simple harmonic motion at 60 Hz with a constant peak acceleration of 30 g's for 32 hours in each of three orientations (X_1 or X_2, or Y_2 and Z_1, or Z_2) for a total of 96 hours.

Discussion. Induced failures are detected during subsequent testing.

4.2.5.9 Vibration Variable Frequency (Unmonitored)

The elements of vibration-variable-frequency (unmonitored) tests are as follows:

Failure Mechanisms

Package defects

Die bond

Wire bonds

Cracked substrate

MIL STD. 750A, Method 2056.

Deleterious Effect. None.

Position of Test in Testing Sequence. This test is conducted after acceleration or shock if these tests are performed; otherwise, after the last thermal test.

Relative Cost. Expensive.

Effectiveness. Fair.

Example of Detailed Stress Levels. 30 G's, through frequency range of 10 to 2000 Hz and back to 10 (logarithmic sweep) for 4 minutes minimum each orientation: X_1, Y_1, Z_1. Total of 12 times for 48 minutes.

Discussion. Sinusoidal vibrators are relatively inexpensive but less effective (because of the restriction in frequency range) than the very expensive random vibrators. The VVF (unmonitored) is preferred over vibration fatigue and is an important test if the devices are to be used in space applications. The environmental conditions simulated in this test are similar to those of space launch.

The test is usually performed on more than a single axis, as indicated above. However, more than one axis of vibration does not seem to contribute significantly to test effectiveness, if the axis is chosen properly. Cost increases significantly as the number of axes is increased. Induced failures are detected during subsequent failure-detection screens.

4.2.5.10 Vibration Variable Frequency (Monitored)

The elements of vibration-variable-frequency (monitored) tests are as follows:

Failure Mechanisms

Package defects	Lead Dress
Die bond	Intermittent lead opens
Wire bonds	Cracked substrate
Conducting particles	

MIL STD. 750A, Method 2057.1.

Deleterious Effect. None.

Position of Test in Testing Sequence. This test is conducted after acceleration or shock if these tests are performed; otherwise, after the last thermal test.

Relative Cost. Very expensive.

Example of Detailed Stress Levels. 30 G's, through frequency range of 10 to 2000 Hz and back to 10 (logarithmic sweep) for 4 minutes minimum each orientation: X_1, Y_1, Z_1. Total of 12 times for 48 minutes. In addition, continuous electrical measurements are made.

Discussion. Monitored vibration reveals that excessive flexibility results from overlong leads or unsupported substrates, evidencing itself as a modulation in the output signal. This assumes there was no end-point testing after acceleration. The variable frequency vibration causes loose particles or bonding wires to move within the package. Such movement is evidenced by an electrical pulse in the monitored output signal.

Output voltages should be monitored, in accordance with circuit topology and the schematic diagram, with the measuring circuit delineated in the detail specification. All inputs and outputs should be connected to power sources through current-limiting resistors. In this way terminal voltages will be free to change as a function of internal shorts or opens. A noise output of 100-mV peak-to-peak or greater constitutes a failure. Rejection at this level will usually preclude triggering of the next in-line device by random vibration noise. It is essential that a true peak-to-peak reading instrument be provided since only one pulse may be sufficient to overcome the noise margin of succeeding devices.

The frequency spectrum of 10 to 2000 Hz is a good range for VVF because it includes rocket noises, which have most of their energy at the low end of this frequency spectrum, while the internal resonances of integrated circuits have most of their energy at the high end of this spectrum.

There is no agreement within the industry about the effectiveness of this test for detecting loose particles. A number of tests have been performed with VVF (monitored) where various-size conducting particles are purposely placed inside the package. While the results are inconclusive, they indicate that VVF (Monitored) is not an effective screen for detecting such failure modes.

Monitored VVF does not seem to be an effective means of locating particles, but it is probably as effective as monitored random vibrations, a much more expensive approach. Neither one is as effective as visual inspection or X-ray.

Monitored vibration tests are part-dependent; they are more effective for particle detection on some devices than on others. Low-voltage devices will rarely fail because of loose particles. High-voltage devices on the other hand will frequently fail if conducting loose particles are inside the package.

4.2.5.11 Random Vibration (Unmonitored)

The elements of random vibration (unmonitored) tests are as follows:

Failure Mechanisms

Package defects

Die bond

Wire bonds

Cracked substrate

Deleterious Effect. None.

Position of Test in Testing Sequence. This test is conducted after acceleration or shock if these tests are performed; otherwise, after the last thermal test.

Relative Cost. Expensive.

Effectiveness. Slightly more effective than VVF (unmonitored).

Example of Detailed Stress Levels. Similar to Section 4.2.5.9 except that vibration is random.

Discussion. This test is similar to Section 4.2.5.9 except that it simulates vibration modes that may occur in actual equipment operation. This test is much more expensive than VVF and is not appreciably more effective. Failures are detected during subsequent failure-detection screens.

4.2.5.12 Random Vibration (Monitored)

The elements of random vibration (monitored) are as follows:

Failure Mechanisms

Package defects

Die bond

Wire bonds

Cracked substrate

Conducting particles

Lead dress

Intermittent opens

Deleterious Effect. None.

Position of Test in Testing Sequence. This test is conducted after acceleration or shock if these tests are performed; otherwise, after the last thermal test.

Relative Cost. Very expensive.

Effectiveness. Poor.

Discussion. The difference between this test and that described in Section 4.2.5.11 is the addition of the electrical-parameter monitoring capability. It is very expensive. Again, particle detection is the primary failure mode to be detected. This is accomplished during monitoring or during a subsequent screen.

Because of its cost and relative ineffectiveness, this test is not recommended.

4.2.5.13 X-Ray

The elements of X-ray tests are as follows:

Failure Mechanism. See Discussion.

MIL STD. 750A, Method 2076.

Deleterious Effect. None.

Position of Test in Testing Sequence. After thermal and mechanical tests.

Relative Cost. Moderate.

Effectiveness. Generally good except for materials, such as aluminum, that are transparent to X-rays. See Discussion.

Discussion. Typical effects that can be uncovered by X-ray screening are:

Misbranded covers (upside-down marking)
Offset covers

Voids in the die-to-case adhesive

Foreign material in the package

Loose weld splatter

Excess gold from gold preforms

Weld particles on wires

Double bonds on the leads and substrate pads

Bonds near the edge of bonding pads and leads

Excess slack in the bonding wires

Excess wire at bond terminations (pigtails)

Extraneous wires in the package cavity or trapped within the package seal

Process nonuniformities

Conducting loose particles

This screen provides an opportunity to inspect a device internally after it has been sealed. Two approaches are possible:

(1) An X-ray vidicon system that displays the X-ray image on a TV screen. A parts manipulator can be used with this system to allow viewing in the X, Y, and Z directions.

(2) Radiographic equipment used in conjunction with developing facilities and ultimate microscopic examinations of the X-ray films.

The first is more expensive than the second. This screen should follow the thermal and mechanical environmental tests to detect induced failures or possible failures such as lead-dress defects, dislodged conducting loose particles, etc.

This test is effective but may be time-consuming and is expensive, especially when conducting loose particles are the primary failure mode to be detected. The transparence of silicon and aluminum to X-rays limits the effectiveness of this screen.

4.2.5.14 Hermeticity Tests, General

Hermeticity tests are performed to detect leaks in any portion of the surface area or seal of the package. Typically a leak rate greater than 10^{-8} Atm cc/sec is considered a failure. In actual practice, devices that do not meet this criterion are sometimes used because it is difficult to say exactly what leak rate is acceptable. For devices in the "gray" area, potting and conformal coatings are used to improve their leak rate. In general, leak testing is inexpensive and effective if performed correctly. Molded plastic packages are not amenable to leak testing because there is no cavity.

4.2.5.15 Helium Leak Test

The elements of helium leak test are as follows:

Failure Mechanism

Fine package leaks in the range of 10^{-6} to 10^{-10} Atm. cc/sec.

MIL STD. 202C, Method 112, Condition C.

Deleterious Effect. None.

Position of Test in Testing Sequence. This is a failure-detection test; it is performed after all physical and thermal environmental tests except those that would result in plugging of device leaks, such as oil-to-oil thermal shock.

Relative Cost. Moderate.

Effectiveness. Good.

Example of Detailed Stress Levels. Devices to be tested are placed in a sealed chamber at a high helium pressure (50 psi) for a specified period of time (typically 4 hours). The devices are removed from the pressure chamber and are tested in a mass spectrometer to detect helium that may have entered a faulty package during the pressure treatment.

Discussion. This screen is effective for fine leaks and has the distinct advantage of introducing no incipient failure mechanisms, which may occur in certain pressure-bomb or bubble tests.

4.2.5.16 Radiflo Leak Test

The elements of radiflo leak tests are as follows:

Failure Mechanism

Fine package leaks in the range at 10^{-8} to 10^{-12} Atm cc/sec.

MIL STD. 202C, Method 112, Condition C.

Deleterious Effect. None.

Position of Test in Testing Sequence. This is a failure-detection test; it is performed after all physical and thermal environmental tests except those that would result in plugging of device leaks, such as oil-to-oil thermal shock.

Relative Cost. Moderate.

Effectiveness. Good. This test can detect finer leaks than the helium test.

Example of Detailed Stress Levels. A radioactive tracer gas is used instead of helium. Any gas trapped in a faulty package is detected with a calibrated ionization gauge.

Discussion. This test is usually performed in place of the helium test if performed at all. The considerations discussed for the helium test apply, as does the procedure under which it is performed.

This is an effective test when properly performed and will not introduce incipient failure modes. Either this screen or the helium screen is mandatory for high-reliability devices.

4.2.5.17 Nitrogen Bomb Test

Failure Mechanism

Package leaks in the mid-range between gross and fine leaks (10^{-5} to 10^{-8} Atm cc/sec).

Deleterious Effect. None.

Position of Test in Testing Sequence. After fine-leak tests.

Relative Cost. Inexpensive.

Effectiveness. Good but operator-dependent.

Example of Detailed Stress Levels. See Discussion.

Discussion. The devices are placed in a sealed nitrogen gas chamber under a pressure of 150 psi for a minimum of 10 hours. When removed, they are placed in an alcohol bath and observed under a microscope. The criterion for failure is the observation of a continuous or intermittent stream of bubbles from the package.

This is a relatively inexpensive test. It is usually performed in connection with the fine-leak test.

Its effectiveness is limited by the necessity of having an operator interpret the results.

4.2.5.18 Gross-Leak Test

The elements of gross-leak tests are as follows:

Failure Mechanism. Package leaks greater than 10^{-5} Atm cc/sec.

MIL STD. 202C, Method 112, Condition A.

Deleterious Effect. See Discussion.

Position of Test in Testing Sequence. This screen is performed after the fine-leak test or the nitrogen bomb test if these tests are performed.

Relative Cost. Inexpensive.

Effectiveness. Good if inside volume is large enough. This test is operator-dependent.

Example of Detailed Stress Levels. See Discussion.

Discussion. The purpose of the gross-leak test is to detect poor seals or cracks in the package. It is performed after all thermal and physical-

stress screens have been performed. It is a detection screen rather than a stress screen.

There are a number of ways in which this test is performed. One method is as follows: The device is immersed in glycerine at 125°C or mineral oil at 125°C (as called out in MIL STD 202C). Defects are detected by careful observation of the device while it is immersed in the liquid. A poor seal is indicated by a continuous stream of bubbles emanating from the device.

A somewhat different approach for gross-leak testing is described below. Experience has indicated that the volume of gas in small flat packages is insufficient, in many cases, to be detected when immersed in hot glycerine. A two-step leak test was developed with fluorocarbons used as the liquid.

The device is first immersed in Freon TF (boiling point of 46.6°C) at about room ambient at 90 psi for about an hour. It is then placed in another fluorocarbon (FC-75, boiling point 100°C) at 70°C and ambient pressure. If any of the Freon TF has penetrated a leaking device, it volatilizes and expands sufficiently to eject easily discernible bubbles.

A fluorocarbon is used as the liquid because it is not flammable and will not trap water vapor, which could affect the metal parts of the package.

There are hazards involved in gross-leak testing, such as introducing moisture. The test is difficult to perform since the indication of a leak may last for only a short time and is subject to interpretation by the person performing the test. The ambiguity arises in differentiating between bubbles trapped on the outside of the package and those which originate inside. It is a messy test, and it may damage device markings. The larger the enclosed cavity of the device, the more effective the test. Gross-leak testing, when performed in a way that precludes introduction of moisture into the package that may not be detected, is an effective screen.

4.2.5.19 Electrical Tests (General)

The electrical tests performed (both ac and dc) are included in the detailed specification for a particular device. These are failure-detection tests and are performed after all physical and thermal environmental tests have been conducted. Automatic test equipment is generally used. Testing is designed to detect out-of-tolerance parameters and various defects induced during previous stress tests or inherent in the manufacturing process. Tests may be performed after each stress test, but this is expensive and is usually not done. Such an approach could be economical as part of a sample test procedure. Some of the

tests that may be performed include the following:

D-C Parameter Tests

Input Characteristics

Output Voltage Levels

Noise Immunity Tests

Worst-case d-c

Pulsed Noise

A-C Parameter Tests

Switching Characteristics

Propagation Delay Time

Power Dissipation Tests

Threshold Tests

Input Capacitance

These are only illustrative of the types of tests that can be performed. Others are equally applicable. Regardless of which tests are performed, they should be conducted under stress conditions that assure proper operation over the full temperature range (typically -55°C to +125°C) and load and power-supply variations as specified in the detailed specification.

Performing electrical tests over the full temperature range of -55°C to 125°C is difficult and expensive. Parasitics introduced by the test fixture and the close temperature control required introduce additional complications, especially when high-frequency devices are being tested.

Electrical tests are absolutely necessary, and the degree of testing will depend on the reliability level required relative to the expense involved. Electrical tests are effective, and the cost will depend on the degree of testing and the stress levels involved.

4.2.5.20 High-Voltage Test

Failure Mechanism. Oxide defects.

Deleterious Effect. None.

Position of Test in Testing Sequence. This test is performed after physical environmental and leak tests.

Relative Cost. Inexpensive.

Effectiveness. This test is not widely used, because its effectiveness has not been determined since it is still in the development stage. It

is not appropriate for dielectrically isolated IC's since it is intended to detect shorts through the oxide to the substrate.

Example of Detailed Stress Levels. A high-voltage pulse of about 50 volts is applied between substrate and intraconnections through a current-limiting resistor. Current is monitored with an appropriate meter.

Discussion. This is one test that is not a carry-over from transistor testing. Various oxide defects such as pinholes and cracks lying under the metalization are detected by this test.

4.2.5.21 Isolation Resistance

The elements of isolation-resistance tests are as follows:

Failure Mechanisms

Shorts

Internal wire-to-metal-lid or -case shorts

Lead shorts to die

Contamination between pins

Deleterious Effect. None if voltages are held to safe levels.

Position of Test in Testing Sequence. This test is performed after physical and thermal environmental tests but before electrical parameter tests.

Relative Cost. Inexpensive.

Effectiveness. Effective.

Example of Detailed Stress Levels. A voltage of about 15 volts dc is applied between case and all leads. Resistance measurement should be 15 megohms minimum.

4.2.5.22 Operating Life Tests (General)

An operating life test (also referred to as burn-in testing) is simply operating the device under extreme operating conditions. The stresses usually include temperature and power. The purpose of such tests is to eliminate those devices that would fail early in equipment operating life. This type of test is based on the assumption that the devices have a decreasing failure rate (DFR) with increasing time. This is generally true for semiconductors, at least during the initial period of their life.

This type of test is highly effective for a rapidly decreasing failure rate and less effective for devices that have slowly decreasing failure rates. The choice of stress levels, which affect the failure rate, depend on the type of circuit, how it is processed, its design, the materials used, and the types of tests that preceded the operating life tests.

Various types of operating life tests are discussed below. The duration of such tests depends on the device type, who is doing the testing, funding, and considerations such as those mentioned in the previous paragraph. Typical test durations are 168, 250, and 1000 hours. These choices are usually arbitrary. While the initial choice may be arbitrary, once an organization has selected a test duration, it is reluctant to change, especially to a shorter time.

Of the operating life tests used for integrated circuits, a-c operating life with or without temperature is the most effective and most widely used for high-reliability applications. (High-temperature reverse-bias testing is not as valuable as it once was, because channeling and inversion defects are not as common as they once were.)

4.2.5.23 Intermittent Operating Life

The elements of intermittent-operating-life tests are as follows:

Failure Mechanism. Parameter drift or failure with time due to inversion, channeling, surface contamination, oxide pinhole shorts, or design.

MIL STD. 750A, Method 1036.2.

Deleterious Effect. None.

Position of Test in Testing Sequence. This test is conducted after thermal and physical environmental tests.

Relative Cost. Expensive.

Effectiveness. Probably no better than a-c operating life

Example of Detailed Stress Levels. The operating portion of this test is usually ac. The device is subjected, intermittently, to operating and nonoperating conditions as specified in the detailed specification.

Discussion. This test is intended to detect parameter drift or failure with time and temperature. Temperature acts as an accelerating force for chemical reactions. Drift or failure may be caused by metalization defects, inversion or channeling, diffusion defects, pinholes, surface contamination, marginal parameter design, and poor wire bonding.

A-c voltages are typically applied and interrupted with electronic clocking or manual controls. Presumably more temperature cycling is produced by this method than by the regular a-c life screen. It is probably no better than that screen except for the possible exception of devices with a high-power density.

4.2.5.24 A-C Operating Life

The elements of a-c operating life tests are as follows:

Failure Mechanism. Parameter drift or failure with time due to inversion, channeling, surface contamination, oxide pinhole shorts, or design.

MIL STD. 750A, Method 1026.2

Deleterious Effect. None.

Position of Test in Testing Sequence. This test is performed after the physical and thermal environmental tests.

Relative Cost. Expensive.

Effectiveness. Operating life (ac) tests are very effective. There is disagreement regarding whether this test should be performed at the device or system (subsystem) level. There is probably no single answer since the quality of the device will determine at which level it is most economical to perform this test.

Example of Detailed Stress Levels. Operation at 25°C for from about 100 to 1000 hours with an a-c signal applied.

Discussion. A-c operating life tests are expensive but will remove early failures and are therefore efficient from a reliability standpoint. Failures are detected by appropriate electrical-parameter measurement usually made before and after the operating life test. However, performance is sometimes monitored continuously, and data on time-to-failure are obtained for reliability analysis.

The stress levels and tests to be performed are detailed in appropriate device specifications. Typically a clock or ringing circuit is used to simulate use operating conditions.

4.2.5.25 D-C Operating Life

The elements of d-c operating life tests are as follows:

Failure Mechanism. Essentially the same as described in Section 4.2.5.23.

MIL STD. 750A, Method 1026.2.

Deleterious Effect. None.

Position of Test in Testing Sequence. This test is conducted after thermal and physical environmental tests.

Relative Cost. Expensive.

Effectiveness. No mechanisms are activated that could not be better activated by a-c life tests.

Example of Detailed Stress Levels. This test can be performed at ambient or elevated temperature (125°C). Its duration may be from 100 to 1000 or more hours. Only rated d-c bias voltage is applied to the device under test.

Discussion. The a-c operating life test is considered to be more effective.

4.2.5.26 High-Temperature A-C Operating Life

The elements of high-temperature a-c operating life tests are as follows:

Failure Mechanism. Essentially the same as described in Section 4.2.5.23.

MIL STD. 750A, Method 1026.2.

Deleterious Effect. None.

Position of Test in Testing Sequence. This test is performed after the physical and thermal environmental tests.

Relative Cost. Very Expensive (probably the most expensive device test).

Effectiveness. Excellent.

Example of Detailed Stress Levels. An a-c signal is applied to the device usually for a period of 100 to 1000 hours at 125°C.

Discussion. The addition of temperature stress accelerates the various failure modes that are to be detected. Except for the temperature characteristic, this screen is the same as a-c operating life discussed in Section 2.4.5.24.

This is a widely used test and is the most efficient operational test from a reliability standpoint. It is more expensive than the a-c operating life test but also more effective. Failures or drift are detected by appropriate electrical measurements. A clock driver or other ring circuit supplies the a-c input for the devices under test at the appropriate (usually 125°C) elevated temperature.

4.2.5.27 High-Temperature Reverse Bias

The elements of high-temperature reverse-bias tests are as follows:

Failure Mechanism. Essentially the same as described in Section 4.2.5.23.

Failure Mechanism. Channeling and inversion.

Deleterious Effect. None

Position of Test in Testing Sequence. After thermal and environmental tests but before hermeticity and final electrical tests.

Relative Cost. Expensive.

Effectiveness. The value of this test is limited because of the improvement in processing that has reduced the incidence of inversion and channeling.

Example of Detailed Stress Levels. An electrical reverse bias is applied to to the device under test in a high-temperature (125°C) ambient, with the facility to remove the temperature but retain the bias.

Discussion. The high-temperature a-c operating life test will induce channeling and inversion and is preferred because it induces higher temperature gradients.

4.2.5.28 Moisture-Resistance Test

The elements of moisture-resistance tests are as follows:

Failure Mechanism. Junction leakage caused by moisture entering through defective package.

MIL STD. 750A, Method 1021.1.

Deleterious Effect. May be destructive.

Position of Test in Testing Sequence. This test is sometimes used in conjunction with high-temperature reverse-bias tests.

Relative Cost. Inexpensive.

Effectiveness. Unknown but probably no more effective than reverse bias alone.

Example of Detailed Stress Levels. The IC device should be subjected to one bend of the lead fatigue test before being placed in the chamber. The sample unit is then subjected to a specified number of continuous cycles of temperature, humidity, and time.

Discussion. When used in conjunction with reverse-bias tests, it is intended to accelerate device failures caused by leakage or electrolysis caused by moisture forced into the package.

4.2.5.29 Physical Dimensions

The elements of physical dimensions tests are as follows:

Failure Mechanism. Incorrect physical dimensions.

MIL STD. 750A, Method 2066.

Deleterious Effect. None.

Position of Test in Testing Sequence. Usually performed as part of lot-acceptance sampling tests.

Relative Cost. Inexpensive.

Effectiveness. Effective in determining if physical dimensions are in accordance with a particular specification.

4.2.5.30 Solderability

The elements of solderability tests are as follows:

Failure Mechanism. Poor solderability of leads.

MIL STD. 750A, Method 2026.1.

Deleterious Effect. Destructive.

Position of Test in Testing Sequence. This test is conducted after the physical-dimensions test.

Relative Cost. Inexpensive.

Effectiveness. Good.

Example of Detailed Stress Levels. The leads of the IC device should be immersed into a flux (25% by weight of ww resin and 75% by weight of 99% isopropyl alcohol) for a minimum of five seconds to a point approximately 1/16 ±1/32 inch from the body of the device. The fluxed leads should be immersed for 10^{+2}_{-0} seconds to the same depth in molten solder (60% Sn and 40% Pb) maintained at 230°C ± 5°C. There should be no mechanical damage, and the cooled coating of solder on the leads should be uniform over a minimum of 95% of the area. The remaining 5% may contain only pinholes or surface roughness visible under 10× magnification.

4.2.5.31 Soldering Heat

The elements of soldering-heat tests are as follows:

Failure Mechanism. Device resistance to high temperature encountered during soldering.

MIL STD. 750A, Method 2031.1.

Deleterious Effect. Destructive.

Position of Test in Testing Sequence. This test is performed on a sample basis. It is usually the first thermal environmental test performed.

Relative Cost. Inexpensive.

Effectiveness. Good.

Example of Detailed Stress Levels. The device leads are immersed in molten metal at about 260°C for approximately 10 seconds.

4.2.5.32 Lead Fatigue

The elements of lead fatigue tests are as follows:

Failure Mechanism. Poor lead-to-package seals.

MIL STD. 750A, Method 2036.3, Condition E.

Deleterious Effect. Destructive.

Position of Test in Testing Sequence. This test is performed after soldering heat.

Relative Cost. Inexpensive.

Effectiveness. Good.

Example of Detailed Stress Levels. Each lead on each device up to a maximum of ten (corner leads on 14-pin flat packages not included) is subjected to three 90° arcs with weights attached. Use 8 ounce weights for TO-5 packages and 2 ounce weights for flat packages.

4.2.5.33 Terminal Strength

The elements of terminal strength tests are as follows:

Failure Mechanism. Defective device leads, welds, and seals.

MIL STD. 750A, Method 2036-3, Condition A.

Deleterious Effect. Destructive.

Position of Test in Testing Sequence. Conducted after lead fatigue.

Relative Cost. Inexpensive.

Effectiveness. Good.

Example of Detailed Stress Levels. One axial pull of 1 pound for 30 seconds on all leads on each device up to a maximum of 10 (corner leads on 14-pin flat packages not included).

4.2.5.34 Salt Atmosphere

The elements of salt-atmosphere tests are as follows:

Failure Mechanism. Package and lead corrosion.

MIL STD. 750A, Method 1042.1.

Deleterious Effect. Destructive.

Position of Test in Testing Sequence. This test is usually one of the last tests to be performed.

Relative Cost. Inexpensive.

Effectiveness. Good.

Example of Detailed Stress Levels. A salt atmosphere fog having a temperature of 35°C is passed through a chamber containing the IC devices for a period of 24^{+2}_{-0} hours.

4.3 SCREENING TESTS

4.3.1 Introduction

The purpose of reliability screening of integrated circuits is to select and remove from a group of devices those having inferior reliability. This is accomplished by subjecting every device to various electrical, thermal, and environmental stresses for the purpose of failing "weak" devices. These weak devices are culled from the lot by subjecting each device to a failure-detection screen such as hermeticity tests, electrical tests, X-ray inspection, and visual inspection.

The reason for screening is that if the tests and stress levels are correctly selected, inferior devices will fail but superior devices will pass, with no degradation resulting from the tests and stresses applied to the devices.

The screen tests performed are typically some combination of the nondestructive tests discussed in Section 4.2. Ideally, the screen selected for a particular application would cull out all potentially unreliable devices and leave only the highly reliable devices for equipment use. In practice, however, this is not the case. While models have been developed for approaching the ideal, they are only an aid for coping with this problem; an ideal screen has not yet been developed.

4.3.2 Developing a Screen

The criticality of the application has an important bearing on the actual number of tests, the stress levels, and the replication of testing (i.e., how often a particular test is repeated during the screening procedure) that will be included in the screening procedure. The screen chosen is also dependent upon the time and funding available.

Various steps can be taken to develop an effective screening program. The first of these is a thorough knowledge of the screens performed by the manufacturer, including their effectiveness and the competence of the manufacturer in performing these screens. User confidence in these screens will require evaluation of the manufacturer and his procedures. This confidence is usually acquired through vendor qualification.

If the vendor performs the screens in an acceptable manner, it behooves the user to develop his screens to complement those performed by the manufacturer. However, if the vendor screens are not performed in a satisfactory manner, the user will be required to develop a comprehensive screen.

A screening program is intended to provide rapid identification of incipient failures. When failure patterns are recognized, it is advisable to provide this information in the form of feedback so that corrective action required in design, processing, or testing can be initiated to correct the problem.

The second step in a screening program is to identify and define IC failures in terms of failure modes and their related failure mechanisms and determine quantitatively their relative occurrence. Failure modes and mechanisms that are typically found in integrated circuits are presented in Table 6-1.

When particular failure modes or mechanisms are known or suspected to be present, a particular screen should be evolved to detect these unreliable elements.

A detailed understanding of the device characteristics, materials, package, and fabrication technique is essential to developing a meaningful screen at a reasonable cost. Devices that perform the same function may be fabricated with different materials (i.e., aluminum leads instead of gold). The effectiveness of a screen is material-dependent. The stress level that is effective for gold may be ineffective for aluminum because of the difference in mass. The X-ray screen is effective for gold, but aluminum and silicon are transparent to X-rays. Some screens are effective for p-n-isolated IC's but ineffective for dielectrically isolated devices. Only a thorough knowledge of the device to be screened and the effectiveness and limitations of the various tests can produce a useful and reliable screening procedure.

4.3.3 Screening Costs

Screening costs depend upon the number of tests conducted, the stress levels involved, and the particular tests performed. The cost may vary somewhat from manufacturer to manufacturer for the same test at the same stress level. In general, the more expensive the equipment required to induce the stress, the more costly the screen. Random vibration and high temperature burn-in are very costly because of the complex equipment required.

The actual screening cost per integrated circuit will depend upon the number of devices being screened and the extent of the screen program. Screening may be performed during any or all of the following equipment buildup stages:

- Vendor screening
- User device screening
- User module screening
- Subsystem screening
- System screening

Since screening can be done at various stages in equipment buildup, the question arises about where it is most economical to perform a particular screen. The one most often discussed is burn-in. The purpose of burn-in is to cull out those devices that fall into the infant mortality portion of the device's failure/time characteristic. While there is agreement that such screens are required for high-reliability devices, there is disagreement about the duration of such tests and in what state of equipment buildup they should be performed.

Many systems manufacturers believe burn-in should be performed at the device level, while others believe the most effective and economical time is at the subsystem or system level. Ryerson* indicates that emphasis on physical-inspection screening during part manufacture results in failure to detect those potential early failures that result from damage during final stages of part assembly or from mismatched tolerances and undetected material defects that can constitute incipient failures. Such an approach may result in a higher total cost for the same reliability because of the need for repair or higher cost to perform similar screens at a higher equipment level.

A device screen incorporating temperature cycling, acceleration, burn-in, hermeticity, and electrical testing currently costs between $3 and $15 per device**.

Perry† reports that a-c burn-in at high temperature for 250 hours will double the basic price of a semiconductor device; if the data are recorded for delta calculations on three parameters, the basic cost will triple -- as compared with only a 15- to 20-percent increase in unit price for temperature cycling with a three-parameter electrical screen.

Monitored shock testing is about five times as costly as unmonitored shock; and X-ray is about six times as expensive as visual inspection.

Electrical testing, using an automatic tester, costs about $0.03 per parameter at 25°C and $0.10 per parameter at 125°C.

For the user with limited funds and perhaps time, it is necessary to select carefully the screens to be performed. There are trade-offs between costs and effectiveness for an expected reliability level. The tabulation in the next section is intended to provide a starting point for the user in evaluating and developing a screen that requires such trade-offs.

---

* C. M. Ryerson, "Relative Costs of Different Reliability Screening Techniques", Proc. Symposium On Reliability (1967), Washington, D.C.

** T. J. Nowak, "Reliability of Integrated Circuits by Testing", Proc. Symposium on Reliability, (1967), Washington, D.C.

†James N. Perry, "Reliability Screening Techniques", Proc. Symposium on Reliability, (1967), Washington, D.C.

4.3.4 Comparison of Screening Tests

The data contained in Table 4-1 have been abstracted from Section 4.2. Only those tests that can be used for screening are included -- destructive tests are excluded. The tabulation presents common quality defects found in high-usage integrated-circuit devices and the screening tests most likely to detect these defects. In most cases defects are grouped under a generic heading. For example, metalization defects include thin metalizations, scratches, voids, tool marks, etc.

Also presented is an estimate of the relative cost and the effectiveness of the particular screen to detect the particular defect. When modifications of the same tests are included, only those defects to be detected by the modification are presented. However, the modified test in most cases will also include or detect failure modes listed for the basic unmodified screen. For example, monitored vibration is intended to detect conducting loose particles; this failure mode is so indicated. However, it will also induce the same stresses as vibration (unmonitored) and thus is effective in testing for those failure modes.

4.4 TESTING IC ELECTRICAL CHARACTERISTICS

For evaluating the electrical characteristics of the IC, the "specified point" and "black box" test methods can be employed.

4.4.1 Specified-Point Testing

Specified-point testing involves checking the electrical characteristics of the IC between its external leads, or between each lead and the substrate. The responses can be analyzed to determine the IC's ability to perform its various electrical functions. This technique is always used in IC failure analysis. The test fixture consists of a switching arrangement (Figure 4-5) that places various combinations of the collector, base, and emitter terminals of a curve tracer between the external leads of the IC under test.

TABLE 4-1
COMPARISON OF IC SCREENS

Screen	Defects	Effectiveness	Cost	Comments
Internal visual inspection	Lead dress Metalization Oxide Particle Die bond Wire bond Contamination Corrosion Substrate	Excellent	Inexpensive to moderate	This is a mandatory screen for high-reliability devices. Cost will depend upon the depth of the visual inspection.
Infrared	Design (thermal)	Very good	Expensive	For use in design evaluation only.
X-ray	Die bond Lead dress (Gold) Particle Manufacturing (Gross Errors) Seal Package Contamination	Excellent Good Good Good Good Good Good	Moderate	The advantage of this screen is that the die-to-header bond can be examined and some inspection can be performed after encapsulation. However, some materials are transparent to X-rays (i.e., Al and Si) and the cost may be as high as 6 times that of visual inspection, depending upon the complexity of the test system.
High-temperature storage	Electrical (Stability) Metalization Bulk silicon Corrosion	Good	Very inexpensive	This is a highly desirable screen.
Temperature cycling	Package Seal Die bond Wire bond Cracked substrate Thermal mismatch	Good	Very inexpensive	This screen may be one of the most effective for aluminum lead systems.
Thermal shock	Package Seal Die bond Wire bond Cracked substrate Thermal mismatch	Good	Inexpensive	This screen is similar to temperature cycling but induces higher stress levels. As a screen it is probably no better than temperature cycling.
Constant acceleration	Lead dress Die bond Wire bond Cracked substrate	Good	Moderate	At 20,000-G stress levels, the effectiveness of this screen for aluminum leads is questionable.
Shock (unmonitored)	Lead dress	Poor	Moderate	The drop-shock test is considered inferior to constant acceleration. However, the pneupactor shock test may be more effective. Shock tests may be destructive.
Shock (monitored)	Particles Intermittent short Intermittent open	Poor Fair Fair	Expensive	Visual or X-ray inspection is preferred for particle detection.
Vibration fatigue	Lead dress Package Die bond Wire bond Cracked substrate	Poor	Expensive	This test may be destructive. Except for work hardening it is without merit.

TABLE 4-1 (continued)

Screen	Defects	Effectiveness	Cost	Comments
Vibration variable frequency (unmonitored)	Package Die bond Wire bond Substrate	Fair	Expensive	
Vibration variable frequency (monitored)	Particles Lead dress Intermittent Open	Fair Good Good	Very expensive	The effectiveness of this screen for detecting particles is part-dependent. See Section 4.2.5.10.
Random vibration (unmonitored)	Package Die bond Wire bond Substrate	Good	Expensive	This is a better screen than VVF (unmonitored) especially for space-launch equipment, but it is more expensive.
Random vibration (monitored)	Particles Lead dress Intermittent Open	Fair Good Good	Very expensive	This is one of the most expensive screens; when combined with only fair effectiveness for particle detection, it is not recommended except in very special situations.
Helium leak test	Package Seals	Good	Moderate	This screen is effective for detecting leaks in the range of 10^{-5} to 10^{-10} Atm cc/sec.
Radiflo leak Test	Package Seals	Good	Moderate	This screen is effective for leaks in the range of 10^{-3} to 10^{-12} Atm cc/sec.
Nitrogen bomb test	Package Seals	Good	Inexpensive	This test is effective for detecting leaks between the gross- and fine-leak-detection ranges.
Gross-leak test	Package Seals	Good	Inexpensive	Effectiveness is volume-dependent. Detects leaks greater than 10^{-5} Atm cc/sec.
High-voltage test	Oxide	Good	Inexpensive	Effectiveness is fabrication dependent. See section 4.2.5.10.
Isolation resistance	*Lead dress* Metalization Contamination	Fair	Inexpensive	
Intermittent operating life	Metalization Bulk silicon Oxide Inversion/channeling Design Parameter drift Contamination	Good	Expensive	Probably no better than a-c operating life.
A-c operating life	Metalization Bulk silicon Oxide Inversion/channeling Design Parameter Contamination	Very Good	Expensive	
D-c operating life	Essentially the same as intermittent life.	Good	Expensive	Some mechanisms are activated that could not be better activated by a-c life tests.
High-temperature a-c operating life	Same as a-c operating life	Excellent	Very expensive	Temperature acts to accelerate failure mechanisms. This is probably the most expensive screen and one of the most effective.
High-temperature reverse bias	Inversion/channeling	Poor	Expensive	See Section 4.2.5.10.

An example of specified-point testing is shown in the functional schematic of Figure 4-6. The circuit is a fully integrated power amplifier (Darlington

COURTESY OF ELECTRONIC DESIGN NEWS

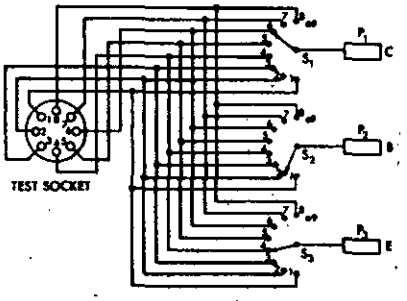


FIGURE 4-5
INTEGRATED-CIRCUIT TEST FIXTURE

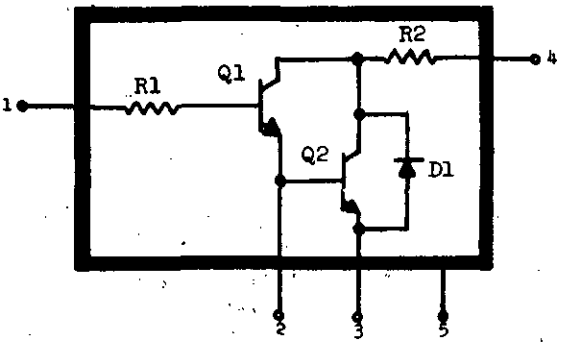


FIGURE 4-6
FUNCTIONAL SCHEMATIC OF POWER AMPLIFIER

configuration) containing two resistors, two transistors, and one diode. Examination of the circuit schematic reveals that both transistors (Q_1 and Q_2) are accessible. This situation would normally indicate that accurate measurements could be made of such parameters as current gain, breakdown and saturation voltages, and leakage currents. However, fully integrated circuits have associated d-c parasitics which often cause parameter variations. Suppose, for example, that a measurement of the current gain of Q_1 is desired. The collector, base, and emitter terminals of the curve tracer (common NPN) are applied to IC leads 4, 1, and 2, respectively. The resultant display on the curve tracer is a combination of Q_1 current-gain characteristics, the leakage currents of the parasitic transistors, and the collector-base leakage of Q_2 (see Figure 4-7). Though these leakage currents are usually small, they can become a problem in power-circuit measurements.

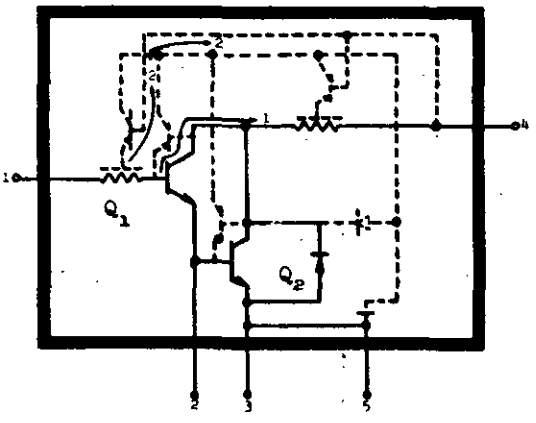


FIGURE 4-7
DC EQUIVALENT SCHEMATIC OF POWER AMPLIFIER

Another test that could be performed by the specified-point technique is the measurement of the collector-to-base breakdown voltage (V_{CB}) of Q_1. This test is performed by switching the collector and emitter terminals of the curve tracer (common NPN) to IC leads 4 and 1, respectively. The display on the curve tracer shows V_{CB} plus all other parasitic junction characteristics. If during this test a breakdown occurs in the Q_1 junction, resistors R1 and R2 provide current limiting for leakage path 1. However, there is no current limiter in the parasitic-junction leakage path 2.

If the engineer knows the equivalent circuit of an IC (including parasitics), he can apply the specified-point test method for an effective d-c analysis of simple circuits. For complex circuits, however, this task becomes difficult, if not impossible, because of the large number of d-c parasitics involved. As mentioned earlier, specified-point testing is generally reserved for IC failure analysis.

4.4.2 Black-Box Testing

Because integrated circuits are generally too complex for the specified-point method of testing, the black-box technique is used. This technique tests the circuit in its entirety. The device is connected to the required power supplies; and the input and output terminals are monitored. The circuit is then accepted or rejected by its circuit characteristics (input, output transfer) rather than by the values of the individual components of the circuit.

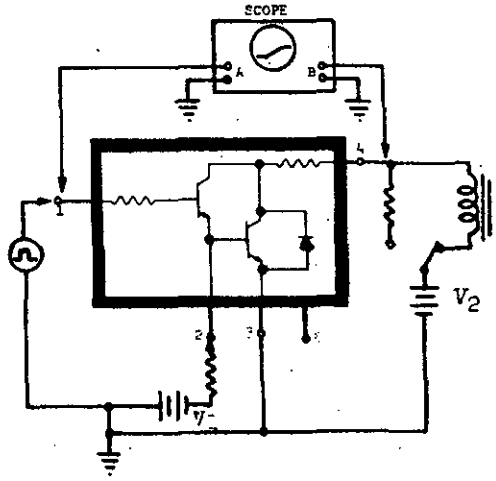


FIGURE 4-8

BLACK-BOX TESTING OF POWER AMPLIFIER

Figure 4-8 illustrates the black-box technique as applied to a power amplifier. Bias voltages (V_1 and V_2) are applied to the IC, and the device is loaded inductively to simulate actual circuit conditions. The input is driven by a square-wave generator, and the input-output response is monitored by a two-channel oscilloscope. Using this test technique, one can determine values for such parameters as dynamic impedance, minimum/maximum operating levels, drain currents, and transient responses. In effect, all necessary operational characteristics of this IC can be determined without a detailed investigation of the package contents (as was necessary with the specified-point technique).

The black-box technique can also be applied to digital integrated circuits. One important measurement in digital IC's is output voltage versus input voltage. Figure 4-9 shows a typical test set-up for this measurement, the digital device being a TTL logic gate. Note that the output of this circuit is leveled. Output loading is necessary for circuits that do not have unilateral characteristics, or whose output resistance is comparable to input resistance over part of the operating range. (In the case of the TTL, both output voltage and current are dependent on a load state.) Although this test can be performed on an ordinary oscilloscope with direct input to the horizontal plates, an oscilloscope with a horizontal-amplification capability (Tektronix type 536 or equivalent) will permit more accurate measurements.

To measure the inverse-voltage transfer characteristics, the test leads shown in Figure 4-9 are reconnected, with the second-stage input voltage going to the Y-deflection terminals of the oscilloscope and the second-stage output going to the X-deflection terminals of the scope and to the variable-voltage source. In addition, bias voltages must be provided at the input of the first stage to provide one-input and zero-input logic levels. Two waveforms can be obtained in this manner -- one for an input "one" condition, and one for an input "zero" condition.

Figure 4-10 shows a standard test arrangement for determining input current versus input voltage (using the same TTL as discussed previously). As in the case of the output-voltage versus input-voltage characteristic, a horizontal drive amplification is required. In the present case, however, the vertical deflection must be proportional to the voltage drop across the current-sensing resistor (R). This measurement must be performed with caution on gate circuits with high input impedances, since the current-sensing resistors have high values; thus, the input impedance of the oscilloscope can cause current-shunting effects that tend to make the input impedance of the circuit under test appear lower than it actually is. As in the previous example, a load stage is required to test the circuit under appropriate system conditions.

Figure 4-11 shows the test circuit for evaluating output current versus output voltage. As in the case of the inverse characteristic, curves are obtained for the "one" and the "zero" input conditions. A differential input to the oscilloscope is required to sense the current through resistor R.

Probably the most difficult black-box measurement is the current-transfer characteristic, since R2 (see Figure 4-12) must be large enough to measure current accurately, but not so large as to have an appreciable effect on gate operation.

Other combinations of input and output voltages and currents can also be measured by the foregoing technique. These techniques apply not only to logic gates, but also to flip-flops, Schmitt triggers, and other direct-coupled circuits (Figure 4-13).

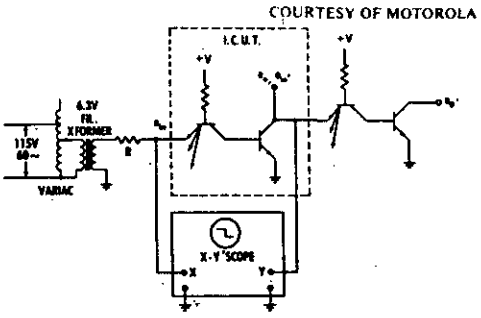


FIGURE 4-9

TEST SETUP FOR MEASUREMENT OF

θ_{OUT} VS θ_{IN} CHARACTERISTICS (T^2L)

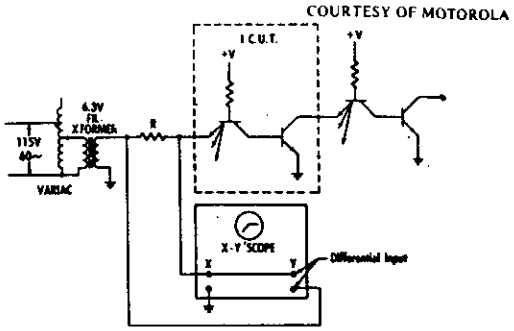


FIGURE 4-10

TEST SETUP FOR MEASUREMENT OF

i_{IN} VS θ_{IN} CHARACTERISTICS (T^2L)

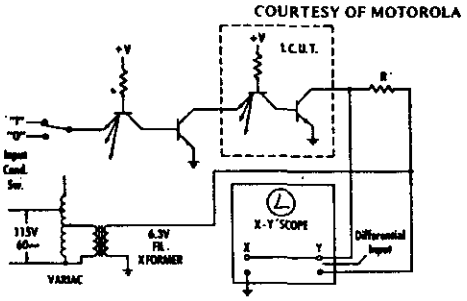


FIGURE 4-11

TEST SETUP FOR MEASUREMENT OF

i_{OUT} VS θ_{OUT} CHARACTERISTICS (T^2L)

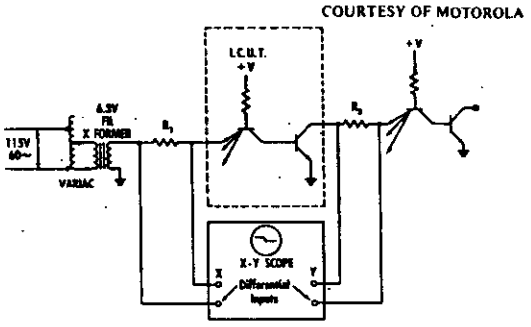


FIGURE 4-12

TEST SETUP FOR MEASUREMENT OF

i_{OUT} VS i_{IN} CHARACTERISTICS (T^2L)

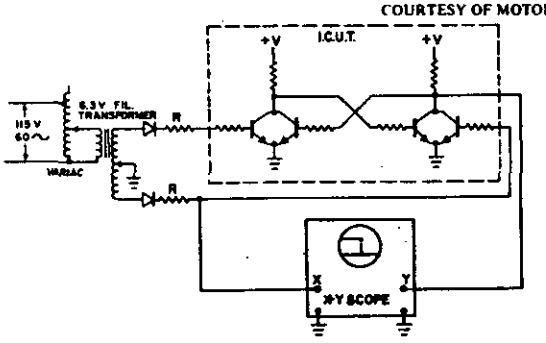


FIGURE 4-13

TEST SETUP FOR MEASUREMENT OF

θ_{OUT} VS θ_{IN} TRANSFER CHARACTERISTICS

OF AN R-S FLIP-FLOP (RTL)

The switching or transient response of the IC device can be evaluated by a variety of methods. Figure 4-14 shows an arrangement for measuring average propagation-delay time. This arrangement allows a large number of samples to be

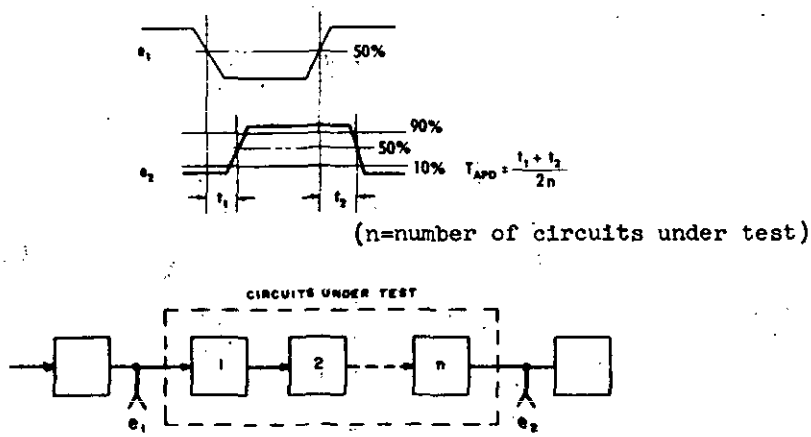


FIGURE 4-14

CIRCUIT FOR MEASURING PROPAGATION-DELAY TIME

checked simultaneously, but this practice has one drawback: the devices could collectively meet the specification requirement although some might be fast and some slow. Therefore, it is usually preferable to test units individually for propagation-delay time. The driving and loading gates can be of either IC or discrete-component form.

4.5 PROBLEMS ASSOCIATED WITH TESTING INTEGRATED CIRCUITS

The most common problem in testing IC's is the variation in normal device operation because of test-fixture capacitance and inductance. All fixtures should be constructed on ground planes with fully bypassed power supplies and minimal lead lengths.

For IC's, all high-frequency measurements (such as roll-off characteristics) are difficult and time-consuming. Oscilloscopes with fast response times are needed to check roll-off distortion during the voltage-swing tests.

Other problems associated with IC testing are as follows:

(1) A small impedance mismatch between the driving source and input of the circuit may cause an excessive amount of ringing and greatly distort the driving waveform.

(2) The sequence of application of power supplies is important; if one bias supply is connected before another in some IC's, an internal junction may be forward-biased and excessive current flow may destroy the device.

(3) In many linear circuits, slight variations in bias supplies drastically affect the circuit output. Bypassing is necessary on all power supplies.

(4) Many circuits are required to have good tracking ability with respect to temperature and drift. The test fixtures used for these temperature tests necessarily have long leads, which makes the measurement of transient response extremely difficult.

(5) Noise is always a problem, and discrimination and rejection circuits may be needed.

(6) It is difficult to limit the number of tests to be performed on an analog circuit and still be confident of its performance. Tests such as phase shift, pulse-width modulation, and frequency response are time-consuming but frequently necessary.

The tests for evaluating IC's are quite similar to those for testing discrete-device circuitry at the module level. IC testing differs in that a designer cannot always probe certain portions of the circuit to determine why the output or input functions changed during a test.

4.6 TEST FIXTURES

A number of fixtures for securing and testing TO-5 and flat packages are available. Unfortunately, many of these fixtures have serious shortcomings. In some, the IC must be welded or soldered into the holder area, and unsoldered or cut upon removal. In others, the IC is clamped in place with inadequate contact pressure, such that test results can be compromised. Some fixtures are not polarized for proper IC insertion. Some do not allow sufficient exposure of IC surfaces for identification markings. Many fixtures cannot be grouped or stacked readily for storage.

Figure 4-15 shows test sockets for use with the modified TO-5 package. The major difference between these sockets and those designed for the standard TO-5 is the number of terminals. Some of the sockets are especially designed for ease of insertion and can accept lead lengths of 0.5 to 1.5 inches. The sockets are polarized and are available with 5, 8, or 10 leads.

COURTESY OF BARNES CORP.

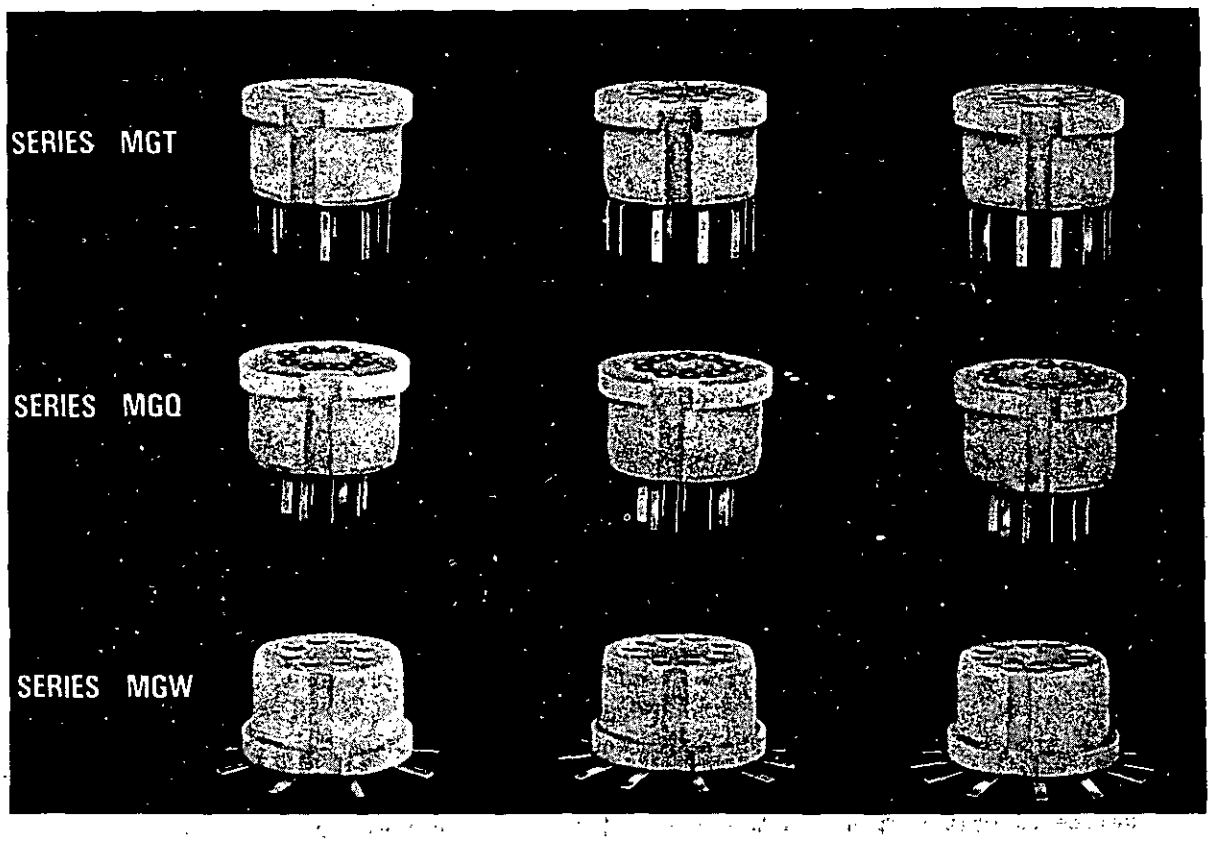


FIGURE 4-15

TEST SOCKETS FOR TO-5 PACKAGE

A carrier for the flat-pack configuration is illustrated in Figure 4-16. This carrier accepts a 1/4 x 1/4 or 1/8 x 1/4 flat pack, and converts the IC pin configuration into an arrangement for plugging into a printed-circuit board (0.050 x 0.100 centers). The IC must be soldered into this fixture.

Figure 4-17 shows another test board. In this test fixture, the depression of two buttons raises spring-type fingers. The IC is placed under these fingers, the buttons are released, and the package is held in place by the spring-loaded fingers. This fixture accepts the 1/8 x 1/4 flat pack with 10 or 14 leads. The test board accepts standard printed-circuit card connectors. One shortcoming of this fixture is the difficulty of aligning the IC leads directly beneath the spring fingers; the resulting poor contact can lessen the accuracy of test measurements.

COURTESY OF WALKIRT

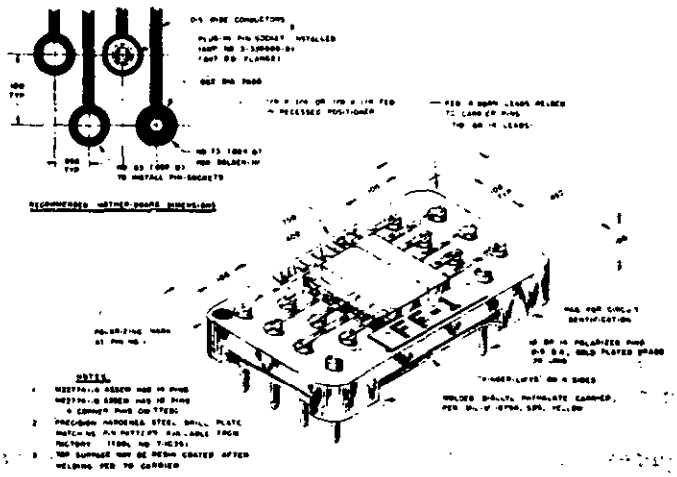


FIGURE 4-16

CARRIER FOR FLAT PACKAGE

COURTESY OF MODULAR ELECTRONICS

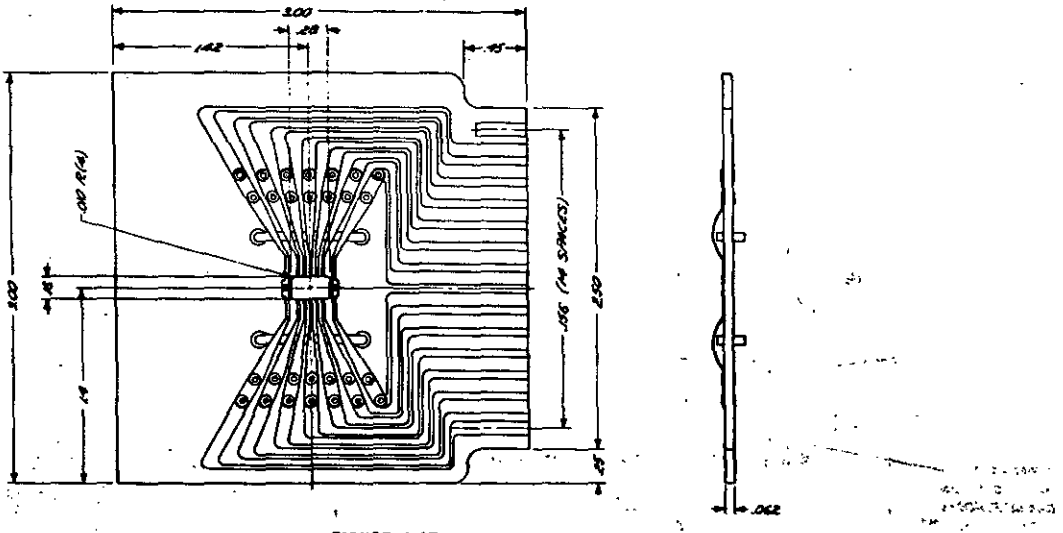


FIGURE 4-17

INTEGRATED-CIRCUIT TEST BOARD

Several variations of the "flip-top" carrier are available; one such is shown in Figure 4-18. The flip-top carrier has the following advantages:

(1) Device leads are positioned automatically by the fixture cavities.

(2) Uniform and adequate pressure is exerted on all leads.

(3) A variety of package configurations, with 10 to 14 leads, can be accepted.

(4) The enclosure can be used for shipping as well as testing, since the fixture completely surrounds the integrated circuit.

(5) The terminals are large enough that wires can be soldered directly to them.

The fixture has terminals that accept the Barnes RD-24 socket. The price of the fixture is about $4 in quantities of 100.

The Auto-Pak carrier consists of a test socket, cover, and base, as shown in Figure 4-19. An IC device can be tested in this carrier and then shipped to the customer. The recipient can perform the same tests, utilizing the carrier with a test socket. The carriers cost 15 cents apiece in quantities of 5,000, and can be returned to the vendor for partial credit after the IC's are removed. The test socket and retainer clip must be purchased separately.

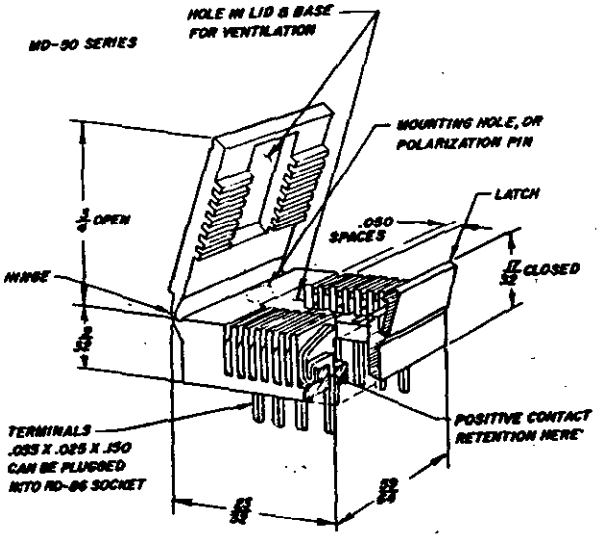


FIGURE 4-18

FLIP-TOP CARRIER

COURTESY OF WESTINGHOUSE

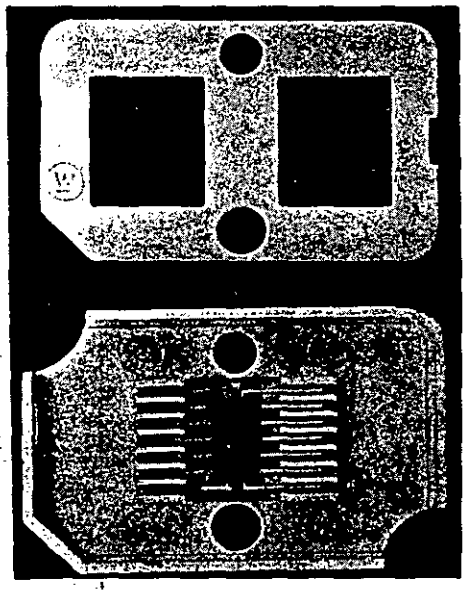


FIGURE 4-19

AUTO PAK CARRIER

The IC socket assembly shown in figure 4-20 has a double-sided printed-circuit board for use with a standard double-row printed-circuit connector. Two or more units can be installed side-by-side in one double-row connector.

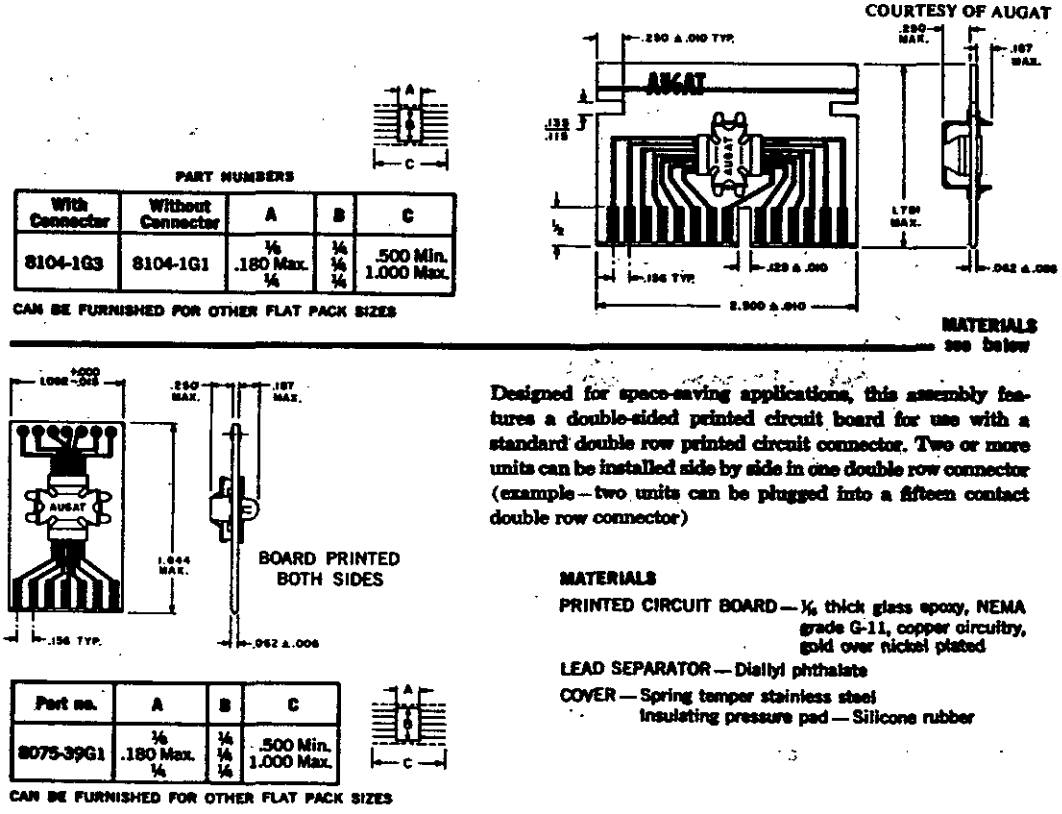


FIGURE 4-20

IC TEST SOCKET

Several types of multiple-unit test boards are available for testing and bread-boarding integrated circuits. Figure 4-21 shows a test board for DIP's; the board in Figure 4-22 is for flat packs.

COURTESY OF AUGAT

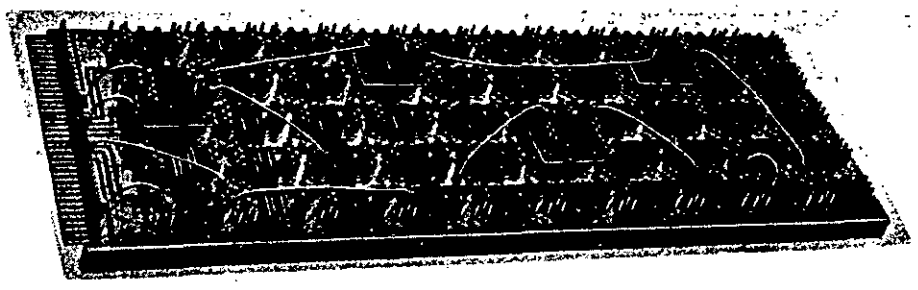


FIGURE 4-21

DIP TEST BOARD

COURTESY OF AUGAT

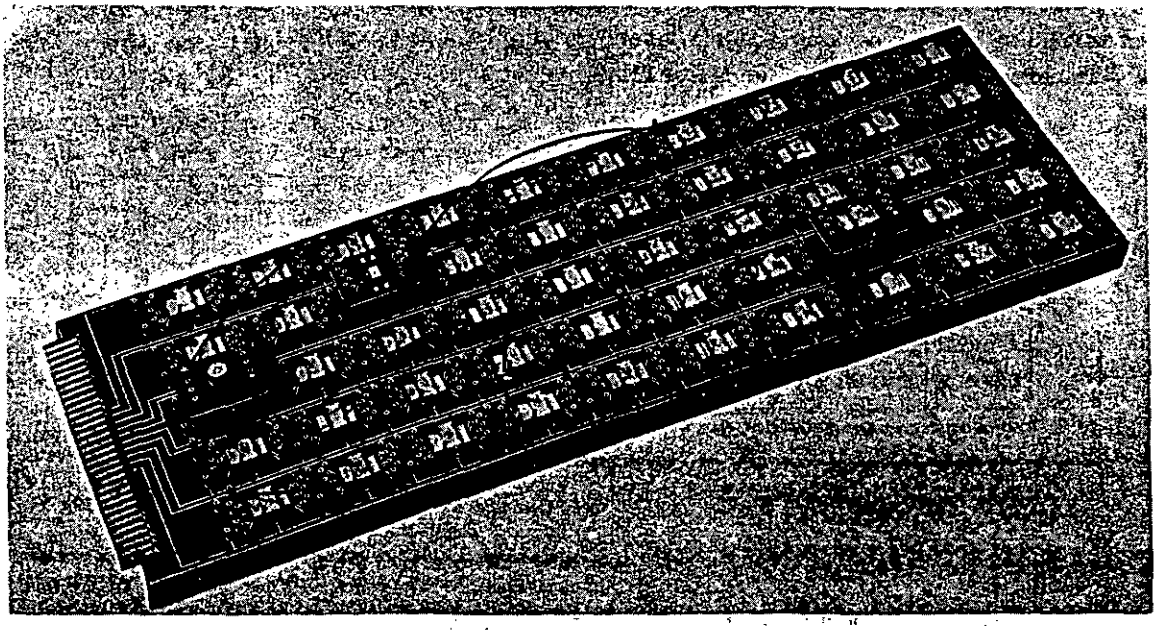


FIGURE 4-22

PRINTED-WIRING BOARD FOR FLAT PACKS

As with test sockets, there is a large selection of test boards.

One of the problems associated with test sockets and fixtures is that IC manufacturers are continually producing new packages, while the socket manufacturers are still trying to make effective low-cost holders for the packages that are already standardized.

4.7 TEST EQUIPMENT

Integrated-circuit test equipment is used for engineering, quality assurance, and production control. The equipments must be adaptable electrically and mechanically to a variety of circuits and must be easily programmed.

Quality-assurance test equipments are employed by both manufacturer and user to verify the quality of integrated circuits. The complexity of these equipments varies from relatively simple to very complex. The more stringent the reliability requirement the more complex the test equipment.

The manufacturer's production-control equipment is usually intricate and expensive, since it must perform complex test programs rapidly and accurately. A few manufacturers are implementing on-line testing and computer evaluation during various steps of fabrication.

Some IC users are building their own test equipment. In doing so they must consider the following:

(1) Whether the device to be tested is digital or linear. Where both types of devices are to be tested, complex equipment will be required.

(2) Whether the device must be exposed to unusual environments (high temperature, shock) while under electrical test. This requirement could necessitate sophisticated fixtures and testing techniques.

(3) The sequence of testing. An inappropriate test sequence could have an adverse effect on time, cost, yields, and failure information.

(4) The method of measuring transient response. The requirements for fast switching speeds may necessitate several fixture cards rather than a single universal fixture.

Many manufacturers and users are purchasing commercial test equipments, which are becoming widely available.

Factors* that should be considered in the purchase of commercial IC testers include the following:

(1) Semiconductor test-equipment repeatability is the most important factor for evaluation, reliability, and quality control. Repeatability requirements are typically specified in millivolts and hundreds of picoamps. Repeatability is affected by the complexity of the test equipment. Generally, the more complex the test equipment, the more difficult it is to obtain close tolerance repeatability.

(2) Considerations that affect the speed of the tester are as follows:

- Automatic handling equipment
- Multiplexing capabilities
- Environmental chambers

Type of test and test rate are key factors when automatic handling equipment is being considered. If handling equipment is required and the package to be tested is nonstandard, delivery of the tester could be delayed for several months. Test speeds of 16 msec per test are readily available, and high-speed options will cut test time to 5 msec per test.

(3) Multiplexing increases the effectiveness of a tester by allowing the equipment to test more than a single IC at the same time. Consider a tester that costs $80,000; if a multiplexing option with two additional test fixtures that costs about $10,000 is added, three times as many devices can be tested during the same period of time for only a modest increase in cost. Another obvious advantage of multiplexing is the saving in floor space.

* Robert G. Hart and Francis J. Kelliher, "Eight Pointers to Remember When Buying IC and Semiconductor Testers", Evaluation Engineering, May/June, 1966, p. 20.

Multiplexing equipment does not require that similar devices be tested at the same time. Different types of devices may be tested simultaneously if the programs are stored on random-access memories such as magnetic discs. In addition, the tests are not restricted to the same environment; one group may be tested in an environmental chamber, while another group may be in an ambient environment.

Devices are often required to be tested under a wide range of temperature conditions. Such measurements require automatic handling through the environmental chamber. The device must be stabilized at the various temperatures; frost at low temperatures must be avoided or readings will be inaccurate.

(4) Computer control is used on many IC testers. Serious consideration should be given to adaptation flexibility that permits the use of more than a single type of computer. Actual and anticipated test needs must be determined to avoid overbuying or underbuying of computer controlled testers.

A low-cost computer may require much more programming than a more expensive computer over the anticipated life of the tester. The life-cycle costs for the inexpensive computer may be much higher than those of the expensive computer. When computer-controlled testers are used, it is not necessary that the computer and the tester be in close proximity.

(5) The data-recording equipment used with the IC tester should be compatible with other data-processing equipment currently being used. Requirements for test rates must be carefully evaluated because data recording is usually the limiting factor in determining test speed. Magnetic tape is the fastest data-recording medium currently in use.

(6) Equipment flexibility is an important consideration because the IC technology is changing. Testers that can handle up to 40 leads are available; many can be adapted to handle more. This is an important consideration in the light of expected developments, especially LSI. Such devices may have up to 160 leads.

BIBLIOGRAPHY: CHAPTER 4

TESTING

Berger, W. "Infrared Detection Reveals Microcircuit Metalization Failure Mechanisms," Evaluation Engineering, (1966).

Bureau of Naval Weapons. Naval Weapons Data; Reliability Engineering, Reliability Testing. 1962.

Burrus, J.C. Shock and Vibration Bulletin; Combined Random Vibration and Extreme Temperature Testing of Integrated Circuits. 1966.

Carlson, H.G., W.L. Gill, and J.E. Hall. Investigation of Reliability Testing and Prediction Techniques for Integrated Circuits. 1966.

Diaz, G. "Computer Methods for Analyzing Test Data," Electro-Technology, 1963.

Fraser, R.M. Thermal (Infrared) Radiometers as Instruments for Nondestructive Reliability Testing. 1966.

Fountain, J.H. A General Computer Simulation Technique for Assessments and Testing Requirements. General Dynamics/Convair Report.

Giles, J.W. "How to Measure Linear-IC Performance," EEE, (1966), p.62.

Hart, R.G., and F.J. Kelliher. "Eight Pointers to Remeber when Buying IC and Semiconductor Testers," Evaluation Engineering, (1966).

Jelinek, H.J., and H.T. Breen. "Proper Test-Point Allocation," Electro-Technology, (1966), pp.50-53.

Jones, A.B., Jr. "Programmable Module Tester," Computer Design, (1966), pp. 42-43.

Kashlwabara N., Siliconix DTL Integrated Circuits Adder/Subtracter Evaluation. 1964.

Lindsay, G.W. "Experiences In The Use of MIL-R-26667, With Comments on Contractual Reliability Test Requirements and MIL-STD-781," Analysis of Reliability and Maintainability, (1965), pp. 28-30.

Lombardi, J., L. McDonough., and H. Padden. High Reliability Screening of Semiconductor and Integrated Circuits Devices. Grumman Final Report. 1966.

Morkowitz, C. "Integrated Circuits in Action: Part III, Getting Your Money's Worth," Electronics, (1966), pp. 56-62.

_____. "Pattern of Tests Changing for Integrated Circuits," Electronics, (1966), 93-101

Mulcahy, D.L. "Selecting and Specifying Hi-Rel Integrated Circuits," Proc. Annual Symposium on Reliability, (1967), pp. 359-364.

Nowak, T.J. "Reliability of Integrated Circuits by Screening," Proc. Annual Symposium on Reliability. (1967), pp. 365-375.

O"Connell, E.P., and T.W. Walsh. Development of a Microelectronics Capability and Facility at RADC. RADC-TR-65-439.

Peralta, B.C. "Screening Silicon Integrated Circuits," Proc. Annual Symposium on Reliability, (1967), pp. 337-348.

Perry, J.N. "Reliability Screening Techniques," Proc. Annual Symposium on Reliability, (1967), pp. 783-788.

RCA. Final Report on Functional Testing of Digital Microelectronic Circuits. Operations Research Government Services. 1966.

Rome Air Development Center. RADC Specification 2867, Quality and Reliability Assurance Procedures for Monolithic Microcircuits.

Schink, P. "Evaluating Integrated Circuit Performance," Electronics, (1963), p, 73.

Sundeen, G.E., Jr. "Review of Test Equipment Used in Integrated Circuit Manufacturing," SCP and Solid State Technology, (1966), pp. 53-55.

Walsh, T., et al. Quality Defects in Electronic Parts: Phase I Final Report. General Electric 66D282. 1966.

SPECIFICATIONS AND PROCUREMENT

F I V E

5.1 INTRODUCTION

Since the primary purpose of a specification is to inform the supplier of the purchaser's needs, it may consist of nothing more than a simple description of the product. Depending on the nature of the product and its application, such a specification may be entirely adequate.

On the other hand, a specification may be a complex document that requires the manufacturer to spend substantial sums to qualify as a supplier; it may contain detailed descriptions of every known characteristic of the desired item, making the cost of verification alone many times the cost of the manufacturing operation. There are few such specifications; however, many military and aerospace specifications -- particularly those for semiconductor devices -- approach such complexity.

In addition to informing the supplier of the purchaser's needs (which will be referred to herein as performance characteristics), the more comprehensive specifications set forth requirements that the manufacturer must meet before his product will be accepted by the purchaser. These additional requirements generally consist of procedures for qualification, acceptance testing, and preparation for

shipment. Preparation for shipment is covered by Military Specification MIL-M-55565, Microcircuits, Packaging of. The basic functions with which most specifications for military and aerospace electronic devices are concerned are as follows:

(1) Performance characteristics
- Electrical parameters
- Environmental stress levels
- Reliability
- Physical dimensions and form factor

(2) Assurance-test procedures
- Sample plan
- Test sequence
- End points

(3) Qualification procedures
- Management and administrative requirements
- Process requirements
- Product test requirements

5.2 PERFORMANCE CHARACTERISTICS

Performance characteristics and their numerical tolerances are generally determined by the application for which the device is intended. If the application is not reflected in the specification, the procured devices may fail to function properly over at least a portion of the desired operational range. Specified performance characteristics must also be compatible with limitations of the associated manufacturing technology and not imply a requirement to advance the state of the art.

5.2.1 Electrical Parameters

The manner in which electrical parameters are specified for integrated circuits (IC) is essentially identical to that of discrete semiconductors. However, while it is possible to specify all of the electrical parameters considered in the design of the discrete component, this is not generally feasible with integrated circuits, because many IC parameters are inaccessible to measurement. The designer is usually restricted to specifying only those IC parameters involving the input and output terminals of the device and the power supplied in terms of voltage and current.

It is possible for an integrated circuit to be manufactured in such a manner that, even with the terminal parameters well within the specific limits, one or more of the "buried parameters" is either unstable or far removed from its design value, causing the device to fail early. If measurement were possible, the device would be rejected. The only known protection against such a possibility that can be incorporated in an integrated-circuit specification is the test-group concept.

During preparation of the wafer masks, selected areas on the wafer can be reserved wherein active and passive devices are fabricated with open terminals. The assumption is that elements incorporated in adjacent circuits will possess similar electrical parameters. Those elements within the test groups are accessible to measurement, and their electrical parameters can be specified in a manner identical to that for discrete devices.

There are exceptions to the rule of specifying performance characteristics in accordance with the device application. Some electrical parameters of semiconductor devices can indicate device quality or stability. Although the exact numerical range depends on the device and technology involved, there is generally a range of values for device parameters outside of which the quality or stability of the device becomes suspect. For example, the silicon planar-passivated transistor commonly incorporated into integrated circuits typically exhibits extremely low leakage currents, usually orders of magnitude lower than required by most applications. If such a device exhibits leakage merely approaching the maximum allowable for many applications, it is likely to be an early failure. Although such a device could be rejected because of other requirements given in the specification, the best opportunity for rejection is lost if this parameter is specified according to the requirements imposed by its application. Thus electrical parameters should be limited to the numerical range commonly associated with quality and stability irrespective of application requirements.

5.2.2 Environmental Requirements

Environmental characteristics specified for IC devices should reflect the conditions to which the devices will be exposed. Additionally, however, all environmental conditions that are related to known failure mechanisms of the devices can and should be specified (regardless of whether the devices will encounter such conditions in use), in an attempt to screen possible future failures.

The environmental requirements specified for integrated circuits are generally the same as those specified for transistors. Applicable tests are described in MIL-STD-750, Test Methods for Semiconductor Devices. It may be noted that these tests differ from those which military electronic systems are required to undergo. While including all the types of mechanical and environmental stresses required for systems, they are particularly adapted to such low-mass devices as

integrated circuits. Frequently, system manufacturers will specify mechanical and environmental tests for integrated circuits on the basis of the test requirements of their completed systems. It is not particularly desirable to perform these types of tests on integrated circuits, because they will not sufficiently accelerate the potential failure modes.

5.3 ASSURANCE TESTING *

It should be generally assumed that devices procured to a specification will possess only those characteristics that are verified by the test requirements of the document. This assumption is seldom precisely correct, but it is always advisable because it will usually prevent inadequate devices from finding their way into systems -- whether their inadequacy is due to failure to comply with specifications or to misapplication of the device by the system designer.

The test requirements alone define the device. The manner in which they are specified is of vital importance, particularly for military and space applications.

5.3.1 Sampling Plans

In assurance testing, samples of the population are tested in a trade-off of confidence against cost and time. Before a sample-selection scheme is chosen, the population from which the sample is to be drawn must be precisely and correctly defined, since conclusions of the experiment are limited to that population from which the samples were drawn.

The principal requirement of any sampling procedure is that it yield representative samples. A representative sample is a miniature of the population. To make inferences about the population from the sample results, the sample design must have an element of randomness. A random sample results when every item in the population has an equal and independent chance of being chosen for the sample. (Random refers to the selection method and not to the sample items chosen.) A completely random sample (often referred to as an unrestrictive random sample) does not necessarily result in a representative sample, however. For example, if the population is a group of ±10% resistors of a specified nominal value, the unrestircted random sample is selected in essentially the following manner: All the resistors in the population are identified in some way, usually by number. The sample is drawn by picking a group of numbers in some random manner, i.e., drawing slips of paper out of a hat or using a random-number table. If the resistance values are normally distributed about the nominal value, it is remotely possible through an unlikely sequence of events, that the items chosen for the sample are all close to the ± 10% value. Certainly the sample is unrepresentative, but it is still unbiased since the procedure by which it was obtained was random. Two means of protecting against this occurrence are increased sample size and the introduction of a certain amount of nonrandomness

in the sample procedure -- if something is known about the population. The non-randomness is concerned with an attempt to secure representativeness by dividing the population into more homogeneous segments (known as stratified sampling). No matter how many population divisions are made, the final choice of the sample items (chosen from each category or stratum) must be a random one. This not only protects against bias but allows the experimentor to make inferences with a measurable uncertainty.

Adequacy in a sample refers to the number of items selected, i.e., the sample size. A sample size is adequate for the particular sampling procedure chosen if it satisfies the requirements of representativeness and precision. The optimum sample size, and hence the optimum number of replications in an experiemental design, is dictated by the following factors:

- Inherent variability of the manufactured parts
- Degree of accuracy required
- Degree of precision required
- The experimental design resulting from problem definition
- Economic considerations

Currently in the semiconductor industry, some form of the AQL and the LTPD sampling plans is used. The AQL (Acceptance Quality Level) plan specifies the maximum deviation from perfect quality that the consumer is willing to accept (assuming that the producer will consistently submit lots of this quality). The AQL is usually considered as that quality of a lot which has a probability of acceptance of .95. An inspection level is specified with the AQL which determines the number of samples to be tested. Although the number of samples increases with lot size, the ratio of sample size to lot size decreases.

There are some advantages and disadvantages in the AQL plan. In normal sampling, a lot at AQL quality will have a high probability of acceptance, usually near .95. Thus a producer has good protection against rejection of submitted lots from a process that is at the AQL or better. On the other hand, this type of classification does not specify anything about the protection the consumer has against the acceptance of a lot worse than the AQL. The AQL is therefore referred to as a producer's risk plan.

The other plan is a consumers' risk plan. The LTPD (Lot Tolerance Percent Defective) plan refers to that incoming quality above which there is a small chance that a lot will be accepted; the LTPD value is usually near 10%. In this plan, there is a small probability that the consumer will accept lots submitted from a process that is at the LTPD or worse. However, the plan does not specify the risk the producer has against the rejection of lots better than LTPD.

Different AQL and LTPD sampling plans are shown in Figure 5-1. Plans 1 and 2 have AQL's of A; the LTPD of Plan 1 is D; that of Plan 2 is E. Plan 3 has an AQL of B and an LTPD of E, and Plan 4 has an AQL of C and LTPD of F. Each of these plans has merit for individual applications since the AQL's and the amount of discrimination vary.

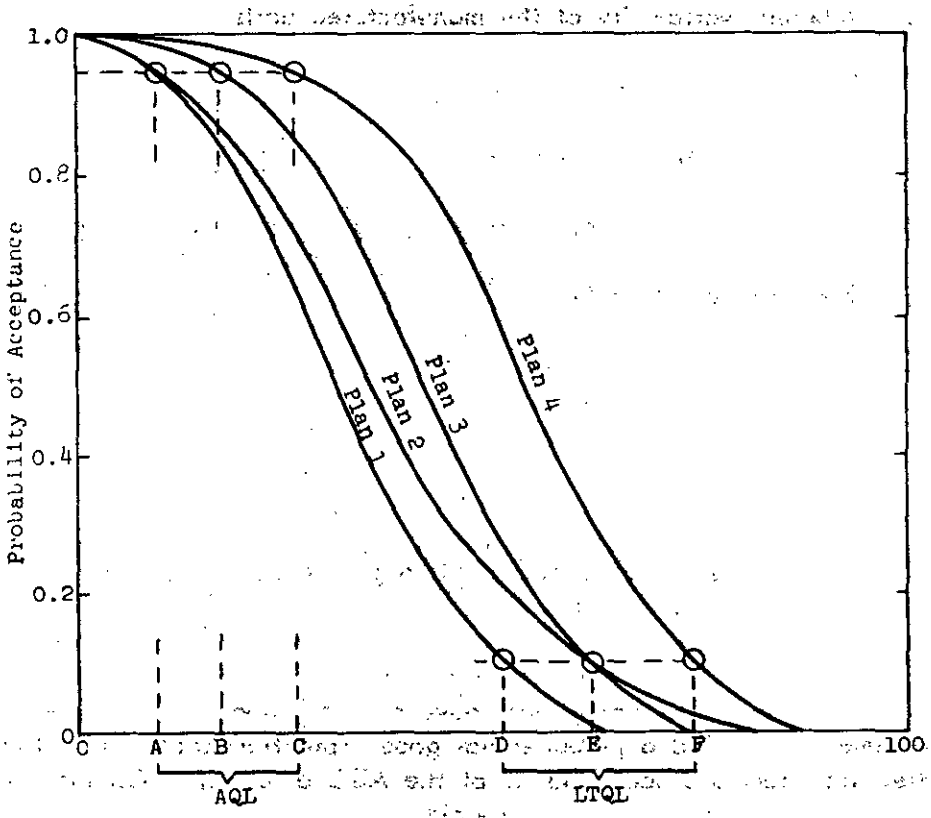


FIGURE 5-1

SEVERAL HYPOTHETICAL OC (OPERATING CHARACTERISTIC) CURVES HAVING DIFFERENT AQL (ACCEPTABLE QUALITY LEVEL) AND LTQL (LOT TOLERANCE QUALITY LEVEL) VALUES

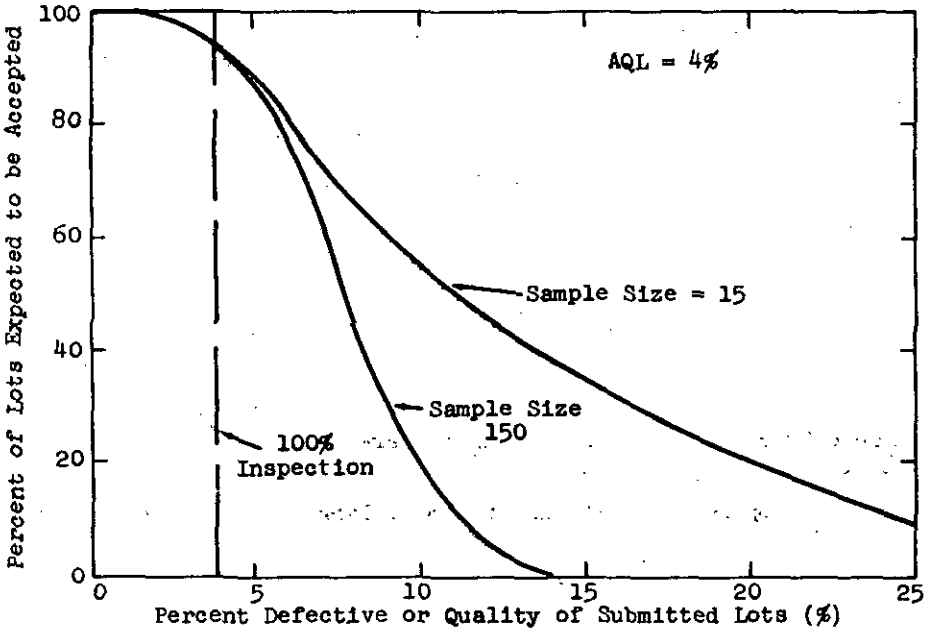


FIGURE 5-2

OPERATING CHARACTERISTICS FOR A 4% AQL

Figure 5-2 shows the effect of sample sizes of 15 and 150 in a plan with an AQL of 4%. As shown, the ideal OC (Operating Characteristic) curve is achieved by 100% inspection. The decreasing sample sizes show the OC varying further from the ideal. The sample size of 150 will permit one lot out of 10 with 11% defectives to pass, while the sample size of 15 will permit a lot of 25% defectives to pass 10% of the time. Since the LTPD controls the lower end of the OC so that only 10% can pass if the specified value is exceeded, then the sample sizes of 15 and 150 assure LTPD's of 25% and 11%, respectively.

For the LTPD plan, sample size is independent of lot size but the percent defective allowed to assure a specified LTPD increases as the sample size approaches 100%. Tables of sample sizes and number of defectives permitted for various LTPD values are included in MIL-S-19500.

Current military specifications frequently use a modified LTPD plan for added consumer protection against the possibility of receiving defective or out-of-tolerance devices, which is greater with the small lot and small sample sizes of the AQL system. The modified plan specifies a minimum-rejection number or a maximum-acceptance number (which is the minimum rejection number minus one) and a maximum size of the sample. Therefore, the lot may be accepted based on test results on smaller sample sizes, but the number of defectives permitted for the sample is lower.

As an example, military semiconductor specifications call for an LTPD value of 5% with a maximum acceptance number of 4 for electrical characteristics. Life test specifications usually do not include the maximum acceptance number because the sample size for the 1,000-hour life test will be limited due to test cost considerations. As pointed out earlier, the test plans usually evolve as a result of trade-offs of confidence against cost and time.

5 3.2 End Points

Generally, the criterion of failure for environmental testing is determined by the effect the stress condition has on the electrical characteristics of the device. These characteristics and their associated limits are referred to as end points.

When only one or two characteristics are used as end points, they should be the device characteristics that are the most sensitive to the particular environmental stress applied. Specifying all or a large number of end points is time-consuming, costly, and generally unnecessary.

Three common methods of specifying end-point limits are as follows:

(1) Maintain initial values. For example,
Gain = 50 minimum, 100 maximum -- initial test value
Gain = 50 minimum, 100 maximum -- end-point limit

(2) Relax initial value. For example,
Gain = 50 minimum, 100 maximum -- initial test value
Gain = 30 minimum, 140 maximum -- end-point limit

(3) Limit shift of initial values. For example,
Gain = 50 minimum, 100 maximum -- initial test value
Initial value ± 20% -- end-point limit

For a variety of practical reasons, method (2) is most widely used although the other two have advantages for specific situations. In any case, careful consideration should be given to the characteristic being specified so a realistic set of intial and end-point values are established based on requirements of accuracy of measurement during test.

5.3.3 Electrical Performance Testing

It has been a common practice for years in the semiconductor industry to perform electrical performance tests at room ambient condition (25°C). Only for some specific high temperature operational requirements were tests such as high-temperature current gain included. The justification for this testing policy is based on the ract that the performance of transistors and diodes over a reasonable range of temperatures is well documented and is predictable. However, neither the same data or test policy can be applied to the more complex integrated circuits. All semiconductor devices exhibit changes in electrical performance that are a function of the temperature of the device. The relationship of these temperature-dependent characteristics in an integrated circuit is hard to predict.

During the early phases of procurement of a given integrated circuit it may be necessary to perform measurements at other than room temperature. As data are accumulated, it should be possible to reduce testing costs by reducing the number of parameters measured at the extreme temperature.

5.3.4 Reliability Screening Tests

Although the process control during production of integrated circuits is essential for reliability, the use of screening tests can be a valuable tool to assure delivery of reliable devices. The screening tests are special tests designed to be performed on the entire population to eliminate devices that are potential failures even though they have passed the standard quality control tests. Because of the cost and nature of the screening tests, they must be carefully designed and used. A detailed knowledge of the physics of failure of the device and the relationship of device design and process steps is required. The effectiveness of a screening test is not constant but varies with design and manufacturing processes.

A screening often specified for integrated circuits in which reliability is of paramount importance is the power burn-in test -- operating each circuit under an accelerated life-test condition for a period between 24 and 250 hours. This operating life test, performed on a 100% basis, will remove early-life failures from the total population and assure a high degree of performance-characteristic stability.

Table 5-1 lists some of the common defects found in integrated circuits and the screening tests that can aid in eliminating devices with these defects. Although the screens are listed in the order of decreasing effectiveness, the difference in effectiveness between tests in the same group are considered minor. For a detailed discussion of each test see Section 4.2.4 and Table 4-1.

5.3.5 JAN-TX (Testing-Extra) Specifications

The TX specification is an effort by DoD to improve semiconductor device reliability and simplify device specifications. It is in limited use for transistors, diodes, and rectifiers and to a lesser extent for integrated circuits. It is expected to eliminate many of the costly device specifications that are invariably developed for each new system. In the past an engineer who wanted to write a high-reliability specification for devices to be used in a new system often collected as many existing specifications as possible that seemed appropriate. The most stringent requirements were then abstracted and combined with a few original contributions to form the new specification. Such an approach has a major effect on cost with no guarantee of a commensurate improvement in reliability. The TX specification should eliminate many of these custom specifications.

TABLE 5-1
IC DEFECTS AND DETECTION SCREENS

Defects	Detection Screen
Lead Dress	Pre-Cap Visual Inspection X-ray (cannot be used for aluminum) Constant Acceleration Vibration Variable Frequency (Monitored) Random Vibration (Monitored) Shock Vibration Fatigue Isolation Resistance
Die Bond	Pre-Cap Visual Inspection X-Ray Temperature Cycling Thermal Shock Constant Acceleration Random Vibration Vibration Variable Frequency Shock Vibration Fatigue
Wire Bond	Pre-Cap Visual Inspection Temperature Cycling Thermal Shock Constant Acceleration Random Vibration Vibration Variable Frequency Shock Vibration Fatigue
Metalization	Pre-Cap Visual Inspection High-Temperature Storage High-Temperature A-C Operating Life A-C Operating Life D-C Operating Life Intermittent Operating Life High-Voltage Test Isolation Resistance
Oxide	Pre-Cap Visual Inspection High Temperature A-C Operating Life A-C Operating Life Intermittent Operating Life D-C Operating Life High-Voltage Test
Conducting Loose Particles	Pre-Cap Inspection X-Ray Vibration Variable Frequency (Monitored) Random Vibration (Monitored)
Contamination	Pre-Cap Inspection High Temperature A-C Operating Life A-C Operating Life Intermittent Operating Life
Substrate (cracked)	Pre-Cap Visual Inspection Temperature Cycling Thermal Shock Constant Acceleration Vibration Variable Frequency Random Vibration Shock Vibration
Corrosion	Pre-Cap Visual Inspection High-Temperature Storage
Package & Seals	X-Ray Temperature Cycling Thermal Shock Random Vibration Helium Leak Test (fine leak) Radiflo Leak Test (fine leak) Nitrogen Bomb Test (mid-range leak test) Gross Leak Test Shock Vibration Variable Frequency Vibration Fatigue
Bulk Silicon	High-Temperature Storage High-Temperature A-C Operating Life A-C Operating Life D-C Operating Life Intermittent Operating Life
Thermal Mismatch	Temperature Cycling Thermal Shock
Parameter Drift	High-Temperature A-C Operating Life A-C Operating Life D-C Operating Life Intermittent Operating Life
Inversion/Channeling	High-Temperature A-C Operating Life High-Temperature Reverse Bias A-C Operating Life Intermittent Operating Life D-C Operating Life
Design	Operating-Life Tests Infrared (Thermal)

The TX specification does not introduce new tests but simply specifies that certain tests be performed on a 100% basis. In addition, it requires identification of each device that is tested and requires repeat measurements of key parameters. The TX specification outlines a standard screening procedure for devices before they are subjected to Group-A electrical tests and Group-B mechanical and environmental tests where the rejection criteria are more stringent than for JAN types. Provisions for Group-C destructive tests on a sample basis are also made.

The following 100% tests (screens) are required by JAN-TX specifications. Each device is identified and the variable test data are recorded and maintained.

- High-Temperature Storage: 48 hours at 200°C
- Temperature Cycling: -55°C and +200°C
- Constant Acceleration: 20,000 G in Y_1 plane
- Hermetic-Seal Test
- Parameter Measurement and Data Recording
- Operating-Life Test 168 hours
- Parameter Measurement 10% Lot Jeopardy

Devices that fail catastrophically during screening are discarded, as are those which drift beyond specified operating limits.

5.3.6 Reliability-Assurance Testing

In reliability-assurance tests, samples of integrated circuits from production or customer lots are subjected to stresses -- usually greater than those stresses that the IC's will encounter in normal operation. Mechanical, environmental, and operating-life tests in a typical IC reliability-assurance test plan are shown in Table 5-2. It should be noted that many reliability-assurance test plans are destructive, that is, the circuits that undergo the tests cannot be used again. Further details on reliability-assurance testing are given in Section 6.3.

Reliability-assurance tests are costly because they frequently destroy the units tested and because they usually take a great deal of time to complete. The latter reason is also disruptive to delivery schedules, as well as being expensive. Costs of reliability-assurance plans could be decreased by standardizing the tests, permitting larger sample sizes and hence greater confidence in test results, and distributing the test costs across several customers.

The delivery-time element of a high-reliability procurement specification is often disregarded. Frequently, a reliability screen, followed by performance and assurance-test schemes, will be included in the procurement specifications.

TABLE 5-2

INTEGRATED-CIRCUIT RELIABILITY ASSURANCE PLAN

Examination or Test	Conditions		LTPD	Max. Acc. No.	Symbol	Limits		Unit
	Mil-Std-750 Method	Specific Conditions				Min.	Max.	
Subgroup 1								
Physical dimensions	2066	--	20	5	--	--	--	--
Subgroup 2								
Soldering heat	2031	1 cycle			--	--	--	--
Temperature cycling	1051	Test Cond.B,T(High)= 175°C			--	--	--	--
Thermal Shock (glass strain)	1056	--	20	5	--	--	--	--
Moisture resistance	1021	--			--	--	--	--
Endpoints: (Same as subgroup 7)								
Subgroup 3								
Shock	2016	5 Blows, X_1, Y_1, & Y_2, 1,5000 $\frac{1}{2}$msec (Total 15 Blows)			--	--	--	--
Vibration fatigue	2046	20G (nonoperating)	20	5	--	--	--	--
Vibration variable frequency	2056				--	--	--	--
Constant acceleration	2006	20,000G, X_1, Y_1, Y_2			--	--	--	--
Endpoint: (Same as subgroup 7)								
[1,2] Subgroup 4								
Terminal strength*	2036 Cond. E	3 leads at random	20	5	--	--	--	--
[1] Subgroup 5								
Salt atmosphere [3] (corrosion)	1041	--	20	5	--	--	--	--
Subgroup 6								
High temperature life(nonoperating)	1031	T_A = 175°C	λ=20	--	--	--	--	--
Endpoints: (Same as subgroup 7)								
Subgroup 7								
Steady state operation life	1026	T_A = 125°C[4] normal operating Bias applied 60 cps sine wave input	λ=20	--	--	--	--	--
Endpoints: D-C output levels as indicated on the individual data sheet.								

[1] Tests listed in these subgroups are considered destructive.

[2] At the conclusion of the testing in subgroup 4, the device shall be examined for evidence of mechanical damage.

[3] The device shall be examined for destructive corrosion and illegible marking.

[4] If the normal power dissipation requirements of the circuit do not elevate the junction above the maximum rated 175°C condition. Otherwise special considerations must be taken for each circuit.

*For Flat Package: Weight - 3 oz.

If the screen is at all non-standard, the testing processes cannot begin until after the order is received. This pushes the delivery date far into the future. Also, since the entire sequence of tests must be repeated if a lot fails, there is a good possibility that a promised delivery date could be missed by a wide margin. Table 5-3 outlines the interrelationships of reliability, cost, and delivery time.

TABLE 5-3

RELIABILITY/COST/DELIVERY TRADE-OFFS

Test	Test Level*	Additional Cost	Normal Delay In Delivery Time	Value For Reliability
Power Burn-In	100%†	Moderate	Moderate	High
Drift	Sample 100%	Moderate High	Moderate Long	Medium Medium
High-Temperature Storage	Sample 100%	Low Low	Short Short	Low Low
Environmental	Sample 100%	Low High	Short Long	Medium Medium
Hermetic Seal	Sample 100%	Low Low	Short Moderate	High High

*Sample = 30% or less. †Normally.

5.4 INTEGRATED CIRCUIT PROCUREMENT*

5.4.1 Developing the IC Specification

Many types of analog and digital IC's are described by commercial specifications. The system designer can frequently use these commercial devices as off-the-shelf items (i.e., without modification). However, in many programs the general trend has been toward in-house specifications, especially for a new product. In-house specifications can be simple, consisting, for example, of a burn-in requirement imposed on a commercial specification; or they may be complex, perhaps describing a custom-type circuit. If the designer cannot use one of the standard commercial integrated circuits, he must prepare a specification and submit it to the solid-state industry for price quotations. This procedure might involve the following steps:

(1) The circuit must be analyzed to determine if it can be integrated by standard techniques, and whether it is to be monolithic, thin-film, or hybrid. These decisions often require close coordination between the device user and manufacturer.

* Material in this section is abstracted from Motorola's Integrated Circuits Design Course (1964).

(2) The decision must be made whether to specify parameters in a black-box or specified-point manner. The parameters discussed thus far are considered the former type, since they describe the sum of the individual components in the IC device. In some IC applications, the characteristics of one or more of the individual components (transistors, resistors) are required by the device user, in which case the specification must include specified-point as well as black-box information. Generally, however, IC's requiring specified-point parameter descriptions do so only for a few components in the silicon chip. The black-box description is usually preferred by both vendor and user.

(3) The minimum number of tests needed to describe the IC adequately must be determined. For analog IC's, some parameter tests can be quite complicated, and should therefore be completely described in the specification, including test parameters, minimum and maximum limits, a schematic description of the test circuit, types of equipment to use, and waveform and component values.

(4) The specification should classify the various tests according to inspection level. These levels establish the number of IC's to be tested in various cases. Some tests (e.g., screens) must be imposed on every IC. Other tests may be concerned with noncritical but still important parameters, in which case testing can be conducted on a sampling basis. Careful consideration should be given to which tests are to be performed on a 100% inspection level, since testing adds substantially to the cost of IC devices.

(5) Schematic and logic diagrams (if applicable) should be included. The schematic should indicate only the basic function, not the parasitics.

(6) The package configuration should be presented, its dimensions having wide enough tolerances to facilitate the selection of alternate manufacturing sources.

5.4.2 IC Availability

The decision to use one device rather than another or to use custom design rather than off-the-shelf products will depend upon the designer's judgment and in most cases will be subjective -- especially when devices are being compared that have not yet been manufactured. This may often be the case because of rapid changes that are taking place in IC technology.

It is easier to be objective when electrical parameters are tabulated on a specification sheet than it is when predicted parameters are being compared. In trying to validate predicted performance, the engineer must rely on such subjective aspects as the manufacturer's reputation, past performance, and what he expects as a reasonable advance in the near future. The degree of subjectivity

can be reduced if a tabulation of important parameters relating to each possible IC is made. A tabulation of this type would usually compare only actual or predicted electrical parameters. Questions concerning past performance and the possibilities for the future must by their very nature be somewhat subjective; they are integrated into the final decision by the decision maker.

In the tabulation, it is possible to weigh each characteristic considered or assume that all characteristics are of equal importance. Usually one or two characteristics will be more important that the rest; and the ultimate weighing will be dependent upon the specific application and its constraints.

In any comparison of specified parameters, care must be used when devices made by different manufacturers are involved. Often characteristics that are ostensibly identical are defined differently or measured under different test conditions or test-circuit configurations.

5.4.3 IC Interchangeability

Although a study of commercial catalogs might lead the reader to believe that many available IC devices are interchangeable, this is not the case. A good general rule is to design with a single family of IC devices from a single manufacturer unless it is defintely known that another manufacturer is a compatible second source.

CUSTODIANS
Army - EL

Navy - EC
Air Force - 11

REVIEW ACTIVITIES
Army - EL, MI, MU
Navy - EC, SH
Air Force - 11, 17

USER ACTIVITIES
Army - ME
Navy -
Air Force -

FEDERAL CIVIL AGENCIES
NASA
FAA

PREPARING ACTIVITY
Army - EL

Project Number 5962-0009

BIBLIOGRAPHY: CHAPTER 5
SPECIFICATIONS AND PROCUREMENT

Blair, K. Getting More Value Out of An Integrated Operational Amplifier Data Sheet. Motorola Report AN-273.

Dertinger, E. F. "Status of Reliability Requirements in Government Contracts," Proc. 11th National Symposium on Reliability and Quality Control. Miami Beach, 1965.

Jones, L. "Specifying Semiconductor Reliability," Electro-Technology. 1966.

McLachland, H. F. "Using Specifications," IEEE Transactions on Reliability. 1966.

Ryerson, C. M. "Relative Costs of Different Reliability Screening Techniques," Proc. Symposium on Reliability. Washington, D.C., 1967.

EIA. Registration Data: Semiconductor Integrated Logic Gating Circuit. 1965.

EIA. Registration Data: Semiconductor Integrated Linear Amplifier. 1966.

EIA. Registration Data: Semiconductor Integrated Bistable Logic Circuits. 1967.

RADC. Quality and Reliability Assurance Procedures for Monolithic Microcircuits. RADC Spec. 2867A. 1967.

Zierdt, C. H., Jr. "Procurement Specification Techniques for High-Reliability Transistors," Proc. Symposium on Reliability, Washington, D.C., 1967.

RELIABILITY AND PHYSICS OF FAILURE

S I X

The reliability of any particular device is dependent upon the quality of materials and the manufacturing process, design, operational requirements, the type and level of testing (including screening), and the ability to perform failure analysis. Each of these facets affect ultimate reliability and are inseparable. However, it will be necessary to discuss each individually here.

Design, testing, and specification were discussed in previous chapters. This chapter deals primarily with physics-of-failure concepts and reliability. While each topic is important, the concepts and data presented in Chapter 4 on testing are particularly germane to failure analysis and reliability.

In fact, testing is in competition with failure analysis for reliability verification and prediction. Components and devices as received from the manufacturer represent a set with essentially two subsets, namely, those components or devices which will operate for a required time interval (reliable devices) and those which will fail within the time interval (unreliable). Massive testing has been and still is widely used to separate these subsets. The extent of testing depends upon the required reliability level. The philosophy, then, is to test reliability into the system.

The other approach for obtaining reliable systems is based on failure-analysis techniques -- an approach that attempts to eliminate any mechanism of failure and thus reduce subsequent failures. It includes analysis of failed devices, limited testing, and accelerated testing.

In theory, the failure-analysis approach appears to be the most attractive for microelectronics because of the difficulty of obtaining valid and sufficient test data. Reasons for these difficulties include the enormous number of microcircuits available, the relatively low failure rates, which require a large number of device testing hours to achieve a reasonable confidence; the variety of manufacturers and methods used to fabricate microcircuits; and the expense of performing such tests. Yet there are difficulties associated with relying solely upon failure analysis. These include the validity of accelerated testing, the timing and quantity of feedback data, competence and validity of analysis, and the inability of the manufacturer to control precisely every step in the manufacturing process (typically 300 to 500 distinct operations).

Some IC users believe that failure mechanisms are never permanently eliminated. When a particular failure mechanism is identified by the user and communicated to the device manufacturer, the manufacturer tightens up his processing to eliminate the mechanism. However, the mechanism gradually reappears as the pressure is eased.

Integrated-circuit reliability data are summarized in Section 6.4. They show a wide range in failure rates for different systems and different preconditioning. These data are intended to provide general information and indicate the range of failure rates currently being reported. The use of absolute numbers is meaningless without knowledge of the type and level of testing, the preconditioning allowed, and the extent, if any, to which the data were purged.

When such numbers obtained from this or similar tabulations or from the manufacturer are the best information that is available to the user, he should approach them with caution. Such numbers may be manipulated* at will, to provide seemingly concrete data to verify predicted device failure rates.

6.1 SYSTEM RELIABILITY IMPROVEMENT

The potential for improved system reliability offered by semiconductor integrated circuits has been a major factor in the widespread introduction of microelectronics into aerospace programs. The effect of the microelectronic device on interconnections, resistance to environment, manufacturing, test, human factors, and system design will contribute greatly to improved

*For example, see J.B. Brauer, "The Numbers Game - Who Wins?", IEEE/ASQC Reliability Symposium, Washington, D.C., 1967.

system reliability. Investigations by both Government and industry indicate that the following system improvements will be achieved through the use of microelectronics:

- Reduction in the number of individual parts and connections
- Improved intraconnection and interconnection techniques
- Increased resistance to environment
- Improved design
- Improved manufacturing processes
- Improved testing techniques
- Fewer circuit configurations
- Fewer types of environmental effects
- More effective use of redundancy techniques
- Lower power consumption (in many cases)

6.1.1 Reduction in the Number of Parts and Connections

The reduction in the number of individual parts, with the attendant reduction in the number of interconnections, contributes substantially to system reliability improvement.

The reliability of a conventional circuit (nonredundant) is predicted on the basis of the failure rates of its constituent parts. Generally, the lower the complexity level, the higher the reliability. For example, the failure rate of a particular NAND gate in discrete form is the sum of the failure rates of its 34 components (13 resistors, 7 capacitors, and 17 transistors) and 215 interconnections. A monolithic integrated circuit performing the same circuit function has one failure rate, which is lower than the sum of the transistor failure rates. Thus system reliability improves in proportion to the complexity of the discrete circuit being replaced by an integrated circuit.

6.1.2 Improved Connection Techniques

Connections will be considered herein as either intraconnections (conductors and connections within the integrated circuit or its discrete equivalent) or interconnections (conductors and connections that provide electrical continuity between the circuit and the outside world).

6.1.2.1 Intraconnections

The integrated circuit usually has less than half the number of intraconnections required by its discrete equivalent. In addition, the quality of the integrated-circuit intraconnections is superior to that of the discrete-circuit intraconnections. Two types of intraconnections must be considered: the highly reliable aluminum evaporation intraconnection and the intraconnection resulting from thermocompression bonding.

The interface resulting from the deposition of a conductor on a compatible surface under ideal conditions is virtually equivalent to a continuous conductor. A failure within such an interface is highly unlikely in an operating device if strict process control is maintained during the deposition phase (see Failure Modes and Mechanisms).

Unfortunately, the current state of the art limits thin-film technology (used for the aluminum evaporation) to a uniform, smooth surface. Therefore, leads from the silicon chip to the package leads must be connected by another method, usually thermal-compression bonding. While this process is far from being perfected, it has been in use in transistor fabrication for several years.

Thermocompression bonds have no known failure mechansims when compatible materials are being bonded in an ideal environment, a circumstance that does not appear practical at the present. However, industry is making a determined effort to minimize the failure mechanisms that occur in the absence of such favorable conditions.

6.1.2.2 Interconnections

The quality of interconnections will also improve with the increased use of microelectronics. The importance of interconnection reliability can be seen if it is assumed that each integrated circuit with 10 leads will involve about 100 interconnections. So that the interconnections will not experience more failures than the devices, each interconnection must exhibit a failure rate at least two orders of magnitude lower than that of an integrated circuit. In other words, there is a distinct possibility that interconnections, rather than microelectronic devices, will be the major cause of unreliability in future systems. The importance of interconnection reliability in microelectronic systems makes it essential that industry devote considerably more attention to methods for improving reliability in this area.

Because of the small size and large number of leads of the integrated circuit, it is frequently advisable to use a multilayer etched-circuitry interconnection technique. The multilayer approach can provide a high level of reliability if careful design and stringent process controls are exercised.

The conventional circuit may include evaporated, thermocompression, welded, crimped, and soldered connections -- involving a number of different materials with different thermal-expansion, solid-state stress-strength, chemical, and manufacturing characteristics. This variance in characteristics leads to uncertainties in design, and in lower reliability than experienced in microelectronic systems, which require fewer connection techniques.

6.1.3 Increased Resistance to Environment

The integrated circuit offers greater resistance to its environment, primarily because of its small mass and size. It is less susceptible to forces of acceleration, i.e., shock and vibration; shielding from radiation, heat, etc., is not as great a problem.* Hermetic sealing reduces corrosion.

*However, as discussed in Chapter 5, small-size and encapsulation present some particular problems also.

6.1.4 Improved Design

It is often said that most failures experienced in system operation were designed into the system. In other words, a significant portion of potential system unreliability can be avoided by careful design. If this is true, the effect of microelectronics on system design may well be the "dark horse" key to system reliability.

At the circuit level, the likelihood of improved design is obvious. Trained circuit designers are responsible for the design of each circuit. Since the economics of the circuit manufacturer depends on a large-volume market, he is willing to spend considerable time and effort proving the design of each circuit to protect his investment.

Not so obvious is the impact of the integrated circuit on system design and, consequently, on system reliability. Microelectronics in system design offers the following advantages, which directly or indirectly influence reliability:

(1) The design engineer has a wider job scope because of the availability of large building blocks: For the design of a given function, fewer engineers are required. This in turn reduces the communications problem, which should result in improved reliability. (An example of the effect of using larger building blocks is a ground checkout computer recently redesigned by an organization in the aerospace industry. The original, discrete version contained 26,000 parts; the microelectronic version of the same computer contained 3000 parts. If the microelectronic design were converted back to a discrete version on a part-for-part basis, the resulting discrete computer would contain about 12,000 parts -- a better-than-50% reduction from the number of parts in the original computer. Designers engaged in this project said that because two systems designers using microelectronics could replace the ten designers required by the original discrete approach, they were able to find functions that could be combined in a single circuit, drastically reduce interfacing, and condense other functions.)

(2) The designer has more time available to concentrate on potential reliability problems because design begins at the circuit level.

(3) Designers need to be familiar with the idiosyncrasies of a few different circuits as compared with 500 to 600 device characteristics, which should lower the probability of misapplication.

(4) Integrated circuits encourage the maximum use of digital design, which will usually improve system reliability by minimizing the number of out-of-tolerance failures.

6.1.5 Improved Manufacturing

From a manufacturing viewpoint, the reliability of a part is proportional to the continuity of production and the simplicity of the part's construction. The manufacture of an integrated circuit can be considered as two operations: pre-assembly and assembly. In the pre-assembly portion -- i.e., material preparation, masking, and diffusion -- the manufacturer, by necessity dependent on yield, must constantly improve and monitor his process control. Since the reliability of an integrated circuit depends on process control, the manufacturer is also obligated to improve and monitor reliability. In the assembly portion of the operation, i.e., when the silicon chip is mounted in a package and leads are made to the outside world, simplicity is the byword. As in the case of transistor assembly, automation or semi-automation will become commonplace.

In the manufacture of a discrete circuit, many suppliers have a role, but only the circuit designer is directly responsible to the customer. In the case of the integrated circuit, a single manufacturer is responsible for accepting raw material, processing the entire circuit, and assuring that the completed circuit performs in accordance with the specifications. Thus the single integrated-circuit manufacturer is better able to coordinate and control the factors that may influence circuit reliability.

It is generally acknowledged that process control is the key to reliability. Since the number of processes represented by an integrated circuit is far smaller than the number involved in a conventional circuit, a given degree of control on these relatively few processes results in a much higher effective control of processes and thus in higher reliability. When the different processes are totaled for a complete system, the relatively few processes required for microelectronics will have a very significant effect on the reliability.

6.1.6 Improved Testing Techniques

Since the integrated circuit is more complex than the transistor, more effort is expended on functional testing before the circuit leaves the manufacturer. While testing alone does not improve reliability, it does strengthen quality control, as evidenced in both increased failure-free operation and improved process controls.

Generally, qualification testing of integrated circuits is more stringent than that of their discrete counterparts; that is, the testing is concentrated at the circuit level rather than at the component level, making the test results more meaningful.

At the system manufacturer's facility, incoming inspection is simplified because the quantity and variety of parts to be tested is reduced. After the system prototype is completed, the "debugging" test is simplified if integrated circuits are used. (In one case, test time was reduced from six months to two weeks by microelectronic design.)

System performance evaluation by the user is also easier because of the greater simplicity of system design. Built-in test and fault-isolation techniques are more productive in microelectronic systems, resulting in higher mission reliability.

6.1.7 Fewer Circuit Configurations

The variety of circuit configurations employed in a single microelectronic system will be quite small compared with that of a discrete-component system. The causes of unreliability can therefore be more easily predicted and prevented. Other reliability influences enhanced by the reduced circuit variety include human factors and system design.

6.1.8 Fewer Types of Environmental Effects

Excessive drift is a major factor in the failures of discrete systems. Designing to a single network's drift characteristics is considerably less complex than designing to the many drift characteristics of various transistors, diodes, resistors, and capacitors made of many different materials -- as is necessary with discrete-component circuits.

The ease of matching characteristics such as temperature coefficients of integrated-circuit elements can also be used to reduce the effects of environmental changes.

6.1.9 More Effective Use of Redundancy Techniques

The reduction in weight, volume, and cost associated with integrated circuits makes the use of redundancy practical even at the circuit-element level. The application of various forms of redundancy, e.g., circuit or functional, majority voting logic, and adaptive techniques, has started only recently, since microelectronics has become an accepted technology. For a more comprehensive discussion on redundancy, see Section 3.8.

6.1.10 Lower Power Consumption

The low power requirements of the typical integrated circuit frequently allow minimum electrical stresses to be imposed on the microelectronic system. Since reliability is inversely proportional to stress, an increase in reliability is expected when the integrated circuits are used within specification.

6.2 FAILURE MODES AND MECHANISMS

The failure modes and mechanisms of integrated circuits are similar to those of transistors, but, as would be expected, the distribution of these modes and mechanisms is different. The failure-mode distribution of integrated circuits differs according to manufacturer, device type, and even lot.

The failure-mode distribution and prevalence of occurrence for two consecutive months as determined by a device user were:

Month A (127 IC Failures)		Month B (108 IC Failures)	
Devices mismarked	3%	Devices mismarked	40%
Electrically overstressed	18%	Electrically overstressed	24%
Bad bonds	37%	Bad bonds	18%
Good devices	14%	Good devices	6%
Other	28%	Other	12%

While this characterization of failures is so general as to seem useless, it does point out that many failures have little relationship to the materials or processes used in fabricating the silicon circuit itself. Many failures are associated with packaging the die (i.e., mismarking and bonding) and testing and handling the packaged device. Typically, bulk and surface defects associated with the silicon die represent less than 30 percent of all failures and may be less than 10 percent. Mechanical defects associated with bonding (chip and pads), handling, encapsulating, and marking are responsible for the preponderance of IC failures.

Table 6-1 is a compilation of defects that are typically found in integrated circuits. As used in the table, failure mode is an electrical or mechanical manifestation of a failure; it is the particular manner in which failure occurs and is independent of why failure occurred. Examples of failure modes are open metalizations or shorted lead wires. Failure mechanism is the fundamental physical or chemical process responsible for a particular failure mode. It is the explanation of the physical and chemical changes leading to a failure. The failure mechanism associated with open metalization might be a scratch.

The manufacturer should analyze the device design to determine critical failure areas and remove susceptibility to such failures. Each assumed failure point should be categorized as to its probable effect on performance.

TABLE 6-1

TYPICAL MICROCIRCUIT DEFECTS

Point at Which a Reliability-Influencing Variable is Introduced	Failure Mechanism	Failure Mode	Failure Detection Method
Slice Preparation	Dislocations and stacking faults	Degradation of junction chracteristics	Initial electrical test; operational-life tests
	Nonuniform resistivity	Unpredictable component values	Initial electrical test
	Irregular surface	Improper electrical performance and/or shorts, opens, etc.	Initial electrical test; operational-life tests
	Cracks, chips, scratches (general handling damage)	Opens, possible shorts in subsequent metalization	Initial electrical test; visual (pre-cap); thermal cycling
	Contamination	Degradation of junction characteristics	Visual (pre-cap); thermal cycling; high-temperature storage; reverse bias
Passivation	Cracks and pin holes	Electrical breakdown in oxide layer between metalization and substrate; shorts caused by faulty oxide diffusion mask	High-temperature storage; thermal cycling; high-voltage test; operating-life test; visual (pre-cap)
	Nonuniform thickness	Low breakdown and increased leakage in the oxide layer	High-temperature storage; thermal cycling; high-voltage test; operating-life test; visual (pre-cap)
Masking	Scratches, nicks, blemishes in the photo mask	Opens and/or shorts	Visual (pre-cap); initial electrical test
	Misalignment	Opens and/or shorts	Visual (pre-cap); initial electrical test
	Irregularities in photo-resist patterns (line widths, spaces, pinholes)	Performance degradation caused by parameter drift, opens, or shorts	Visual (pre-cap); initial electrical test
Etching	Improper removal of oxide	Opens and/or shorts or intermittents	Visual (pre-cap); initial electrical test; operational-life test
	Undercutting	Shorts and/or opens in metalization	Visual (pre-cap); initial electrical test
	Spotting (etch splash)	Potential shorts	Visual (pre-cap); thermal cycling; high-temperature storage; operational-life test
	Contamination (photo-resist, chemical residue)	Low breakdown; increased leakage	Visual (pre-cap); initial electrical test; thermal cycling; high-temperature storage; operational-life test; reverse bias
Diffusions	Improper control of doping profiles	Performance degradation resulting from unstable and faulty passive and active components	High-temperature storage; thermal cycling; operational-life test; initial electrical test

(continued)

TABLE 6-1 (continued)

Point at Which a Reliability-Influencing Variable is Introduced	Failure Mechanism	Failure Mode	Failure Detection Method
Metalization,	Scratched or smeared metalization (handling damage)	Opens, near opens, shorts, near shorts	Visual (pre-cap); thermal cycling; operational-life test
	Thin metalization due to insufficient deposition or oxide steps	Opens and/or high-resistance intraconnections	Initial electrical test; operational-life test; thermal cycling
	Oxide contamination; material incompatibility	Open metalization due to poor adhesion	High-temperature storage; thermal cycling; operational-life test
	Corrosion (chemical residue)	Opens in metalization	Visual (pre-cap); high-temperature storage; thermal cycling; operational-life test
	Misalignment and contaminated contact areas	High contact resistance or opens	Visual (pre-cap); initial electrical test; high-temperature storage; thermal cycling; operational-life test
	Improper alloying temperature or time	Open metalization, poor adhesion, or shorts	Initial electrical test; high-temperature storage; thermal cycling; operational-life tests
Die Separation	Improper die separation resulting in cracked or chipped dice	Opens and potential opens	Visual (pre-cap); thermal cycling; vibration; mechanical shock; thermal shock
Die Bonding	Voids between header and die	Performance degradation caused by overheating	X-ray; operational-life; acceleration, mechanical shock; vibration
	Overspreading and/or loose particles of eutectic solder	Shorts or intermittent shorts	Visual (pre-cap); X-ray; monitored vibration; monitored shock
	Poor die-to-header bond	Cracked or lifted die	Visual (pre-cap); acceleration; shock, vibration
	Material mismatch	Lifted or cracked die	Thermal cycling; high-temperature storage; acceleration
Wire Bonding	Overbonding and underbonding	Wire weakened and breaks or is intermittent; lifted bond; open	Acceleration; shock; vibration
	Material incompatibility or contaminated bonding pad	Lifted lead bond	Thermal cycling; high-temperature storage; acceleration, shock, vibration
	Plague formation	Open bonds	High-temperature storage: thermal cycling; acceleration shock, vibration
	Insufficient bonding-pad area or spacings	Opens or shorted bonds	Operational-life test; acceleration, shock, vibration; visual (pre-cap)
	Improper bonding procedure or control	Opens, shorts, or intermittent operation	Visual (pre-cap); initial electrical test; acceleration, shock, vibration

(continued)

TABLE 6-1 (continued)

Point at Which a Reliability-Influencing Variable is Introduced	Failure Mechanism	Failure Mode	Failure Detection Method
Wire Bonding (continued)	Improper bond alignment	Opens and/or shorts	Visual (pre-cap); initial electrical test
	Cracked or chipped die	Open	Visual (pre-cap); high-temperature storage; thermal cycling; acceleration, shock, vibration
	Excessive loops, sags, or lead length	Shorts to case, substrate, or other leads	Visual (pre-cap); X-ray; acceleration, shock, vibration
	Nicks, cuts, and abrasions on leads	Broken leads causing opens or shorts	Visual (pre-cap); acceleration, shock, vibration
	Unremoved pigtails	Shorts or intermittent shorts	Visual (pre-cap); acceleration, shock, vibration; X-ray
Final Seal	Poor hermetic seal	Performance degration; shorts or opens caused by chemical corrosion or moisture	Leak tests
	Incorrect atmosphere sealed in package	Performance degradation caused by inversion and channeling	Operational-life test; reverse bias; high-temperature storage, thermal cycling
	Broken or bent external leads	Open circuit	Visual; lead fatigue tests
	Cracks, voids in kovar-to-glass seals	Shorts and/or opens in the metalization caused by a leak	Leak test; electrical test; high-temperature storage; thermal cycling; high-voltage test
	Electrolytic growth of metals or metallic compounds across glass seals between leads or between leads and metal case	Intermittent shorts	Low-voltage test
	Loose conducting particles in package	Intermittent shorts	Acceleration; monitored vibration; X-ray; monitored shock
	Improper marking	Completely inoperative	Electrical test

6.2.1 *General Observations*

Cracked substrates are seen in relatively high numbers -- improper support is the most common cause. However, the bonding stress levels (time and pressure) may also cause such defects. Typically, the bonding stress levels are automatically controlled; if they are properly adjusted before each shift, defects caused by improper stress levels and operator error can be minimized.

Eutectic soldering may leave ridges that do not provide a proper support for the die. Such ridges may cause cracking during wire bonding. Oxide films at the solder interfaces also affect die-bond integrity.

Many mask flaws are seen, especially in the epitaxial region around the emitter. If the emitter region is misaligned, the metalization window of the base will partly cover the emitter region. Misalignment may also cause shorts when wires are bonded to the pad.

Oxide-layer defects are not as prevalent as they were a few years ago. Some of the improvement is no doubt due to the use of double photoresist.

Accidental etchant spotting is seen frequently. This results in a hole or near-hole in the oxide. Such defects are usually time/voltage dependent and result in an eventual short if a metalized area lies over the fault. Failure mechanisms of this type are difficult to detect during testing and affect the long-term reliability of the device.

Channeling is still seen but not as frequently as in the past. It is dependent upon the metalization path but is essentially an oxide problem. Closely related to this failure mechanism is the unwanted solid inclusion of silicon carbide on the surface or forming a part of the surface. If such inclusions are near a junction, the result is a soft junction. If a metalization covers the inclusion, a short circuit is possible.

Surprisingly, incomplete oxide removal is uncommon. When it is seen, it is usually manifested as a high-resistance ohmic contact.

Oxide steps are a problem; as IC's become more complex (requiring many more oxide steps on the same die), such failure mechanisms may be critical. These steps cause material transport because of the thinness of the metalization at the step. This is seen quite commonly. Unlike tool marks that cause failure immediately or have no serious effect, oxide-step failures are time dependent. The steeper the step, the more serious the problem. Therefore, ohmic contact steps are the most vulnerable. A thicker metalization is not necessarily the solution to such a problem because it introduces additional problems such as a higher susceptibility to plague and peeling. A metalization thickness between 6000 and 7000 Å seems optimum.

Misalignment of metalizations is common. Such defects also increase the susceptibility to inversion because the metalization is not in an optimum place. Poor adhesion of metalization is still seen but not as frequently as it was a few years ago. Poor adhesion is usually due to poor cleaning of oxide surfaces before aluminum deposition or other metalization.

Although plague problems still occur, their lower incidence recently is due in large part to an understanding of the mechanisms that cause plague. Improvements in process control affecting bonding temperature and pressure, lidding temperature, and stabilization bake temperature have reduced the occurrence of these defects.

Particles can be introduced into the device any time before lidding. One source of particles is the solder used for die bonding. When the die is attached to the substrate with a eutetic solder, an oxide may form that affects bond integrity. To improve these bonds the die is scrubbed against the header in the pool of solder. Excess solder will ball up around the edge of the die; each particle so produced is a potential failure hazard. Lidding boats may also introduce particles, especially if they are made of graphite.

Lead dress is a serious and common reliability problem and is almost entirely operator dependent. Defects such as sagging, shorts, or poor cutting are seen frequently. Proper location of the bonding pads is important to proper lead dress. Filling the inside of the package to eliminate such problems is not ideal, because such material introduces other problems. X-ray is an effective method of examining lead dress after encapsulation, but some materials (e.g., aluminum) are transparent to X-rays.

Hermeticity failures are still common. Lead bonding and package handling can introduce fractures in the glass seals. If package leads have not been properly heat-treated, the lead may break off from the package. In all-aluminum systems, the leads are gold plated and unless care is exercised in processing, plague defects can be induced.

Failures resulting from metallurgical defects constitute the largest segment of integrated-circuit reliability problems. Such defects include those caused by plague, aluminum intraconnection problems, overbonding, and underbonding.

6.2.1.1 Plague

Plague is a term used to describe a time-dependent formation of a chemical compound at semiconductor-metal or metal-metal contacts that increases contact resistance and weakens bonds. It is one of the most significant integrated-circuit failure mechanisms. Plague has been the subject of many detailed physics-

of-failure studies (not all in agreement) that are beyond the scope of this chapter. The following types of plague are most commonly encountered:

(1) Purple Plague. The time/temperature formation of the gold-aluminum $AuAl_2$. (Some studies argue that purple plague is not created during extended bakes at 200°C or less.)

(2) Black Plague. The time-temperature formation of the ternary Au-Si-Al compound formed at about 300°C. The silicon apparently acts as a catalyst. Figure 6-1 shows a device which failed because of black plague, causing opens in the narrow regions of the interconnect pattern near the ball bonds.

(3) White Plague. Aluminum hydroxide.

(4) Silver Plague. Tin migration along the bond wire.

(5) Periphery Plague. A situation in which a small amount of one of the plagues described above is present and the aluminum pad area is too small. Normally neither condition alone would cause a failure, but together they are likely to create an open. It is quite apparent that the bonding pads in Figure 6-2 are too small to accommodate the large ball bond, an example of periphery plague.

Plague is generally the result of a poorly controlled bonding procedure and is common to the transistor. It often can be screened either by visual inspection or by mechanical stress -- e.g., by a high-linear-acceleration test after the capping.

6.2.1.2 Aluminum Intraconnection Problems

Aluminum intraconnection problems include the following:

(1) Poor Adherence. The result of applying aluminum metalization to a surface that has not been properly cleaned to permit good alloying of

COURTESY OF UNIVAC

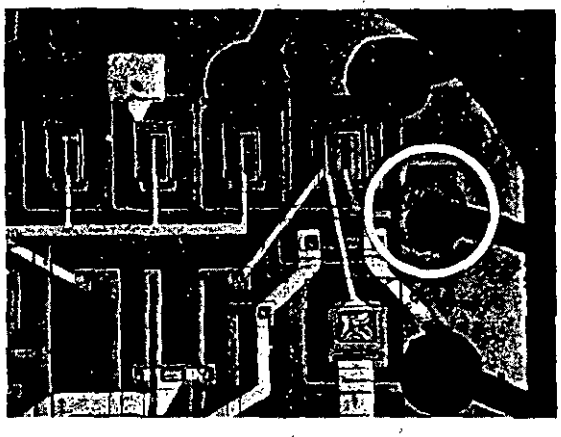


FIGURE 6-1
OPEN CIRCUIT CAUSED BY BLACK PLAGUE

COURTESY OF UNIVAC

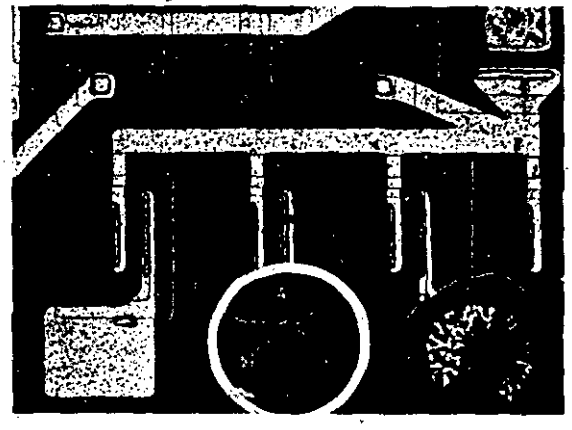


FIGURE 6-2
OPEN CIRCUIT CAUSED BY PERIPHERY PLAGUE

the metalization to the silicon oxide. This failure mechanism is triggered by mechanical stress such as pulling, probing, or thermal shock.

(2) Overetched Metalization. The result of etching the metalized intraconnection to the extent that the conducting path is too narrow to carry required currents.

(3) Electrolysis. The result of the inclusion of ionic materials on the surface of a device and the effect these contaminants have on metalization when a potential is applied.

(4) Melted or Vaporized Metalization. Excessive currents can cause the metalized intraconnection to be vaporized or melted, particularly where the cross-sectional area is reduced. This reduction generally occurs where the metalization crosses an oxide cut.

(5) Insufficient Aluminum Metalization - Insufficient thickness of the deposited aluminum conductor over the passivation oxide steps can make it difficult to achieve adequate intraconnections. These steps result from the windows that are etched through the oxide for diffusions and contacts. The height of these steps also vary depending on the number of diffusion and oxidation processes. Thin plating across the steps may result in an open or a high resistance because of high current densities in the thin layer of metal. This problem can be aggravated when elevated temperature combines with the heat of the high-current density.

Metalized leads must always be considerably thicker than the oxide step height to assure continuity across the step.

Even when the metalization is considerably thicker than the step height, thin cross-overs occur if the step is very steep and the metal deposition is directional. This problem can be minimized by the use of multifilament metal-evaporation systems during deposition to assure an adequate nondirectional buildup of the deposited metal. Figure 6-3 illustrates the type of directional metalization that results in the failure modes discussed above. Figure 6-4 is an example of the phenomenon.

(6) Other opens in Al Intraconnections: (a) Hydrated alumina (Al_2O_3) that has formed at dissimilar-metal contacts in the presence of excessive moisture (may be formed at room temperatures but is accelerated by power operation or baking); (b) hydrated alumina caused by faulty wash and dry techniques; and (c) aluminum corrosion at scratches.

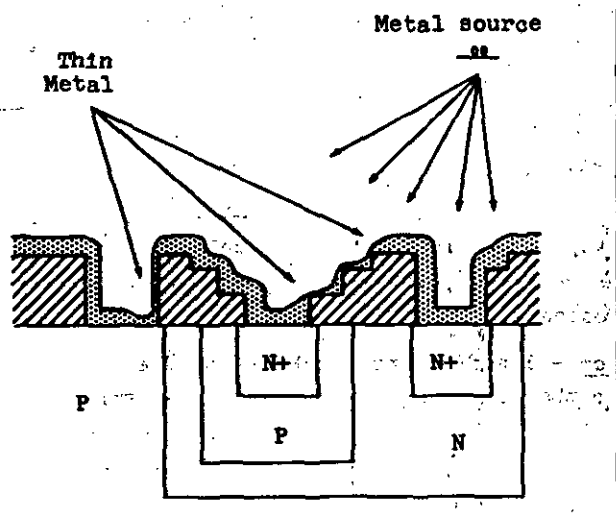


FIGURE 6-3

THIN METALIZATION CAUSED BY DIRECTIONAL DEPOSITION

COURTESY OF UNIVAC

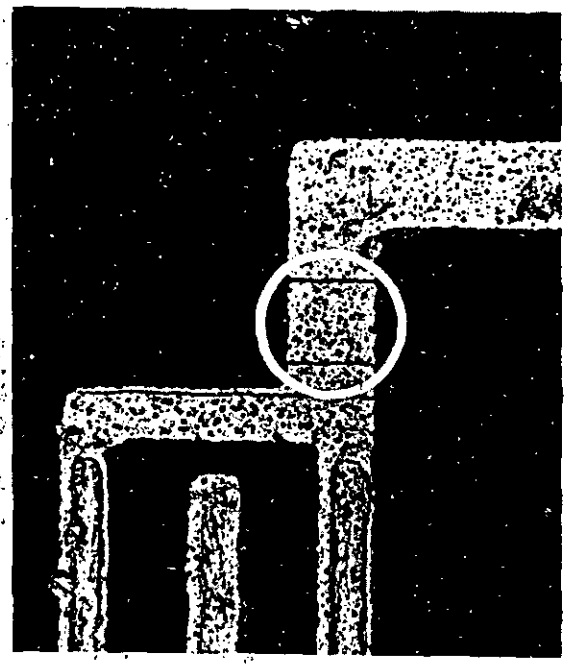


FIGURE 6-4

THIN METALIZATION ON INTEGRATED CIRCUIT

6.2.1.3 Overbonding

Overbonding, resulting from excessive temperature or pressure, or both, is a non-time-dependent failure mechanism. Under such conditions, SiO_2 pulls up, causing the aluminum to peel or lift, exposing raw silicon. A crack in the substrate, resulting in a short circuit, may also be caused by excessive pressure.

6.2.1.4 Underbonding

Underbonding, the result of insufficient temperature or pressure, or both, during the bonding process, is another non-time-dependent failure mode that results in an inadequate metallurgical bond between the gold and the aluminum. This failure mode can be triggered by mechanical stressing.

6.2.2 Surface Effects

Surface effects also contribute significantly to integrated-circuit unreliability. Failure mechanisms in this category include the following: pinholes and other shorts through the SiO_2, corrosive etching, insulation layers, ionic surface effects, loss of hermetic integrity, and effects of environmental stresses.

6.2.2.1 Pinholes

One of the major surface-instability problems results from flaws in the passivation caused by deficiencies in the photoresist process. Breakdown or the inclusion of foreign particles in photoresist masks inhibits the proper doping of the silicon wafers, with the result that columns of improperly doped materials may appear in a critical area of the integrated circuit. This type of flaw appears as a pinhole in the surface of the finished device when examined under dark-field illumination; it permits shorting between interconnecting films and other portions of the circuit. In the case of aluminum conductors, the high-density current resulting from the short circuit causes the aluminum to melt. This problem can be minimized by the use of additional or more rigid process controls for mask inspection, mask usage, dark-room dust, and contamination.

6.2.2.2 Other Shorts Through the SiO_2

Other shorts from the aluminum metalization to the silicon through the silicon-dioxide are caused by entrapped impurities in the SiO_2 or by the coating of the SiO_2 with substances of poor dielectric strength. These are voltage-dependent failure modes.

Pinholes and the other shorts mentioned above generally can be detected during parameter testing. A particular test for pinholes is the application of over-voltage to the various diodes while current is stringently limited. Unfortunately, complete shorting through the oxide does not always result. Partial oxide damage is often not detectable (except possibly by infrared techniques); consequently, this failure mode can be classified as time-dependent.

6.2.2.3 Corrosive Etching

Surface instability can result from the residue of corrosive contaminants. Integrated-circuit manufacturing techniques require the use of several extremely active etchants. Failure to completely remove these etchants can result in open aluminum conductors, as illustrated in Figure 6-5. This time-dependent failure mechanism is, of course, the result of inadequate process control.

COURTESY OF UNIVAC

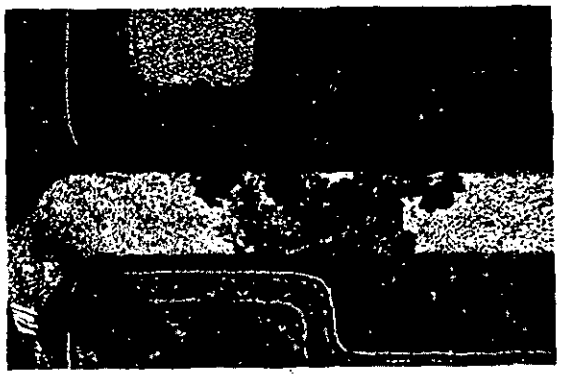


FIGURE 6-5

OPEN CIRCUIT CAUSED BY CORROSION

6.2.2.4 Insulation Layers

The formation of electrical insulation (dielectric) layers between the aluminum film and the silicon causes an interface at the window in the oxide. This is the result of

faulty oxide removal at the windows. Failures or incipient failures of this type are most rapidly detected after baking or operation. The dielectric layer can be broken through by microprobe pressure or voltage, temporarily healing the device; however, the unwanted insulation can be recreated by baking. Therefore, this failure mode is considered time-dependent.

6.2.2.5 Surface Effects

Historically, surface effects have been one of the major limiting factors in the reliability of integrated circuits. Device failures such as low-voltage breakdowns, high leakage currents, instability, and low gain can all result from surface effects. Although improvements in surface passivation techniques have greatly reduced these problems in the last few years, some still arise because of ionic effects associated with the passivating layer.

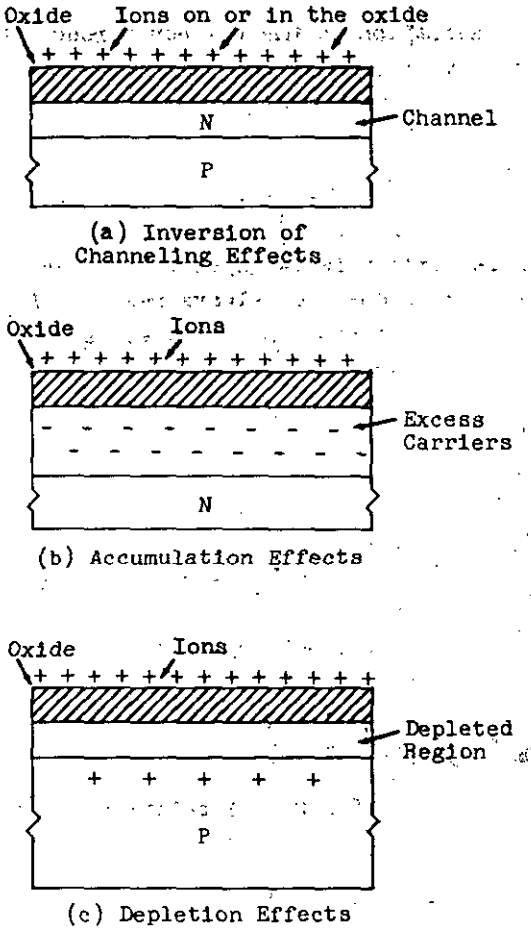


FIGURE 6-6

EXAMPLES OF IONIC SURFACE EFFECTS

Positive and negative impurity ions could exist (at the silicon/silicon-dioxide interface, in the silicon dioxide, or on the silicon dioxide surface), become mobile, and migrate to selected areas to form any of three phenonmena (Figure 6-6): inversion or channeling, excessive accumulation of impurities on similarly doped material, and depletion of normal impurities. Mobility and direction of these ions is dependent on any or all of the following: concentration of ions, resistance of the silicon, the applied field, and ambient conditions outside the oxide.

These effects are time and temperature dependent, but they are often difficult to detect. Some effects will not appear until after several hundred hours of operating life and others only after severe stress-testing is performed. In some instances these effects will not appear in normal environments, but will appear when subjected to nuclear radiation.

One interesting fact about ionic surface effects is that they are usually reversable. Some devices will

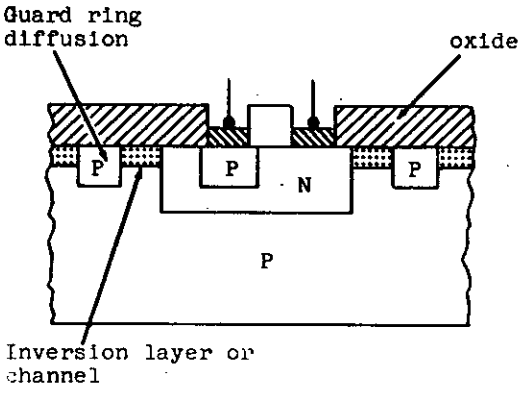


FIGURE 6-7
PNP TRANSISTOR WITH GUARD RING

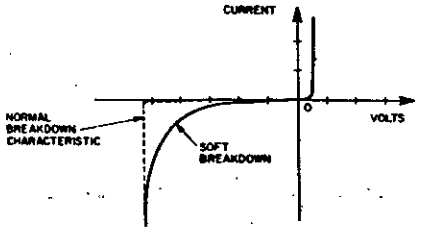


FIGURE 6-8
ILLUSTRATION OF A SOFT BREAKDOWN CHARACTERISTIC

recover in a few minutes by heating to approximately 200°C without bias. This temperature provides enough thermal energy to redistribute the ions to form a neutral surface. Others require the addition of external fields to produce neutral surfaces. A classic example of a corrected surface effect under ideal moisture and voltage conditions was the famous correction of a Telstar transistor malfunction. The Telstar malfunction resulted from exposure to nuclear radiation while in orbit.

Some manufacturers use a "guard ring" diffusion (Figure 6-7) to limit the effect of channeling. The high-concentration diffused ring encircles the device, thereby preventing the channel from spreading.

These time-dependent failure mechanisms can be accelerated at elevated temperatures and high-temperature operational-life testing. They can be minimized by proper cleaning and washing steps, obtaining the purest and and densest passivating layer possible, and by carefully controlling the gas inside the package (the presence of water inside the package is of particular concern).

6.2.2.6 Loss of Hermetic Integrity

Moisture in the package can cause shorts at surface scratches or in the surface passivation where the current densities in the conduction pattern are high (hot spots). Contamination on the top of the silicon can cause high leakage or soft breakdown characteristics (see Figure 6-8). Sources of this contamination might be moisture; weld gasses; gas desorption from other parts; and mobile contaminants migrating to the junction area from within the sealed package, or from outside if the seal is broken.

6.2.2.7 Effects of Environmental Stresses

Ionic contamination, neutron radiation, and large local mechanical stresses can result in time-dependent failure. The first mechanism was discussed under Ionic Surface Effects; the latter mechanisms will be discussed under Bulk Defects.

6.2.3 Mechanical Imperfections

Mechanical imperfections, usually the result of workmanship errors, are a major source of failure in integrated circuits. Often an incipient failure resulting from one of these imperfections can be detected through careful inspection by the device manufacturer. Failures in this category result from improper scribing, improper handling of tools, bonding and lead dress problems, internal lead wire problems, unwanted residues, improper die attach, and other quality defects.

6.2.3.1 Improper Scribing

Three relatively minor problems can arise in the scribing process used to separate the individual dice from the parent wafer. The scribing can be misregistered, causing the crystal to be severed too closely to the circuit-lead bonds. The bond then overlaps the edge of the chip and may eventually short to the substrate.

Improper scribing can cause severe edge chipping of the oxide, exposing the substrate and permitting shorting (Figure 6-9). Chipping is a function of lattice orientation with respect to the scribing. Lattice orientation is, in turn, a function of the desired device characteristics. The problem can be circumvented if sufficient margin is left for expected chipping. Cracking may also be initiated by the scribing process (Figure 6-10). The cracks may be propagated by environmental factors, resulting in a time-dependent failure.

6.2.3.2 Improper Handling of Tools

A careless smear from a tool can remove a portion of the aluminum metalization (Figure 6-11) and create a situation similar to that of the insufficient metalization deposition, previously discussed. Cases in which smeared aluminum has bridged various circuit elements (Figure 6-12) have also been reported.

Scratches on the aluminum-metalization conduction pattern often cause opens. The scratch reduces the cross-section of the interconnection, resulting in electrical overheating and a subsequent open. If the scratch is not detected as

COURTESY OF UNIVAC

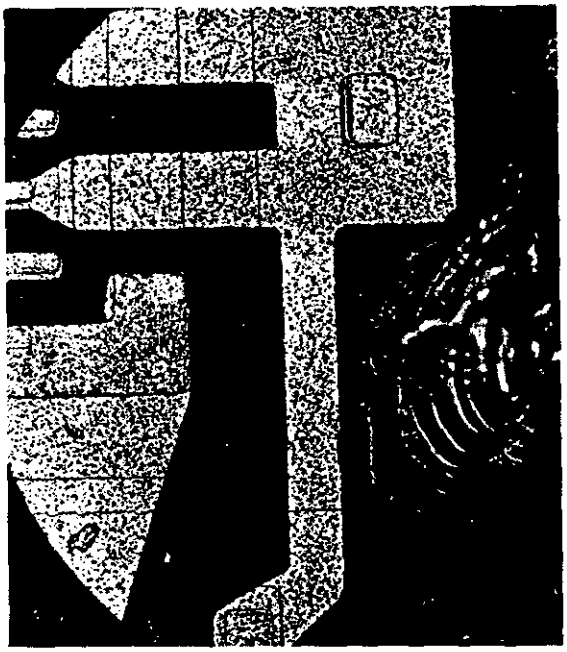


FIGURE 6-9

EXCESSIVE CHIPPING CAUSED BY IMPROPER SCRIBING

COURTESY OF UNIVAC

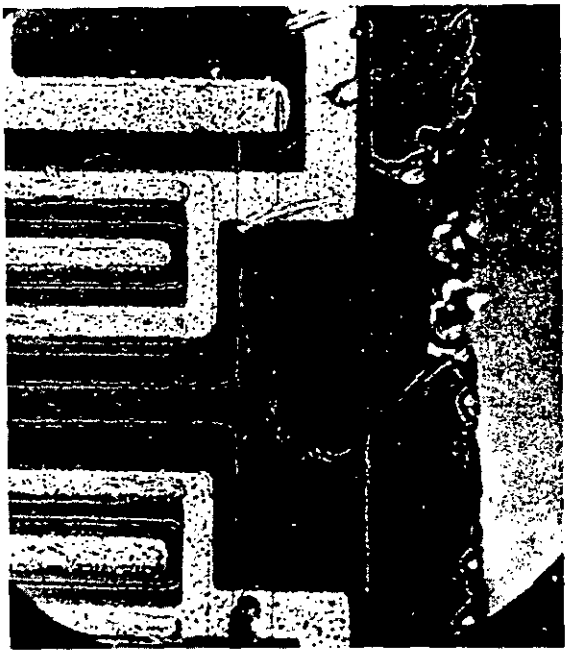


FIGURE 6-10

CRACKING CAUSED IMPROPER SCRIBING

COURTESY OF UNIVAC

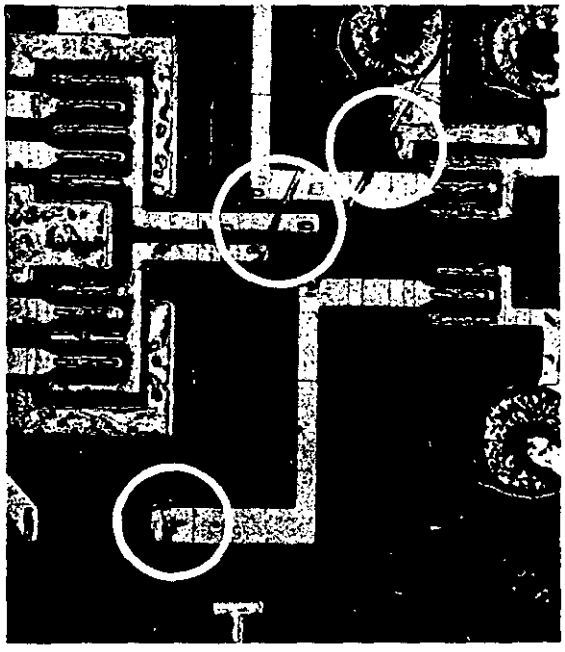


FIGURE 6-11

DAMAGED METALIZATION CAUSED BY HANDLING

COURTESY OF UNIVAC

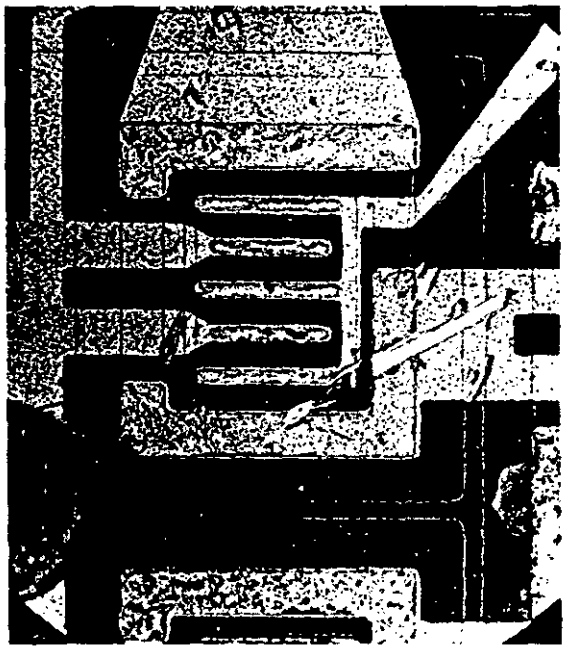


FIGURE 6-12

SHORT CIRCUIT CAUSED BY SMEARED ALUMINUM

an open before the circuit is used in a system, the circuit's tolerance to overstress pulses is markedly decreased; hence the probability of a latent failure is significant.

A scratch on the chip surface can also cause a base-to-base type of shorting that would permit the integrated circuit to pass most operational tests as well as operate in a ring-oscillator circuit (series life test), and yet not perform properly in a computer circuit, where logic operations are required.

Because the ratio of the aluminized area to total crystal area is small in the case of the transistor, it is almost safe to handle individual transistor chips during their manufacture with well designed forceps. However, the high density of circuit elements on integrated circuits makes it clear that such liberties can no longer be tolerated. Still, a large proportion of failed circuits that are opened show evidence of some tool damage.

There is some tendency to overlook tool damage, apparently based on the feeling that the true problem units have opens or shorts that are identificable at inspection and therefore are not installed in hardware. Experience indicates, however, that this assumption is over-simplified; often tool marks appear to be harmless, only to result in a failure in system operation. Figure 6-13 is a photograph of a device that has three tool marks on the deposited aluminum lead. At 1500X magnification (Figure 6-14) it is not easy to discern just where the path finally opened, but there are two points in the scuffed area where the path that remained after the accident is tenuous.

COURTESY OF UNIVAC

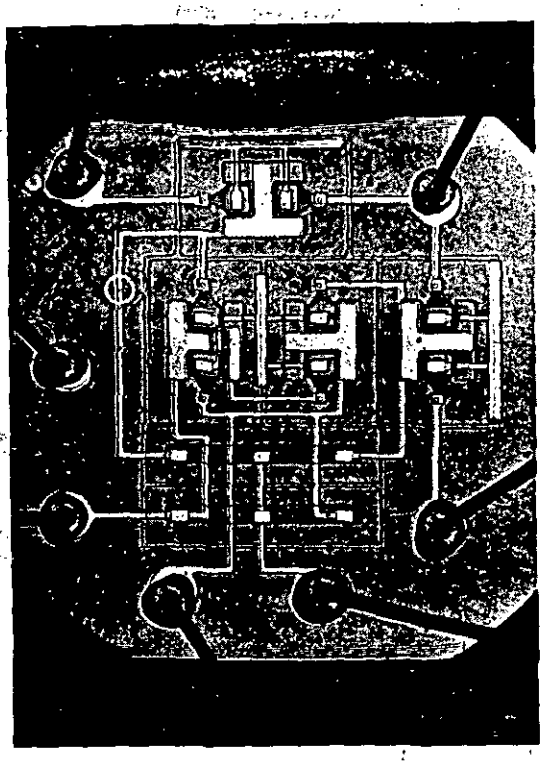


FIGURE 6-13

DAMAGED METALIZATION THAT MIGHT PASS TESTS

6.2.3.3 Bonding and Lead Dress Problems

Improper thermocompression bonding can induce cracks in the silicon substrate under nail-head bonds, which may be further propagated by the appliation of thermal stress. Also, the die bonding operation can intercede to accelerate a crack that would otherwise have been harmless. This is especially true if the solder preform used to hold the substrate to its associated header is distributed unevenly or fails to wet well to either the substrate or the header.

COURTESY OF UNIVAC

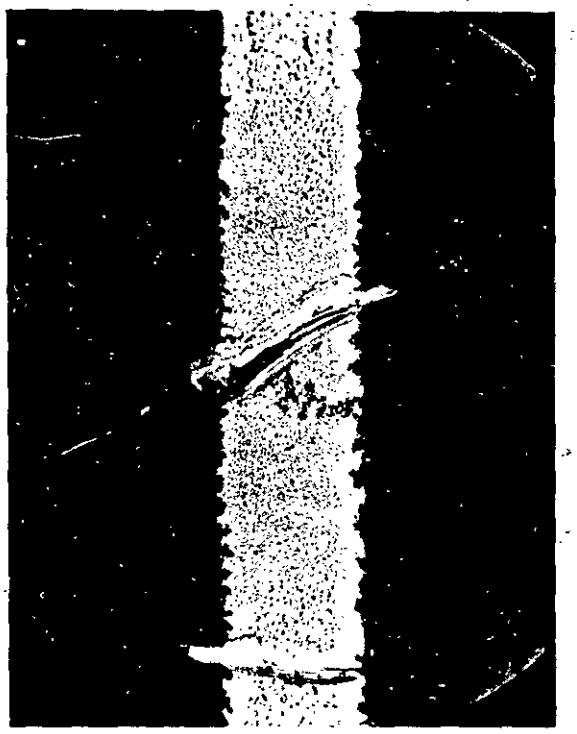


FIGURE 6-14

MAGNIFIED (1500X) VIEW OF AREA FROM FIGURE 6-13, SHOWING SEVERE DAMAGE

COURTESY OF UNIVAC

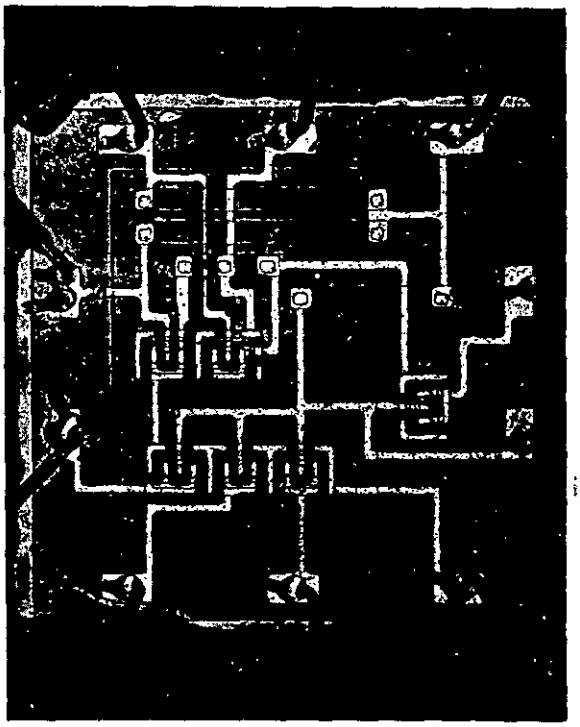


FIGURE 6-15

CRACKED CRYSTAL THAT PASSED ALL ACCEPTANCE TESTS

Figure 6-15 is a photograph of a device that passed the vendor's functional test and the system manufacturer's incoming inspection but after some handling and short operation in a laboratory experiment failed catastrophically because of a progressive crack.

In some situations it is necessary to fold a stitch-bonded lead back over itself, which may result in a severed lead or loosened bond.

In the case of ball bonds, it is very important to have the proper relationship between the ball size and the pad size as well as to apply the correct pressure to avoid cracking the passivation and causing a circuit short. Another type of failure is also attributed to ball bonds. The formation of oxide before the bond is made, or the presence of other types of contamination on the passivation layer, will cause an open circuit at the bond.

There have been several cases in which the pigtails of the ball-bond were shorted to the header.

Overheating during the thermocompression bonding process can cause embrittlement of the bonding wire (see Internal Lead-Wire Problems).

Some of the above-described problems can be classified as workmanship errors; others involve registration or the centering of leads in the glass holes of the header.

Poor pattern layout that requires bonding too close to the oxide edge on the surface of the chip causes failures. Shorts can occur when the gold overlaps the edge and goes across the SiO_2 to the silicon or shorts through the SiO_2, which tends to be thin near the edge of the substrate, thereby presenting a decreased dielectric. If a failure does not appear immediately, added strains on the device may rupture the dielectric.

COURTESY OF UNIVAC

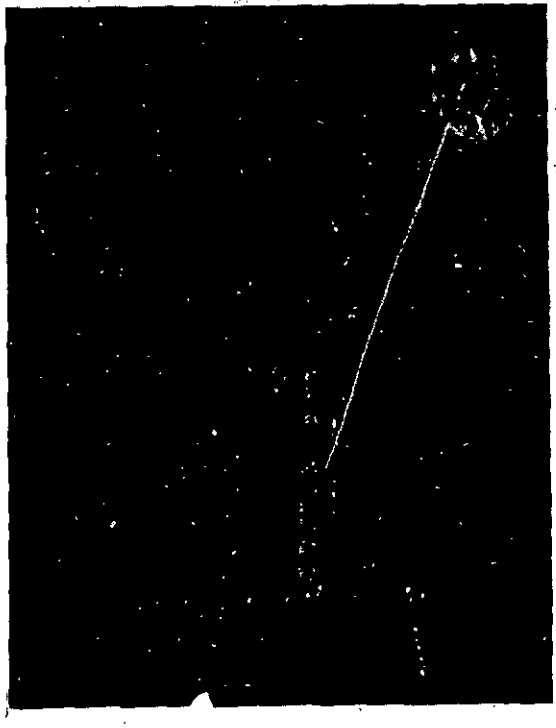


FIGURE 6-16

DAMAGE RESULTING FROM EXCESSIVE LEAD TENSION

6.2.3.4 Internal Lead-Wire Problems

Too little or too much tension on the internal leads can cause failure when the integrated circuit is operated in an environment of vibration (Figure 6-16). The probability of the leads shorting is increased when the package post is at a lower level than the surface of the chip, which is the case in many flat-packs. Excessively long lead wires may short to the lid, to the bottom of the package, to the surface of the chip, or to one another. Close proximity of leads results from either poor pattern layout or improper mounting in a package.

Overheating during thermocompression bonding can cause embrittlement of the bonding wire adjacent to the bond. Subsequent mechanical or thermal stress thus can more readily cause such wires to break. Aluminum wire seems to be slightly more susceptible to this difficulty than gold. Several of the open-lead field failures have been attributed to this problem.

Opens can also occur in lead wires as the result of faulty handling, bending, and processing. Nicks and cuts may break during mechanical stressing. If bending causes extreme thinning of the wire or if the wire is constricted, it is susceptible to breaking open under minor stress. It is particularly noticeable at the post, where wedge bonds are made.

The above-described failure mechanisms can usually be accelerated and failures detected if the devices are subjected to the centrifuge test.

6.2.3.5 Unwanted Residues

Unwanted residues such as metallic fragments, etching residues, etc., can cause failures.

Loose materials and particles have been identified as the cause of failure in some integrated circuits. Broken pigtails and package tabs are two common examples of free conducting materials that have caused intermittent shorts. The X-ray of Figure 6-17 shows a relatively large foreign object inside the flat-pack of an integrated circuit. A photograph of the device (Figure 6-18) made after the package was opened shows the object to be a scrap of metal apparently chipped or torn from an external tab connector. Nonconducting material can cause mechanical damage.

COURTESY OF UNIVAC

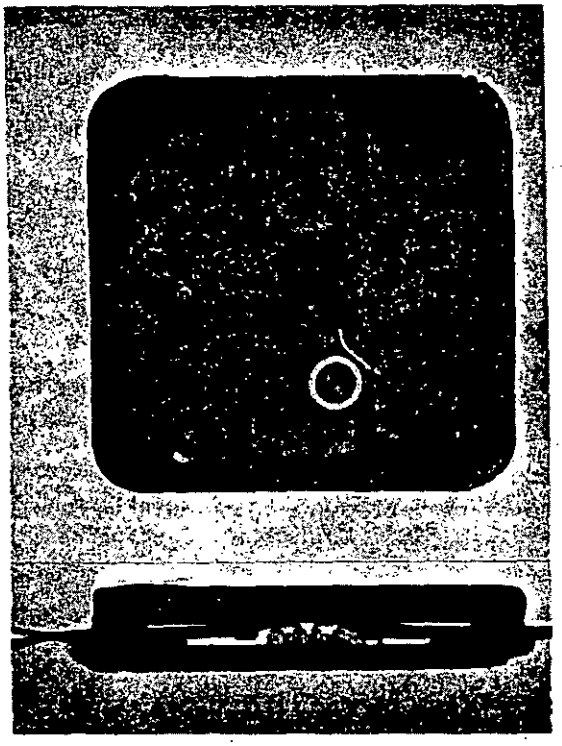


FIGURE 6-17

X-RAY VIEWS SHOWING LARGE FOREIGN OBJECT IN PACKAGE

COURTESY OF UNIVAC

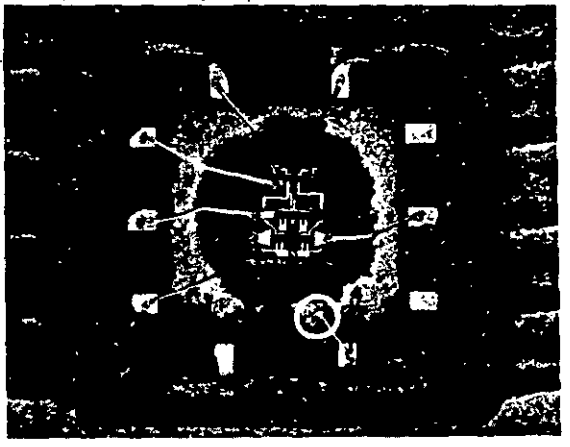


FIGURE 6-18

CIRCUIT OF FIGURE 6-17 AFTER OPENING

An impurity in the substrate can create a small diode. If the diode is within the area of an active circuit element, the circuit fails.

It might logically be assumed that since integrated circuits are relatively expensive and require great care in production, adequate measures are taken to protect them from penetration by impurities. Yet it is not uncommon to see gold

and glass particles adhering to the substrate surfaces, and larger items are occasionally found.

6.2.3.6 Improper Die Attach

Strains set up during die attach, particularly when a silicon-gold eutectic has not been reached, may result in cracking of the silicon die. Excessive pressure on the substrate side of a flat pack, or thermal propagation of minute surface cracks of a silicon die mounted in any package, can induce total cleavage.

6.2.3.7 Electrical Overstressing and Improper Marking

Electrical overstressing by the test equipment and improper marking of the completed package have also caused some difficulties, in spite of numerous 100% inspections.

6.2.3.8 Other Mechanical Problems

Other mechanical problems, such as misregistration (poorly defined geometry, creating hot spots), masking flaws, packaging solder residue, package leakage, insufficient lead plating, and photolithographic-process deviations also cause failures of integrated circuits.

6.2.4 Bulk Defects

Bulk defects are responsible for a small percentage of the operational failures of integrated circuits in systems, and will be treated only briefly here. Failure mechanisms that fall in this category include dislocations (crystal-lattice anomalies), impurity diffusions and precipitations, and resistivity gradients resulting from mechanical, nuclear, or thermal stresses. These defects can lead to diffusion spikes, which in turn cause hot spots, voltage breakdown, and other deviations from the desired electrical characteristics. Such defects are usually induced in the crystal-preparation process.

The steep concentration gradients found in epitaxial diffusion result in crystal-lattice strain. This strain is subsequently released by the formation of dislocation structures that contain edge components perpendicular to the concentration gradient. The chip is structurally weaker at the dislocation fault plane; thus bulk failure can be triggered by mechanical stress.

Deviations in epitaxial growth, resulting in impurity diffusion, are another source of bulk failures. Impurity diffusion is more likely along edge dislocations, particularly along the arrays of edge dislocations that form small-angle grain boundaries. The precipitation of impurities at the resulting crystal-lattice-orientation fault planes is believed to lower the reverse breakdown voltage.

Resistivity gradients caused by a heat differential between the center and the outer surface of the chip can result in secondary breakdown.

Large local stresses can cause changes in resistivity and, hence, in electrical characteristics. These local stresses can be caused by mechanical shock or vibration, which would generally result in microphonics. The stress levels usually would have to be so high as to be destructive in other areas (seals, bonds, etc.). Sufficiently large stress concentrations can crack the die.

Thermal shocks resulting from normal processing steps can cause cracks in the silicon slices or oxides. These defects are normally the result of too thick an oxide layer or too fast a cooling rate. Although most of these faults are detected by visual inspection, some are not detected until further mechanical and electrical tests are performed.

Secondary breakdown that leads to localized alloying or diffusion in semiconductor junctions depends critically on collector voltage and mode of operation. In some transistor structures, the breakdown current does not flow through the hot-spot regions, indicating that a different mechanism, associated with the presence of local defects, exists and manifests itself in a "soft" collector-junction characteristic.

The failure mode usually associated with bulk defects is bulk shorting, which results from secondary breakdown or uncontrolled PNPN switching when the circuit design is such as to allow floating internal junctions.

There are punch-through effects between collector and emitter if the base region is narrow. In this case the collector-depletion region extends to the emitter at high collector voltages. This occurs most frequently in high-frequency transistors, since these devices have a minimum distance between emitter and collector to reduce recombination effects. This condition also occurs in transistors that have high base resistivity ($>300\Omega/\square$) and relatively narrow base regions.

Carriers entering and leaving a region through floating junctions are subject to multiplication within the junction. This effect lowers the breakdown level of the junction and leads to negative-resistance regions. This action is enhanced at higher currents and is evidenced in the isolation junctions of integrated circuits.

6.2.5 Failure-Analysis Tools and Procedures*

One procedure for failure analysis is shown graphically below. As indicated earlier, devices that are actually good may be identified as failures. Thus the first step is usually to verify that a device is actually defective. Once the device has been verified to be defective, it is processed through a logical series of failure-analysis steps. The steps are ordered such that all required data are obtained and recorded before the next step, which would destroy the potential for obtaining desirable information.

Instrumentation is important to proper failure analysis. However, imaginative analysts who are capable of innovation are equally important. Outfitting a failure-analysis laboratory with the most modern analysis instruments does not assure high-quality failure analysis.

Table 6-2 is a tabulation of instruments commonly used for failure analysis. Associated with each instrument is the physical manifestation of the abnormality, related failure modes, and probable failure mechanisms.

*The material contained in this section has been abstracted from "Integrated Circuits in Action: Part 4, Post Mortems Prevent Future Failures," S. Schwartz, Electronics, (January 23, 1967).

(COURTESY ELECTRONICS, McGRAW-HILL ©)

TABLE 6-2
FAILURE-ANALYSIS INSTRUMENTS

Instrument	Abnormality Observed
	Physical
Dye penetrant	Cracks, pores
Etching and microscopy	Dislocation distribution
	Pits, cracks, and chips
	Pinholes and cracks in oxide insulating layer
Binocular microscope (3-120X magnification)	Non-homogeneity
	(Oxidation, contamination, intermetallics or rub marks)
	Opened bonds
	Cracked dice
	Cracks in package lead seals
	Pits and pyramids on dice
	Poor registry or masking
	Scratches on dice or intraconnects
	Current concentrate NS
Phase contrast microscope	Stacking faults
	Transparent contaminants
	Improper oxide metalization topography
Dark field microscope	Photoresist residues, dust
	Bubbles in sealant glass
Interferometer	Oxide and surface topography varies from design
	Oxide thickness varies from design
	Metal thickness varies from design
Electron microscope (magnifies to 100,000 X)	Distribution of dislocation and stacking faults
	Contaminant and corrosion location
	Etch pits, scratches, dust, and deposit roughness
	Pattern alignment and topology errors
	Undercut etched edges
	Weld porosity
	Porous diffusion products
	Scribe cracks
Radiographic equipment	Voids in welds
	Metal migration
	Contaminant particles
	Cracks
	Long wires
	Misalignment of metal parts
Controlled etching	Depth of inversion charge
X-ray diffraction	Identity of contaminant compounds or corrosion product compounds
	Identity of materials
Metallograph	Observes same abnormalities as binocular microscope
	Poor adherent interconnects or bond structures
	Abnormal junction depth (by angle lapping)
	Voids in thermo-compression bonds
	Incomplete or poor welds
	Thin or nonadherent plating
Electron back-scatter thickness meter	Thickness of oxide, plating, or photo-resist
	Chemical
Hot-stage metallograph	Contaminant melting point and reactivity (for identification)
	Interdiffusion of metals, for example, Al-Au, Ni-Au, Ti-Au, Kovar-Au
	Whisker growth
	Strain growth
	Surface diffusion of metal films
Electron diffraction	Contaminants, crystal structure, and identity of intermetallic compounds
	Deposit crystallinity
Low-energy electron diffraction	Presence of absorbed substances in surface
Electron microprobe (elemental chemical analysis)	Contaminant or rub-mark residue identity and map
	Dopant and dopant concentration
	Intermetallic analysis
	Corrosion product identity
	Deposit topography map by chemical element
	Deposit thickness map by chemical element
	Crack, pit, and pinhole maps
	Junction misalignment or movement
	Opens, short, and current and/or voltage nonuniformity map
Gas chromatograph	Abnormal package ambients
	Leaks (by presence of air or test fluid)
Mass spectrograph	Leaks: Also: reaction of device to ambient changes in the spectrograph (Absorption, then charge migration causes inversion)
Infrared absorption spectrograph	Thin oxide
	Abnormal oxygen content in silicon
	Abnormal epitaxy thickness
	Impure photoresist
	Water
Emission ultraviolet and visible spectrograph	Analysis and identity of contaminants
Visible and ultraviolet absorption spectrographs	Analysis and identity of contaminants
Neutron activation analyzer	Same as emission spectrograph but at lower concentrations
	Mechanical
Strain gauges	Loose headers or dice
	Poorly brazed headers or dice
Bubble tester	Gross leaks in packages
Helium leak tester or Radiflo tester	Small leaks
Thermal plotter	Hot spots due to voids
	Hot spots due to thin base areas
	Hot spots due to dislocation or stacking faults
	Electrical
Curve tracer	Shorts
	Opens or intermittent contacts
	Soft junctions
	Abnormal resistance
	Leakage currents and inversion

6.2.6 Multilayer-Board Failure Modes

Multilayer boards were discussed in Chapter 3. As indicated at that time, these are difficult to produce reliably. In this section some of the failure mechanisms and reliability aspects of multilayer boards are discussed.*

The data presented here were obtained from approximately 40 failed boards. These boards had passed screening and final inspection, and their components were already mounted on them.

6.2.6.1 Chemical Corrosion

One source of chemical corrosion is incomplete removal of plating or etching solutions. Another is related to incomplete applications of photoresist to protect metal surfaces, especially copper plate, in the holes. Corrosive etch may be trapped in internal pockets within the board, resulting in time-dependent failures. Figure 6-19 illustrates a section of a plated-through hole corroded from from process etchants that penetrated voids in the tin-lead plating.

6.2.6.2 Drilling Defects

Improper or poor control of drilling can introduce defects into multilayer boards. Figure 6-20 illustrates "epoxy smear" that prevents contact between hole plating and terminal areas. Another drilling defect is illustrated in Figure 6-21. Incorrect drilling procedure has torn out the internal pad. Figure 6-20 also illustrates unsatisfactory through-hole plating. Such defects may be caused by pieces of glass, epoxy, or metal that remain after the drilling operation.

COURTESY OF NEP CONFERENCE

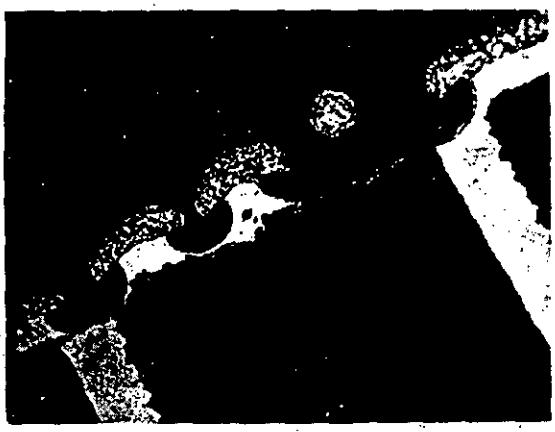


FIGURE 6-19

CORRODED PLATED-THROUGH HOLE

COURTESY OF NEP CONFERENCE

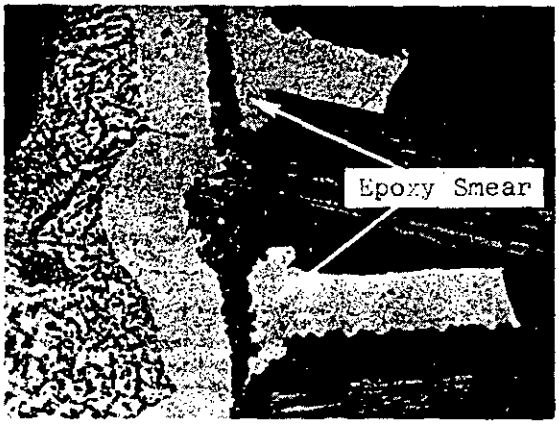


FIGURE 6-20

EPOXY SMEAR, PLATED-THROUGH HOLE

*This section was abstracted from a paper by John E. McCormick, "Multilayer Reliability", NEP Conf. 1967 (West), Long Beach, California.

COURTESY OF NEP CONFERENCE

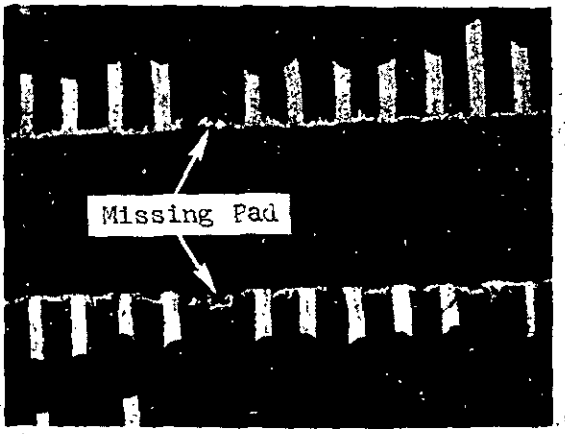


FIGURE 6-21

DAMAGE CAUSED BY DRILLING, PLATED THROUGH HOLE

COURTESY OF NEP CONFERENCE

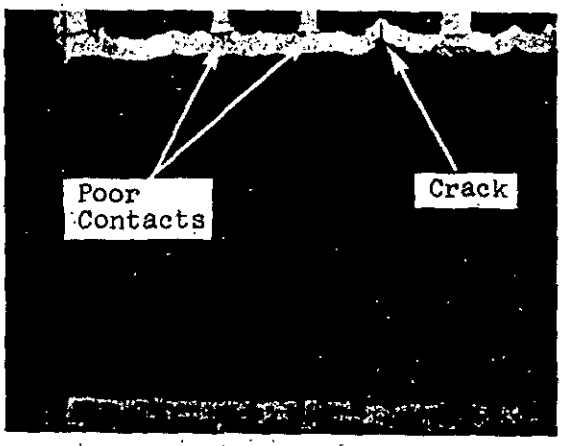


FIGURE 6-22

CRACKED PLATE, PLATED-THROUGH HOLE

COURTESY OF NEP CONFERENCE

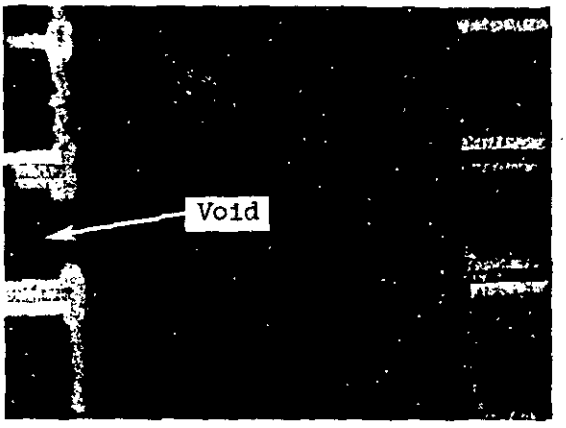


FIGURE 6-23

VOID IN PLATE, PLATED-THROUGH HOLE

COURTESY OF NEP CONFERENCE

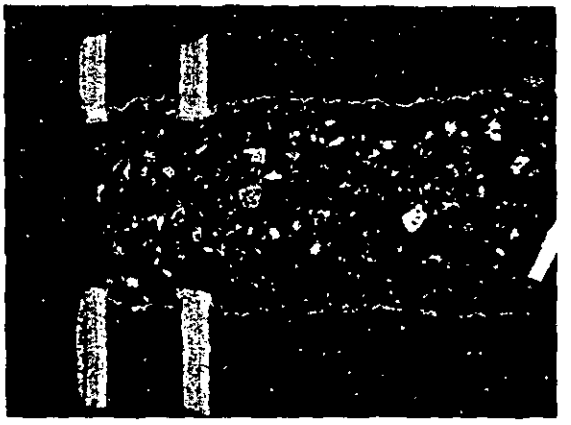


FIGURE 6-24

INADEQUATE PLATE, PLATED-THROUGH HOLE

6.2.6.3 Stress and Voids

Stresses induced during processing may cause metal fracture within the board conductors as illustrated in Figure 6-22. Voids or inadequate plating may also induce board failures. A void is illustrated in Figure 6-23. The metal has not deposited on the wall, possibly because of contamination at that point or possibly because of the formation of a gas bubble during plating. An example of inadequate plating is shown in Figure 6-24. It was probably caused by depletion of the plating solution in the hole or by inadequate agitation during the plating operation.

COURTESY OF NEP CONFERENCE

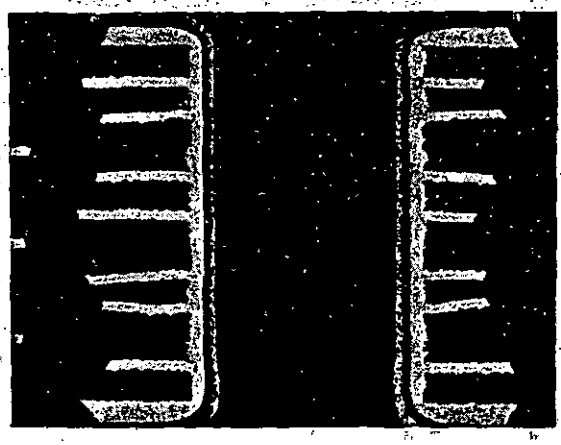


FIGURE 6-25

EXCELLENT PLATED-THROUGH HOLE

A good plated-through hole is illustrated in Figure 6-25. The following are characteristic of a reliable plated-through hole:

- Uniformity of wall thickness of tin, lead, and copper
- Absence of voids and pinholes
- Continuity between hole plating and all terminal areas
- Good drilling characterized by straight, smooth walls free of epoxy smear

Approximately 25 percent of board failures were caused by shorts on the surface of the boards due primarily to solder bridging between contact pad or conductors. There were five internal shorts and five open or intermittent open failures. Internal shorts are most probably caused by improper drilling or slippage or misalignment of the layers during lamination. The open failures are related to insufficient wetting of the bond area with solder. This is an assembly rather than manufacturing defect and is identified as a connection error. The adjacent tabulation summarizes the failures associated with the boards that were analyzed.

Board Failures	88%
Connection Failures	12%
Board Failures	
Opens (internal)	60%
Shorts (surface)	30%
Shorts (internal)	10%
Connection Failures	
Opens	100%
Shorts	0%

6.3 SOURCES OF QUANTITATIVE RELIABILITY DATA ON INTEGRATED CIRCUITS

Reliability data for integrated circuits are available from two broad categories of observations: device test and system operation. Within the device-test category are operating life tests, accelerated life tests, step-stress tests, environmental tests, and screening tests -- all performed by the device manufacturer -- as well as tests that are performed by the system manufacturer. Within the system-operation category, reliability data are available from observations of prototype test and system demonstration by the system manufacturer and from observations of system operation by the user.

6.3.1 Device Testing

6.3.1.1 Operating-Life Tests

Operating-life tests are performed on devices primarily to ascertain the probable (absolute) reliability level (MTBF and failure rate) of the lot of devices that the devices on test represent. A secondary objective is to generate failure-mode information. One of two stress conditions is generally used:

(1) Normal electrical conditions and 25°C free air ambient.

(2) Maximum rated electrical and thermal (85°C or 125°C) conditions. In some instances, power is applied but switching operations are not performed.

Operating-life tests can be divided into three categories: static, parallel switching, and series switching. Of the three, most integrated-circuit manufacturers use the series-switching test because it enables them to accumulate the greatest number of a-c test hours (although it can be expected to result in lower observed failure rates than a parallel-switching test would show).

6.3.1.2 Static Bias Life Tests

The static life test is a d-c test similar to transistor life tests. D-c voltages are applied to the circuit and maintained. A test of this type is suitable for bias drivers and similar circuits. Its basic advantage is that it requires a minimum of connections. Its disadvantage is that it fails to stress the circuit adequately in that no signals are applied to the network. Because of this, the results of static life tests are not a good indicator of the reliability of the circuits under test.

6.3.1.3 Parallel-Switching Life Tests

Each driver is given an independent switching signal (usually 60 cps) when a network is subjected to a parallel-switching life test, although the circuits have a common bias. This type of life test provides the best indication of circuit reliability because the devices operate independently and are adequately stressed. Because this test requires more connections than the static and series life tests and requires external signal sources, the number of circuits that can be tested simultaneously is limited because of the higher costs involved. Therefore, the quantity of reliability data available for analysis from parallel testing is smaller than desired.

6.3.1.4 Series-Switching Life Test

In a series-switching life test, each circuit is connected in series; i.e., the output of one circuit drives the input or the following circuit, etc. The frequency achieved (usually in the megacycle range) depends on the propagation

delays of the circuits. The advantage of such a test, often called a ring-counter or ring-oscillator test, is one of cost and time. Because of the smaller number of connections and the requirement for only a single signal source, many circuits can be tested simultaneously. For this reason, by far the largest quantity of reliability life-test data available today is based on this type of test. Its major disadvantage is that the performance of each circuit is dependent on the performance of each preceding and succeeding circuit. Critics of this type of test claim that the test does not stress the devices adequately, while proponents argue that the environment is similar to that experienced by devices in most systems.

Many of the life tests are conducted for only 1000 hours, although in some cases some of the specimens in the short-term tests are left on test for a long-term reliability analysis. Consideration is often given to the relationship of junction temperature with the combined effects of ambient temperature and dissipated power. These factors are then controlled to maintain a given junction temperature (usually the maximum rated).

6.3.1.5 Accelerated Life Tests

Accelerated life tests accelerate the failure mechanisms of a circuit in order to reduce the number of samples and test time required to obtain useful reliability data. Usually the acceleration factor is a thermal or electrical stress greater than that expected for the circuit in system operation. A detailed knowledge of the device failure mechanisms is necessary before the test is designed. If an accelerated life test introduces new failure mechanisms or does not accelerate all failure mechanisms, the test is not valid for predicting reliability at lower stress levels.

Integrated-circuit manufacturers usually run accelerated life tests only at elevated temperatures. This philosophy is carried over from the transistor technology, in which the majority of failure modes of a semiconductor device are influenced by temperature. Storage tests are the most economical accelerated tests since they do not require elaborate test set-ups. Unfortunately, however, several known failure mechanisms are not significantly accelerated without the presence of electrical stimuli.

Accelerated life tests are used for the following:

(1) To predict failure rates at lower stresses

(2) To compare relative failure rates

(3) To assess potential problem areas

When circuit manufacturers have accelerated tests performed at three or more stress levels, failure rates are generally plotted on a logarithmic scale against stress (usually temperature, °C) on an absolute scale. The curve is then extrapolated to lower stress levels to estimate the failure rate at any lower stress

level. Using the acceleration factors to predict failure rates, however, opens the door to many uncertainties concerning the validity of the predicted rate; a small error in calculation of the true failure rate, or incomplete knowledge of the effect of other stresses, can cause significant errors in the estimates. For this and similar reasons, failure rates estimated from accelerated-test results are not universally accepted.

The possible errors introduced by the uncertainties discussed above can be minimized if the results of an accelerated test on a new circuit can be compared with those of an identical test on a circuit with a known reliability level. For this reason, accelerated life tests can be beneficial, indicating the relative reliability of a new design or process when a quick reaction time is required. Accelerated life tests can also provide a continuous indication of the quality of a given production line's output.

6.3.1.6 Step-Stress Tests

A step-stress test is a reliability test in which a sample of a population is subjected to discrete stress levels of successively increasing severity until all or nearly all units fail. After the device is subjected to each discrete stress level, it is generally returned to its normal operating stress level. Its critical parameters are then measured and the number of rejects determined. The purpose of the step-stress-to-failure test is generally one of the following:

(1) To determine the safety factor inherent in the design with respect to a given stress

(2) To accelerate failure mechanisms or uncover failure modes in devices

(3) To predict failure rates at lower stress levels in a shorter calendar period and with fewer samples than required by operational life testing

(4) To compare the relative reliability of two or more integrated-circuit designs, manufacturing processes, etc.

(5) To check for unannounced changes in design, materials, and processing by the vendor or to evaluate the effect on device units of announced changes.

The distribution of failures with respect to the applied stress can be plotted and used to indicate the strength of a particular device with respect to a given stress. The probability of failure with respect to this given stress can be estimated by comparison of the applied stress with the anticipated or actual operating stress.

The causes of failure can be reduced or eliminated if design and manufacturing procedures are improved through the use of information obtained by accelerating the failure mechanisms so that the cause of failure can be identified. The cumulative percentage of failures is plotted on a Gaussian probability scale and the stress level on an absolute scale. If the line that joins the points has a sharp discontinuity, it can be assumed that a new failure mechanism has been introduced

at the stress level indicated by the point of discontinuity. Failure analysis is then performed to determine failure modes. Failure-mode information can be used to take corrective action to eliminate the cause of failure. Another step-stress test can then be run to measure improvements.

Assumptions are made regarding failure distributions and acceleration factors. Failure rates at lower temperatures are estimated from extrapolations of the step-stress-tests data. These estimations can be criticized for the same reasons for which estimates based on accelerated life tests are criticized. For the reasons given in the discussion of accelerated life testing, comparative failure rates can be more useful than absolute failure rates.

As is the case with accelerated life testing, temperature is the most common stress used for step-stressing of integrated circuits, although electrical stresses have been used.

6.3.1.7 Environmental Tests

Environmental tests are used to determine the capability of an integrated circuit to withstand certain operational stresses. These tests usually follow MIL-S-19500 in accordance with procedures in MIL-STD-750 and MIL-STD-202. Some manufacturers have run tests exceeding the requirements of these military documents and have also conducted special environmental tests dictated by specific customer requirements, e.g., radiation tests. These tests are discussed in Chapter 4.

6.3.1.8 Preconditioning and Screening Tests

Preconditioning and screening tests are tests performed on the integrated circuits prior to their installation in an operational system in order to identify potential failures. Screening tests usually do not include inspections. Particular screening tests and inspections are discussed in Chapter 4. Screening tests can be performed by the integrated-circuit vendor or the system manufacturer, or by both.

Reliability data generated by vendors' device tests are readily available to interested parties; similar data generated by system manufacturers are much more difficult to obtain. Generally, the latter are preferred since they can be assumed to be less biased. Vendor data may be very useful in assessing a circuit's potential reliability, but analyses from such data must be carefully scrutinized since they can easily be misinterpreted. The wise customer will request sufficient data to perform an independent analysis.

6.3.2 Reliability Data Obtained from System Operation

Integrated-circuit reliability data are collected because a system manufacturer or user wants to know how reliable a proposed system using these circuits will be; what can be done to make the system more reliable, and at what cost; what system

failure modes can be expected, and their frequency; and what the spares-provisioning and maintenance requirements will be. The primary objective, then, is to predict how the integrated circuit will perform in the system environment. If available to the analyst, the data most relevant to the system studies are data drawn from experience with similar systems.

While a significant quantity of data on microelectronic system operation has been accumulated, it is not always available to the general technical public in sufficiently detailed form. The data are often related to classified or proprietary programs and are thus more difficult to obtain than device data. Nevertheless, system and equipment engineers using integrated circuits should make every effort to obtain this type of information.

Reliability data based on system operation are of greater value than those based on device tests because the observed results may reflect system environment factors not reflected in device testing. The most important of these factors may be the failures that are not the direct result of a device failure but that can be attributed to the fact that a device is being used.* For example, an interconnection failure between the device package and the circuit board, or within the board, would not be observed or, if observed, certainly not counted during a device test. Other factors include the following:

(1) The failures resulting from the interdependence of device operation, such as drift, noise, etc.

(2) The design problems associated with the implementation of the microelectronics technology

(3) The actual physical environment (rarely reflected exactly in device testing)

(4) The effect of failures, drift, etc., of other circuits or subsystems

(5) The effects of maintenance and built-in test equipment

(6) The system characteristics resulting from the use of particular logic types

Because the number of sources of system data is a function of the number of systems using microelectronics, there are considerably more such sources than there are sources of device test data, although the quantity of device operating hours from any given system source can be expected to be modest. Because of this and the problem of accessibility to the data, the engineer seeking system data usually will have to limit the number of data sources he interrogates. The following should be considered in the selection of systems as potential sources of integrated-circuit reliability data to be used in system development:

(1) Functional similarity

(2) Environmental similarity

*For example, see J.B. Brauer and D.F. Barber, "Are Reliabile Microconnections Really Attainable?" Electronic Packaging and Production, (June 1966), p. 91-95.

(3) Common logic

(4) Packaging commonality

(5) Interconnection similarity

Microelectronics reliability data can be obtained from observations of field operation, demonstration, and prototype test.

6.3.2.1 Field Operation

Reliability data acquired in an analysis of field operations should theoretically be the most useful to the reliability engineer. Because the devices are operating in a "real" system, exposed to the actual physical and maintenance environment, the data include many of the factors that other data sources do not.

Practically, however, these data may have some drawbacks, including the following:

(1) Until more microelectronic systems become operational, there is insufficient operating time to draw confident conclusions in many areas.

(2) Failure reporting is rarely complete. This problem is compounded by retrofits, where the total operating time is often not known. Incomplete failure reporting or the absence of any reporting is common, and space-borne microelectronic systems are particularly difficult because telemetry cannot monitor every circuit and cannot determine the failure mechanisms.

(3) Maintenance personnel are still in the early part of a learning curve. This lack of experience significantly contributes to the failure rate when a new technology is introduced.

(4) Many new systems are still not debugged; failures may reflect lack of design experience, which would not normally be evident in follow-on systems.

6.3.2.2 System Demonstration

Observations of system demonstration minimize the effects of the first three factors listed above. To date, a significant amount of system demonstration time has been accumulated. However, most of the pertinent data are not centrally located; they are in the possession of the manufacturers who performed the demonstrations. The biggest drawback to these data is that they do not reflect the operational environment. Some systems do undergo testing in a simulated environment, but generally the percentage of operating hours in environmental test is low.

6.4 RELIABILITY DATA

Table 6-3 summarizes reliability data that have been reported by equipment manufacturers and users to give an indication of failure rates. Although reliability data are becoming more plentiful, it can be reasonably argued that additional data of this type would be of little value without details of preconditioning, screening, testing, and data-recording rules (e.g., no failures are counted during the first 100 hours of subsystem testing). Because of the wide range of failure rates being reported at the device level and the questionable validity of such information for use in a particular application, device failure rates should be used with care.

Data such as those in the table are useful only for broad generalizations, and an expansion of the number of samples in these tables would have little utility. These data represent failure rates that have been obtained on specific samples; the reader is reminded that the source selected and specifications enforced for a particular item will determine the reliability for that particular item.

TABLE 6-3

TYPICAL MICROCIRCUIT FAILURE DATA

Level of Test	Test Hours	Failures	Failure Rate* (Multiply by 10^{-6})	Date	Remarks
Airborne Navigational Set	10,000,000	7	0.7	1/67	No failures counted during first 100 hours. Navigational computer system field experience. Four types of IC's for a total of 1000 per system.
UHF Radio Set	304,000	1	3.3	4/67	MIL STD-781A, Test Level F -- 3 hours off, 5 hours on. Temperature range -54°C to +71°C. Failure occurred 330 hours into test due to open bond.
Airborne Navigational Set	2,500,000	28	11.2	9/66	Level-2 Agree.
A System	15,800,000	1	0.063		Preconditioning: 4000 hours system operation at 250°C before failures are counted.
A System	600,000	0	1.7	6/65	Two failures caused by error in testing.
Digital Computer	13,600,000	0	0.052	1/67	Operational data. Operating temperature believed to be in excess of 50°C.
Radar Indicator	1,000,000	2	2.0	7/66	
A System	2,000,000	0	0.5	5/65	Time-code generating system.
ECM	800,000	1	1.2	6/65	Failure cause not known.
A System	6,000,000	36**	6.0	6/66	Actual flight results.
System	3,500,000	0	0.29	6/65	T.I. Series 51 for "EPIC" GCE on BOMARC prototype.
Digital Computer	220,500,000	51	0.23	6/65	Includes 17 failures due to particles. Such failure modes have been subsequently eliminated. Operational data.
A System	66,000,000	18	0.27	10/66	DTL circuits used primarily in aerospace systems. Operational data.
Computer	20,000,000	7	0.35	6/65	
Apollo Computer	319,000,000	6	0.018	5/66	Extensive screening tests are performed. Operating data for Block I and Block II computer.
Minuteman	155,000,000	33	0.213	4/66	System testing after screening, etc.
Minuteman	100,000,000	0	0.001	4/66	Field operation (after system testing).

*Failure rates are estimated by dividing the number of failures by the test hours (one failure is assumed where none is reported).

**Removals only are listed; failures were not verified.

BIBLIOGRAPHY: CHAPTER 6

RELIABILITY AND PHYSICS OF FAILURE

Adams, J. D., and P. L. Holden. "What is Meaningful Integrated Circuit Reliability," Electronic Design, (1965), pp. 54-58.

ARINC Research Corporation. Reliability Handbook. Annapolis, 1966.

Beading, D., and C. Casey. Evaluation Procedures for Reliability, Maintainability, and Effectiveness of Microelectronics. RADC-TR-65-477. 1966.

Becker, P. W., and R. E. Warr. "Reliability vs. Component Tolerances in Microelectronic Circuits," Proceedings of the IEE, (1963).

Bell, J. E. "Weibull Distribution in Reliability Analysis of Certain Electronic Components," Semiconductor Products, (1963).

Bernay, R. A. "Looking at Integrated Circuits," Electronic Industries, (1965), pp. 76-80.

Best, G. E., G. R. Bretts, and H. M. Lampert. "Physics of Failure and Accelerated Testing," Electro-Technology, (1965), pp. 81-94.

Borofsky, A. J., and D. C. Fleming. "Design and Process Contribution to Inherent Failure Mechanisms of Microminiature Electronic Components for Minuteman II," Fourth Annual Symposium on Failure in Electronics, November 1965.

Brauer, J. B. "Microelectronics and Reliability Assessment," Electronic Industries, (1966), pp. 67-71.

_____. "Poor QC Yields Bad IC's," Electronic Engineer, (1966).

_____. Reliability in Microelectronics. RADC-SP-66-3. 1966.

Benesch, S. W., and C. R. Gates. Model of Failure in Components. Jet Propulsion Laboratory. ND.

Christensen, D. F., and M. E. Nelson. "Improving Reliability," Electronic Industries, (1961)

Cummings, D. G. "Identification of Thermal Compression Bond Failures," IEEE Wescon Technical Papers, (1965).

Cummings, D. G. "Thermal Compression Bond Matrix Study," IEEE Wescon Technical Papers, (1965).

Cunningham, J. A., and J. G. Harper. "Semiconductor Reliability Focus on the Contacts," Electronic Engineer, (1967).

DaSilva, T. Failure Mode Tabulation and Apportionment Documentation, Reliability Engineering Report, Signetics Corporation. Sunnyvale, California 1966.

deAtley, E. R. Reliability Report on Planar Epitaxial Microcircuits. Philco Corporation, Lansdale, Pennsylvania.

de Miranda, W. R. Rodrigues. "Reliability of Integrated Circuits Analysis of a Survey," Proc. 1966 Symposium on Reliability.

Dertinger, E. F. Status of Reliability Requirements in Government Contracts. Raytheon Report.

Dworkin, E. J., and R. P. Misra. "A New Method for Predicting Reliability," Electronic Industries, (1966).

Easterday, J. L., and R. G. Bowman. Reliability of Integrated Circuits Used in Missile Systems. U. S. Army Missile Command. RSIC-220. 1964.

Finocchi, A. J. "Interpreting MIL SPECS for System Reliability," Electronic Products.

_____. "Reliability Has Failed to Meet Its Goals," Analysis of Reliability and Maintainability, (1965), pp. 28-30.

Flood, J., and K. R. Mackenzie. Integrated Circuit Reliability, Motorola AN-148.

Fruci, N. A. "Pessimistic Circuit Design," Electro-Technology, (1963).

Go, H. T. "Studying and Controlling IC Reliability," Electronic Industries, (1966).

_____. Integrated Circuits: How and Why They Fail. EEE.

_____. Integrated Circuit Failure Mode Data and Cost Analysis. Westinghouse Defense and Space Center. N62462-6-1687.

Goldberg, F. M. Physics of Failure In Electronics: Volume 3. ITT Research Institute. 1964.

Gorton, H. C., and K. P. Duchamp. "Research Toward A Physics of Aging of Silicon P-N Junctions," IEEE Transactions on Component Parts.

Grose, V. L. "Reliability Can be Predicted," Analysis of Reliability and Maintainability, (1965), pp. 28-30.

Handlye, L. D., J. Partridge, and E. C. Hall. The Application of Failure Analysis In Procuring and Screening of Integrated Circuits. MIT Instrumentation Laboratory.

Jenson, P. A. "Introducing Redundancy in Analog Systems," Electro-Technology, (1965).

Karmiol, E. D., W. Weir, W. Thomas, and S. J. Youtcheff. "A Practical Method of Demonstrating Reliability," Electronic Industries, (1966).

Kemp, J. C. "Optimizing Reliability in Digital Systems," Computer Design, (1963).

Lawrence, J. G. "Simplifying System-Reliability Calculations," Electro-Technology, (1964).

Lawton, A. T. "An Equipment Manufacturer's Problems in the Application of Microelectronics" Microelectronic Reliability, (1965).

Levine, J. I. "Failure Prevention Through Design Optimization Utilizing Monte-Carlo Simulation Techniques," Solid State Design, (1965).

Maloney, R. Setting Reliability Incentives Using Linear Programming. New York

McConnell, Harold V., "Failure Effect and Failure Mode Analysis Using a Digital Computer," Evaluation Engineering, January-February 1966.

Motorola, Inc. Current Reliability Status.

NASA Microelectronics Reliability Program. Information Dynamics Corporation. 1966

Partridge, J. H. Progress Report on Attainable Reliability of Integrated Circuits for Systems Applications. MIT Instrumentation Laboratory. Report E-1679.

Gorton, H. C. RADC Reliability Physics Notebook. RADC-TR-65-330. November 1965.

Pieruschka, Erich. Specification and Measurement of the Reliability of Parts of Guided Missiles and Space Vehicles. Lockheed LSMC-800920. Sunnyvale, California, 1961.

Rapp, W. K. "Reliability With Standby Units," Electro-Technology, (1965).

Ravenis, J. V. "Estimating Weibull-Distribution Parameters," Electro-Technology, (1964).

Rome Air Development Center. Failure Mechanisms in Microelectronics. RADC-TDR-63-425. 1964.

Royal, E. L. "Practical Reliability-Engineering Program," Electro-Technology, (1963).

Santeler, D. J. "Effect of Vacuum Environment on Thin-Film Component Reliability," SCP and Solid State Technology, (1965), pp. 17-22.

Schwartz, S. "Integrated Circuits in Action: Part 4, Postmortems Prevent Future Failures," Electronics, (1967), pp. 92-106.

Shuda, D. G. "The Design Engineer Applies Reliability Techniques," Solid State Design, (1964).

Siegal, M., and R. Crosby. "Defect Analysis by Device Sectioning," SCP and Solid State Design, (1966), pp. 35-41.

Skinner, S. M., E. R. Pemsel, W. J. Lytle, J. W. Dzimianski, and A. E. Javitz. "Improve Device Reliability With Physics-of-Failure Techniques," Electronic Design, (1965), p. 299.

Soltau, R. Failure Modes and Mechanisms in Microelectronic Devices, Philco Corporation. Lansdale, Pennsylvania, 1965

Stansbury, A. P., and R. A. Struble. "Application of Noise Measurements to the Reliability Analysis of Semiconductor Devices," SCP and Solid State Technology, (1965).

Thomas, H. R. "Comments on MIL-HDBK 217 and MIL-R-22256, 22973, and 23094A", Analysis of Reliability and Maintainability, (1965), pp. 28-30.

Toye, C. "Extrapolating Component Life Tests", Electro-Technology, (1964)

UNIVAC. Integrated Circuit Study. Sperry Rand Corporation, 1964. (Bureau of Ships PX3266.)

Young, M. A. "Predelivery Steps Toward Reliable Computers," Analysis of Reliability and Maintainability, (1965), pp. 28-30.

GLOSSARY

GLOSSARY

The following terms have been coordinated with the Electronics Industry Association, Aerospace Industries Association, Department of Defense, and NASA.

Microelectronics. That area of electronic technology associated with or applied to the realization of electronic systems from extremely small electronic parts or elements.

Element (of a microcircuit or integrated circuit). A constituent of the microcircuit or integrated circuit that contributes directly to its operation. (A discrete part incorporated into a microcircuit becomes an element of the microcircuit.)

Substrate. The supporting material upon which or within which the elements of a microcircuit or integrated circuit are fabricated or attached.

* Microcircuit. A small circuit having a high equivalent-circuit-element density, which is considered as a single part composed of interconnected elements on or within a single substrate to perform an electronic-circuit function. (This excludes, for example, printed wiring boards, circuit-card assemblies, and modules composed exclusively of discrete electronic parts.)

* Integrated Circuit. A microcircuit consisting of interconnected elements inseparably associated and formed in situ on or within a single substrate to perform an electronic circuit function.

* Monolithic Integrated Circuit (Monolithic Microcircuit). An integrated circuit consisting of elements formed in situ on or within a semiconductor substrate with at least one of the elements formed within the substrate.

* Multichip Microcircuit. A microcircuit consisting of elements formed on or within two or more semiconductor dhips that are separately attached to a substrate.

Semiconductor Chip. A single piece of semiconductor material of any dimension.

* Film Integrated Circuit (Film Microcircuit). A circuit consisting of elements that are films all formed in situ upon an insulating substrate.

* Hybrid Microcircuit. A microcircuit consisting of elements that are a combination of the film circuits and the semiconductor circuits or a combination of one or both of these types with discrete parts.

*These definitions appear in MIL-STD-1313, Microelectronic Terms and Definitions

☆U.S. GOVERNMENT PRINTING OFFICE: 1969 - 604-833/10030

STANDARDIZATION DOCUMENT IMPROVEMENT PROPOSAL

(See Instructions – Reverse Side)

1. DOCUMENT NUMBER	2. DOCUMENT TITLE

3a. NAME OF SUBMITTING ORGANIZATION	4. TYPE OF ORGANIZATION *(Mark one)*
b. ADDRESS *(Street, City, State, ZIP Code)*	☐ VENDOR ☐ USER ☐ MANUFACTURER ☐ OTHER *(Specify)*: _____

5. PROBLEM AREAS

a. Paragraph Number and Wording:

b. Recommended Wording:

c. Reason/Rationale for Recommendation:

6. REMARKS

7a. NAME OF SUBMITTER *(Last, First, MI)* – Optional	b. WORK TELEPHONE NUMBER *(Include Area Code)* – Optional
c. MAILING ADDRESS *(Street, City, State, ZIP Code)* – Optional	8. DATE OF SUBMISSION *(YYMMDD)*

(TO DETACH THIS FORM, CUT ALONG THIS LINE.)

DD FORM 82 MAR 1426

PREVIOUS EDITION IS OBSOLETE.