In Silicon Engine

exhibit

6.19.09

SEMICONDUCTOR INTEGRATED CIRCUITS

by G. E. Moore

FAIRCHILD SEMICONDUCTOR

A Division of Fairchild Camera & Instrument Corporation

to be published by McGraw-Hill as

Chapter 5 of

PRINCIPLES OF MICROELECTRONIC ENGINEERING

Edited by E. Keonjian

(1962)

FAIRCHILD

SEMICONDUCTOR

545 Whisman Road · Mountain View, California

ROBERT K. WAITS

SEMICONDUCTOR INTEGRATED CIRCUITS

by G. E. Moore
FAIRCHILD SEMICONDUCTOR
A Division of Fairchild Camera & Instrument Corporation

to be published by McGraw-Hill as
Chapter 5 of
PRINCIPLES OF MICROELECTRONIC ENGINEERING
Edited by E. Keonjian "MICROELECTRONICS"

(1962)

PREFACE

Rather than present an exhaustive survey of the semiconductor integrated circuits which have been reported to date, this chapter attempts to view the capabilities and limitations of the underlying technology and the device structures using specific integrated circuits only as examples of the broad capability inherent in the expanding technology. The emphasis has been in trying to point out the unique features associated with integrated circuitry. For example, considerably more space is devoted to the possible structures to achieve isolation than is devoted to the transistor itself. The only justification for this relative emphasis is that the transistor has been discussed in detail elsewhere.

Much of the data and information in this chapter is previously unpublished. It represents work done by many contributors at Fairchild Camera and Instrument Corporation. While to single out a few is probably unfair to many others, I would like to give credit to D. Farina and N. Gault for supplying much of the data and to V. H. Grinich and R. N. Noyce for critically reading the manuscript. I would like to thank Mrs. Ruth Ann Cameron and William Jimenez for their work on the figures and Mrs. Helen Bonfadini for extracting and preparing the manuscript.

TABLE OF CONTENTS

SECTION	TITLE	PAGE
5-1	INTRODUCTION	1
5-2	THE TECHNOLOGY	6
	Materials Preparation	7
	Diffusion	10
	Epitaxial Growth	31
	Surface Geometry Control..............	42
	Metallization for Intraconnections.......	52
	Assembly and Packaging	55
5-3	DEVICE STRUCTURES	63
	Transistors	64
	Diodes	66
	Isolation	66
	Intraconnections	72
	Resistors	74
	Capacitors	82
	Other Structures	85
5-4	CIRCUIT CONSIDERATIONS	93
	Digital Circuitry	96
	Linear Circuits	103
5-5	AN EXAMPLE	105
	Manufacturing Steps	109
	Properties and Characteristics	117
5-6	COST AND RELIABILITY	120
5-7	SUMMARY AND CONCLUSIONS	139
	REFERENCES	141
	FIGURES	147

5-1. INTRODUCTION

If miniaturization were the sole justification for microelectronics, an improvement of the order of a factor of 10^3 in volumetric efficiency could be achieved merely by elimination of the packing voids in conventional electronics. The cost of this type of assembly can be easily justified (see Chapter 3) for special applications where volume and weight are at a premium. The excitement created by microelectronics, however, extends far beyond size and weight alone. The potential achievements in the reduction of cost and in the improvement of reliability point the way to an entirely new realm of allowable system complexity. The actual achievements to date, while considerably more modest than the publicity might lead one to believe, suggest that microelectronics will be employed extensively from now on. An important approach to this exciting field is provided by semiconductor integrated circuitry. This should not be viewed as a completely independent entity from other approaches, for example thin films, but rather as a complementary approach evolving toward the same general goal but arising from a different technological background. The origin of this evolutionary process sprang naturally from the transistor and diode technology. It is being extended to include more and more complex combinations of operations.

Before we proceed further, some definition of semiconductor integrated circuitry is required because this term or similar terms have been applied to a wide variety of levels of sophistication. There are two extremes in semiconductor integrated circuitry. The first of these is the chip approach wherein individual components, such as transistors, resistors and diodes, are produced on separate pieces of material, then these separate components are mounted and interconnected in a single package to produce a circuit function by what is in reality a microassembly technique. The second extreme makes the entire electronic function in and upon a single piece of semiconductor material having many components or regions that are isolated or interconnected electrically, as the circuit requires. In general in this latter approach all of the intraconnections within the functional block itself will be done by batch processing on large numbers of circuits. The only individual assembly operations are associated with mounting the final circuit function in a package so that it can be connected conveniently to the outside world. All levels between these two extremes are practiced. The level of integration will, in general, be determined at any time by the degree of development of the technology and such economic considerations as the number of identical circuits needed, the time scale, etc.

The approach employing individual components offers the advantage of complete flexibility analogous to that obtainable in conventional electronics. Individual chips are interconnected by lead bonding techniques essentially identical with those employed in transistor manufacture. More elaborate schemes wherein the individual components are mounted in precise locations on a substrate and where the separations between individual components are filled with a material suitable for thin film interconnections have been proposed. None of these schemes have been developed adequately that they can be seriously considered at the present. This chapter will be concerned primarily with the other extreme of semiconductor integrated circuits -- those circuits which are produced entirely in a single monolithic block of material. Examples of the latter are Solid Circuits produced by Texas Instruments, Inc., and Micrologic Circuits produced by Fairchild Camera and Instrument Corporation. Intermediate cases can usually be considered as the single chip approach applied to smaller monolithic blocks; for example, the Molecular Electronics used by Westinghouse in the AN/ARC-63 transceiver[1]. The only semiconductor materials which need be considered for their applicability to integrated circuits at present are germanium and silicon. Since the trend toward integrated circuits arose after the silicon technology was well developed, and

indeed was to a considerable extent stimulated by this technology, germanium has not been an important material. We will accordingly restrict our discussion to silicon. It is unlikely that other semiconductor materials, such as, for example, gallium arsenide will be of importance in the field of integrated circuitry for other than a few highly specialized applications.

The large scale manufacture of silicon mesa transistors by batch processing had been accomplished prior to the rise of semiconductor integrated circuits. Such mesa transistors are made in an array covering a slice of silicon material of the order of one inch in diameter. Typical arrays of double diffused silicon transistors are shown in Fig. 5-1. Such a wafer may contain as many as 1,000 transistor structures. These mesa transistors are completed electrically while still in wafer form. They are positioned in a precise geometric array. The manufacturing operation beyond this stage consists of separating these structures, mounting them and selling them to a customer who proceeds to reconnect them to form circuits. Since this dicing and assembly is the major expense in making small transistors, one is lead rather naturally to consider performing interconnections between the structures without loss of index in order to decrease total system manufacturing costs. Several problems arise when this is considered.

First the surfaces of mesa transistors are extremely sensitive to ambient, resulting in relatively low yields from devices at the wafer stage to good final products. This problem is aggravated by increasing the number of devices or by attempting to run interconnections over the surfaces. Interconnections by bonding of leads can be accomplished successfully. However, this is a unit-by-unit operation and reinserts the expensive portion of semiconductor device manufacture while decreasing yields because of the increased number of possibilities for imperfections as the complexity of the interconnected unit increases. The next important step in making semiconductor integrated circuitry practical was achieved with the development of the planar transistor structure[2]. The schematic cross-sections of a single planar transistor structure and a mesa transistor are shown in Fig. 5-2. The advantages of the planar structure are intimately tied to the silicon dioxide layer covering the region where the junctions intersect the surface. The oxide layer is an effective barrier to the deleterious effects of the ambient on the junction surfaces. In addition, it supplies a relatively flat surface upon which thin metal intraconnections can be applied. While this oxide layer is actually a micron or less in thickness, it has a dielectric strength ranging to several hundred volts, allowing metal films to pass over

the junction regions without effect. Thus, in the planar structure the interconnection problem for the individual components is solved. In order to make complete circuits, it is necessary to add the capability of making other elements, such as resistors and capacitors as well as to achieve the required electrical isolation of these components. This chapter will discuss the technology and device structures available for the manufacture of semiconductor integrated circuits and will try to define the limitations and potentialities of the technology.

5-2. THE TECHNOLOGY

The critical role played by the available technology in determining the applicability of microcircuitry in a given application cannot be over-estimated. It determines what can be done at all as well as what can be done economically. Using known physical phenomena integrated solid state structures can be conceived which would perform almost any useful electronic function presently performed by conventional electronics. Only a small subclass of this class of structures is accessible because of the limitations imposed by the available technology. Thus, the problem in making functional electronic blocks does not reside merely in the conception of structures which, if made, would perform the function but rather in the conception of a structure which is economically accessible

through the available technology. The technological capability determines into which class the structures will fall -- if it will be a manufacturable item suitable for design into operating equipment; if it will be a structure whose "feasibility" can be established by the construction of a few samples by highly trained laboratory technicians; if it will be a one-of-a-kind structure made through the extensive labors of engineers adjusting and selecting until operation was finally achieved, suitable for description in a technical publication or contract report; or if it will fall into the class of gedanken structures, unmakable, but with principals of operation consistent with established physical laws.

Each major addition to the pool of technologies has a tremendous fan-out in that many more circuit functions can be made. As well as major additions, such as epitaxial growth, important contributions can be made by many processing changes which result in appreciable increases in yield. A decrease of a factor of two in the shrinkage at a particular processing operation can result in the doubling of the complexity of circuits which can be achieved economically. In this section, some of the more important technology areas associated with semiconductor integrated circuitry will be described.

Materials Preparation

The material used for microcircuitry is essentially

identical with the silicon ordinarily employed in transistor manufacture. It is available commercially as pure polycrystalline silicon, as doped single crystals and as epitaxial wafers. Since the technology for preparing crystals of silicon suitable for microcircuitry is well established, it will not be discussed in detail here. Sufficive it to say that both Czochralski crystals pulled from a crucible of molten material and zoned melted material can be used. Slicing with a diamond impregnated saw and lapping with silicon carbide abrasive is straightforward and wide spread throughout the semiconductor industry. Any differences in material preparation which exists that are significant occur subsequent to this lapping operation. It is, of course, generally agreed that a high degree of cleanliness is required for the high temperature processing operation, such as diffusion. There is disagreement as to which method of surface preparation results in wafers most suitable for the various processes. It is likely, in fact, that the most suitable surface for a given process will differ from that for a different step in the operation. For example, an etched surface very desirable for diffusion might have sufficient surface irregularities that the precision of the masking operations would be influenced, both by affecting the coating operation and by making poor contact with the mask during exposure. Essentially the choice comes down to

that between etched surfaces and mechanically polished surfaces. Etched surfaces are generally considered to be more nearly perfect on a microscopic scale, while on a macroscopic scale a well mechanically polished wafer is superior. As usual with two alternatives of this sort, one can think of combinations of the two processes which he hopes will supply the advantages of each. For example, electrochemical etching can be employed using an extremely thin film of electrolyte[3]. This results in selective dissolution of the high points on the wafer, resulting in a flat surface comparable with that achieved with mechanically polishing, while the surface itself should have the perfection expected from an etched surface. However, this technique is relatively tedious and involves such undesirable problems as attaching electrodes to the wafer so that electropolishing current can be passed. At the other extreme, mechanically polished wafers can be etched slightly either directly or by such techniques as oxidation, followed by removal of the oxide, a technique which results in an extremely uniform removal of the damaged surface layer preserving the flat surface. The simplest surface preparation employed consists merely of etching the wafers after they have been cleaned subsequent to lapping in a polishing chemical etch, usually consisting primarily of nitric acid and hydrofluoric acid, occasionally with an addition of a

moderator, such as acetic acid. For compositions relatively rich in nitric acid, a non-selective etching occurs which results in a shiny surface blemished only by a relatively uniform orangepeel effect.

For epitaxial growth the surface preparation is perhaps even more critical than for the diffusion and masking operations. Here again, one is torn between etching and mechanical polishing. In order to allow the reader some latitude in which to exercise his creativity, we will recommend neither method above the other.

Diffusion

This is presently the most convenient, flexible and well controlled method to obtain the impurity concentration distributions necessary for device fabrication. An extensive literature has developed concerning diffusion in semiconductors[4]. Study of diffusion in semiconductors has lead to a greatly improved understanding of the details of the diffusion process itself.

For device fabrication in silicon, one is primarily concerned with the diffusion of impurity elements from the third and fifth groups of the periodic table. These impurities, in general, enter the lattice substitutionally at the regular silicon lattice sites. The diffusion mechanisms consists of the jumping from one lattice site to an adjacent vacancy. Such substitutional diffusion in crystals

like silicon requires temperatures approaching the melting point to proceed at reasonable rates. Some other impurities of interest diffuse by an interstitial mechanism. These in general are much more rapid diffusers. Examples of the latter class are gold and nickel in silicon. In order for diffusion to occur it is necessary that a concentration gradient or more precisely a chemical potential gradient be established. Several treatises on the diffusion mechanism itself exist[5].

In device fabrication one is usually concerned with the diffusion from a surface of the semiconductor material. Two boundary conditions are of special interest in the diffusions employed for device manufacture. The first of these is where the concentration of the impurity on the surface is maintained constant throughout the diffusion process. In this case, a useful approximation to the concentration profile obtained is given by the complementary error function distribution.

$$C(x) = C_o \left[1 - \frac{2}{\sqrt{\pi}} \int_0^{x/2\sqrt{Dt}} e^{-\lambda^2} d\lambda \right] = C_o \operatorname{erfc}(x/2\sqrt{Dt}) \quad (5\text{-}1)$$

where C(x) is the concentration a depth x below the surface, C_o is the surface concentration, D, the diffusion constant, is the constant of proportionality between the diffusion flux and the concentration gradient, and t is the

time. The other case of special interest is where a fixed amount of impurity Q is deposited on the surface at the beginning of the diffusion and the total quantity Q is held constant throughout the diffusion. Under these conditions, the useful approximation is the gaussian distribution given by:

$$C(x) = Q(\pi Dt)^{-\frac{1}{2}} \exp(-x^2/4Dt) \qquad (5-2)$$

The above relations hold fairly well for deep diffusions greater than several microns or for low concentrations. With shallow diffusions and in high concentrations deviations often appear. Reasons for these deviations can be understood qualitatively, but because they are often dependent on the specific parameters of the diffusion system under consideration it is desirable to allow for them empirically. The reasons for these deviations include such factors as the dependence of D upon concentration, the time required to obtain C_o or Q at its final value, the fact that the surface is usually moving during the diffusion process, either by evaporation or oxidation, and interactions between impurity atoms.

Diffusion Systems. Three different types of diffusion systems must be considered. In closed tube diffusion the semiconductor wafers are sealed in an ampule with the impurity source. In vacuum diffusion a controlled pressure of the diffusant impurity is established in a vacuum

chamber and diffusion proceeds inward while simultaneously the semiconductor material evaporates off. In open tube diffusion an atmosphere is established to produce the desired boundary conditions by flowing gas over the semiconductor wafers.

With silicon in the closed tube system, one usually encapsulates the wafers and source in a fused silica ampule. The concentration of impurity on the silicon surface can be controlled in a variety of ways. One can limit the total quantity of impurity added, one can use with an alloy of the impurity in some neutral diluent such as lead or tin to reduce the equilibrium vapor pressure of the impurity, or one can use a doped silicon source. In the last case, the impurity concentration is determined by the relative surface areas of the sample to the total surface area of silicon, assuming that the volume of the ampule is relatively small and that its surface does not act as a sink for the impurities. The advantage of the closed tube system is flexibility. A single diffusion furnace can be used for a large number of different diffusions with no problems of cross-contamination. The disadvantages of the technique for device manufacture, however, greatly outweigh this advantage of flexibility. It is difficult by this technique to process large batches of material; the sealing process is annoying at best; and

it does not allow one to take full advantage of oxide masking techniques for geometry control.

Vacuum diffusion has not been used extensively for device manufacture. Fig. 5-3 shows an apparatus employed for the simultaneous vacuum diffusion of phosphorus and gallium into evaporating silicon[6].

A dynamic vapor pressure of gallium and phosphorus is established by controlling the temperature of the reservoir of the elements. These are adjusted to give the desired surface concentration. The entire jig is heated by the r.f. field to the diffusion temperature. A vacuum is maintained with the liquid nitrogen trap and pumps beyond the trap. The tantalum jig as well as supporting samples contains heat shields to help maintain a uniform diffusion temperature. The vapor of gallium and phosphorus pass from the sources up past the silicon wafers establishing a concentration dissolved in the surface layer of the silicon. The surface of the silicon evaporates at a constant rate when all the oxide has been eliminated.

An interesting feature of vacuum diffusion is that because of the constant evaporation rate of the semiconductor material one approaches a steady state condition wherein the impurity profile is exponential with depth and independent of time. This suggests a process wherein control of time and temperature need not be especially

precise to obtain a reproducible profile. However, the expense of vacuum diffusion apparatus combined with the disadvantage that oxide masking is not possible has limited it to a laboratory technique, useful in the study of the fundamental diffusion processes.

Open tube diffusion is the generally employed method for the fabrication of diffused silicon device structures. The remainder of this section will be concerned with this diffusion technique. The principal advantages of the open tube technique are its applicability to large-scale batch processing and the use of oxide masking to control the geometry of the diffused regions. Fig. 5-4 shows a photograph of open tube diffusion furnaces used for production of silicon devices.

There are many variations to the open tube process depending upon the impurity source and the desired impurity concentrations. It is usually convenient to employ a two-step diffusion process wherein the impurity is deposited on the surface or in a shallow surface layer during the first step and diffused to the desired depth during a subsequent operation. The reason for the preference of the two-step process is that surface damage to the silicon wafers results at the high diffusion temperatures necessary to get deeper diffusion in a non-oxidizing atmosphere. On the other hand, if the atmosphere is

made oxidizing the formation of SiO_2 masks at least partially against all common group three and group five impurities, except gallium. The simultaneous oxidation competing with the source deposition results in a relatively uncontrollable process. The lower temperatures ordinarily employed in the first step or predeposition process are below those where appreciable surface evaporation takes place and, accordingly, can be carried out in a non-oxidizing atmosphere. In general, the idea during predeposition is to establish an atmosphere containing the desired impurity surrounding the surface of the semiconductor to be diffused. Ideally one has an equilibrium between the volatile species in the surrounding atmosphere and the dissolved species on the silicon surface. The atmosphere can be established in a variety of ways which will be discussed below for the specific impurities. Ordinarily, as shown in Fig. 5-5 the silicon wafers stand in a silica boat, although there are cases where laying down flat on a silica plate seems to improve the uniformity of the diffused layer.

Donor Impurities. Phosphorus, arsenic and antimony are the important donor impurities in silicon. The diffusion constants[7] and solubilities[8] of these impurities as a function of temperature are shown in Fig. 5-6 and Fig. 5-7, respectively. Arsenic and antimony

are relatively slow diffusers compared with phosphorus. All three impurities are commonly employed, although phosphorus is the most common.

Phosphorus Diffusion: Phosphorus diffusion is done most conveniently by a two-step process. Red phosphorus, P_2O_5, $POCl_3$ and PCl_3 are the usual sources to consider. Red phosphorus and P_2O_5 are solids at the source temperatures employed. Fig. 5-8 shows a schematic diagram of a diffusion system suitable for use with P_2O_5 or red phosphorus. This is a two-zone system where the silicon wafers are held at the temperature desired usually of the order of 1000°C to 1200°C and the source is held at a considerably lower temperature, in the vicinity of 200°C for P_2O_5. The temperature increases monatomically from the source furnace to the silicon wafers so that material evaporated at the source will not condense on the furnace walls. A carrier gas such as argon or nitrogen passes over the source, ideally becoming saturated with the vapor of the source at the source temperature, and carries through a quartz wool filter to remove any particles picked up by the flowing gas stream and passes over the silicon wafers and out the exhaust vent. The quantity of phosphorus deposited on the silicon wafers is roughly an exponential function of source temperature and depends upon the silicon

wafer temperature and the time of deposition. P_2O_5 is a convenient source and works well if kept dry. The principal problem concerning P_2O_5 is moisture pick-up while loading the boat and transferring it to the source furnace. Relatively small quantities of moisture appreciably effect the apparant vapor pressure from the source. By controlling the source temperature with dry P_2O_5 one can perform phosphorus diffusions with surface concentrations from below 10^{16} to above 10^{20} impurity atoms/cc. This wide range of available concentrations makes P_2O_5 an extremely flexible dopant source. Red phosphorus is unreliable as a source, since at the temperatures required to obtain appreciable vapor pressure it gradually converts to another form of phosphorus with a corresponding decrease in the equilibrium vapor pressure. $POCl_3$ and PCl_3 are low boiling liquids. A suitable system for the use of these source materials is shown diagramatically in Fig. 5-9. An inert gas is saturated with the vapor or liquid and introduced into the diffusion furnace. The amount of impurity introduced can be controlled either by controlling the temperature of the impurity source or by controlling the fraction of the gas stream that gets saturated from the source. With the halogens present, there is always the danger of pitting of the silicon surface by chemical reaction to form silicon tetrahalide.

This is an especially important problem with PCl_3. Many other compounds of phosphorus can be employed but results obtained with P_2O_5 or $POCl_3$ are adequate for device manufacture. After predeposition the diffusion to achieve the desired depth can be performed in any oxidizing atmosphere depending upon the amount of oxide necessary to prevent surface damage by evaporation and upon the thickness of the oxide layer desired for subsequent processing. Dry oxygen, wet oxygen, wet nitrogen and pure water vapor are common oxidizing atmospheres.

Arsenic Diffusion: As_2O_3 is a convenient source for the diffusion of arsenic. It is used in a manner very similar to P_2O_5.

Antimony Diffusion: For antimony Sb_2O_3 is a convenient source material, although it has appreciably lower vapor pressure than either P_2O_5 or As_2O_3. Source temperatures using Sb_2O_3 are generally 700°C to 1000°C. Where separate source and diffusion furnaces are employed, one must be especially careful to maintain a monotonically increasing temperature. A more convenient system for antimony employs a single furnace with separately controlled source. At high concentrations of antimony there appears to be a new phase formed on the surface that acts as a strong source. This makes it inconvenient

to control antimony diffusions above about 10^{19} impurity atoms/cc. There is also a tendency to get pitting during antimony diffusions that can be very troublesome. Like arsenic it is an extremely slow diffuser. This is often the principal reason for choosing arsenic or antimony instead of phosphorus.

Acceptor Impurities. Boron and gallium are the most important acceptor impurities in silicon. The diffusion constant of these and of aluminum are included in Fig. 5-6 as a function of temperature[7]. The curve for aluminum is also included. Occasionally because of its extremely high diffusion coefficient, aluminum is worth considering for special structures. The solid solubility versus temperature is also shown for these impurities in Fig. 5-7[8]. Boron is the only one with adequate solubility for use as an efficient emitter in transistors or for other applications requiring extremely high impurity concentrations. Boron diffuses at about the same rate as phosphorus. Gallium diffuses somewhat more rapidly.

Boron Diffusion: Boron is the most extensively employed donor diffusant because of its high solid solubility and because it is oxide masked. There is, however, no general agreement on the system and impurity sources between various groups. B_2O_3 when used in a system similar to that shown for the group V oxides in

Fig. 5-8 results in a variety of problems, most of which are associated with the fact that B_2O_3 and SiO_2 are missible. After a period of time, the fused silica tube gets extremely sticky from the B_2O_3 resulting in difficulty in removing sample boats and the cracking of furnace tubes if they are cooled to room temperature. The dissolution of the boron in the silica also results in establishing a reservoir of boron in the system that ends up as the controlling dopant source. Successful use of B_2O_3 has been achieved in the box technique[9] shown in Fig. 5-10. In this method, the wafers and dopant source, usually a mixture of SiO_2 and B_2O_3, are placed in a box made of platinum or graphite which is closed on top with a loose fitting cover. The box is then placed in an ordinary open tube diffusion furnace. This method is somewhat similar to the closed tube technique, although it avoids the sealing and evacuation operations.

BCl_3 and BBr_3 are employed as sources. BCl_3 is a gas at room temperature and can be purchased as the liquid under pressure in cylinders. BBr_3 is a low boiling liquid. BCl_3 can be metered directly into the gas stream and apparatus similar to that used for phosphorus halides shown in Fig. 5-9 is suitable for boron tribromide. Here again both of the halides have a tendency to pit the silicon wafers. This is especially pronounced

in the case of BCl_3. This effect can be minimized by adding a small quantity of both hydrogen and oxygen (of the order of 1%) to the inert carrier gas. When there is too little hydrogen, severe pitting occurs while with inadequate oxygen, black deposits, often found during boron diffusion, form on the surface of the wafers. The addition of water vapor is not a substitute for adding both hydrogen and oxygen since the water vapor will hydrolize the boron halide at low temperatures resulting in B_2O_3 deposits on the furnace tube. The nature of the black deposit which forms with inadequate oxygen or excess BCl_3 is somewhat variable. It contains a large percentage of elemental boron but also can involve silicon and oxygen. The layer is chemically very inert. When it is necessary to find a material from which to make a container for the universal solvent, it should be a strong contender. Actually, in the boron halide diffusions using hydrogen and oxygen it may well not be the halide by the time the gas stream arrives at the silicon wafers. Upon looking down a boron furnace operating to achieve relatively high concentrations of boron, one can see solid particles or liquid droplets as a fog in the furnace tube. These are likely to be particles of B_2O_3.

Another method of introducing a boron source is by the controlled oxidation of methyl borate. A small quan-

tity of $(CH_3O)_3B$ is metered into the gas stream with a quantity of oxygen. These react before reaching the wafers to produce a smoke that is probably boron oxide. The other combustion products such as CO_2 do not interfere with this process. Many other boron compounds have been tried, such as diborane, B_2H_4 and penta borane, B_5H_{10}. To the best of the author's knowledge none are completely satisfactory.

Gallium Diffusion: Gallium diffusion is unique because it is best done as a one-step process[10]. This occurs because at the diffusion temperature SiO_2 has no measurable effect upon the diffusion of gallium into the underlying silicon. The only important source for gallium diffusion is the refractory oxide Ga_2O_3. Gallium diffusion is done in a two-zone furnace where the Ga_2O_3 source is held between about 800°C and the diffusion temperature. The gas stream contains a quantity of hydrogen necessary to reduce the Ga_2O_3 to form a volatile species. The reaction at the source is thought to be $H_2 + Ga_2O_3 \rightleftharpoons Ga (\text{or } Ga_2O) + H_2O$. It is relatively easy to approach equilibrium at the source in this system. Hence, the pressure of the volatile species, Ga or Ga_2O, is controlled by the ratio of the pressures of hydrogen to water vapor. This results in an excellent method of source control. Combined with the fact that silicon

dioxide does not mask the diffusion, gallium diffusions are extremely uniform and well controlled. The inability to employ oxide masking, however, is a serious drawback in many applications to microelectronics.

Oxide Masking. Frosh and Derrick[11] found that SiO_2 is an effective barrier for preventing many impurities from reaching the underlying silicon. This is an extremely important invention with respect to the construction of complex structures. A fraction of a micron of SiO_2, which is easily grown thermally, prevents phosphorus, arsenic, antimony and boron from penetrating. In this manner, the extent of diffused areas on the surface of a silicon wafer can be controlled by oxidizing the wafer and exposing only those areas where diffusion is desired. The oxide for masking can be grown conveniently in a furnace similar to the diffusion furnaces. Either oxygen or water vapor is a suitable oxidizing agent. The properties of the oxides in both cases are very similar although minor differences have been reported[12]. The growth rate is parabolic in time; that is, dx/dt is proportional to $x^{-\frac{1}{2}}$, where x is the oxide thickness. Fig. 5-11 shows oxide thickness versus time with temperature as a parameter for pure dry oxygen at one atmosphere pressure and similar data for pure water vapor at one atmosphere is shown in Fig. 5-12.

The rate of oxidation is essentially independent of the doping in the silicon except at the very highest impurity concentrations such as those which occur on the silicon surface in the emitter regions of transistor structures. At these extremely high impurity concentrations the rate of oxidation is somewhat higher. When a multiple oxide masking is desirable, as for example when one wishes to make a planar transistor by diffusing in a base through a hole in the oxide and follow this by the diffusion of an emitter into a portion of the area of the base, it is usually convenient to adjust the oxidizing atmosphere in the previous diffusion to achieve a regrown oxide of the desired thickness for masking against the next diffusion. This will be more clearly illustrated in Section 5-4. Silicon dioxide layers can be prepared by other methods which for some special applications offer advantages. Anodic oxidation results in a relatively porous oxide whose principal advantage is that it can be prepared at room temperature. Oxide films formed by the pyrolytic decompositions of siloxanes, such as $Si(OC_2H_5)_4$, offer the advantage of not removing material from the silicon substrate. Since the pyrolysis proceeds readily at temperatures of the order of 750°C the oxide can be deposited with negligible diffusion of impurities in the silicon[13]. It is difficult to achieve as good a uniformity with the pyrolytic oxide as the anodic oxide

although its masking problems are similar to those of the thermally grown oxide. The rate of removal of pyrolytic oxide and also of anodic oxide in fluoride-containing etches is considerably greater than for the thermal structure. Oxidation in high pressure steam can be made to proceed at relatively low temperatures[14]. It requires an autoclave capable of operation to above the critical pressure of water. Under these conditions, the growth rate of the oxide becomes linear with time and, accordingly, should be good for the production of thick oxides. Since the temperature is relatively low, the problem of relative expansion coefficients of the oxide and silicon which results in cracking of thick oxides prepared in other ways should be minimal. In general, oxygen or wet inert gas are the preferred methods for making the oxide. One can make adherent layers up to approximately 2 microns thick before cracks and other flaws appear. If an oxidized wafer has part of the oxide removed and then is reoxidized the new oxide which grows against the original layer is well bonded and serves as an effective diffusion mask. The thickness of oxide necessary to mask against the various impurity diffusions is certainly somewhat system dependent. As an example, Fig. 5-13 illustrates the oxide thickness required to prevent the penetration of phosphorus[15].

Evaluation of Diffused Layers for Impurities. For

diffusions of donors into p-type or acceptors into n-type obeying one or the other of the distributions given by equations 5-1 and 5-2, one need make only two measurements to evaluate the diffusion length, $(Dt)^{\frac{1}{2}}$, and either the surface concentration C_o or Q, the total quantity of impurity atoms/cc. The most convenient measurements to make are those of sheet conductivity of the diffused layer, and the concentration at one depth, usually the concentration at a junction where the concentration of the diffused species is equal to the original doping of the substrate. The sheet conductivity is conveniently measured by means of a conventional four point probe. In the case of diffused layers the probe spacing in general is large compared with the thickness of the layers leading to the approximation where the measured voltage drop divided by the current, V/I, is independent of probe spacing[16]. This measured resistance is related to the sheet resistivity σ of the diffused layer by $\sigma \cong 4.3$ V/I. Junction depth measurements are made conveniently by angle lapping throughout the layer or by grinding a cylindrical groove[17]. In either case, the junctions are delineated by one of the variety of techniques, the most common of which is staining with a solution of hydrofluoric acid containing a fraction of a percent of nitric acid. This mixture, when applied under strong illumination, preferentially stains p-type material

(and very heavily doped n-type). The low angle bevel or the cylindrical groove gives a mechanical magnification which can be employed to convert the apparent length of the stained areas to junction depth from simple geometric considerations. However, it is very convenient to employ an interference microscope which superimposes upon the stained pattern an interference pattern so that the depth from the original surface can be determined by counting interference fringes. Fig. 5-14 shows a photomicrograph of a stained device structure taken through a vertically illuminated microscope employing a sodium lamp as a light source with an ordinary microscope slide inserted over the specimen. Fringes obtained with this sample setup are adequate for most diffusion evaluation. In this particular sample, diffusion was performed from an etched surface. The unevenness of such a surface results in an uncertainty of the order of one fringe in the measured junction depths. For more precise measurement it is necessary to employ more nearly flat original surfaces. By this technique one fringe corresponds to one-half of the wavelength of the light or with sodium light to approximately 3000 Å.

Diffusion of Impurities Other Than Group Three and Group Five. Occasionally it is desirable to diffuse other impurities than the common donor and acceptor elements. For example, it is common to diffuse gold into silicon

device structures in order to increase the carrier recombination rate. This is important to improve diode recovery times and to decrease the storage of minority carriers in the collector regions of npn transistors. Diffusion of gold, which goes by interstitial mechanisms, is rapid but does not follow one of the simple approximations[18]. It is preferable to establish the desired diffusion schedule for gold by empirical measurement of the parameters one wishes to control. The usual way of diffusing such impurities is to deposit them on the surface of the wafer either the front or the back by evaporation or chemical plating techniques. The wafers are then diffused for the appropriate time-temperature cycle. Since the solubility of these impurities at the diffusion temperatures is usually considerably higher than at lower temperatures it is necessary to cool relatively rapidly to prevent precipitation on the surfaces or at defects in the semiconductor material.

Miscellaneous Effects. A variety of phenomena occur when various combinations of oxidations and diffusions are performed. These are often specifically related to the device structures being considered. As a few examples of the effects with which one must contend, we give the following. This list is by no means exhaustive. When oxide is being grown on a wafer doped with one of the common donor elements there is evidence that the growing oxide rejects the impurity

originally contained within the bulk semiconductor. This results in a concentration of impurity near the surface of the semiconductor. This so-called "snow-plow effect" results in the lowering of breakdown voltages on the surface of structures diffused in under the oxide. It is possible, although usually not worthwhile, to eliminate the affects of this effect by making guard ring structures, where a first diffusion forming a higher breakdown voltage is done in the form of a ring surrounding the diffused area where the desired junction structure is subsequently prepared. As an example of the interaction of impurities, Fig. 5-15 shows a cross-section through an npn structure made with a gallium base diffusion. One notices that the base collector junction has penetrated more deeply under the emitter than over the rest of the surface. It can be shown by spreading resistance measurements under the emitter that this additional diffusion is not the result of rejection of impurities by the region of high phosphorus concentration, but rather is the result of an increased diffusion constant for gallium in the region near the collector junction. This is very similar to the result achieved upon high energy electron bombardment of a diffused layer[19]. The explanation presented in the case of the electrons was that bombardment increased the vacancy concentration beneath the surface and hence increased the diffusion constant. It is possible

that strain from the extremely high phosphorus concentration at the surface results in an increased vacancy concentration below the emitter. The same enhanced diffusion under the emitter can be observed in structures with boron as the base impurity. However, because of the lower diffusion constant of boron it is less pronounced. In pnp structures, where heavy boron emitters are employed, no such enhanced diffusion at the collector junction is apparent.

In the case of boron in particular during diffusion in an oxidizing atmosphere the total quantity of impurity atoms in the silicon is decreased. This is because the boron is soluble in silicon dioxide so that boron on the surface is incorporated into the growing oxide interface. This loss of boron by outdiffusion results in large deviations from the simple diffusion models, especially when trying to predict the concentration profiles of deep layers from the depth and sheet conductance of shallow predepositions. The converting effects of oxide growth outdiffusion and indiffusion are too complex to be taken into account analytically for routine control. Empirical curves again are strongly recommended.

Epitaxial Growth

As we will use here the term epitaxial growth refers rather restrictively to the process by which a film of semiconductor material is deposited upon a substrate of the same

semiconductor by a gas or vapor process. While this is not really consistent with the original idea of epitaxy, it can be considered such from the point of view that the doping in the deposited film is usually considerably different from that in the substrate for structures of interest in microcircuitry applications. Epitaxial growth is a relatively new addition to the semiconductor technology and as such has not yet been exploited to the extent of such well developed technologies as solid state diffusion. Microcircuits containing epitaxially grown films are just at the laboratory stage at present. This technique offers what appears to be an extremely flexible method of obtaining impurity concentration profiles which are not accessible by any of the other established techniques. As such, it certainly has an important future in microcircuitry. The technology is in a state of very rapid development. In this state of flux, any description of its present standing will be obsolete in a short period of time. Accordingly, we will present a rather brief and general review of the methods and usefulness of the technique.

As in the case of diffusion, several different systems have been employed for epitaxial growth. It has been done by vacuum evaporation of silicon onto a silicon heated substrate[20]; it has been done in a closed tube system wherein a halogen is used to transfer semiconductor from a low temperature load to the substrate at a higher temper-

ature load to the substrate at a higher temperature[21]; and open tube systems have been employed wherein silicon is deposited by the reaction of silicon tetrachloride or other chlorinated silanes with hydrogen upon the surface of a heated substrate[22]. The reaction will proceed readily in any of a variety of systems providing only that the temperature is in the correct range and that the composition of the gas mixture is appropriate. Single crystal epitaxial growth of silicon on silicon has been reported over the temperature range from 930°C to 1400°C. In general, in order to obtain single crystal growth at the lower temperatures it is necessary to employ relatively low growth rates. At temperatures above about 1200°C growth rates of the order of one micron per minute up to 5 microns a minute result in useful single crystal films. Growth at higher rates leads to increasingly imperfect films and finally to polycrystalline deposition. The growth rate is determined primarily by the concentration of chlorinated silane in the gas feed. Usual mole fractions in the gas are of the order of 0.01. The temperature dependence of deposition rate is system dependent. According to Tung[23], the temperature dependence of the reaction corresponds to an activation energy about 25 kilocalories per mole, while Corrigan[24] claims that in his apparatus no apparent dependence upon temperature is observed. In most systems there seems to be a broad

range of temperature in the vicinity of 1150 to 1300°C where the effective growth rate is independent of temperature to a good approximation. Since this is also the desired temperature range in which to work to grow films of good quality, the importance of careful control of the substrate temperature is minimized.

Both hot-wall, resistance heated furances[25] and cold-wall rf heated reactors have been employed[26]. These include a wide variety of geometry and flow configurations. A fairly typical system is diagrammed in Fig. 5-16. It shows an rf heated reactor where the silicon substrate wafers are placed upon a silica coated graphite block to which the rf is coupled The silicon tetrachloride is contained in a thermo-stated silica flask through which hydrogen can be bubbled to saturate it with the vapor. Part of the flow of pure, dry hydrogen goes through the silicon tetrachloride reservoir with the majority going directly to the reactor. By adjusting the relative flows the partial pressure of the silane can be controlled to any desired value. This particular apparatus also has provisions for other silane bottles which are doped with boron tribromide and phosphorus trichloride. By putting a portion of the hydrogen flow through these doped sources of silane,.it is possible to prepare films having n or p-type impurities. At the exhaust, provision is made to assure a positive flow through the system by inclusion of an oil

bubbler trap. A water spray is shown for disposal of the silane. The hydrogen should be adequately vented, since relatively large quantities are employed. Both silicon tetrachloride and trichlorosilane hydrolyze violently. Care must be exercised in handling these materials to be sure that they do not come into contact with any appreciable quantities of water.

In general, the properties of the film with which one must be concerned can be divided into three categories; thickness, doping profile and perfection. The variables influencing each of these properties can be considered independently. The thickness of the film is primarily determined by the growth rate which depends upon gas composition and, to some extent, the temperature of the substrate and of the gas stream itself. In order to achieve a uniform growth rate over the entire area of silicon exposed, it is necessary that this area be subjected to a uniform gas composition. Because the reaction of hydrogen with the silicon tetrachloride proceeds rapidly and in several steps, this requirement of uniform gas composition is difficult to maintain. In fact, starting with silicon tetrachloride through a single small reaction zone, it has been shown that a large fraction of the affluent gas, of the order of 20%, can be converted to trichlorosilane with appreciable concentrations of other less heavily

substituted silanes as well[22]. This being the case, it is important that for uniformity only a small fraction of the total available starting material be consumed during its residence time in the reactor. In addition, it is necessary that the flow be as uniform as possible across the surface area. Achieving the desired flow geometry is pretty much a matter of trial and error. Not only is it necessary to concern oneself with the uniformity from wafer to wafer, but it is equally important to see that adequately uniform growth is obtained across the surface of a given wafer. Depending upon the flow pattern, it is possible to get a thicker deposit in the center of the wafer than at the edges or even the opposite condition of thicker at the edges of the wafer than in the center. A very common type of non-uniformity is a wedging of the film from one side of the wafer to the other in cases where the gas flows across the wafer surface. When all of the flow parameters are optimized, it is possible to obtain growth uniform to better than 10% over an entire group of twenty wafers or more.

The impurity concentration in the films can be controlled by doping the silane source as is suggested in Fig. 5-16. For some of the structures of interest in microcircuitry this control will prove to be quite satisfactory. In the case of growing relatively lightly doped films on heavily doped substrates, however, there is a

considerable transfer of dopant from the substrate into the growing films[27]. It is possible to grow n-type films with impurity concentrations in excess of 10^{16} atoms/cc by using intrinsic silanes upon very heavily doped substrates, of the order of .005 Ω cm. This transfer of dopant from the substrate to film can also involve the remainder of the system. In particular, any silicon or exposed graphite can act as a reservoir of impurity. The affect can often be minimized by inserting a backing wafer under the substrate wafer itself. In this manner silicon from the backing wafer transfers to the substrate wafer during heating in the halogen containing atmosphere and serves to lock the impurity atoms in the substrate. An alternative method employs a quartz plate which fits tightly to the silicon surface on the back of the substrate. This prevents appreciable gas diffusion from the back and hence, minimizes transfer. One can also use a thermally grown oxide layer on the silicon itself to minimize the effect of transfer from the substrate. This gas-phase transfer can be appreciably more important than straight diffusion of the substrate impurity into the growing film. In all cases where heavily doped substrates are employed, one should test for this transfer of dopant.

Other ways of introducing desired impurities into the growing film have been tried. Small quantities of PH_3

or B_2H_6 can be added to cylinders of hydrogen for use as doping materials[28]. One of the principal advantages of this direct gas feeding is that it results in a low inertia system wherein one can change dopants rapidly, which is very important in the preparation of multi-layer structures. Another method of feeding consists of the metering of the liquid silane into a flash evaporator where it is picked up by the hydrogen gas stream. This is a positive method and eliminates the change in doping level of a reservoir of silane as it is fractionally distilled by hydrogen bubbling through it in the sort of system shown in Fig. 5-16.

The concentration of impurity in the film is not, in general, the same as in the original silane, since the impurity compounds have reaction rates different from the silane itself. In general, however, over a wide range of compositions, the concentration in the film is proportional to the concentration in the silane if the growth rate is held constant.

Many defects can appear in the growing film. These can usually be related to their cause from their general appearance. "Chevrons" as shown in the left portion of Fig. 5-17 indicate too rapid a growth rate for the particular temperature being employed. "Pyramids" as shown on the right side of this figure are a defect which results from a number of causes including impurities in the

system and flaws on the substrate. The number of dislocations in the film is never less than that in the original substrate. On occasion, it can be several orders of magnitude higher. There are usually stacking faults which can be observed by lightly etching the surface of the film with a pitting etch. These have been proposed as a useful feature for the measurement of film thickness[29]. Polycrystalline film regions indicate the temperature is much too low for the particular growth rate employed or that gross contamination exists.

The preparation of the surface prior to epitaxial growth is an extremely important consideration. Both mechanically polished and chemically etched substrates can be used. In either case, it is usually preferable to give an additional *in situ* etching either by soaking at high temperature in a pure hydrogen atmosphere or by gas-etching with hydrogen chloride. The surface of the film reproduces quite well that of the original substrate. Scratches come through as scratches, although after the growth of a relatively thick film, the sharp features have been smeared out. Any orangepeel effect on the substrate resulting from chemical polishing is reproduced in detail.

The thickness of epitaxial films can be measured by weight gain, providing the system is so arranged that growth occurs only on a known surface area; by beveling

and staining, since the junction between the substrate and the film can usually be delineated by light acid etching stains; and in the case of relatively highly doped films on heavily doped substrates of resistivity of about 0.01 ohms/cm or lower, the film thickness can be measured by observing Fabry-Perot fringes in the infrared[30]. These fringes result when the light reflected from the interface interferes with that reflected from the surface. This last measurement technique will usually differ from that obtained by beveling and staining, since it really measures to the region in which the refractive index of the silicon is changing rapidly. Usually because of diffusion this is a micron or so into the film itself. There is also the direct method of measurement wherein one measures the increase in thickness of the wafer with a precision mechanical measuring device. Of these several techniques only the bevel and stain is destructive. For the cases where there is a junction between the substrate and film, four point probe measurements on the film give information concerning the average doping. However, where no such junction occurs, that is where n-layers are grown on n-substrates or p upon p, this technique cannot be employed. A good measure and one that is directly relatable to the electrical properties desired, is obtained by measuring the breakdown voltage

and capacitance of a diode junction structure made near the top of the film. This technique is, however, time consuming and destructive. It has been proposed to use measurements of the Seebeck voltage in order to obtain doping levels[31]. This, however, has not been reduced to a standard technique. In any case, there is virtually always an impurity gradient in the films. This can be evaluated by measuring the capacitance of a diode at a function of voltage or by an interesting technique proposed wherein one employs the diffusion of an impurity of known diffusion constant from a surface where a bevel has been cut through the film. By means of a second bevel and staining one establishes a doping profile of the film in the region where the gradient is steep.

Silicon dioxide masks against epitaxial growth at certain temperature ranges. With proper control, one can get the deposit to go down selectively on the exposed areas of the silicon while leaving the silicon dioxide film free of deposits and intact. As greater ability to make oxide masked structures develop and as diffusion and epitaxial growth are more closely interwoven, the range of structures which will be accessible for microcircuitry will expand tremendously. The importance of epitaxial growth in the microcircuitry will certainly be much more than in the case of ordinary transistors, where the prime

function is that of reducing the effect of the large superfluous body of collector material necessary to give mechanical rigidity and for ease in handling in the ordinary double diffused transistors.

Surface Geometry Control

The existence of semiconductor integrated circuitry depends upon the ability to make fine scale patterns in all three dimensions. Diffusion and epitaxial growth are used to control the depth of the structures. The surface geometry is controlled by masking operations. These are generally important in two areas. First for making patterns in the oxide layer to serve for diffusion masks and secondly, for the deposition of the intraconnection metallization pattern. It is also important that good geometry control be maintained for any thin film elements deposited upon the substrate surface. The technology used to achieve this surface geometry control was well developed for the manufacture of multiple diffused transistor structures. In the double diffused mesa silicon transistor shown in Fig. 5-2 two precise indexed masking steps are required. The oxide removal step prior to emitter diffusion in the figure must be indexed with the metal removal step. In addition, the masking for mesa etching must conform to the previous steps. However, this operation is usually less precisely controlled and operates with larger tolerances. In going

to the planar transistor structure, the number of indexed masking operations increases to four. As shown in Fig. 5-2 holes are first opened in the oxide prior to base diffusion. Indexed holes above the base regions are then opened prior to emitter diffusion. Additional holes are opened to expose the emitter and base contact region and finally the metallization pattern necessary to achieve contact to the various layers is etched in index with previous maskings. For the high frequency transistors presently in production, the spacing between adjacent structural features is often as small as 0.0005". This means that alignment errors from one pattern to the next cannot exceed a few ten thousandths of an inch. In extending further to make semiconductor integrated circuitry the number of indexed masking operations increases as does the complexity of the patterns which must align. A simple integrated circuit might require five masking steps while more complicated ones often employ six or seven. In the foreseeable future, as more pieces of the available technology are incorporated into single structures, the number of masking steps will undoubtedly rise still further. Thus, surface geometry control represents an extremely important part of the microcircuitry technology.

The most important techniques employed to achieve this geometry control are those of photolithography.

These employ a photosensitive material which can act as a mask against chemical etchants. These are usually low molecular weight organic compounds which polymerize when subjected to ultraviolet radiation. The most common are commercially available compositions sold under such trade names as Kodak KPR, KMER and KPL. Extensive application data on these materials is usually available from the manufacturer. For example, the spectral sensitivity of KPR[32], one of the most common resist materials, is shown in Fig. 5-18. The other method used in achieving geometry control is vacuum evaporation through a mask. Several variations of this exist. It is common in germanium mesa transistor technology to employ double evaporation through a single aperture at different angles to achieve fine scale structures. Other examples of evaporation through masks are given in Chapter 4 on thin film microcircuitry. The remainder of this section will be concerned primarily with the technology of the photolithographic process.

Process

Masks: The first requirement for making fine scale photolithographic patterns is the original mask that will be duplicated on the device surface. These are usually made from photographic emulsions either on glass plates or on flexible film bases. The art work is done as a single

pattern or small number of patterns at adequate magnification so that errors in drawing are minimized. This pattern is then reduced usually through at least two steps to the final dimensions desired. During this reduction, the pattern is transformed into an array of patterns either by a step and repeat procedure or by a multiple lens camera. Fig. 5-19 shows the original art work and the final working masks for a typical pattern employed in microcircuitry. The requirements placed upon the photographic equipment to achieve the desired precision of reduction are extreme. For example, in a pattern one-tenth of an inch across, a slight tilting in the plate holder corresponding to 1/2% variation in the focal length can result in a 1/2 mil size error, which is of the order of the maximum error band allowable from all considerations. Accordingly, rigid and precise equipment is needed for the production of quality masks. Fig. 5-20 shows a typical copy camera employed in the preparation of photographic masks for microcircuitry. Several commercial suppliers for masks presently exist. Care must be exercised to assure that masks have high contrast and minimum pinholes in the dark areas or scratches and other blemishes in the clear regions. Ordinarily these masks are employed to make contact exposure of the photolacquers. They must be protected from damage and discarded when damage becomes

appreciable.

Coating: After making the masks, the first step in performing the photolithographic process consists of coating the surface to be photoengraved with the unpolymerized photosensitive material. The general requirements of the coating are that to be uniform, adherent and pinhole free. The most desirable thickness varies with the particular engraving problem. For highest resolution, thin coatings are desirable. However, thinner coatings are more readily penetrated by etchants and are more inclined toward pinhole-itis, a serious disease. The usual range over films employed in microcircuitry is about 0.5 to 2 microns. This is much thinner than employed in conventional metal etching applications, such as in the preparation of printed circuit boards. The film may be applied by spraying, dipping or by spreading with centrifugal force. For best results the photosensitive materials should be filtered immediately before use. With the more viscous materials, such as KMER, a pressure filter is necessary. After coating, the film is dried thoroughly and baked under infrared lamps or in a ventilated oven. These operations must be carried out in an environment with very little blue and ultraviolet light because of the photosensitivity of the material. Strong yellow light, however, can be employed.

Indexing and Exposure: The exposure of the photosensitive film can be done either by contact printing or by projecting the mask pattern. In either case, it is necessary to have a means for precise alignment of successive patterns. Two methods are employed. Either the masks are aligned precisely in a jig system including index points against which the semiconductor wafer makes contact or the system is arranged so that it is possible to observe the pattern on the wafer surface with light of wavelength outside the range of sensitivity of the film and to adjust the position of the wafer for precise alignment before exposure to ultraviolet light. Fig. 5-21 shows a mask clamped in a frame suitable for contact printing. The mask is one of a set which has been aligned one with another. The two index points on the bottom of the jig frame make contact with a flat edge of the silicon wafer while the circular arc on one side of the circle touches the flat side of the jig frame, thus assuring three point contact. The wafer is pressed against the mask either by a vacuum chucking arrangement or by other mechanical means. By this system any damage to the index surfaces of the wafer, such as chipping during processing result in loss of index. As in all contact printing systems, the necessity of close contact between pattern and film result in damage, both to the film and to the mask from any

dust particles or surface irregularities on either.

Alignment by adjusting the wafer to the pattern can be accomplished by jigging under a microscope or vertically illuminated projector. Means must be established for separating the wafer and mask slightly during the time the alignment adjustment is being made to minimize film damage. This method has the advantage of allowing broken wafers and wafers with chipped edges to continue through the process. This is an especially important advantage in developmental work in microcircuitry where there is a large number of operations increasing the probability of damage to the wafers. It is worth considering as a production technique. It is, however, appreciably slower than a well-jigged mechanical alignment scheme. In either the mechanical or the optical alignment schemes, the wafer and mask can be clamped together and removed to an exposure fixture. It is, however, convenient to build the exposure arrangement into the apparatus. This is readily done in the mechanical jigging by having an ultraviolet source provided with a shutter on the oppostie side of the mask from the wafer. A convenient arrangement has the optical axis at about a 45° angle to the horizontal. In this manner, gravity helps hold the wafer against index points during exposure. In the case of optical alignment under a microscope, it is possible to replace the usual vertical illuminator with the

necessary high intensity ultraviolet source for exposure. This is usually best done by inserting an additional mirror into the system. It is, of course, necessary to remove the objective lens of the microscope in order to cover the entire wafer area. It is usually possible to modify the nose piece of the microscope sufficiently to obtain uniform illumination over a circle at least one inch in diameter.

In the case of projection printing, one images the mask through a high quality lens onto the silicon wafer which is again either located by preset mechanical index points or an arrangement is made so that the image can be observed through an appropriate filter while the alignment is adjusted. Projection printing has the important advantage of preventing those problems caused by contact of the mask illumination and photosensitive coating on the wafers. It does, however, require a very high quality lens to obtain adequate resolution over the entire pattern.

The high intensity ultraviolet light needed for the photo lacquers can be obtained either from a carbon arc lamp or from a high pressure mercury arc. If the light from the source is collimated or if it comes from essentially a point source, one obtains somewhat better definition than with diffused light especially if contact between mask and wafer is not close. However, as in the case of conventional photographic enlarging, collimated light, while improving

the resolution, increases the number of imperfections in the negative which reveal themselves in the print. The exposure time should be adjusted to achieve clean development in the unexposed regions and no lifting in those areas which have been exposed. These photo lacquers are all extremely high contrast materials. There is essentially no gray scale.

Development: As well as proprietary developed solutions sold by vendors of the photosensitive materials, common laboratory solvents can be employed. These include any of the solvents for the monomers, such as methyl cellosolve acetate, trichlorethylene and xylene. These are used by dipping, spraying or in a vapor degreaser arrangement. Best definition with KPR has been obtained by employing a two-stage development process wherein the film is first developed in an active solvent, such as trichlorethylene, followed by a less active one, such as xylene or isopropyl alcohol. Supplier of the photosensitive lacquers are aware of the problems in the semiconductor industry and can offer considerable application information.

Etching: For etching patterns in silicon dioxide, the usual etchant is buffered HF. A typical solution is ten molar NH_4HF_2 brought to a pH of 5 by the addition of HF. The exact composition can be varied appreciably, yielding faster etching rates for more acid compositions. Too low a pH can, however, result in lifting of the

polymerized resist film. For metal films, other etchants in general are needed. With aluminum, a solution of 10% soldium hydroxide used with KPR is satisfactory. With other materials, KMER is usually more suitable since it is more resistant to acid etches and can even stand *aqua regia* for short times.

Stripping: For a material as prone to lift accidentally during etching, it is magnificent in its adherence when no longer desired. Stripping procedures include softening the film by boiling in TCE followed by scrubbing in acetone or by boiling in hot mineral acids such as HNO_3 or H_2SO_4. Hot chromic acid, the classic "cleaning solution" for organic residues will remove the film. However, it is well known that treatment with this solution leaves absorbed chromium on glassware that is extremely difficult to clean off. There are proprietary pattern strippers of unknown composition which are somewhat effective in stripping the polymerized resist. These are not recommended because of the uncertainty in their composition and in the residues which might remain upon the etched surface. When stripping resist films from soft metal films, such as aluminum, care must be exercised so as not to damage the metal film. This is especially true when one employs the scrubbing technique.

Use of Resist Materials as Evaporation Masks: Occasionally for films it is more convenient to put the resist pattern

on the wafer prior to deposition of the film. This would be the case, for example, with a metal that was difficult to etch, such as nichrome. In this case evaporation covers the film and the exposed areas of the substrate. In the subsequent stripping operation the resist film is removed with that metal which is on top of it, while those areas which were exposed during evaporation remain coated, hopefully, with the metal. The principal disadvantage to this technique is that the resist film limits the outgassing that the substrate can receive prior to evaporation. This can result in an adhesion problem.

Capabilities and Limitations. Definition by the photolithographic techniques of the order of 0.1 mil (2.5 microns) over an entire pattern is about the present state of the art. This means that a .001" wide resistor is ± 10% at best. Clean 0.1 mil lines have been made[33]. However, these employed a mask which was appreciably narrower than the resulting stripe.

Metallization for Intraconnections

The thin metallic films employed for intraconnections over the silicon dioxide layer must meet several general requirements. It must adhere to the silicon and to the silicon dioxide; it must not penetrate the silicon dioxide; it must make ohmic contacts to both n and p-type silicon; it must be of adequately low sheet resistance and, in addition,

it must supply a means for connecting leads for interconnection of the circuit blocks. Evaporated films of aluminum similar to those commonly used in transistor contact areas fulfill these requirements, providing that all of the n-type areas to be contacted are heavily doped. Film thicknesses employed are usually in the range of 0.1 microns to 1 micron. The sheet resistivity of such films is of the order of 10^{-2} to 10^{-3} ohms/square. The possible problem of reaction of aluminum with silicon dioxide because of the greater affinity of the aluminum for the oxygen does not prove to be a real problem at the operating temperatures experienced by the microcircuitry. This reaction at the interface does, however, yield excellent adherence. At temperatures above the melting point of aluminum, 660°C, the oxide is penetrated rapidly. In fact, at all temperatures above the aluminum silicon eutectic temperature at 577°C, where a liquid phase can form, the rate of penetration is appreciable. Extended life tests, however, at 300°C with thin films of SiO_2 show that penetration at these temperatures is extremely slow. The principal disadvantage to the aluminum metallizing is that the soft aluminum film is very susceptible to mechanical damage. It can be scratched easily during the normal processing operations -- a problem which is much more severe in the microcircuitry than in ordinary transistors.

The aluminum is usually deposited in a conventional evaporator by heating on a refractory metal filament. Good adherence requires a well cleaned, grease free surface. Ionic bombardment with an inert gas, such as argon, prior to evaporation is helpful in obtaining reproducible results. After photoetching the pattern desired with sodium hydroxide etch, employing a photoresistive film, it is usually desirable to heat treat the structure to the vicinity of the aluminum silicon eutectic temperature briefly in order to assure a firm bond between film and substrate. The principal problem with aluminum aside from the ease with which it can be damaged mechanically is that it limits the temperature at which the microcircuit can be subjected. This limitation prevents the inclusion of certain potentially desirable packaging techniques. However, the well established reliability of thermo compression bonds properly made to the aluminum film often overweighs this objection. Other types of metallization have been considered. An interesting development in this direction was reported by J. Langdon, _et al_[34] who used silver chromium metallization which was subsequently covered with a glass whose expansion coefficient matched that of the underlying silicon. Contacts were made by using holes through the glass to contact points where appropriate metallic buttons were raised in place. Assembly techniques such as this offer

the advantage of complete mechanical protection with essentially no external package. By interconnecting such structures directly on a mating metallized pattern -- a sort of micro printed circuit board -- one can hope to achieve very high packing densities at minimum expense.

Assembly and Packaging

The techniques employed to assemble and package the microcircuits have an important bearing upon their usability and reliability. No single factor is more important in influencing the reliability of planar semiconductor integrated circuits than the packaging of these circuits since the residual failure modes are generally found to be mechanical. The close relation of packaging configuration and usability is obvious. While the input of the system manufacturer is necessary to achieve proper external configurations suitable for efficient system assembly, at least equally important in choosing a package design is careful consideration of the various materials and materials assembly techniques which go into its construction.

The basic requirements any package must fulfill are first that the circuit be secured mechanically; second, that the necessary connections required in the circuit's use be made available; and third, that the circuit be protected as required from its environment. The various packaging schemes considered to date include the considerations of

other boundary conditions depending upon the particular use to be emphasized. For example, some are oriented toward ease of systems assembly while others emphasize small size and light weight. Two packaging schemes employed at present for semiconductor integrated circuits are worth describing in some detail. The technologies used for assembly in these cases are somewhat different, so will be described in conjunction with a particular package configuration. A modified version of the standard transistor package has been employed where maximum reliability and usability are desired, since such a package makes maximum use of proven transistor technology and printed circuit board assembly. An integrated circuit in such a package prior to sealing the "can" is shown in Fig. 5-22. The small, flat "bug" package is suitable where size and close packing are of paramount importance. Fig. 5-23 shows such a structure. The multi-lead transistor package consists of a Kovar eyelet with a matched glass-to-metal seal through which the leads are sealed, often with one lead welded to the bottom of the metal eyelet. These are usually of the same general configuration as the TO-5 transistor header with a 0.200" pin circle. Up to ten leads can be used, although a maximum of eight is more common. On occasions the smaller TO-18 size header with eight leads maximum is employed. In either case, the silicon die is attached to the gold plated Kovar header

using a gold solder or a gold alloy if electrical contact is desired.

To make contact to n-type silicon gold with about 1/2% antimony is appropriate. For p-type material, it is more difficult to make ohmic contact, although gold with several percent gallium works with relatively heavy p-doping in the silicon itself. The solder-down operation is then in an enert atmosphere either by a hand operation holding the silicon in tweezers or through a jig operation employing a belt furnace. The operation proceeds easily, providing that the silicon surface is free of oxide. The temperature during this operation is in the vicinity of 400°C. Gold solder preform thicknesses of the order of 1 mil are typical. Since the penetration into the silicon during the alloying is to a maximum of about 20% of the preform thickness when penetration is uniform, the circuit structure built within the top fraction of the die is undamaged. The solder joint is strong and of low resistance. Since the Kovar eyelet matches the thermal coefficient expansion of silicon quite closely, there is a minimum amount of thermal strain introduced by this joint. Such a soldering with gold silicon alloy is well established in the transistor industry. The principal problems one is likely to encounter are the occurrence of voids and carrier injection into high resistivity material. Since microcircuits in general do not

handle high currents these are not particularly important problems here.

Bonding of leads in this packaging is usually by thermo compression bonds. To make such a bond the structure is heated to above 300°C but below the gold silicon eutectic temperature of approximately 370°C. About 350°C is the preferred operating temperature. Soft metal, usually gold but occasionally aluminum or silver, is compressed against the metallization. The tool employed for compression can be a wedge but best bonds are obtained when a capillary tube is employed to make a nailhead bond. In the latter case the gold wire is fed through a small glass capillary tube. The hole in the capillary should be just large enough to let the wire pass freely. A ball is made on the end of the wire by melting a portion of the wire with a small flame or heated platinum wire. Typically for two mil gold wire, the diameter of the gold ball is approximately 4 or 5 mils. The ball is positioned over the bonding pad over the device metallization and pressed firmly against the pad so as to result in about a 50% deformation of the ball. Under these conditions a firm bond between the aluminum metallization and the gold bond is achieved. The other end of the wire can then be thermo-compression bonded to the gold plated Kovar leads or it can be resistance welded.

The device is hermetically sealed in an inert atmos-

phere by resistance welding of a flange on the Kovar eyelet to a metal cap. Well made seals of this type show no leakage to the most sensitive leak testing techniques available today. By pressurization with radioactive krypton and counting under a scintillation counter, devices of this type have been shown to be producible on a production basis with leak rates less than 10^{-12} standard cc/sec. Reproducible achievement of this seal has been the result of considerable development in the glass-to-metal sealing techniques as well as the precise geometry and conditions for achieving reproducible welds.

Advantages of this type package are that it employs only proven techniques carried over directly from transistor technology. The package and techniques have been evaluated through extensive life testing. In addition conventional printed circuit board assembly techniques can be employed, although for best efficiency it is necessary to employ two layers of conductors. The disadvantages of this package are that it limits the number of leads to eight or a maximum of ten, and it does not use space efficiently, since the integrated circuit die represents only a small fraction of the total volume of the package. This latter limitation can be alleviated appreciably by going to a multi-lead variation of the TO-47 package. In this case, the Kovar eyelet is replaced with a solid Kovar slug and overall package heights of 0.060"

can be achieved.

In order to get around these shortcomings the flat package with co-planar leads has been used. The one illustrated in Fig. 5-23 is a 1/4" x 1/8" rectangle approximately 0.040" thick. As can be seen in the photomicrograph, this configuration can have a maximum of 14 leads -- five coming out each side and two out each end. The leads are ordinarily on 0.050" centers. Such a package obviously makes much more efficient use of volume.

Assembly in this package varies. The die can be attached to a metallized ceramic base plate with metal solder similar to the die attach operation in the modified transistor package; or, in cases where electrical contact to the bottom of the die is not desired, a low melting glass can be employed to glue the integrated circuit to the package base. Lead attachment again is usually by thermo-compression bonding, although because of the limited vertical space available use of the superior ball bonding technique is severely limited. Here a wedge shaped tool is more common to compress the wire from the side upon the metallized areas. Because of the limited vertical dimension, care must be exercised to assure that no shorting occurs between wires or between a wire and an undesired portion of the semiconductor die or package.

The sealing of these packages presents a rather

formidable problem. Attempts to make such packages with flanges for resistance welding are thwarted by the lead positions. Solder sealing of the top has never proven to be a reproducible operation. The best technique available at present appears to be sealing with a low melting glass. Unfortunately the internal structure limits the maximum temperature to which the structure can be subjected, so the choice of glasses for sealing is severely restricted. Both metal and ceramic oovers have been used. The metal cover has the disadvantage that it can short the internal wires if extra insulation is not provided. The ceramic top adds more to the overall package thickness but at the present appears to be the best all-around solution.

The advantages of this package are small size, light weight and suitability for stacking into modular configurations. The disadvantages are that it is often expensive and that the new assembly techniques required introduce real and potential reliability problems, each of which must be investigated extensively before one can assure the same reliability as is attainable with the transistor type package. System assembly techniques to take full advantage of the size and configuration of this package are expensive and appear useful only for small special purpose systems. However, increased acceptance of this general configuration suggests that appreciable progress in its use is being made. Addi-

tional disadvantage to this particular package is that the 1/4" x 1/8" size is too small for the largest microcircuits available today. As our ability to make still larger integrated circuits improves it will soon outstrip the available package area. Of course, this technique is not limited to the 1/4" x 1/8" dimension. Other flat package configurations are useful for special purposes. For example, a circular flat package with radial leads offers the advantage that the further one goes from the package center, the greater the separations between adjacent leads.

Other schemes for packaging integrated circuits exist. For example, the inclusion of integrated circuits in Micromodule wafers (see Chapter 3) has been considered. Such a combination of integrated circuitry with the Micromodule assembly technique would result in an extremely high effective high component density.

The ultimate scheme for packaging and interconnection involves the complete elimination of separate packages for individual components or circuits and the packaging of the system as a whole. Such efforts as mentioned in the end of the section on metallization wherein the structure is coated with an insulator through which holes are formed at the appropriate spots for interconnection is moving in exactly this direction.

5-3. DEVICE STRUCTURES

The performance of microcircuitry is determined by the structural features employed to achieve the desired electrical functions. This section will be devoted to examining various of these structures and discussing how they might be made from the existing technology as well as the limitations and compromises which must be considered.

In general, semiconductor integrated microcircuitry is made from discrete regions which are associated with particular electrical functions. While these are not necessarily a one for one correspondence between the regions of an integrated circuit and the components one might employ to make such a circuit by conventional means, such a correspondence can often be made. Some distributed effects can be employed. For example, distributed resistor-capacitor networks are sometimes used. In addition, one is forced to consider many parasitics which occur as basic parts of the structure itself. These can be used to advantage, although one is usually more concerned with minimizing their effect upon performance. As well as active and passive elements, one must also achieve within the integrated circuit the functions assigned to the printed circuit board in conventional electronics; that is, those functions of interconnection, insulation and mechanical support. An appreciable part of this section will be

devoted to these benign electronic functions.

Transistors

Since the whole concept of semiconductor integrated circuitry arose from and is significant because of the transistor, we will discuss transistor structures first. Fig. 5-2 has shown schematic cross-sections of conventional mesa and planar transistors. In such a schematic cross-section, the geometry is considerably distorted in order to show the emitter and collector regions. Actually, the emitter and base layers penetrate only a few percent of the thickness of the wafer. The oxide and metal films are also of the order of 1% of the overall thickness. All of the drawings of transistor-like structures will have this distortion in order to illustrate the important features while still showing the entire wafer thickness. The steps necessary to make the two transistors shown in Fig. 5-2 are shown in Fig. 5-24 and Fig. 5-25. In the case of the mesa device, the original wafer is diffused over the entire top surface with the desired impurity element -- for an npn transistor, either boron or gallium. The oxide growing during diffusion is photoetched to form a pattern of apertures through which the emitter is then diffused. The oxide is removed from the wafer, metal is evaporated, a second mask is applied and indexed with respect to the emitter dots and metallized base and emitter contacts are made. The

mesa is then cut by a third indexed operation, although usually with looser tolerances than with the first two. For the planar transistor structure, the first step is an oxidation followed by the opening of holes in the oxide for the base diffusion. An oxide masked base diffusant such as boron is used, followed by an indexed oxide removal to open the aperture for the emitter diffusion. After emitter diffusion holes are opened above the base and emitter contact areas and subsequent metallization and photoengraving leaves the desired metallizing pattern in these areas. The advantages of the planar structure for microcircuitry are of sufficient magnitude that only the planar structure need be seriously considered for microcircuitry transistors. These advantages are associated with the fact that junctions are all covered by an insulating oxide layer and that the surface is relatively flat. For integrated circuitry the four index masking operations and two diffusion steps can be made to yield a transistor structure identical with any isolated planar transistor. The differences in the performance one might get with such a transistor in an integrated circuit as compared with the identical structure as an isolated component are associated with the environment into which the transistor is placed and the resulting effects upon the means which might be employed to achieve electrical contact to the various regions.

Other applications of the technology to make transistor structures can also be considered, although they have not yet been employed. For example, planar transistor structures can be achieved by epitaxial growth of the base layer followed by diffusion through the grown layer to limit the extent of the base region, followed in turn by a second diffusion to make the diffused emitter. In the final analysis, the structure itself is the important consideration rather than the technology by which it is achieved.

Diodes

Diode structures follow easily from the transistor structures. Both planar and mesa diodes are made as isolated components and again the planar structure is more applicable for integrated circuitry. Either the emitter base or the collector base diode of a transistor can be used for the two terminal structure. In addition, it is possible to make Zeners using, for example, the emitter base diode of a transistor.

Isolation

Only a few very simple circuits can be made from transistors with all collectors common. In the usual array of transistors, such as that shown in Fig. 5-1, this common collector configuration occurs naturally. This fact accounts for the popularity of Darlington amplifiers in integrated circuitry. To get flexibility in circuitry design

it is necessary to achieve electrical isolation of the individual component structures. In conventional electronics this is done by cutting the device apart mechanically. In the chip approach similar isolation by separation is also employed. In the integrated structure achieving adequate isolation is one of the principal problems. Many schemes are possible. Nearly all depend upon the inclusion of extra junctions so that regions which must be isolated are separated in operation by a reversed biased diode. The simplest scheme in principal is to produce a series of islands of one conductivity type material in a wafer of other conductivity type and to insist that the substrate wafer is always so biased with respect to the rest of the circuit that the junctions separating the islands from the wafer are never forward biased. This can be achieved in several ways each with its own advantages and disadvantages. An example of how this might be achieved in shown in Fig. 5-26. An original wafer of p-type material is oxidized and holes photoengraved through the oxide layer. The diffusion of n-type islands is performed with a surface concentration held adequately low so that the subsequent diffusion of device structures into the islands can be accomplished. In the particular example in Fig. 5-26, the steps necessary to build a transistor in the island are shown. Again the base and emitter diffusions of the transistor can correspond to

those of a conventional planar device. One can appreciate the effects of this type isolation by considering the limitations it introduces in the performance of the transistor structures. First, in order to achieve the diffusion of the isolation region, it is necessary that a concentration gradient be established resulting in relatively light doping near the junction compared with that near the surface. Since all the collector current must flow laterally through this lightly doped region to the collector contact made in the surface, a very large collector spreading resistance is added. This might well run of the order of thousands of ohms and can be reduced to the order of hundreds of ohms only with considerable difficulty. In addition, the isolation junction acts as a capacitor between collector and substrate which couples this region to every other isolated region in the structure. Thirdly, because of the necessary surface concentration for the isolation diffusion, the concentration gradient across the base collector junction at the surface is appreciably higher than would be necessary for the isolated device, resulting in a lower base collector breakdown voltage than would otherwise be required for the equivalent transistor. Hence, the electrical parameters associated with the separate transistor structure itself have been considerably degraded. The structure, however, would still be

useful for relatively low voltage, low current circuitry. An example of circuitry employing this type of isolation is the Series 51 Solid Circuit family produced by Texas Instruments, Inc.[35].

An alternative method of achieving isolation is shown in Fig. 5-27. The two versions of this basic technique are illustrated. In this case, the isolated islands in which the components are subsequently constructed consist of the original wafer doping. Isolation is achieved by diffusion of the opposite polarity completely through the wafer so that the diffusion fronts intersect.

At the left side of Fig. 5-27 a scheme is shown by which masking on both sides of the wafer is employed, yielding isolated islands completely penetrating the wafer, while on the right side a structure is shown wherein diffusion takes place from the entire back surface, resulting in an array of isolated islands which penetrate roughly half way through the wafer. The advantages of this isolation technique over that described above are that the transistor is made in the original bulk doping as it would be for a separate component. This results in breakdown voltages comparable to the separate device; and, since the doping under the collector junction of a transistor can be appreciably higher, it is easy to obtain lower collector spreading resistances. This is especially true in the

structures shown on the left, if one is willing to make collector contact on the opposite side of the wafer. Disadvantages of this structure are that the diffusion through the wafer requires thin wafers and long diffusion times with high surface concentrations, resulting in high isolation capacitance. Also, the large junction area associated with these isolated structures makes it somewhat difficult to achieve good diode characteristics resulting in relatively poor isolation. However, in the cases where capacitance and degree of isolation are important, it is possible to employ a multiple grid, such as that shown in Fig. 5-28, wherein more junctions are employed in series to decrease both the leakage and the capacitance at the expense of an increase in area necessary for a given isolated structure.

By employing epitaxial growth, it is possible to make an isolated structure combining the advantages of both of the structures described above. Such a structure is shown in Fig. 5-29. By starting with a high resistivity p-substrate one can minimize the capacitance per unit area of the isolation junction. The n-type film can be grown to a desired concentration or, to obtain maximum advantage, can be grown with a concentration somewhat heavier initially and lighter near the surface. The limitation on the concentration at the bottom of the film is that one must be able to diffuse through it to achieve isolation. In Fig.

5-30 typical impurity profiles for the three different isolation techniques described are compared, each with the same transistor structure included. In the diffused collector, one sees that there is relatively low doping in the collector region corresponding to the high collector spreading resistance described above. With the diffused isolation appreciably more doping is available in the collector region, although the isolation junction has high capacitance per unit area because of the necessarily high diffusant concentrations necessary to completely penetrate the wafer. In the case of the epitaxial collector with diffused isolation, however, conductivity in excess of that for the diffused isolation can be achieved, combined with the isolation capacitance per unit area of the diffused collector structure. The price one must pay for those advantages is the inclusion of epitaxial growth in the processing.

A variety of other schemes for achieving isolation have been tried. These include mesa etching in various configurations and some attempts to remove the silicon between structures mechanically without the loss of index. For example, structures have been made wherein the regions of semiconductor are isolated by etching through from the backside with selective etch that does not attack the silicon dioxide on top of the wafer. While the performance of such isolation is excellent, the process

difficulties are still beyond the state of the art. As the technologydevelops other isolation structures will arise. For example, selective growth of regions by epitaxy similar to oxide masking of the diffusion will open a variety of possibilities. Similarly, integrated circuits wherein the original substrate constitutes a common emitter rather than a common collector are likely to become producible structures. A large fraction of the improvements to be made in integrated circuitry in the next few years will be the direct results of improved schemes for achieving electrical isolation within a single block of semiconductor material.

Intraconnections

Equally important with the ability to isolate structures from one another is the ability to connect these structures in the desired circuit configurations. In the earliest integrated circuitry, such as the example shown in Fig. 5-31, this connection was accomplished by means of thermo compressionbonding of fine wire, similar to that employed in standard transistor fabrication, wherever it was necessary to use something other than the contiguous silicon substrate itself. This type intraconnection discards a large portion of the potential cost advantage adherent in the integrated circuit approach over the separate components and, in addition, introduces considerable opportunity for failure,

since the most common failure mode of modern semiconductor devices is the opening of a lead. It was to a large extent, however, dictated by the mesa structure. The advent of the planar structure made it possible to intraconnect by means of thin films passing over the silicon dioxide insulating layer, making contact to the underlying substrate where appropriate. With oxide layers of the order of one micron in thickness, these intraconnections have a capacitance of the order of 0,016 picofarads/square mil. Aluminum is the most commonly employed intraconnection metal. It sticks well, does not penetrate the oxide at operating temperatures and can be made to make ohmic contacts to the underlying silicon. The evaporated films are usually a few tenths of a micron thick and have a sheet resistance < 0.01 ohms/square. Therefore, the resistance contributed by an intraconnection lead across even a large integrated circuit contributes only a fraction of an ohm resistance, even for stripe widths as small as 1 mil. In the intraconnection of complicated circuits it is often necessary to cross leads. This can be accomplished readily by utilizing the underlying silicon. Circuits of considerable complexity can be designed where all of the lead cross overs are done making use of structural features necessary from other considerations, For example, it is often convenient to run leads across a diffused

resistor. Such features as split collector contacts can also accomplish the lead crossing function. If it is necessary to add a completely separate lead crossing, one can make a transistor-like structure in a separate isolated region and make use of the relatively low spreading resistance in the emitter diffusion. The 2 ohms/square sheet resistance usually employed for this diffusion allows one to make lead crossing with only a small resistance contribution, while maintaining the size reasonable.

Intraconnection in the chip approach is the major problem. The usual technique makes use of thermo compression bonding to each of the components. This is actually a microassembly job of circuit components without their normal packages. Considerable work has been done in an attempt to locate the separate chips precisely and to fill the voids between these chips in a manner so that the resulting agglomerate is a suitable structure for evaporated metal intraconnection. When the technology is adequately developed to allow this or a similar scheme to be done economically, the single chip approach will offer a new degree of flexibility with minimum interference from parasitics.

Resistors

In choosing a resistor structure one must consider the range of nominal values of resistance accessible,

the precision with which a nominal value can be achieved, the possibility of matching sets of resistors, the variation of the resistance values with time and temperature, the heat dissipation properties of the structure and the increase in construction cost necessitated by any additional steps in the processing. All of the technologies appropriate at present make use of a resistive film of one type or another. The resistance value is accordingly the product of the length to width ratio of the structure and the sheet resistivity. Because of the very high cost of the substrate in semiconductor integrated circuitry compared with that for other types of resistor networks an additional consideration is introduced. It is important that the resistive structures be as small as is compatible with the performance requirements. For this reason, it is necessary to use high sheet resistivities and very fine structures to achieve reasonably high resistance values. We will consider both structures made within the semiconductor material and by films upon a semiconductor substrate.

Semiconductor Resistors. The bulk resistivity of the semiconductor can be employed to make resistors. This is easily done when the grid type isolation shown in Fig. 5-27 is used. In this case, the nominal resistance is given by

$$R = \frac{\ell}{wd} \rho$$

Where ℓ, w and d are respectively the length, width and thickness of the layer and ρ is the resistivity of the semiconductor material. A typical structure of the type shown in Fig. 5-27(b) where a wafer might be of the order of 80 microns thick, resulting in a resistor with d $\sim$ 40 microns, gives sheet resistances of the order of 250 ρ ohms/square. Because the surface geometry is controlled by the difference between the original separation of grid lines and twice the lateral diffusion depth during the isolation diffusion, it is necessary to employ relatively wide lines in order to achieve much precision by this technique. If high resistivity material is used in order to get high value resistors one can easily run into incompatibility problems with transistors in the same structure. In fact, the sheet resistivity of the resistors employing this structure can be traded with the collector saturation resistance of adjacent transistors. The temperature coefficient of such resistors is usually high and positive. By employing the epitaxial isolation structure as shown in Fig. 5-29, high sheet resistivities can be obtained by the use of relatively thin structures but the same problem with respect to other components occurs. Resistors made from the bulk semiconductor are not, in general, very useful.

A very convenient technique for making resistors is to employ diffused stripes, since this can often be accomplished

with no additional processing steps. For the diffused collector isolation structures shown in Fig. 5-26, one can use either the collector diffusion or the base diffusion, while in the grid diffused structure the base diffusion is usually convenient to employ. In either case, the sheet resistivity is of the order of 100-200 ohms/square. Because of the oxide masking capabilities, excellent geometry control can be maintained. Common line widths for such resistors are 1 or 2 mils with equal spacings. By employing 1 mil lines with 1 mil spaces with the aforementioned sheet resistivities, a resistor in the 20-40 kilohm range can be made in a 20 mil square. By decreasing the line width to 1/2 mil with 1/2 mil spacings this value can be quadrupled. Because of the diffusion profile the doping in such a resistor is not constant. Most of the conductivity is contributed by the relatively highly doped region near the surface. When the base diffusion is employed this doping is usually of the order of 10^{18} impurities/cm^3. At such a doping level, the resistivity has a high temperature coefficient. The resistance vs. temperature curve for such a resistor is shown in Fig. 5-32. For temperatures above 25°C, the temperature coefficient is about 0.2% per degree and somewhat less for lower temperatures. The temperature coefficient is strongly dependent upon temperature, concentration and the specific impurity.

By using gallium, for example, instead of boron a resistor like the one shown would have a negative coefficient below about 0.°C[36]. It is possible to effect the temperature coefficient appreciably by the addition of deep level impurities. However, these usually interfere sufficiently with the remainder of the microcircuitry that they have not been employed for this purpose. The different behavior obtained with gallium diffused structures results from its relatively high ionization energy.

The precision attainable is effected both by the variations in geometry and in the uniformity of the diffusion. Geometry effects usually appear as a general increase or a general decrease in the line widths in a given area; that is, as a scale factor variation rather than as a random deviation about the desired line width. Inhomogeneties in temperature or flow pattern during diffusion or variations in the original semiconductor wafer account for variations in the sheet conductivity. For both reasons resistors in closely adjacent areas should be expected to show similar deviations from the desired nominal value. It is difficult to achieve absolute resistor precision better than about 10% with high yields. A typical curve of resistor value distribution designed for a nominal value of 600 ohms is shown in Fig. 5-33, representing a large number of diffusion runs over an extended production period[37]. The medium value

of 570 ohms misses the design value by 5%. The standard deviation in this case is 35 ohms or about 5%. However, adjacent structures are matched more accurately than if they had been selected at random from such a distribution. Fig. 5-34 shows a distribution curve of the ratio of pairs of resistors[38]. One sees that about 50% of such pairs are matched to within 1%. In spite of the large temperature coefficients associated with these diffused resistors, pairs of adjacent resistors track very well with temperature. Fig. 5-35 shows the tracking of such resistor curves.

In cases where very high values of resistance are required, one can consider the use of the base spreading resistance by making a structure as shown in Fig. 5-36. This is similar to an ordinary diffused resistor using the base diffusion, except that the emitter diffusion is included over all of the resistor except near the contacts. Typical sheet resistivities from this structure are 5-50 kilohms/square. Because this is achieved by taking the difference between two large numbers, it is not easy to control precisely. Fortunately, by using the emitter diffusion and the base diffusion, one gets a degree of control by monitoring the current gain of any transistor structures made at the same time. For the usual spreads necessary to obtain useful yields the sheet conductivity cannot be controlled much better than to within a factor of 2 from the desired value.

In addition, resistors of this type are not truly ohmic since there is some self-pinching effect and they are limited in applied voltage to the emitter base breakdown voltage of the transistors. Fig. 5-37 shows a voltage current plot on such a resistor[39].

Film Resistors. Many of the techniques described in Chapter 4 on thin films can be combined with the semiconductor circuitry. It is usually more convenient to make film resistors for semiconductor integrated circuitry by the use of resist-masking techniques rather than to evaporate through a shadowing mask because of its compatibility with the remaining technology. Resist techniques take two forms: coating the entire wafer with the desired film and etching to leave the desired pattern; and putting a resist pattern on the wafer before applying the resistive film so that the excess metal will come off when the resist is stripped. Nichrome, for example, is useful up to approximately 200-300 ohms/square. It is fairly stable especially when sealed in an inert atmosphere and it adheres nicely to the silicon dioxide layer. Fig. 5-38 shows a photomicrograph of an integrated differential amplifier using Nichrome resistors. The refractory metals are also applicable. The advantages of the thin film resistors in this application over the semiconductor resistors are lower temperature coefficients, lower shunting capacitance, some possibility of saving

area and better isolation. The disadvantages are the inclusion of extra processing steps, at least one evaporation and one masking, and lower power capability.

Other films than metals have been considered. Structures have been made using tin oxide[40]. The inclusion of cermets by simultaneous evaporation of metal and insulators appears interesting[41]. The achievement of a high sheet resistance, stable, controllable film resistor compatible with the rest of the processing technology will be an important extension of the semiconductor integrated circuitry.

The tolerance of present thin film resistors is of the same order as that for the semiconductor resistors, although one can reasonably anticipate that use in production will result in an appreciable improvement. Because of the ability to monitor the sheet conductance during evaporation variations from this source can be minimized. In addition, with thin film resistors, one has the possibility of adjusting the individual structures subsequent to their preparation by any one of a variety of techniques which have been described[42]. This can be important for making circuits requiring precise matching of resistance values.

The capacitance distributed along the film resistors between film and substrate is of the order of 0.02 picofarads/square mil for the usual oxide thicknesses of

approximately 1 micron. This is about an order of magnitude less than the capacitance per unit area usually associated with the diffused silicon resistors. The power dissipation of the film resistors is seldom a problem, since the overall power dissipation is usually limited by the entire package. Silicon is an excellent thermal conductor and the thin silicon oxide layer increases the thermal resistance relatively slightly. Each thin film structure, however, must be carefully evaluated with respect to its stability when dissipating power since the deterioration of such thin film structures has been shown repeatedly to be more severe under operating conditions than when subjected to what is designed to be an equivalent thermal environment.

Capacitors

A most obvious structure to use to achieve the function of a capacitor in semiconductor integrated circuitry is the pn junction itself. While such a structure follows rather naturally out of the technology used to produce the other components, it does not, in general, have especially good electrical characteristics. The junction capacitance is a strong function of the applied voltage and depends upon the impurity concentration profile. For a typical emitter base junction, one gets approximately 0.4 picofarads/square mil at 5 volt bias. Because of the rectifying nature of these junctions, the capacitors can only be used with voltage

applied so as to reverse bias the junction. Somewhat less voltage dependence and operation in either polarity can be achieved by two junctions back-to-back, although this results in a factor of two decrease in the capacitance attainable in a given area. In Fig. 5-39 several possible configurations of junction capacitors are shown. In Fig. 5-39 (a) the simple junction structure is depicted. Fig. 5-39 (b) and (c) show two alternatives for the back-to-back junction configurations, while Fig. 5-39 (d) shows how the equivalent base and collector base junctions in a transistor structure can be combined so as to increase the capacitance achievable in a given area. In all cases, one terminal has a fairly high resistance contributed by the low sheet conductivity of the semiconductor layer. If this series resistance is important in a given application, it can be minimized by going to many-fingered, interdigitated structures, such as those employed in high frequency power transistors. As in other capacitor structures, the maximum useful voltage can be traded for capacitance per unit area. In contrast to some other techniques, however, these junction capacitors can be broken down and recover their original characteristics when the voltage is decreased, providing that their power handling capacities are not exceeded.

Another type capacitor structure is shown in Fig. 5-40. In this structure the dielectric is made from the silicon

dioxide layer grown upon the semiconductor. The underlying silicon forms one electrode while the other electrode is the conducting film. These capacitors have good linearity and can be made with low series resistance. Again the capacitance achievable in a given area is limited to about 0.4 picofarads/square mil. Fig. 5-41 shows the capacitance per unit area and the dielectric breakdown voltage for metal-oxide-silicon capacitors versus the thickness of thermally grown silicon dioxide layers. Working voltages should, of course, be restricted to something appreciably less than the dielectric breakdown. The incidence of pinholes in the silicon dioxide film limits the areas that can be employed economically in this type capacitor. Values up to a few hundred picofarads are as large as can be combined with other semiconductor components at present with adequate yields.

Work to get higher dielectric constant films has been reported. By combining titanium dioxide with the silicon dioxide, dielectric constants as high as 30 have been achieved[43], but such structures have not yet been shown to be adequately stable. Extension to high capacitance values in semiconductor integrated circuitry seems to require the use of thin film structures, such as those described in Chapter 3. Until these have been successfully integrated high capacitors should be avoided wherever possible or

included upon a separate substrate where absolutely required.

Other Structures

Many other structures are possible, some of which will be extremely important. In this section we will describe a few of these, suggesting some of the advantages they might be employed to achieve.

As a first example, consider the unipolar field effect transistor[44] which is desirable in several applications requiring high input impedance or as a switch with very low offset voltage. Field effect structures can be made in a variety of ways. In Fig. 5-42 a structure similar to the resistor employing the base spreading resistance is illustrated that will act as a field effect device. Referring to the characteristics of the pinch-off base spreading resistors shown in Fig. 5-37, one sees that a pinch-off voltage of the order of two volts can be expected for this structure. By employing a reasonably long narrow channel consistent with usual processing technology, say, for example, a channel that has a length to width ratio is 20, one can reasonably expect to get source-to-drain saturation currents of the order of 1000 μa. Combined with the 2 volts pinch-off voltage, this suggests transconductances in the vicinity of 500 μmhos. By further adjusting the geometry and diffusion schedule one can get transconductances above 1000 μmhos and on impedances less than 1000 ohms. In addition, the technology by which

this structure is prepared is completely consistent with the transistor technology, making circuitry requiring a combination of unipolar and bipolar transistors relatively straightforward.

Field effect structures where the gate is separated from the channel by insulating layer can also be made. Typical cross-sections for these surface controlled field effect transistors are given in Fig. 5-43. In the top structure n-type contact regions are diffused into a p-type substrate to form the source and drain contacts. A gate contact overlaying the oxide completely covering the region between the two n-type contact areas is employed. When the gate potential in this polarity structure is adequately positive, electrons are attracted to the surface of the silicon resulting in the formation of an inversion layer, which connects the two n-type contact regions. Such a field effect device is normally non-conducting, although the effects of surface states at the silicon oxide-silicon interface can result in the formation of an inversion layer at zero gate voltage. The dc input impedance of such an insulated gate is extremely high, of the order of 10^{15} ohms for a typical structure. The other variation of the surface controlled field effect shown in Fig. 5-43 is constructed to have a channel connecting the contacts at zero gate bias. In this case where a junction exists to the channel the semiconductor body can also be used as a gate electrode as with the ordinary field effect transistor.

Here the top gate overlying the oxide need not extend completely from source contact region to drain contact in order to control the channel conductance. The future of such surface controlled structures looks extremely interesting. However, there are problems associated with the stability of the surface states that can result in appreciable drifts of the characteristics. Additional understanding of this interface is required before wide use of this type of structure can be considered.

Another device suitable for integration is the surface control transistor tetrode, described by Sah[45]. A cross-section of this structure is shown in Fig. 5-44. Essentially it consists of a planar transistor, either NPN or PNP, modified by the addition of a grid electrode overlying the region where the emitter base junction intersects the semiconductor surface. By varying the potential on the surface at this point the current gain of the transistor can be modulated. The effect is the result of modulating the surface recombination velocity and in some instances the formation of a channel over either the base or the emitter under the grid electrode. This structure also offers extremely high input impedance, of the same order as that for the surface controlled field effect transistor. In addition, it is relatively easy to achieve transconductances of several thousand micromhos. Again, the technology by which it is produced is

completely compatible with that employed in making ordinary integrated bipolar transistors. Certainly a device with the properties possessed by this structure will have applications at least in special situations.

One of the earliest examples of a true integrated circuit was given by D'Asaro[46] when he described a stepping switch employing pnpn structures. By making use of the voltage drops which occur in an asymmetric four-layer device structure, he was able to control if the next stage would trigger or not upon the application of properly timed clock pulses by the condition of the first stage. A schematic representation of the circuit he used as a pulse counter is shown in Fig. 5-45. The operation of this structure depends upon the fact that in the common n-layer (as drawn) that the voltage is higher under the emitter of the structure that is conducting than it is under the opposite end of the individual p-regions. Thus, the common p-n junction is more heavily forward biased in the next structure to the right than it is in the structure to the left. For a correct combination of currents the n + 1 device will conduct when switch S is closed because the bottom junction will be sufficiently forward biased to remove the high impedance state of the four-layer structure. Upon opening switch S the on state moves to position n + 2. Thus in this structure each time switch S is opened and closed, i. e., each time line B

is pulsed, the conducting state moves two positions to the right. The structures described by D'Asaro would trigger the device to the right and not trigger the one to the left for a range of currents through a structure from less than three to greater than thirty millamperes.

Such a structure represents an extremely simple technique for achieving a shift register function, something which by means of conventional circuitry using only bipolar transistors requires a considerable number of components. While this is one of the earliest examples of a true integrated circuit, it still remains as one of the best examples of capitalizing on a unique interaction in order to achieve a desirable electronic function.

Another structure of considerable interest, although not really employing devices other than those discussed previously, is that where both npn and pnp transistors are included in the same circuit. In many instances in analog circuitry, configurations such as that shown in Fig. 5-46 are useful in order to use differential input and to obtain output not displaced with respect to ground potential[47]. Many other examples where complementary circuitry can be used to advantage exist. A simple method of achieving npn and pnp transistors in the same structure is shown in Fig. 5-47. Here we see the case where diffused isolation is employed for the npn transistor while the pnp transistor

is prepared by using the base diffusion of the npn as its emitter and the collector diffusion of the npn as the base for the pnp. The use of these diffusions for the pnp transistor cannot be expected to result in an optimum structure since they must be adjusted in order to achieve reasonable operating characteristics for the npn. In situations where low beta can be tolerated in the pnp device, this represents a straightforward technique for achieving complementary circuitry. In cases where it is important that the characteristics of the two devices be separately optimized, it is necessary to resort to more complicated processing schemes. A wafer having both p-regions and n-regions suitable for the collectors of the pnp and npn transistors respectively can be obtained in a variety of ways. For example, in Fig. 5-48 we have shown two methods by which such a structure might be achieved by epitaxial growth. In one of these structures, the growth covers only a portion of a p-type substrate. This can be done either by masked growing techniques or by etching through the grown layer. In the second configuration, the epitaxial growth is used to grow a layer over the entire surface of the substrate, layer and substrate being of opposite conductivity type. In the second case the pnp transistor will be produced on one side of the wafer while the npn transistor is made on the opposite side. In the first case, of course, these transistors can be

prepared side by side. Other structures employing diffusion at low surface concentration can be used to form a similar grid. In order to separately optimize the npn and pnp transistors it is necessary to first perform those diffusions which can be done with extremely slowly diffusing impurities. For example, the first diffusion might use antimony or arsenic as the n-type impurity in order to produce the base region for pnp transistors. A second diffusion might diffuse boron as the base of the npn transistor and simultaneously it could be diffused into the previously prepared antimony or arsenic region in order to form an emitter for the pnp device. Since this boron will probably be too lightly doped to be an efficient emitter, it might be followed with a relatively heavy boron predeposition over the pnp emitter and a phosphorus predeposition on the surface of the npn. The base width and hence, the final beta of the pnp, will be determined primarily by the difference in the diffusion depths of the first antimony or arsenic layer and the first boron diffusion. Since the impurity gradient at the emitter base junction in this structure is relatively small, it will not diffuse much during the time that a phosphorus emitter is diffused into the npn structure. On the other hand, the heavy boron concentration placed near the surface of the emitter for the pnp transistor will assure adequate emitter efficiency.

One sees that in order to prepare such a structure several additional diffusion and masking operations are required. The circuit design should be considered carefully before one decides that the best solution to his problem resides in making high quality npn and pnp devices in the same semiconductor substrate.

As a final example of an interesting structure useful in microcircuitry, we would like to consider the thermal feedback amplifier[48]. In this device, the temperature variation of the emitter base voltage of a transistor is used as a thermal sensor to control the amount of current being passed through a transistor structure and, hence, the amount of power dissipated in the collector spreading resistance. It has been shown that in this manner one can actually thermostat the piece of silicon to within a few degrees centigrade in spite of ambient variations from -55°C to + 100°C. In effect, one has succeeded in making a miniature thermostated oven. A diagram of a circuit capable of performing this function is shown in Fig. 5-49. As the temperature of the circuit falls, the emitter-base voltage of the first transistor necessary for conduction increases, resulting in lower conductance of the first transistor and, accordingly, making the base of the second transistor more positive. As the second transistor turns on, relatively large current flows dissipating power in the internal resistances

of the structure. This heats the silicon block until the first transistor conducts and shuts off the second transistor.

The usefulness of this sort of structure in stabilizing temperature sensitive circuits such as precision amplifiers or precise voltage reference sources is apparent. A voltage reference stabilized in this manner has shown stability to one part in 10^5 over an ambient temperature range of 150°C[49]. Here is an example of improved performance achievable by a structure uniquely possible in semiconductor integrated circuitry.

The structures described in this chapter are by no means all inclusive. Many others will be of importance. In particular the use of unique interaction effects available only when the various circuit elements are formed within parts of the same monolithic block will increase.

5-4. CIRCUIT CONSIDERATIONS

The same general considerations that go into the design of conventional circuits are required in the design of integrated circuits. However, some additional boundary conditions must also be considered. As with conventional circuitry, the prime considerations are that the resultant design will perform the required function with adequate reliability and that it will do this at the lowest possible cost. The principal changes in boundary conditions which occur when one considers the present technology for circuit integration

are the following: First, the relative cost of components change. Cost of components are now primarily determined by the area they consume on the semiconductor wafer. Thus, a transistor is approximately as expensive as a diode, which in turn corresponds to a resistor of about 4 kilohms ± about 30% or one-quarter that value ± 20%. Second, the values of available components are restrictive. Values of resistors exceeding about 50 kilohms are prohibitive and capacitors above a few hundred picofarads must be assembled separately. Third, the absolute values of resistors are poor, although the ratios between resistors in the same structure can be held fairly closely and track well with temperature, as was discussed in more detail in the subsection covering resistors. Fourth, all the components of the integrated circuit are coupled by various parasitic capacitances and conductances. This is a function of the close spacing and the techniques employed for isolation.

In general, in the design of integrated circuitry, prime consideration must be given to the size of the resulting die and to the yield which can be expected, for die size and yield are of prime importance in determining the economics of integration. In order to assure highest possible yields, the voltages to which the circuits are subjected should be kept as low as possible. The effect of voltage upon yield will be discussed more completely in Section 5-6.

A useful technique to aid the circuit designer in taking into account the various peculiarities imposed by integration makes use of special separate components for circuit breadboarding. The various component structures available in the integrated circuits can be prepared separately by the technologies used for the final integrated version. Thus, for example, transistors are made within isolated regions as they would appear in the final structure. In addition to the emitter, base, and collector leads, the isolation region is also made available. These transistors can be packaged in regular four lead packages. Similarly, diffused resistors or metal film resistors can be made as separate components surrounded by the environment which simulates what will surround them in the integrated circuit. In this manner, a circuit can be assembled wherein each component appears in the environment it will occupy in the final circuit and all isolation regions can be connected in an appropriate manner. By making combinations of transistors in the same isolation region and by making more than one resistor structure in a given block of silicon, the effects appropriate to common collector transistors and to pairs of resistors made simultaneously on the same substrate can be included. Circuits breadboarded with such components can be evaluated prior to laying out the masks and constructed in fully integrated forms. Such prior evaluation has proven

to be quite representative of the final integrated structure, although usually slightly on the conservative side, since additional parasitics are introduced by having separate packages and the advantageous effects of the uniform temperature which exists in the common integrated circuits are not shown at full advantage with the separate integrated circuit components. This technique results in a procedure by which integrated circuits can be designed and integrated with confidence that they will perform the desire function. Such a procedure is of considerable aid in giving the circuit designer the ability to include the effects of those peculiararities of semiconductor integrated circuitry.

Digital Circuitry

Digital circuits present the ideal instance for integration. The binary operation of such circuits places minimum strain upon the tolerances of individual components. The transistor itself is used in its optimum application when it is employed as a switch. Logic systems often employ large numbers of identical circuit functions important in the most economic use of integration, and the low voltage levels usually employed are consistent with the requirements for high yields. It is not surprising that the first useful applications of integrated circuitry have been in the digital area.

Most of the common types of logic circuitry are suitable for integration. However, the effects of the new

boundary conditions may change previous ideas about which circuit schemes are most appropriate. Both the type of logic and the circuitry desired must be considered in deciding on the most appropriate approach to a given problem. Some circuitry is most useful for high speeds, other for lowest power, and still other configurations might be most appropriate in situations where the highest possible logical gain is important. As examples of the tradeoffs which might be considered in choosing the circuit configuration, the basic three input gates for several types of logic circuitry presently used in digital computers is shown in Fig. 5-50. By comparing these, some idea of the relative usefulness can be made. For example, in Fig. 5-51 a plot of the relative speeds that can be expected for several of these configurations as a function of the power per node is plotted. Such a plot assumes that all of the gates are made with the identical technology and that a transistor of relatively low storage time is available. As the technology is altered the positions of the curves is altered. For example, if a transistor with appreciably lower storage time was included, the advantage of current mode logic at high power levels over direct coupled logic would largely disappear. The component values included in Fig. 5-50 are representative of those which might be employed when the particular type circuitry is used somewhere near its most advantageous condition. Of course,

these values change appreciably as the power level and speed of these circuits is altered.

A comparison of the TRL gate in Fig. 5-50 (a) and the DTL gate in Fig. 5-50 (b) is a good illustration of the effect of relative component costs upon circuit design. The TRL gate uses resistors in series with the input, while the DTL gate employs diodes. With conventional circuitry the resistors are often cheaper than the diodes, making TRL somewhat less expensive than DTL. This was especially true in the past. In the integrated case, however, the diodes required in the input of the DTL actually consumed less are than do the resistors required for TRL. Hence, DTL is a cheaper circuit to make than TRL, Since diode transitor logic also offers higher speed and greater digital gain than does TRL, there is no reason to even consider TRL seriously for integrated logic circuitry. DTL offers large voltage swings with corresponding high noise immunity. It is, however, quite slow as can be seen from the relative speed plot in Fig. 5-51. The pull-down resistor is relatively large and the need for both a positive and a negative power supply can be a considerable disadvantage, especially in a densely packed system, where one does not wish to have any unnecessary non-signal-carrying conductors. In addition, DTL ordinarily employs relatively high voltage power supplies which, as was mentioned above, is not especially desirable in integrated circuitry.

The low level logic NAND gate shown in Fig. 5-50 (c) offers the advantage of extremely high fan-in. With the two diodes in series, as shown, it has good noise immunity. This can be further improved by the substitution of a Zener diode. However, this is at a sacrifice in power and speed. It is difficult at present to make an adequately low voltage Zener diode in integrated circuitry, so the structure shown employing forward diode drops is more appropriate. The large resistor in the pulldown circuit is a problem and again the requirement of two power supplies with the accompanying disadvantages of extra power buses throughout the system is a drawback.

The T^2L gate shown in Fig. 5-50 (d) offers the same NAND function offered by low level logic, with some of the drawbacks relative to its integration removed. The input transistor is drawn to represent the manner in which it would ordinarily be made in integrated circuitry. Three emitters would be placed in a common base region, since the load transistors are connected in a common base, common collector manner. This circuitry is extremely simple, fast and easy to integrate. The gate shown requires only one relatively small resistor in addition to the two transistors, including the one with multiple emitters. Unfortunately, this circuitry has some disadvantages. Because of the additional drop across the gating transistor, the noise immunity

is less than one gets even for straight DCTL since the "down" voltage on the base of the inverter is always held higher than the saturation voltage of the previous inverter by the voltage drop across the gating transistor. It also suffers from a current hogging problem, considerably limiting fan-in and fan-out capabilities. If the additional components necessary to cure these problems are added, the advantages of this type circuitry disappear rapidly.

The modified DCTL shown in Fig. 5-50(c) wherein a resistor is inserted in the base of each transistor to prevent current hogging is simple and fast. All resistors are small, voltages are low, and only one power supply is required. The logical gain of such circuitry can be traded for noise immunity. Proper choice of base resistances with low saturation resistance transistors gives good noise margin over a wide temperature range with fan-in of 3 and fan-out of 5. While such circuitry with conventional components is quite expensive because of the large number of transistors employed per logical operation, it is quite suitable for integrated circuits.

The current mode logic gate shown in Fig. 5-50 (b) is most useful where speed is of paramount importance. With the technology one can expect to have available in the foreseeable future, such non-saturating, current steering circuitry appears that it will always give the minimum average propagation times at high power. As the gate is drawn in the

figure there is a level-setting problem which can be solved by using alternate stages of complementary polarity or by the inclusion of a Zener diode-resistor network. Neither of these techniques is especially desirable for integrated circuitry. A possible solution to the level-setting problem employs emitter followers[50]. This results in an appreciable increase in the area required and also requires a separate reference bus to each circuit. Another advantage of this CML circuitry is that both the OR and NOR outputs are available. The use of both outputs in perhaps 20% of the logic nodes in a system can be advantageous. The circuitry can be so designed that the relative values of resistors in a particular logic node are important rather than their absolute values, which again is compatible with the integrated circuitry boundary conditions. This type circuitry is presently restricted to high power, since at low power the magnitude of the resistors exceeds those readily available. The level-setting is a severe limitation. All solutions to this level-setting problem have resulted in a considerable increase in the area required.

Of course, consideration of just the three input gate does not give the whole story. If only gate structures are available as integrated circuits the advantages inherent in the technique are hardly apparent. Only when larger logical assemblies are considered, do the real advantages become apparent. As one considers larger functions, the effects of

area and yields become of increasing importance. This should be considered carefully when choosing the logic form desired. For example, using modified DCTL of the type shown in Fig. 5-50 (e), a complete shift register stage as shown in the photomicrograph in Fig. 5-52 can be made in an area 0.075 x 0.075 inches. The equivalent function using CML, even at much higher power level, takes approximately three times the area when the required level setting is included.

With evolution of the technology, perhaps the relative advantages of one structure over another in this respect will change. For example, with the inclusion of a good technology for making high values of resistances in small areas, the disadvantages to those circuits employing high resistors will disappear and such structures as low level logic must again be considered.

In any case, all of the logic forms in Fig. 5-50 are suitable for integration. When new systems are under consideration, the logic forms should be considered from the point of view of taking maximum advantage of semiconductor integrated circuitry. For inclusion into existing system, however, it may be advantageous to make integrated circuitry using logic forms which existed in the system when originally constructed from conventional components, even though these may differ from those that take maximum advantage

of integration. In these instances, one can still achieve the advantage of size, weight and elimination of interconnections. The use of the integrated circuit components for breadboarding of any of these logic configurations is a useful aid in designing and evaluating a circuit before the investment necessary to make it in a fully integrated form is committed. One can expect larger and larger logic blocks to be available in the future, as well as other digital circuitry, such as coding and decoding matrices.

Linear Circuits

The limitations imposed by the restricted values and precision of passive elements become extremely important in the area of linear circuitry. The lack of quality reactive components, especially the almost complete lack of inductance available in integrated circuits is a severe limitation on any structure requiring sharp tuning. The lack of precision available in the resistors is a limitation on many analog computing functions. While a considerable amount of work has been supported in the area of linear integrated circuits, very little of it has resulted in the construction of practical, useful configurations. The single real advantage which has been exploited to date in the area of integrated analog circuitry has made use of the fact that components built in the same piece of semiconductor material are at very close to the same temperature at all times and track well with tempera-

ture. Thus, it has been possible to make differential amplifiers with excellent performance. The thermal feedback amplifier described in the subsection on other device structures also makes use of this property.

Another area where existing technology can be used to make linear circuits which give advantages not available with conventional circuitry is in those applications where size is important. For example, it is possible to build an entire integrated sense amplifier into the head of a magnetic pickup, thus considerably reducing the noise problem inherent in running leads from the head to a remote amplifier. Also, the inclusion of some amplification in such structures as transducers close to the sensing elements themselves is often advantageous. Special applications, such as hearing aids, where size can be important and maximum fidelity is of relatively little concern are natural places to consider linear integrated circuitry.

Such applications as high quality, i. f. stages are severely limited by the lack of an appropriate compatible tuning element. The RC tuning that can be obtained with present resistor-capacitor networks is considerably inferior to that easily achieved employing LC networks in conventional electronics. The inclusion of small, sharply resonant structures, such as piezoelectric crystals, may eventually result in good tuning in a small volume. However, anything done

in this area to date has been more nearly the isolated chip approach than a truely integrated structure.

The economic application of truely integrated circuits to applications in linear electronics will be restricted to relatively special applications for the present. Considerable development, both with respect to the manner in which circuit functions are performed and the technology available will be required before the majority of linear circuitry is accessible to economic integrated circuits.

5-5. AN EXAMPLE

As an example of the application of the considerations in the previous sections concerning the technology, structures and circuits it is worthwhile to examine in detail the design, manufacture and evaluation of a semiconductor integrated circuit. For such an example, consider the logic circuit shown in Fig. 5-53. This is the circuit diagram for a gated flip-flop. It is one of a family of circuits called Micrologic Elements manufactured by Fairchild Camera and Instrument Corporation.[51] This family consists of a buffer element, a counter adapter, a flip-flop, a three input gate and a half adder in addition to the gated flip-flop called a half shift register. The logic diagrams for the various circuits are shown in Fig. 5-54.

The circuit form used in this family is modified DCTL as was shown in Fig. 5-50(e). All the integrated circuits are

made with the same technology and structural features, only the surface geometry is varied from circuit to circuit.

One sees that the gated flip-flop we have chosen for our example consists of nine transistors and fourteen resistors. The transistors exist as four common collector pairs and one separate device. Hence, the transistors must fall into five separate isolated regions. In order to make transistors of relatively low saturation resistance required in the chosen circuitry and in order to achieve high speed performance, the diffused grid isolation structure shown in Fig. 5-27 will be selected. The epitaxial isolation would be equally suitable but the design of these integrated circuits preceded the epitaxial technology. The five collector load resistors required can be made by having several diffused stripes of one conductivity type material in a region of the opposite conductivity type. Since with only positive voltages applied to the circuit, this structure which always separates resistors by two junctions will assure that these junctions do not get forward biased. Hence, all five collector resistors can be included in a single isolation region. In addition the nominal value of these resistors of around 600 ohms each is nicely compatible with the sheet resistance desirable for the base diffusion of approximately 150 ohms per square used in making the transistors. Accordingly, all five collector load resistors will be made

in a single region simultaneously with the base.

The nine base resistors could similarly be placed in an isolation region. However, if they were concentrated in a single region it becomes cumbersome to interconnect them to the various bases spread throughout the remainder of the circuit. Since the base diffusion sheet resistance is convenient to employ for these resistors also, an obvious possibility is to distort the geometry of the base region so as to obtain the desired number of squares between the base contact and the edge of the emitter. Care, however, must be exercised in proceeding in this direction. If the straightforward structure of a large rectangular base is made, additional effects come into play which have an important bearing on the operation of the device. In this case, because of the relatively large base resistance between contact and emitter, when the transistor is driven into saturation, the region of the base under the base contact becomes heavily forward biased. In this configuration the transistor acts as if it had a diode shunting from the base contact to the collector. This could result in an additional current path negating the effect of the base resistor to prevent current hogging by lowering the base voltage at which appreciable current flows. In order to make this "overlap diode" effect insignificant, it is necessary to insert in series with this shunting diode a resistance large compared to the base

resistor itself. This can be done by making the overlapped diode extremely small increasing "collector" spreading resistance directly under the contact area. Accordingly, a contact dot as small as consistent with the available state of the technology is chosen for the end of the base resistor. This accounts for the peculiar transistor geometry to be seen in the photomicrographs of Fig. 5-56.

Next, we must choose the transistor surface geometry with an emitter size consistent with the current level at which the circuit will operate. The resistor length-to-width ratios is determined from the sheet conductivity of the base diffusion. The remaining problem is to lay out a geometric configuration which makes efficient use of the area while resulting in as simple an intraconnection pattern as possible. A considerable simplification in layout can be achieved by running all grounded emitter connections to the isolation region itself. In this manner, the back isolation diffusion acts as a common ground plane. This greatly simplifies the lead crossing problem.

The photomicrograph in Fig. 5-56(a) shows a configuration of the isolation region suitable for this circuit. This consists of an array of 0.001" lines. The five collector resistors will be prepared in the upper right hand rectangle, while the other five rectangles will contain the transistors and their associated base resistors. The

circle at the lower right hand corner of the isolation pattern is included so that a common ground contact might be made to the isolation region from the top surface of the device for packaging techniques where it is inconvenient to employ a connection to the bottom of the semiconductor die.

Manufacturing Steps

Table 5-1 shows the process steps required to complete the micrologic element. After the crystal is grown and sliced, the wafers are lapped to a uniform thickness of 120 microns to remove the surface damage introduced by diamond sawing. A controlled etching step reduces the thickness of the wafers to 80 ± 5 microns. This thickness is chosen as the best compromise between difficulty in handling the wafers without breakage through the remainder of the process and complications caused by the long isolation diffusion required to completely penetrate the wafer. This 80 micron dimension determines the 120 micron dimension after lapping, since approximately 20 microns of material must be removed from each surface of the wafer to assure that all of the damage introduced by the lapping operation has been removed. Step 5, oxidation of the wafers, takes place in a wet oxygen atmosphere at 1200°C. It is necessary to build up a relatively thick oxide at this point since the high boron concentration and long diffusion time which will be employed to achieve

Table 5-1

Steps in Micrologic Process

1. Grow Crystal
2. Slice Crystal
3. Lap Wafers
4. Etch Wafers
5. Oxidize Wafers
6. Photoengrave Oxide (to expose isolation pattern)
7. Deposit Boron (in isolation pattern)
8. Diffuse Boron (through wafer)
9. Photoengrave Oxide (to expose base and resistor patterns)
10. Deposit Boron (in base and resistor pattern)
11. Diffuse Boron (to desired base-collector junction depth)
12. Photoengrave Oxide (to expose emitter and collector contact patterns)
13. Deposit Phosphorus (in emitter and collector contact pattern)
14. Diffuse Phosphorus (to desired emitter junction depth)
15. Strip Oxide From Back of Wafer
16. Evaporate Gold on Back of Wafer
17. Diffuse Gold Throughout Wafer
18. Photoengrave Oxide (to expose regions for inter-

connection contacts)

19. Evaporate Aluminum (over entire surface)

20. Photoengrave Aluminum (to the desired interconnection pattern)

21. Heat Treat Wafer (to assure good bonding and ohmic contacts)

22. Scribe and Break Wafer to Dice

23. Sort Dice (optically and electrically)

24. Attach Die to Header

25. Bond Leads

26. Weld Leads

27. Clean Subassembly

28. Vacuum Bake Subassembly

29. Seal Micrologic Element

30. Tumble Element

31. Temperature Cycle Element

32. Leak Test Element

33. Centrifuge Element

34. Test Element

isolation puts a severe strain on its masking ability. Approximately 0. 8 microns of oxide is prepared.

Step 6, the photoengraving of the oxide to expose the isolation pattern, is a straightforward masking operation employing contact printing of the mask pattern on to the wafer surface and etching with fluoride-containing oxide removal etch. Step 7, boron deposition, is performed at 1200°C using a boron trichloride or boron tribromide source. It is desirable to achieve very high surface concentration at this time so a maximum concentration of boron is deposited consistent with no formation of the inpenetrable black skin. The oxide has been completely removed from the back of the wafer during process step 6. The boron deposition is done with the wafers in a standing position in the boat so that both sides receive the deposition. Diffusion of boron in step 8 then proceeds for 24 hours at 1300°C in a slightly oxidizing atmosphere of dry oxygen. This time-temperature combination is adequate to assure that the diffusion fronts from the isolation grid pattern on the wafer top and from the entire backside meet and overlap significantly. It is important that the overlap be relatively large, since we depend upon the conductivity from the grid on the top side to the layer on the back surface for our emitter grounding connection. If these regions to not overlap sufficiently, it is possible to introduce large emitter

resistance values, which interfere with the operation of the circuit.

The next operation, step 9 in the process, is another masking step to expose the bottle-shaped transistor bases and the resistor patterns. The same oxide etch is used as before. In this case it is necessary that the mask pattern be positioned precisely with respect to the initial pattern. Boron is then deposited in step 10 and diffused in step 11 to the desired base collector junction depths. The photomicrograph in Fig. 5-56(b) shows a structure on the wafer at this stage of the operation. The accompanying diagram shows a cross-section through one of the resistors and a transistor. Since the boron diffusion, step 11, was performed in an oxidizing atmosphere a layer of SiO_2 has reformed over the originally exposed areas. This layer is of the order of 0.6 microns in thickness.

Step 12 is another oxide photoengraving operation. In this case, holes through the oxide are opened for the emitter and in those regions where collector contacts will be made to the transistors. Step 13 is the deposition of phosphorus from a P_2O_5 source followed by the diffusion of the emitters in step 14 to the desired depth. For the transistor structures employed here, we desire an emitter junction depth of about 1.5 microns with a base width of the same magnitude. The inclusion of the heavy emitter diffusion

corresponding to a sheet resistivity of about 2 ohms/square over the collector contact areas, serves a two-fold purpose. First, it facilitates the later preparation of ohmic contacts to these regions and secondly, it will decrease the effective collector saturation resistance of the devices. This is especially true since some of these regions will be used for lead crossings; for example, the one in the lower right isolation region is so used. The photomicrograph in Fig. 5-56 (c) shows the device structure, subsequent to this emitter and collector contact diffusion.

The next three steps, 15, 16 and 17 in Table 5-1 are those required to control the storage time of the transistors to a value consistent with the speed requirements of the circuit. In this case, gold is diffused from the back side of the wafer throughout the entire structure and, in particular, into the collector body where it forms recombination centers for any holes injected during the operation of the transistors in saturation. This is a relatively common technique of minority carrier storage control in silicon switching devices. The gold is evaporated on the bare back of the wafer and diffusion proceeds at temperatures of the order of 1000 - 1050°C for times of the order of 10 minutes. Gold is extremely rapid as a diffuser and this time-temperature cycle is adequate to assure appreciable concentrations in the desired regions of the structure,

while not interfering with the diffusion profiles of the ordinary donor and acceptor impurities established previously.

The next step, number 18, is another oxide photoengraving operation that is employed to expose those points where contact must be made to the silicon for the thin film intraconnections. The photomicrograph in Fig. 5-56 (d) shows the structure after this operation. As well as emitter, base, and collector contacts to the transistors, holes have been opened over the resistor stripes. Next, in step 19, the entire surface of the wafer is coated with an evaporated layer of aluminum which is photoengraved in step 20 employing a sodium hydroxide etch to leave the desired intraconnection pattern. The last photomicrograph in Fig. 5-56 shows the wafer subsequent to this aluminum photoengraving operation. While aluminum adheres readily to both silicon dioxide and silicon, it is desirable to give the film a heat treatment to improve this adherence and to assure that good ohmic contacts are obtained. This heat treatment is in the vicinity of the aluminum silicon eutectic temperature. If the eutectic temperature is exceeded, the time cycle must be held short so that the aluminum does not significantly penetrate the silicon dioxide layer.

At this stage, the circuits are complete and need only be separated from one another, packaged and tested. The rest of the operations in the flow table suggests the sequence

of events employed for micrologic elements packaged in a standard Kovar eyelet-type header similar to the TO-5 transistor package. In the case of other packages, for example a flat ceramic package, the sequence of events might be changed slightly.

In summary, one sees that the operations required to make this particular semiconductor integrated circuit include some nine high temperature furnace operations, two metal film evaporations and five indexed masking operations. If each of these major steps is broken into the sub-steps, which must be considered in production, for example, if each masking operation is divided into cleaning, coating, exposing, developing and stripping steps as is commonly done in specifying these processes for production workers, the total number of process steps prior to dicing of the wafer is well over 100. A sequential materials processing operation of this sort can function only if most of the individual steps have yields very close to 100%. These operations can be this well controlled. In other circuit structures which it will be desirable to make in the near future, for example, those including epitaxial growth, thin film resistors and capacitors, and npn and pnp transistors in the same substrate, the total number of processing operations will increase by an appreciable amount. Even so, it appears certain that the increased processing complexity will more than justify the

additional production steps by offering greater electronic function for a given cost.

Properties and Characteristics

Let us now examine some of the properties of the microcircuit which we have made. Perhaps to simplify discussion, it will be more appropriate to discuss the gate or "G" element whose circuit diagram is shown in Fig. 5-57. First consider the isolation junction containing the three transistor structures. Fig. 5-58 is a distribution curve of the breakdown voltages of isolation regions on micrologic gates prepared in this manner measured at 10 μa current. One sees that the breakdown voltage is typically above 100 volts. Since the circuitry was designed to operate at 3 volts, it is obvious that adequately high isolation has been achieved. It can be confirmed that the collector resistances are near the nominal value and experience over extended production runs suggest that ± 20% is a reasonable estimate of the long term accuracy which can be achieved by this production process. The production distribution obtained for this resistor was shown in Fig. 5-33. Fan-in and fan-out measurements over the required temperature range show that the base resistor is performing the desired function to prevent the current hogging problem and that the small base contact has so reduced the effect of the overlap diode that it does not interfere with the proper operation of the base

resistor. Propagation delay times measured at room temperature with fan-in and fan-out both equal to unity show that average propagation delay per gate runs about 20 nanoseconds. The isolation capacitance for the three input gate turns out to be about 60 picofarads at a bias of a few tenths of a volt, the level it experiences in operation. Thus, one estimates that of the 20 nanosecond propagation delay approximately half is contributed by the charging of the parasitic isolation capacitance. Since the 20 nanosecond propagation delay is completely consistent with the speed objective of one megacycle clock rate for a system over the entire temperature range -55°C to + 125°C, the effect of the isolation is compatible with the desired operation.

It should be pointed out that when such circuits are being produced in quantity at reasonable yield that the production capability greatly exceeds that to make a gate or a family of six microcircuits. What exists is a technology capable of making a large variety of circuits, limited only in that they must use only those structures which can be achieved within the limits of the established technology and that the complexity does not exceed some limit which is controlled by yield or other cost considerations. Thus, there is a capability to make custom circuits to fulfill many specific requirements which may arise.

As an example to demonstrate this capability, the dif-

ferential amplifier circuit shown in Fig. 5-59 was breadboarded employing the special integrated circuit components described in Section 4, made with the Micrologic technology. Using the design established during the breadboarding, a set of masks was prepared and integrated versions constructed. Three weeks after the breadboard was firmed up, completed differential amplifiers in wholly integrated form were produced, having the characteristics shown in Table 5-2.

Table 5-2

Typical Performance Data on Darlington Differential Amplifier Shown in Fig. 5-59

Offset voltage referred to input (both inputs grounded)	2mv, -55° to 125°C
Common mode rejection	> 80db, -55° to 125°C
Gain (single ended, 10mv rms input)	25-28db, -55° to 125°C
Band width	500 kc

The performance of the integrated structures was actually somewhat superior to that of the breadboard itself. The process was held constant in this example. Even the gold diffusion for the limiting of lifetime in the collector bodies of the transistors was included. Here we have an example of a technology established primarily for its usefulness of the preparation of a family of digital circuits being used

to make a useful linear amplifier. As more bits and pieces of technology are added to the available production capability, the range and performance of circuits accessible in fully integrated form will expand rapidly. The idea of using special components consistent with the technology in the design of future integrated circuits will certainly decrease lead times and result in rapid turnaround to achieve the lowest cost optimum, special integrated structures.

5-6. COST AND RELIABILITY

Since it is generally agreed that the major promises of the semiconductor integrated circuit approach to microcircuitry are reduced systems costs and improved reliability rather than merely decreased size and weight, it is important to examine in more detail why improved reliability and lower system cost can be expected. First, let us consider cost. This can be divided into the cost of the individual integrated circuits, i.e., the "component" cost, and those costs of assembling these to make a system. The cost of the microcircuits themselves can be further subdivided into three areas: the production cost, the tooling cost and the engineering cost necessary to design the circuit initially.

Consider the production cost. For this it is useful to divide the manufacturing process into two parts which we will call fabrication and assembly. Fabrication will take

in all the batch processing steps necessary to make the circuit die and will include a first electrical test corresponding to step 23 in Table 5-1 for Micrologic. At this test a check for gross electrical problems such as shorts and opens is made on the individual circuit structures by probing on the separate dice. The assembly part of the process will be that from this die sort operation through assembly and including final test. The total microcircuit production cost, C_t, can be considered as made up from the cost of a good die plus the cost of assembly as modified by the yield at final test. That is,

$$C_t = (C_d + C_a)Y_f^{-1} \qquad (5\text{-}3)$$

where C_d is the cost of a good die, C_a is the assembly cost, which equals the total cost of all the operations in this portion of the process divided by the number of units finally tested, and Y_f is the yield at final test, defined as the ratio of the number of good devices to the total tested. But C_d can be expressed in terms of the cost of any die, C'_d, and the die sort yield, Y_d, as follows:

$$C_d = C'_d Y_d^{-1} \qquad (5\text{-}4)$$

Define further $\underline{a}$ as the area of the die, A as the area of the wafer and C_p as the cost of processing a wafer up through the die sort operation. Then

$$C_d \cong C_p \underline{a} (AY_d)^{-1}, \qquad (5\text{-}5)$$

since for $\underline{a} << A$, the usual case, the number of dice

obtained from a wafer is $A/\underline{a}$. Accordingly, the cost of the completed unit is given by

$$C_t \cong \left[C_p \underline{a} (AY_d)^{-1} + C_a\right] Y_f^{-1}$$
$$\cong C_p \underline{a} (AY_d Y_f)^{-1} + C_a Y_f^{-1}. \qquad (5\text{-}6)$$

For small transistors the second term in Eqn. (5-6) usually dominates. That is, the assembly cost exceeds the fabrication cost. The crossover point where the contribution of these two terms is equal occurs somewhere in the range of $\underline{a}$ = 1000-5000 square mils; that is, for a die size of about 30 - 70 miles on an edge. The exact point of crossover varies with the particular product and with the detailed specification which the device must meet. For microcircuits, both C_p and C_a are increased over the transistor case. The increase in C_p results from the increased number of processing steps required during fabrications, while C_a reflects the increased number of leads which in turn affects the price of the package, the lead bonding operation and final test. C_p increases somewhat faster than directly proportional to the number of process steps, since the more process steps introduce additional opportunity for loss of material during fabrication. Even if only the increased probability of dropping batches of wafers on the floor is considered, the rate of increase exceeds direct proportionality. As an indication of how the number of steps for microcircuit compares with that

for conventional transistors, the process flow diagram in Table 5-1 has approximately twice the number of operations as would be included in a similar flow chart for a standard planar transistor. Hence, the processing costs for a microcircuit wafer is something in excess of twice that for the transistor. C_a, on the other hand, generally increases at a rate less than proportional to the number of leads, since only a few of the operations during the assembly portion of the process are directly affected by number of leads and these usually only in approximately direct proportion. Accordingly, the crossover point for microcircuits where the fabrication costs equal the assembly costs moves in the direction of smaller die size. On the other hand, the dice useful in microcircuitry are usually appreciably larger than those associated with ordinary low power transistors. This results in the dice cost being the dominant cost for semiconductor microcircuits.

From Eqn. (5-5) we see that die cost is inversely proportional to the die sort yield, Y_d. To appreciate what influences Y_d it is helpful to examine the cause for shrinkage at die sort testing. This test point is, in general, a place where no precise parameter measuring is done. It is not ordinarily employed as a point where the tails are cut off of normal distribution curves. Rather it is used as a monitor for catastrophic problems. Sometimes, it

can find entire batches which are rejects, but even in good batches one finds that there is a certain occurrence of inoperable structures. In general, this occurrence looks very much like it is tied to the probability of inclusion of a "bad spot" in a sensitive portion of the circuit structure. Those "bad spots" can occur from a variety of mechanisms; for example, a pin-hole in the silicon dioxide layer under a metallized intraconnection can result in a short to the substrate. Similarly, one of the diffusion defects commonly called "pipes" which occurs in the region of a transistor structure can result in an emitter-collector short. Problems of this kind represent major causes for shrinkage at die sort. In general, the problem can be represented by assuming that there is a certain density of "bad spots" spread more or less at random over the surface of the semiconductor wafers. Inclusion of one of these in an area of the structure sensitive to this particular type "bad spot" results in an inoperable device. Hence, assuming random density of the spots, the chance of inclusion of one is increased with the area of device. Accordingly, Y_d is a strong inverse function of device area. For die sort yields less than 50%, which are typical of microcircuits and of large transistor structures, for that matter, the loss of yield due to "bad spots" inclusion varies more rapidly than in inverse proportion to area.

There is the possibility in cases where most of yield loss is because of parameter distribution problems that this picture can be altered appreciably. For example, a microcircuit depending upon an extremely precise ratio between a pair of large resistors might actually show increased yield with an increase in the width of the resistor stripes, so that minor variations in the line widths due to the masking operations would have less influence on the resistor values. In a case like this, however, it is still necessary to consider the total effect on die costs rather than die sort yield alone. Unless increasing the resistor width increases the yield faster than it decreases in the total number of dice per wafer, the total die cost will increase.

Using the above estimate that Y_d varies more rapidly than inversely proportional to device area, we see from Eqn. (5-5) that the cost of a good die varies faster than the square of the die area. This strong dependence upon die size should be borne in mind in any microcircuitry design considerations. Since area becomes the important contributor to cost, components consuming large areas are very expensive. For example, large resistance values made with low sheet resistances should be carefully considered. If one uses the base diffusion to make a 40,000 ohm ± 20% resistor it consumes an area the equivalent of four com-

pletely isolated transistor structures. If greater precision is required in the same resistor so that the line width must be increased from one mil to two mils, another factor of four is introduced into the required area.

On the other hand, extra processing steps required to extend the available technology in the construction of microcircuitry are relatively inexpensive. For example, in order to add thin film resistors to the Micrologic technology, one need add only three additional processing steps -- an evaporation, a masking and a heat treatment. This results in a total increase in the number of fabrication process steps of about 12%. If this additional technology can result in even a small decrease in the area required for a given circuit, in this case if the die size could be reduced by as little as 10% on an edge with no effect on die sort yield, the total cost of producing the microcircuit would be decreased. Since the addition of some of the other technological capabilities can be expected to appreciably decrease die size for a given circuit function, the increased processing complexity which will be incorporated in integrated circuitry production as time goes on will further decrease the production costs.

Of course, one way to assure high die sort yield is to perform no die sort testing, assuring that the yield is

always unity. Since the total cost of the microcircuit is generated by multiplying all of the costs up through final test by the final yield, however, it is more important that Y_f be high than that Y_d be maximized, since Y_f also multiplies the assembly costs. Accordingly, it usually pays to perform tests that can be done easily on the unpackaged dice at the die sort operation if these result in appreciable yield loss. Since the planar techniques employed in making the microcircuits under consideration result in structures which show negligible degradation during the usual assembly operations, the die sort test is an efficient cost reducing operation.

It cannot be overemphasized, however, that the most important way of achieving low cost microcircuits is to design the circuits themselves for low cost. One must limit the component values as much as possible to those achievable in small areas. He must accept large tolerances on individual components and must avoid such luxuries as high voltage requirements wherever possible. To illustrate the importance of these considerations, curve 1 of Fig. 5-60 shows the voltage at 10μa current through the isolation diodes from a production run of Micrologic elements, measured at room temperature. One sees that 70% of the isolation regions have voltages of at least 100 volts for this particular current, Hence, for

individual components one could specify 100 volt ratings and suffer only a 30% yield loss. However, if he is interested in a microcircuit involving several of these isolation regions, the situation can be altered appreciably. The curves labeled 2 to 4 and 8 in Fig. 5-60 are generated from curve 1 for 2, 4 and 8 isolation regions respectively, assuming that each is chosen at random from the distribution of curve 1. For eight such regions one sees from curve 8 that only 10% of the groups would have all their voltage ratings in excess of 100 volts, while 50% exceed 20 volts and 70% exceed 10 volts. Accordingly, a specification on a microcircuit involving eight isolation regions requiring 100 volts isolation capabilities would result in a seven-fold decrease in yield over the identical circuit requiring only 10 volt ratings. While this, in some respects, is a pessimistic case since no allowance was made from the possibility that there may be area correlations so that the high leakage regions are not completely randomly distributed, this correlation is relatively small. The qualitative conclusion that an unrealistically high voltage rating results in an extremely severe cost penalty is certainly correct. It has been shown, in addition, that the planar, protected junctions of the type shown in Section 3 are very stable. Even when they show excessive leakage initially, they show no tendency to drift. Accordingly, for operation at 10 volts structures good to

100 volts are not necessarily more reliable than those good only to 10 volts initially. Specifications should be established realistically on the basis of need. Unrealistic specifications are much more expensive in the case of microcircuits than they are for individual components. Adjusting circuit design toward low voltage is, in addition the correct direction in which to go to achieve greater reliability from other considerations as well.

In order to evaluate the promise of the integrated circuitry approach with respect to cost, it is useful to compare directly with transistors. A transistor in an isolated region similar to one shown in Fig. 5-27 including all contacts made to the top surface requires an area of about 100 square mils, assuming that one restricts himself to one mil line widths and separations and 0.5 mil total indexing tolerances for successive masks -- the tolerances allowed in Micrologic. A conventional separate transistor on the other hand requires a 20 mil square die as about a practical minimum for ease of handling. If one allow a 10 mil border completely around the edge of a microcircuitry die for lead attachment and for separation of one microcircuit and the next, he can get 64 of the isolated transistors mentioned above in a tenth inch square area. Alternatively, in place of a transistor he can have four kilohms resistor. On the other hand, one gets only 25 separate transistors

from an area this size. The microcircuit assembly cost would be much less than that for the 25 transistors, since it will undoubtedly have very many fewer leads; and the package which contains it will certainly be less expensive than 25 separate transistor packages. Therefore, one can expect the microcircuit equivalent to a 64 transistor circuit to cost much less than the 25 transistors if the yields are at all comparable. Fortunately, for the microcircuits, yields are defined somewhat differently than for the individual device structures. In the case of an individual transistor one is faced with the problem of specifying the parameters of this device so that it will operate when used in any one of thousands of circuits which can be built from this particular transistor type. This requires some rather arbitrary numbers concerning the various device parameters to define a good unit. In the case of the transistor incorporated into a microcircuit, however, the definition of yield becomes more closely related to its particular operation. This transistor is good if when used with exactly those other components in the circuit it will perform the specific function required of the circuit itself. This is a much less stringent requirement than the former and results in yields of entire microcircuits being of the same order as those for individual components. Indeed, this order of magnitude equivalence of yields is necessary for

the production of microcircuits to be feasible. Fortunately, it is the case. Actually the example considered above is a worse case in the respect that all of the transistors in the microcircuit were considered as requiring separate isolation regions. In actual practice in such a complex circuit there would usually be several cases where common regions could be used, resulting in the possibility of even more transistors in the given area. Also, the inclusion of additional technologies, for example, the addition of thin film resistors or of some of the epitaxial isolation techniques, can further increase the amount of circuitry per unit area. Even at the circuit complexity level of the micrologic gate -- three transistors, four resistors and six lead bonds -- the cost of producing the integrated circuit is approximately equal with the cost of producing the individual components in equivalent quantities. As the complexity is increased, microcircuits are favored more and more strongly until one reaches the point where the yield because of the complexity falls below a productionworthy value. The point at which this occurs will continue to push rapidly in the direction of increasing complexity. As the technology advances the size circuit function practical to integrate will increase rapidly for a variety of reasons. The problems associating yield and area will be better understood and the yield improvements which will result will allow the use of larger areas.

More circuitry will be possible in a given area through the use of finer scale structures. For example, changing from the one mil line width with 1/2 mil indexing tolerances to 1/2 mil line width with 1/4 mil indexing tolerance is probably within the present state of the art. This decreases the area of a given circuit by a factor of four. Additionally other structures, such as some of those described in Section 3 and others still to be conceived, will increase the available electronic function per unit area.

Another contribution to the cost of microcircuitry that must be considered is the tooling cost required to make a new circuit. If one considers that a standard production technology exists within which one can make a large variety of circuits by changing only the set of masks employed and the electrical testing, the tooling cost to be considered is that of masks and test equipment. Two approaches to the mask problem have been considered. One involves making a complete set of optimum masks for each new circuit. The other considers the standard item to be a wafer containing a variety of individual component structures such as transistors and resistors, rather than an available technology, and makes only one mask -- the intraconnection mask -- which is superimposed on this "master wafer" to intraconnect the already existing components into the new circuit[52]. This single mask technique offers the

advantage of rapid turn-around time because it circumvents a delay of up to several weeks in the early fabrication steps inherent in a smoothly flowing production line with adequate in-process inventory. A single mask requires less time to prepare than a complete set of masks and can be made for a few hundred dollars. This single mask, master wafer technique, however, obviously makes extremely non-optimum use of the available area. Accordingly, it is suitable only for very small quantities of microcircuitry, probably less than 100, and where short turn-around time is important. An optimum set of masks appears the more economical approach for any microcircuit requiring more than several hundred finished structures and for any very complex circuit. A set of masks optimized for a particular circuit probably costs two to four times the cost of an individual mask. Since even in this case the mask amortized over the first several hundred circuits contributes only a small fraction of their cost, it is not an overriding consideration except for extremely small usage items where the desirability of integration must be especially carefully considered anyhow.

Testing costs are a variable that is difficult to estimate in general since they are so dependent upon detailed specifications. It is possible to conceive of quite flexible microcircuitry test equipment having an assortment of available

forcing functions and measuring schemes that can be programmed for a specific circuit. Testing a microcircuit with such a tester should be less expensive than testing the equivalent number of separate components.

At least for the present the third contribution to the cost of producing microcircuits, the circuit engineering expense, is not likely to decrease since the circuit design engineer experienced at designing with standard components must familiarize himself with an entirely new set of boundary conditions. After this initial educational period, which can certainly be compressed through maximum early consultation with the integrated circuit manufacturer, the circuit engineering expense should be no higher than with conventional electronics. In fact the use of the integrated technology will simplify the job of component selection. An important part of the circuit designer's job will be the consideration of appropriate functional tests to assure proper circuit performance.

While detailed analysis of system cost is beyond the scope of this chapter some general arguments can be made. Consider as an example the logic portion of a computer. Beyond the components and other parts cost, the cost to complete the system can be divided into engineering, assembly and checkout. Engineering costs with microelectronics should be greatly reduced because only the

layout of the logic need be considered. Because of many fewer parts and, accordingly, simpler layout, assembly cost is also decreased. Beyond that, checkout is simpler and less expensive, since there are fewer available test points and fewer possibilities for error. Each of the circuits has been completely evaluated prior to insertion. It has been estimated that the cost of designing, assembling and checking out a first prototype logic section of a digital computer using Micrologic would be only 20% of that using standard components[53]. Early experience confirms this prediction at least qualitatively[54].

In considering the impact of semiconductor integrated circuitry upon system reliability one must appreciate that each integrated circuit replaces many components and interconnections compared with a system made from conventional components. In order to improve overall system reliability it is necessary only that the integrated circuit exceed the reliability of the total of the features which it replaces. Not enough experience with systems employing integrated circuits has been accumulated to date to demonstrate that the anticipated improvement in reliability has been realized, although what experience does exist is not in contradiction with the prediction. In order to estimate the magnitude of the improvement to be expected it is useful to look at the failure mechanisms for these circuits

and to compare them with devices of established reliability. The most obvious comparison to make is with silicon planar transistors, for which a large amount of data has been accumulated.

The semiconductor integrated circuits employing identical pn junction structures with the planar devices and are dependent upon the same silicon dioxide insulating layer for passivation. Even the feature of running the metallization over the silicon dioxide is used in some transistor types and has been the subject of extensive reliability evaluations. The leadbonding and packaging for microcircuits can be identical to that for individual transistors. For example, the package usually employed for Micrologic elements is identical with the transistor package except for the number of leads. Not all packages proposed for microcircuits are directly comparable with conventional transistor packaging. Those which differ must be considered independently. The principal additional failure modes which one introduces as possibilities in such integrated circuitry as Micrologic are those associated with the long metal intraconnections. The possibility of the occurrence of breaks in these intraconnections does not exist in the same degree in conventional transistors. Of course, the addition of new technologies such as film resistors adds additional failure modes. No single change in technology, however,

introduces more possibilities of new failure modes than does a change in the basic packaging scheme.

On ultrareliable transistors the residual failure mode is generally mechanical rather than any drift in the semiconductor structure itself. This has been confirmed both for the case of large scale life testing, such as that to which the planar transistors for the Minuteman Missile System were subjected, and for accelerated testing wherein a relatively small quantity of units are stressed until failure. For the Minuteman device failure rates less than 10^{-8}/hour were established. The residual mode was the opening of lead bonds. One sees no *a priori* reason why microcircuit reliability should differ significantly from the transistors providing the packaging is equivalent since there are no other important additional failure modes expected. Considerable life test data has been accumulated on Micrologic elements. More data is available on these than on any other integrated circuits. Both operating life and high temperature storage data exists. Over 6,400,000 hours of operating life test at 125°C ambient, the maximum operation temperature for these structures, has been accumulated representing some 2,070 elements. These were production units, mostly gates although several flip-flops were also included. Most of the units were rejects from the ordinary final test specifications for such problems as

high isolation leakage or low available node current resulting from high collector load resistance values. Accordingly, if borderline units represent an increased tendency toward failure, this life test should be a worse case. One failure was observed in these tests. This resulted from a lead bond opening on one device. A portion of these units were monitored in detail for parameter drift. No measurable drift was observed.

The storage life data showed no failures at 150°C storage in 266,000 unit-hours. At 300°C storage, however, 16 failures were observed in 840,000 unit-hours, resulting in an estimated mean-time-to-failure of 52,500 hours. All 16 failures were opens or high resistance in the lead connection. This mean-time-to-failure at 300°C storage can be compared directly with a mean-time-to-failure of 206,000 hours under identical storage conditions measured for the Minuteman planar transistors. The ratio of a factor of four in mean-time-to-failure is somewhat greater than the ratio of lead bonds of 5/2. If this difference is significant is difficult to say since the two products represent a different degree of production maturity. It does establish, however, that the microcircuit demonstrates stability about comparable to an individual component. Accordingly, a system composed of a given number of microcircuits should be about as reliable as a system

containing that number of components; or stated another way, a given system made from microcircuits should demonstrate an appreciably improved reliability compared to the equivalent system made from conventional components, since fewer components would perform the same function.

5-7. SUMMARY AND CONCLUSIONS

The structures and technology described above are reasonably representative of the present state of semiconductor integrated circuitry. This concept of functional electronics wherein completed electronic functions are produced within and upon a single monolithic block of semiconductor material is rapidly assuming a leading role in advanced electronic systems. Its promise of improved reliability at decreased cost is being realized. Just as the performance capabilities of transistors is seldom limiting the performance of systems made by conventional electronics today, the microcircuits will soon deliver as high performance as is useful for the entire system. With decreased cost and improved reliability, larger and more complex systems than ever before practical can be considered. The amount of circuit function economically to put in a single functional electronic block will increase rapidly. This, in turn, will further contribute to cost and reliability improvements in systems of a given size. While we have used digital com-

puters and logic circuits as examples because the advantages are most dramatic in these cases where large numbers of repetitive circuits are employed, the advantages will carry over into all of electronics. We can fully expect to find integrated semiconductor functional electronic blocks in consumer products when the technology has assumed a higher degree of maturity.

References

1. "Molecular Electronics Receiver," Semiconductor Products, April 1962, p 46.

2. J. A. Hoerni, IRE Electron Devices Meeting, Washington, D. C. (Oct. 1961); Fairchild Semiconductor, Mountain View, Calif. Technical Article TP-14 (1960).

3. D. L. Klein, G. A. Kolb, L. A. Pompliano and M. V. Sullivan, The Electrochemical Society Meeting, Indianapolis (May 1961).

4. See H. Reiss and C. S. Fuller, Semiconductors, Edited by N. B. Hannay, Reinhold Publishing Corporation, New York (1959) for an extensive bibliography.

5. See the first reference cited in Reiss and Fuller, above.

6. R. L. Batdorf and F. M. Smits, J. Ap. Phys. 30, 260 (1959).

7. C. S. Fuller and J. A. Ditzenberger, J. Ap. Phys. 27, 544 (1956).

   R. C. Miller and A. Savage, J. Ap. Phys. 27, 1430 (1956).

   C. T. Sah, H. Sello and D. A. Tremere, J. Phys. Chem. Sol. 11, 288 (1959).

8. F. A. Trumbore, Bell System Tech. J. 39, 205 (1960).

9. L. A. D'Asaro, Sol. State Electronics 1, 3 (1960).

10. A. D. Kurtz and C. L. Gravel, J. Ap. Phys. 29, 1456, (1958).

11. C. J. Frosch and L. Derrick, J. Electrochemical Soc. 104, 547 (1957).

12. M. M. Atalla, E. Tannenbaum and E. J. Scheibner, Bell System Tech. J. 38, 749 (1959).

13. J. E. Sandor, The Electrochemical Soc. Electronics Division Abstracts 11, 228 (May 1962).

14. M. M. Atalla, Proceedings of the Conference on Elemental and Compound Semiconductors, H. Gatos, Editor, Interscience Publishers, New York (1960).

J. R. Ligenza, J. Electrochem. Soc. 109, 73 (1962).

15. C. T. Sah, H. Sello and D. A. Tremere, J. Phys. Chem. Solids 2, 228 (1959).

16. L. B. Valdes, Proc. IRE 42, 420 (1954).

17. W. W. Happ and W. Shockley, Bull. Am. Phys. Soc. Series II 1, 382 (1956).

B. McDonald and A. Goetzberger, J. Electrochem. Soc. 109, 141 (1962).

18. W. R. Wilcox and T. J. LaChapelle, The Electrochemical Soc. Electronics Division Abstracts 11, 194 (May 1962).

19. P. Baruch, C. Constantin, J. C. Pfister and

R. Saintesprit, Disc. Farraday Soc. 31, 76 (1961).

20. A. P. Hale and B. D. James, IRE Electron Devices Meeting, Washington, D. C. (Oct. 1960).
21. J. C. Marinace, IBM J. Research and Development, 4, 248 (1960).
22. H. C. Theurer, J. Electrochem. Soc. 107 , 29 (1960).
23. S. K. Tung, Metallurgy of Semiconductors, Vol. 15, 87, Interscience Publishers, New York (1962).
24. J. Corrigan, Metallurgy of Semiconductors, Vol. 15, 103, Interscience Publishers, New York (1962).
25. B. E. Deal, J. Electrochem. Soc. 109, 514 (1962).
26. R. Glang and E. S. Wajda, Metallurgy of Semiconductors, Vol. 15, Interscience Publishers New York (1962).
27. H. Basseches, S. K. Tung, R. C. Manz and C. O. Thomas, Metallurgy of Semiconductor Materials, Vol. 15, Interscience Publishers, New York (1962).
28. R. M. Warner, Electronics 35, 49 (1962).
29. W. C. Dash, J. Ap. Phys. 33, 2395 (1962).
30. W. G. Spitzer and M. Tannenbaum, J. Ap. Phys. 32, 744 (1961).
31. C. C. Allen and E. G. Bylander, Metallurgy of Semiconductor Materials, Vol. 15, J. B.

Schroeder, ed. Interscience, New York (1962), p. 113.

32. KODAK Photosensitive Resists for Industry, Eastman Kodak Company Publication No. P-7, Rochester, 4, New York (1962).

33. H. Kressel, H. Veloric, A. Blicher and D. Rauscher, The Electrochemical Soc. Electronics Division Abstracts Vol. 11, 190 (May 1962).

34. J. L. Langdon, W. E. Mutter, R. P. Pecorarv and K. K. Schuegraf, IRE Electron Devices Meeting (Oct. 1961).

35. Solid Circuit Type SN510-515, Bulletins No. DL-S 612015-8, Texas Instruments, Inc. , Dallas, Texas (Oct. 1961).

36. A. Evans, Research on Semiconductor Single Crystal Circuit Development, Electronic Technology Laboratory, Contract Nr. AF33(616)6600, Wright Air Development Division, Air Research and Development Command, Technical Note 61-25, United States Air Force, Wright-Patterson Air Force Base, Ohio (Aug. 1961).

37. R. E. Crippen, Private Communication.

38. W. M. Lafky, Private Communication

39. R. E. Craig and D. E. Farina, Private Communication.

40. Second Quarterly Report Compatible Techniques for

Integrated Circuitry, U. S. Air Force Contract Nr. AF33(616)8276, prepared by Motorola Semiconductor Products, Inc., for U. S. Air Force Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio, p. 33 (1961).

41. W. J. Ostrander and C. W. Lewis, Transaction of the Eighth Vacuum Symposium, Luther E. Preuss, Editor, Pergamon Press, Oxford(1961).
    M. Beckeman and R. E. Thun, ibid.
42. G. W. A. Dummer and J. W. Granville, Miniature and Microminiature Electronics, John Wiley and Sons, Inc. New York (1961) p. 247.
43. Ref. 40, p. 19.
44. W. Shockley, Proc. IRE 40, 1365 (1952).
45. C. T. Sah, Proc. IRE 49, 1623 (1961).
46. L. A. D'Asaro, IRE WESCON Convention Record (1959) Part 3, p. 37 (1959).
47. W. F. deBoice and J. F. Bowker, Electronics 35, 37 (1962).
48. Texas Instruments, Inc., Contract No. AF33(616)-8339, Wright-Patterson Air Force Base, Ohio.
49. J. Kilby, Presented at IRE 4.10 Committee Meeting, Seattle, Wash. (Aug. 1962).
50. J. A. Narud, W. C. Seelbach, N. Miller, Computer Design, 1, 26 (1962).

51. "Micrologic," Fairchild Semiconductor, A Division of Fairchild Camera & Instrument Corp., Mountain View, Calif. (Aug. 1961).

52. Electronic Equipment Engineering 10, 14 (Sept. 1962).

53. J. Nall, R. Anderson, D. Farina, R. Norman, J. Campbell, LA Section, AIEE Microsystems Conference (May 1961).

54. B. Miller, Aviation Week and Space Technology, (May 21, 1962).

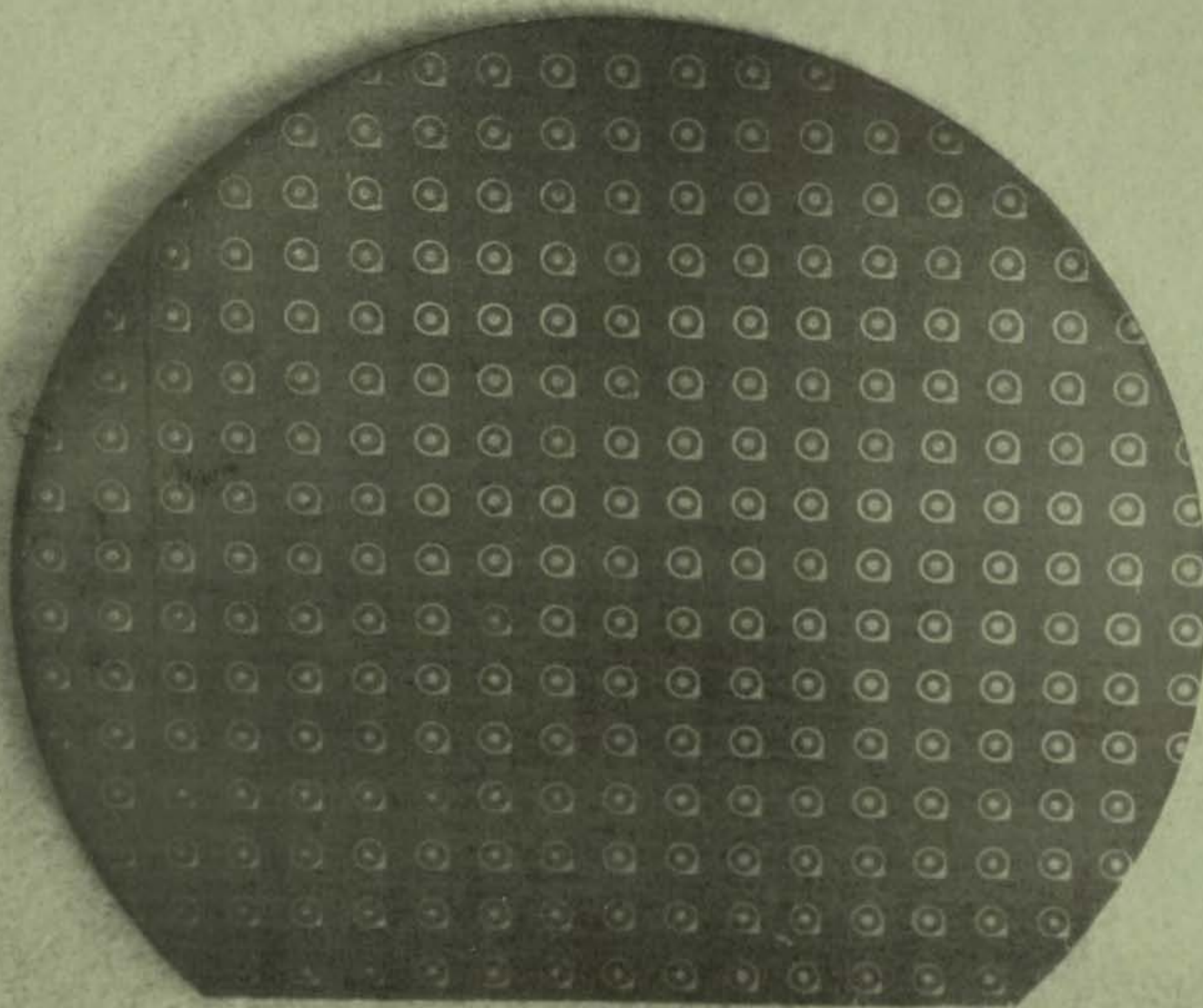
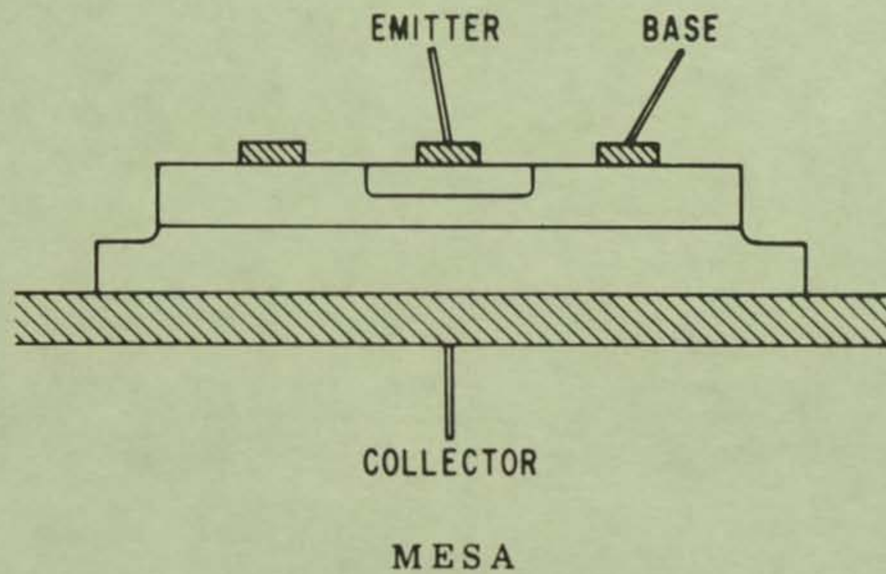


Fig. 5-1. A typical array of double diffused silicon transistors. Each ring and dot pattern is a transistor structure. Only the metallized emitter and base contacts are visible in this photomicrograph. The silicon wafer is approximately one inch in diameter. The flat edge is used for indexing during process.



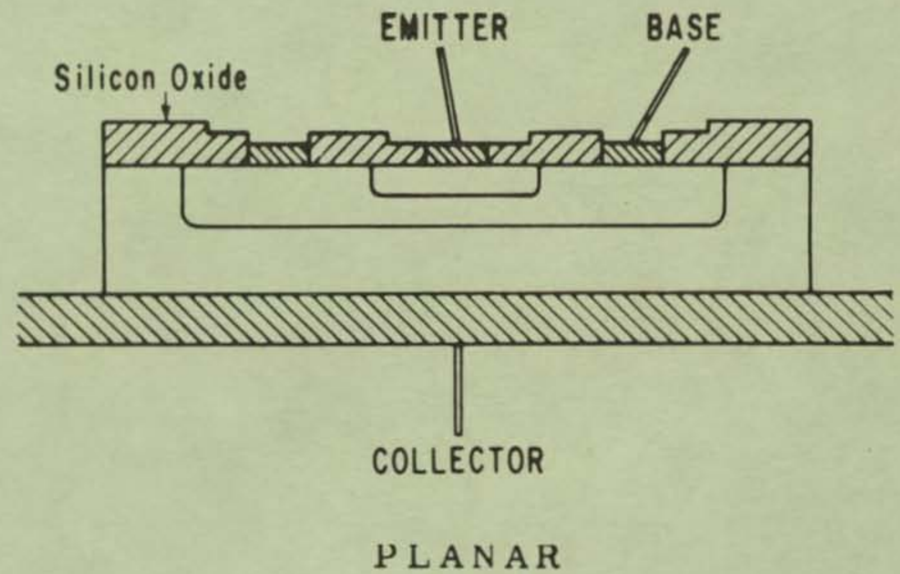


Fig. 5-2. Schematic cross-sections of mesa and planar transistors. The metal emitter base and collector contacts normally employed are shown. Note the silicon oxide layer covering the intersection of the junctions with the surface in the planar case.

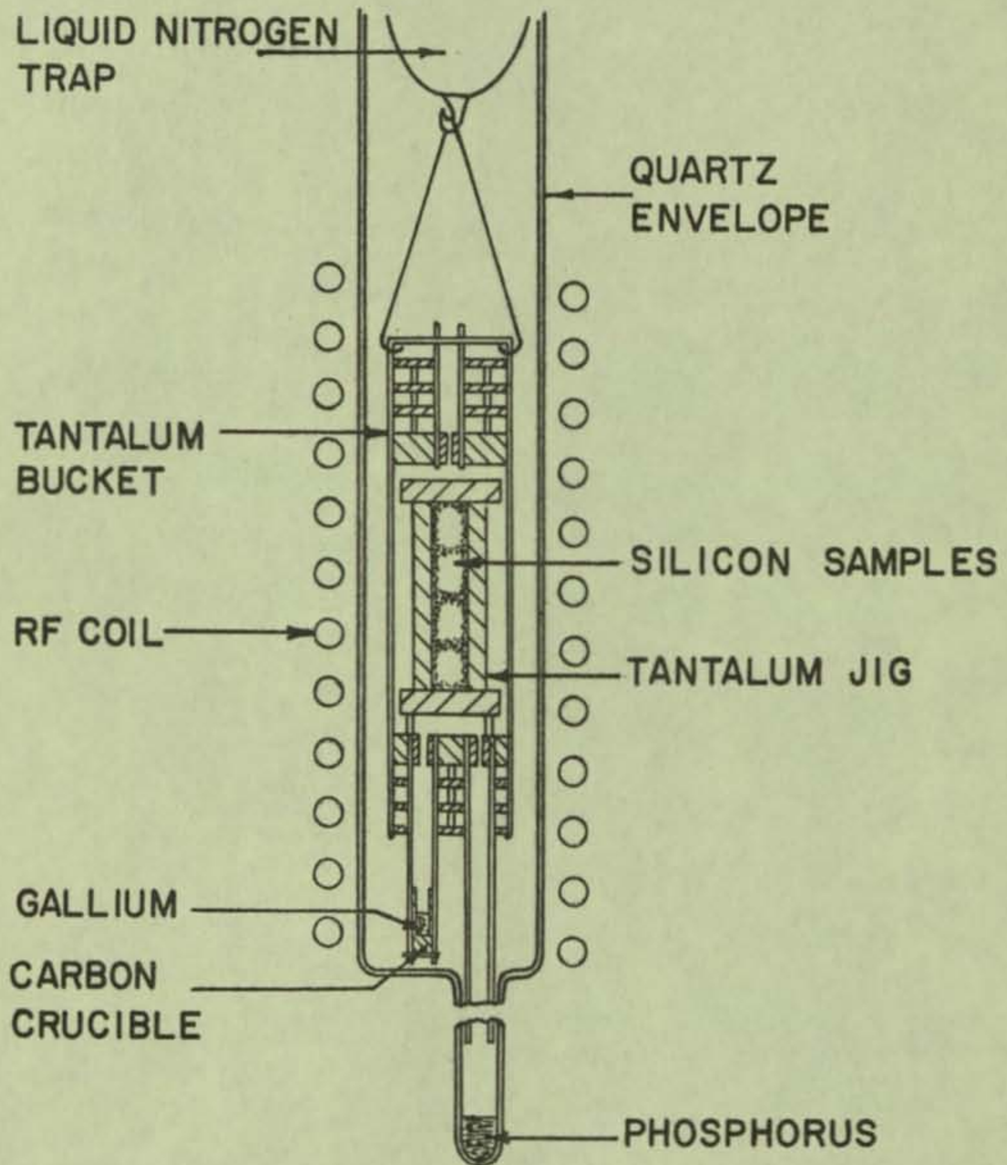


Fig. 5-3. Cross-section of a vacuum station used to diffuse phosphorus and gallium simultaneously into evaporating silicon (from Batdorf and Smits, J. Ap. Phys. 30, 260 (1959).

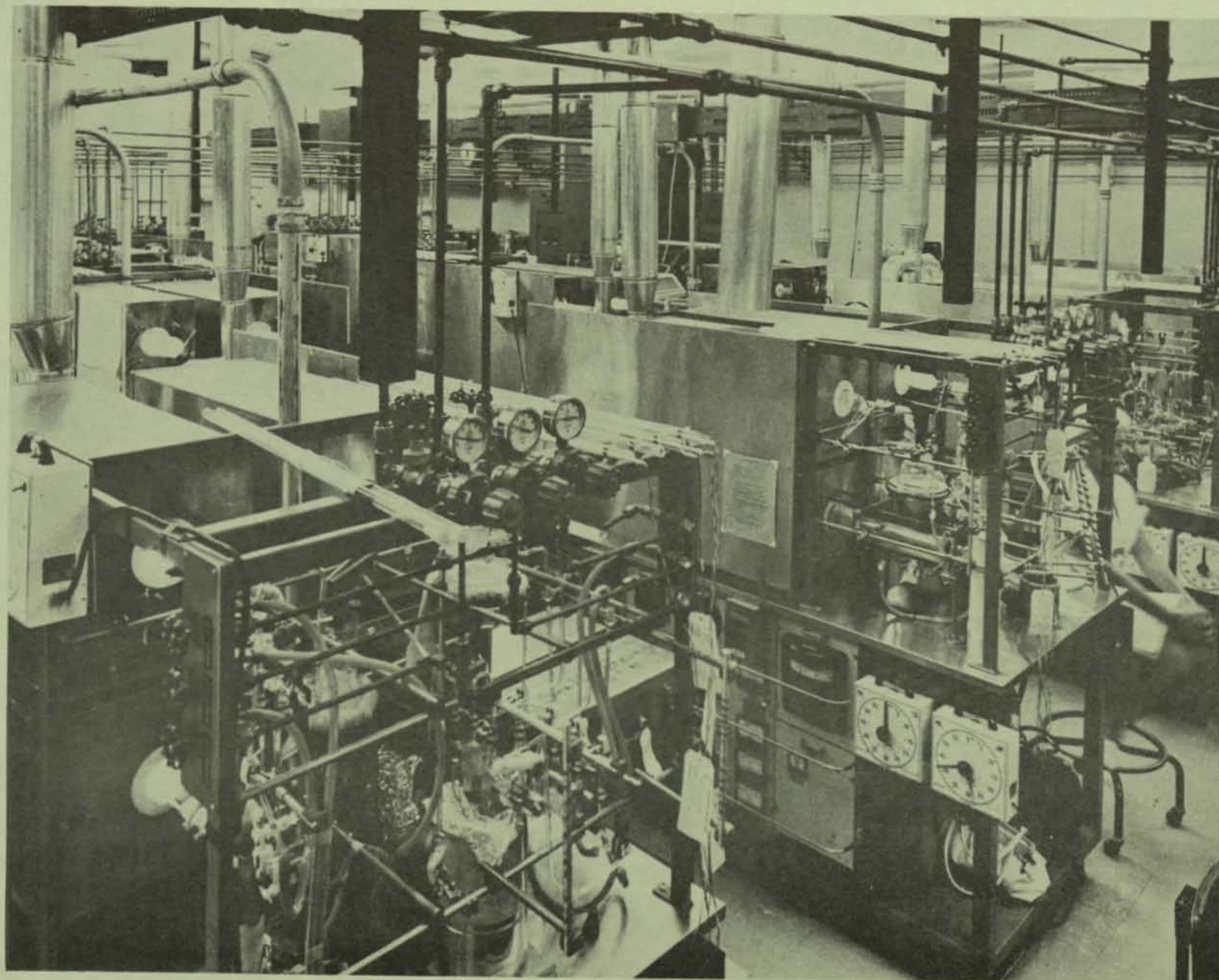


Fig. 5-4. An array of open tube diffusion furnaces used in the production of silicon diffused devices.

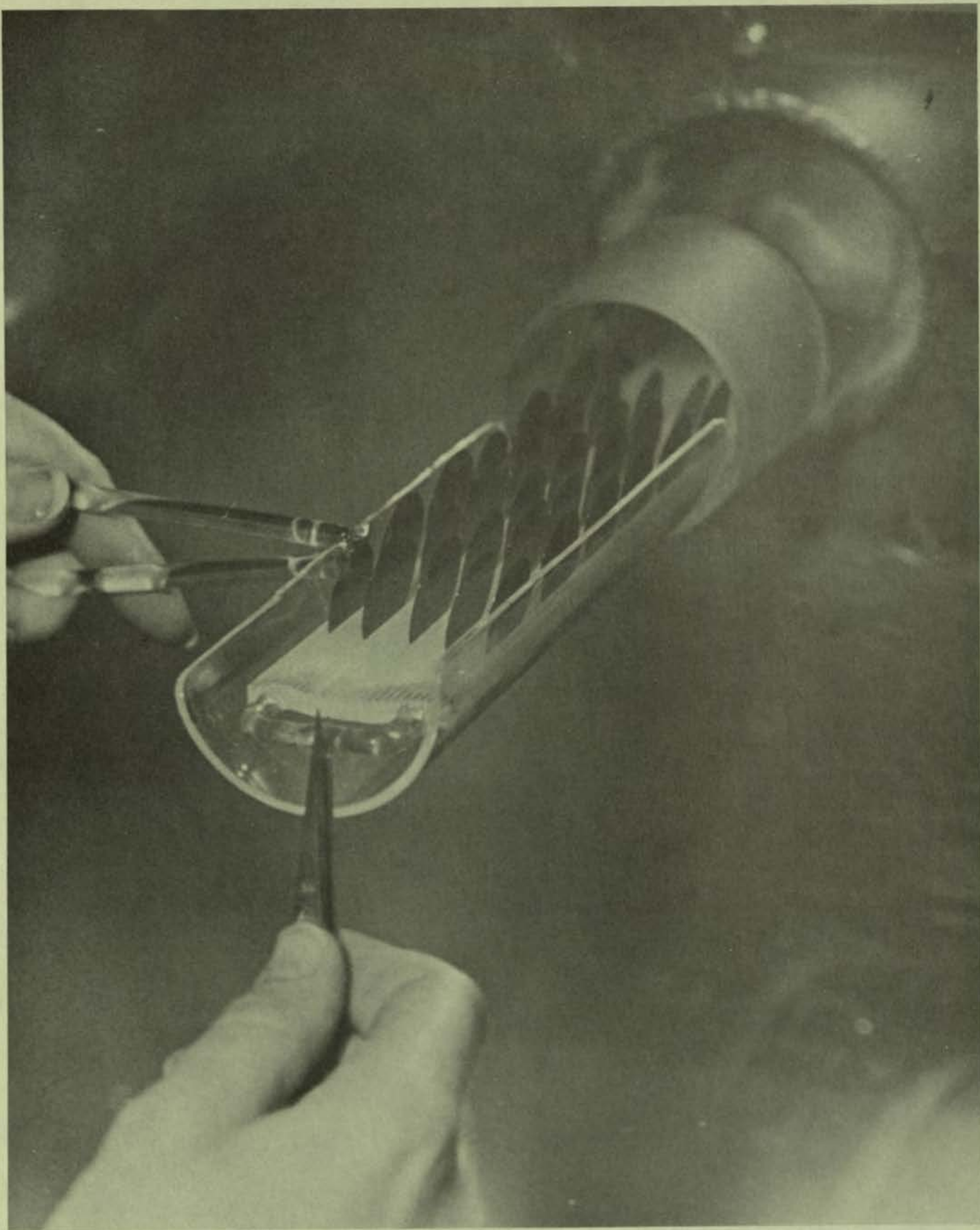


Fig. 5-5. Silicon wafers in a silica boat ready for insertion into a diffusion furnace.

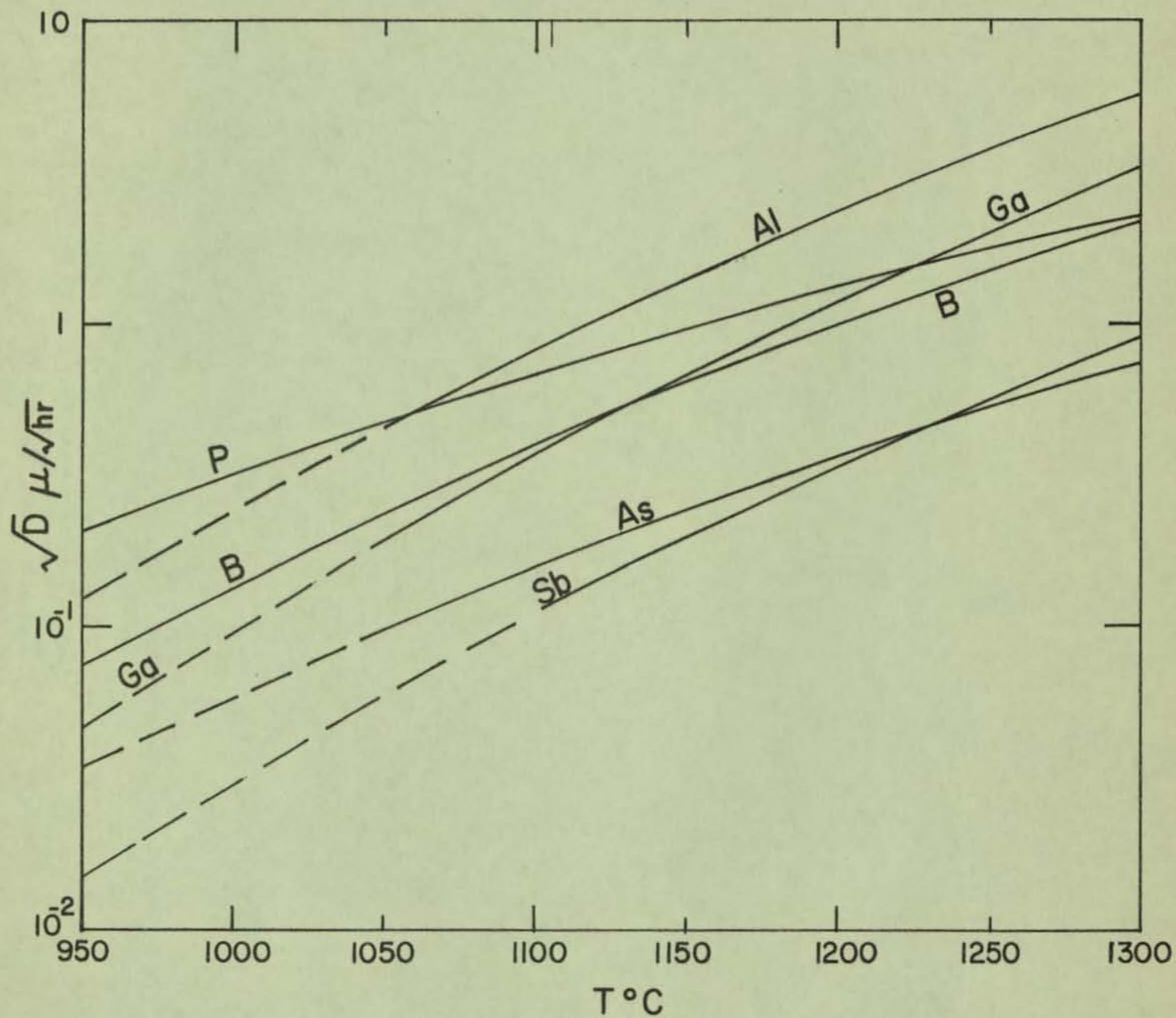


Fig. 5-6. Square root of the diffusion constant versus temperature for the common donor and acceptor impurities in silicon.

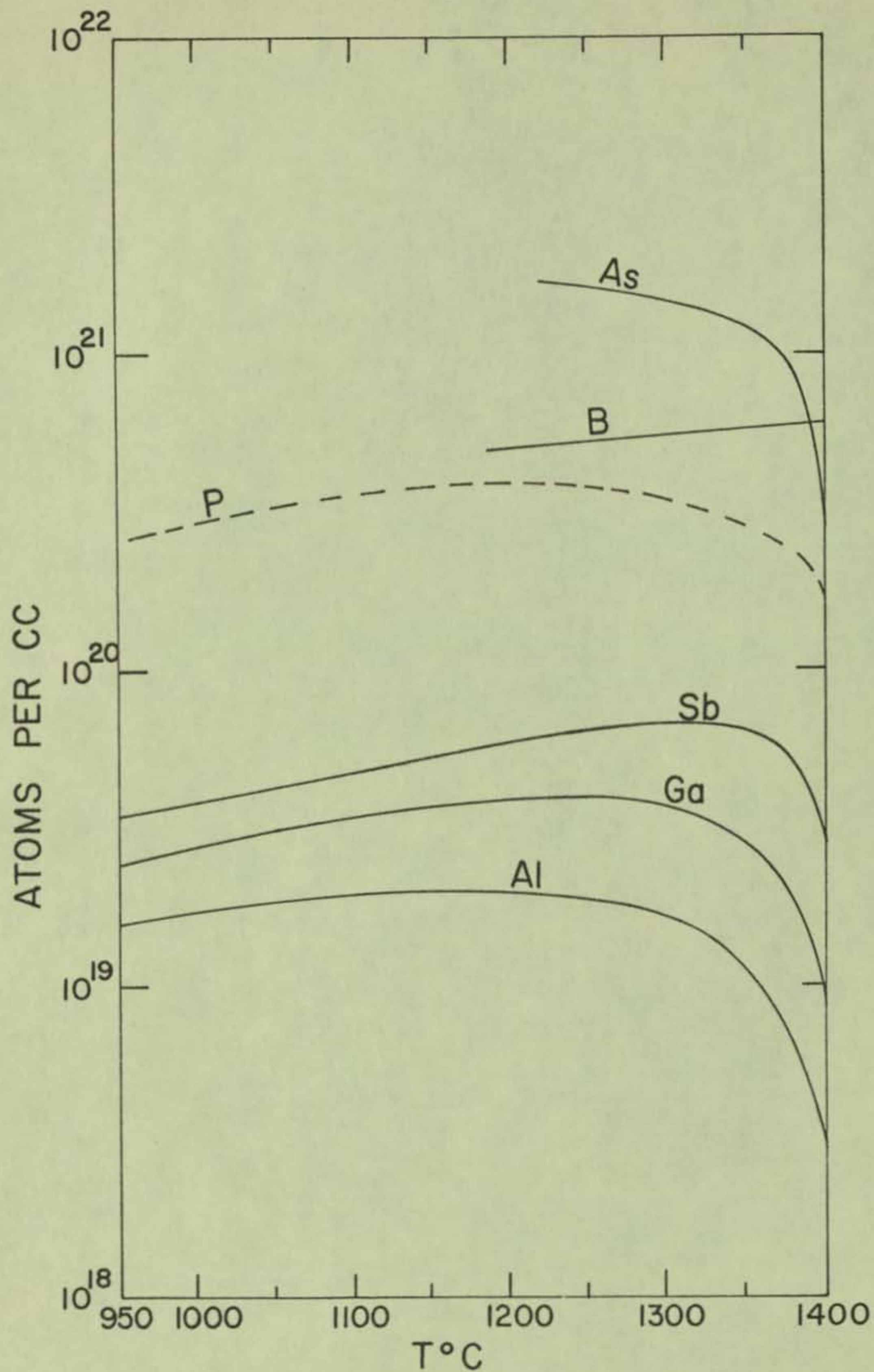
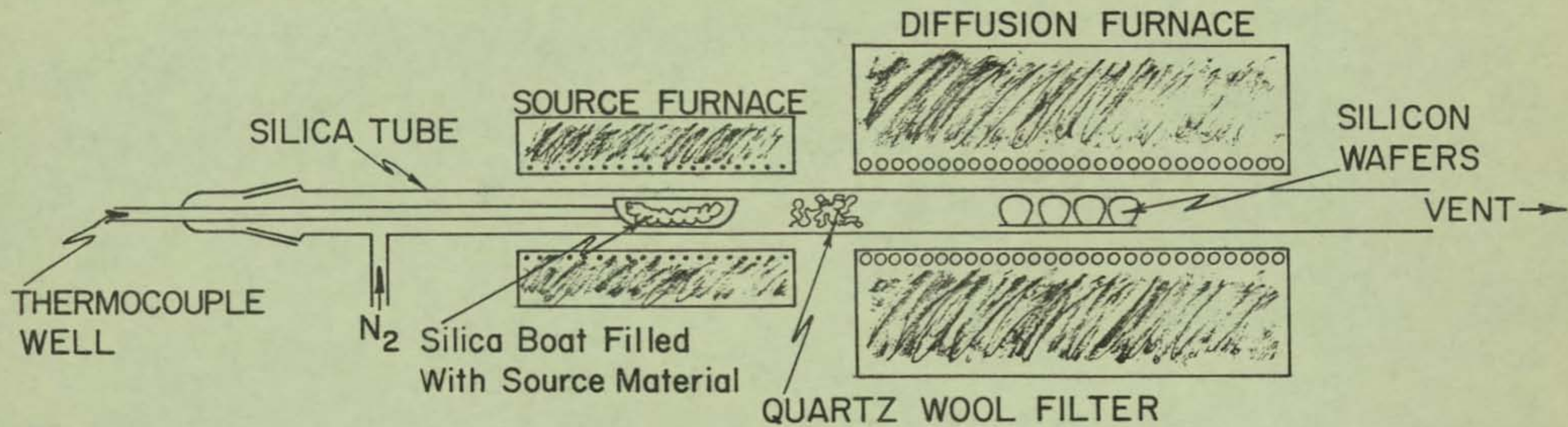


Fig. 5-7. Solid solubilities of common donor and acceptor impurities in silicon vs. temperature (after Trumbore, Bell System Tech. J., 39, 205 (1960).



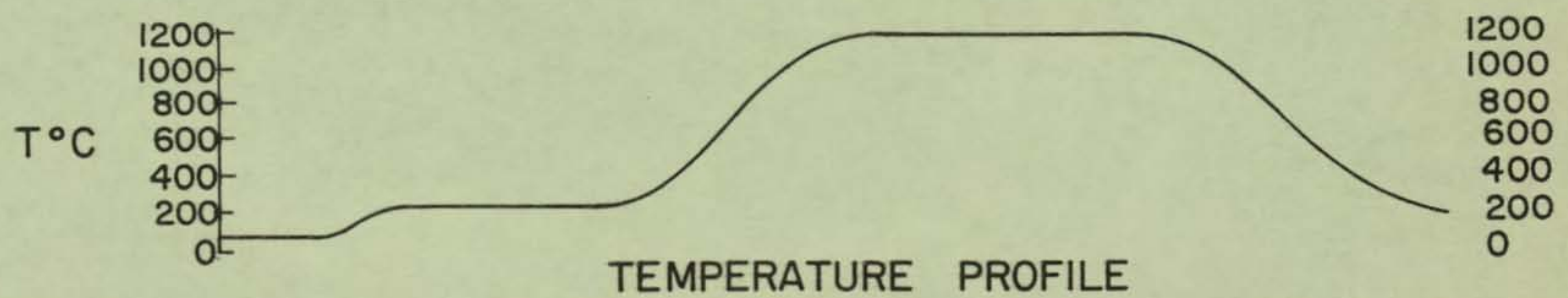


Fig. 5-8. Schematic diagram of a diffusion system suitable for use with P_2O_5 or red phosphorus source material. A typical temperature profile is also shown.

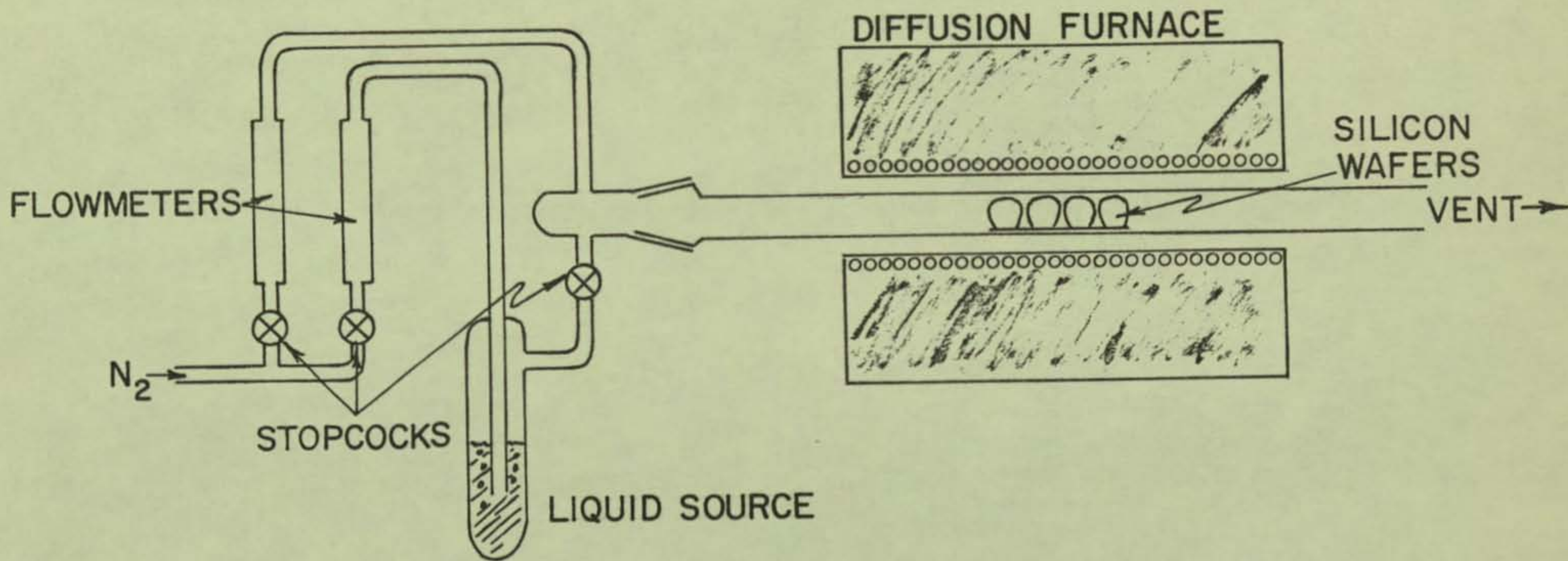
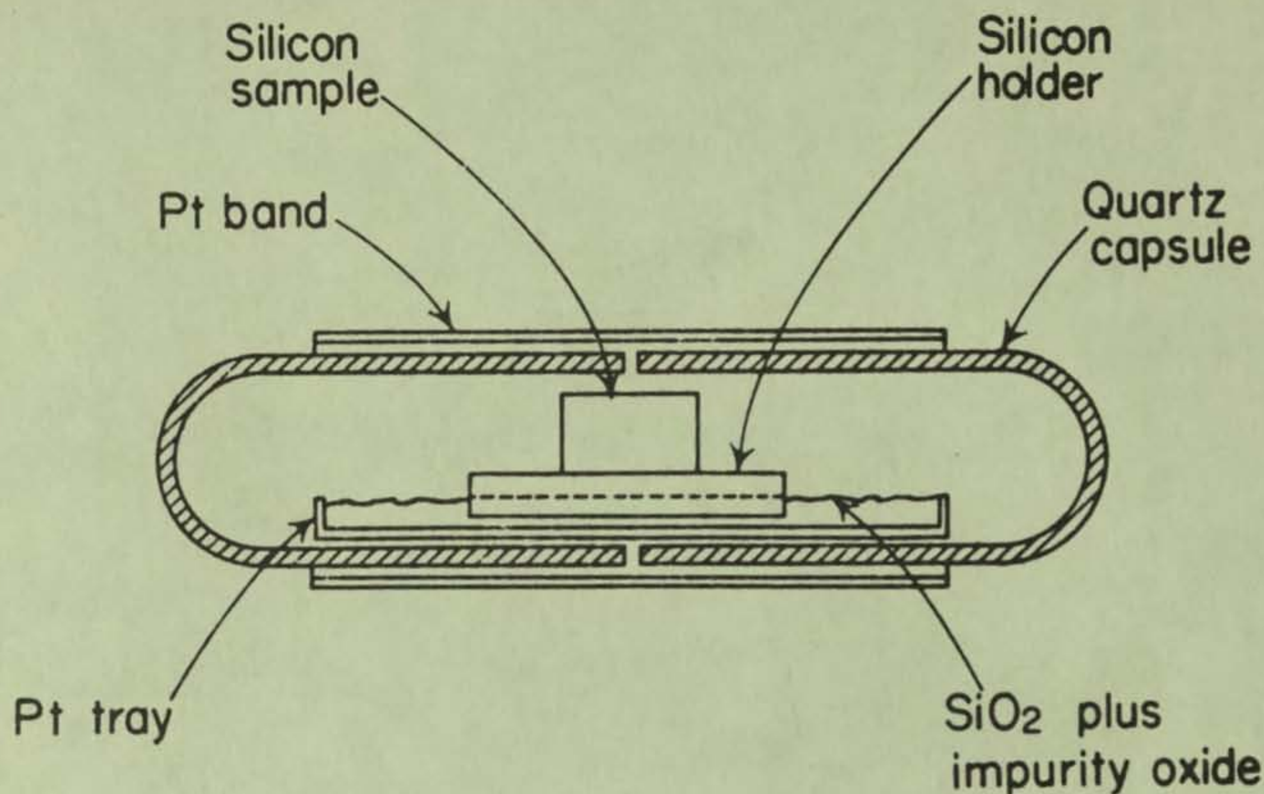


Fig. 5-9. Schematic diagram of a system suitable for the use of liquid source material, such as $POCl_3$.



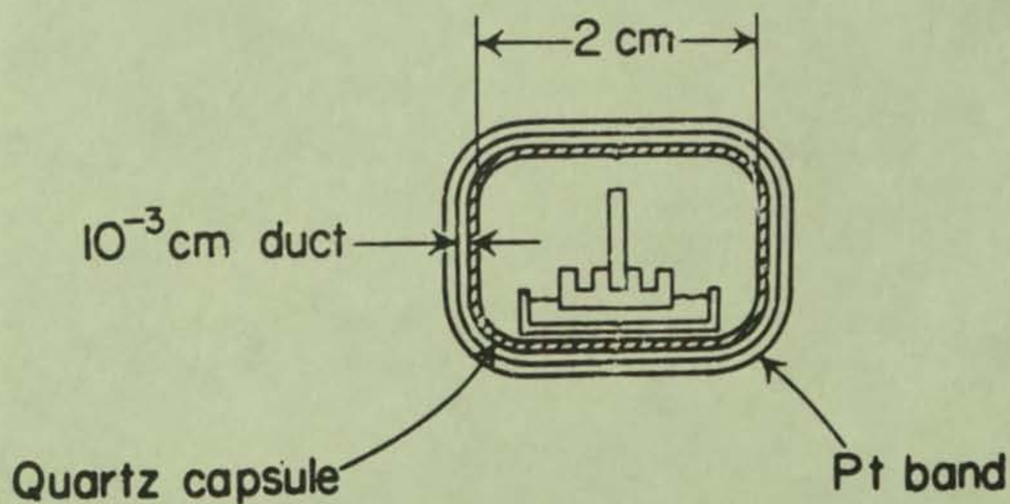


Fig. 5-10. Schematic diagram of apparatus used for boron diffusion by the box technique. (From D'Asaro, Solid State Electronics, 1, 3 (1960).

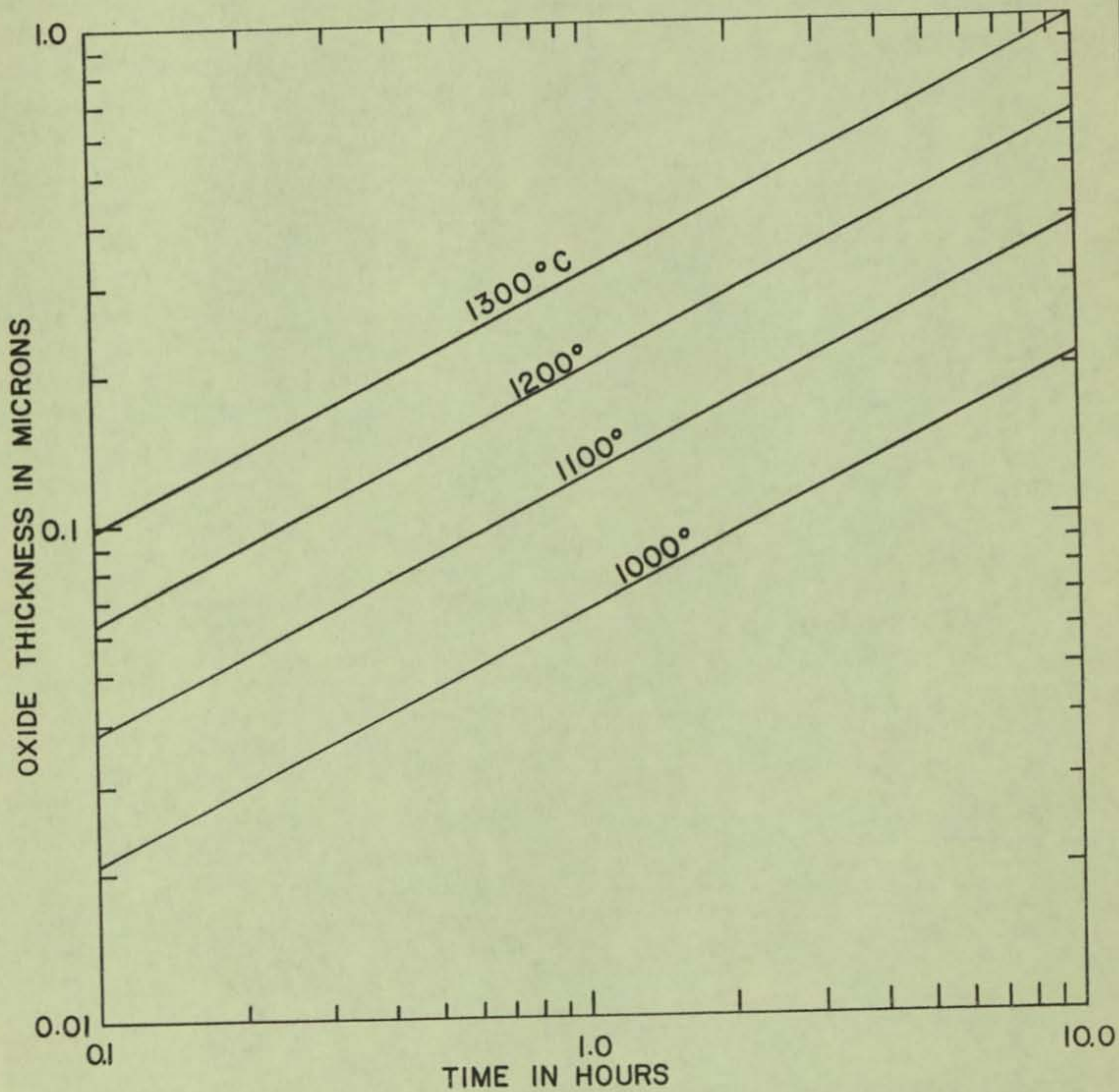


Fig. 5-11. Oxide thickness vs. oxidation time in pure dry oxygen with various temperatures for relatively pure silicon. (Flint, The Electrochemical Society, Electronics Divn. Abstracts, Vol. 11, Spring Meeting, p. 222 (May 1962).

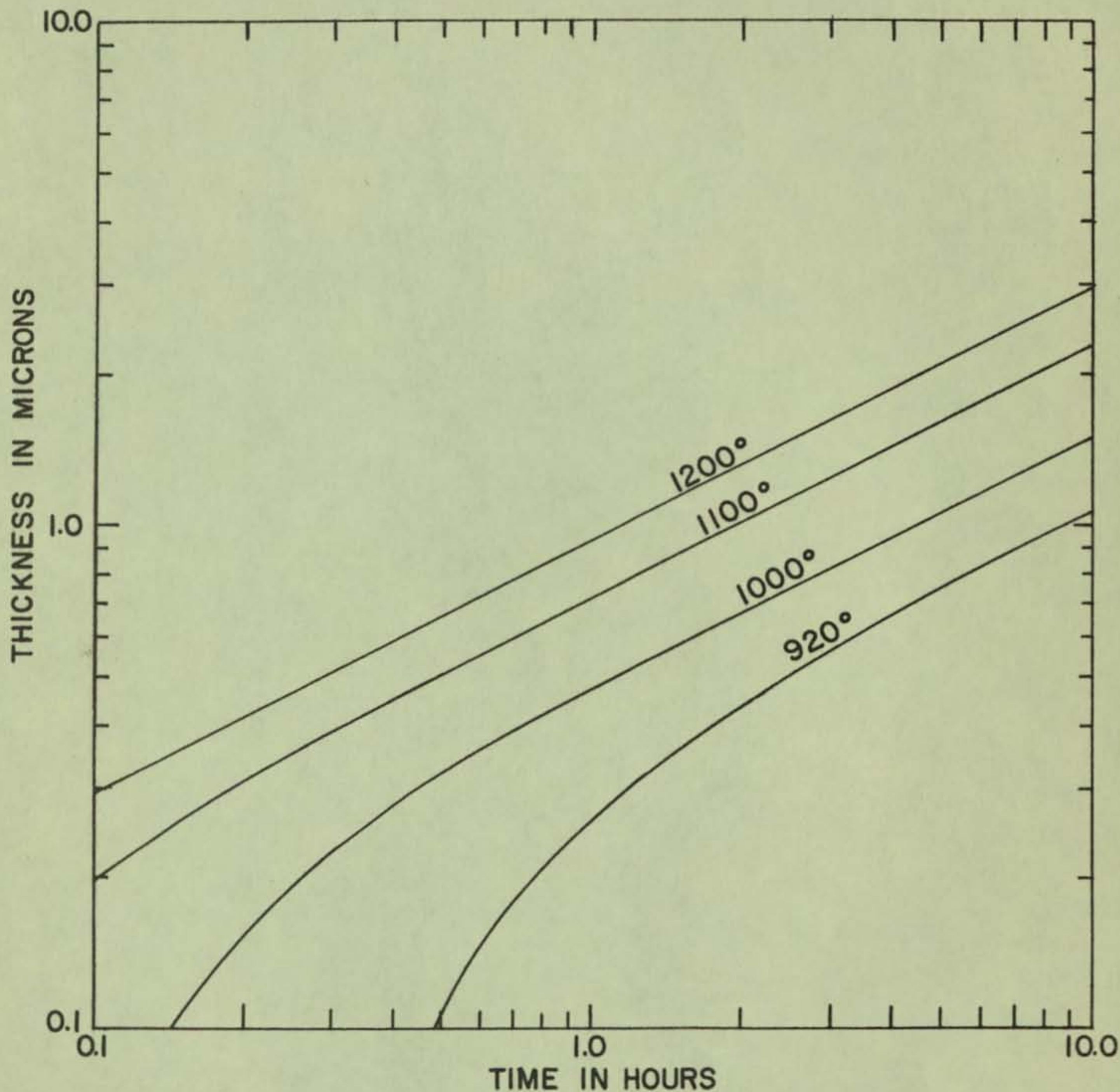


Fig. 5-12. Oxide thickness vs. time for various temperatures in pure water vapor at atmospheric pressure. (Flint, The Electrochemical Society, Electronics Division, Abstracts, Vol. 11, Spring Meeting, p. 222, (May 1962).

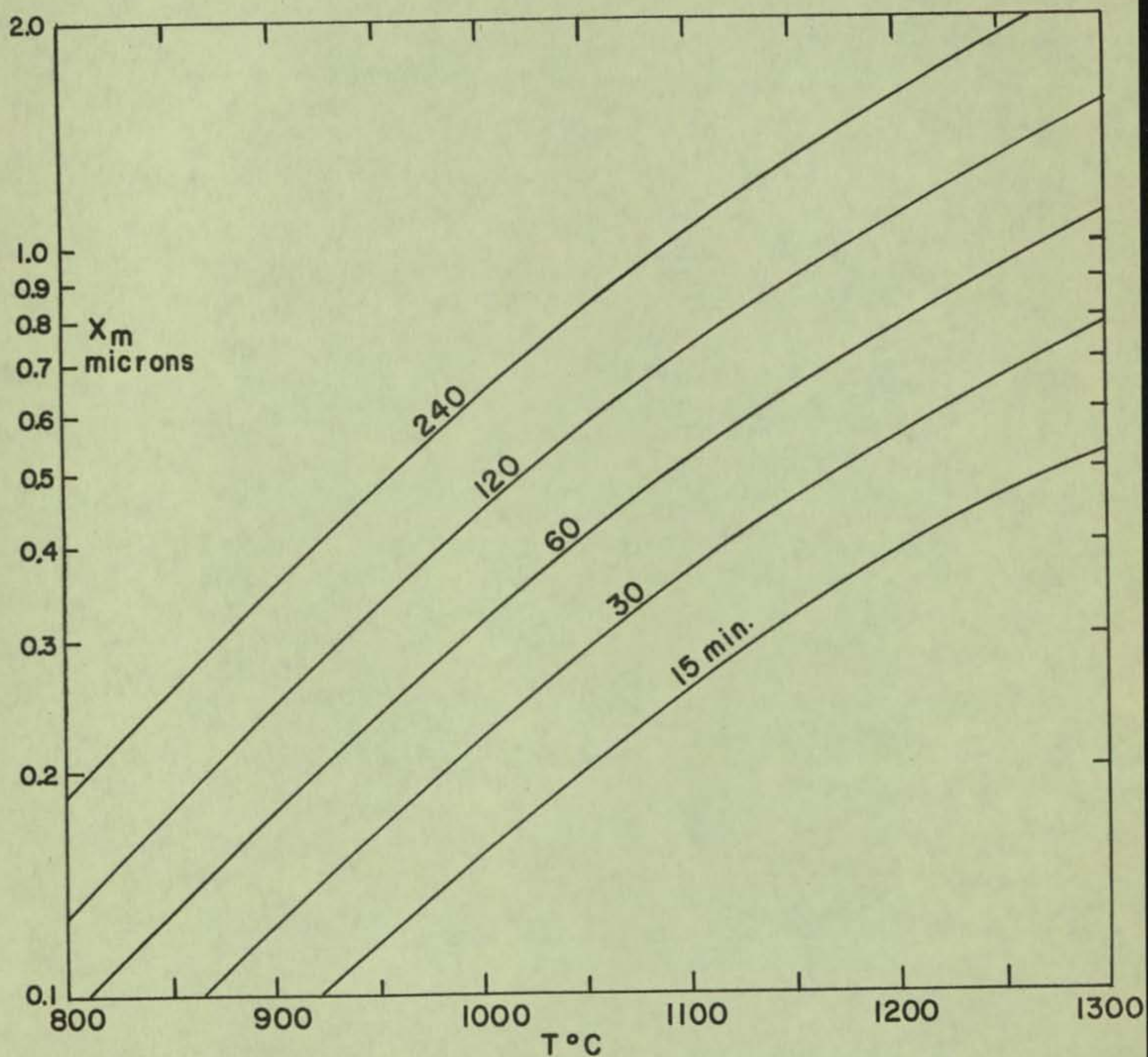


Fig. 5-13. Oxide thickness required to mask against high concentration phosphorus diffusion for various times (From Sah, Sello and Tremere, J. Phys. Chem. Solids 2, p. 228 (1959).

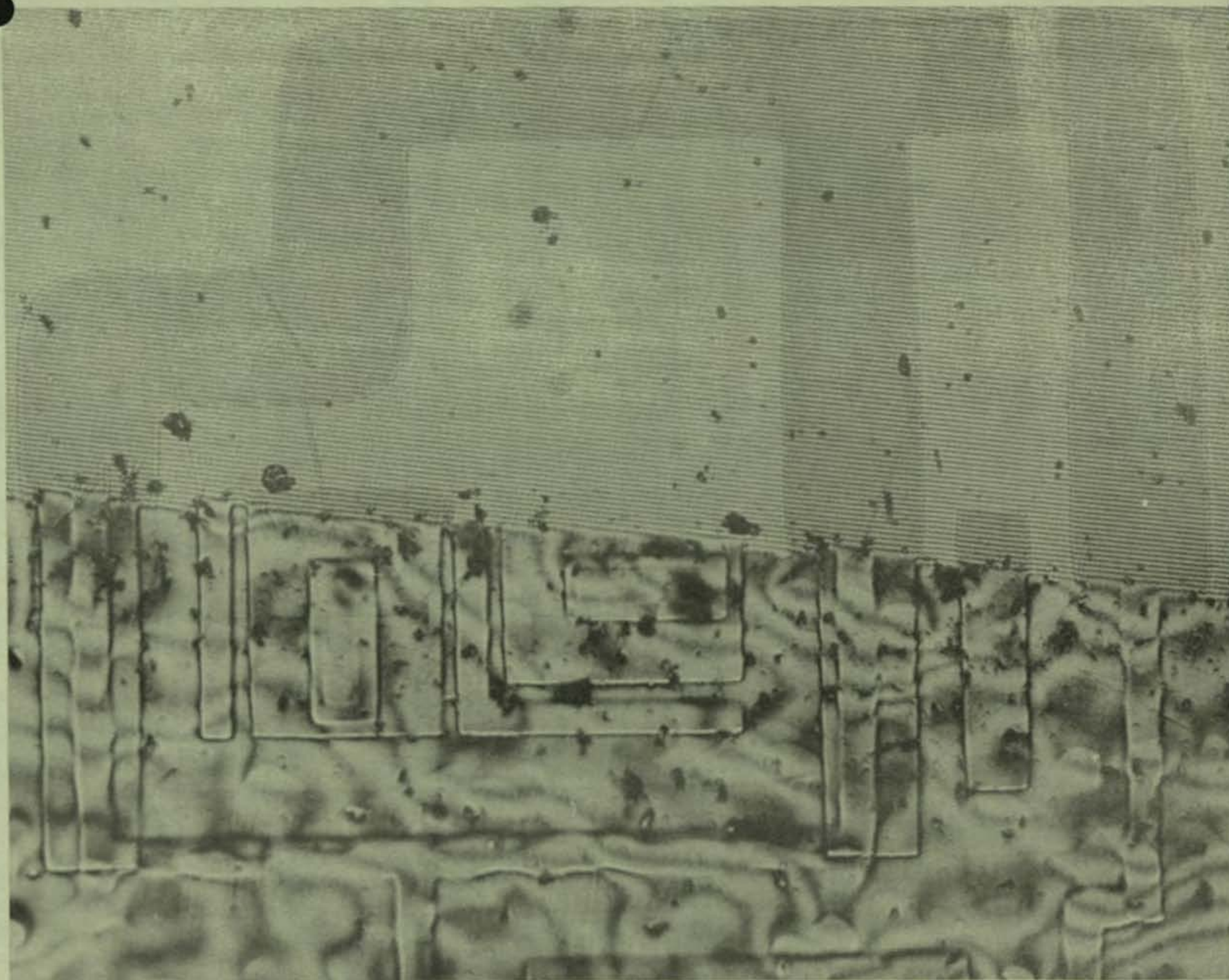


Fig. 5-14. Photomicrograph of a stained device structure taken through a vertically illuminated microscope employing a sodium lamp as a light source with an ordinary microscope slide inserted over the specimen. The originally etched surface shows the outline of the isolation and components. The isolation grid is stained dark on the side of the bevel. The narrow parallel lines along the bevel are interference fringes. Each corresponds to approximately 0.3 microns below the original surface.

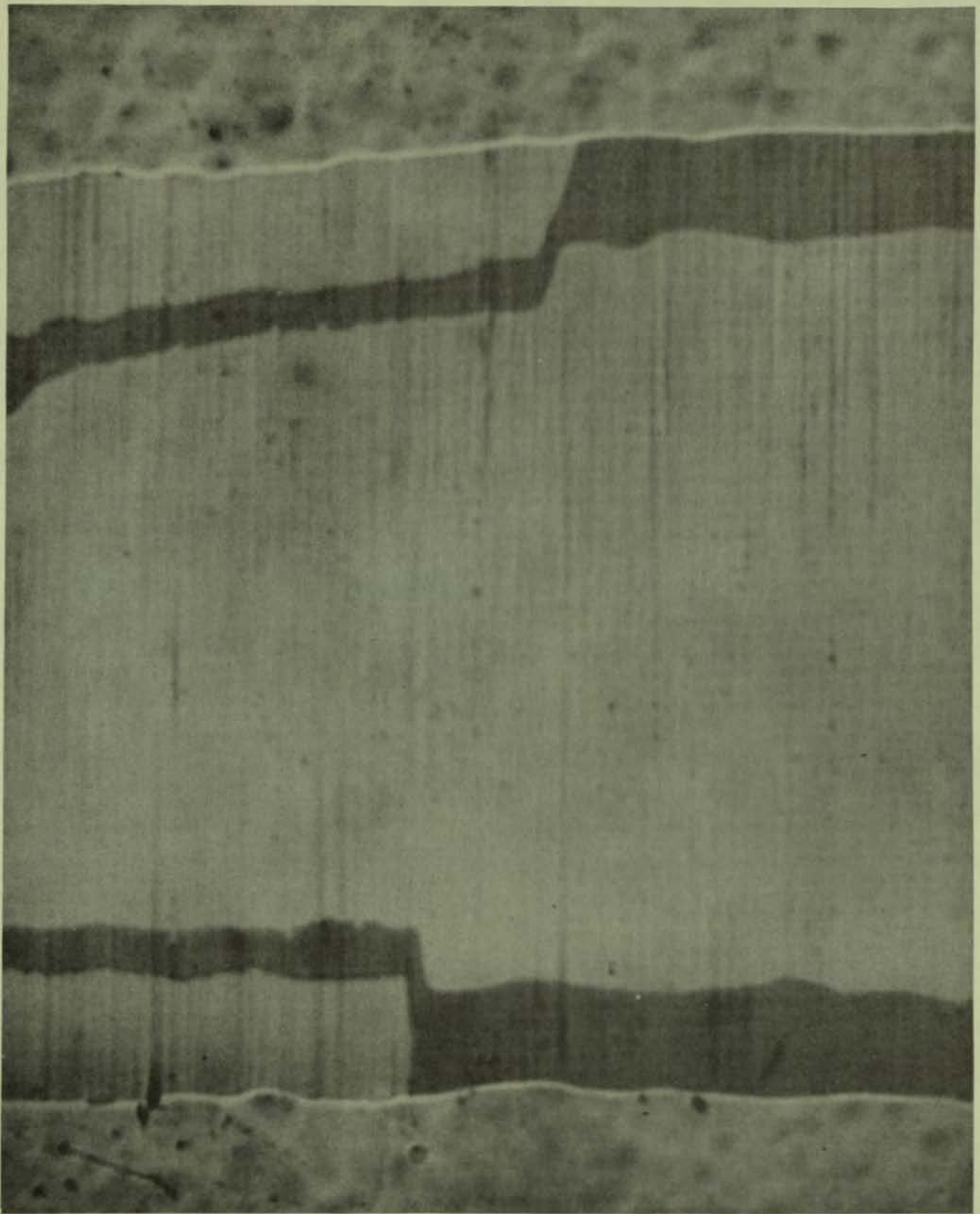


Fig. 5-15. Cross-section through an npn structure with gallium base diffusion. The section was prepared by grinding a cylindrical curve in the surface of the etched wafer. The p-type material is stained dark. The bumpy area is the original wafer surface. Without the heavy phosphorus emitter diffusion, the gallium junction depth would have been uniform across the wafer. In this case where the emitter diffusion is about equally deep to the base diffusion away from the emitter, the enhanced diffusion under the emitter stands out clearly.

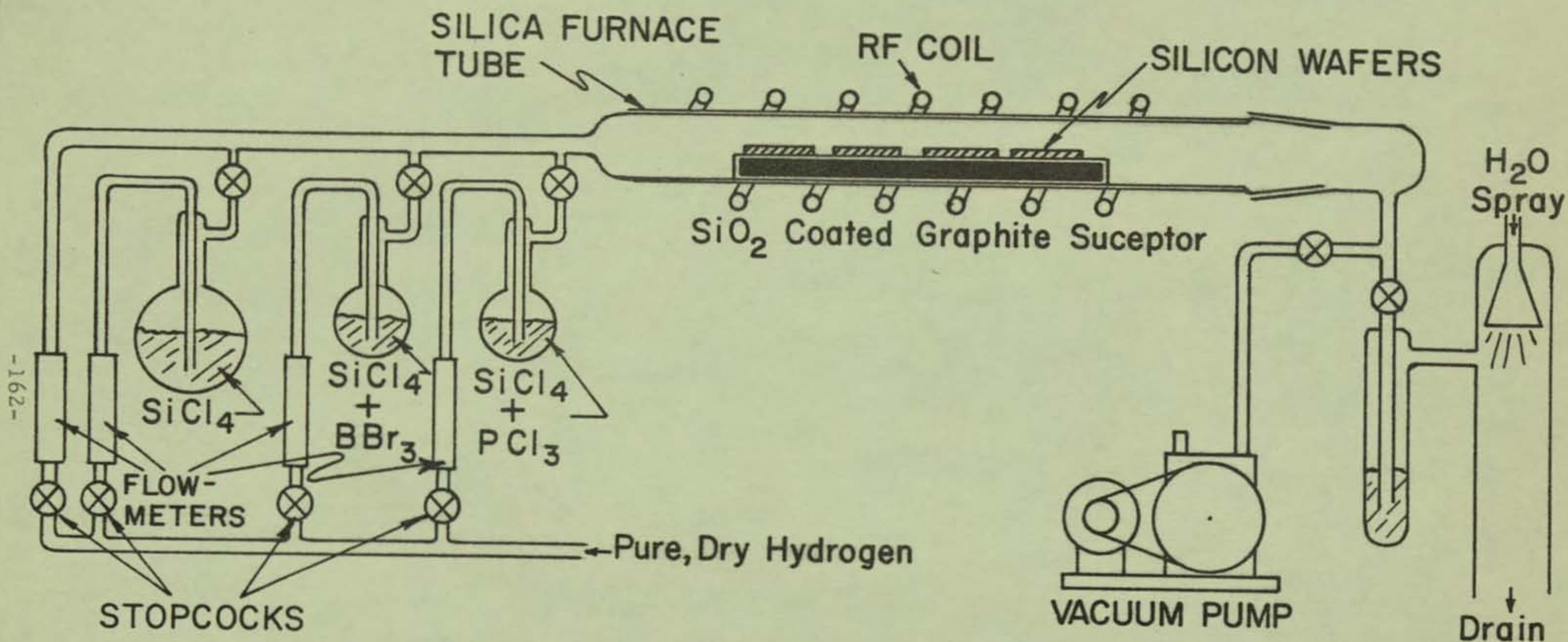


Fig. 5-16. A typical epitaxial system suitable for the preparation of doped silicon layers.

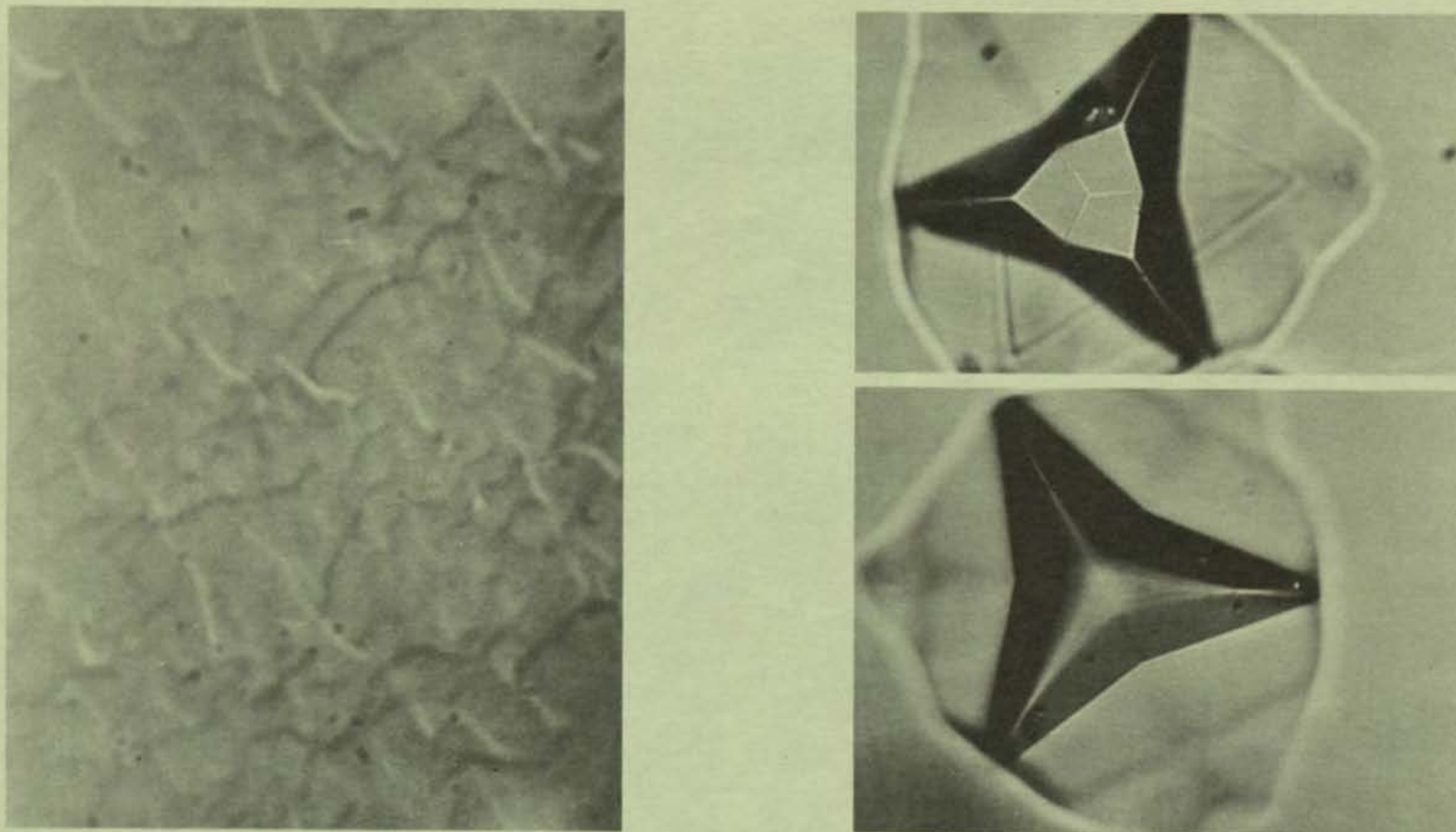


Fig. 5-17. Two typical defects often seen in silicon epitaxial films. "Chevrons" on the left result from too rapid growth. "Pyramids" on the right nucleate at the substrate-film interface.

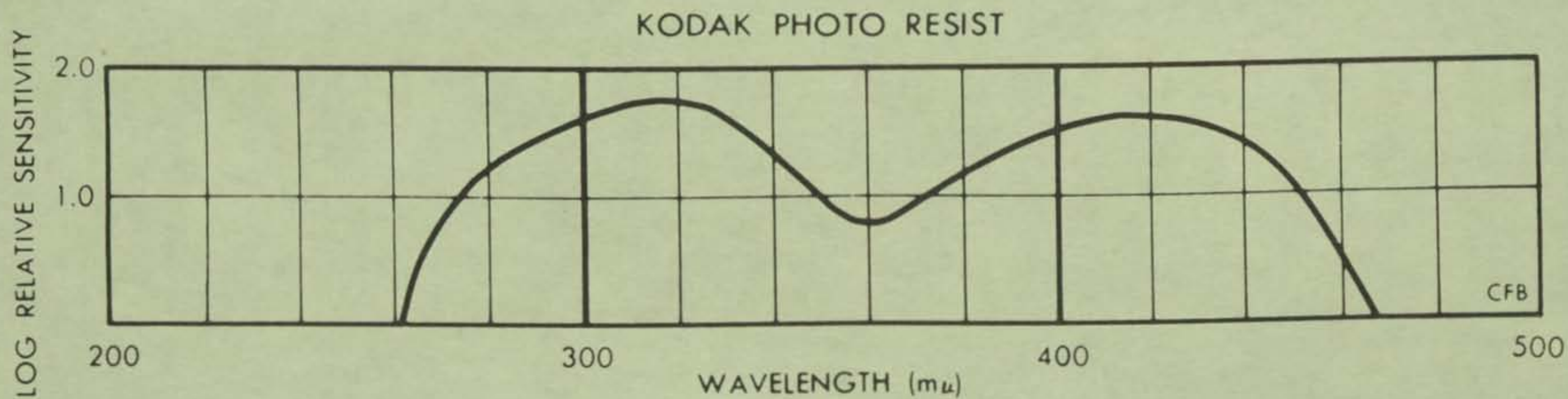
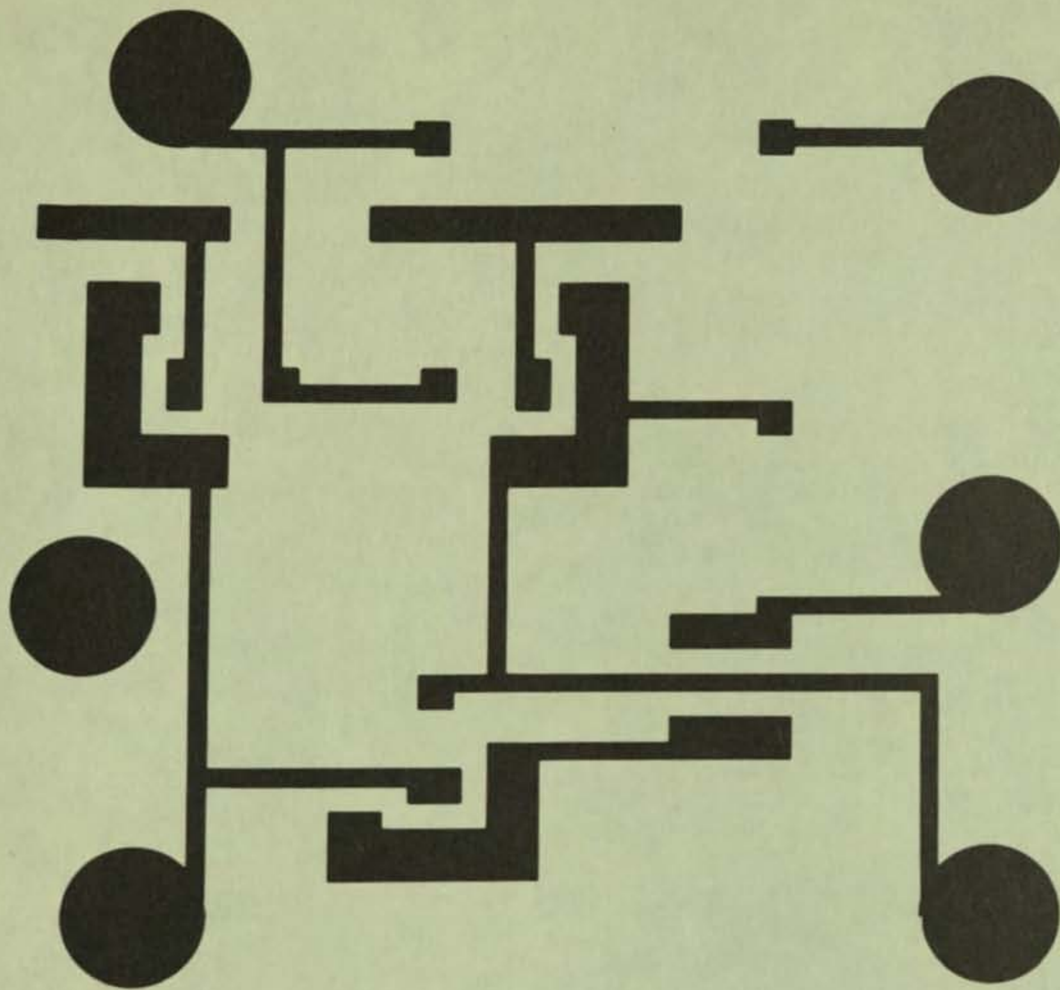


Fig. 5-18. Spectral sensitivity of KPR (From Kodak *Photo Sensitive Resists for Industry*, Eastman Kodak Company, Rochester, N. Y., (1962).



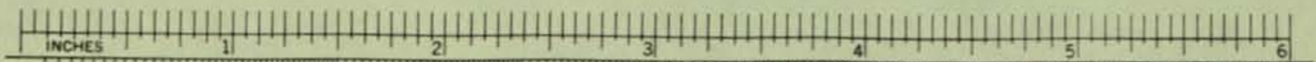


Fig. 5-19. An example of the original art work and final working mask for a typical pattern employed in microcircuitry.

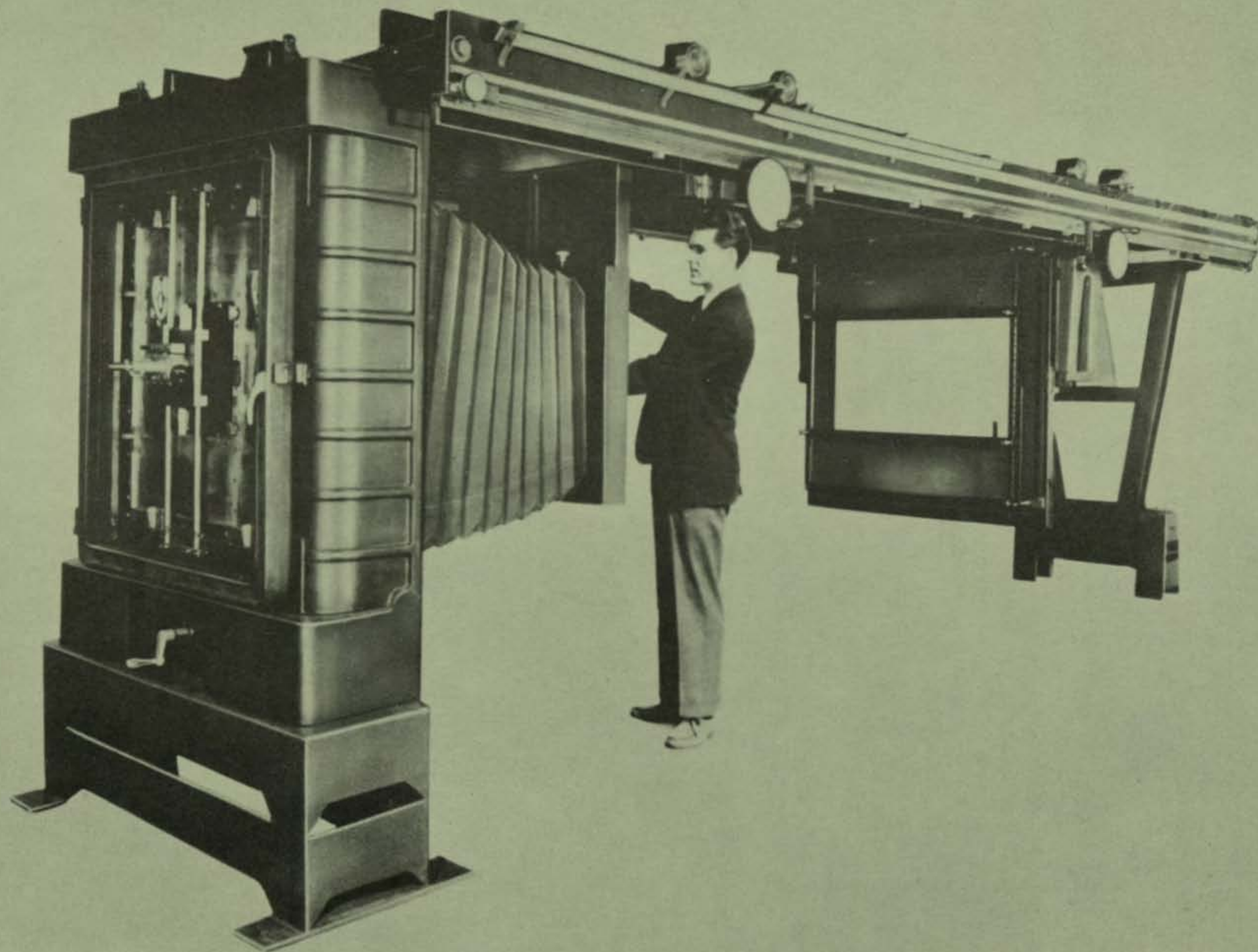


Fig. 5-20. A copy camera of the type useful in making high precision photographic masks.

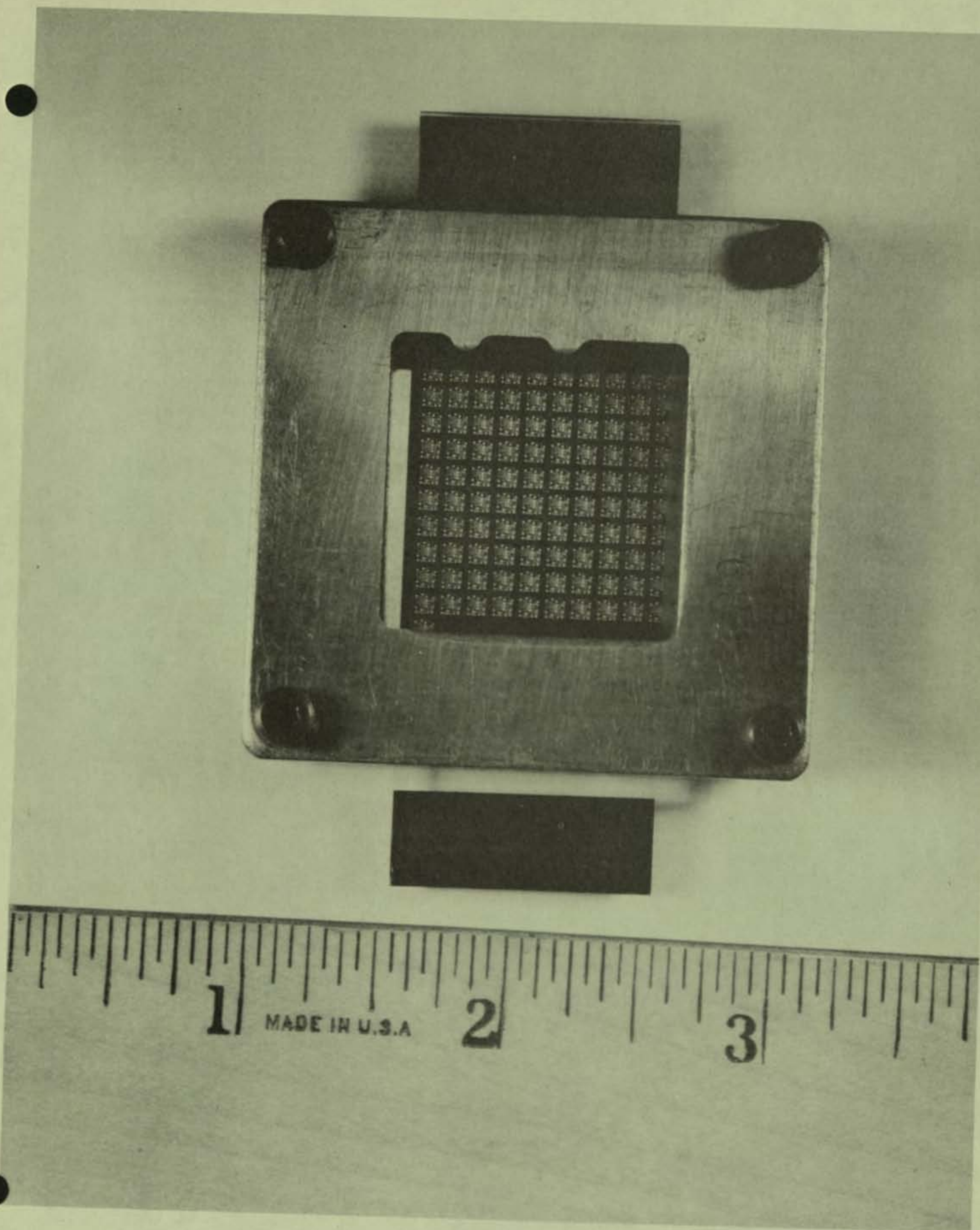


Fig. 5-21. A photographic mask in an indexed frame. The silicon wafer makes contact with the mask frame at the two semicircular bumps and at one flat edge.

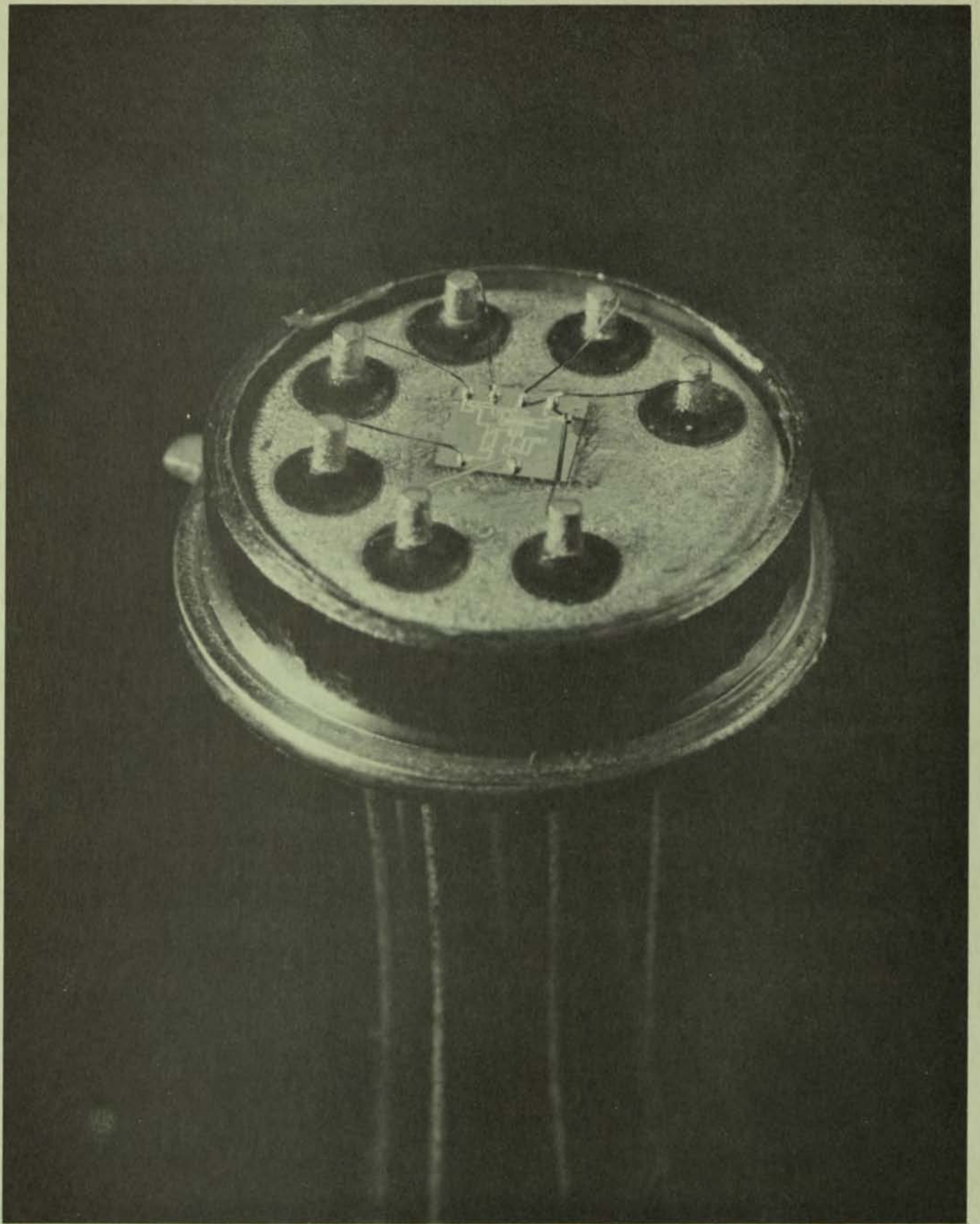


Fig. 5-22. Microcircuit mounted upon an eight lead transistor type header. The eighth lead is spot-welded to the bottom of the metal eyelet. The top of the can has been cut away.

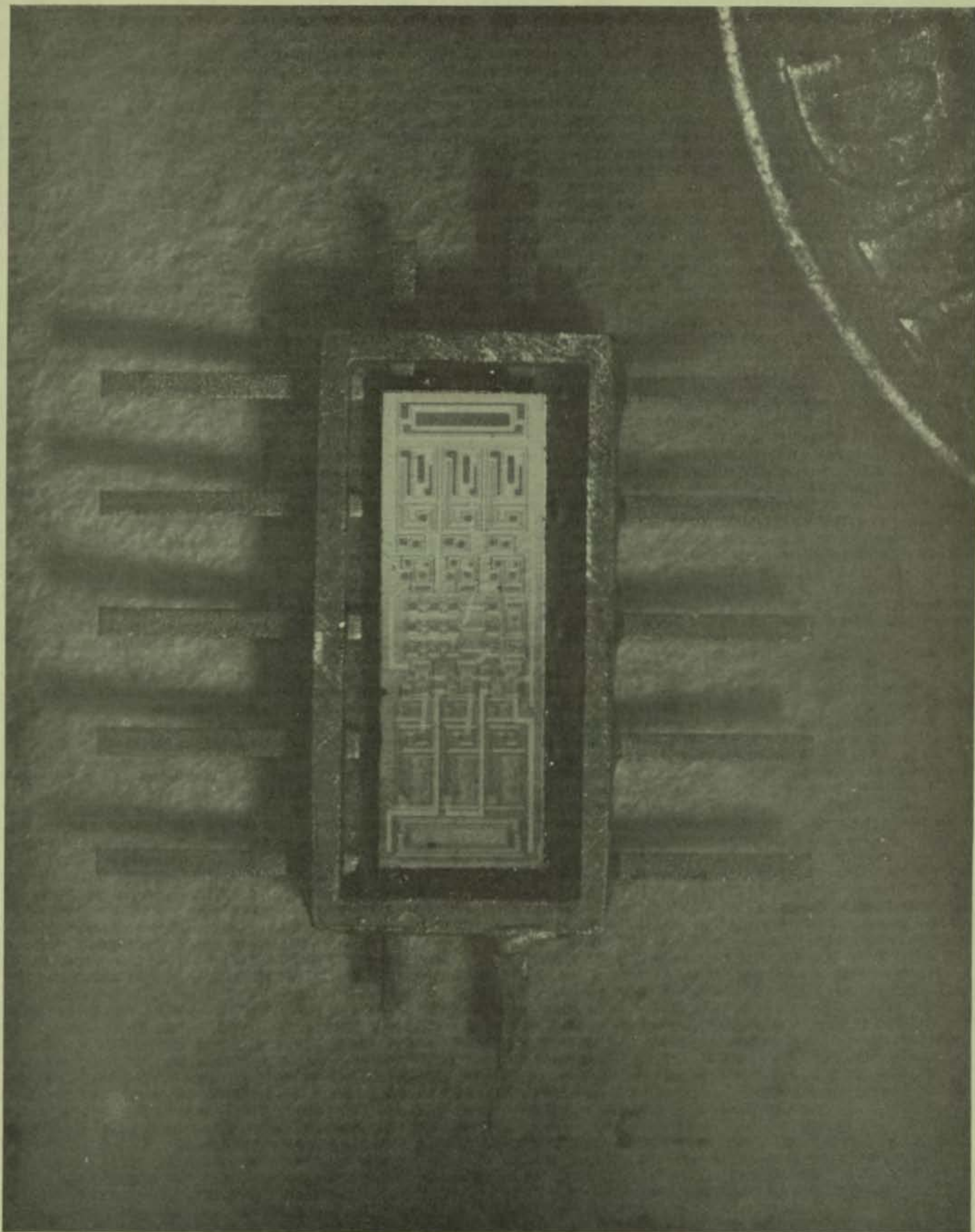


Fig. 5-23. A semiconductor integrated circuit in a flat package before lead bonding and sealing. This package is 1/8" wide by 1/4" in length. (Courtesy of Texas Instruments, Inc.)

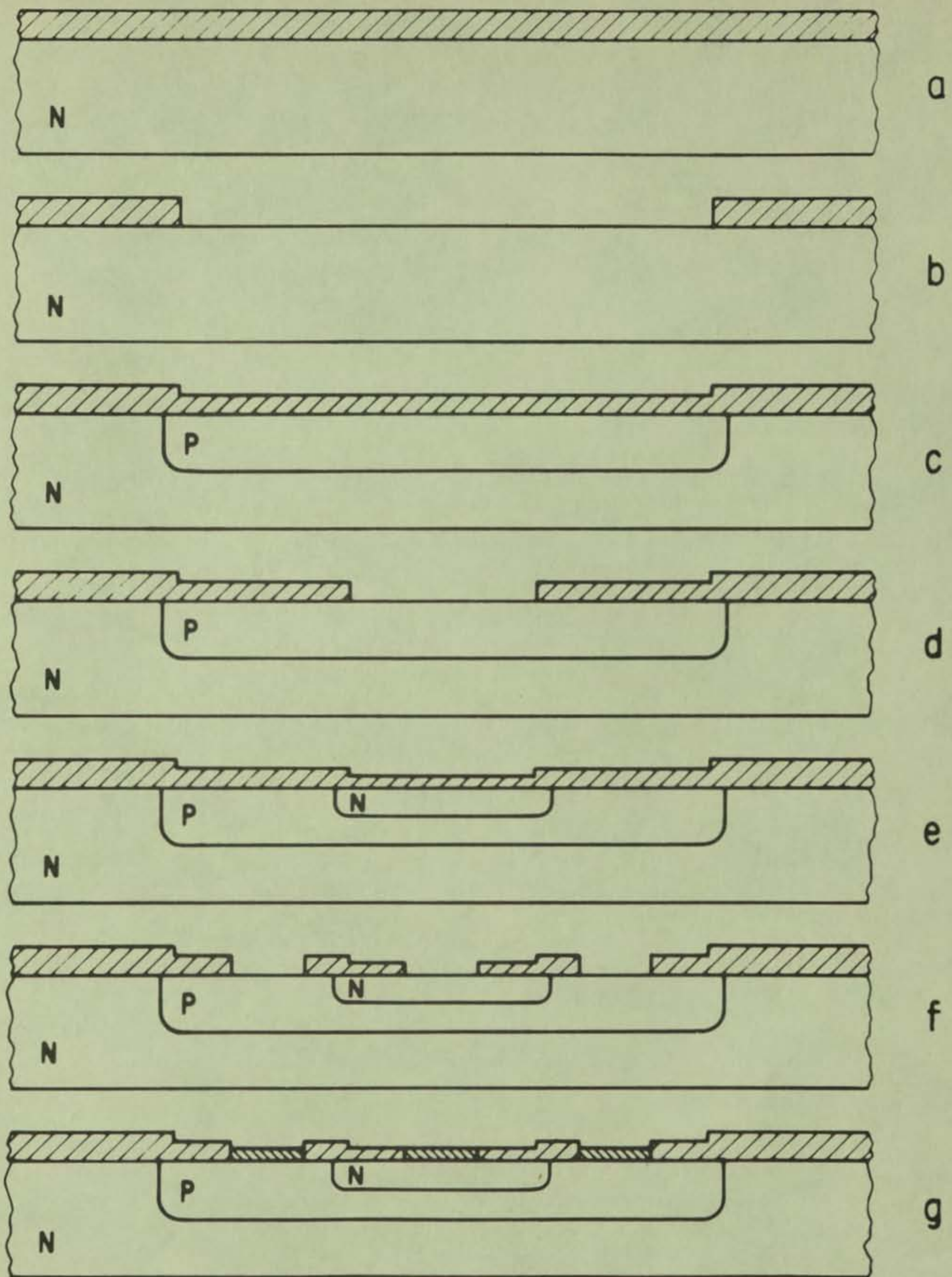


Fig. 5-24. Schematic cross-sections of a single planar transistor structure after various processing steps. (a) after oxidation; (b) after base oxide removal; (c) after based diffusion; (d) after emitter oxide removal; (e) after emitter diffusion; (f) after contact oxide removal; (g) after contact metallization.

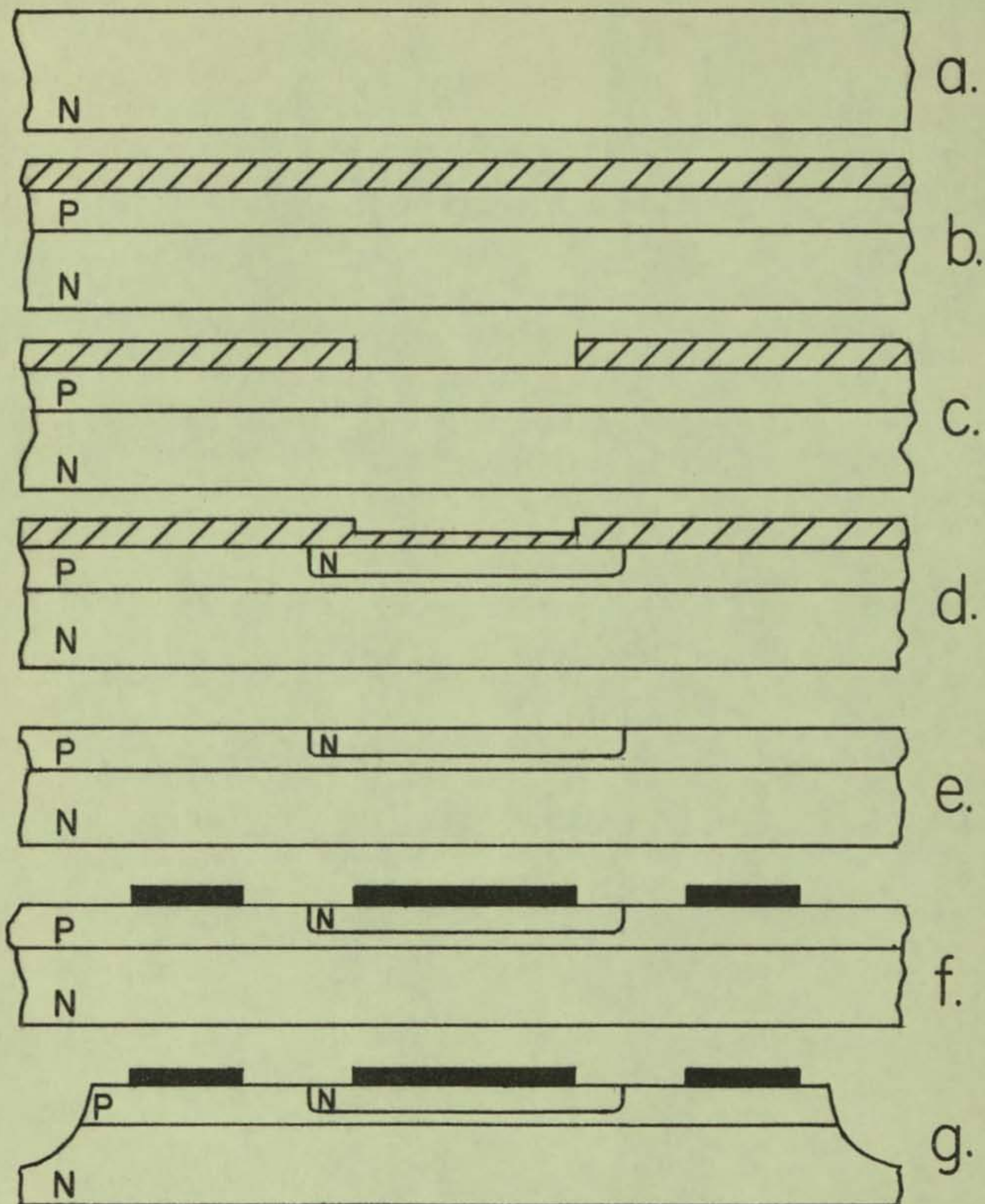


Fig. 5-25. Schematic cross-section of a single double diffused mesa transistor after various processing steps. (a) original wafer; (b) based diffusion; (c) emitter oxide removal; (d) emitter diffusion; (e) oxide stripping; (f) contact metallization; (g) mesa formation.

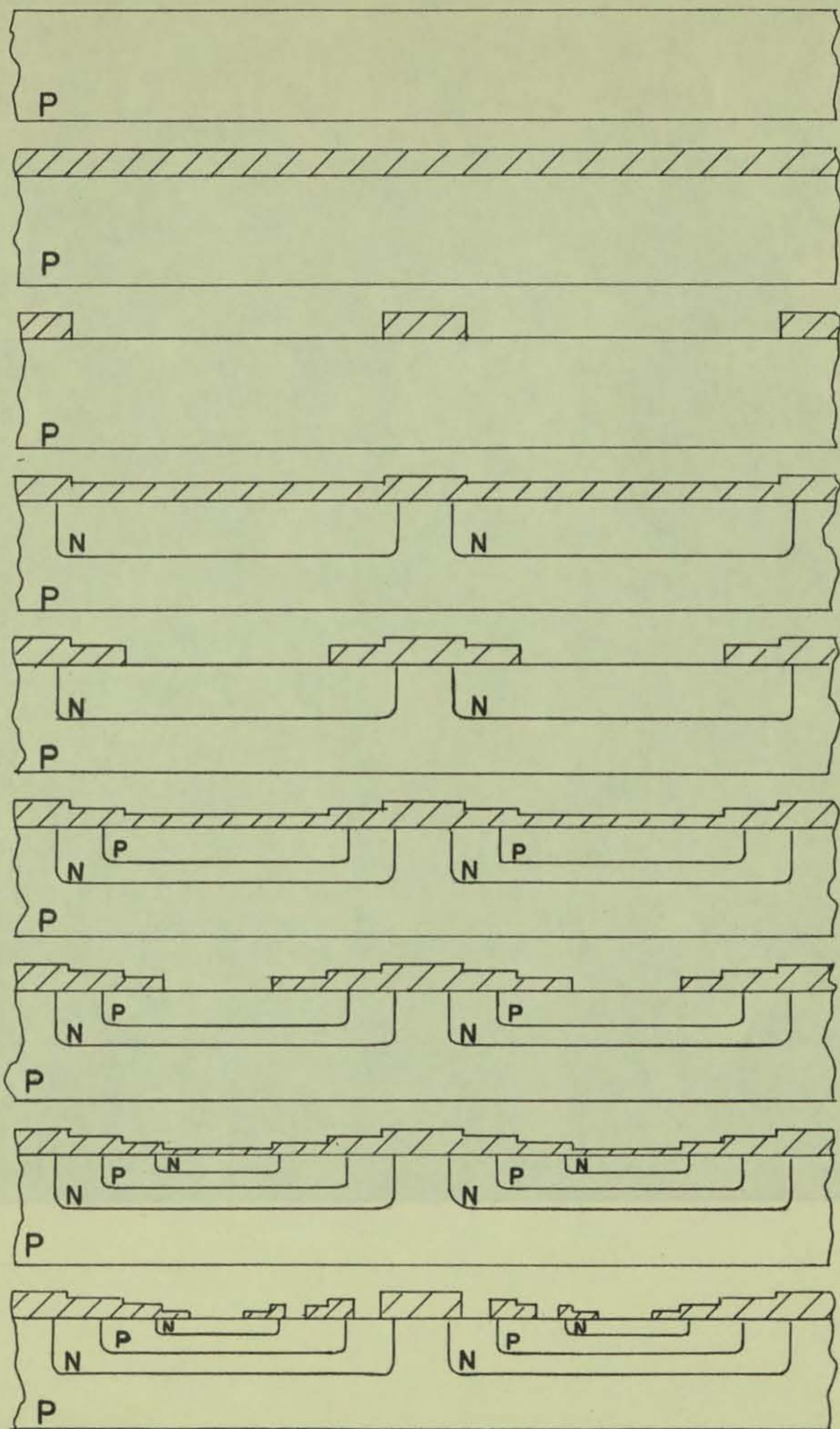


Fig. 5-26. Schematic cross-section showing an example of how two transistor structures isolated by reverse bias junctions can be achieved in a single block of silicon by the use of three diffusion steps. The crosshatched area is silicon dioxide.

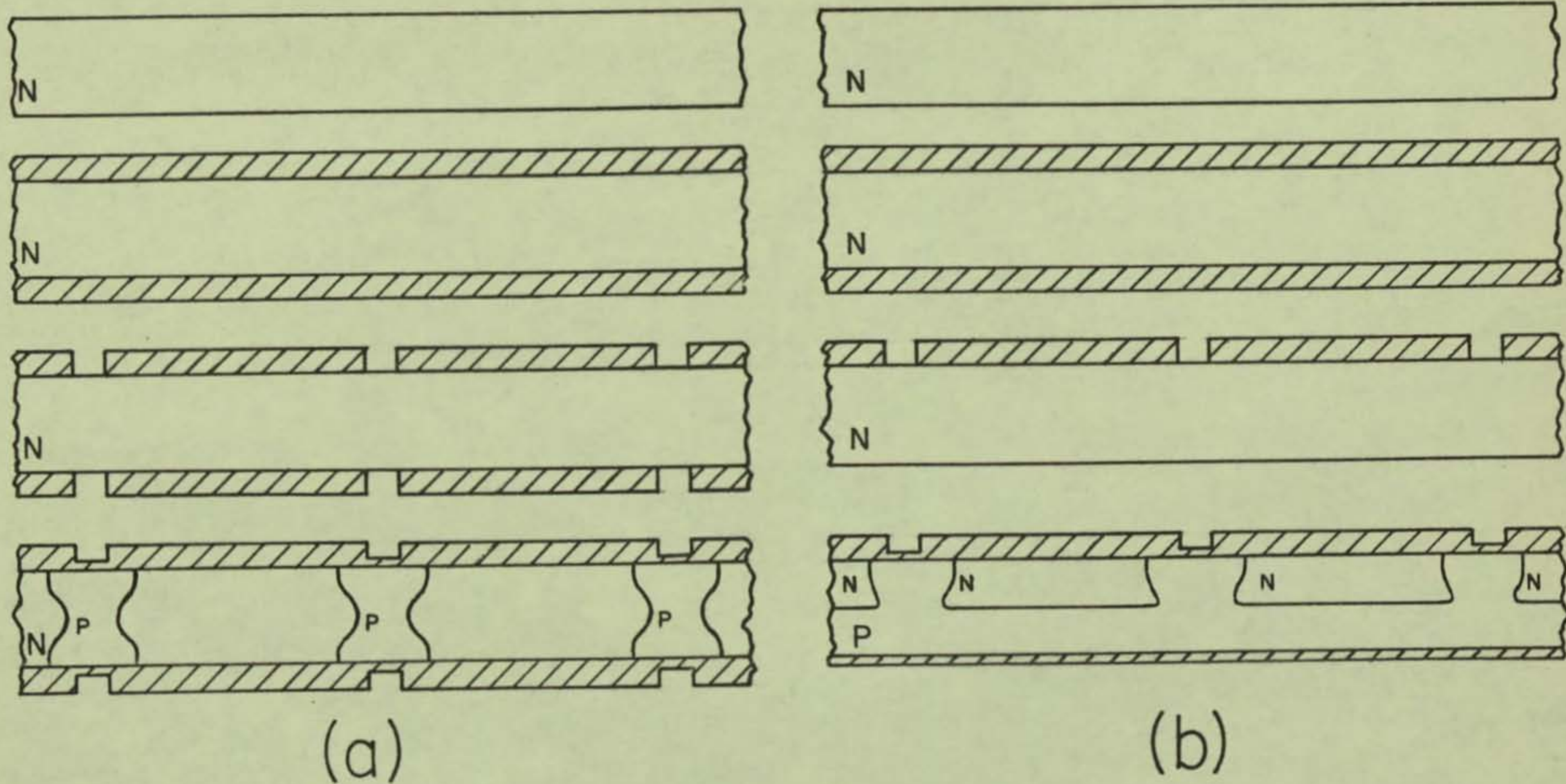


Fig. 5-27. Alternative methods of achieving isolation. The left side shows a scheme involving oxide masking on both sides of the wafer for preparation of a grid through the wafer. In the right portion, the oxide masking is employed only on the top surface, while diffusion takes place from the entire back surface. These isolated regions can then be used to produce the desired device structures.

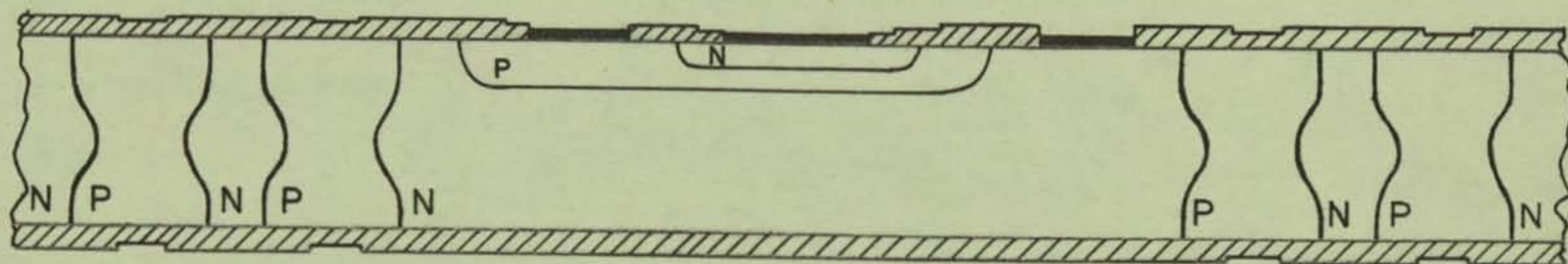


Fig. 5-28. Multiple grid isolation suitable for the reduction of isolation leakage and capacitance.

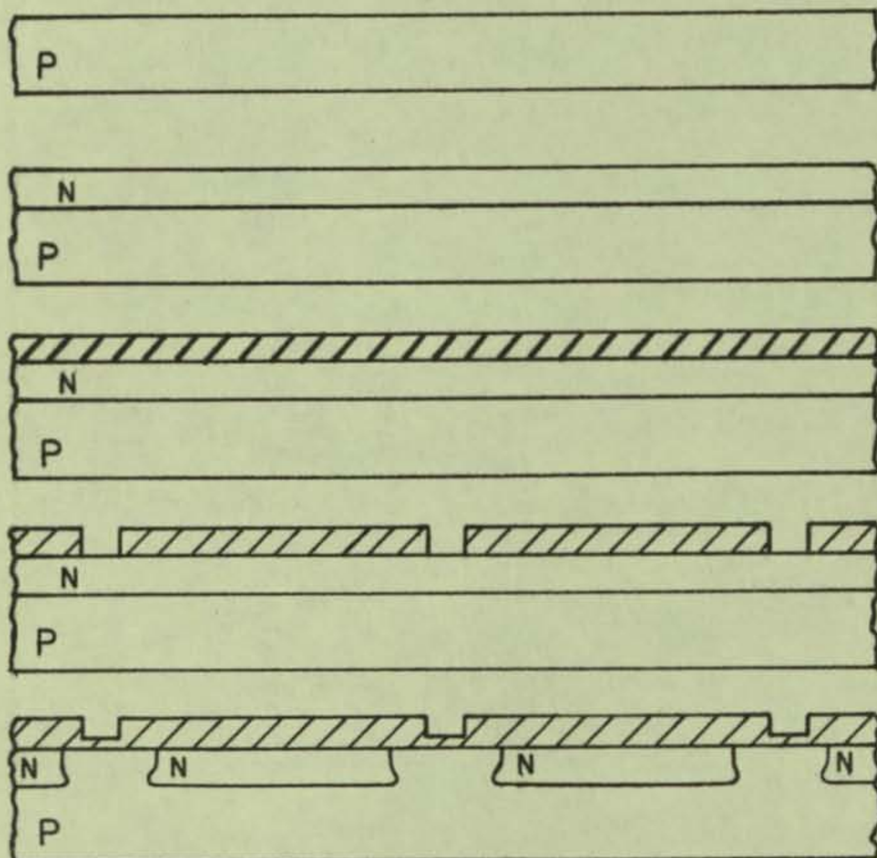


Fig. 5-29. Isolation structure employing an epitaxially grown layer. The five steps top to bottom show schematic cross-sections after substrate preparation, epitaxial film growth, oxidation, oxide etching and isolation diffusion, respectively.

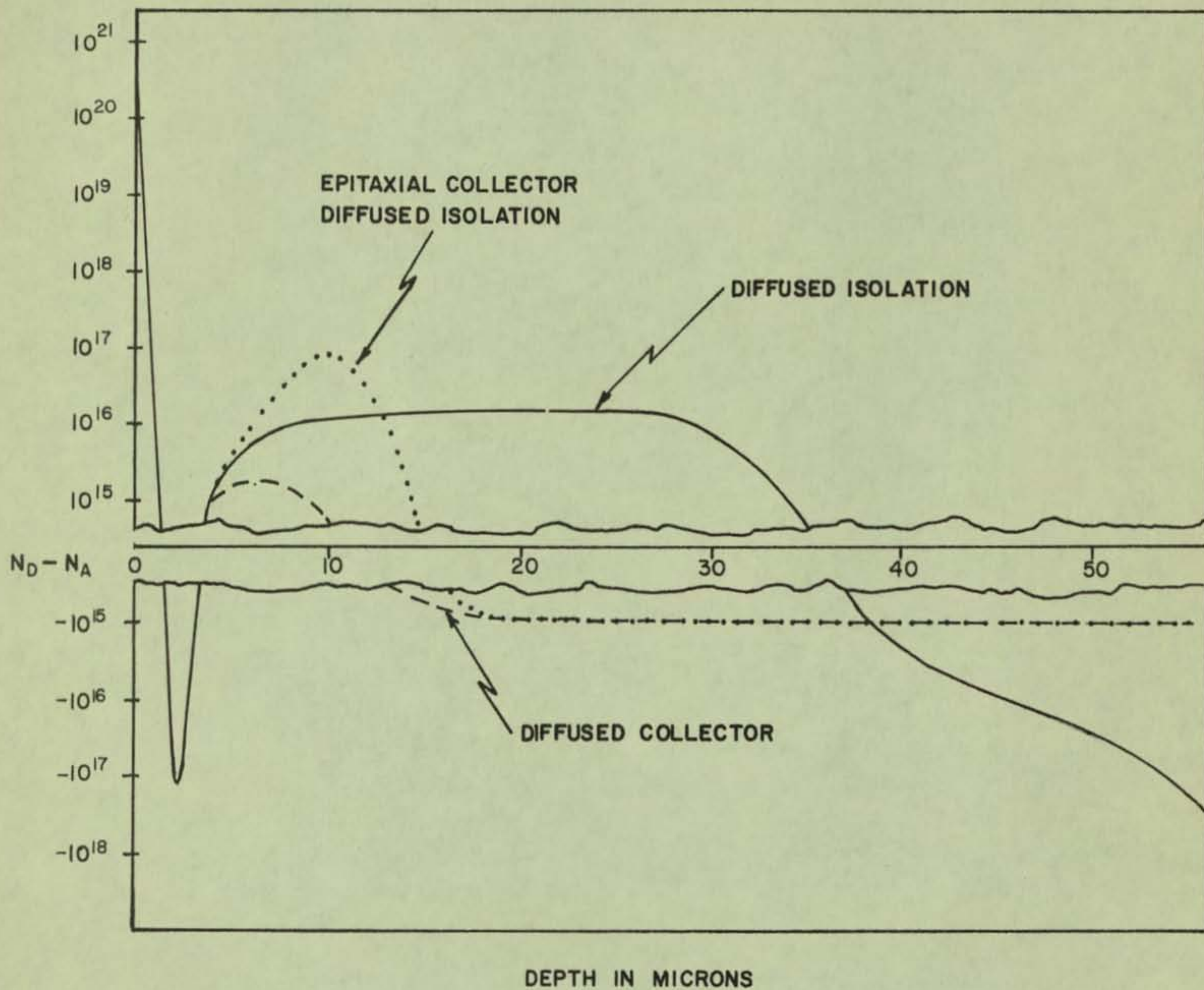


Fig. 5-30. Comparison of typical impurity profiles for three different isolation techniques.

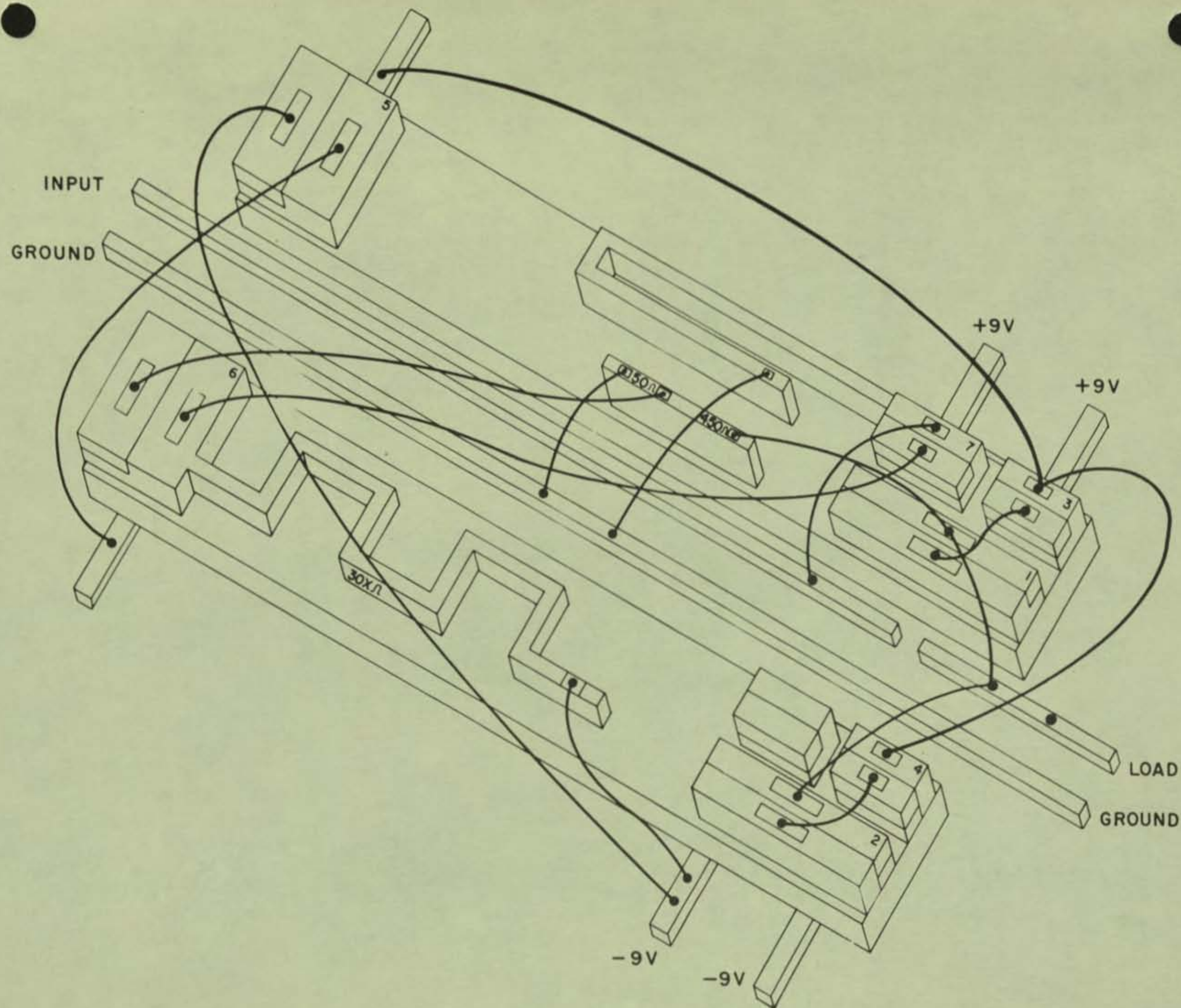


Fig. 5-31. Diagram of an early integrated circuit employing thermo compression bonded wire for intraconnection.

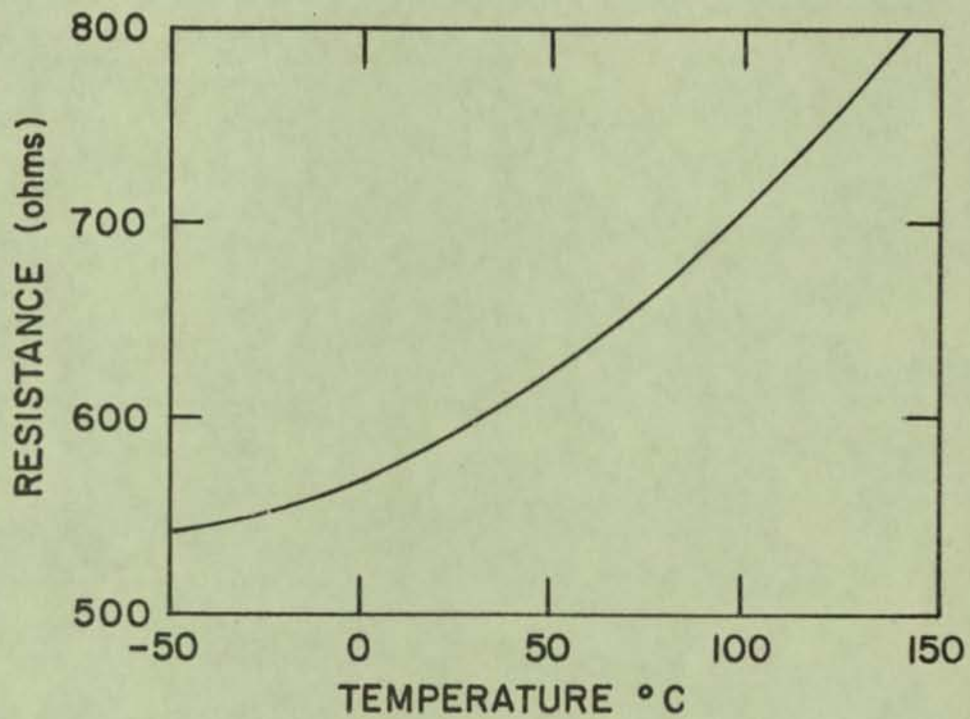


Fig. 5-32. Resistance versus temperature for a typical boron diffused resistor.

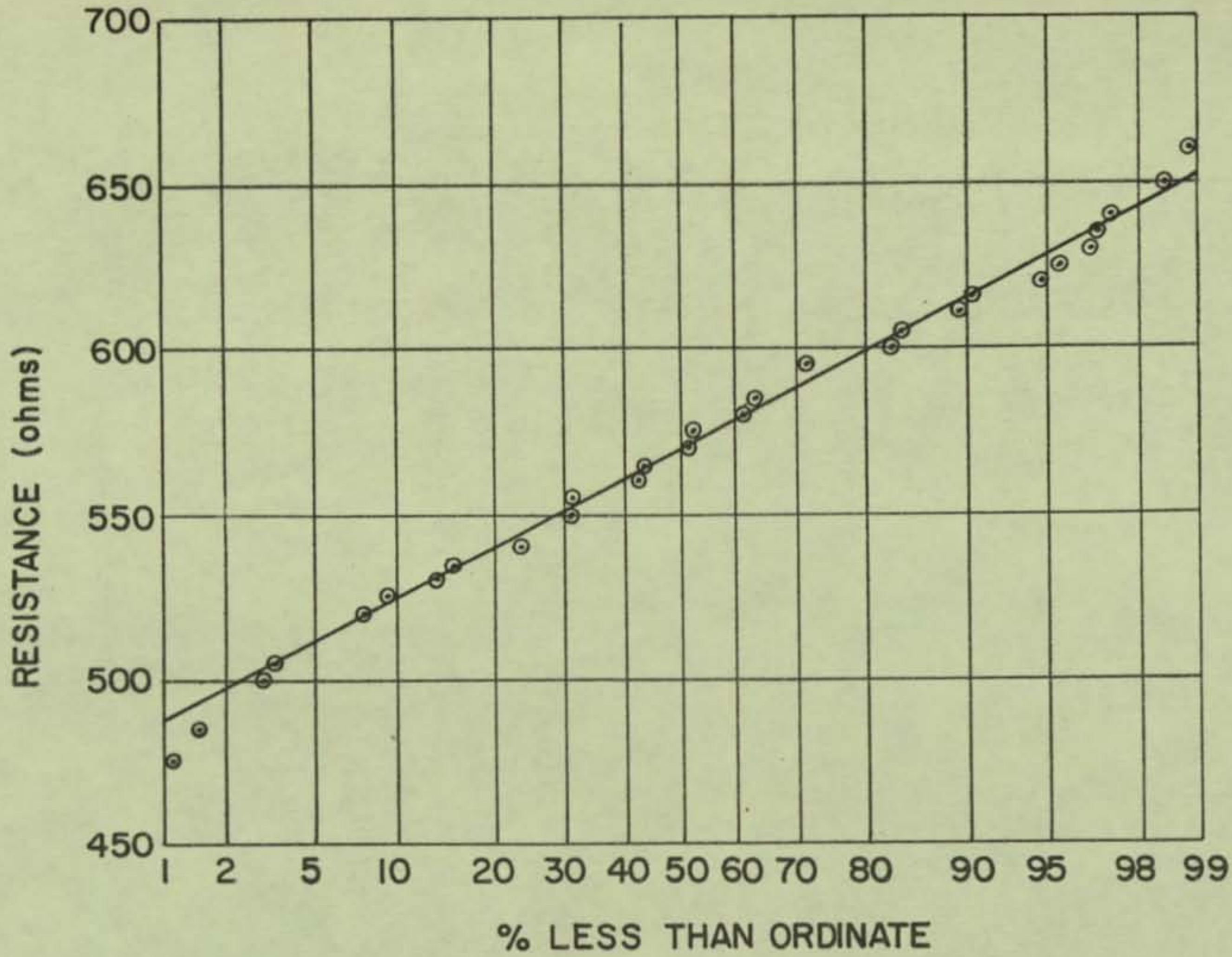


Fig. 5-33. Distribution curve of a diffused resistor representing an extended period of time in production. The nominal design value was 580 ohms.

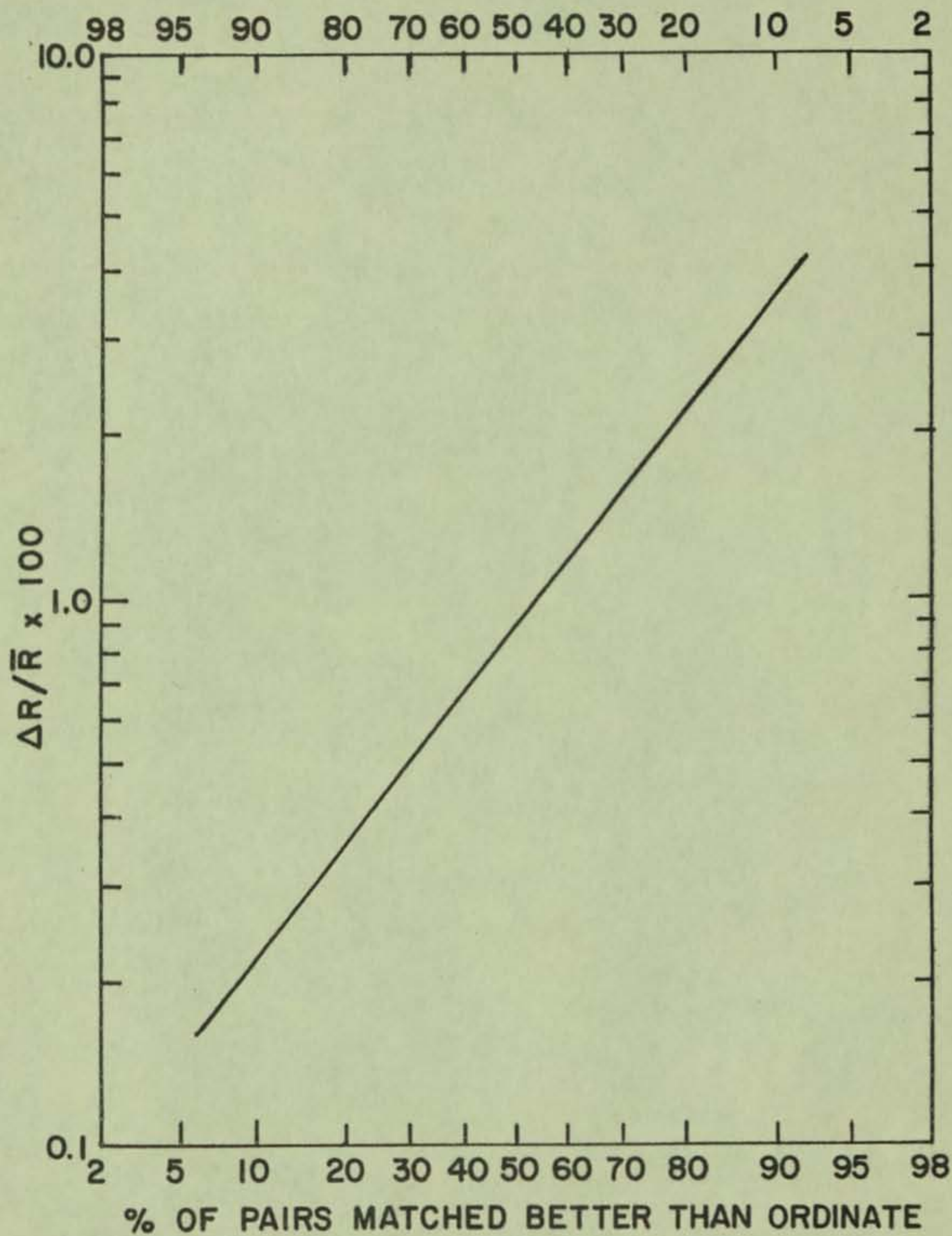


Fig. 5-34. Distribution of the ratio of pairs of adjacent diffused resistors.

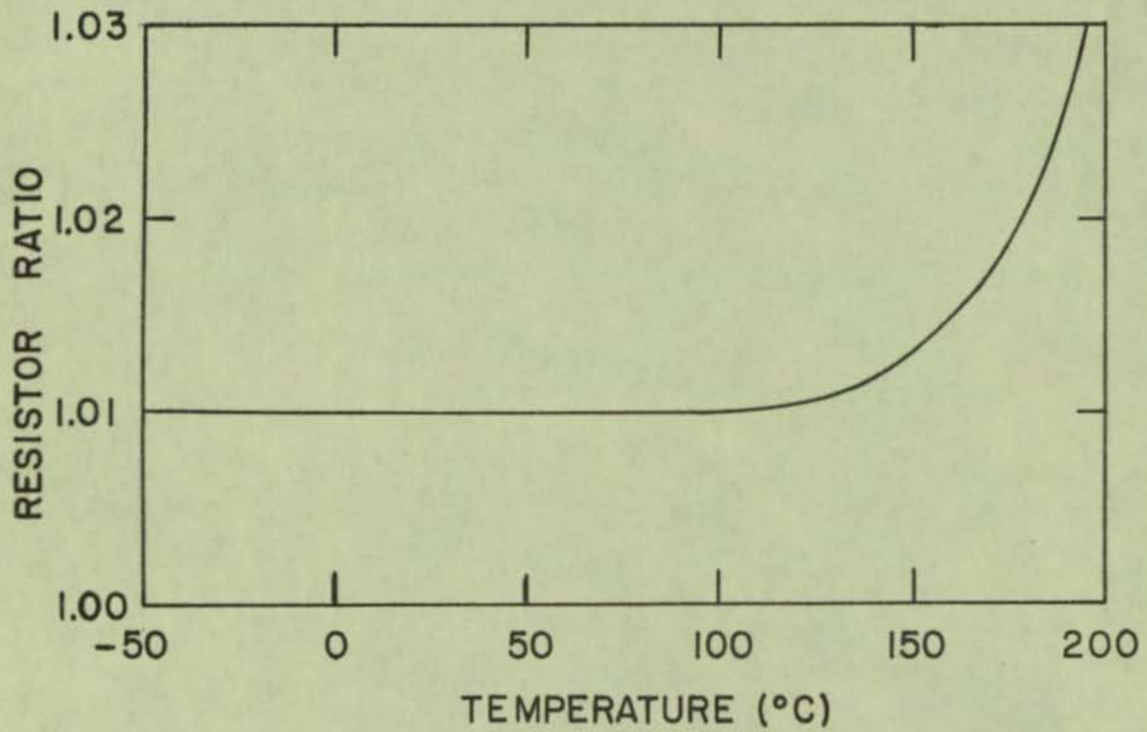


Fig. 5-35. Ratio of resistance vs. temperature for a typical pair of diffused resistors.

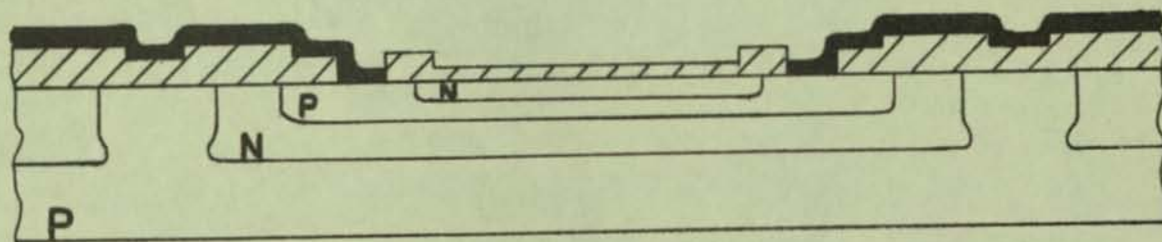


Fig. 5-36. Schematic cross section through a resistor structure making use of the equivalent of the transistor base spreading resistance.

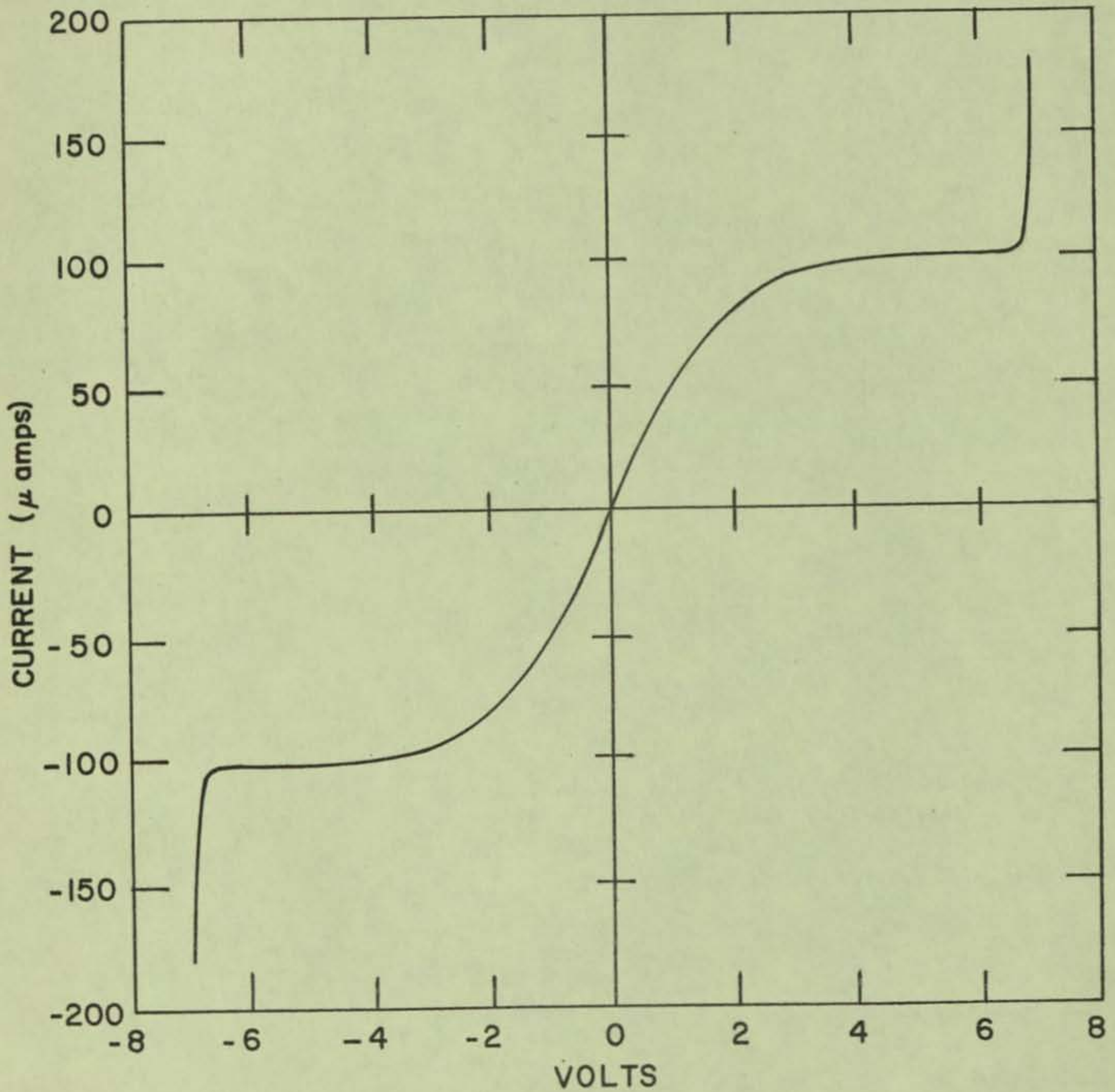


Fig. 5-37. V-I characteristics of a resistor, such as shown in Fig. 5-36. The resistance at V = 0 is approximately 10 kilohms.

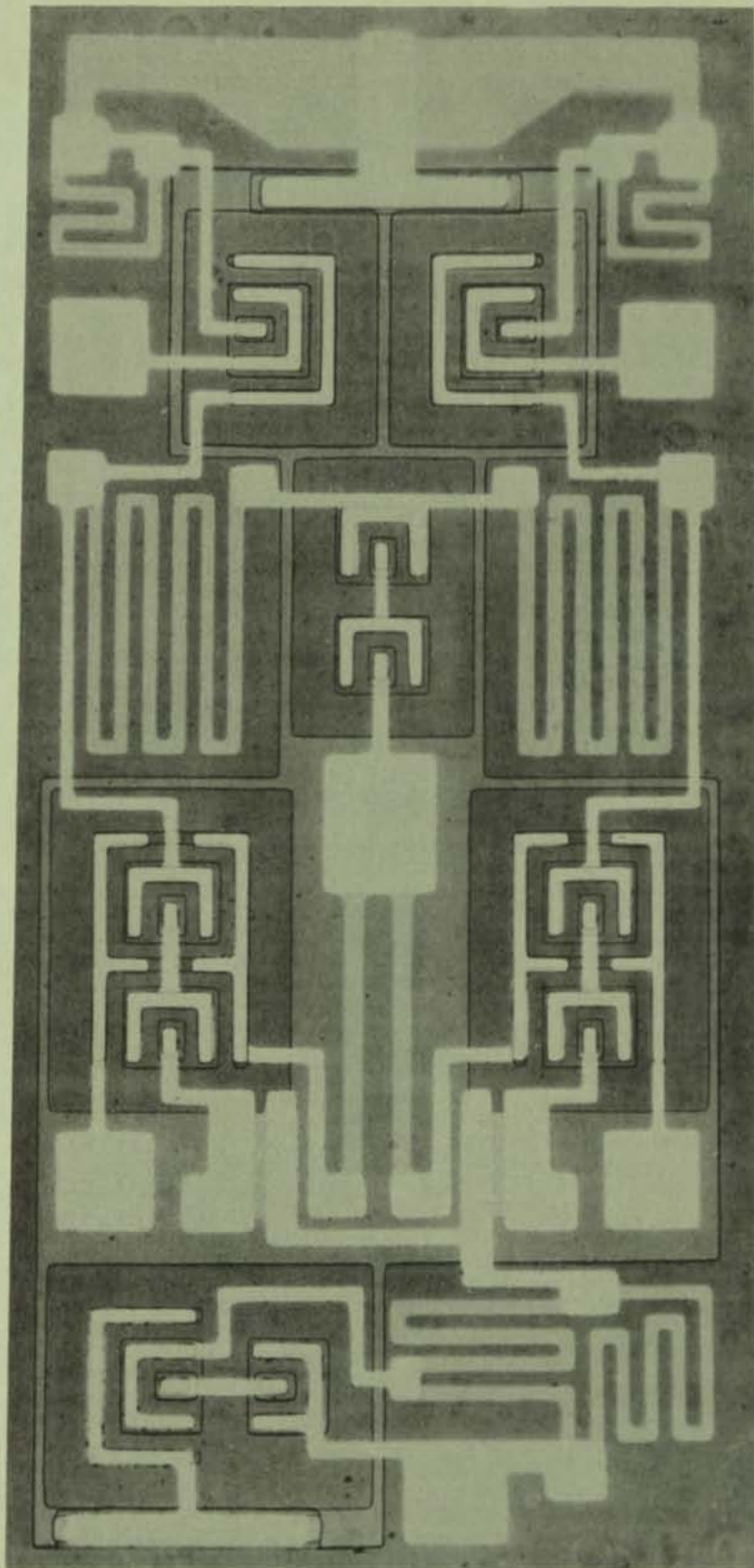


Fig. 5-38. Photomicrograph of an integrated differential amplifier circuit employing thin film resistors. The white metallization is aluminum interconnections while the gray stripes are Nichrome films.

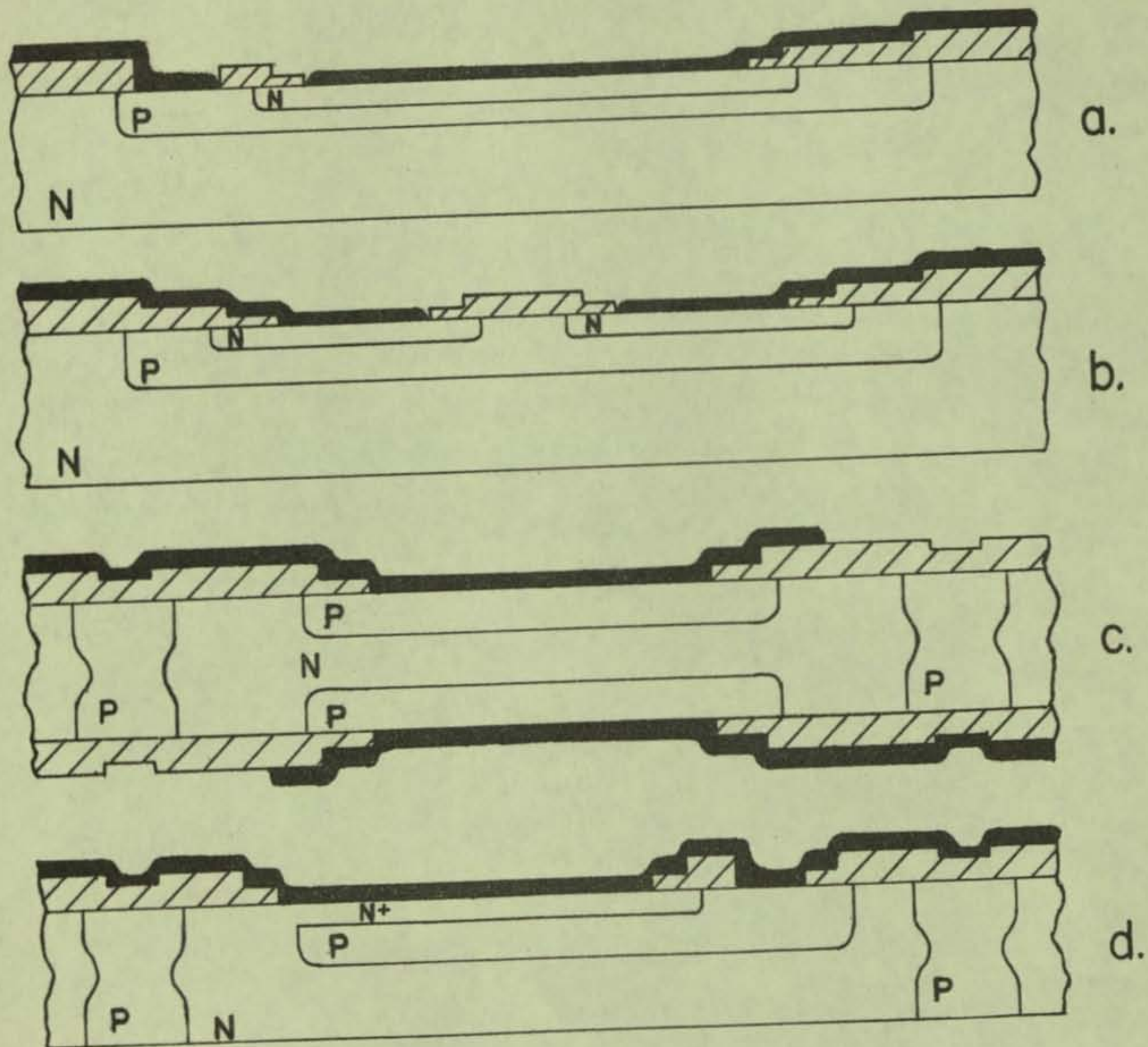


Fig. 5-39. Junction capacitor structures. (a) double diffused; (b) double diffused back-to-back; (c) back-to-back bilateral; (d) combined emitter-base and collector-base junctions.

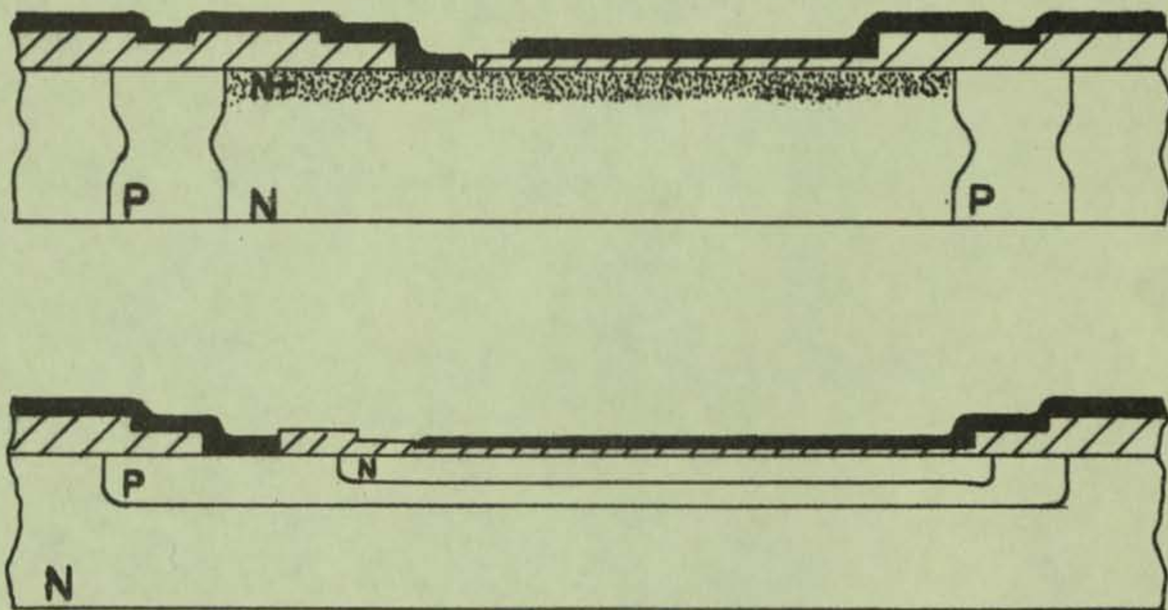


Fig. 5-40. Metal-oxide-semiconductor capacitor structures.

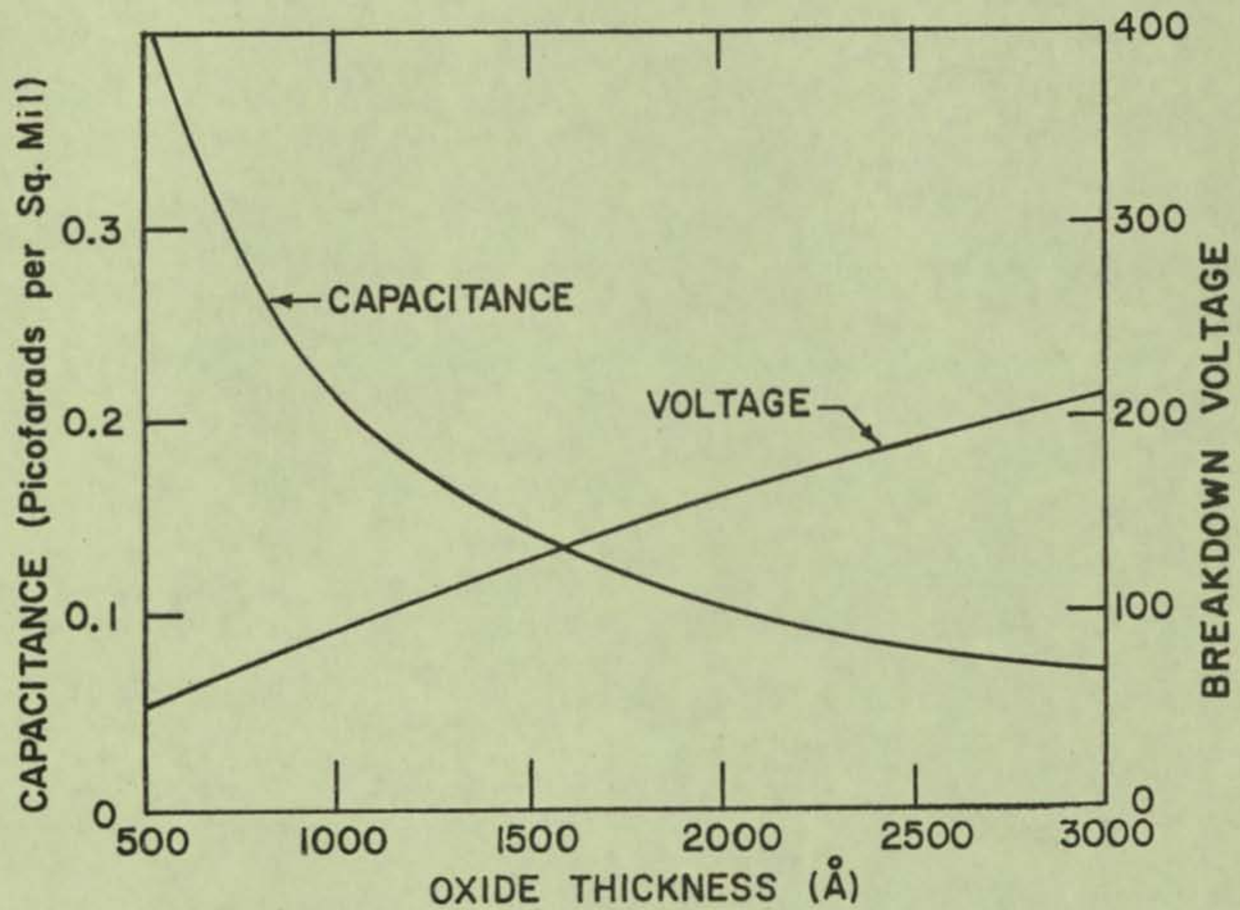


Fig. 5-41. Capacitance and breakdown voltage versus oxide thickness for metal-SiO_2-silicon capacitor structures.

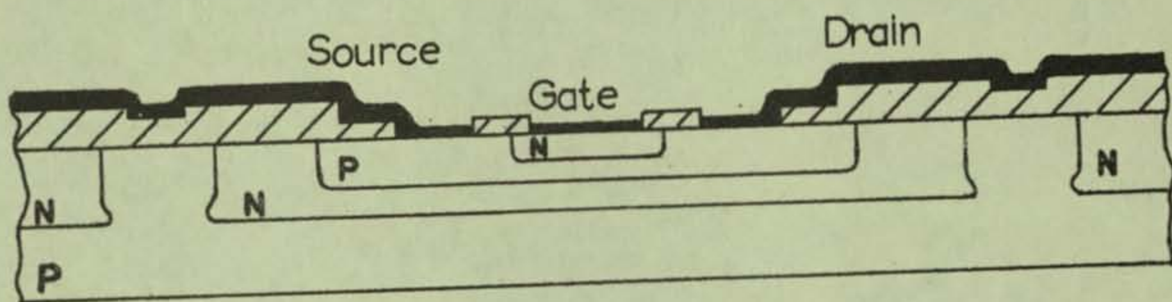
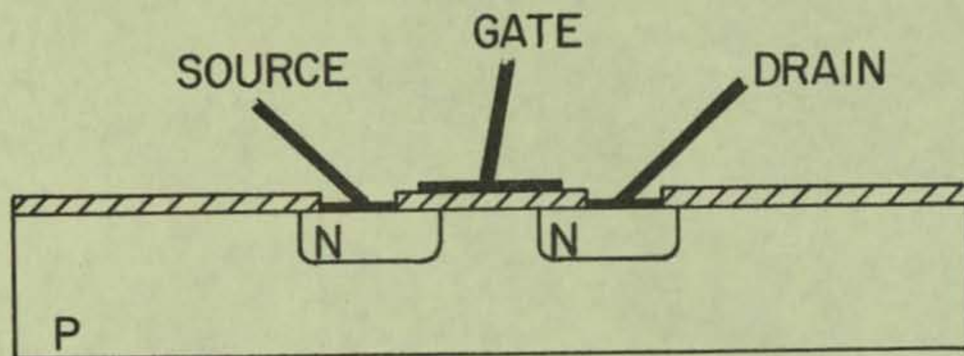
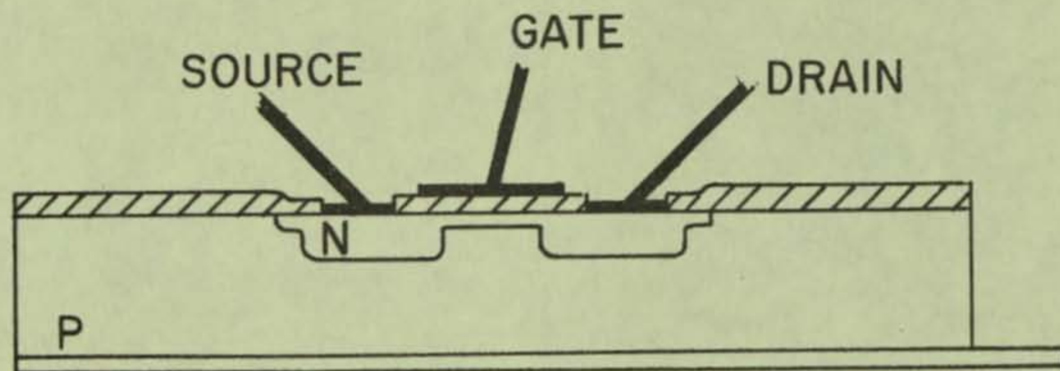


Fig. 5-42. Isolated unipolar field effect transistor structure.



a) NORMALLY OFF



b) NORMALLY ON

Fig. 5-43. Schematic cross-sections of surface controlled field effect transistors. In the normally OFF structure a channel is induced between the n-type regions by applied gate voltage. The normally ON configuration is usually used with gate voltages to decrease the channel conductance.

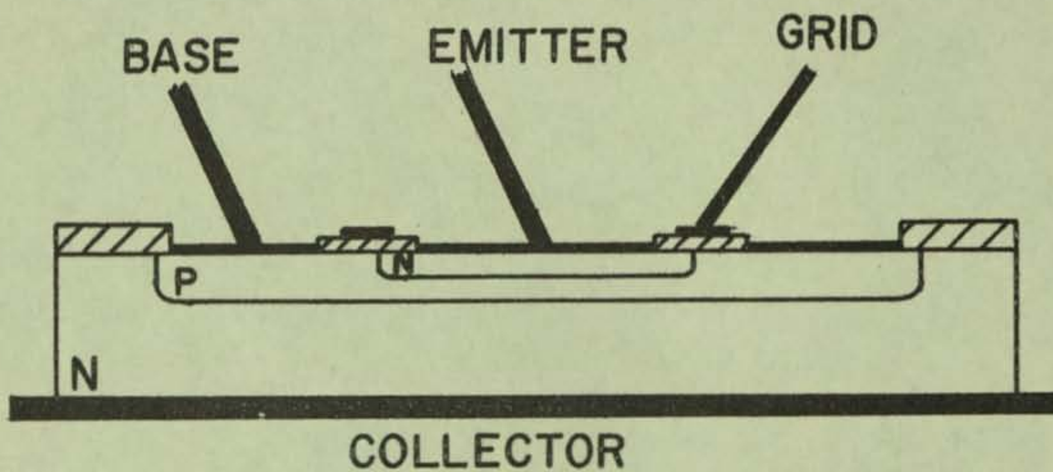


Fig. 5-44. Surface controlled field effect transistor. The grid electrode acts as a high impedance input which modulates the current gain of the transistor structures.

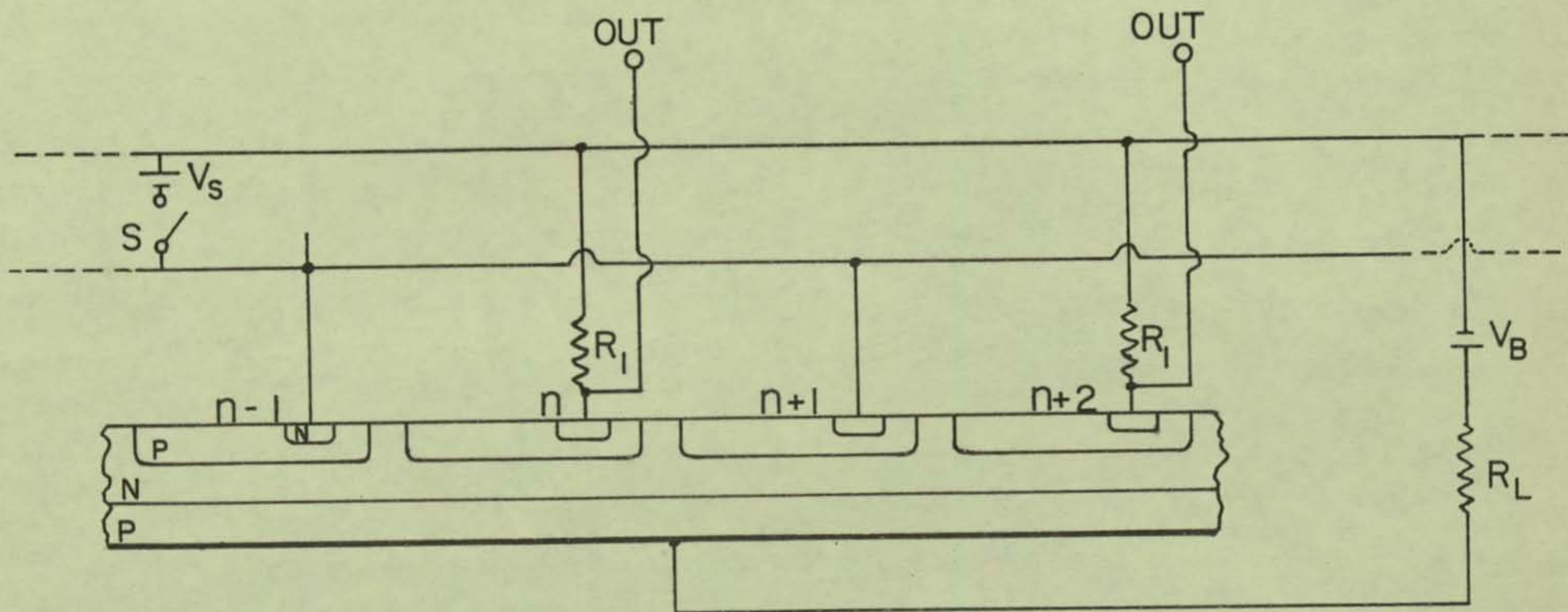


Fig. 5-45. Schematic representation of the connection of an integrated pnpn structure to construct a pulse counter (after D'Asara Ref. 46).

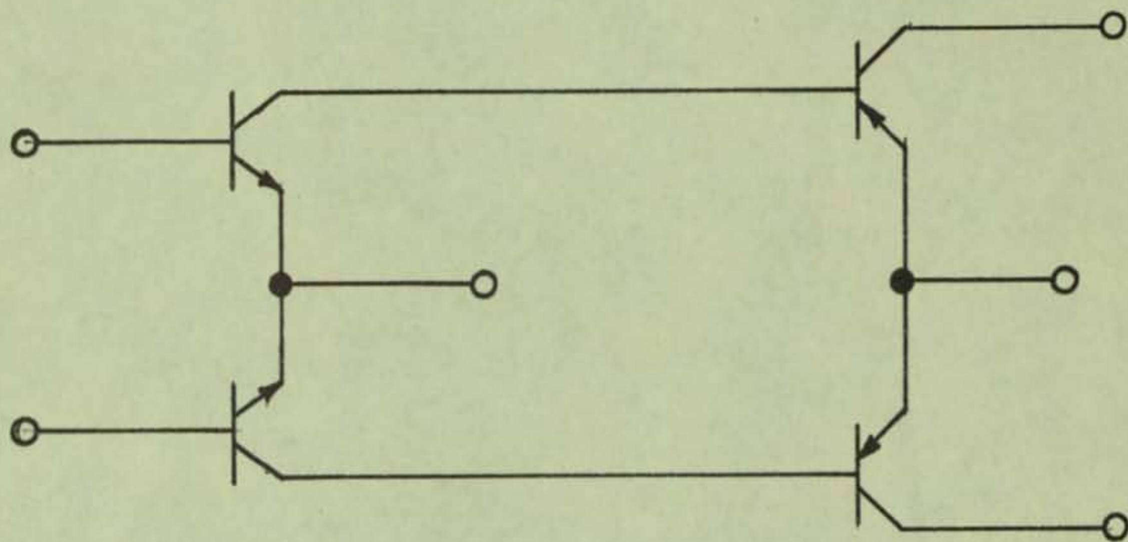


Fig. 5-46. Complementary circuit configuration useful in analog circuitry.

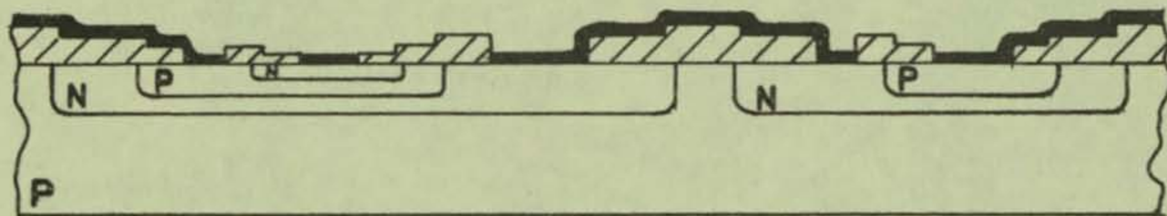


Fig. 5-47. Schematic cross-section of a structure with npn and pnp transistors in the same substrate.

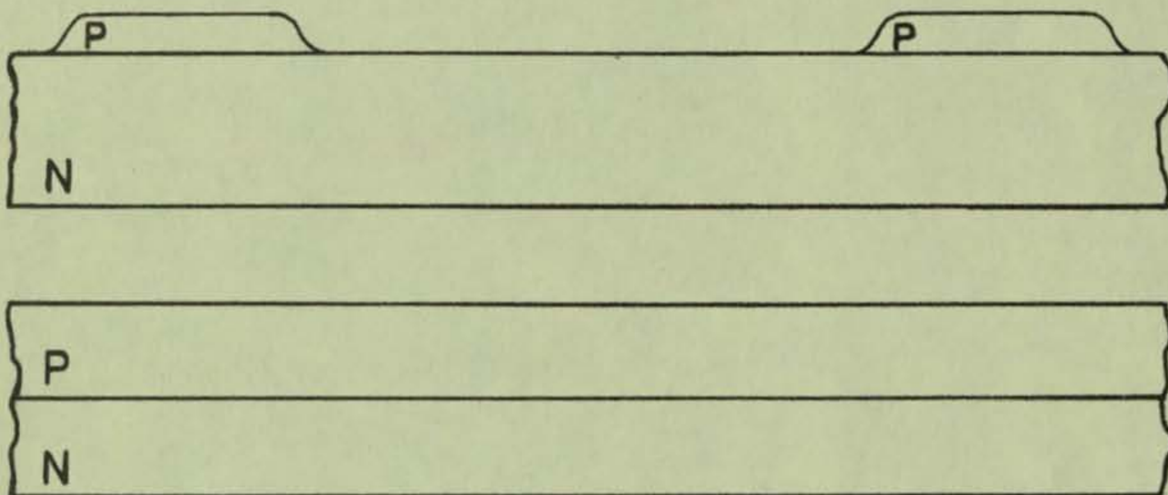


Fig. 5-48. Schematic cross-section of two structures achievable by epitaxial growth suitable for the construction of optimum npn and pnp transistors in the same semiconductor substrate.

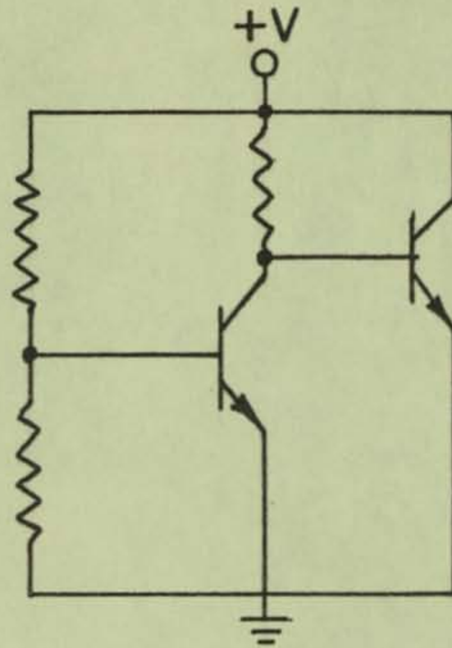
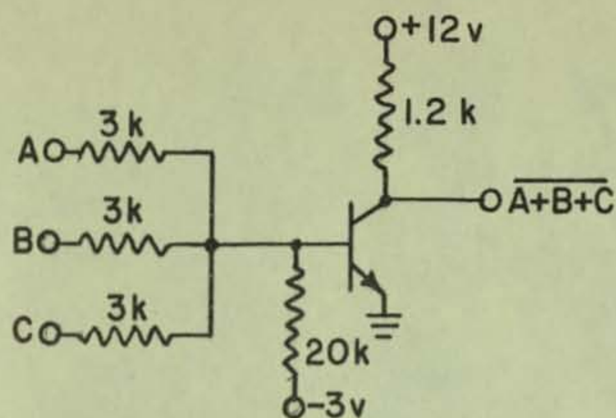
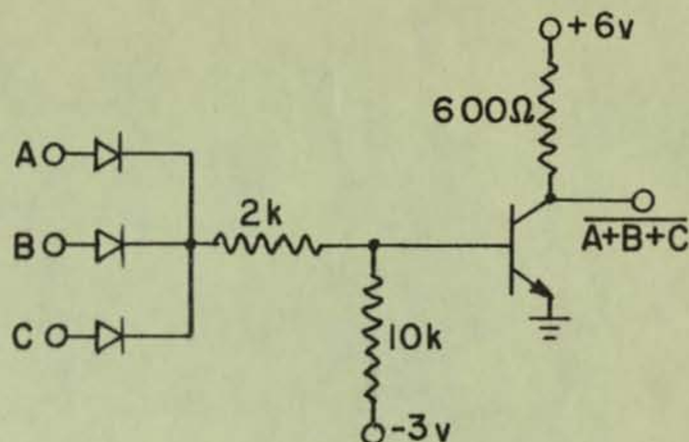


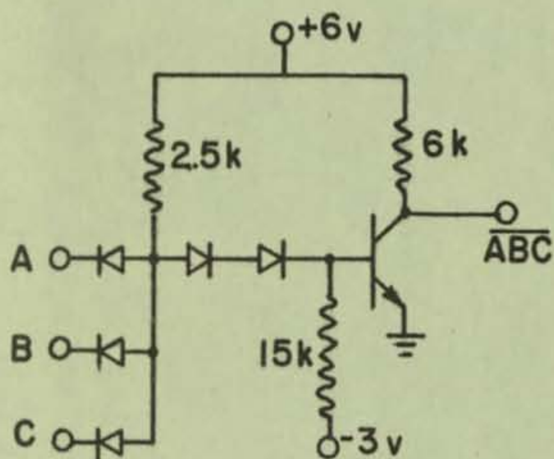
Fig. 5-49. Circuit diagram of structures suitable for maintaining constant temperature in an integrated circuit.



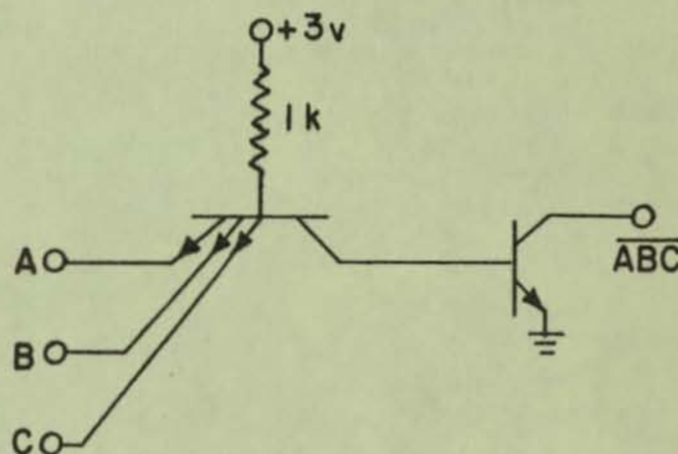
a) Transistor–Resistor Logic NOR (TRL)



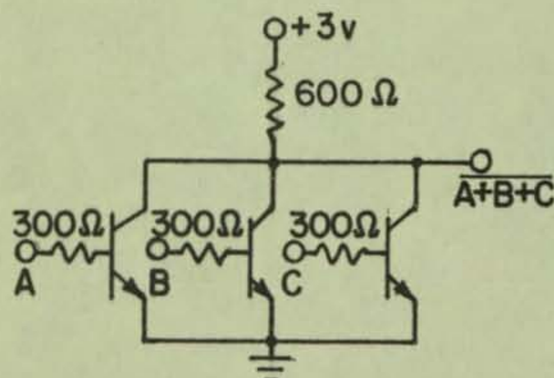
b) Diode–Transistor Logic NOR (DTL)



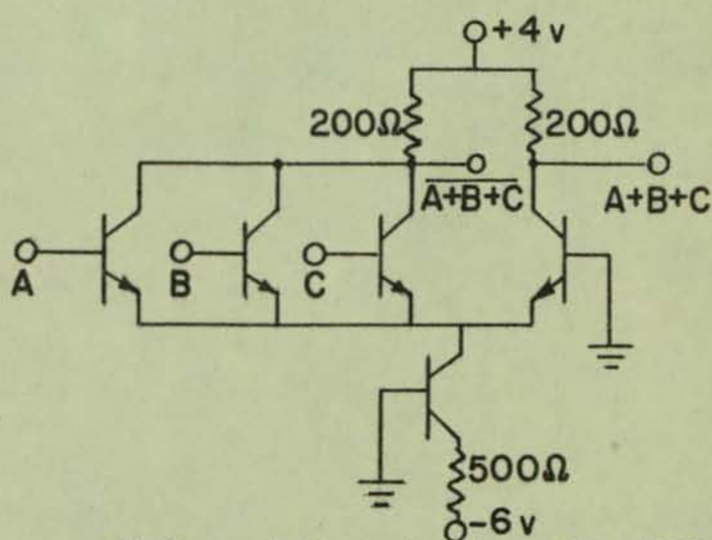
c) Low Level Logic NAND (LLL)



d) Transistor–Transistor Logic NAND (TTL) or (T^2L)



e) Modified Direct Coupled Logic NOR (DCTL)



f) Current Mode Logic OR–NOR (CML)

Fig. 5-50. Various forms of logic gates suitable for integration. Values given are typical of those commonly employed.

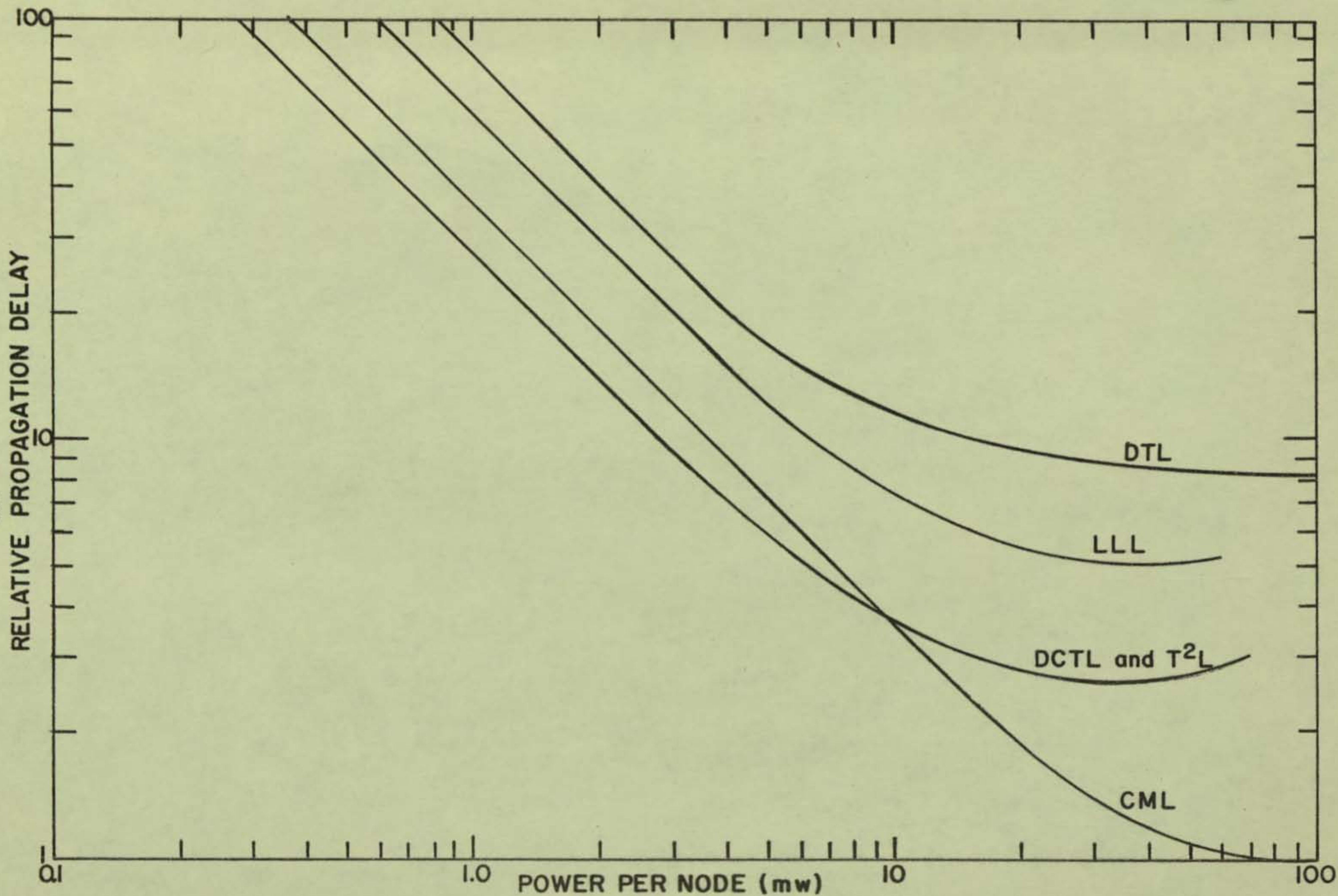


Fig. 5-51. Relative propagation delay versus power for fan-in and fan-out of unity, comparing various integrated circuits made with a particular technology.

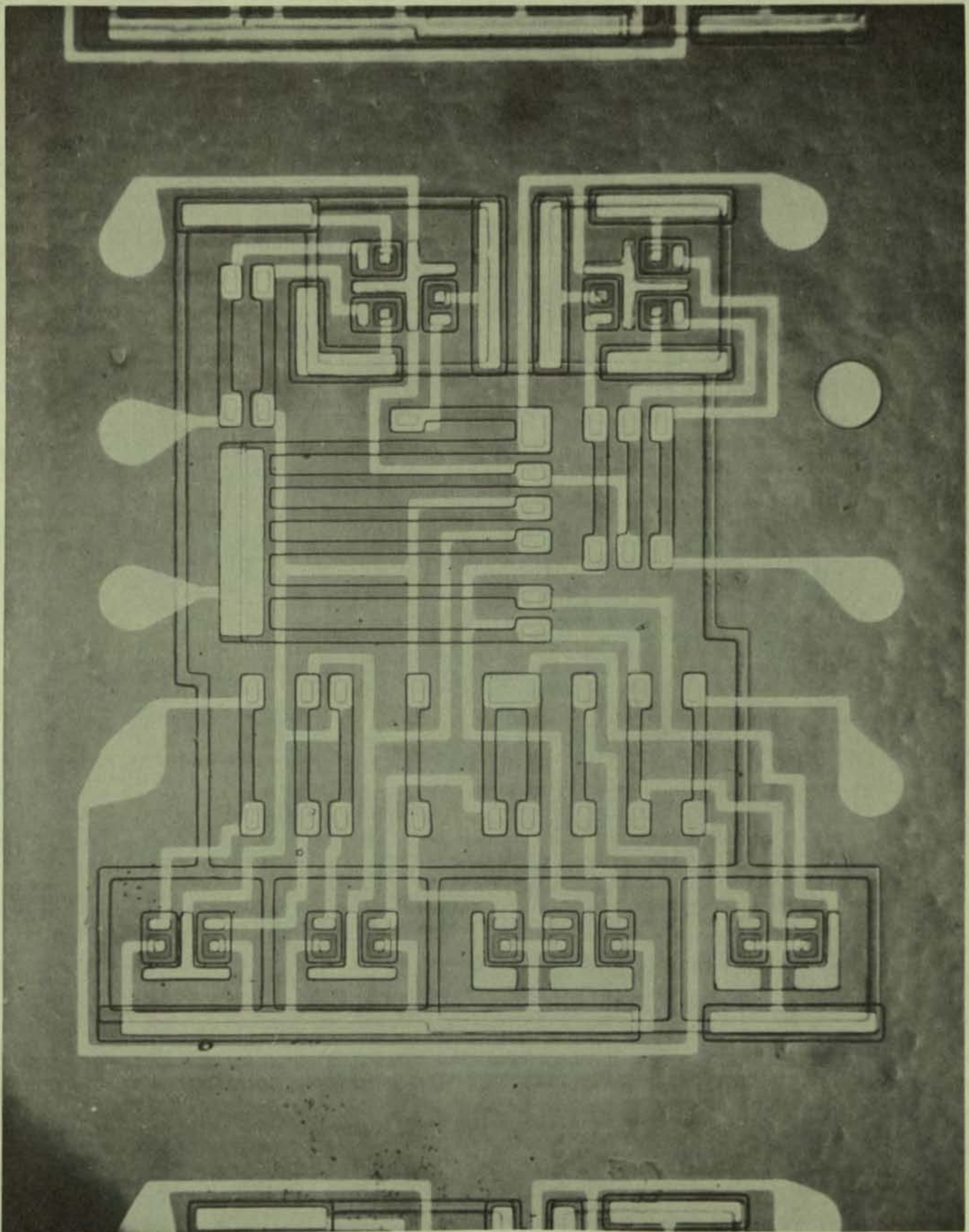


Fig. 5-52. Photomicrograph of a complete shift register stage using diffused resistors. This circuit has 15 transistors and 21 resistors. It employs grid type diffused isolation.

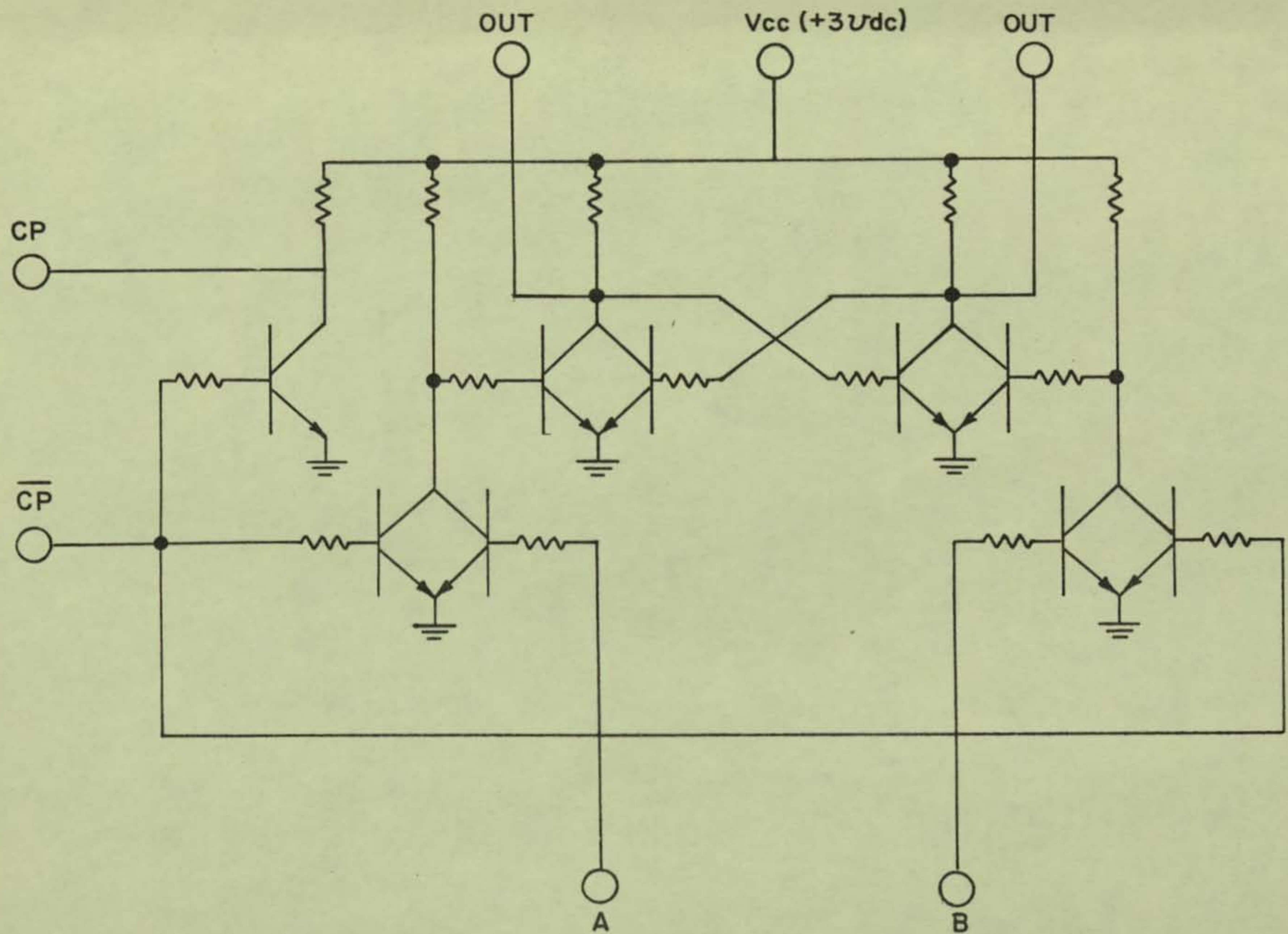
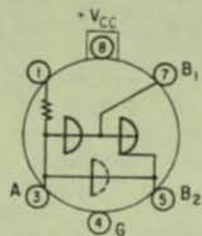
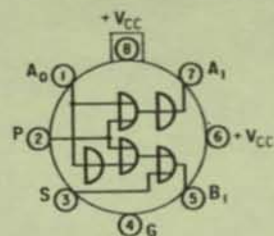


Fig. 5-53. Circuit diagram corresponding to a Micrologic "S" element.



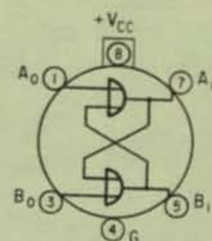
MICROLOGIC ELEMENT "B"
BUFFER

$B_1 . B_2 = \bar{A}$



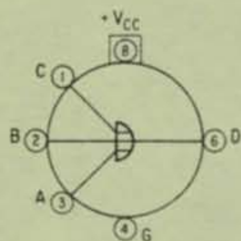
MICROLOGIC ELEMENT "C"
COUNTER ADAPTER

$\bar{A}_1 = \bar{A}_0 \bar{P}$
$\bar{B}_1 = A_0 \bar{P} + S$



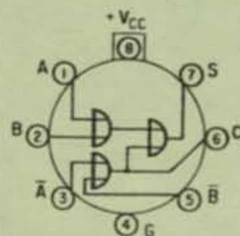
MICROLOGIC ELEMENT "F"
FLIP - FLOP

$\bar{A}_1 = B_1 + A_0$
$\bar{B}_1 = A_1 + B_0$



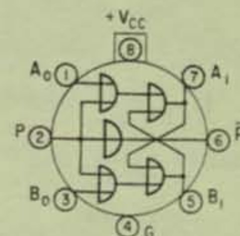
MICROLOGIC ELEMENT "G"
GATE

$D = (\overline{A + B + C})$



MICROLOGIC ELEMENT "H"
HALF ADDER

$S = A\bar{B} + \bar{A}B$
$C = AB$



MICROLOGIC ELEMENT "S"
HALF SHIFT REGISTER

$\bar{A}_1 = B_1 + \bar{A}_0 \bar{P}$
$\bar{B}_1 = A_1 + \bar{B}_0 \bar{P}$

Fig. 5-54. Logic diagram and Boolian functions of Micrologic elements. Each "D" shaped symbol represents a nor gate. The dotted symbol in the buffer element represents an emitter follower.

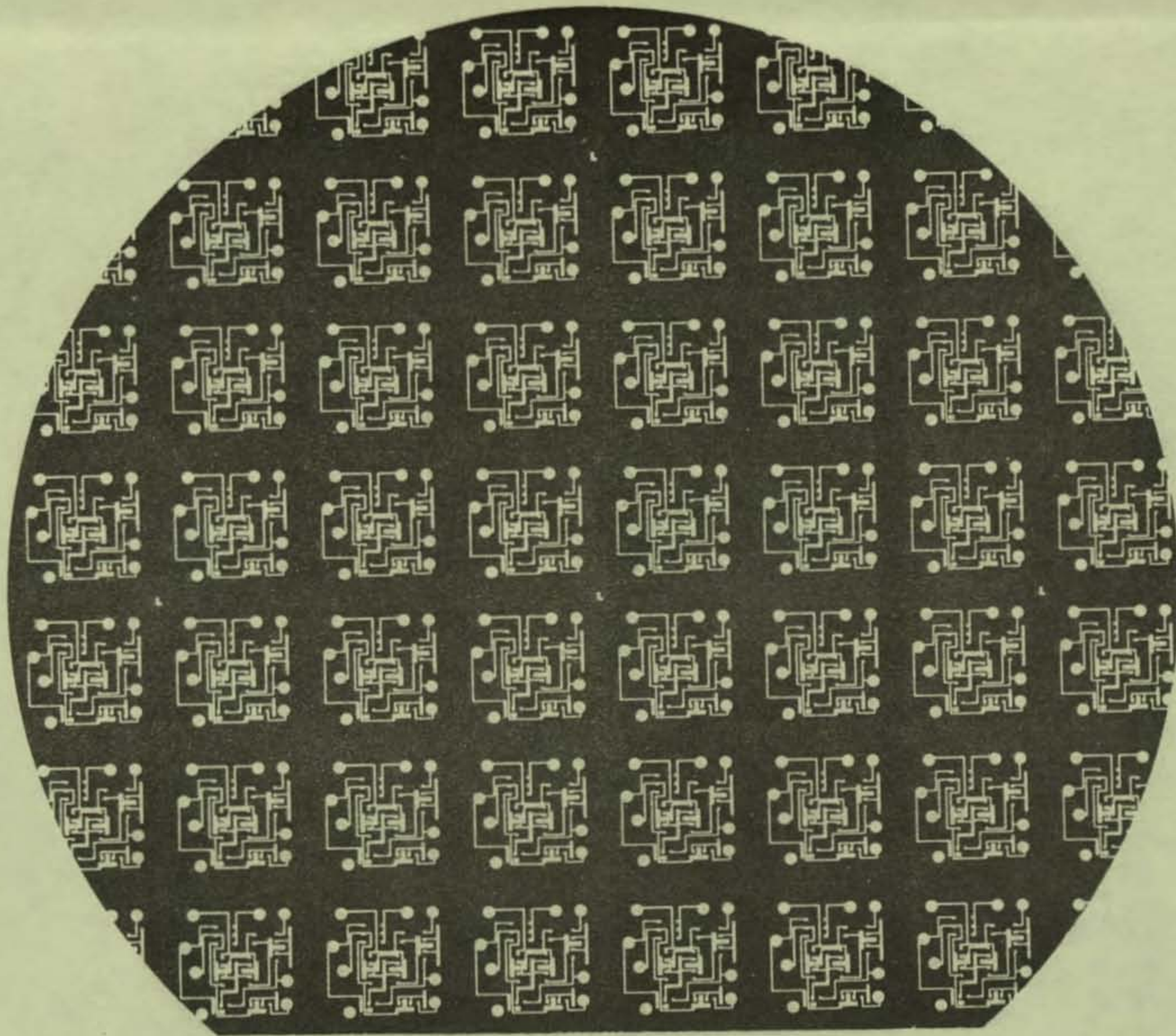


Fig. 5-55. A wafer of a completed microcircuit structure. Only the aluminum intraconnections for lead bonding pads are visible with this illumination. Each of the structures is the nine transistor, fourteen resistor Micrologic "S" element. The wafer is approximately 3/4" in diameter.

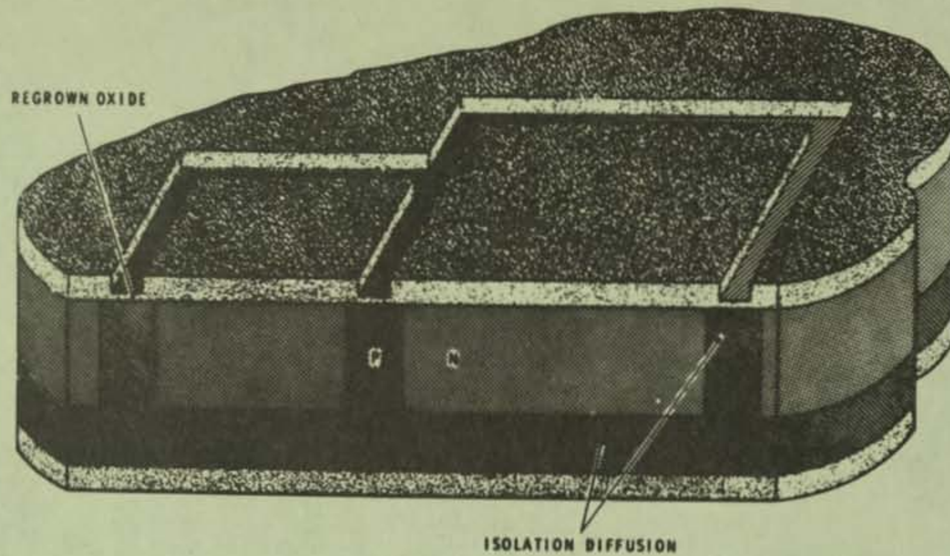


Fig. 5-56. Photomicrograph and schematic cross-section of a Micrologic element after various process steps. (a) after isolation diffusion;

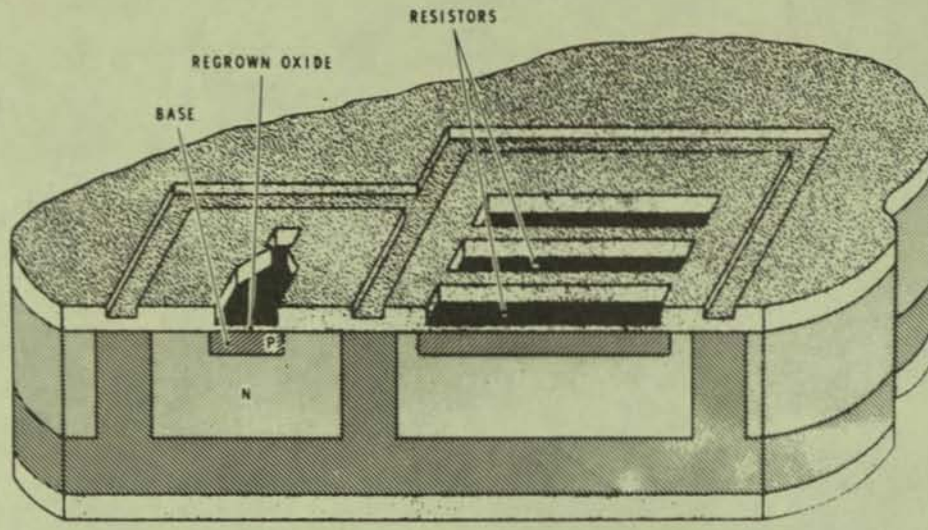


Fig. 5-56(b). after base and resistor diffusion

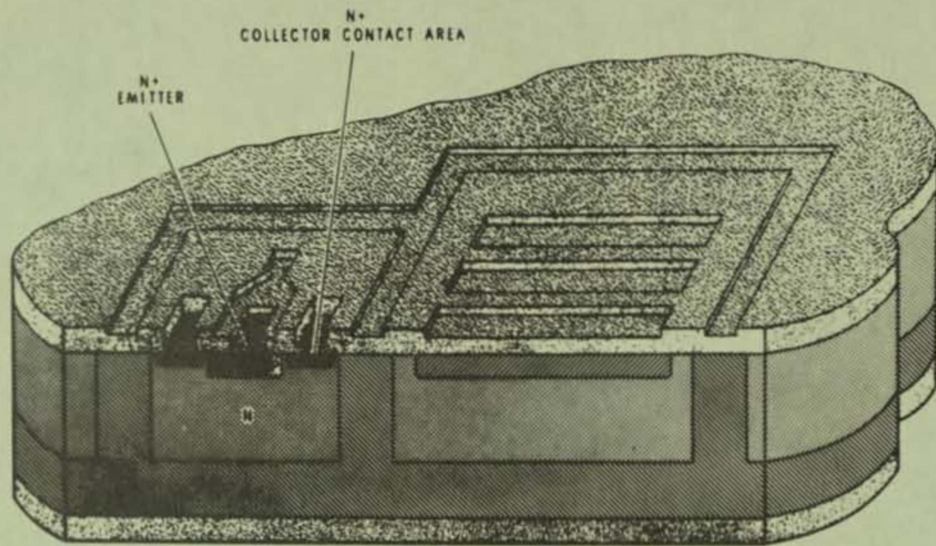


Fig. 5-56(c). after emitter diffusion

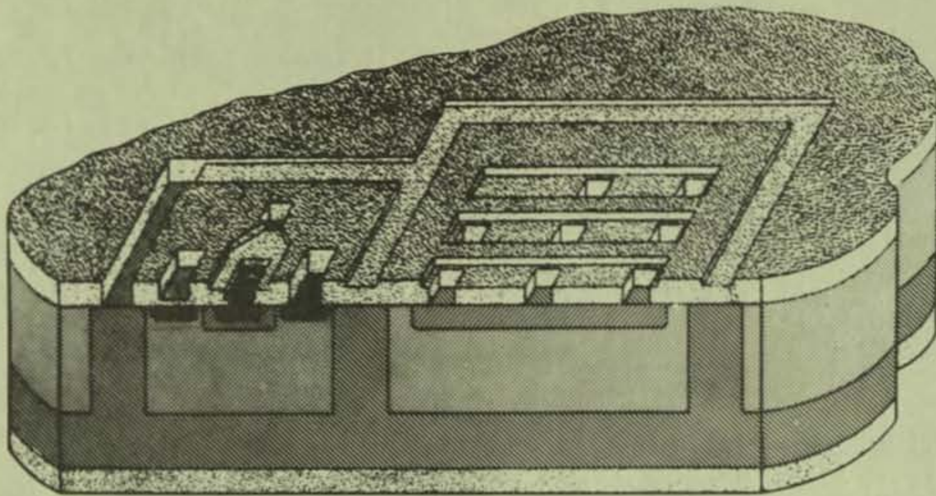


Fig. 5-56(d). after oxide removal for contacting

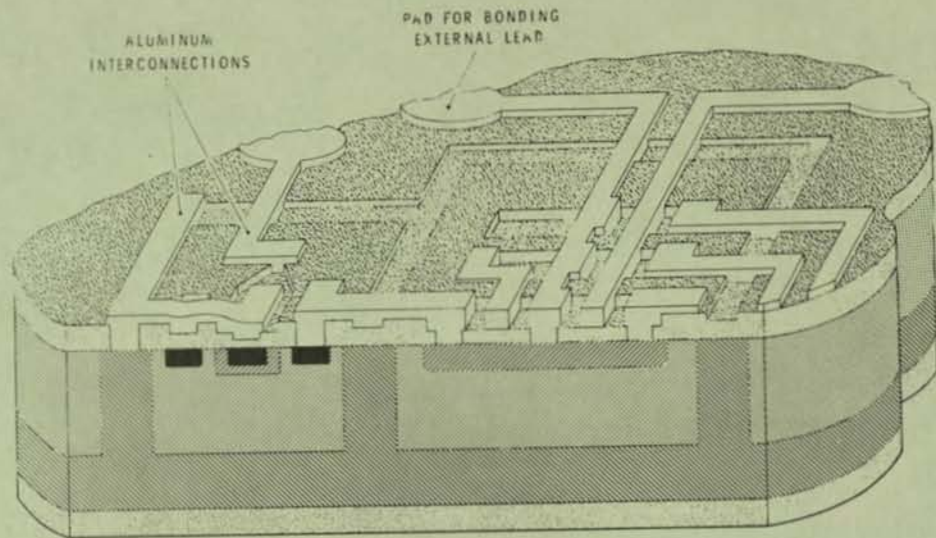


Fig. 5-56(e). after intraconnection metallization.

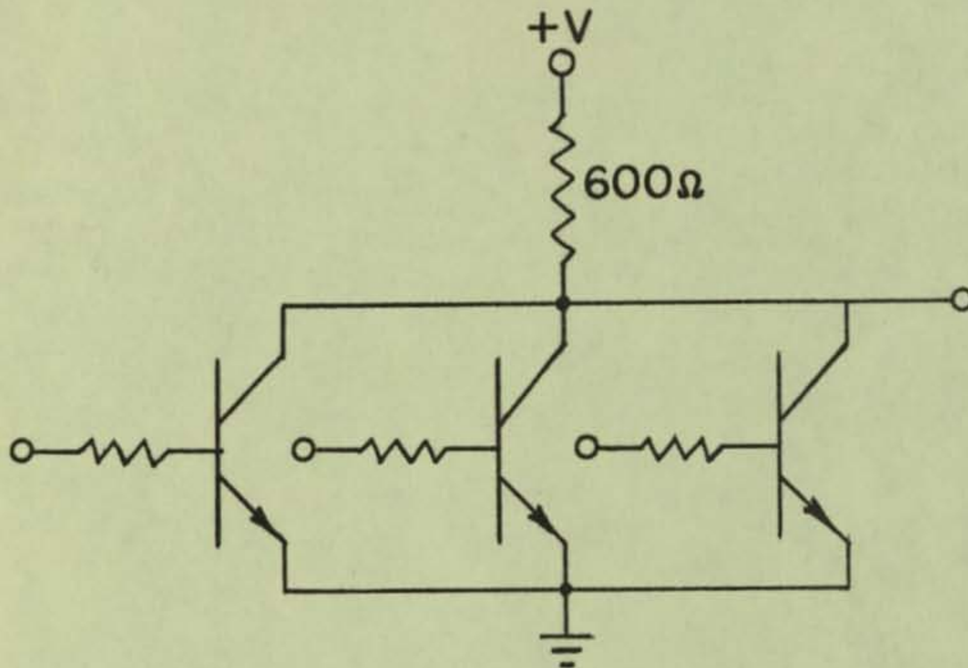


Fig. 5-57. Circuit diagram of a Micrologic "G" element.

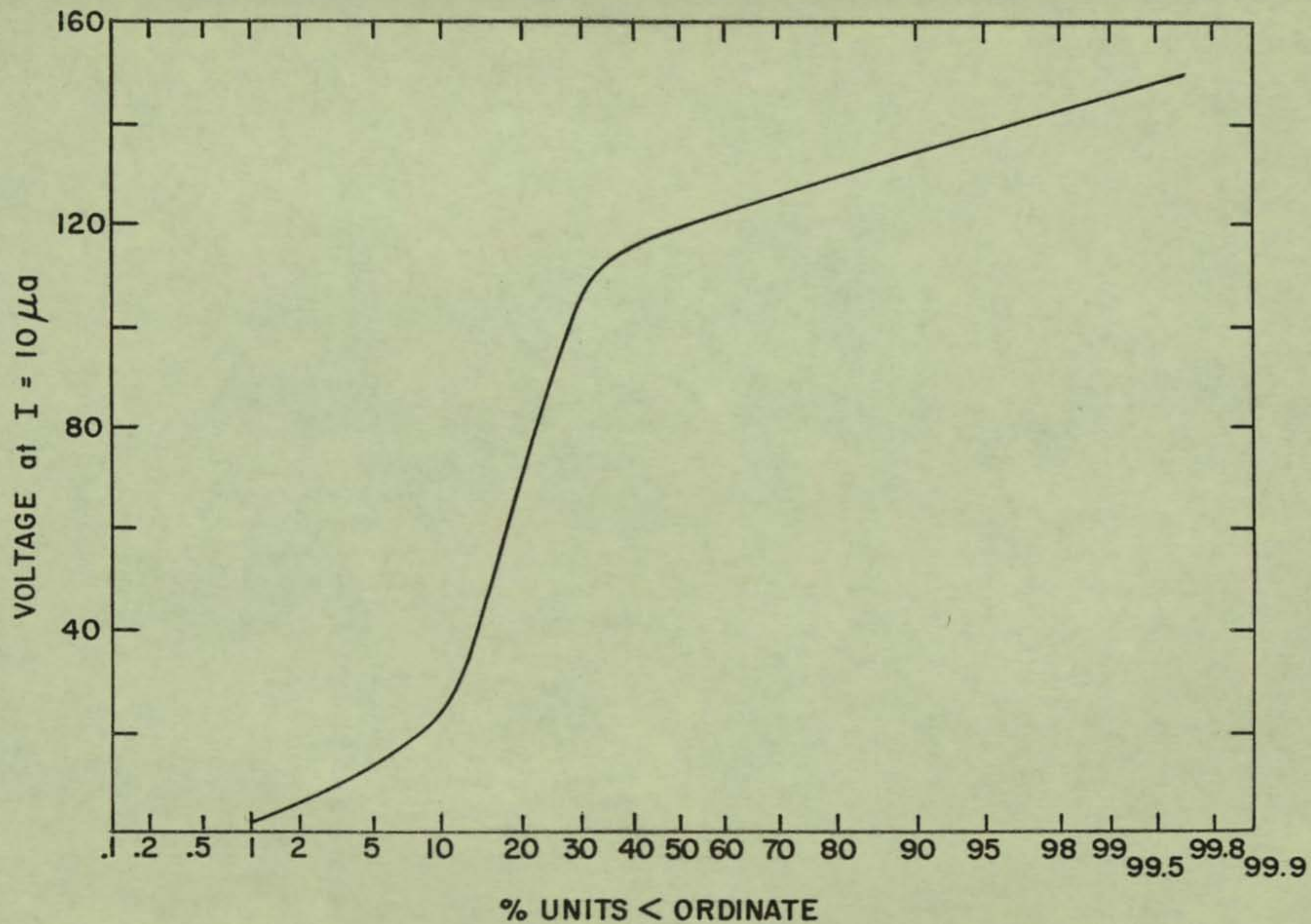


Fig. 5-58. Distribution of voltage across isolation region corresponding to 10 μa leakage current for a Micrologic gate element.

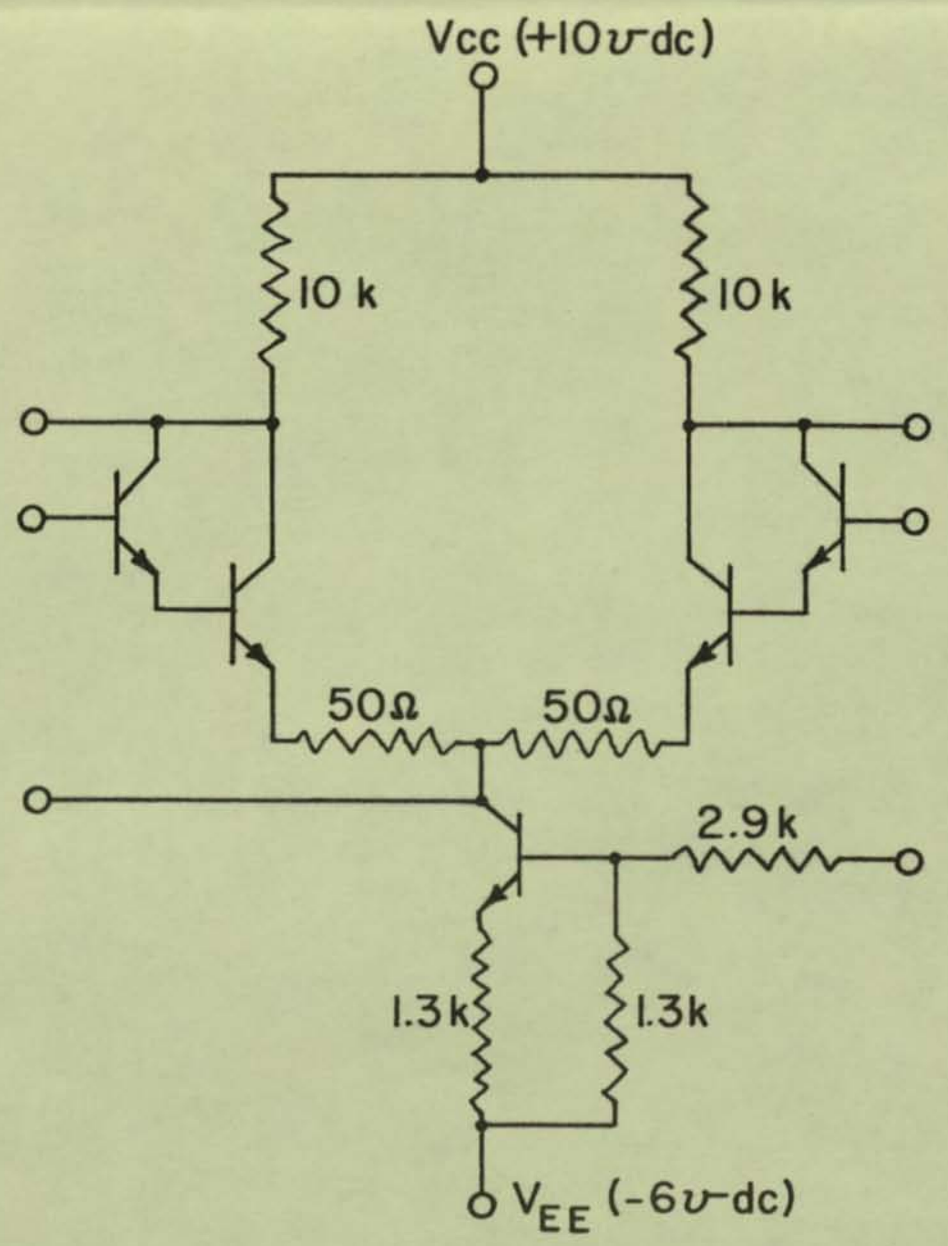


Fig. 5-59. Circuit diagram of Darlington differential amplifier structure.

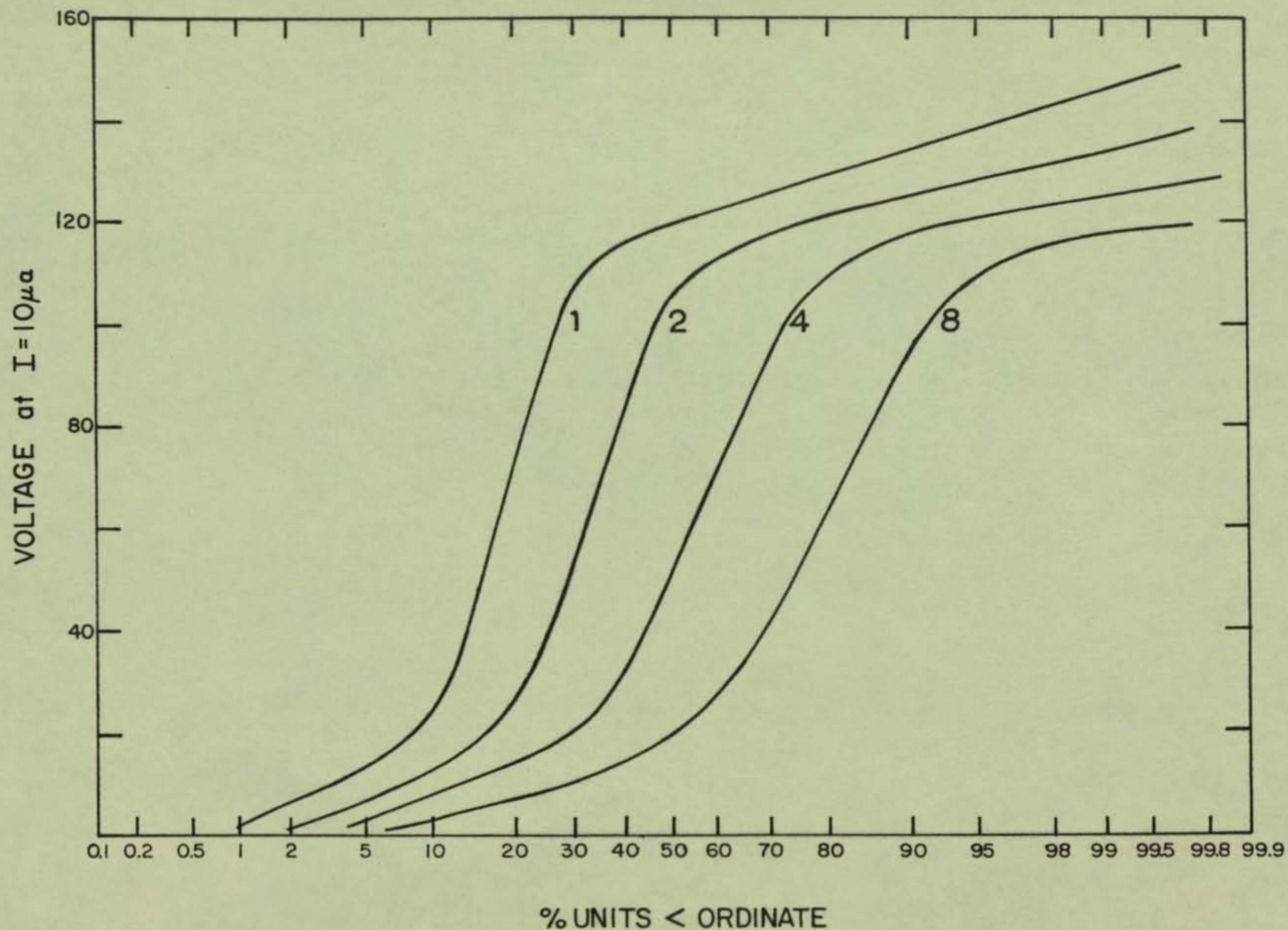


Fig. 5-60. Distribution of isolation breakdown voltage for 1, 2, 4 and 8 isolation regions respectively. The curves with several regions were generated from the curve for one region, assuming a random distribution.