FABRICATION OF BIPOLAR JUNCTION TRANSISTORS

by

SUKHDEV SINGH DHALIWAL

B.Sc. Engineering (Hons), Punjab University, India, 1969

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Electrical Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1972

Approved by:

M. L. Lucas

Major Professor

THIS BOOK CONTAINS NUMEROUS PAGES WITH THE ORIGINAL PRINTING BEING SKEWED DIFFERENTLY FROM THE TOP OF THE PAGE TO THE BOTTOM.

THIS IS AS RECEIVED FROM THE CUSTOMER.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

TABLES

FIGURES

CHAPTER I

INTRODUCTION

The transistor has been one of the single most significant inventions to catalyze industrial growth. Since its advent in 1948 there has been a phenomenal rise in its application to every sphere of technology. It has made possible for man to implant his footprint on the moon. Time is not far off when man will not be able to imagine life without the transistor. It has given way to a new technology - that of integrated circuits which promises to have a greater effect on electronic systems than the transistor itself.

After the invention of bipolar junction transistor (BJT) other kinds such as field effect transistors (JFET's and IGFET's) have come into being. Potentially these are considered to be cheaper (especially IGFET's) than the BJT's owing to the fewer fabrication operations required but the latter excels considerably in performance. Their input threshold voltage is much lower and sharper than most of the IGFET's. More over some new processes like base diffusion isolation[10] have come a long way to achieve simplification of the fabrications processes involved.

The purpose of this thesis is to study the various techniques involved in the fabrication of an integrated circuit bipolar junction transistor. It has been tried to demonstrate that the fabrication of silicon integrated circuits can be carried out using the "not-so-sophisticated" equipment available in the laboratory. To this end, the fabrication of bipolar junction transistors beginning with the preparation of the artwork for the photomasks to the metallization has been carried out in the laboratory.

CHAPTER II

REVIEW OF LITERATURE

The contemporary silicon integrated circuit is an offspring of semiconductor technology. Before the advent of integrated circuits, circuits were built from discrete components. In complex circuits this results in numerous interconnections and their associated parasitic capacitances. The need for smaller and smaller size and high speed circuits gave birth to the idea of an integrated circuit.

Since p-n junctions have capacitances and semiconductor materials have resistances it probably occurred to many people early in the transistor history that the fact could be exploited to fabricate the entire circuit on a single semiconductor chip. However the actual growth started with the invention of p-n junction isolation technique by Lehovac* and the planar diffusion process by Hoerni**. The Lehovac patent resulted in the achievement of "logical synthesis of complex electronic functions by the interconnection of separate components in a single block of semiconductor material".[5] The planar process made the increased fabrication yield possible. Isolation between different components is necessary in an integrated circuit. Depending upon the type of isolation and the way it is achieved it gives rise to various processes as follows:

---

*Patent number 3,029,366 Sprague Electric Company, Massachussets.

**Patent number 3,025,589 Fairchild Camera and Instrument Corporation, New York.

(A) P-N junction isolation - It can be further subdivided as:

1. Triple diffusion process.
2. Double diffusion epitaxial.
3. Standard buried collector (S.B.C.)
4. Collector diffusion isolation (C.D.I.)
5. Base diffusion isolation (B.D.I.)

(B) Silicon dioxide isolation.

(C) P-I-N isolation.

(D) Beam lead technology.

A.1 Triple Diffusion Process[5]

This was the first scheme to exploit the p-n junction isolation. A p-type silicon wafer (typically 10 μm) is used as a substrate. In it, n-type collector areas are selectively diffused. After this step a p-type diffusion and an n-type diffusion is made selectively for base and emitter areas respectively. The substrate is connected to the most negative voltage point in the circuit. The transistors are thus isolated by two reverse biased diodes.

The process has some disadvantages as follows:

(i) In order to have lower collector series resistances the impurity concentration in the collector region has to be high. But this results in a lower collector to base breakdown voltages.

(ii) The impurity concentration in the collector region is graded. It is highest at the collector-base junction and lowest at the collector-substrate junction. This results in a lower collector to base breakdown voltage.

(iii) The presence of thermally grown oxide tends to invert the surface characteristics in the p-type isolation regions resulting in a short circuiting channel.[5]

A.2 Double Diffused Epitaxial:

In this process the starting material is a lightly doped p-type silicon wafer. A thin (5-15 μm)[5], n-type layer is grown epitaxially over it. Isolation between different components is achieved by selectively diffusing p-type impurity in this epitaxial layer to make contacts with the p-type substrate. Transistors are fabricated in the resulting isolated pockets of n-type material. The process has some advantages over the triple diffusion process as listed below:

(i) The epitaxial layer need only be thick enough to fabricate the devices. Therefore the time required for isolation diffusion assumes manageable proportions.[2]

(ii) Since epitaxy results in a relatively abrupt doping profile, the choice of layer resistivity may be based only upon device design considerations.[5]

However achieving lower collector series resistances without sacrificing collector to base breakdown voltages is still a problem. This is overcome by using a low resistivity buried collector layer.

A.3 Standard Buried Collector Process (S.B.C.)[2,5]

It is the mostly used single process. A buried layer of very low resistivity (5-15 ohms/square)[5] is employed to get lower collector series resistances. Since this layer is isolated from the collector-base junction, the collector to base breakdown voltage is relatively high.

Collector Diffusion Isolation (CDI) process:[10]

In this process the isolation diffusion is carried out simultaneously with the collector diffusion. The sequence of operations is as follows:

1. A buried n^+ layer is selectively diffused in a high resistivity p-type wafer.
2. A thin, p-type, layer is grown, epitaxially, over the entire surface.
3. A heavily doped n^+ diffusion is made selectively to make contacts with the buried collector layer. This diffusion is carried out around the periphery of every transistor so that the resulting p-type islands can be used as base layers for transistors.
4. A p-type unmasked diffusion is made into the entire surface. It has the following advantages:
   a. It inhibits the formation of surface inversion layer due to thermally grown oxides.
   b. It minimizes the edge injections from emitters thus increasing the current gain.
   c. It provides a built in field at the surface that reduces the flow of electrons to the surface resulting in a lower effective recombination velocity.
5. Emitter diffusion and ohmic contacts are made as usual.

The process is simpler as compared to standard buried collector process since there is no need for a separate isolation diffusion. Secondly, only shallow diffusions are required to achieve isolation due to thin epitaxial layers employed.

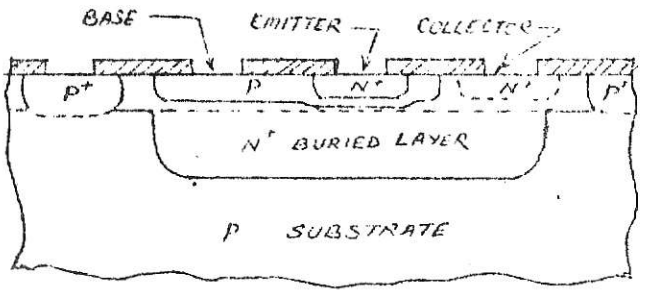
THIS BOOK CONTAINS NUMEROUS PAGES WITH DIAGRAMS THAT ARE CROOKED COMPARED TO THE REST OF THE INFORMATION ON THE PAGE.

THIS IS AS RECEIVED FROM CUSTOMER.

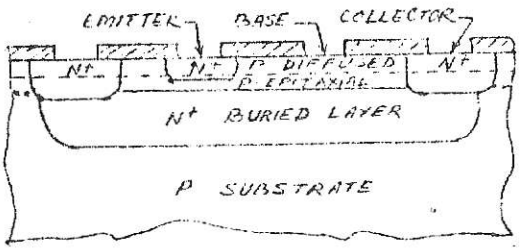
ILLEGIBLE DOCUMENT

THE FOLLOWING DOCUMENT(S) IS OF POOR LEGIBILITY IN THE ORIGINAL

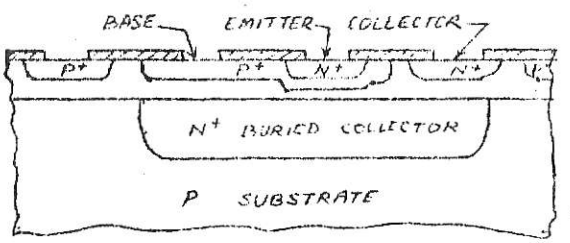
THIS IS THE BEST COPY AVAILABLE



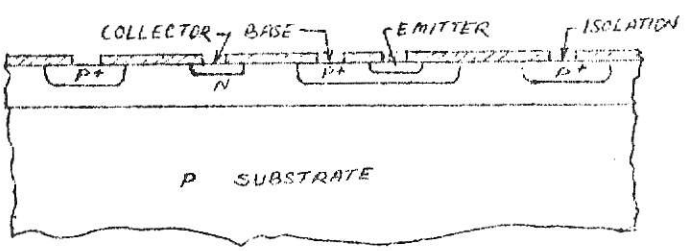
STANDARD BURIED - COLLECTOR



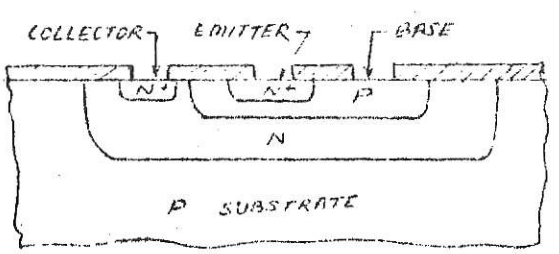
COLLECTOR DIFFUSION ISOLATION



BURIED COLLECTOR B.D.I.



LOW POWER B.D.I. STRUCTURE



TRIPLE DIFFUSION STRUCTURE

Figure (2-1) Different structures employing p-n junction isolation

To have a minimum movement of the buried n^+ layer during later diffusion or oxidation cycles, dopants like As or Sb are used. The diffusion coefficients of these dopants are lower than those of the conventional dopants (boron and phosphorus) by an order of magnitude. Since arsenic is toxic in nature, antimony is preferred. But due to lower solid solubility of Sb in silicon, sheet resistances only as low as 15-20 ohms/square can be achieved. With arsenic sheet resistances up to 5 ohms/square are possible.[5] The process is described in detail as follows:

1. A p-type substrate is used as substrate. A n^+ buried layer is selectively diffused in it.
2. A thin (10-25 μm) n-epitaxial layer is grown over the entire surface.
3. P-type isolation diffusion is carried out to diffuse through the epitaxial layer to make contact with the p-substrate.
4. Base and emitter diffusions are carried out as usual.

In order to achieve good ohmic contacts with the lightly doped n-regions, a heavily doped n-type layer is diffused under the collector contact areas along with the emitter diffusion.

The standard buried collector process requires four separate diffusions and one epitaxial deposition. However, if a thin epitaxial layer is used, the isolation diffusion can be carried out simultaneously either with the collector or base diffusion. These processes described by Murphy et. al.[10] are termed as collector diffusion isolation (CDI) and base diffusion isolation (BDI) respectively.

This helps in reducing the area required for each transistor. A 4:1 area advantage can be achieved as compared to the standard buried collector process. However the transistors obtained in the process have lower collector to base breakdown voltages (typically 6 to 8 volts).

Base Diffusion Isolation (B.D.I.) Process:[10]

In this process the isolation diffusion is carried out simultaneously with the base diffusion. The essential operations of the process are as follows:

1. A p-type substrate (typically 10 μm) is used as a starting material. A heavily doped n^+ buried layer is selectively diffused in it.
2. A thin n-type epitaxial layer of sufficiently low impurity doping is deposited over the entire surface.
3. A heavily doped p-type diffusion is selectively carried out to get the base and isolation areas.
4. A heavily doped n^+ impurity diffusion is selectively made in the base regions to get the emitter areas.
5. Ohmic contacts are made as usual. Contacts are also made to the isolation stripe.
6. The isolation is achieved by applying a large enough negative voltage to the p^+ isolation stripe, which is punched through the substrate by the depletion layer around it.

The epitaxial layer must be sufficiently thin and/or sufficiently lightly doped so that punch-through occurs below the breakdown voltage. Typically, for a 0.5 ohm cm n-type layer of 2μm thick and a base diffusion depth of 1μm, the punch through voltage is 10 volts, whereas bulk breakdown voltage is 50 voltages.

The base diffusion isolation described above requires the same number of operations as the collector diffusion isolation process. However, for low power circuits, further simplification in the process can be achieved. If the circuits are to be operated at sufficiently low current levels, say, at collector currents of <<1mAmp, the buried collector layer can be eliminated.

The buried diffusion isolation structure requires an extra power supply for isolation. However, very little current flows in this network. The isolation junctions should not extend to the chip edge to avoid leakage currents.

B. P-I-N Isolation

In this structure described by Vore, et. al.[14] an intrinsic region replaces the conventional space charge region in the p-n junction isolation. The thickness of the intrinsic region (I-region) can be controlled, unlike the space charge width. This lowers the substrate to collector capacitance by an order of magnitude. The process for achieving an intrinsic region between the p and n type regions is explained as follows:

1. A high resistivity n-type epitaxial layer is deposited over a heavily doped p^+-type substrate. The thickness of this layer determines the thickness of the I-layer and hence the collector-to-substrate capacitance.
2. A p^+ isolation diffusion is made after growing an oxide and selectively etching it.
3. The n^+ collector regions are diffused.
4. Oxide is etched from the entire surface and a second n-type epitaxial layer is deposited. The impurities from the p^+ and n^+ regions out diffuse into this layer. But the arsenic used for n^+

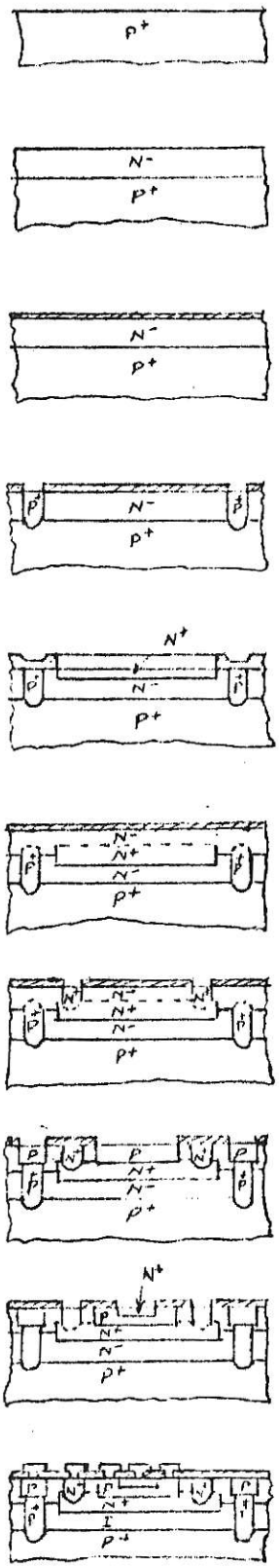


Figure (2-2) P - I - N Isolation structure

collector regions has a much lower diffusion coefficient than the boron used for the p^+ isolation diffusion.

5. Heavily doped n-type rings are diffused around the collector regions.
6. P-type impurity rings are diffused simultaneously with the base diffusion to meet the out diffusion from the p^+ ring into the n^- region.
7. A heavily doped n^+ diffusion is made in the emitter areas as well as in the regions under collector contacts.
8. Gold is evaporated on the back side of the wafer. The wafer is later heated to diffuse the gold. This converts the n^- regions into intrinsic regions.

The various operations carried out during the process are explained in Figure (2.2).

C. Silicon dioxide Isolation

At the first thought the most forward approach towards isolation seems to be the use of an insulating layer between different islands of single crystal semiconductor. The process described by Warner, et. al.[15] is described as follows:

1. A silicon wafer of appropriate impurity doping is taken. Channels are etched in this wafer selectivaly.
2. A sufficiently thick layer of silicon dioxide is grown over the entire surface of this wafer.
3. A polycrystalline silicon layer is grown on the top of the oxide layer.
4. The crystal is lapped on the single crystal side resulting in pockets of single crystal silicon in polycrystalline substrate,

isolated by silicon dioxide. These pockets are used for component fabrication. The polycrystalline material serves as a mechanical base.

Ideally this is a better isolation as compared to the p-n junction isolation, but the processes involved are cumbersome. Lapping has to be carried out very carefully since a slight mechanical misalignment can destroy the shallow (typically 10 μm) single crystal pockets.

Another process described by Allison et. al.[1] uses an anisotropic etchant that form V-shaped areas. The directional etchant etches 30 times more rapidly along the 100 crystal plane--where atoms are less densly packed-- than it does along the 111 plane. Thus the etchant forms a V-shaped groove, the depth of which depends only on the width of the photolithographically produced isolation mask. Since etching virtually stops when the 100 plane is no longer exposed - that is, when the tip of the V is reached - the process is easily controlled.

Conventional etchants attack the crystal equally in all directions, producing grooves at least twice as wide as they are deep. Therefore the components can not be densely packed.

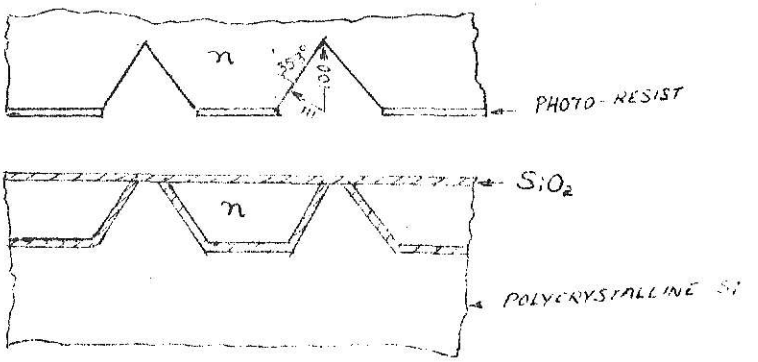


Figure (2-3)

V-type grooves along 100 plane

D. Beam-lead Construction:

In this structure described by Hause, et. al.[7] gold leads are used for interconnection and isolation. Here interconnections between isolated regions are formed by electroplating additional gold (to about 15 microns thickness) on a composite metal contact and interconnect pattern. Isolation is then achieved by etching from the back (unmetallized) surface of the wafer through to the surface carrying the gold interconnects. This etching is restricted by photoresist techniques to narrow regions surrounding each area to be isolated. The etching is stopped when all silicon has been removed from between the islands. The interconnections in the isolation regions and other regions where the attachment of the circuit to the package is to be made are kept more thick (15 microns) as compared to other regions. The islands remain hooked together electrically and physically by the relatively thick gold patterns. The physical gap between the silicon islands provides isolation of the active device regions. The size of the gap is determined by the initial wafer thickness and etching process and usually uses more silicon area than either junction or solid dielectric methods.

CHAPTER III

OXIDATION

Oxidation plays a vital role in contemporary silicon device fabrication technology. An important property of the silicon-dioxide layer is its ability to mask against the commonly employed impurities in the fabrication process. The diffusion coefficients of boron and phosphorus in silicon-dioxide are orders of magnitude lower[16] than those in silicon. This characteristic is exploited to make selective diffusions and obtain specific geometries of n or p-type impurity regions.

The thermal coefficient of expansion of silicon-dioxide matches that of silicon[2]. It is a good dielectric and thus overlay metallization contacts are possible. It can be easily etched giving finely etched areas. The aluminum metal which is invariably used for metallic contacts adheres well to silicon-dioxide surface. All these properties added together make silicon-dioxide the best achievable masking layer in the silicon device technology.

Methods of Oxide Formation

Various techniques are employed to grow oxide layers on silicon wafers. The major processes are listed below:

(1) Thermal Oxidation

It can be further subdivided into:

a. Steam Oxidation

b. Oxygen Oxidation

(2) Chemical Deposition

(3) Evaporation

(4) Electrolytic Oxidation

(5) Reactive Sputtering

(6) Spin-on Oxides

Thermal Oxidation, the most widely used single process, refers to the oxides formed by the "thermally activated" reaction of silicon with steam, oxygen or combination of both. The oxidation of silicon by steam or oxygen is possible at room temperatures but due to very low rate of reaction the method is impracticable. So usually temperatures in the 900^{o} to 1300^{o} C range are employed to achieve workable thick layers in reasonable oxidation time.[12]

The reaction involved in the thermal oxidation can be described by a parabolic relationship between time and oxide thickness[12] as follows:

$x^2 = kt$, k is the rate constant

k varies with impurity doping and

crystal orientation. Thermal oxides are relatively dense (2.3 gm. / $cm.^3$) and adhere well to the silicon surface. Thermal Oxidation can be further subdivided according to the process or oxygen bearing media employed

(1) Steam Oxidation:

a. Open tube steam oxidation

b. High pressure steam oxidation

(2) Oxygen Oxidation:

a. Dry oxygen method

b. Wet oxygen method

Each of these methods is described in detail as follows:

(3) Evaporation

(4) Electrolytic Oxidation

(5) Reactive Sputtering

(6) Spin-on Oxides

Thermal Oxidation, the most widely used single process, refers to the oxides formed by the "thermally activated" reaction of silicon with steam, oxygen or combination of both. The oxidation of silicon by steam or oxygen is possible at room temperatures but due to very low rate of reaction the method is impracticable. So usually temperatures in the 900° to 1300° C range are employed to achieve workable thick layers in reasonable oxidation time.[12]

The reaction involved in the thermal oxidation can be described by a parabolic relationship between time and oxide thickness[12] as follows:

$x^2 = kt$, k is the rate constant

k varies with impurity doping and

crystal orientation. Thermal oxides are relatively dense (2.3 gm. / cm.^{3}) and adhere well to the silicon surface. Thermal Oxidation can be further subdivided according to the process or oxygen bearing media employed

(1) Steam Oxidation:

a. Open tube steam oxidation

b. High pressure steam oxidation

(2) Oxygen Oxidation:

a. Dry oxygen method

b. Wet oxygen method

Each of these methods is described in detail as follows:

Open Tube Steam Oxidation

Open tube steam oxidation is the easiest to handle and is most widely employed. Apparatus used is similar to that employed in open tube diffusion process.

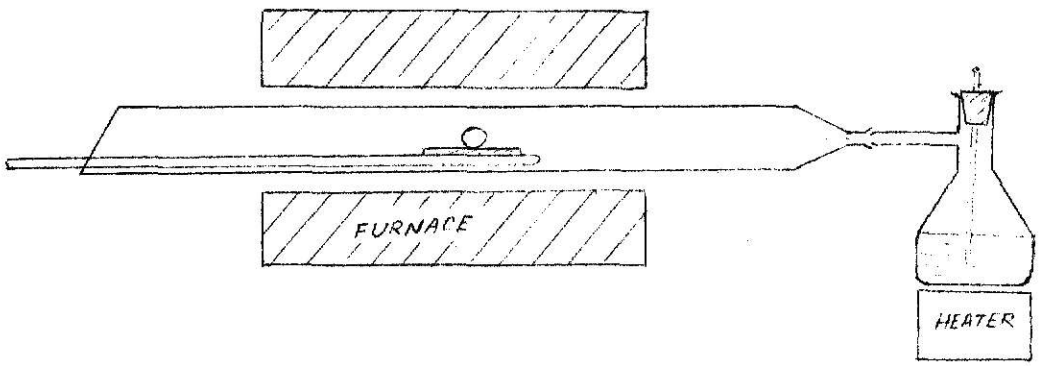


Figure 3.1 Open tube steam oxidation

Highly pure water is heated in a flask and the water vapor is allowed past the silicon wafers placed in a quartz boat assembly. Steam reacts with Si at high temperature to give SiO_2. The water bath temperature does not have any accountable effect on the rate of oxidation[2]. Higher temperature increases the flow past the wafer but does not increase its concentration.

An empirical relation between temperature, time, and oxide thickness is given below:[2]

$$x^2 = t10^{(.861-4030/T)} \qquad (3.1)$$

where x = oxidation thickness in microns

t = oxidation time in minutes

T = temperature in ^{o}k.

The equation is especially valid at high temperatures. At lower

temperatures a square term in 't' must be added to this equation. The relationship between oxide thickness, time and temperature is presented in the form of curves in Figure (3.2).[2]

High Pressure Steam Oxidation

To achieve higher rates of oxide growth, steam at high pressure is used. The rate of oxide growth is directly proportional to the pressure over certain range of temperature, time and pressure. Ligenza, J. R.[9] gives the relation for high pressure steam oxidation as follows:

$$\frac{dx}{dt} = \frac{P}{T} .10^{\,8.28 - \frac{5190}{T}} \quad \text{(Angstroms/minute)} \qquad (3.2)$$

where P is the pressure in atmospheres

and T is the temperature in ^{o}k

The oxidation is carried out in a tight, constant volume metallic enclosure. The silicon wafer along with highly pure water is placed in the enclosure. It is then heated along with its contents and the pressure inside rises. The pressure and temperature of oxidation can be computed using the PVT relation.

This method is cumbersome and usually is not employed. Open tube steam oxidation process is comparatively much simpler and easier to control.

Wet Oxygen Process:

In this process dry oxygen is bubbled through a pure water bath before entering the furnace. The rate of flow of oxygen and the water bath temperature determine the contents of the mixture. The apparatus used is the same as for dry oxygen process.

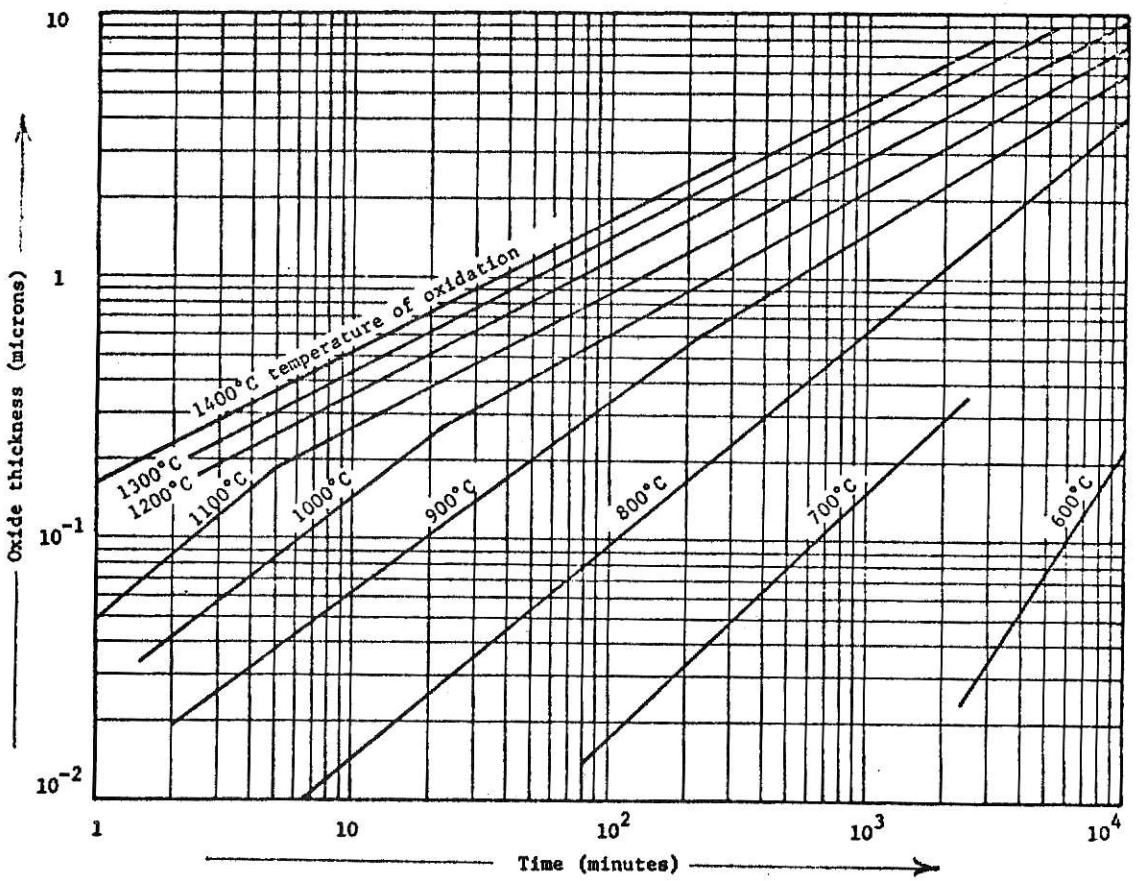


Figure (3-2) Oxidation of silicon in steam at atmospheric pressure[2]

If an inert gas is used in place of dry oxygen the oxidation rate decreases, the decrease being equivalent to the oxidation rate of dry oxygen. This shows that both dry oxygen oxidation and water vapor oxidation proceed independent of each other.[2]

Chemical Methods:

Chemical methods can be employed to grow silicon dioxide films at low temperatures (less than 700°C). Many compounds which contain the Si-O group react irreversibly to form silicon dioxide and hydrocarbons. If the temperature is raised to a higher value than 750°C the hydrocarbons decompose, depositing a layer of carbon on the wafer.

A commonly employed chemical is tetraethylorthosilicate (TEOS) in an oxygen atmosphere[2]. The following reactions take place

$$Si\ (O\ C_2H_5)_4 \xrightarrow[600^oC]{Heat} SiO_2 + 2\ (C_2H_5)_2\ O$$

$$2\ (C_2H_5)_2\ O + 12O_2 \xrightarrow[600^oC]{Heat} 10H_2O + 8CO_2$$

Reactive Sputtering:[2]

In this method an electrical discharge is passed between a cathode made of Si and an anode with the wafer mounted on it, at low pressures. At high voltages and low pressures a glow discharge occurs between the electrodes. Due to the bombardment of the cathode by ionised gas molecules, cathode material is released as free atoms. The oxidation of these released silicon atoms can take place at the cathode, during transportation or at the substrate surface. This depends upon the ratio of oxygen to the inert gas, in the residual gas atmosphere and the gas pressure.

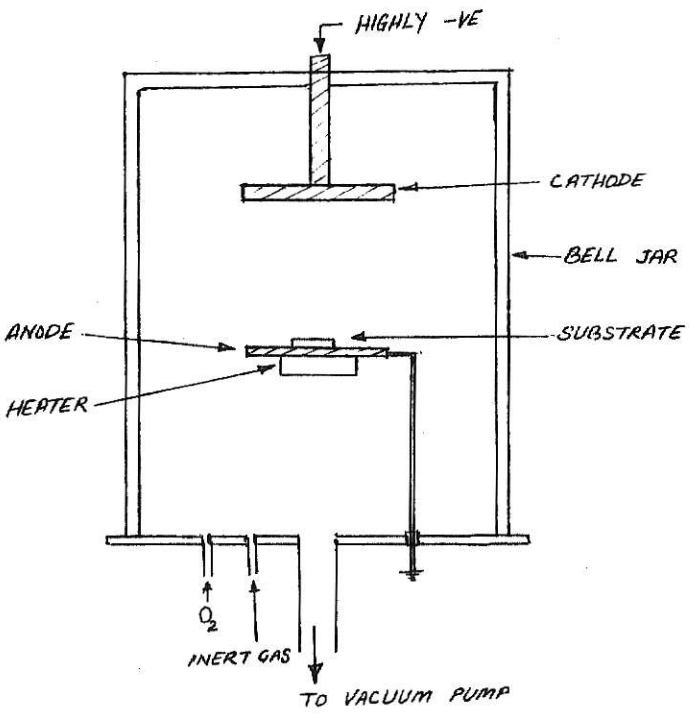


Figure (3-3) Reactive sputtering system

Figure 3.3 shows a reactive sputtering system. Here the silicon to be sputtered and oxidized is mounted on the cathode and the silicon wafer is mounted on the anode. The electrodes are mounted in a vacuum chamber. The pump used is a high speed oil diffusion pump. The usual pressures employed are 0.02 - 0.2 Torr. range. The voltage used is 1000 - 3000 volts with the current density of 1 mA/Cm^2 [2].

The deposition rate, using oxygen alone, is about 75 Å/minute. But a mixture of argon and oxygen gives a much higher deposition rate of 300 Å/minute.

Evaporation Process:[2]

Vacuum deposited oxides are obtained by a variety of methods employing various types of sources (silicon, silicon monoxide, silicon dioxide). In all these methods the source is heated in an oxidizing atmosphere. The silicon wafer is kept at a temperature of about 300°C.

The physical, chemical and electrical properties of the deposited layer depend upon the partial pressure of oxygen, rate of evaporation, system geometry, source composition and the temperature. The films deposited by this method seem to consist of silicon dioxide plus the other oxides of silicon.

Silica Emulsions:

Silica emulsions available from the 'Emulsitone Co.' can be used to get fine oxide layers. The emulsion is applied to the wafer by pouring a few drops and spinning it on a spinner. This layer later heated at a temperature above 200°C, for a few minutes, gives a hard oxide layer. Different film grades are available for different thicknesses. Otherwise the thickness

can be controlled by applying two or more thin layers of oxide. For a specific grade film the thickness depends upon the speed and spinning time.

This process has some advantages over thermal oxidation, for example, there are no built-in charges in the oxide layer that are characteristic of the thermal oxides. Moreover the wafer need not be heated to higher temperatures since that disturbs the previous diffusions.

Some important points about oxidation:

Oxidation is carried out several times during the fabrication process of BJT's. The effect of heating during oxidation upon the previously diffused emitter or base must be taken into consideration. The heating cycle during oxidation must not overdrive the junctions. One way to achieve this aim could be to underdiffuse the junctions so that these get diffused up to the correct depth during the oxidation process. But the best way is to adjust the heating cycles so as not to appreciably affect the existing diffusions. The use of silica emulsions in place of thermal oxides can eliminate this difficulty.

During oxidation about 40% of the oxide thickness is derived from the silicon substrate. This point is important where shallow diffusions are involved. The curves showing the etching rate, minimum oxide thickness requirements and ratios of diffusion coefficients of various impurities in silicon and silicon dioxide are shown in Figure (3.4), (3.5) (3.6) and (3.7) respectively.[16]

Oxide Thickness Measurement Techniques:

Due to samll magnitudes involved, optical interference methods are employed. Other methods such as measurement of increase in the weight of

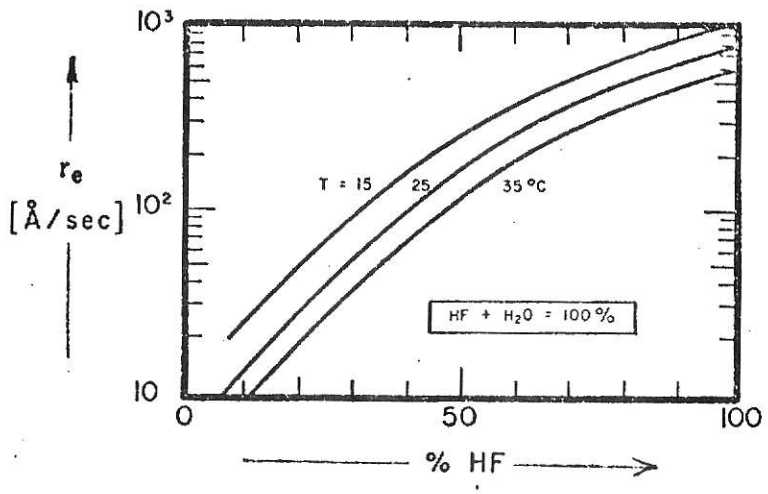


Figure (3-4) Oxide etching rate for different concentrations of HF[16]

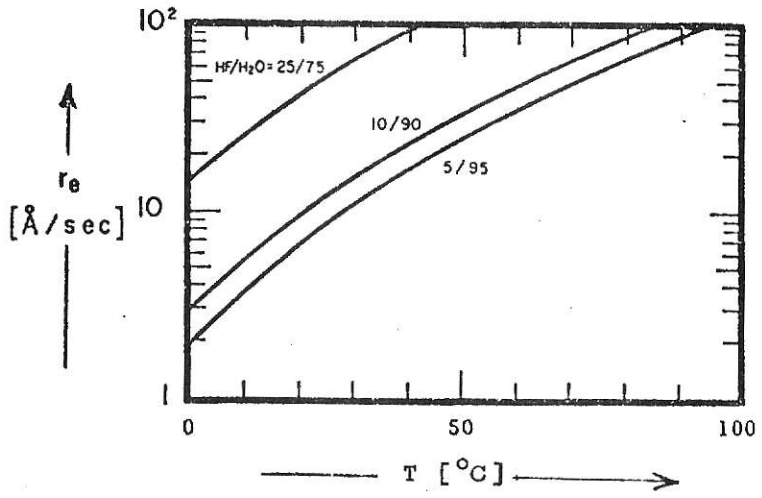


Figure (3-5) Oxide etching rate in HF at different temperatures [16]

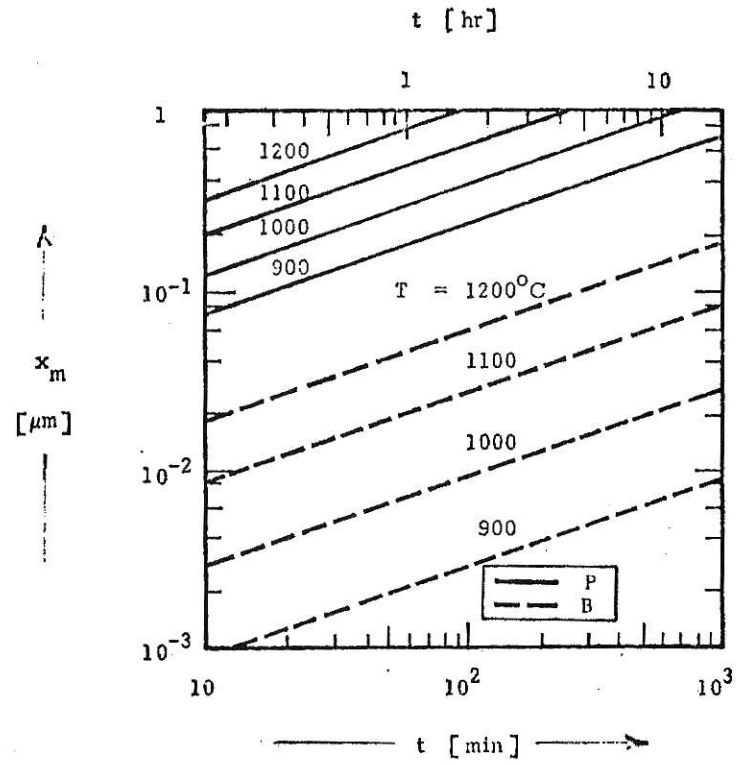


Figure (3-6) Minimum oxide thickness for masking against boron and phosphorus [16]

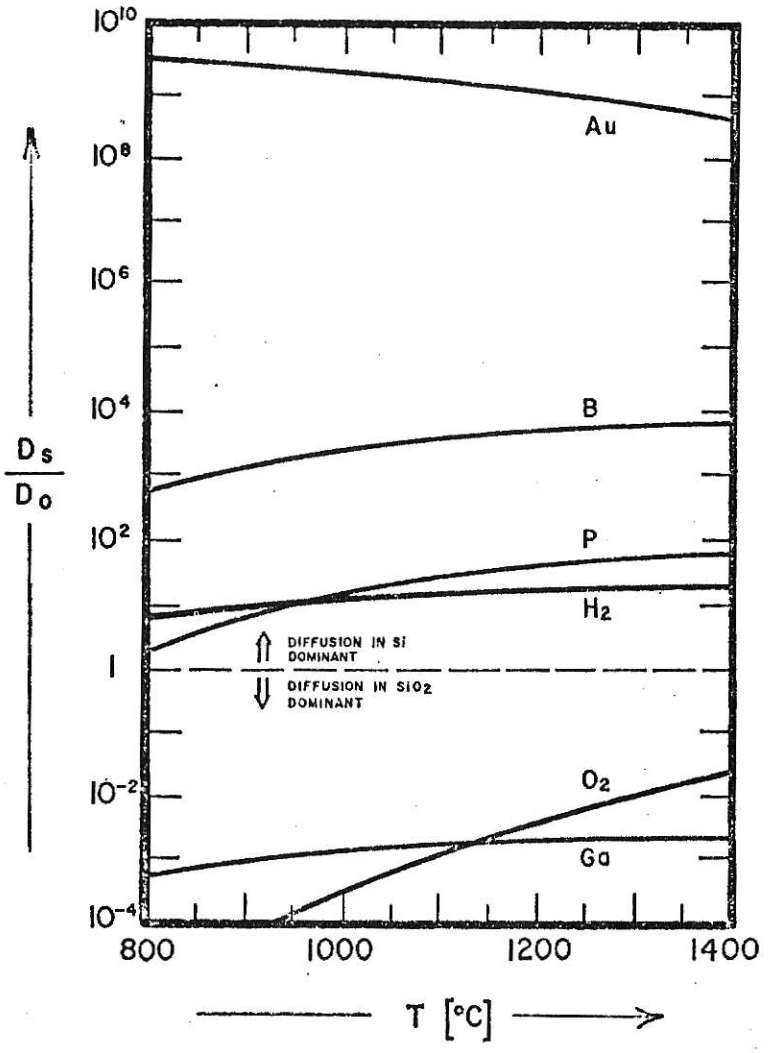


Figure (3-7) Ratio of diffusion coefficients of various impurities in silicon to silicon dioxide[16]

the specimen and measurement of breakdown voltage are useful only for comparison purposes. Because absolute determination requires the accurate knowledge of oxide density and dielectric strength respectively. The following methods are usually employed:

1. Fizeau's fringe method
2. Newton's rings method
3. Color method

Fizeau's Fringe Method:

When two slides are placed so as to form a small wedge between them, and they are viewed under monochromatic light, interference fringes observed are called Fizeau's fringes. The phenomenon is used for measuring oxide film thicknesses.[2]

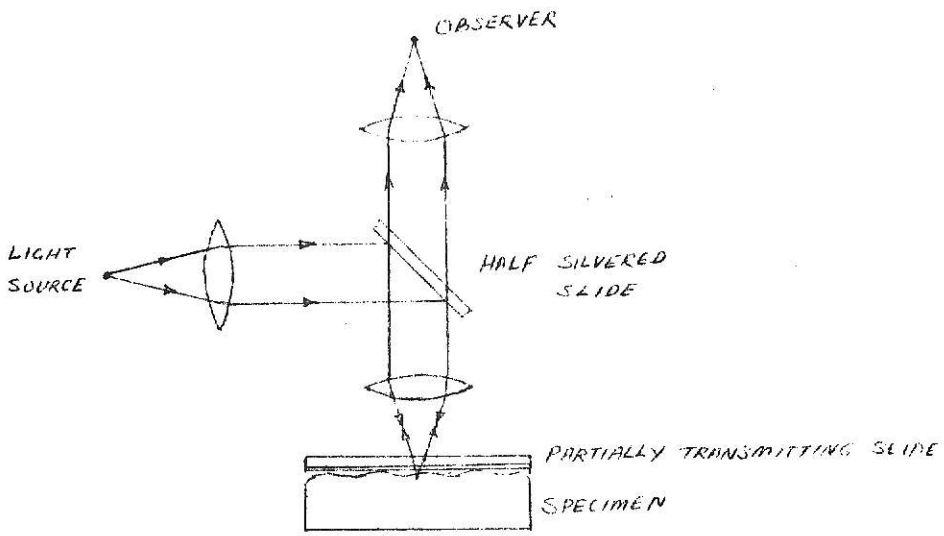


Figure (3-8)

Figure 3-8 shows such an arrangement. Here one slide is replaced by oxidized silicon wafer with the oxide etched from the part of the wafer surface, and the other by a partially transmitting thin film of low absorption.

Part of the oxide layer is etched from the wafer surface using 48% HF and photoresist techniques. This creates a wedge of silicon dioxide. (The 48% HF is used instead of the usual buffered HF etch, which gives a step etch.)[11] The number of fringes in the wedge portion gives the oxide thickness as shown in Figure (3-9).

Oxide film thickness = Number of fringes x $\lambda/2$

where λ = wavelength of the monochromatic light.

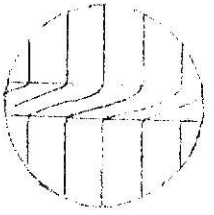


Figure (3-9) (Displacement of fringes due to oxide wedge)

If the partially transmitting plate is removed, the fringes will still be visible in the wedge portion. In this case interference occurs between the beams reflected from the oxide surface and the oxide silicon interface respectively.[2]

Here thickness = $\underline{\text{number of fringes in the wedge}} \times \frac{\lambda}{(\text{refractive index of } SiO_2 \text{ layer})}$

(Refractive index of thermally grown silicon dioxide is 1.45)[5]

This method is simpler than the first one. A metallurgical microscope with monochrometic light source is all that is required. Moreover in the first case there is danger of damaging the specimen and the partially transmitting plate.

Newton's Rings Method:

This method is used for very thin oxide layer measurements.[2] The

partially transmitting plate used in Fizeau's fringe method is replaced by a convex lens. (The radius of curvature of the lens lies between 50 to 100 cms.) The center of the convex lens is positioned above the step formed by etching the silicon dioxide. The etching is carried out with buffered HF to get a vertical step.[11]

The step appears as discontinuity in the interference rings. The portion of the ring formed on the silicon surface is displaced towards the center of the ring pattern as shown in the Figure (3-10).[2]

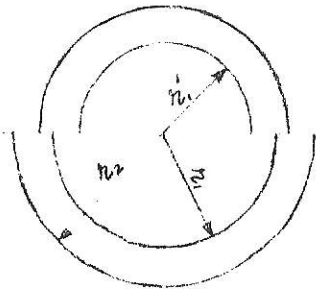


Figure (3-10)

The various radii r_1, r_2 and r_1' are measured. The oxide thickness is given by the formula:[2]

$$S = \frac{r_1^2 - r_1'^2}{r_2^2 - r_1^2} \times \frac{\lambda}{2} \qquad (3.3)$$

where S = oxide thickness in angstroms

λ = wavelength of the monochromatic light used in angstroms

Color Method:

Suppose a light ray is incident on a layer (of refractive index n_1) formed on a substrate (of refractive index n_2). Then if the refractive

index n_1 of the layer lies between the refractive index of air n_o and that of the substrate n_2 ($n_o<n_1<n_2$), the phase change on reflection at each surface is π radians. Further if the path difference between the rays reflected from the two surface is an integer multiple of the wave length, then the reflected light will show a maxima for that particular wave length and the film will seemingly posses that characteristic color.[11]

This phenomenon can be applied to measure the thickness of the oxide films. The refractive index of the silicon dioxide (1.45)[5] lies between the refractive index of the air (1) and that of silicon (3-42).[5] If the layer is observed at an angle of 90°, the thickness is given by the relation.[11]

$$S = \frac{k\lambda}{2n_1}$$

where S = Oxide thickness in angstroms

λ = wave length of the color observed

n_1 = Refractive index of silicon dioxide = 1.45

k is an integer

The value of k can be determined by knowing the approximate value of the oxide thickness. The approximate oxide thickness can be found out, if the temperature and time of oxidation is known. Table (3-1) relates the oxide thickness, for various values of k, to the color of the layers. To eliminate the errors arising out of color observation, National Bureau of Standards color chart can be referred to.[2]

Table (3-1)[5]

Color	Thickness in Angstroms			
	k=1	2	3	4
Gray	100			
Tan	300			
Brown	500			
Blue	800			
Violet	1000	2800	4600	6500
Blue	1500	3000	4900	6800
Green	1800	3300	5200	7200
Yellow	2100	3700	5000	7500
Orange	2200	4000	6000	
Red	2500	4900	6200	

CHAPTER IV

DIFFUSION

Present day commercial methods for fabricating integrated circuits rely heavily on the process of solid state diffusion[4] for forming p-n junctions. It provides the means of introducing controlled amounts of impurities into the silicon lattice to alter its electronic properties. The present day diffusion process is adaptable to batch processing.

Diffusion takes place whenever there is an impurity gradient. The diffusion process is governed by the well known Fick's laws. The first law states that, "particle flux is proportional to the concentration gradient".[15]

$$J = - D \nabla N \qquad (4\text{-}1)$$

where J = particle flux density in atoms per square centimeter per second.

D = diffusion coefficient, in square centimeters per second

For planar diffusion technology the diffusion is assumed to take place along one dimension only.

$$\therefore \quad J = - D \frac{\partial N}{\partial x} \qquad (4\text{-}2)$$

The negative sign indicates that diffusion takes place in the direction of decreasing concentration.

The second law which is derived from the first law is more important. It states that, "the time rate of change of volume concentration of impurities is directly proportional to the second derivative of the volume concentration with respect to distance".[15]

$$\frac{\partial N}{\partial t} = D \frac{\partial^2 N}{\partial x^2} \qquad (4\text{-}3)$$

Solution of this equation, depending upon two different boundary values, gives two different kinds of diffusion distributions.

1. Gaussian distribution
2. Complementary error function distribution

Gaussian distribution: Here the total amount of impurity atoms is finite therefore the surface concentration decreases as the diffusion proceeds. The corresponding boundary values applicable in the solution of equation (4-3) are:

$$N(x,o) = No, \qquad o<x<\Delta x$$

$$N(x,o) = o, \qquad o>\Delta x$$

Applying these boundary conditions we get[5]

$$N(x,t) = \frac{Q}{\sqrt{\pi Dt}} \, \epsilon^{-\frac{x^2}{4Dt}} \qquad (4\text{-}4)$$

where Q = total amount of impurity atoms per square centimeter of the surface, placed prior to diffusion.

D = diffusion coefficient in square centimeters per second.

x = diffusion depth in centimeters.

t = diffusion time in seconds.

This kind of distribution is simulated by exposing the specimen to an infinite impurity source for some definite time in order to deposit a finite amount 'Q' and then removing the impurity source. It is ideal for base diffusion where relatively low value of impurity concentration with high diffusion depth is required.[2]

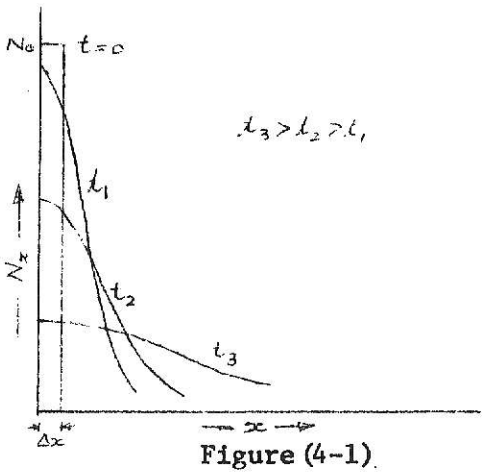


Figure (4-1)

Figure (4-1) shows the Gaussian distribution profile for different diffusion times.

Complementary Error Function Distribution:

This type of distribution results when the specimen is exposed to an impurity source throughout the diffusion process. The corresponding boundary conditions are:

$$N(o,t) = N_o \text{ for } t>o$$

$$N(x,o) = o \text{ for } x>o$$

Applying these conditions to solve equation (4-3) we get[5]

$$N(x,t) = \frac{N_o}{2} \operatorname{erfc} \frac{x}{2\sqrt{Dt}} \qquad (4\text{-}5)$$

where N_o = surface concentration (constant) in atoms per cubic centimeter

D, x and t are the same as defined for equation (4-4)

This type of distribution profile is approximated by either exposing the wafer to the impurity source in gaseous form or coating the wafer with the impurity. The process is used during emitter diffusion where we require shallow diffusion with high surface concentration.[2]

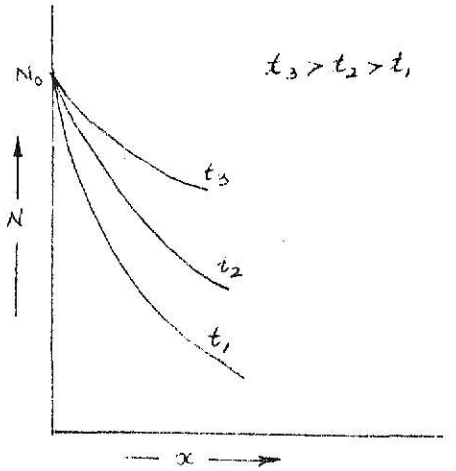


Figure (4-2)

Figure (4-2) shows complementary error function distribution with different diffusion times.

Two-step Impurity Diffusion:

Usually the impurity diffusion is carried out in two steps. The specimen is exposed to the impurity for a certain period to deposit an amount Q of the impurity given by:[5]

$$Q_1 = 2\, N_{o1} \sqrt{\frac{D_1 t_1}{\pi}} \qquad (4\text{-}6)$$

where N_{o1} = the surface concentration in atoms/Cm^3

t_1 = deposition time in seconds.

D_1 = diffusion coefficient in Cm^2/sec. at the deposition temperature.

The impurity source is then removed and the specimen is subjected to a higher temperature. This is called the drive-in phase.

The final impurity concentration at any point at a distance x from the surface is given by[5]

$$N(x,t_2) = \frac{Q_1}{\sqrt{\pi D_2 t_2}} e^{-\frac{x^2}{4D_2 t_2}} \qquad (4\text{-}7)$$

$$\text{or} \quad N(x,t_2) = \frac{2\,No_1}{\pi} \sqrt{\frac{D_1 t_1}{D_2 t_2}}\, e^{-\frac{x^2}{4D_2 t_2}} \qquad (4\text{-}8)$$

To account for the heating cycles during later diffusions and oxidations an effective "Dt" product should be used in the above equation.[5]

The two step diffusion has the following advantages:[2]

1. The impurity is present only for short time-during low deposition temperature therefore it makes oxide masking more effective.
2. Gaussian distribution can be simulated by this process.
3. Corrections are possible in second step for errors in the first step.

Diffusion Variables:

The diffusion process involves a number of variables. In order to achieve the desired results it is important to know the effect of each one of these. The important variables are:

1. Diffusion coefficient
2. Diffusion time
3. Solid solubilities of the impurities employed.

1. The diffusion coefficient varies with the nature of the diffusant and the diffusion temperature. Surface concentration also seems to have some effect on the value of diffusion coefficient. The following equation relates the diffusion coefficient to the temperature.[15]

$$D = D_o\, e^{-\Delta E/KT}$$

where D_o = Apparent value of diffusion coefficient at infinite temperature. It is constant.

ΔE = Activation energy. It is constant for certain temperature ranges.

T = Diffusion temperature in degrees Kelvin

The values of D_o and ΔE for various impurities for certain temperature ranges are given in the Table (4-1).[2]

Table (4-1)

Diffusion Coefficients in Silicon

Element	D_o (cm^2/Sec.)	ΔE (ev)	Temperature Range (^oC)
Aluminum	4.8	3.36	1050-1380
Antimony	12.9	3.98	1190-1398
Arsenic	68.6	4.23	1100-1350
Boron	10.5	3.69	950-1275
Gold	1.1×10^{-3}	1.12	800-1200
Phosphorus	10.5	3.69	950-1235
Bismuth	10.30	4.64	1220-1380
Gallium	3.6	3.51	1105-1360
Indian	16.5	3.90	1105-1360

Curves relating the diffusion coefficient to diffusion temperature for commonly employed impurities are given in Figure (4-3).[4] From the curves we see that the temperature critically effects the diffusion coefficient. Therefore it should be controlled to within $\pm 0.5^oC$ of the predetermined value.

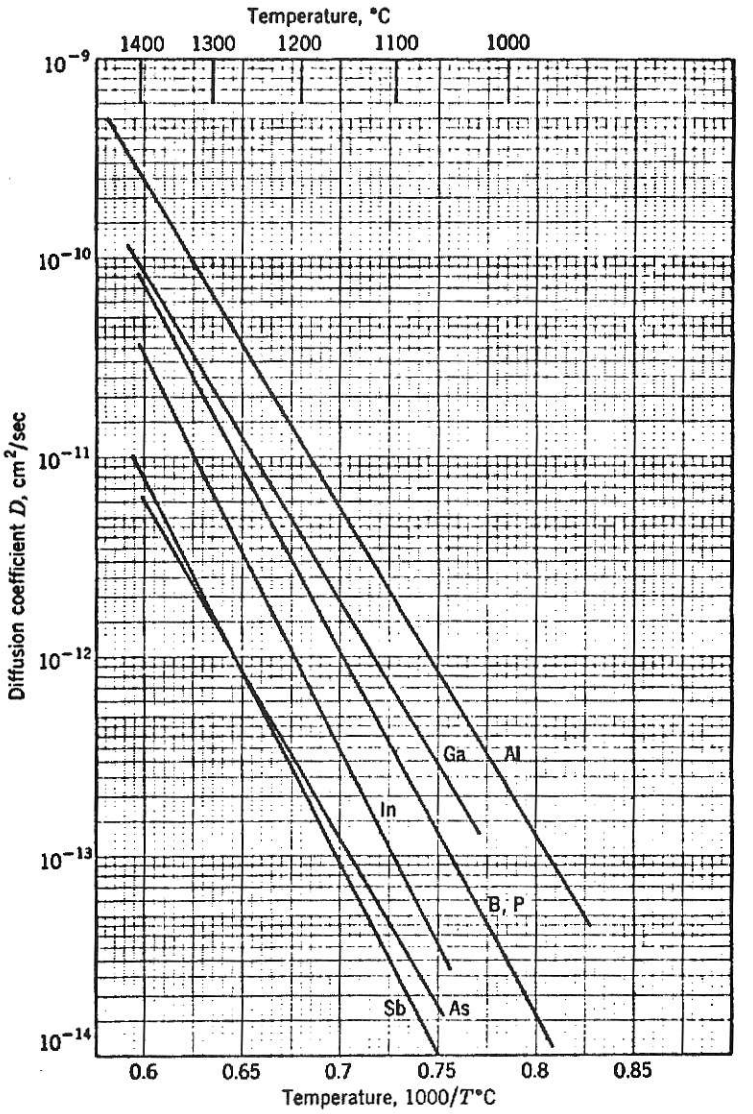


Figure (4-3) Diffusion coefficient of various impurities in silicon[4]

2. Diffusion time - This is an easy variable to control. The changes in diffusion time has the same effect on diffusion depth as the diffusion coefficient since it appears in the product 'Dt'. The diffusion temperatures and hence the diffusion coefficients should be such that reasonable times are required for the required diffusion depth. For short diffusion times the time taken by the specimen to reach the specific temperature must be taken into account.[5]

3. Solid solubility - This is defined as the maximum impurity concentration that can be obtained with a certain impurity. It also depends upon the diffusion temperature. Maximum solid solubilities for commonly employed impurities in silicon technology are given in Table (4-2).[2]

Table (4-2)

Impurity	Maximum Solid Solubility ($atoms/cm^3$)	Temperature for Maximum Solid Solubility (^oC)
Phosphorus	$1.3x10^{21}$	1150
Boron	$5x10^{20}$	1200
Arsenic	$2x10^{21}$	1150
Antimony	$6x10^{19}$	1300
Aluminum	10^{19}-10^{20}	1150
Galium	$4x10^{19}$	1250
Indium	10^{19}	1300
Gold	10^{17}	1300

Figure (4-4) gives the curves relating solid solubility to the temperature for various impurities.[13]

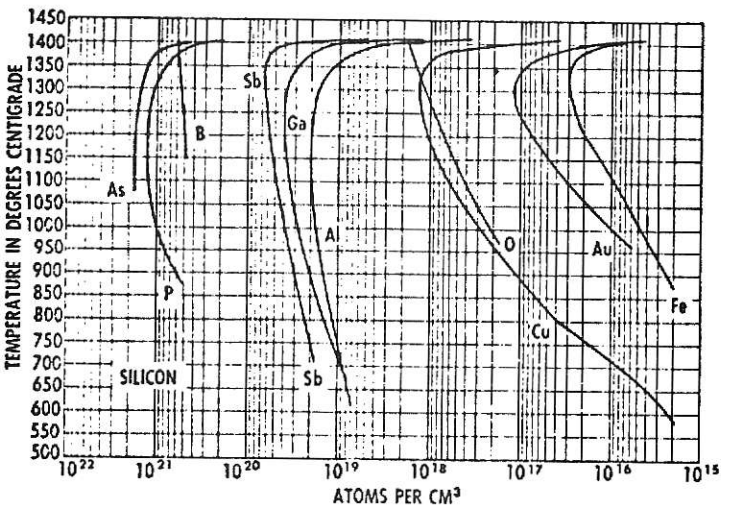


Figure (4-4) Solid solubilities of various impurities in silicon[13]

Diffusion Systems

Various types of diffusion systems are employed as follows:

1. Vacuum diffusion
2. Solid tube diffusion
3. Open tube diffusion

These can be further subdivided depending upon the form of impurity employed.

(i) liquid

(ii) gaseous

(iii) solid

Open tube system is the one mostly used. The process derives its name from the fact that one end of the tube is open to atmosphere.[2] Figure (4-5) shows an open tube tube diffusion system.

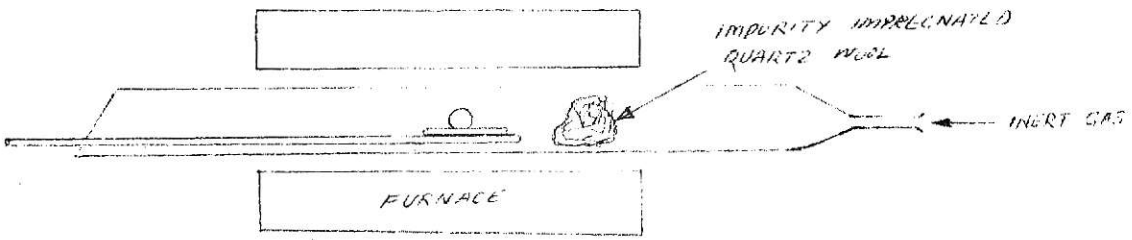


Figure (4.5)

The wafer is placed on a quartz rod and boat assembly. The impurity can be applied by any of the following ways:

(i) Depositing a thin layer of impurity silica-emulsion

(ii) Passing the impurity over the wafer in the gaseous form

(iii) Impregnating the impurity in quartz wool and heating it. An inert gas is used to transport the impurity from the quartz wool to the silicon wafer.[2]

Diffused Layer Analysis:

The diffusion profile can be established by determining the impurity concentration at various points along the depth. This is achieved by removing successive layers using chemical etching methods and determining surface concentration after every step.[2]

Second method employed is to diffuse the radio-active isotope of the impurity along with the impurity itself. Integrated radiation intensity is measured after successively etching thin layers.[2] This is based on the assumption that radio-active isotopes distribute themselves uniformly among the normal impurity atoms and have the same diffusion coefficient. There are some disadvantages to this method. First being that it gives the entire impurity content and not the electronically active part. Secondly it has to be conducted separately.

To determine the impurity concentration profile it is important to know the sheet resistance and the junction depth.

Measurement of Sheet Resistance:

It is carried out using the four point probe. The apparatus consists of four equally spaced sharp needle electrodes. A constant current is passed through the two outer electrodes and the corresponding voltage drop is measured across the two inner electrodes.[15]

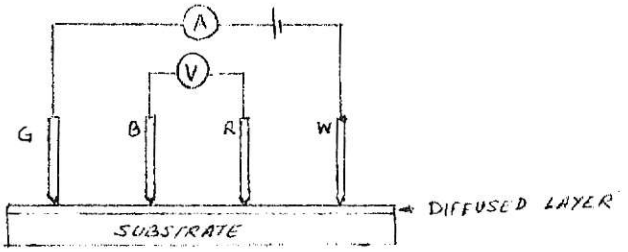


Figure (4-6)

(Four point probe)

Figure 4-6 shows a four point probe arrangement. G and W are the current electrodes, B and R are the voltage electrodes. The sheet resistance R_s is given by:

$$R_s = C \frac{V}{I} \text{ ohms/square}$$

where C is a constant and depends upon the orientation of the probe over the wafer. Average value of C is about 4.5.[15] A calibration table giving values of C for various orientations of the probe over the wafer are available with the equipment.

Measurement of Junction Depth:

Analytically, a p-n junction is located at the surface where the diffused impurity concentration is equal to the background impurity. Experimentally the junction depth is measured using interference methods. A part of the wafer, 1/8" x 1/4" in size is mounted one end of a lapping piston. It is lapped at an angle, usually 1^o to 5^o, to obtain a wedge shaped surface. To delineate the p-n junction it is stained by using a solution consisting of 100 ml of 48% HF and five drops of HNO_3.[11] This makes p-type region appear darker than the n-type region.

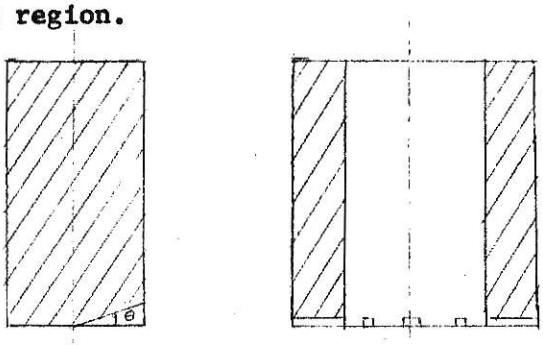


Figure (4-7)

Wafer grinding piston and cylinder

The specimen is fixed on the end of the piston using wax. The piston

is then put in the cylinder to keep its axis vertical to the lapping plate. Lapping is carried out using 1000-grit silicon carbide with liberal amounts of detergent and water. Polishing is carried out on a plate covered with plastic using 1000-grit silicon carbide. The assembly. The p-n junction is then delineated by applying a few drops of the solution described previously.

The speciman is then placed under the microscope with the interference objective in place. The position of the wafer is adjusted until the fringes on the top of the wafer are perpendicular to the wedge shaped surface.[2] The number of fringes in the upper region (p or n) gives the junction depth as follows:[11]

$$\text{Junction depth} = \text{number of fringes} \times \frac{\lambda}{2}$$

where λ = wave length of the monochromatic light

CHAPTER V

Photo-masks and Photo Resist Techniques:

Photoetching is the fundamental tool in the fabrication of semiconductor devices. It consists of the following processes:

1. Artwork or masterdrawing.
2. First reduction.
3. Final reduction and step & repeat.
4. Photo resist techniques.

Artwork or Masterdrawing:

The original artwork is generally cut on a peelable mylar, using a coordinatograph which can locate a point within its working area to one mil. The original artwork is drawn 100 to 1000 times the actual size to minimize the errors arising out of drawing work. The size of the master drawing depends upon the total size of the final circuit, smallest line width in the final size, the precision on which the original artwork can be drawn etc. The original artwork can be either negative or positive depending upon the total number of reduction steps, convenience in drawing the artwork and whether the final image should be positive or negative.

Peeling of the lacquer from between regions less than 15 mils wide is tedious. Therefore the reduction ratios should be adjusted, wherever possible so that lines less than 15 mils in width are not required in the stencil.[15]

First Reduction:

The artwork is photo-reduced to a size suitable for stepping. Reduction

ratios of 5x to 20x are usually employed. Reduction ratios larger than 20x require distances between the lens and object that are not practicable in ordinary rooms. Reduction ratio is given by the formula:[12]

$$R = \frac{U-f}{f}$$

where R = reduction ratio

U = object distance

f = focal length of the lens

The light source should adequately and uniformly illuminate the copy material. A provision should be made for instantaneous turning on and off the light source for timed exposures. Special kinds of lamps usually xenon, have been developed for this purpose.[12] This avoids the need for using a lens shutter since it may introduce vibrations. The artwork can be illuminated either by reflected or transmitted light. Transmitted light gives relatively better contrast with the artwork on mylar sheet the contrast with transmitted light between opaque and transparent regions is 1000:1.

Step & Repeat and Final Reduction:

In the fabrication of transistors and silicon integrated circuits, where a batch process is used, large arrays of identical masks are required. In order to avoid the enormous labor needed to draw the whole array on the original artwork, a step and repeat camera is used. The process derived its name from the fact that the first reduction is repeatedly contact printed at accurate spacings.

Two processes can be used to reduce the whole array of artwork to its final size. In the first process, the first reduction is contact printed repeatedly at accurate spacings. The final mask is obtained by photo-reducing this array of masks to the required size.

In the second process the stepping and final reduction is carried out simultaneously. Special systems like an array of apertures or lenses are employed to eliminate the step and repeat operation. The multiple pin-hole or lens camera introduces no distortion if the image and the object are normal to and centered on the optic axis. For an array of lenses only one lens can be centered on the object. The images for the off-center lenses will exhibit trapezoidal distortion.[12] But the distortion will be the same for every mask.

Rudge et. al.[12] have described a mask generating system using an array of simple convex-plano lenses. An array of such lenses is termed as fly's eye lens.

Photo-resist Techniques:

Photo-resists are light-sensitive organic compounds[5]. The use of photo-resist in conjunction with the photomasks provides the necessary patterns required on the silicon wafer. There are two kinds of photo-resists:

1. Negative acting.
2. Positive acting.

Negative acting photo-resists give a negative impression of the photomask. Here the regions that are not exposed to light are dissolved during the developing process. Kodak photo-resists are the negative acting type.

Positive acting type:

In this case the regions of the photo-resist exposed to light are dissolved away during the developing process. Shipley resists are of this kind.

The fundamental requirements for a good photo-resist are that it should adhere well to the silicon dioxide or aluminum surface, it must give a sharp resolution, it should be dimensionally stable with respect to temperature, it should be easy to remove and should not be affected by metal and oxide etches. Following operations are carried out during the photo-resist process.

1. Coating - A few drops of the photo-resist are placed on the wafer and it is spun at 2000-5000 r.p.m. for a certain time to get a thin uniform layer. Usually two thin layers are used in place of a single thick layer to avoid the possibility of pin-holes.[5]
2. Mask Alignment and Exposing - The photomask is aligned over existing pattern on the photo-resist coated wafer. The system is then exposed to intense ultraviolet light for a definite time.
3. Developing - The wafer is developed in the appropriate developer. This washes away the regions exposed to light or the portions not exposed to light depending upon whether the photo-resist is a positive acting or negative acting type respectively.
4. Etching - The wafer is then placed in the appropriate etching solution to etch away the oxide or the metal as the case may be.
5. Stripping the photo-resist - The remaining photo-resist is then etched away. Shipley resist are easily dissolved in acetone. Kodak photo-resists may have to be mechanically agitated after putting in either hot concentrated H_2SO_4 or trichloroethylene.[5]

CHAPTER VI

Metal contacts

Aluminum is the most widely used metal for contacts in the silicon integrated circuit technology. It is preferred due to the following reasons:[5]

1. It can be readily deposited.
2. It forms a eutectic with silicon at 577°C. Thus on heating at about 550°C for five minutes, it alloys to the silicon to form a strong bond.
3. At the alloying temperature it reacts with the oxide surface on which it is evaporated and bonds strongly to it.
4. It is a relatively good conductor. A typical aluminum interconnection 5000 Å thick has a sheet resistivity of about 0.025 ohms per square.
5. It welds readily to gold attachment leads by thermocompression bonding. In addition aluminum leads can be easily welded by ultrasonic bonding.
6. It can penetrate through the thin (up to 100 angstroms thick) oxide layers that develop immediately on silicon regions when exposed to atmosphere.

The process is carried out as follows:

1. An oxide film is grown over the entire surface of the wafer.
2. Oxide is selectively etched from the regions where contacts are to be made.
3. A thin layer of aluminum is deposited on the entire surface. Aluminum makes contacts to the silicon at regions from where oxide

has been etched, everywhere else, it is insulated from the silicon by a silicon dioxide layer.

4. The wafer is heated at 550°C for five minutes to alloy the aluminum to the silicon.[5]

5. The aluminum is selectively etched from the undesired portions to give a final contact pattern. A suitable etch for aluminum is:[11]

75 ml of 85% H_3PO_4

15 ml of 99% acetic acid

5 ml of 3% H_2O_2

3 ml of 70% HNO_3

The above solution heated to 120°F acts very well. Another etch[5] that can be used is described as follows:

phosphoric acid: nitric acid: water: 70:3:27 (by volume)

The aluminum layer can be deposited on the wafer by the following methods:

1. Chemical deposition
2. Sputtering
3. Vacuum evaporation

A thin film of aluminum may be deposited on the wafer by passing organic compounds of aluminum over it. The compound decomposes to give a thin layer of aluminum at the wafer surface.[11]

Sputtering is the process involving bombardment of an aluminum target by energetic positive ions of heavy mass.[12] Usually an inert gas like argon is used. The bombarding ions knock out arrays of source material from its surface. If a silicon wafer is placed in the path, a fine, thin coating of aluminum is obtained on the surface.

The most widely used process for this purpose however is the vacuum evaporation. In this process aluminum is heated in a high vacuum. This results in emission of rays of aluminum particles in all directions. If a silicon wafer is placed in the path of these rays a fine layer of aluminum gets deposited on it.

To prevent the collision of the aluminum atoms with gas molecules, the pressure in the vacuum chamber should be very low. Pressures in the range of 10^{-5} mm of Hg or less yield satisfactory results. The mean free path (average distance between collisions) of aluminum atoms increases with the vacuum. Percentage of aluminum atoms reaching the wafer surface without suffering a collision is given by:[11]

$$\phi(x) = \frac{-x}{\varepsilon L} \qquad (5\text{-}1)$$

where x is the distance between the source and wafer, L is the mean free path of aluminum atoms.

L is given by:

$$L = \frac{1}{\sqrt{2}\pi n d^2} \qquad (5\text{-}2)$$

n: is the number of gas molecules per c.c. in the vacuum chamber

d: is the diameter of the aluminum molecules in cms

Aluminum is evaporated by placing a small piece of wire on a tungsten spiral and passing a current through it. During early phases of evaporation, the wafer should be isolated using a shutter. The melting point of aluminum is 600°C.[5] A slow evaporation rate gives a more uniform coating. Heating of the wafer at 300°F during evaporation gives an aluminum layer which adheres well. Aluminum layers deposited on cold wafers get undercut during its selective etching to form connecting lead patterns.

The aluminum layer should be sufficiently thick to give low resistance connecting patterns, but it should not be so thick that the alignment of the connecting lead pattern over it becomes impossible. Layers in the thickness range 0.5 micron to 1 micron serve the purpose very well.[2]

Aluminum acts as an acceptor impurity for silicon. Therefore if alloyed to low impurity n-type silicon region, it forms a p^+-n junction instead of a simple ohmic contact. To avoid this problem a n^+ layer is diffused in the lightly doped n-regions (usually collects regions) under the areas where ohmic contacts are to be made. This diffusion is carried out simultaneously with the emitter diffusion.

CHAPTER VII

DETAILED FABRICATION PROCEDURE

Layout and Preparation of Artwork:

The original artwork was cut on "rubylith" sheet 100 times the actual size. The size of the areas for base, emitter, collector and isolation diffusions was determined taking into account the limitations of resolution of the available photo reduction equipment and alignment tolerance. The actual size of one complete transistor was calculated to be 400 x 360 μm. The various masks for isolation, base, emitter, ohmic contacts and final metal contact pattern are shown in the figures (7-1) through (7-6).

Since the step and repeat facility was not available in the laboratory the whole array consisting of about 240 patterns was cut on the original artwork. To scribe the artwork on the "rubylith" sheet the "micromarker" or the coordinatograph was employed. Alignment marks were drawn on each sheet in order to facilitate the alignment of the photo-masks above the existing patterns on the wafer.

First Reduction:

The original artwork was photoreduced to 1/10th of its size using a high-contrast low-resolution Kodak photographic artfilm. The data for exposure and developing are given below:

exposure time	:	4 seconds
aperture	:	f 16
developing time	:	3 minutes
stop bath (water)	:	10 seconds
fixing	:	5 minutes

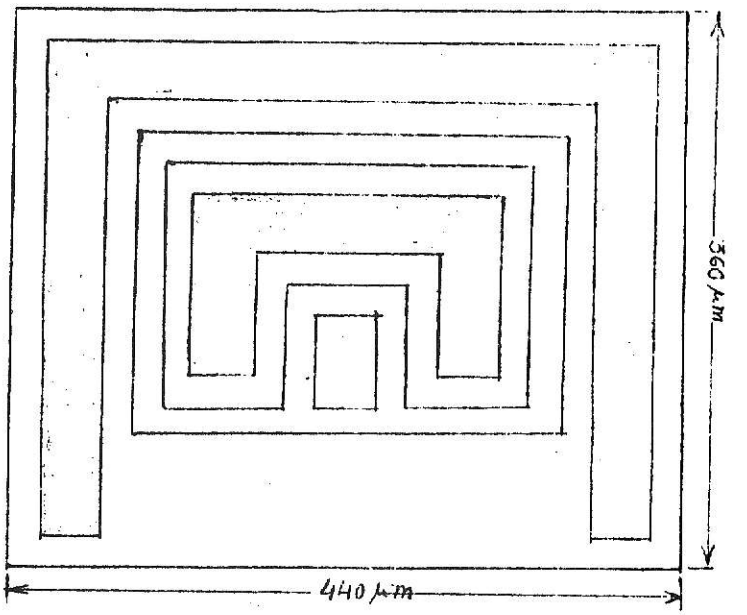


Figure (7-1) Complete layout with contacts

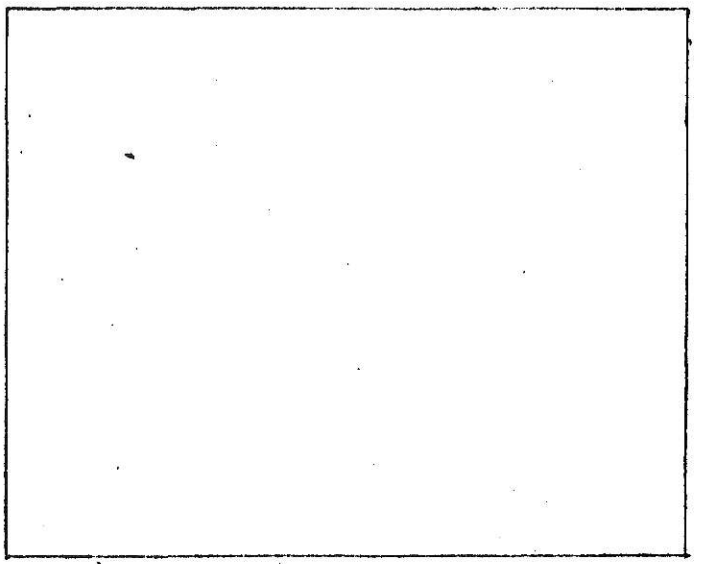


Figure (7-2) Isolation mask

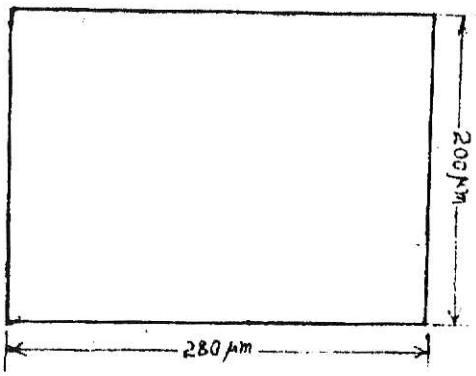


Figure (7-3) Base diffusion mask

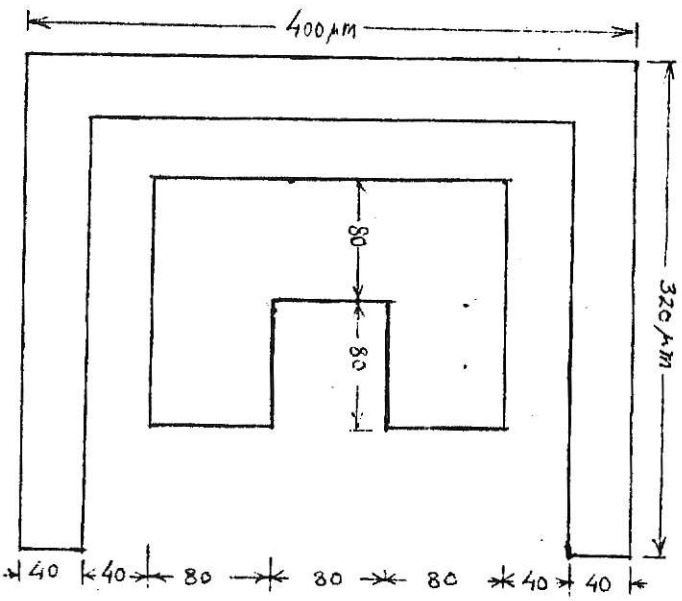


Figure (7-4) Emitter diffusion mask

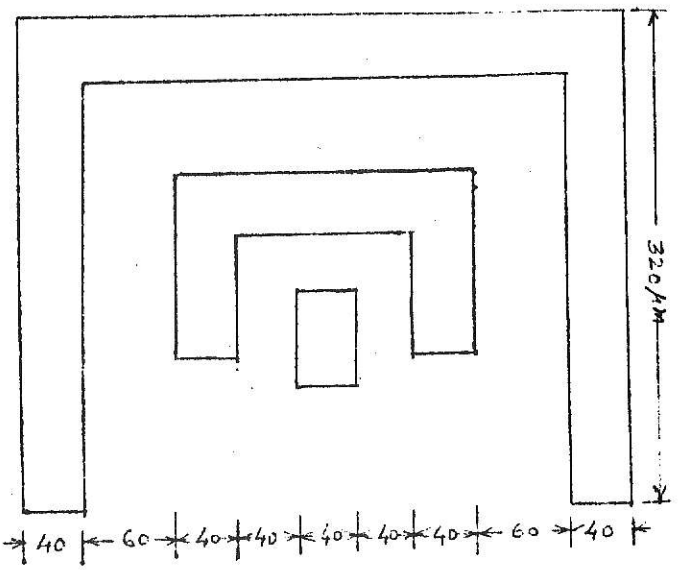


Figure (7-5) Pre-ohmic contact mask

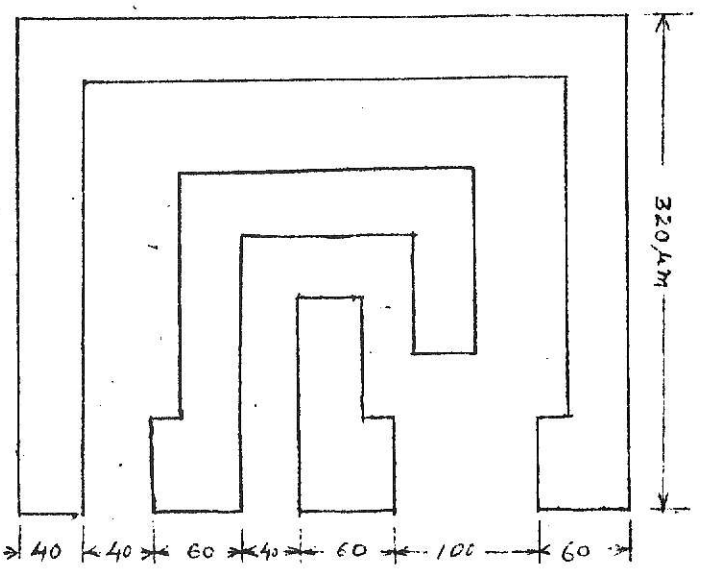


Figure (7-6) Final metal contact mask

The results obtained were quite good. The pattern obtained after first reduction are shown in Figure (7.7).

Final Reduction:

The final reduction was carried out using high-resolution glass plates. This reduction was 10:1. It was carried out using the following data:

exposure time	:	1 minute
aperture	:	f:16
developing time	:	3 minutes (using D19, developer)
stop bath	:	10 seconds
fixer	:	5 minutes

The final masks were thus obtained on glass plates 2½" x 2½".

The starting material was n^+ silicon substrate with n type epitaxial layer on it. The wafers were obtained from the Western Electric. The data for the wafers are:

Substrate : 006--0.0/3 Ω Cm, n type

Epitaxial layer: 5 Ω Cm, 12 μm thick, n type

The wafers were already polished when received.

Cleaning the wafers before oxidation:

The following procedure was used to clean the wafers:

1. Place wafer in hot (85^oC) trichloroethylene for five minutes.
2. Agitate ultrasonically, while still in trichloroethylene for five minutes using ultrasonic vibrator.
3. Rinse the wafer in methanol for two minutes to remove the trichloroethylene.
4. Rinse in double distilled deionized water for five minutes.

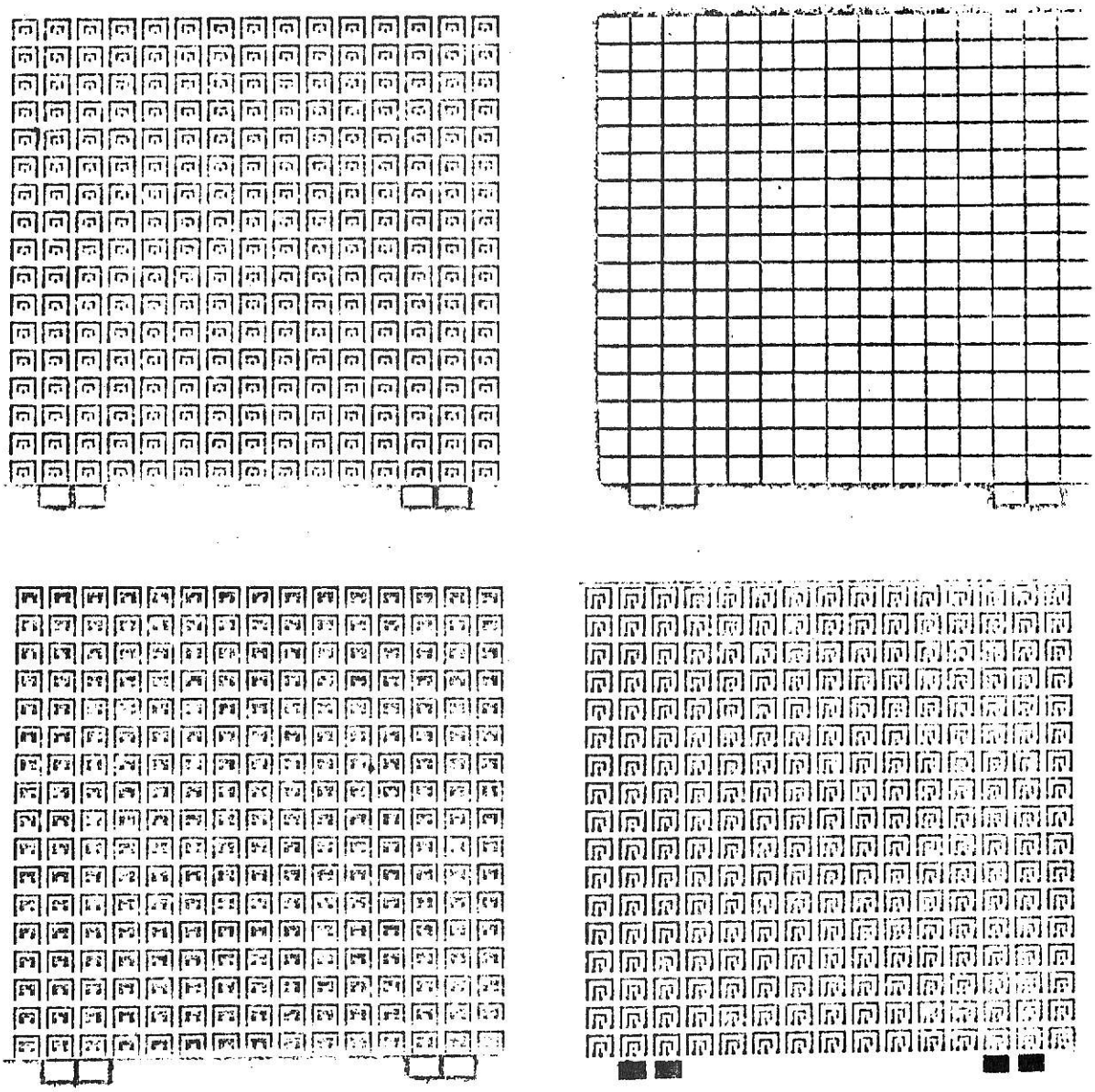


Figure (7-7) Various patterns after first reduction

5. Put in hot sulphuric acid for five minutes.
6. Agitate ultrasonically while still in H_2SO_4.
7. Rinse in double distilled deionized water for two minutes.
8. Place in hot mixed acid I (50: 50:: H_2SO_4: HNO_3) for ten minutes.
9. Vibrate ultrasonically while still in mixed acid for five minutes.
10. Rinse in double distilled deionized water for two minutes.
11. Place in hot mixed acid II (50: 50:: HNO_3:H_2SO_4) for ten minutes.
12. Agitate ultrasonically while still in mixed acid for five minutes.
13. Rinse in doubled distilled deionized water for four minutes.
14. Put in buffered HF for sixty seconds.
15. Rinse in double distilled deionized water for five minutes.
16. Dry under the infra red heat lamp and inspect visually under the microscope for any dirt particles or defects.

Oxidation:

Thermal oxides were grown using the oxidation furnace available in the laboratory, where the open tube steam oxidation process was used. The furnace available was not a precision one. The temperatures could be controlled only to $\pm 10^{o}C$ of the required value. However, it was found quite satisfactory for the purpose since the thickness of the oxide was not a critical factor.

Immediately after cleaning, the wafer was placed in the furnace. A quartz platform mounted on a quartz rod was used to position the wafer inside the furnace tube. The rod was pushed slowly (at the rate of about one inch per second) inside. The oxidation was carried out at $1050^{o}C$ for one hour to grow about 5000 Å thick oxide layer.

Base Diffusion:

Application of Photo-resist and Photo Etching:

The wafer was cleaned and dried under the infra red heat lamp for ten minutes to drive away the moisture. This is very important, otherwise the photo-resist will not stick well to the wafer surface. Shipley AZ-1350 positive acting photo-resist was employed. The photo resist was filtered using 'mipore filter' NR WP 02500 of the Millipore Corporation. Fine uniform layers of photo-resist were obtained using a spinner. Model EC 101 Spinner Headway Research Inc. was available in the laboratory. The following procedure was used to apply the photo-resist:

1. Applied six drops of the AZ1350 photo-resist and spun the wafer at 3000 rpm for 30 seconds.
2. Again applied six drops of the photo-resist and spun the wafer at 3000 rpm for 30 seconds.
3. Put the wafer under the infra red heat lamp for 15 minutes to dry the photo-resist. The lamp was placed at 8" from the wafer (at about 50°C).

Exposed the wafer under the sun lamp after aligning the base mask over it. The exposure time was two minutes. The wafer was mounted on the central platform of the jig for alignment and exposure. The wafer was held to the platform firmly by vacuum. The mask was held on the outer platform. The relative movement between the two in the X, Y and angular direction, was possible. After aligning the mask with the help of the microscope the central platform was raised to bring the wafer into contact with the mask. The jig was then pushed under the lamp assembly which automatically opens the shutter for the required exposure time.

Developing:

After exposing the wafer under the ultra violet light it was developed using Shipley AZ-1350 developer diluted 1:1 with distilled water. The developing time was 20 seconds. Immediately after that the wafer was rinsed with doubled distilled deionized water. The portions of the photo-resist exposed to ultra violet light were etched away during the developing process.

Post baking:

The wafer was then placed under the infra red heat lamp at a distance of 8" for twenty minutes. This step helps in getting a photo-resist layer that adheres well to the wafer surface. Otherwise undercutting of the oxide results during etching.

Oxide etching:

The wafer was placed in buffered HF for 6 minutes. It was agitated in the ultrasonic vibrator during the process. (Etching rate of the thermal oxide in buffered HF is about 1000 Å/minute. However the first 600 Å layer which was exposed to the impurity, etches within 5 seconds.[16]) This etches windows in the oxide for base diffusion. The wafer was rinsed with deionized water and dried under the infra red lamp. The photo-resist layer was stripped off by putting the wafer in acetone for ten seconds. It was again rinsed with deionized water and dried under the heat lamp for 10 minutes.

Application of Borosilica Film:

Borosilica film (Emulsitone Co.) was used for p-type impurity doping. It was found difficult to filter it using 'mipore' filter so the film was applied without filtering as follows:

1. Covered 3/4th of the wafer surface area with borosilica film and spun it at 3000 rpm for 30 seconds.
2. Repeated the above step to place a second layer over the top of the first layer. However it was found advisable to dry the first layer before applying the second layer. This results in very fine uniform layer. Otherwise the layer obtained is not uniform.

Base Diffusion:

Base diffusion was carried out in two steps:

1. Pre-deposition
2. Drive-in

Boron pre-depostion:

The wafer was heated in the furnace for 30 minutes at 875°C to diffuse a monolayer of the p-type impurity into base area. This process amounts to complementary error function type distribution (diffusion from an infinite source). A fixed number of impurity atoms 'Q' diffuse per unit area of the wafer surface. This is given by the following relation:

$$Q = 2N_o \sqrt{\frac{D_1 t_1}{\pi}} \qquad (7\text{-}1)$$

where N_o is the concentration of the impurity atoms in the film per cubic centimeter

D_1 is the diffusion coefficient of the boron at the predeposition temperature in Cm^2/Sec

t_1 is the diffusion time in seconds.

The concentration N_o was not definitely known but it was assumed to be 3×10^{20} atoms/Cm^3

$$\text{at } 875^{\circ}\text{ C},\ D_1 = 10^{-15}\ \text{Cm}^2/\text{Sec}$$

$$t_1 = 30\text{x}60 = 1800\ \text{Sec}$$

$$\therefore \quad Q = 2\text{x}2\text{x}10^{20}\ \left(\frac{10^{-15}\text{x}1.8\text{x}10^{-3}}{\pi}\right)^{\frac{1}{2}}$$

$$= 4\text{x}10^{20} \text{ x } 7.59 \text{ x } 10^{-7}$$

$$= 3.04 \text{ x } 10^{14}\ \text{atoms/Cm}^2 \qquad (7\text{-}2)$$

Base Drive-in Cycle:

The wafer was then boiled in deionized water for five minutes. Without this step, sometimes it was found difficult to etch the impurity in buffered HF. The wafer was then placed in buffered HF for two minutes to etch the impurity layer. It was later rinsed in deionized water and dried under the infra red heat lamp.

The wafer was placed in the diffusion furnace for six hours at 1110°C. This amounted to Gaussian distribution (diffusion from a finite impurity source). It was then placed in the oxidation furnace for one hour at 1050°C to get 5000 Å thick oxide layer over the base regions. Calculations for the junction depth are given below:

The concentration of impurity atoms at any point at a distance x from the surface is given by:

$$N_x = \sqrt{\frac{Q}{\pi Dt}}\ e^{-\frac{x^2}{4Dt}} \qquad (7\text{-}3)$$

Here Dt is the effective combined 'Dt' product for drive-in and oxidation cycles.

$$Dt(\text{effective}) = D_2t_2 + D_3t_3$$

$$D_2 = 2.5 \text{ x } 10^{-13}\ \{\text{at } 1100^{\circ}\text{C during drive-in}\}$$

$$t_2 = 6 \times 60 \times 60 = 2.16 \times 10^4 \text{ seconds}$$

$$D_3 = 8 \times 10^{-14} \text{ (at 1050°C during oxidation)}$$

$$t_3 = 60 \times 60 = 3.6 \times 10^3 \text{ seconds}$$

$$\text{Dt(effective)} = 2.5 \times 10^{-13} \times 2.16 \times 10^4 + 8 \times 10^{-14} \times 3.6 \times 10^3$$

$$= 5.4 \times 10^{-9} + 2.88 \times 10^{-10}$$

$$= 5.688 \times 10^{-9}$$

The junction is located at the point where background (n-type) impurity equals the (p-type) diffused impurity. The background impurity for 5Ω Cm resistivity n-type layer is found from the curve described by Irwin[8] to be 10^{15} atoms/Cm^3.

If $N_x = 3 \times 10^{15}$ atoms/Cm^3 then $x = x_j$ (junction depth in Cm from the surface).

∴ From equation (7-3)

$$x_j^2 = 4\,Dt\,\ell n \frac{Q}{N_x\sqrt{\pi Dt}}$$

$$= 4 \times 5.69 \times 10^{-9}\ \ell n \frac{3.04 \times 10^{14}}{3 \times 10^{15}\sqrt{\pi \times 5.69 \times 10^{-9}}}$$

$$= 22.76 \times 15^9 \times 2.3 \log \frac{3.04 \times 10^{-1}}{3 \times 1.32 \times 10^{-4}}$$

$$= 5.24 \times 10^{-8} \log 0.758 \times 10^3$$

$$= 5.24 \times 10^{-8} \times 1.88$$

$$= 9.8 \times 10^{-8}$$

$$x_j = 3.1 \times 10^{-4} \text{ Cm} = 3.1\ \mu\text{m}$$

Emitter Diffusion:

Windows were etched in the oxide layer applying the photo-resist techniques as explained in the case of base diffusion phosphorosilica film

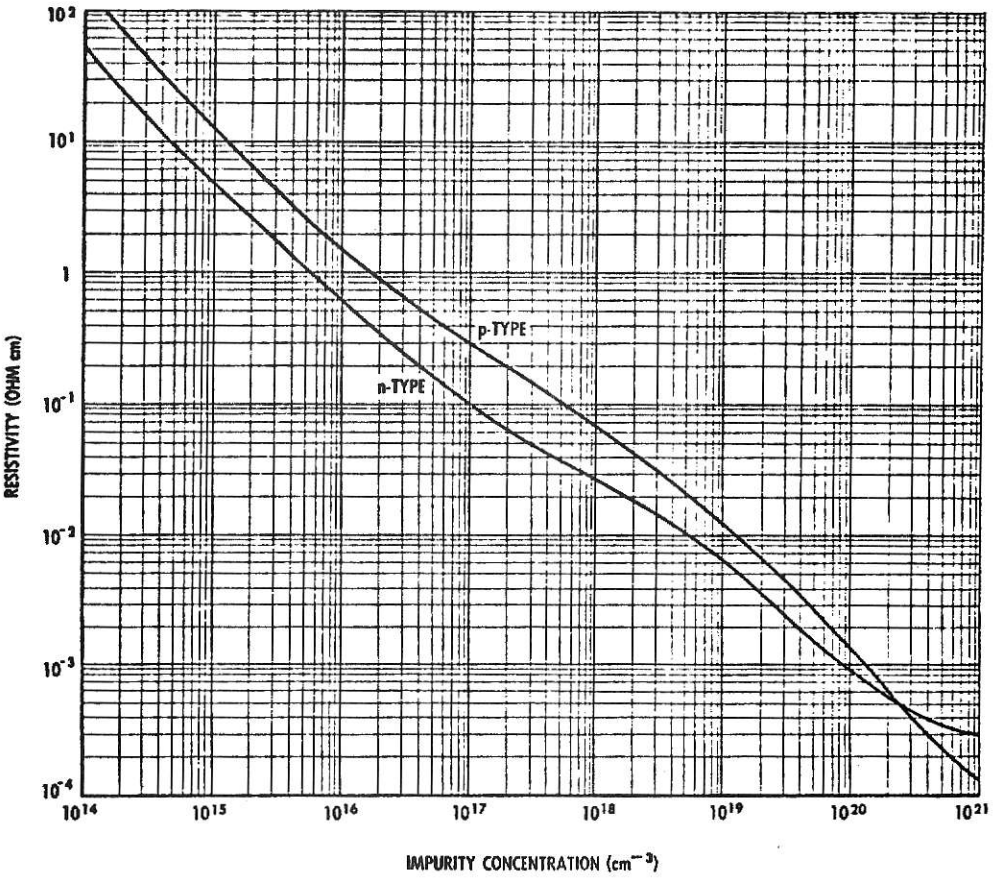


Figure (8-1) Resistivity versus impurity concentration[8]

was used for n-type emitter diffusion. Along with the emitter diffusion n^+ impurity diffusion was carried out under collector contact regions to achieve good ohmic contacts with aluminum.

After etching windows in the oxide layer for emitter areas and in those regions where ohmic contact to the collectors were to be made, phosphorosilica film was applied to the wafer using the spinner. The procedure is the same as for the borosilica film.

The wafer was then placed in the diffusion furnace for 35 minutes at $1050^{\circ}C$. This is supposed to give a base diffusion of about 2 microns. The wafer is then immediately put in boiling deionized water for 5 minutes to remove the phosphoric acid formed on the surface. Without this step the phosphoric acid forms a glass layer if the wafer is kept in air for a few minutes, which cannot be etched in buffered HF.

Oxidation:

After removing the phosphoric acid in the boiling water the wafer was placed in the buffered HF and agitated ultrasonically in the ultrasonic vibrator for six minutes. This etches the phosphorosilica glass that forms on the wafer surface. It also etches the thermally grown oxide over the wafer surface. Then two layers of silica film were applied to the wafer using the spinner. The spinner was rotated at 3000 rpm for 30 seconds each time. The second layer was applied after drying the first layer. The wafer was then placed in the furnace for 20 minutes at $950^{\circ}C$ to get a good hard oxide layer.

The above step was substituted after it was found that the Shipley AZ-1350 photo-resist did not work on the thermally grown oxide after phosphorus diffusion.

Ohmic Contacts:

The wafer was coated with the photo-resist and the ohmic contact mask was aligned over it. The system was then exposed to ultra violet light. The photo-resist was exposed and developed. Afterwards the wafer was placed in the buffered HF for four minutes to etch windows in the oxide layer for ohmic contacts. It was then rinsed in deionized water and the photo-resist was stripped off by putting in acetone for 10 seconds. The wafer was then dried after rinsing it with deionized water.

Aluminum Evaporation:

The wafer was placed in the vacuum system supported by two microscopically clean glass plates with the side of the wafer to be coated facing downwards. It was heated to $300^{o}F$ using a projector lamp. Voltage to the lamp was supplied through a variac and the wafer temperature was monitored by a thermocouple.

The aluminum was evaporated by placing a small piece of wire on a tungsten spiral fixed between two electrodes and passing a heavy current (to 50 amp) through it. The system was allowed to pump down to 10^{-5} Torr. before the aluminum was evaporated. The wafer was shielded by using a shutter during the early phases of evaporation. A sufficiently thick layer of aluminum was deposited on the layer.

Alloying the contacts:

The wafer was heated in the furnace for five minutes at $550^{o}C$ to alloy the aluminum to the silicon to achieve good contacts.

Final Metallization Contacts:

The wafer was then coated with the photo-resist and was exposed to ultra

violet light after aligning the final contact pattern mask over it. It was then placed in hot (95°C) aluminum etch for 15 seconds. Immediately after that it was rinsed with deionized water for three minutes. The photo-resist was stripped off by putting the wafer in acetone for ten seconds. The wafer was rinsed with deionized water. The composition of the etching solution used aluminum etching is as follows:

70 ml of phosphoric acid

3 ml of nitric acid

27 ml of distilled H_2O.

CHAPTER VIII

RESULTS AND CONCLUSIONS

The transistors were tested for the base-emitter and base-collector diodes. A needle probe was used to make contacts with the transistor electrodes. After checking for the diodes the transistors were connected to the Tektronix 576 curve tracer to check the output characteristics. Plate 1 shows typical output characteristics of one of the transistors in common emitter configuration. From the figure it can be concluded that the current gain 'B' is quite low. This can be attributed to wide base thickness. Attempts made to obtain lower base thicknesses did not succeed due to the defective borosilica emulsion employed during base diffusion. The emulsion had been lying for a long time and it had settled in form of small crystals. It is concluded that good results can be obtained by using new emulsion.

In the beginning it was proposed to fabricate isolated transistors and the masks were prepared accordingly. But since the required type of wafers were not available it was decided to fabricate transistors with common collector. These masks can be employed to fabricate integrated circiut bipolar transistors if p-type substrates with n-type epitaxial layer are used.

Cutting of the master drawing with whole array of transistors requires a lot of time and labor. This can be a lot easier if step and repeat cameras were available. The present layout for the transistors was made keeping in mind the convenience in drawing it. With the step and repeat facility

THIS BOOK CONTAINS NUMEROUS PAGES THAT WERE BOUND WITHOUT PAGE NUMBERS.

THIS IS AS RECEIVED FROM CUSTOMER.

PLATE 1

Output characteristics of one of the transistors.

THIS BOOK CONTAINS NUMEROUS PICTURES THAT ARE ATTACHED TO DOCUMENTS CROOKED.

THIS IS AS RECEIVED FROM CUSTOMER.

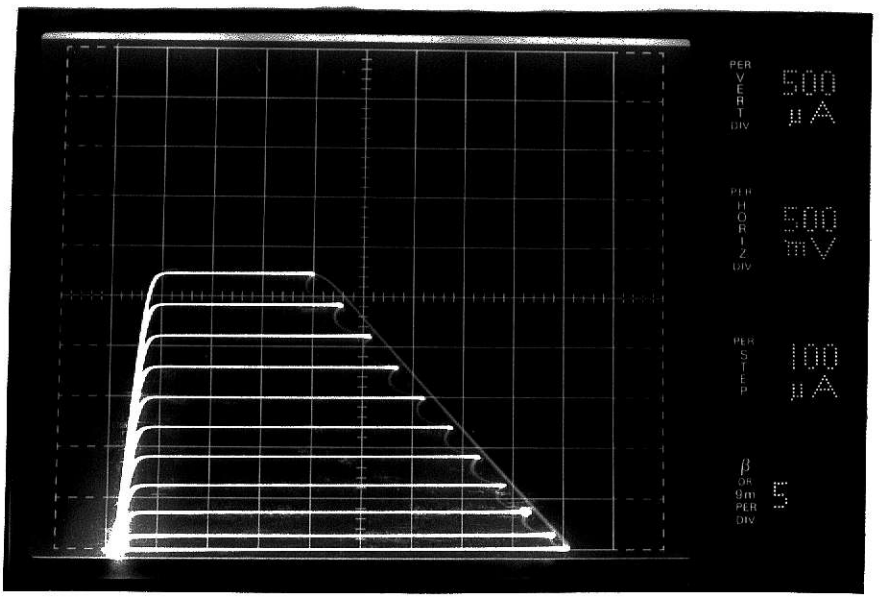

PER VERT DIV
500 µA
PER HORIZ DIV
500 mV
PER STEP
100 µA
β OR 9m PER DIV
5

Plate 1

Plate 2

Microphotgraph of the transistors

Plate 3

Photograph of the needle probe

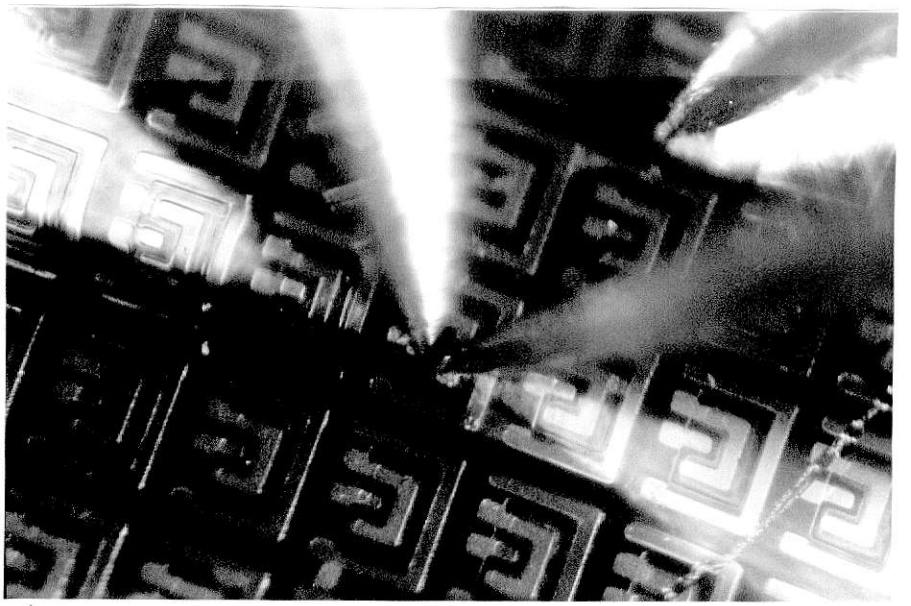


Plate 2

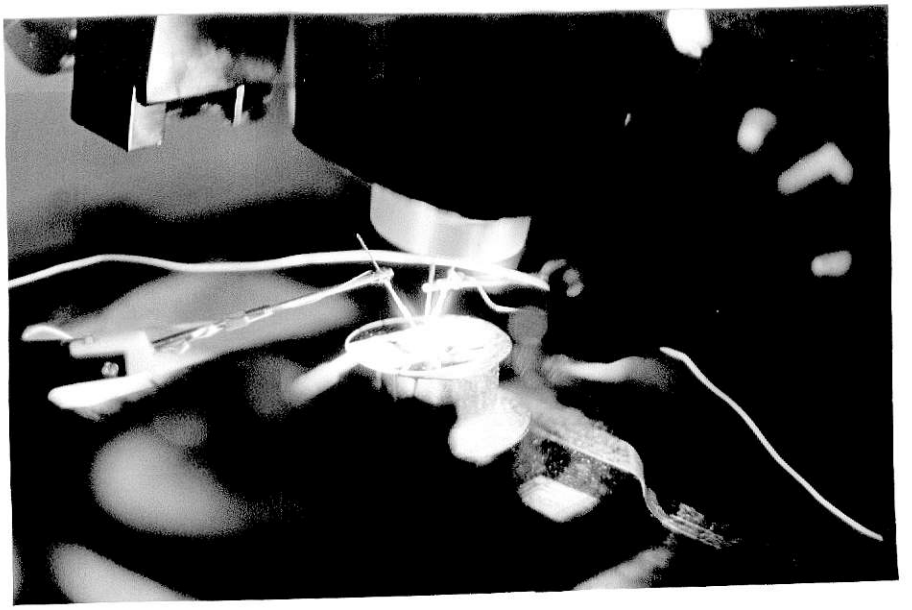


Plate 3

available the layout can be made for better performance rather than convenience in drawing the original artwork.

The diffusion depths were not measured due to the lack of the necessary equipment. If the grinding piston and cylinder were available this could have been done to establish the diffusion profile and some empirical relationship between diffusion depth time and temperatures for the setup in the laboratory. This can be of great help in the fabrication of integrated circuits.

SELECTED BIBLIOGRAPHY

1 Allison, D. F. et. al., "New etchant puts dielectric isolation in the groove", Electronics, Vo. 42, No. 10., 1971.

2 Burger, R. M., R. P. Donovan, "Fundamentals of silicon integrated device technology", Vol I, Prentice Hall, New Jersey, 1967.

3 Burger, R. M., R. P. Donovan, "Fundamentals of silicon integrated device technology", Vol II, Prentice Hall, New Jersey, 1967.

4 Fuller, C. S., J. A. Ditzenberger, "Diffusion of donor and acceptor elements in Silicon", J. Appl. Phys., Vol 27, May 1956.

5 Ghandhi, S. K. "The theory and practice of microelectronics", John Wiley and Sons, 1968.

6 Glinski, V. J., B. T. Murphy, "Bipolar integrated circuits formed in p-type epitaxial layer", IEEE transactions on electron devices, V ED-16, n-2, Feb. 1969.

7 Hause, S. S., R. A. Whitner, "Manufacturing beam lead, sealed junction, monolithic integrated circuits", Western Electric Engineering, V II n-r, Dec., 1967.

8 Irwin, J. C., "Resistivity of Bulk Silicon and diffused layer in Silicon", Bell System Tech., J., Vol 41, March 1952.

9 Ligenza, J. R., "Oxidation of Silicon by high pressure steam", J. Electrodemical Society, 109, Feb., 1962.

10 Murphy, B. T. et. al., "Simplified Bipolar Technology and its application to systems", V SC-5, n-1, Feb., 1970.

11 Owen, W. E., "Development of the techniques for the construction of semiconductor network and devices", Louisiana State University Engineering Research Bulletin, n-92, 1967.

12 Schwartz, S., "Integrated circuit technology", McGraw Hill, 1967.

13 Trumbore, F. A., "Solid solubilities of impurity elements in Germanium and Silicon", Bell System Technical J., V-39, Jan., 1960.

14 Vora, M. B., J. J. Chang, "P-I-N isolation for monolithic circuits", IEEE transactions on electron devices, V ED-15, September 1958.

15 Warner, R. M., Jr., J. M. Fordemwalt, "Integrated Circuits", McGraw Hill, New York, 1965.

16. Wolf, H. F., "Semiconductors", John Wiley and Sons, Inc., 1971.

ACKNOWLEDGEMENTS

The author wishes to express his sincere gratitude to Dr. Michael S. P. Lucas, who offered invaluable guidance and stimulating encouragement throughout the preparation of this thesis.

FABRICATION OF BIPOLAR JUNCTION TRANSISTORS

by

SUKHDEV SINGH DHALIWAL

B.Sc. Engineering (Hons), Punjab University, India, 1969

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Electrical Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1972

ABSTRACT

Various processes involved in the fabrication of an integrated circuit bipolar junction transistor, such as oxidation, diffusion, isolation, photolithography and metallization have been studied in this thesis. Fabrication of BJT's has been carried out in the laboratory using simplified fabrication techniques and "not-so-sophisticated" equipment. Cutting of the original artwork on a nylar sheet and its photo-reduction was carried out to obtain the photomasks for an array of about 250 transistors. Oxidation, diffusion, photo-resist techniques and metallization were accomplished using the available equipment to obtain good working transistors.