DEPARTMENT OF WEAPONS TRAINING
LOWRY AIR FORCE BASE
COLORADO

T R A N S I S T O R S

IN SERVICE TRAINING COURSE

CHAPTER 5

THE FABRICATION OF DIODES
AND TRANSISTORS

2 AUGUST 1961

FOR INSTRUCTIONAL PURPOSES ONLY

CHAPTER 5

Author: A1C Robert J. Widlar

THE FABRICATION OF DIODES AND TRANSISTORS

Abstract - The preparation of germamium and silicon single crystals is described from the chemical purification and zone refining to the crystal growing and doping. Examples of both the techniques and the equipment used are given. Next, the formation of PN junction by direct growing, alloying, and diffusion methods is explained. Special emphasis is placed on the fabrication of practical diodes and transistors. The attachment of ohmic and rectifying metal-semiconductor contacts is also covered. Soldered, bonded, and deposited contacts are included. Lastly, the surface treatment and encapsulation of completed devices is briefly considered.

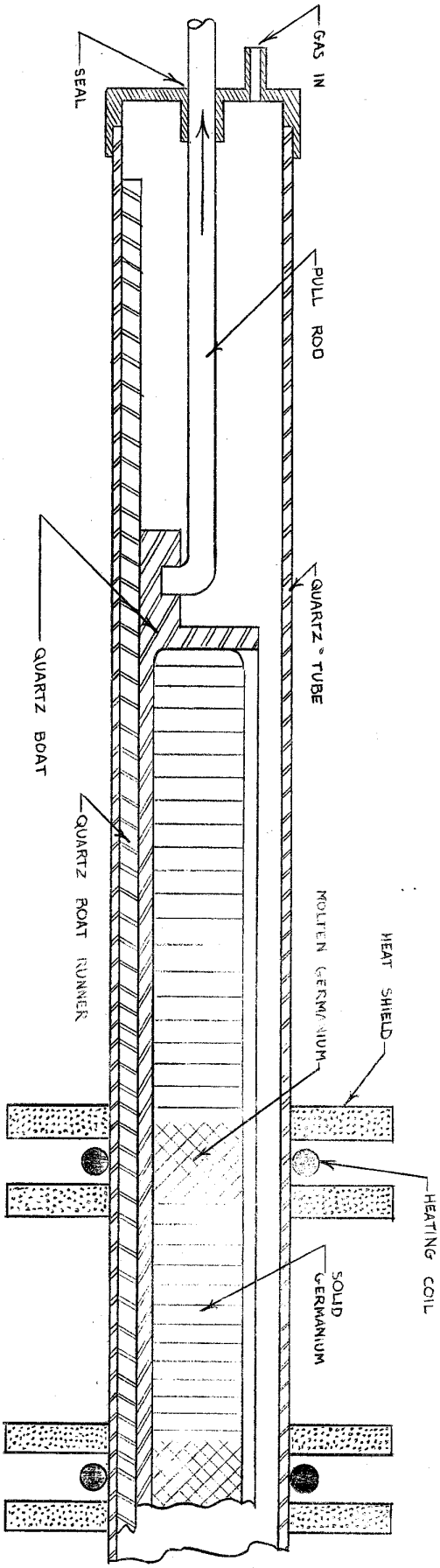


Figure 5.1. Germanium Zone Refiner.

INTRODUCTION

The purpose of this chapter is not so much to describe the exact processes used in manufacturing semiconductor devices, but to show the feasibility of producing extremely pure materials and seemingly complex structures on a production scale. In view of this, the chapter will bring out some of the more important processes used in preparing and purifying semiconductor materials. Moreover, the growing and doping of high quality crystals will be covered, and some of the techniques currently employed in fabricating PN junction devices will be explained. This will include direct growing, alloying, and diffusion techniques as applied to diodes and two junction transistors.

At the present time, the only semiconductor materials used to any appreciable extent are germanium and silicon. Of these, the technology of germanium is fairly well developed, with silicon lagging behind somewhat because of handling difficulties associated with its high melting point. This chapter will be confined to these two materials. Other compounds, i.e. silicon carbide, gallium arsenide, and indium antimonide, have limited application; but it is difficult to obtain good quality crystals so they are of minor importance and will not be covered here.

PREPARATION AND PURIFICATION

Before a semiconductor is suitable for use in a diode or transistor, the impurity concentration must be reduced to less than one part per hundred million . Prior to the transistor, this degree of purity could only be realized using chemical methods on a laboratory scale. However, intensive research has resulted in new techniques which can be applied to the mass production of extremely pure semiconductor materials.

Germanium. Germanium can be purified chemically by the fractional distillation of the volatile liquid, germanium tetrachloride (boiling point: 83°C). Germanium tetrachloride decomposes on contact with water into insoluble germanium dioxide, and the dioxide can be reduced to elemental germanium by passing hydrogen gas over it in a reduction furnace operating at 650°C. Using this process, germanium of extraordinary purity can be obtained. However, before it can be used in semiconductor devices, it must be refined further.

Most of the impurities remaining in germanium after chemical purification are more soluble in molten germanium than in the solid. Therefore, when molten germanium is cooled, these impurities will tend to stay in the melt as the material freezes. This suggests a method for further purification. This is, if a rod of molten germanium is progressively solidified down its length, that portion which solidified first will have a lower concentration of impurities.

ELEMENT	IMPURITY TYPE	SEGREGATION COEFFICIENT		MELTING POINT	BOILING POINT
		Germanium	Silicon	°C	°C
Boron	P	20	0.9	2300	2550
Aluminum	P	0.1	0.004	660	1800
Gallium	P	0.1	0.01	99.8	1600
Indium	P	0.001	0.0004	155	1450
Phosphorus	N	0.12	0.35	44	280
Arsenic	N	0.03	0.3	615	615
Antimony	N	0.003	0.04	630	1380
Zinc	P	0.01	—	420	907
Copper	P	0.00001	—	1083	2300
Gold	P	0.00001	0.00003	1063	2600
Iron	P	0.000001	—	1535	3000

Table 5.1. Physical Constants of Impurities Frequently Encountered in Semiconductor Work.

This portion of the rod could be cut off and the process repeated until a material of the desired purity was obtained.

One disadvantage of this system is that the rod must be removed from the furnace after each successive freezing to crop off the contaminated end. This is complicated by the fact that germanium must be melted in an inert atmosphere because it reacts with oxygen at these temperatures to form germanium dioxide. Hence, this process will be time consuming since the furnace must be purged to remove oxygen before heat is again applied. This difficulty is overcome by the zone refining process.

Zone Refining. At the present time, zone refining is used in the purification of practically all electronic grade germanium. In this process, a narrow molten zone is passed down the length of a germanium bar. The impurities tend to stay in the melt so they too are carried down the bar. An advantage of this system is that many passes can be made without removing the bar from the furnace.

A drawing of a zone refiner is given in figure 5.1. The germanium is contained in a quartz boat which is inserted into a quartz tube. The tube is sealed and purged with nitrogen, and then filled with an inert gas. Gas flow is maintained throughout the process to prevent the seepage of oxygen into the system. Molten zones are created in the germanium by radiation from heating coils. These zones are confined by the use of heat shields between the coils. The regions between the coils is left open to facilitate cooling of the germanium between zones.

After the molten zones have been established, the boat is slowly pulled down the tube. The germainum passing under the coils is then melted, and that passing out from the coils is cooled by radiation and refreezes. Therefore, the molten zones move down the bar, carrying along the impurities. The process can be recycled by removing heater power and returning the boat two its original position. After several passes, the impurities will be concentrated at one end of the ingot. Normally, about 90 percent of the bar will have adequate purity.

The number of passes required is determined by the nature of the impurities. The greater their solubility in molten germanium as compared with the solid, the more effective will be the process. This characteristic is described for a given impurity by the segregation coefficient which is the ratio of its solubility in the solid germanium to its solubility in the liquid. The segregation coefficient of several impurities frequently encountered in semiconductor work are given in table 5-1.

Fortunately, true chemical purity is not always required. Impurities that do not produce current carriers are of little

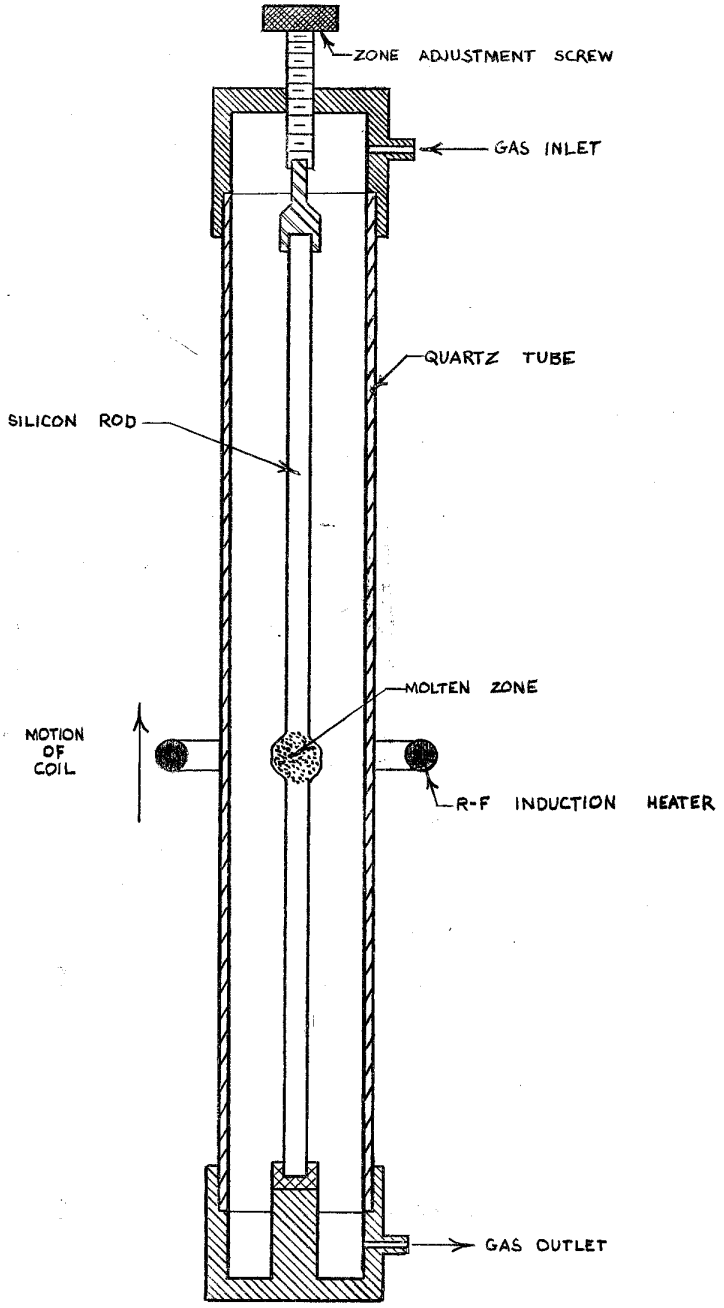


Figure 5.2. Silicon, Floating-Zone Refiner.

consequence as long as their concentration is low enough so that they do not produce excessive distortions in the crystal structure. It is believed that with zone refining some impurities such as boron, which has a greater solubility in solid germanium, are not removed but are converted to their respective oxides. In this state they do not produce current carriers since the outer shell electrons are tied up in producing a single, oxide molecule.

The purity of a semiconductor sample can easily be determined by measuring its resistivity and comparing it with the theoretical intrinsic resistivity of the material. With zone refining, near intrinsic resistivities can be obtained by making a large number of passes.

Silicon. Chemical purification of silicon can be accomplished, as with germanium, by fractional distillation of volatile silicon tetrachloride. The pure silicon tetrachloride is then reduced to elemental silicon by exposure to metalic zinc at high temperatures. The reaction product, zinc chloride, is volatile at these temperatures and is drawn off. The excess zinc can be removed in subsequent operations by evaporation since its boiling point is lower than that of silicon.

Further purification of silicon is difficult so the chemically pure silicon is frequently used "as is". Zone refining has been tried, but it does not give any significant improvement. This is partially due to the high segregation coefficient of most impurities in silicon and the contamination of silicon by its container.

High purity quartz is about the only satisfactory material that can be used to contain molten silicon. The reaction between quartz (silicon dioxide) and molten silicon to produce silicon monoxide proceeds slowly. However, molten silicon tends to take on impurities from the quartz; and it will also stick to quartz upon cooling causing frequent breakage of containers.

A process that has been used to produce high quality silicon is floating zone refining. The apparatus is shown in figure 5.2. A silicon rod is vertically supported at both ends within a sealed quartz tube. A single molten zone is established in the rod by a RF induction heater. The molten zone is contained by surface tension so there is no container problem. In operation, the heating coil is raised slowly, carrying the molten zone and the impurities to the upper end of the bar.

Only one zone can be maintained in the bar at one time so a large number of passes are required to obtain the required purity. Boron, which is frequently found in silicon, cannot be removed in any reasonable number of passes because of its high segregation coefficient. However, it can be removed chemically by passing hydrogen gas saturated with water vapor through the apparatus during operation.

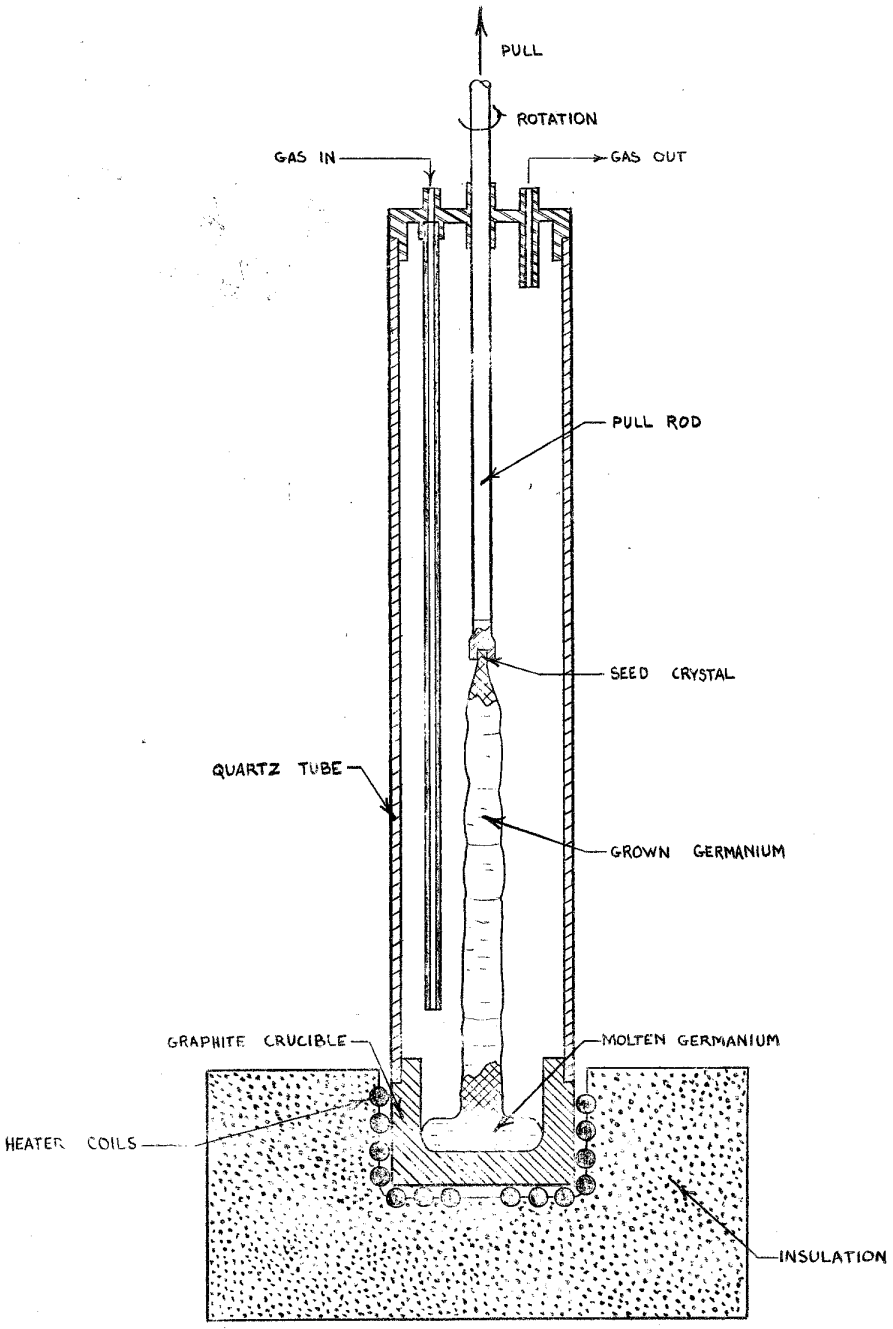


Figure 5.3. Germanium Crystal Puller.

CRYSTAL GROWING

Normally, when a molten semiconductor is cooled, freezing will start at several points; and a number of separate crystals will be formed. This polycrystaline material is useless for the fabrication of diodes and transistors because extensive discontinuities exist at the interface between crystals. However, if the melt is seeded with a single crystal and cooled in such a way that freezing proceeds around the seed, the atoms of the molten material will build onto the seed crystal and produce a single crystal. This process is referred to as crystal growing.

Germanium crystals are usually grown in a crystal puller like that shown in figure 5.3. Similar equipment can be used for growing silicon crystals, except that the graphite crucible must be replaced by quartz.

In operation, a seed crystal is lowered into contact with molten germanium which is heated just above its melting point (960°C). That portion of the melt in contact with the seed will be cooled by conduction to the seed holder so freezing will begin. As the seed is withdrawn slowly (a few inches per hour) a rod of single crystal germanium will be grown.

The diameter of the grown crystal depends on both the melt temperature and the pull rate so these quantities must be carefully controlled. The seed crystal is rotated during the pulling operation to promote uniform growth.

Silicon single crystals can also be grown using the floating zone apparatus illustrated in figure 5.2. Although it has its disadvantages, this system eliminates the container problem associated with silicon. To produce a single crystal, the lower support is seeded before the silicon rod is inserted. Then, as the floating zone is raised, silicon will refreeze on the seed producing the desired single crystal.

Since defects in a semiconductor crystal introduce recombination centers, the quality of a grown crystal can be determined from measurements of the minority carrier lifetime. The lifetime can be measured by creating hole electron pairs in the crystal, i.e., with a pulse of light, and measuring how long these carriers contribute to conduction through the crystal. If, for example, an intrinsic crystal is exposed to light its conductivity will increase. After the light source is removed, conductivity will fall off exponentially because of carrier recombination. The time required for the conductivity change to decrease to one half its value is an indication of the minority carrier lifetime.

Doping. A semiconductor can be doped by the addition of certain impurities to the melt during the crystal pulling operation. The impurity atoms will grow into the crystal taking the place of germanium or silicon atoms in the crystal structure. Boron, aluminum, gallium, and indium will produce P-type crystals while phosphorus, arsenic, and antimony will yield N-type crystals. Other impurities can be used for doping, but they generally produce complex electron structures in forming covalent bonds. This gives rise to recombination centers in the crystal.

The concentration of the impurity in the grown crystal will normally be lower than that in the melt because of impurity segregation. The segregation coefficients given in table 5.1 are for slowly grown crystals. Increasing the growth rate will also increase the segregation coefficient. The amount of increase will depend on the element: impurities with low segregation coefficients are generally more sensitive to growth rate. If the crystal is grown rapidly enough, impurity segregation will not take place so the impurity concentration in the crystal will equal that in the melt.

The best impurity for a particular application is determined by many factors, including segregation coefficient, the maximum solubility of the impurity in the crystal, and its boiling point. For heavy doping, an impurity with a high segregation coefficient and high solubility must be used. For light doping, an impurity with a low segregation coefficient is desirable since it would be difficult to measure the extremely small quantity of impurity required if one with a high segregation coefficient were used. The boiling point is also improtant. If a volatile element such as phosphorus is added to the melt, most of it will be lost by evaporation. This can cause contamination of the crystal puller which could affect subsequent operations.

JUNCTION FORMATION

There are several different methods available for making PN junctions using various doping techniques. For example, a crystal can be directly doped during growth to produce a PN junction; or impurities can somehow be added to a grown crystal producing a junction as with the alloy or diffusion processes. Frequently, as a result of these processes, a particular section of a semiconductor will contain both donor and acceptor impurities. When this is the case, the carrier type will be determined by whichever impurity is present in excess. If donor impurities are present in a crystal containing an excess of acceptor impurities, they will not contribute electron current carriers to the crystal. Instead, the free electrons supplied by the donor atoms will be used to complete the covalent bonds of some acceptor atoms. Thus, the presence of donor impurities in a P-type crystal will reduce the

effective number of acceptor impurities by providing electrons which fill vacancies in the covalent bonds.

The presence of two impurity types in a single crystal will not produce any harmful effects as long as the total impurity concentration is kept sufficiently low. If, however, the total impurity concentration becomes too high both the carrier mobility and the monority carrier lifetime of the crystal will be reduced by the localized fields of the impurity atoms.

Direct Doping. A PN junction can be made by directly doping the melt during crystal growth. In this process the melt is doped with a single impurity and a length of crystal grown. Then an impurity of opposite type is added to the melt in sufficient quantity to reverse the carrier type in the grown crystal. This produces a PN junction at the point of impurity reversal. Several junctions can be formed on a single bar by alternately reversing the impurity type, but the quality of the junctions will be progressively degraded due to the buildup of compensating impurities. This problem can be overcome by placing two crucibles side by side in the crystal growing apparatus. The melt in one crucible is doped with P-type impurity and that in the other with a N-type impurity. The seed crystal is dipped into the first crucible, and a length of P-type crystal is grown. It is then withdrawn and dipped into the other crucible, so a length of N-type material can be grown. This process can be repeated, and a large number of junctions grown.

Direct doping can also be used in the production of transistors. Again the carrier type in the grown crystal is controlled by the addition of impurities to the melt during growth. The collector is grown first and is lightly doped to reduce the amount of impurity required to produce a reversal when the base is grown. This is necessary because the performance of a transistor is greatly affected by the minority carrier lifetime in the base. The emitter, which is grown last, is heavily doped to give good injection efficiency.

After the emitter is grown, the melt is left heavily doped so the process cannot be repeated. Again reversing the impurity type would give an excessive impurity concentration and would deteriorate crystal quality. This shortcoming can be overcome by using the two (or three) crucible system.

A graded junction is usually formed with direct doping because of the slowness of the growth process. Furthermore, a wide range of resistivities can be obtained without complication since this is done merely by the addition of appropriate amounts of impurities. Finally, a large number of devices can be made from a single ingot by slicing out a section containing the junction and dicing it into individual units.

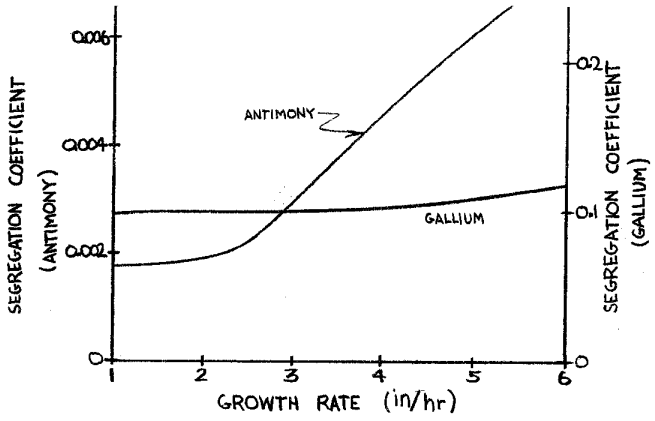
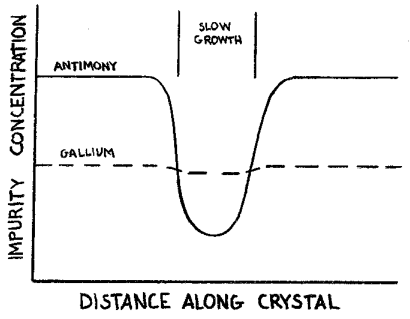
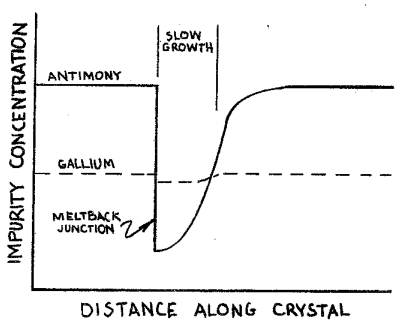


Figure 5.4. Variation of Segregation Coefficient with Growth Rate for Antimony and Galhum (in Germanium).



a. Without Meltback



b. With Meltback.

Figure 5.5. Impurity Concentrations in a Rate-Grown Transistor.

Rate Growing. In the rate growing process, both donor and acceptor impurities are present in the melt. One impurity is made to dominate over the other in the grown crystal by changing the growth rate.

As shown in figure 5.4, the segregation coefficient of antimony increases markedly with crystal growth rate, while that of gallium is practically constant over the range shown. If both these impurities are present in a germanium melt, their relative concentrations in the grown crystal will depend on the growth rate. If the melt is properly doped, an impurity reversal can be obtained by changing the growth rate such that rapid growth will produce an N-type crystal and slow growth will produce a P-type crystal.

This method is particularly applicable to the formation of transistor junctions. An N-type crystal is produced by rapid growth, but the growth rate is intermittently slowed to give a P-type base as shown in figure 5.5a. Using this system many sets of junctions can be produced in a single ingot by periodically slowing the growth rate.

It is advantageous to produce the base during slow growth because the base region is most sensitive to crystal quality. Slow growth reduces both the total impurity concentration and the likelihood of grown-in imperfections, such as missing atoms.

The usual method employed to slow the growth rate in transistor production is to increase the temperature of the melt. This causes the germanium to freeze less rapidly. If the melt temperature is increased enough so that a portion of the grown crystal is remelted, an abrupt junction will be produced as shown in figure 5.5b. This happens because the gradual junction produced as the temperature increased is melted off. The second junction will be formed as the melt cools. Impurity reversal will take place gradually and a graded junction will be produced.

The advantages of the rate grown transistor employing meltback should be noted. A graded impurity distribution is grown into the base giving an accelerating drift field, and the collector junction is graded which results in a lowered collector capacitance and a higher maximum collector voltage.

With rate growing a great degree of control over the base width and resistivity profiles is possible, but it is difficult to realize the low emitter resistivities necessary for high injection efficiencies at high current levels. This is fixed by the segregation properties of the impurities used. For this reason, rate growing is used primarily for the production of low power, medium-frequency transistors.

Rate growing can also be used for making junction diodes. In this case the growth rate is usually altered by changing the rate of

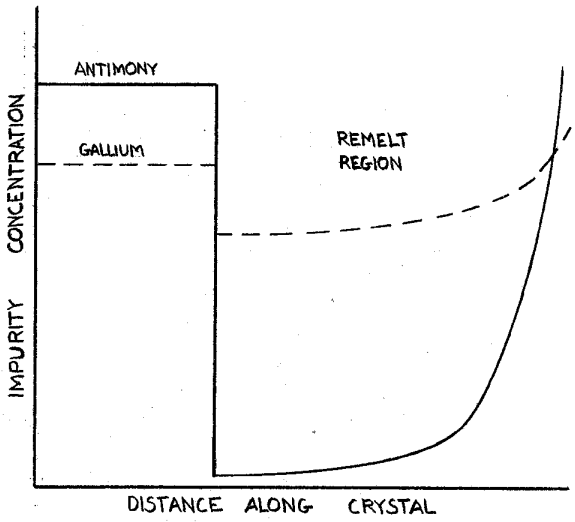


Figure 5.6. Formation of a Remelt Junction.

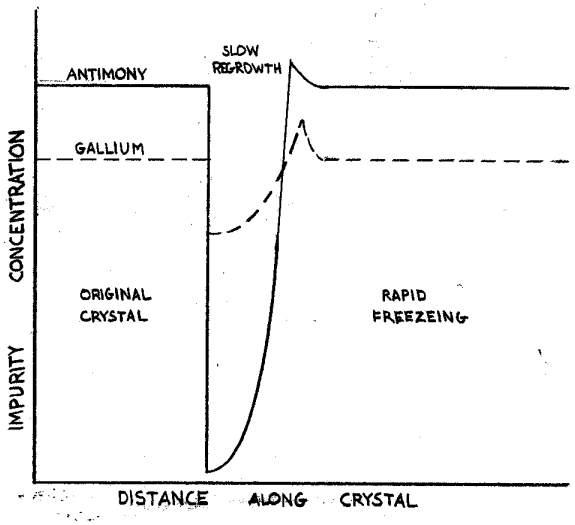


Figure 5.7. Formation of a Remelt Transistor.

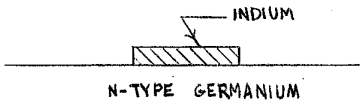
withdrawl although temperature control may be employed. The latter method usually results in alternate junctions having somewhat different characteristics. This is especially true if any amount of meltback takes place. Again a disadvantage of rate growing is the limited resistivity range. The low resistivities necessary for high conductance diodes cannot be realized. However, the great degree of control possible over junction characteristics, particularly in producing graded junctions, makes the system attractive for some applications requiring high reverse voltage or low junction capacitance.

Remelt Junctions. The remelt operation is performed on a good quality germanium crystal that has been doped with both gallium and antimony, but with antimony in excess. One end of this bar is heated causing it to melt. After about half the bar has been melted, the heat source is removed; and the bar is allowed to refreeze slowly. As the slow regrowth proceeds, an impurity reversal takes place because much less antimony will go into the crystal at the slow growth rate. Thus, a PN junction is produced at the furthest penetration of the molten zone as shown in figure 5.6.

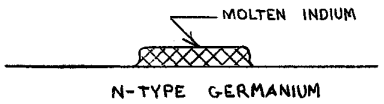
After the regrowth proceeds for some distance, the impurity concentration in the melt will increase greatly since all the impurities do not grow into the crystal. This is especially true for antimony. Therefore, near the end of the bar, which is last to freeze, the antimony will again dominate and a second junction will be produced. There will be a considerable distance between these two junctions so the structure is not suitable for use in transistors. However, if the second junction is cropped off, the remelt junction can be used in the fabrication of diodes.

Remelt transistors can be made by attaching one end of the double-doped bar to a heat sink while the other is heated strongly. After about half the bar has melted, the heat is removed. At first the bar will refreeze slowly, but rapid freezing will soon take over because of conduction to the heat sink. Impurity segregation takes place in the slow regrowth region (about 0.001 inch wide) so an abrupt junction is produced as shown in figure 5.7. However, as rapid freezing takes over, the impurities do not segregate. Instead, they grow into the crystal in their original concentration. This produces a second junction close to the first. As shown in the figure, this junction will tend to be graded.

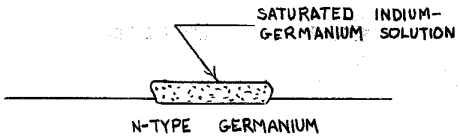
Reasonably narrow junctions can be obtained with reproducable results using this process. Furthermore, a graded base with its associated accelerating field is produced. Because this is essentially a rate-growing process, similar limitations on the range of resistivities obtainable are present. Hence, the remelt technique is used in practice for low power transistors.



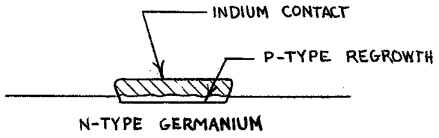
a. Indium dot is placed on N-type germanium and heated.



b. Indium melts (155°C).

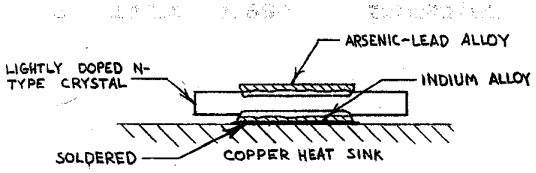


c. Indium dissolves germanium.

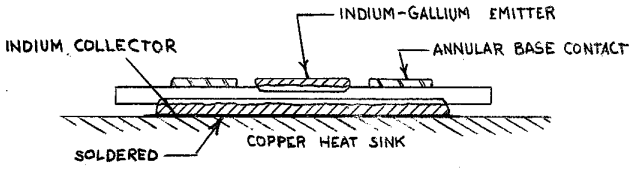


d. On cooling, dopped germanium recrystalizes.

Figure 5.8. Steps in the Formation of an Alloy Junction.



a. High Efficiency PNN+ Diode.



b. Alloy-Junction power transistor.

Figure 5.9. Examples of Alloy-Junction Devices.

Alloying. The formation of an alloy junction is illustrated in figure 5.8. A low melting point P-type impurity such as indium is placed on a lightly doped N-type germanium wafer. The assembly is heated in an oven containing an inert atmosphere. The indium will melt and take some of the germanium into solution. The assembly is then cooled, reducing the solubility of germanium in the molten indium. This causes the dissolved germanium to recrystalize on the N-type wafer. In recrystalizing the germanium will take along a considerable amount of the P-type impurity so a junction will be formed between the recrystalized germanium and the N-type wafer.

The depth of the alloy junction is determined by the maxium temperature used in the process. The solubility of germanium in molten indium increases with temperature. Thus, at higher temperatures more germanium will be dissolved and the penetration will be deeper. Time of exposure at this temperature has little effect on the shape of the junction.

It is necessary for the molten indium to wet the crystal surface for successful alloying. An ohmic contact will be formed between the indium and the crystal in any unwet areas. A soldering flux is usually applied to the crystal surface to remove surface oxides and promote wetting. Only that area beneath the indium is fluxed so that the indium does not spread when heated.

PN junctions can be alloyed to a P-type crystal using N-type alloys. Arsenic and antimony have been used, but these materials are brittle and do not wet germanium easily. Other materials such as lead can be added to the alloy to improve the mechanical properties, and with proper fluxing the alloy can be made to wet the germanium. Nonetheless, better results are usually obtained with indium so P-type alloy junctions are prefered for most applications.

The low resistivity of the recrystalized region makes the alloy junction well suited for high conductance diode rectifiers. The abrupt junction, characteristic of the alloy process, is somewhat of a disadvantage in applications requiring high inverse voltages.

In one type of power diode (figure 5.9a) an acceptor impurity (indium) is alloyed to one side of a high resistivity N-type wafer, producing a PN junction. A donor impurity (arsenic) is alloyed to the opposite side of the wafer producing a NN^+ junction. Rectification takes place at the PN junction, which has a low forward resistance because of the very low resistivity of the recrystalized P region. Current flow across the junction is largely due to the injection of holes from the P region. The high resistivity base crystal serves to widen the depletion region, thereby permitting high inverse voltages. The second alloy junction provides an ohmic contact to the base crystal and lowers the forward resistance by shortening the current path through the high resistivity material.

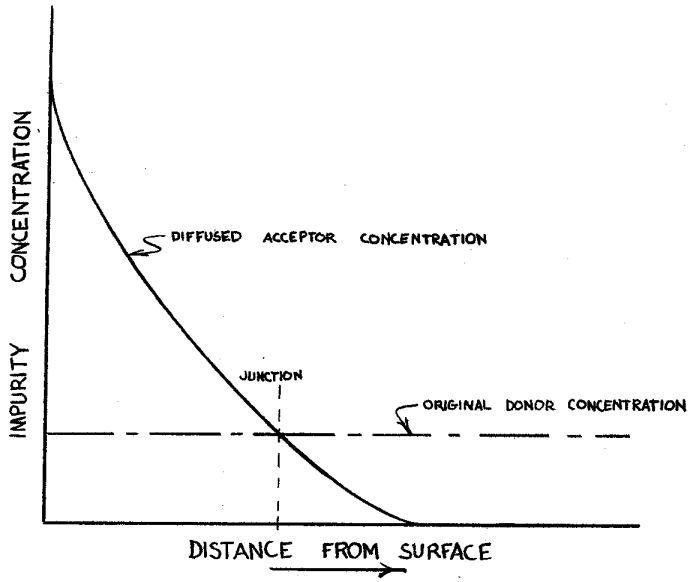


Figure 5.10. Formation of PN Junction by Vapor Diffusion of Acceptor Impurity into Dopped Crystal.

Alloy junction diodes are generally used in low frequency power applications requiring very low forward resistance. When low junction capacitance or high reverse voltage is the important consideration, the graded junctions produced by other methods are more desirable.

An alloy junction transistor is made by alloying junctions to opposite sides of a thin high resistivity wafer (figure 5.9b). The collector junction is usually made about three times larger than the emitter junction to improve collection of carriers crossing the base. In practice, base widths of 0.001 to 0.003 inch can be realized, but the degree of control is not too great. Most alloy junction transistors are the PNP type because of the superior alloying properties of indium.

Alloy junction transistors are used in low level audio amplifiers, audio frequency power amplifiers, and low frequency switching circuits. The design has not found much application at high frequencies. The alloy process is well suited to the fabrication of power transistors for the following reasons:

(1) The emitter can be heavily doped to give good injection efficiency at high current levels. Gallium, which has a higher solubility than indium in solid germanium, can be added to the emitter alloy to give these low resistivities.

(2) With proper techniques it is possible to make uniform, large-area junctions.

(3) Efficient heat transfer from the collector junction can be accomplished by soldering the collector directly to a heat sink. The thermal resistance of germanium is considerably higher than that of indium, and using the alloy process the only germanium in the heat path is the thin recrystalized region.

An additional advantage of the alloy process is that it can be performed by automatic machinery which lowers the device cost considerably.

Alloying techniques have been used on silicon with varying degrees of success. Problems are encountered in finding suitable alloys. Most tend to be very brittle. Furthermore, wetting of silicon with the alloy is hampered by the formation of silicon dioxide on the crystal surface. Alloy junction silicon transistors have not seen much use, but alloy emitters are widely used on silicon diffused base transistors. The impurity is usually vacuum deposited on the crystal surface and alloyed by subsequent heating.

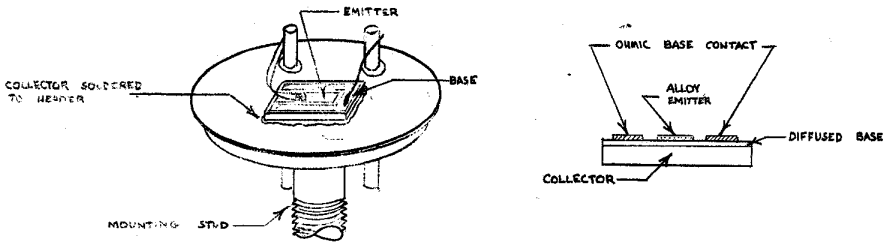


Figure 5.11. Construction of a Diffused Base Power Transistor Including a Section Showing Junctions.

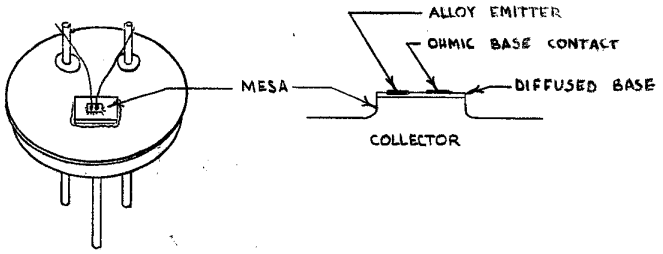


Figure 5.12. Construction of a Diffused Base Mesa Transistor.

Diffusion. Silicon and germanium can be doped by exposing a crystal to the vapors of an appropriate impurity at temperatures below the melting point of the crystal. Activated by thermal agitation, the impurity atoms bombard the crystal, some of them fixing themselves into the crystal structure by displacing the parent atoms. This will not just occur on the surface. The impurities can pentrate some distance into the crystal, but the number reaching the interior will fall off rapidly with distance. A comparatively long time is required for the impurities to diffuse into the crystal which makes the process easily controlled.

Diffusion doping is carried out in an evacuated furnace. The crystal and a volatile impurity are inserted into the furnace. On heating, some of the impurity will evaporate into the chamber and diffuse into the crystal. The depth of pentration will depend on the time of exposure, the temperature, and the impurity used (smaller impurity atoms diffuse more rapidly).

A PN junction can be formed on a doped crystal using diffusion doping. For example, if a donor impurity is diffused into a P-type crystal, an impurity reversal will take place near the surface of the crystal; and a juntion will be formed. The diffused-impurity concentration will fall off exponentially with distance from the surface so a graded junction will be produced. This is shown in figure 5.10. The impurity distribution and junction depth can be closely controlled.

At the present time the most important use of diffusion techniques is in the production of high performance transistors. A diffused base transistor is shown in figure 5.11. Acceptor impurities are diffused into the top of a thin N-type wafer, producing a junction about 0.0005 inch below the surface. Two thin metalic strips are deposited on the diffused P layer. The assembly is then heated. One of the metal strips is a P-type impurity so it alloys into the P-type base forming an ohmic contact. The other strip is an N-type impurity so it forms an alloy emitter on the base. The crystal is then soldered to the transistor case which serves as a heat sink and a collector terminal.

The advantages of this construction should be noted: The collector junction is graded giving a high collection breakdown voltage and low junction capacitance. The impurity distribution in the base is graded giving rise to an accelerating drift field. The graded base extends all the way to the surface so the injected carriers are accelerated away from the surface, and surface recombination is greatly reduced. The base is extremely thin. Base widths of 0.0001 inch can be realized in production. An alloy emitter is used. This gives high injection efficiencies at high

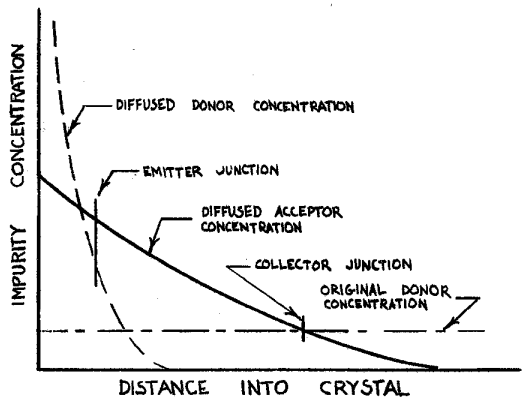


Figure 5.13. Plot of the Impurity Concentrations in a Double Diffused Transistor.

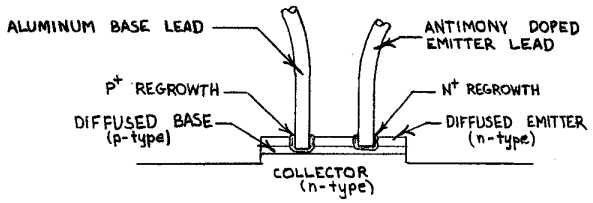


Figure 5.14. Construction of a Double Diffused Mesa Transistor Showing Compression Bonded Emitter and Base Leads.

current levels. A short thermal path is provided between the collector and the heat sink. In addition, the diffusion process is applicable to silicon. The collector junction of silicon transistors can be operated at a higher temperature, greatly increasing maximum power dissipation. Large area planar junctions can be produced without complication. The entire operation can be controlled closely and yields transistors having uniform characteristics.

Another diffused base transistor is shown in figure 5.12. It is similar to the one already described except that special techniques along with the "mesa" construction are employed to greatly reduce size.

In the mesa transistor the emitter and the base contact are deposited on the diffused base. Typically, both are a few ten thousandths of an inch wide. The excess material surrounding these contacts is etched away leaving a small mesa protruding from the base crystal. This greatly reduces junction area. The emitter and base leads are fine gold wires that are compression bonded to the deposited metal.

Significant results are obtainable with the mesa construction. Transistors capable of delivering several watts at frequencies above 250 megacycles have been made. High power mesa transistors that operate at kilowatt power levels at frequencies in the order of 10 megacycles are being developed. And an experimental mesa transistor that operates above 1000 megacycles has been built. In this transistor, the area of the mesa is less than the cross-sectional area of a human hair.

Double Diffusion. If a fast-diffusing acceptor impurity and a high-solubility donor impurity are simultaneously diffused into a P-type crystal, a NPN structure will be produced. This is illustrated in figure 5.13. The acceptor impurity will diffuse furthest into the crystal, but the donor will be present in higher concentration near the surface. As the acceptor concentration falls off with distance, the impurity type will revert to the original N-type. Small base widths can be obtained with this method. Since the process is governed entirely by the physical laws of diffusion, the results are easily reproduced.

Double diffusion techniques are particularly applicable to the fabrication of mesa transistors. The usual method of lead attachment is shown in figure 5.14. Connection is made to the base through the emitter. The bonding process used to make this connection is similar to alloying. Thus, when the aluminum wire is bonded to the mesa, it is surrounded by a P^+ regrowth. A PP^+ ohmic contact is made to the base, but the PN junction formed with the emitter prevents shorting of the base lead to the emitter. Similarly, the antimony doped emitter lead is surrounded by a N^+ regrowth that makes an ohmic contact with the emitter, but will form a PN junction if it penetrates through to the base.

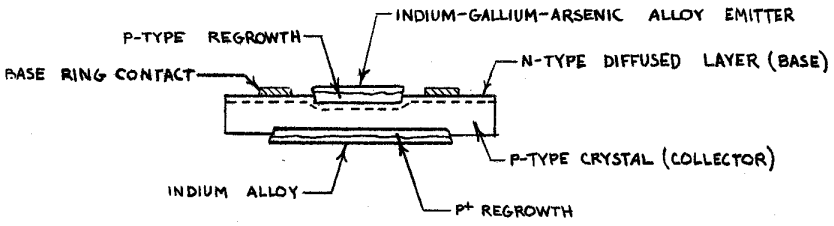


Figure 5.15. The Alloy-Diffused Transistor.

In power transistors of this type, it is desirable to diffuse a high-solubility donor impurity into the bottom of the wafer. This gives a triple diffused N+PNN+ structure. The highly doped region diffused into the collector reduces the thermal resistance between the collector junction and the heat sink, thereby facilitating the removal of heat.

Alloy Diffused Junctions. Multiple junctions can also be formed in a single operation using the alloy-diffusion process. An indium-gallium-arsenic alloy is placed on a P-type crystal and heated. When the alloy melts, arsenic will diffuse into the P-type crystal below the molten alloy and produce an impurity reversal since it diffuses considerably faster than either gallium or indium. When the assembly is cooled, gallium which has the highest maximum content in solid germanium will dominate in the recrystalized region. Hence, a P-type alloy emitter surrounded by a thin N-type layer will be formed. This is illustrated in figure 5.15.

Arsenic is usually evaporated into the furnace during this process to produce a low resistivity N-type layer on the crystal surface. This serves two purposes. First, a low resistance contact can be made to the base on this surface. Secondly, a strong drift field is created near the surface by the impurity gradient. This field turns the injected carriers away from the surface and reduces surface recombination. An indium-alloy contact is also fused to the P-type crystal. This contact does not serve as a collector, but instead it provides a low resistance path for the removal of heat from the collector junction.

Alloy-diffused transistors have found application at high frequencies and as low frequency power amplifiers where collector voltage and operating frequency requirements cannot be satisfied by alloy junction transistors. As with other diffusion transistors, the alloy-diffused transistor has a narrow base, graded base resistivity, and graded collector junction. The design is not adaptable to making very small devices, as in the mesa construction. It is, however, more economical since fewer operations are required in manufacture. The major frequency limitation of this transistor is the large area of the collector junction, but the design is useful at frequencies as high as 100 megacycles.

METAL-SEMICONDUCTOR CONTACTS

When a metal is attached to a semiconductor, either a rectifying or an ohmic contact will be formed. The type will depend on the metal used, the conductivity type of the semiconductor and the surface condition of the semiconductor where contact is made. Many metals will form a rectifying contact on a N-type semiconductor, but few

will produce rectifying contacts on P-type material. Any rectifying contact will be degraded if mechanical defects or contaminants are present near the surface.

Metal-semiconductor contacts can also be classified as injecting or noninjecting. Rectifying contacts will inject minority carriers into the semiconductor while ohmic contacts will not. This distinction points out the necessity of strictly ohmic contacts in some applications. For example, if an injecting contact were made to the base of a transistor, injected minority carriers would be swept across the collector junction. This would increase the base current and reduce the current gain of the device. Injecting contacts have already been mentioned in connection with the PNPM transistor in chapter 4.

Soldering. Almost any soldered contact will be ohmic. Soldering introduces defects into the semiconductor from thermal stresses set up during the operation. Hence, essentially ohmic contacts will be formed even when the metals used normally produce rectifying contacts. Appropriate dopants are usually added to the solder when the requirements for a noninjecting contact are demanding. These dopants are selected so that NN^{++} or PP^{++} junctions are produced on the semiconductor in the soldering operation.

Bonding. A bonded contact is formed by fuzing a metal containing certain impurities to the semiconductor. This is usually accomplished by the simultaneous application of heat and pressure. During the process, a liquid solution of the metal and the semiconductor is created at the interface of the two materials. Upon cooling, doped semiconductor is redoposited around the contact. The junction formed is similar in many respects to an alloy junction, and sometimes no distinction is made between the two processes. The primary differences are the time required for junction formation and the method of applying heat. The time required for bonding is appreciably less. Heat is applied by passing a high current through the contact or by conduction through the metal. Wetting of the contact area can usually be accomplished by the application of sufficient pressure.

By the choice of suitable contact materials, bonded contacts can be made either ohmic or rectifying. For example, diodes have been made by bonding an aluminum wire to N-type silicon. Aluminum is an acceptor impurity so the regrowth around the wire is P-type and a PN junction is formed. The same contact applied to P-type silicon would provide an ohmic PP^{+} junction. Similarly, if a copper wire containing phosphorous were bonded to P-type germanium, a rectifying contact would be produced. Phosphorous would dominate in the regrowth since it has a lower melting point, higher segregation coefficient, and higher maximum solubility than copper. Hence, the regrowth will be N-type.

Bonded contacts are frequently employed as the rectifying junction of medium power,diode rectifiers. The method is less complicated than alloying, particularly for silicon,because wetting of the contact area is difficult when conventional alloying techniques are used.

Another important application of bonding is the making of small area ohmic contacts to transistors. For instance, bonding is used to attach the base lead to grown junction transistors and, as shown in the previous section, to attach the emitter and base leads to double diffused transistors. Compression bonding is usually employed for these small area contacts. Sufficient pressure is applied to the lead to melt the semiconductor. Heat is not applied from an external soruce.

Vapor Deposition. A metal can be evaporated in an evacuated chamber and,deposited on a semiconductor. The presence of surface oxides on the semiconductor degrades the performance of these contacts, but an alloy junction (rectifying or ohmic) can be formed by heating the deposited contact. Therefore, this method is essentially on extention of alloying techniques.

One advantage of vapor deposition is that intricate contact pattern can be reproduced on a semiconductor surface. A contact is deposited over the entire surface. The desired pattern is then registered on the surface using photographic masking techniques. Subsequent etching will remove undersired portions of the contact. Very small patterns (down to the resolution power of a microscope) can be successfully reproduced using this system. The emitter and base contacts of diffused base mesa transistors are applied by vapor deposition.

Electroplating. Good quality rectifying contacts can be electroplated directly onto a semiconductor. The success of this process depends on the cleaning of the crystal surface prior to plating by electrolytic etching. The contacts are then electrodeposited without removing the crystal from the electrolyte, which prevents possible contamination.

The surface barrier transistor is an unusual design that uses electrodeposited contacts for the emitter and collector. The manufacturing process involves the electrolytic etching of depressions into opposite sides of a N-type germanium wafer. This is accomplished by directing small jets of an electrolyte against opposite faces of the wafer and passing a current from the electrolyte into the germanium. The system is shown in figure 5.17. After etching, metalic ions in the electrolyte are plated into the wells by reversing the etching current.

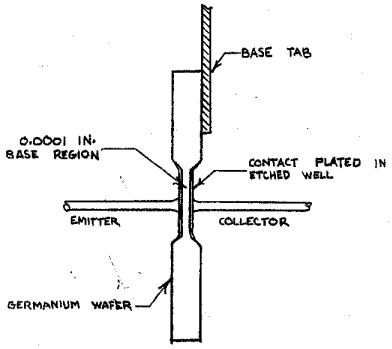


Figure 5.16. Construction of the Surface-Barrier Transistor.

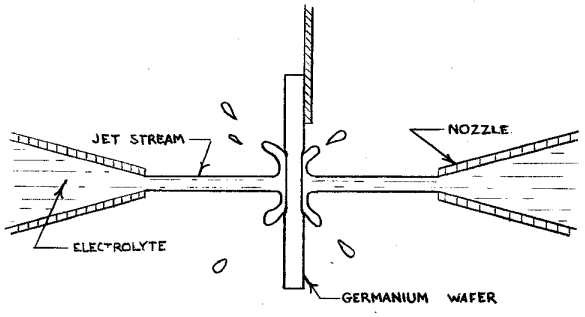


Figure 5.17. Set up for Etching and Plating Operations on a Surface-Barrier Transistor.

An automatic control over base width is inherent in the etching process. During etching, depletion regions are formed near the surface of the crystal in contact with the solution. When the depletion regions on each side of the crystal meet, etching will stop and current will drop off. Thus, flat bottomed wells separated by about 0.0001 inch are formed.

SURFACE TREATMENT AND ENCAPSULATION

The condition of the crystal surface near a junction of a diode or transistor greatly affects the performance of the device. It is desirable to have the surface free from contaminants, defects, and mechanical strain. This is necessary to reduce surface recombination of minority carriers and to eliminate conducting paths across the junction.

The usual method used to clean the surface is chemical or electrolytic etching. Etching removes the damaged portions of the crystal and produces a clean, mirror-like finish.

Even when the surface is properly cleaned, it has been found that the absorption of water vapor and air affects device performance in an unpredictable fashion. These contaminants can cause large changes in characteristics, particularly the reverse current of a back-biased junction or the current gain of a transistor. Therefore, it is usually necessary to outgas the completed units in a vacuum. This should be done at a high temperature, but the temperature is limited by the device. After outgassing, the diodes and transistors must somehow be protected from the atmosphere. The most effective protection is to hermetically seal the device in an inert atmosphere. Coating with some protective compound such as silicone grease and encapsulation in plastic has been shown effective for noncritical applications.

The encapsulation used must also provide for the dissipation of heat from the transistor. In low power units no special precautions are required since adequate cooling can be provided by conduction through the leads. Medium power transistors should be mounted directly to a metal case. This is also true for high power transistors except that the case must be designed so that it can be fastened to a heat sink.