TRANSISTORS: Theory and Application

Point-Contact

Part V

By ABRAHAM COBLENZ and HARRY L. OWENS
Signal Corps Engineering Laboratories
Forth Monmouth, New Jersey

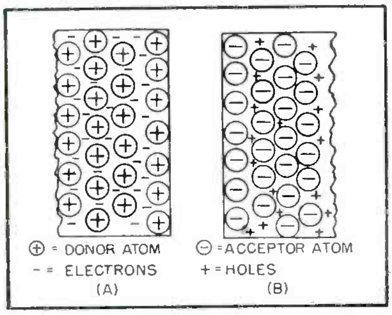

FIG. 1—Donor material (A) and acceptor material (B) show arrays of charged atoms just below surface of germanium

PREVIOUS articles of this series have developed the theoretical concepts associated with *p* and *n*-type semiconductor materials. These principles will be applied to a detailed analysis of the theory of operation of the point-contact transistor.

Temperature Dependence

Much information is obtainable from the covalent bond picture of germanium and silicon. From this picture it is apparent on a qualitative basis that the transistor using germanium or silicon is a temperature-dependent device. An increase in temperature will be accompanied by disruption of some of the covalent bonds thereby freeing electrons to act as carriers.

By definition, the presence of these additional carriers will increase the conductivity. Many of the parameters which characterize a transistor are dependent on conductivity, and these parameters will vary with temperature. If the temperature is increased sufficiently, enough covalent bonds are broken so that additional increases in temperature will have a negligible effect on the number of available carriers. Under such circumstances, the conductivity approaches an upper limit, called the intrinsic conductivity. The word "intrinsic" implies that the conductivity of the semiconductor is essentially dependent upon the properties of the material itself rather than upon the impurities that control its conductivity at lower temperatures.

The intrinsic temperature for germanium normally suitable for transistors is of the order of 100 C. Near the intrinsic temperature, control of the carriers is very difficult because of their very large number and high thermal energy. Since this control is essential to efficient transistor action, it is evident that the intrinsic temperature sets an upper limit for satisfactory transistor operation. In practice, other considerations further limit the maximum operating temperature so that it may be substantially less than the intrinsic temperature.

In Part IV of this series it was shown that impurity atoms occupy lattice sites and by their displacement of germanium atoms give rise to excess electrons or holes. In the conduction process the carriers are the holes or the electrons—the ionized atoms at the lattice sites do not contribute to the conduction.

Conduction in a semiconductor is essentially electronic rather than ionic. In an ionic conduction process, conduction is by atoms that have gained or lost one or more electrons (ions).

At a given instant of time if it were possible to look just below the surface of a semiconductor such as germanium, there would be seen an array of countless germanium and impurity atoms vibrating about their mean lattice positions due to thermal agitation with mobile holes or electrons moving among the atoms. In the case of a donor material, the impurity atoms have a net positive charge and it is convenient for analytical purposes to picture the situation as in Fig. 1A. Here is shown an array of donors, hereafter indicated as a plus sign enclosed in a circle, near the surface of the *n*-type material. For the *p*-type material, Fig. 1B, an array of acceptors with their negative charge is shown. This picture will serve as a useful tool in the analysis of transistor action.

P-N Junction

Figure 2A shows a piece of *p*-type material adjacent to a piece of *n*-type material, such an arrange-

PREVIOUS ARTICLES IN THIS SERIES

Transistors: Theory and Application, Part I, p 98, March 1953.
Energy Levels in Transistor Electronics, Part II, p 138, Apr. 1953.
Physical Properties of Electrons in Solids, Part III, p 162, May 1953.
Transistor Action in Semiconductors, Part IV, p 164, June 1953

Transistor Operation

Detailed discussion of point-contact transistor action is presented, including physical construction, hole and electron movement, potential hills, surface states and relationship of these phenomena to external connections and applied voltages

ment producing a *p-n* junction. It is not possible to make a satisfactory junction by simply taking a piece of *n*-type and a piece of *p*-type germanium and putting them together, regardless of the pressure used to hold them together or how carefully the interfaces are cleaned and polished. Experience has shown that, because an action occurring at a microscopic level among particles of atomic dimensions is involved, it is extremely difficult to get orders of purity and cleanliness or smoothness at the surfaces that are required to obtain a satisfactory *p-n* junction. It is therefore not considered feasible to make a *p-n* junction mechanically. A process for making *p-n* junctions will be described later.

These *p-n* junctions have been used successfully in circuit applications commonly associated with the familiar germanium diode.

This structural arrangement is a convenient device for establishing a general rule for the polarities of applied potentials both for transistors and diodes.

Potential Hills

It might appear at first that under the action of ordinary diffusion, the excess of holes (Fig. 2A) would diffuse into the *n* region and excess of electrons would diffuse into the *p* region so that, in time, the *p-n* junction as such would cease to exist. Of course this does not happen at ordinary temperatures. A very simple analysis, but one which has some very far reaching implications, will show why.

In trying to diffuse into the *n* region, a hole in the *p* region encounters a barrier layer of positively-charged atoms (donors) just across the junction plane. The positive electric field created by the donors opposes the transgression of holes.

Figure 2B shows what the potential variation or distribution looks like to the hole trying to get across the junction. As it approaches the junction, the lines of electric flux extending out impede its motion, and if the hole does reach the junction, it must have had initially a large energy to overcome the opposition of the donor electric field. The increase of energy which a hole must thus possess in order to move against the electric field, is shown in Fig. 2B as a small potential hill. This name is commonly applied to the effect at the junction that prevents the holes from diffusing across it.

At room temperature, for example, the holes will, in general, not have sufficient energy from thermal agitation to climb the potential hill and therefore a significant number of holes do not diffuse from the *p* to the *n* region. Electrons in the *p* region however will, due to the electric field of the donors, diffuse easily into the *n* region sliding down a potential hill. However, the electrons are only minority carriers in the *p* region, and their migration to the *n* region will never annihilate the *p-n* junction because there is already an excess of electrons in the *n* region.

In Fig. 2C is shown the corresponding condition for electrons in the *n* region which would, ordinarily, diffuse into the *p* region under normal kinetic vibration or motion. Analogously for the case of the holes, electrons that approach the junction come under the influence of the negative electric field

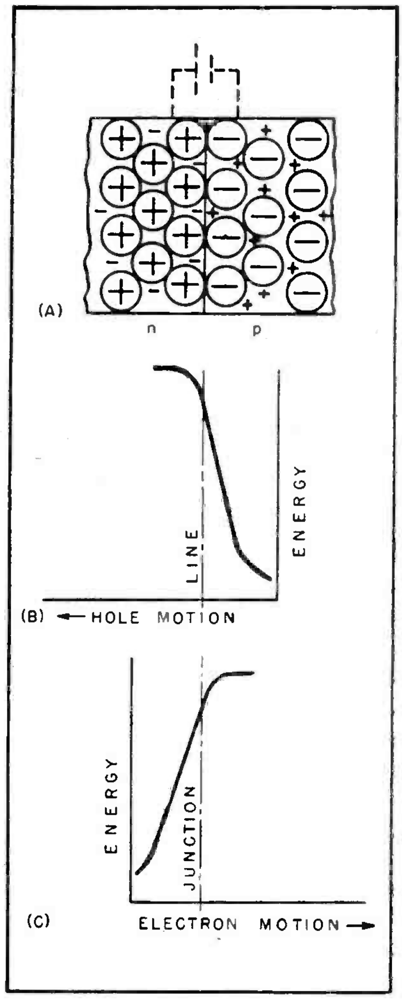

FIG. 2—Donor and acceptor materials arranged as a junction are shown in *A*. Any hole trying to diffuse from acceptor material to donor must climb potential hill due to array of donors near surface of n germanium (B). An electron trying to diffuse from n region to p region must climb potential hill indicated in *C*

due to the array of acceptors and encounter the potential hill that turns back most of them.

The effect due to the array of acceptors and donors is representable as in Fig. 2A by a small battery shown dotted in the figure. For the position shown it will introduce the same potential hills as do the arrays of acceptors and donors. This picture, however, is extremely important because it provides instantly the answer to the polarity problem.

If a battery is connected as in Fig. 3A, it aids the equivalent battery of Fig. 2A and increases the height of the potential hill. With this arrangement the number of carriers that can climb the potential hill is very small and most of the conduction is actually due to minority carriers since this polarity of bias enhances the flow of electrons from the p region to the n and of holes from the n region to the p region. The number of minority carriers is so small that the total current is small and this polarity of bias gives rise to the reverse-current or high-resistance condition.

In Fig. 3B, however, the external battery is connected so as to oppose or flatten out the potential hill or the equivalent battery of Fig. 2A. The flow of holes from the p to the n region and the flow of electrons from the n to the p region is enhanced. This is the low-resistance or forward-current polarity of bias.

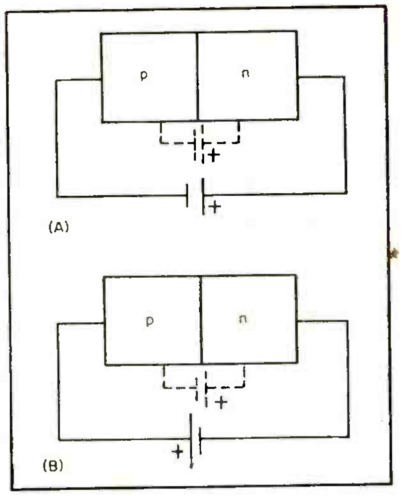

FIG. 3—In A, the potential hill, represented by dashed-line battery, is reinforced or raised by connection of external battery. This is identified as reverse-current (high-resistance) connection. In B, the potential hill, represented by dashed line battery, is overcome or lowered by external battery. This is forward-current (low-resistance) connection

Barrier Potential

The order of magnitude of the potential difference across the junction in the absence of an external battery is in tenths of a volt and it may thus appear that a very small battery connected as in Fig. 3B would be sufficient to completely annihilate the potential hill. This is not so. A semiconductor that has approximately a hundred thousand times the resistance of a conductor such as copper is involved. An important potential drop occurs in the p and n regions without reference to the barrier at all and, of course, as the current increases this drop increases.

The more the barrier is broken down by using the polarity of Fig. 3B the more readily do the carriers move across it. In that sense the resistance of the barrier is decreased. As the current increases the net drop available across the barrier itself for counteracting this contact potential at the barrier decreases, and it is clear that the two actions are opposite in direction. Nevertheless, it is possible to make the potential hill quite small by the use of potentials of the order of one volt. If the battery potential is made too high in the reverse direction voltage breakdown occurs, see Fig. 3A. In the case of the forward connection of Fig. 3B, the application of too high a voltage will permanently damage the junction due to the heating effects of excessively high currents.

State of Theory

Before proceeding to the discussion of the theory of the point-contact transistor it is necessary to make some precautionary remarks. The vacuum tube has been known for over 50 years and vacuum-tube theory is on relatively firm ground. In vacuum tube theory, the reader is accustomed to seeing only the well-worked-out and thoroughly established theories in print, each such theory having had ample time to be checked and cross-checked by many workers in the field. This is in general not true of transistor theory because the entire field is barely five years old.

Authorities in the field today do not unanimously agree upon a theoretical explanation of point-contact transistor operation. To the reader familiar with the calculus this will not appear particularly surprising. The mathematical analysis of physical phenomena occurring at a point such as the pointed tip of the fine wire used for the cat whisker is extremely difficult and involved. Theories that adequately explain all the known phenomena involved in point-contact operation have not been completed. Many of the equations developed are so complex and difficult to verify experimentally that physically significant solutions are not yet available.

There is, therefore, no complete mathematically supported theory of point-contact rectification or transistor action that may be used to interpret observed phenomena or predict future behavior. The theoretical explanation of the action of a point-contact transistor that will be given must be considered to be merely one of the many possible theories. The reader should look upon the explanation proffered with a certain amount of critical reserve. The explanation to be given is, in the opinion of the authors, the best currently available but because the art is so young it is well to bear in mind that it may be considerably modified in years to come.

Theory of Operation

In Fig. 4 is shown the essential arrangement for point-contact transistor operation. The base (or pellet) is a piece of n-type germanium of about 4 to 5 ohm-cm resistivity, approximately 20 mils thick, and about 50 mils in length and width. The cat whiskers are wires of some metal such as phosphor bronze approximately 5 mils in diameter, spaced approximately 2 mils apart with some simple provision such as a bend in the wire, to keep a few grams pressure on the surface of the germanium.

The cat whisker shown on the left is called the emitter for reasons which will appear shortly. The other cat whisker is the collector. The base connection on the underside of the germanium pellet consists of an ohmic soldered connection.

The reader may be concerned with possible deleterious effects of the heating involved in the soldering operation. Actually harmful effects on the electrical properties of the germanium are not observed until temperatures considerably higher than those encountered in the soldering operation are reached. Further, to degrade appreciably the characteristics of the germanium, elevated temperatures must be maintained for times much in excess of those encountered in the soldering process.

It is known that electrons that find their way to the surface of the semiconductor become held or bound in certain conditions or states that are different from the quantum states of the electrons inside the bulk of the material. When the electrons enter into these surface states (surface quantum states) they appear to be bound and do not readily return into the bulk of the material. The result is that a surface layer of such electrons in these surface states is built up on the material. For an n-type material, and mention has already been made that point-contact transistors are usually made with n-type material, this surface layer of electrons, as shown in Fig. 5, combines with the array of donors just below the surface to form a small potential hill, as shown by the dotted battery. This arrangement produces an effect analogous to the p-n junction discussed in connection with Fig. 2.

Rules for Polarity

Two rules may be established for the polarities of connections of transistors, which will apply to all the transistors, point-contact and junction. The reasons for these rules will become apparent during this article and the next.

(1) The emitter is always biased in the forward or low-resistance direction.

(2) The collector is always biased in the reverse or high-resistance direction.

On the basis of these two rules, and if we accept the surface-states electron theory discussed above, it is apparent that the emitter bias or battery E_e of Fig. 4 will be connected to oppose the potential hill at the surface as shown, and the collector battery E_c will be connected to aid the surface potential hill. It is thus seen that without knowing the theory of operation of the transistor the reader by the simple mnemonic outlined, will know exactly how to bias a point-contact transistor, or, as we shall see later, a junction transistor.

Looking at Fig. 4, the surface electrons near the emitter have been removed because of the polarity of the connection, but those a little further away are still present. This is intended to illustrate the concept of bound electrons. If a metal plate, such as in a capacitor, is given a charge the charge will, in general, distribute itself evenly over the surface. If a battery terminal of suitable polarity is connected to the plate, all the charge on the plate can be drained off at once merely by connecting the battery at one point because the charges on a conductor plate are, in general, not bound. This is not true of an insulator or semiconductor. When the cat whisker at the emitter is applied as shown, only the electrons on the surface in the immediate vicinity are whisked away by the battery; the remaining surface electrons remain in their bound states as indicated.

Cat Whiskers

One might wonder why it is necessary to use a cat whisker? Why can't a simple metal-plate connection be used? To answer this question it is necessary first to investigate the general characteristics of a point in electrical work.

An electric field such as may exist between the plates of a capacitor is usually considered to be made up of lines or rays called flux lines. The number of such flux lines depends on the potential, not on the area. Thus, for a given potential, electric flux lines are extremely crowded at a point and the electric flux intensity which by definition is flux lines per unit area is very high.

In a lightning rod, for instance, since the cloud must be discharged into the rod as soon as possible before the cloud accumulates enough charge to cause a lightning flash, the rod is made pointed. By crowding the electric flux lines from the cloud into a point, a sufficient field intensity is developed to cause a current to flow down the electric rod and thus partially discharge the cloud.

Large potentials may not be applied to a transistor because of the possibility of very large currents, which will cause heating or, for the case of reverse connections, voltage breakdown. To break the covalent bonds in the germanium, a high-intensity electric field is needed at the emitter. The only way to get this high-intensity field is to use a point. For this reason the emitter contact is a point contact.

Hole Injection

When the emitter bias E_e is applied, even though this bias is of the order of 1 volt, a high-intensity electric field is created at the point and imparts sufficient energy to the electrons in the valence bonds nearby to raise them into the conduction band and to break these valence bonds. These electrons under the influence of the applied potential immediately flow out of the material and into the emitter. The breaking of the valence bonds creates holes in the immediate vicinity of the emitter as shown in Fig. 4. While the term is somewhat a misnomer, this process of creation of holes is called injection because the effect is the same as if holes had been injected by the emitter. As soon as the holes are created they drift toward the collector under the influence of the electric field between emitter and collector.

Since we are dealing with an n-type material, with free or excess electrons, many of the holes on their way to the collector will recombine with the electrons and cease to exist. As these recombinations are taking place all the time, the number of holes that will combine with electrons increases with the transit time. Since a large number of holes is necessary for maximum effectiveness in transistor action, it is desirable to place the emitter and collector close together. Point-contact transistors of current design have emitter to collector spacings up to approximately 5 mils. If spaced too far apart, the gain and frequency response are adversely effected.

In the region of the collector the

potential hill at the surface limits the current flow and in the absence of holes the current will be of the order of one or two milliamperes. Holes that reach the immediate vicinity of the collector are attracted by the negative charge there and in moving to the collector point tend partially to cancel out or nullify the potential hill at the surface. In this way more electrons are permitted to climb the potential hill and the current at the collector is increased.

This increase in collector current as a result of hole injection is transistor action, and it is observed that small changes in the hole injection such as may be due to a modulation of the emitter bias by an a-c potential result in amplification of the emitter current containing the modulation intelligence.

Resistance Gain

In transistor action the low-resistance connection is used in the emitter circuit and the high-resistance connection in the collector circuit, referring to the biasing potentials applied. The following rules apply:

(1) The internal resistance of the transistor between emitter and base, usually designated by r_{11}, is always a relatively low resistance. The internal resistance of the transistor between collector and base, usually designated by r_{22}, is always a relatively high resistance.

(2) Transistor action is accompanied by a resistance gain in that the ratio of r_{22} to r_{11} is greater than 1.

The low-resistance connection of the battery E_e in Fig. 4 satisfies the emitter circuit requirement of rule 1 but in the collector circuit the action of the holes nullifies the effect of the surface charge and thereby lowers the internal resistance between collector and base. Because point-contact transistor action is always accompanied by this effect of the holes in decreasing the internal resistance of the germanium, it is seen that the collector-to-base resistance requirement of rule 1 cannot be met. If r_{22} is small, satisfactory voltage and power gains cannot be achieved. However, there is an additional factor involved.

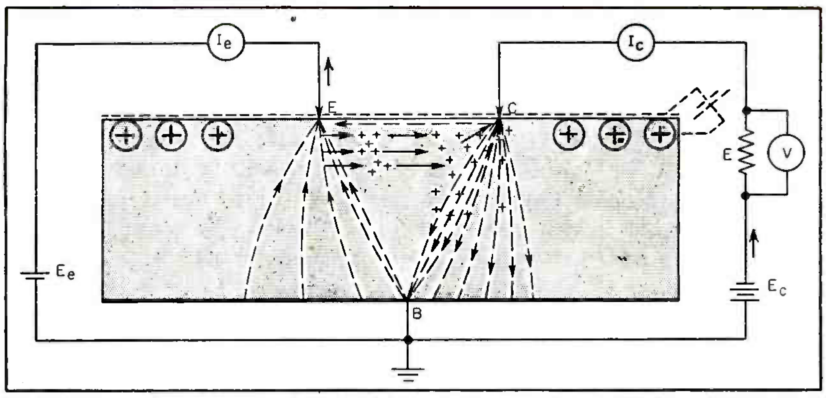

FIG. 4—Drawing illustrating point-contact transistor action shows role of surface states and array of donors with equivalent potential hill denoted by dashed-line battery

Looking at Fig. 4 it is seen that some holes do not follow a straight-line path from emitter to collector but, by the process of diffusion and random motion, travel in a more or less circuitous or indirect path to the collector. These holes form a positive space charge or cloud within the germanium in the region of the conducting path and thereby evoke or attract additional electrons from nearby sites, lowering the resistance of the path.

The significant point here is that the advent of the holes causes a decrease in the collector-to-base resistance due both to the effect of the holes in annihilating the surface charge and due to the positive space-charge effect. The **result is** that the resistance of the collector circuit is significantly decreased and a current flows that may be two to three times the emitter current. However, this clearly does not explain a resistance gain.

The condition of low resistance and high current is shown in section AB of Fig. 6. In the region from A to B the current is seen to be proportional to the voltage in accordance with Ohm's law; an increase in voltage is accompanied by a proportionate increase in current. If the applied collector bias is increased beyond point B the available supply of electrons soon becomes inadequate to sustain an Ohm's-law current. The supply of electrons is controlled by:

(1) The normal number of majority carriers at room temperature. This is a function of the impurity content.

(2) The neutralizing action of the holes on the surface electrons.

(3) The positive space charge due to the diffused holes.

Above point B of the figure, for instance, increases of applied potential V_c are not followed by proportionate increases of the current due to this exhaustion of the available carriers or electrons, and the curve rises steeply as shown in Fig. 6. The slope of the curve in the figure is r_{22}, the collector-to-base resistance.

The transistor is normally operated at approximately point C of the figure where the collector-to-base resistance is high. Summarizing, in the collector circuit the reason for the high resistance is not alone the reverse-current connection of the bias, but also the exhaustion of the available current carriers beyond the Ohm's law limit.

Typical Values

Let $i_e r_{11}$ represent the internal voltage drop from emitter to base and $i_c r_{22}$ the internal voltage drop from collector to base. Since i_c is greater than i_e due to the effect of the holes, and r_{22} is greater than r_{11} due to the carrier-exhaustion effect mentioned, the possible voltage gain of the transistor is the product of current gain and resistance gain. It is important to note that voltage and power gain in the point-contact transistor are due both to the current gain and resistance gain.

The ratio of i_c to i_e is usually denoted by the Greek letter α (alpha), and for typical point-contact transistors now on the market is of the order of 2.5. Typical values for r_{11} and r_{22} are approximately 300 and 18,000 ohms respectively, representing a resistance gain of about 60.

Thus it may be seen that typical voltage gains are of the order of 150 and typical power gains are of the order of 400.

In their migration from the emitter to the collector the holes move in an n-type material, and recombinations of holes and electrons are the rule rather than the exception. In fact, the discussion regarding recombination should explain why a very important effort in solid state physics is directed toward improving and controlling the lifetime of these carriers. Single crystals of germanium that are to be used in the manufacture of transistors are compared and evaluated for their lifetimes as a production control. Lifetime of carriers in single-crystal germanium is an important design parameter. Typical values vary from a few microseconds to two or three thousand microseconds.

Frequency Response

The velocity of holes and electrons in semiconductors is not the same as the velocity of electrons in conductors. The reader accustomed to thinking of electron velocities in orders of hundreds of thousands of centimeters per second may be surprised to learn that in semiconductors velocities are of the order of a few thousand centimeters per second. Typical values for germanium are: for electrons, about 3,600 cm per second, for each volt per cm of potential difference or as is sometimes said, of potential gradient; for holes, about 1,700 cm per second per volt per cm of gradient.

Some very important effects in the transistor ensue as a result of this relatively slow movement of holes and electrons in semiconductors. For instance, in a wire information in the form of modulated waves piling up when fed into one end is not a matter of concern. The extremely high velocity of the electrons in the wire that carry the intelligence makes it almost certain that, for any reasonable frequency, the information fed in will move out of the way before the next bit is fed in. In semiconductors, however, because the velocity is so low, this piling up presents a serious limitation. At relatively low frequencies serious modulation distortion results because of the piling up of intelligence connected with the long transit time of electrons and holes.

Another consequence of the large transit time in transistors is evident in the decrease of the current gain, alpha, with frequency. When the period of one cycle is equal to the transit time, the positive half of a sinusoid may still be within the emitter-collector region when the negative half of the same cycle enters. Under these conditions, the effect on the flow of holes due to the positive and negative halves of the cycle cancel, and current gain is zero. This is for the extreme case when frequency is the reciprocal of the transit time. For lower frequencies, the effects are in proportion and the observed fact is that the current gain decreases with frequency.

The decrease of alpha with frequency is also due to the fact that all the flow lines of the holes from emitter to collector are by no means straight lines. In fact, these flow paths resemble more closely the flux lines observed in pictures of the magnetic field between two poles. Since the transit time for these circuitous paths is greater than for the straight paths, neutralizations of the effects of the holes occur, particularly at the higher frequencies (lower periods).

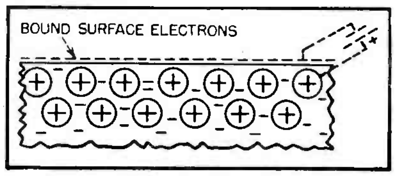

FIG. 5—A layer of electrons may be considered to exist in bound states at the surface of a semiconductor. This layer, in conjunction with the array of donors just below the surface, produces in effect the potential hill represented by the dashed-line battery

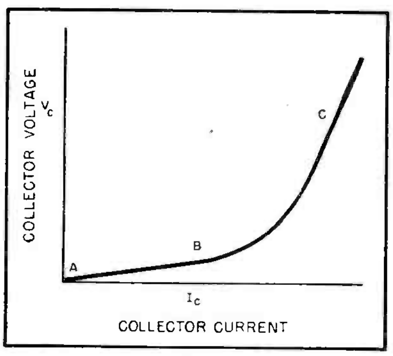

FIG. 6—Typical collector characteristic curve indicating variation of collector current with applied bias voltage. After point *B* insufficient carriers are present to support a proportional current for the applied electric field and the apparent resistance increases rapidly

While figures of frequency cutoff as high as 300 mc have been reported in the literature, these must be regarded as development laboratory values rather than representative limits obtainable from transistors now on the market. For most point-contact transistors currently being produced upper frequency limits are in the neighborhood of 10 mc, typical values around 3 mc. Some remarkable progress is being made along these lines at present and the reader can confidently expect commercially available units with importantly extended frequency limits to appear on the market in the near future.

This frequency limitation does point out, however, the desirability of having the emitter and collector as close together as possible because the frequency response increases rapidly as the spacing is decreased. Mechanical and electrical counter-indications set a lower limit on this spacing.

Resume

(1) The point-contact transistor is capable of current gain due to the action of holes in enhancing the collector current for a given amplitude of emitter current.

(2) Voltage and power gain of the point-contact transistor are due to a combination of current and resistance gain.

(3) The theory of operation of point-contact devices in general is, as yet, not completely understood and it is essential to keep abreast of the literature in order to observe the progress in this field.

(4) The concept of potential hills is essential for the discussion of transistor action and is useful as a convenient mnemonic by which the polarity of applied biases may be remembered.

(5) The frequency response of point-contact transistors is a function of the transit time and the distribution of the flow lines of the carriers from emitter to collector.