Hole Injection in Germanium—Quantitative Studies and Filamentary Transistors*

By

W. SHOCKLEY, G. L. PEARSON and J. R. HAYNES

Holes injected by an emitter point into thin single-crystal filaments of germanium can be detected by collector points. From studies of transient phenomena the drift velocity and lifetimes (as long as 140 microseconds) can be directly observed and the mobility measured. Hole concentrations and hole currents are measured in terms of the modulation of the conductivity produced by their presence. Filamentary transistors utilizing this modulation of conductivity are described.

1. Introduction

THE invention of the transistor by J. Bardeen and W. H. Brattain[1, 2, 3] has given great stimulus to research on the interaction of holes and electrons in semiconductors. The techniques discussed in this paper for investigating the behavior of holes in *n*-type germanium were devised in part to aid in analyzing the emitter current in transistors. The early experiments suggested that the hole flow from the emitter to the collector took place in a surface layer.[1, 2] The possibility that transistors could also be produced by hole flow directly through *n*-type material was proposed in connection with the *p-n-p* transistor.[4] Quite independently, J. N. Shive[5] obtained evidence for hole flow through the body of *n*-type germanium by making a transistor with points on opposite sides of a thin germanium specimen. Such hole flow is also involved in the coaxial transistor of W. E. Kock and R. L. Wallace.[6] Further evidence for hole injection into the body of *n*-type germanium under conditions of high fields was obtained by E. J. Ryder.[7]

In keeping with these facts it is concluded[3] that with two points close together on a plane surface, as in the type-A transistor[8], holes may flow either in a surface layer or through the body of the germanium. For surface flow to be large, special surface treatments appear to be necessary; such treatments were not employed in the experiments described in this paper and the results are consistent with the interpretation that the hole current from the emitter point flows in the interior.

The experiments described in this paper, in addition to any practical implications, serve to put the action of emitter points on a quantitative basis and to open up a new area of research on conduction processes in semicon-

* It is planned to incorporate this material in a book entitled "Holes and Electrons, an Introduction to the Physics of Transistors" currently being written by W. Shockley. This book is to cover much of the material planned for the "Quantum Physics of Solids" series which was discontinued because of the war.

ductors. It is worth while at the outset to contrast some of the new aspects of these experiments with the earlier experimental status of the bulk properties of semiconductors. Prior to the invention of the transistor, inferences about the behaviors of holes and electrons were made from measurements of conductivity and Hall effect. For both of these effects, under essentially steady state conditions, measurements were made of such quantities as lengths, currents, voltages and magnetic fields. The measurement of time was not involved, except indirectly in the calibration of the instruments. Nevertheless, on the basis of these data, definite mental pictures were formed of the motions of holes and electrons describing in particular their drift velocity in electric fields and the transverse thrust exerted upon them by magnetic fields. The new experiments show that something actually does drift in the semiconductor with the predicted drift velocity and does behave as though it had a plus or minus charge, just as expected for holes and electrons. In addition, experiments described elsewhere[9] show that the effect of sidewise thrust by a magnetic field actually is observed in terms of the concentration of holes and electrons near one side of a filament of germanium.

We shall discuss here evidence that holes are actually introduced into n-type germanium by the forward current of an emitter point and show how the numbers and lifetimes of the holes can be inferred from the data. We shall refer to this important process as "hole injection." Discussions of the reasons why an emitter should emit holes are given for point contacts by J. Bardeen and W. H. Brattain[1, 2, 3] and for p-n junctions elsewhere in this journal.[4] There are other possible ways in which semiconductor amplifiers can be made without the use of hole injection into n-type material or electron injection into p-type material.* In this paper, however, our remarks will be restricted to semiconductors which have only one type of carrier present in appreciable proportions under conditions of thermal equilibrium; for such cases the theoretical considerations are simplified and are apparently in good agreement with the experiments.

2. The Measurement of Density and Current of Injected Holes

The experiment in its semiquantitative form is relatively simple and is shown in Fig. 1.[10] A rod of n-type germanium is subjected to a longitudinal electric field E applied by a battery B_1. Collector and emitter point contacts are made to the germanium with the aid of a micromanipulator. The collector point is biased like a collector in a type-A transistor by the battery B_2 and the signal obtained across the load resistor R is applied to the input of an oscilloscope. At time t_1 the switch in the emitter circuit is closed so that a forward current, produced by the battery B_3, flows through the emitter point. At t_3 the switch is opened. The voltage wave at the collector, as

* For example see references 1 and 11.

observed on the oscilloscope, has the wave form shown in part (b) of the figure.

These data are interpreted as follows: When the emitter circuit is closed, the electrons in the emitter wire start to flow away from the germanium (i.e. positive current flows into the germanium). These electrons are furnished by an electron flow in the germanium towards the point of contact. The flow in the germanium may be either by the excess electron process or by the hole process. In Fig. 2 we illustrate these two possibilities. At first

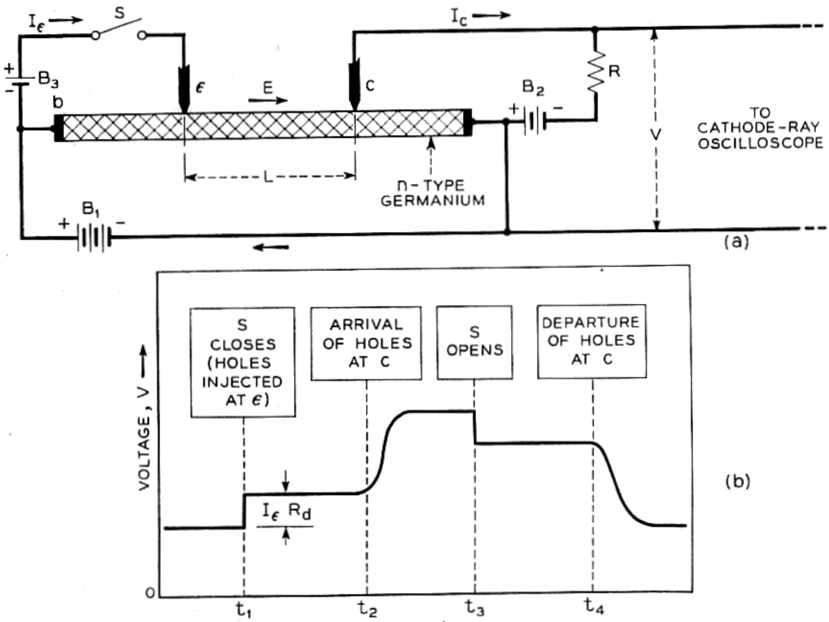

Fig. 1—Experiment to investigate the behavior of holes injected into n-type germanium
(a) Experimental arrangement.
(b) Signal on oscilloscope showing delay in hole arrival at t_2 in respect to closing S at t_1 and delay in hole departure at t_4 in respect to opening S at t_3.

glance it might appear that the difference between these two processes is unimportant since the net result in both cases is a transfer of electrons from the germanium to the emitter point. There is, however, an important difference, one which makes several forms of transistor action possible. In the case of the hole process an electron is transferred from the valence band structure to the metal. After this the hole moves deeper into the germanium. As a result the electronic structure of the germanium is modified in the neighborhood of the emitter point by the presence of the injected holes.

Under the influence of the electric field E, the injected holes drift toward the collector point with velocity $\mu_p E$, where μ_p is the mobility of a hole, and thus traverse the distance L to the collector point in a time $L/\mu_p E$. When they arrive at the collector point, they increase its reverse current and produce the signal shown at t_2.

There are two important differences between the signal produced at t_2 and that produced at t_1. The signal at t_1, which is in a sense a pickup signal, would be produced even if no hole injection occurred. We shall illustrate this by considering the case of a piece of ohmic material substituted for the

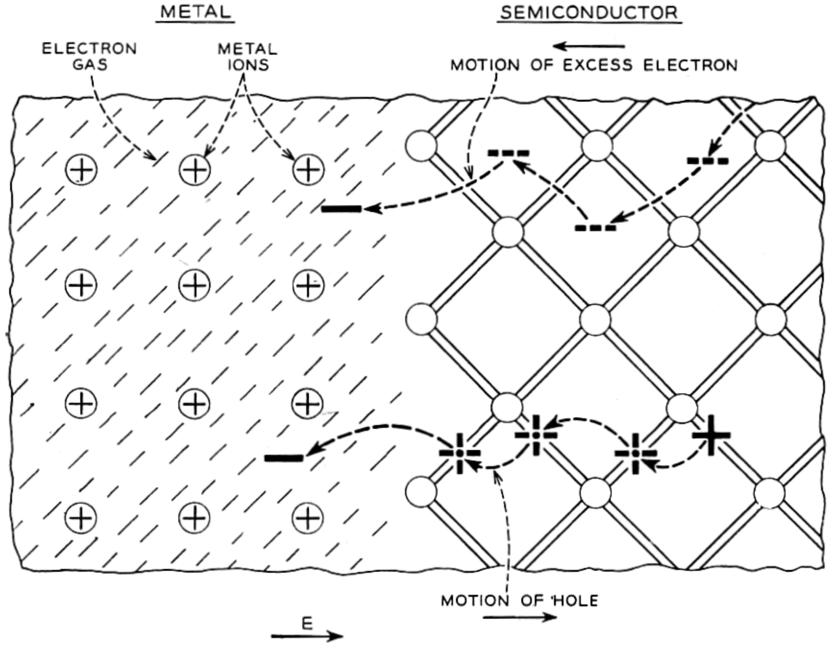

Fig. 2—Electron flow to the metal may be produced by an excess electron moving toward the metal or by bonding electrons jumping (dashed arrows) successively into a hole thus displacing the hole deeper into the semiconductor.

germanium. Conventional circuit theory applies to such a case; however, in order to contrast this purely ohmic case with that of hole injection, we shall also give a description of the conventional theory of signal transmission in terms of the motion of the carriers. According to conventional circuit theory, the addition of the current I_ϵ would simply produce an added IR drop due to current flow in the segment of the specimen to the right of the collector. This voltage drop is denoted as $I_\epsilon R_d$ in part (b), R_d representing the proper combination of resistances to take into account the way in which I_ϵ divides in the two branches. This signal will be transmitted from the emitter to the collector with practically the speed of light—the ordinary theory of signal transmission along a conductor being applicable to it. This high speed of transmission does not, of course, imply a correspondingly high velocity of motion of the current carriers. In fact the rapidity of signal transmission has nothing to do with the speed of the carriers and comes about as follows: If the ohmic material is an electronic conductor, then the withdrawal of a few electrons by the emitter current produces a local positive charge. This positive charge produces an electric field which progresses with the speed of light and exerts a force on adjoining electrons so that they move in to neutralize the space charge. The net result is that electrons in all parts of the specimen start to drift practically instantaneously. They flow into the specimen from the end terminals to replace the electrons flowing out at the emitter point and no appreciable change in density of electrons occurs anywhere within the specimen.*

The distinction between the process just described and that occurring when holes are injected into germanium is of great importance in understanding many effects connected with transistor action. One way of summarizing the situation is as follows: In a sample having carriers of one type only, electrons for example, it is impossible to alter the density of carriers by trying to inject or extract carriers of the same type. The reason is, as described above (or proved in the footnote), that such changes would be accompanied by an unbalanced space charge in the sample and such an unbalance is self-annihilating and does not occur.†

When holes are injected into n-type germanium, they also tend to set up a space charge. Once more this space charge is quickly neutralized by an electron flow. In this case, however, the neutralized state is not the normal thermal equilibrium state. Instead the number of current carriers present has been increased by the injected holes and by an equal number of electrons drawn in to neutralize the holes. The total number of electrons in the specimen will thus be increased, the extra electrons coming in from the metal terminals which complete the circuit with the emitter point. The presence of the holes and the neutralizing electrons near the emitter point modify the conductivity. As we shall show below, this modification of conductivity may be so great that it can be used to measure hole densities and also to give power gain in modified forms of the transistor. We shall summarize this situation as follows: *In a semiconductor containing substantially only one type of current carrier, it is impossible to increase the total carrier concentration by*

* This is a description in words of the result ordinarily expressed in terms of the dielectric relaxation time obtained as follows: $\nabla \cdot I = -\dot{\rho}$, $I = \sigma E = -\sigma \nabla \Psi$, $\nabla^2 \Psi = -4\pi\rho/\kappa = \dot{\rho}/\sigma$ so that $\rho = \rho_0 \exp[-(4\pi\sigma/\kappa)t]$, where I = current density, ρ = charge density, σ = conductivity, E = electric field, Ψ = electrostatic potential, κ = dielectric constant.

† In the case of modulation of conductivity by surface charges,[11] a net charge is produced by the field from the condenser plate. The changed charge density extends slightly into the specimen but should not be confused with the true volume effect of hole injection. Such space charge layers are discussed in other articles in this issue.[4, 12]

injecting carriers of the same type; however, such increases can be produced by injecting the opposite type since the space charge of the latter can be neutralized by an increased concentration of the type normally present.

Thus we conclude that *the existence of two processes of electronic conduction in semiconductors, corresponding respectively to positive and negative mobile charges, is a major feature in several forms of transistor action.*

In terms of the description given above, the experiment of Figure 1 is readily interpreted. The instantaneous rise at t_1 is simply the ohmic contribution due to the changing total currents in the right branch when the emitter current starts to flow. After this, there is a time lag until the holes injected into the germanium drift down the specimen and arrive at the collector. When the current is turned off at t_3, a similar sequence of events occurs.

The measured values of the time lag of $t_1 - t_2$, the field E and the distance L can be used to determine the mobility of the holes. The fact that holes, rather than electrons, are involved is at once evident from the polarity of the effect; the disturbance produced by the emitter point flows in the direction of E, as if it were due to positive charges; if the electric field is reversed, the signal produced at t_2 is entirely lacking. The values obtained by this means are found to be in good agreement with those predicted from the Hall effect and conductivity data. The Hall mobility values obtained on single crystal filaments of n- and p-type germanium[13] are

$$\mu_p = 1700 \text{ cm/sec per volt/cm}$$

$$\mu_n = 2600 \text{ cm/sec per volt/cm}$$

The agreement between Hall effect mobility and drift mobility, as was pointed out at the beginning of this section, is a very gratifying confirmation of the general theoretical picture of holes drifting in the direction of the electric field.

We shall next consider a more quantitative embodiment of the experiment just considered. In Fig. 3, we show the experimental arrangement. In this case it is essential in order to obtain large effects that the cross-section of the germanium filament be small. A thin piece of germanium is cemented to a glass backing plate and is then ground to the desired thickness. After this the undesired portions are removed by sandblasting while the desired portions are protected by suitable jigs consisting of wires, scotch tape, metal plates, etc. After the sandblasting, the surface of the germanium is etched. In this way specimens smaller than 0.01×0.01 cm in cross-section have been produced. The ends of the filament are usually made very wide so as to simplify the problem of making contacts.

Under experimental conditions, a battery like B_1, of Figure 1 applies a "sweeping" field in the filament so that any holes injected by the emitter current are swept along the filament from left to right. In the small filaments used for these experiments, the resulting concentration of holes is so high that large changes in conductivity are produced to the right of the emitter point and, as we shall describe below, these changes can be measured and the results used to determine the hole current at the emitter point. In order to treat this situation quantitatively, we introduce a quantity γ defined as follows:

γ = *the fraction of the emitter current carried by holes.*

Accordingly, a current γI_ϵ of holes flows to the right from ϵ and produces a hole density, denoted by p, which is neutralized by an equal added electron density. A fraction $(1 - \gamma)I_\epsilon$ of electrons flows to the left; these electrons do not, however, produce any increased electron density to the left of the emitter since they are of the sign normally present in the n-type material. The presence of the holes to the right in the filament increases the conductivity σ (as shown in Fig. 3c) both because of their own presence and the presence of the added electrons drawn in to neutralize the space charge of the holes. The mobility of electrons is greater than the mobility of holes, the ratio being[13]

$$b = \mu_n/\mu_p = 1.5 \text{ for germanium}$$

and the electrons are always more numerous than the holes*

$$n = n_0 + p, \tag{2.1}$$

where n_0 is the concentration of electrons which would be present to neutralize the donors if p were equal to zero; consequently, the current carried by electrons is greater than the current carried by holes. The concentration of holes diminishes to the right due to the fact that holes may recombine with electrons as they flow along the filament.

From this experiment the value of γ and the lifetime of a hole in the filament can be determined. The measurements are made with the aid of the two probe points P_1 and P_2. The conductance of the filament between these points is obtained by measuring the voltage difference ΔV and dividing it into the current $I_b + I_\epsilon$, no current being drawn by the probes themselves. The necessary formulae for calculating hole density and hole current,

* The notation used in the equations is as follows: n, p, n_0 = respectively density of electrons, of holes, of electrons when no holes are injected. N_d and N_a are the densities of donors and acceptors, assumed ionized so that $n_0 = N_d - N_a$. I_ϵ, I_b, I_c are as shown on Figs. 3 and 9. (I_c used for the probe collector in Figures 1 and 8 does not enter the equations.) $q = |q|$ is the charge on the electron, used to be consistent with Ref. 4, where e is used for $2.718 \cdots$.

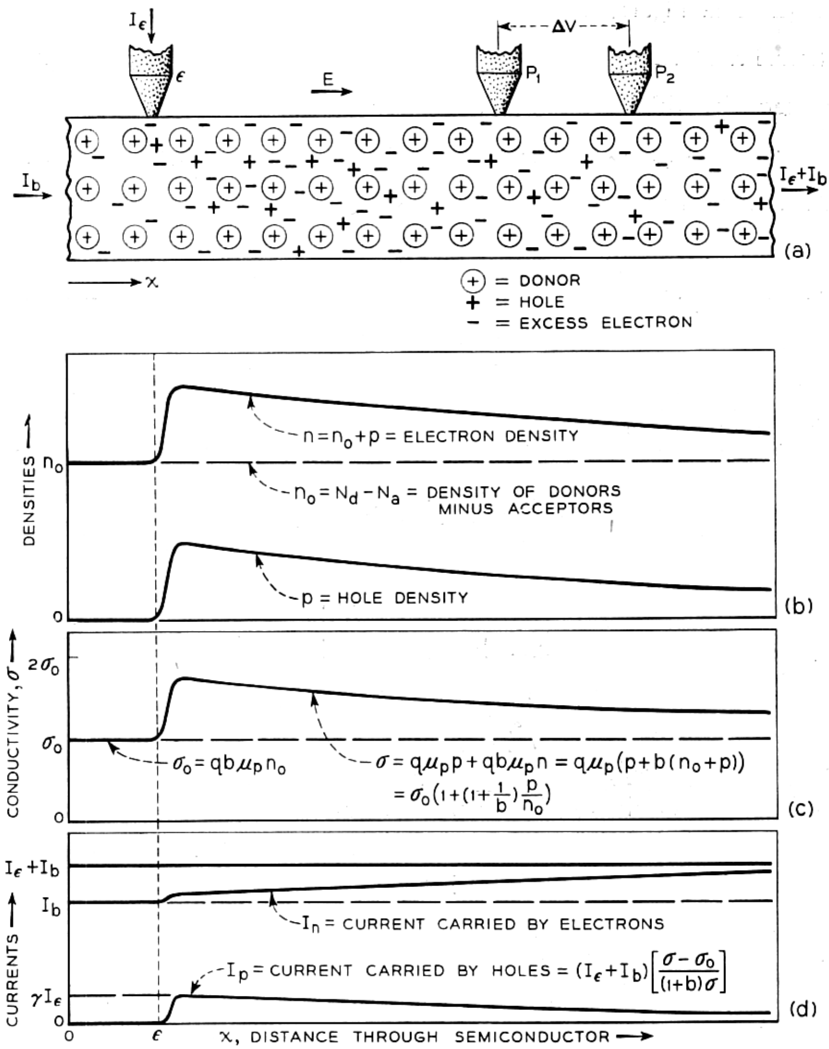

Fig. 3—Method of measuring hole densities and hole currents.

(a) Distribution of holes, electrons and donors. Acceptors, which may be present, are omitted for simplicity, the excess of donor density N_d over acceptor density N_a being n_0.

(b) To the right of the emitter the added hole density p is compensated by an equal increase in electron concentration.

(c) The conductivity is the sum of hole and electron conductivities.

(d) The total current $I_b + I_\epsilon$ to the right of the emitter is carried by I_p and I_n in the ratio of the hole to the electron conductivity.

shown on the Figure, are derived as follows:

Normal conductivity $\sigma_0 = q\mu_n n_0$, (2.2)

conductivity with holes present $\sigma = q\mu_n n + q\mu_p p$

$$= q\mu_n(n_0 + p) + q\mu_p p = \sigma_0[1 + (1 + b^{-1})(p/n_0)]. \tag{2.3}$$

The conductance,

$$G = (I_\epsilon + I_b)/\Delta V,$$

between P_1 and P_2 is proportional to the local conductivity, and hence to

$$1 + (1 + b^{-1})(p/n_0),$$

so that a measurement of the conductance gives a measurement of p/n_0. Letting G and G_0 be the conductances between the points with and without hole injection, we have

$$\frac{G}{G_0} = \frac{\sigma}{\sigma_0} = 1 + (1 + b^{-1})(p/n_0) \tag{2.4}$$

or

$$\frac{p}{n_0} = \frac{\sigma - \sigma_0}{\sigma_0(1 + b^{-1})} = \frac{(G/G_0) - 1}{1 + b^{-1}}. \tag{2.5}$$

The ratio of hole current to electron current is $q\mu_p p/q\mu_n n$ and the fraction of the current carried by holes is thus

$$\begin{aligned} \frac{I_p}{I_n + I_p} &= \frac{q\mu_p p}{q\mu_n n + q\mu_p p} = \frac{p}{bn_0 + (1 + b)p} \\ &= \frac{p/n_0}{b[1 + (1 + b^{-1})(p/n_0)]} = \frac{1-(G_0/G)}{1 + b} \end{aligned} \tag{2.6}$$

Hence from the measured values of G, it is possible to obtain the fraction of the current carried by holes. Multiplying this fraction by $I_\epsilon + I_b$ then gives the actual hole current flowing past the probe points.* If there were no decay, the current past the probe points would be γI_ϵ and since I_ϵ is known, γ could be easily determined. Actually, however, there may be quite an appreciable decay. However, if the current I_b is increased, the holes will be swept more rapidly from the emitter to the probes and less decay will result. Thus by increasing I_b, the effect of recombination can be minimized and the value of hole current can be extrapolated to the value it would have in the absence of decay. This value is, of course, γI_ϵ.

* In these calculations the formulae $n = p + n_0$, corresponding to completely ionized donors and acceptors, has been used. In germanium this is a good approximation. For silicon, however, modifications will be necessary.

In Fig. 4 we show some plots of this sort. The ordinate is I_p/I_ϵ which should approach γ as the value of I_b becomes larger. The theory indicates that a logarithmic plot should be used and that the abscissa should be made proportional to transit time so that the case of no decay or zero transit time

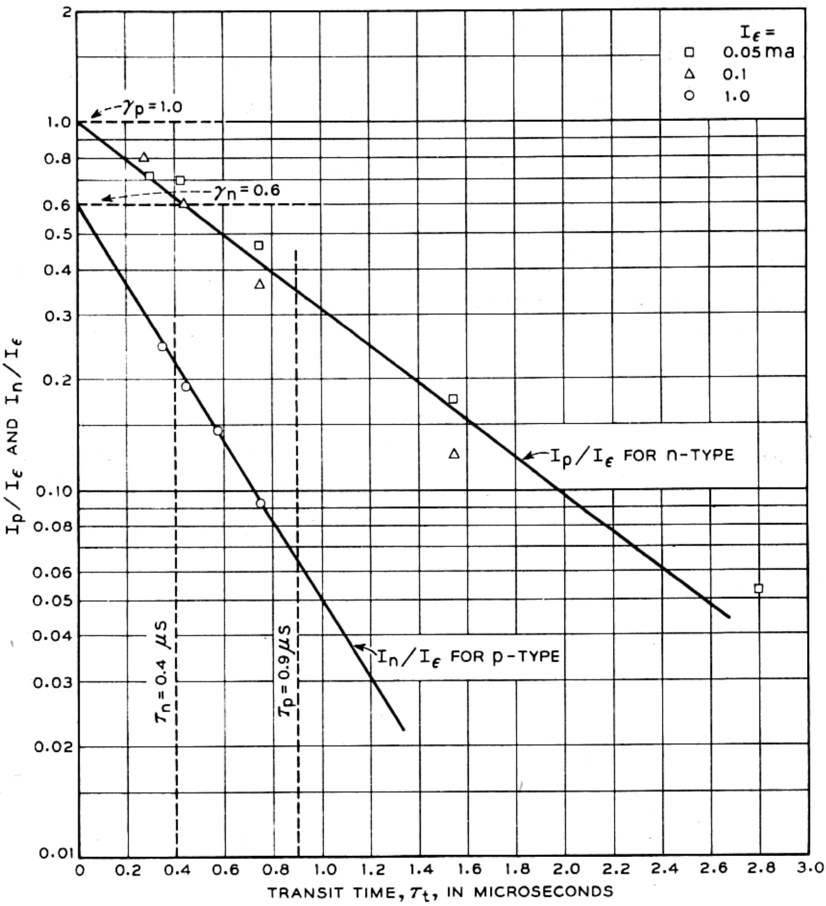

Fig. 4—Extrapolation of measured hole and electron currents to zero transit time in order to determine γ.

comes at the left edge.† The conclusion reached from this plot is that for the case of the n-type sample, the value of γ is substantially unity,—all the

† If the lifetime of a hole is τ, then the hole current at the points is $I_p = \gamma I_\epsilon \exp(-t/\tau)$ where t is the transit time to a point midway between the points, say a distance L from the emitter. If the electric field is $E = \Delta V/\Delta L$, then the transit time $t = L\Delta L/\mu_p \Delta V$. Hence if $\ln I_p$, as determined from the ratio of conductivities is plotted against $t = L\Delta L/\mu_p\Delta V$ a straight line with intercept $\ln \gamma I_\epsilon$ and slope $-1/\tau$ should be obtained.

emitter current is holes. For the opposite case in which electrons are injected into *p*-type material,[14] the corresponding value of I_n/I_ϵ extrapolates to 0.6 indicating that for this case 60% of the current is carried by electrons and 40% by holes. For these particular specimens the lifetimes are found to be 0.9 and 0.41 microseconds respectively. There is a body of evidence, some of which we discuss below, that holes combine with electrons chiefly on the surface of the filament.

3. The Influence of Hole Density on Point Contacts

The presence of holes near a collector point causes an increase in its reverse current; in fact the amplification in a type-A transistor is due to the modulation of the collector current by the holes in the emitter current. The influence of hole density upon collector current has been studied in connection with experiments similar to those of Fig. 3. After the hole current and the hole density are measured, a reverse bias of 20 to 40 volts is applied. The reverse current is found to be a linear function of the hole density. Figure 5 shows typical plots of such data. Different collector points, as shown, have quite different resistances. However, once data like that of Fig. 5 have been obtained for a given point, the currents can then be used as a measure of hole density. This experimental procedure for determining hole density is simpler than that involved in using the two points and much better adapted to studies of transient phenomena. It is necessary in employing this technique to keep the current drawn by the collector point somewhat smaller than $I_b + I_\epsilon$; otherwise the disturbance in the current flow due to the collector current is too great and the sample of the hole current is not representative. Experiments have shown, however, that this condition is readily achieved and that the collector current may be satisfactorily used as a measure of hole density.

The hole density also affects the resistance of a point at low voltage. Studies of this effect have also been made in connection with the experiment of Fig. 3. After the hole density has been determined from measurements of ΔV and $I_b + I_\epsilon$, a small additional voltage (0.015 volts) was applied between P_1 and P_2 and the current flowing externally between P_1 and P_2 was measured. From these data a differential conductance, for small currents, is obtained for the two points P_1 and P_2 in series. As is shown in Fig. 6, this conductance is seen to be a linear function of the hole concentration. The conductance of a point contact arises in part from electron flow and in part from hole flow. From experiments using magnetic fields[9], it has been estimated that under equilibrium conditions the two contributions to the conductance may be comparable. In connection with Fig. 6 it should be noted that the hole concentration on the abscissa is the average hole concentration throughout the entire cross section; the hole concentration may be much less near the surface due to recombination on the surface.

Techniques of the sort described above can be used to measure the properties of collector points. If a collector point is placed between the emitter and P_1 in Fig. 3, then the hole current extracted by the collector can be determined in terms of the hole current past P_1 and P_2. By these means an "intrinsic α" for the collector point can be determined. The intrinsic α is

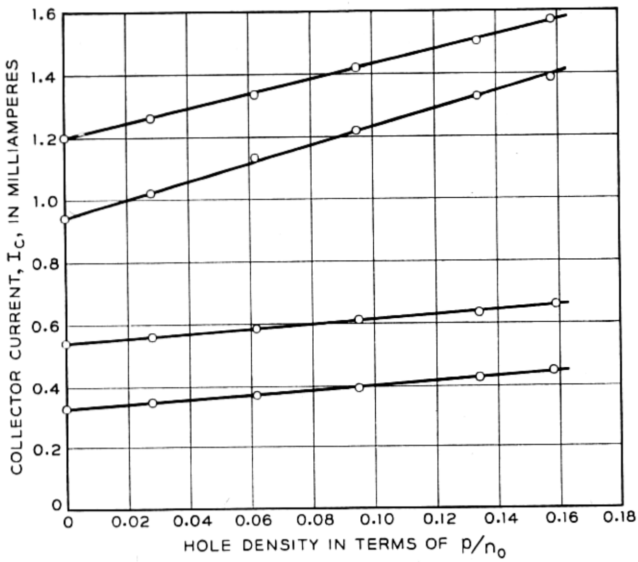

Fig. 5—Dependence of collector current I_c upon average hole density being swept by collector point. Collector biased 20 volts reverse.

defined as the ratio of change in collector current per unit change in hole current actually arriving at the collector.

4. Studies of Transient Phenomena

The technique of using a collector point to measure hole concentrations has been employed in a number of experiments similar to those described in connection with Fig. 1. These experiments give information concerning hole lifetimes, hole mobilities, diffusion and conductivity modulation.

One of the methods employed to measure hole lifetime involves the measurement of the increase in collector current, produced by the arrival of the leading edge of the hole pulse, as a function of the transit time of the holes from emitter to collector. This time is varied by varying the distance between the emitter and the collector points.

In Fig. 7 we show a plot, obtained in this way, from a sample of germanium having dimensions 1.0 × .05 × .08 cm. It is seen that the increase in collector current due to hole arrival decays exponentially with a time constant of 18 microseconds. This time constant increases as the dimensions of the germanium sample are increased so that a time constant of 140 microseconds was measured, using a sample having dimensions 2.5 × .35 × .30

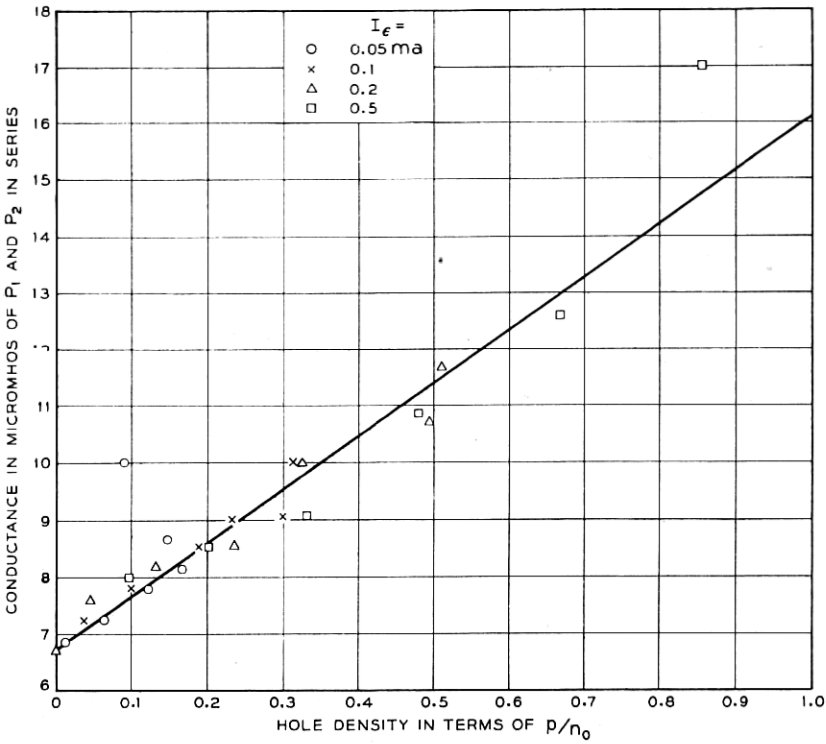

Fig. 6—Conductance of P_1 and P_2 of Fig. 3 in series as a function of p/n_0, showing that conductance depends on hole concentration but not on currents in filament. For each value of I_ϵ the hole density was varied by varying $I_b + I_\epsilon$ from .038 to 0.78 ma.

cm. Since the holes injected into the interior of this sample can diffuse to the surface and recombine in about 100 microseconds, the process may still be largely one of surface recombination. In any event, it may be concluded that the lifetime in the bulk material used must be at least 140 microseconds. Making use of the electron density determined from other measurements, we conclude that the recombination cross section must be less than 10^{-18} cm^2. This cross section, which is less than 1/400 the area of a germanium atom, may be so small because a hole-electron pair has difficulty in satisfying in the crystal the conditions somewhat analogous to conservation of energy and momentum which hinder recombination of electrons and positive ions in a gas discharge. Thus it has been pointed out that a hole-electron pair will have a lowest energy state in which the two current carriers behave something like the proton and electron of a hydrogen atom.[15] Such a bound pair are called an exciton and the energy given up by their recombination is the "exciton energy." In order to recombine they must radiate this energy in

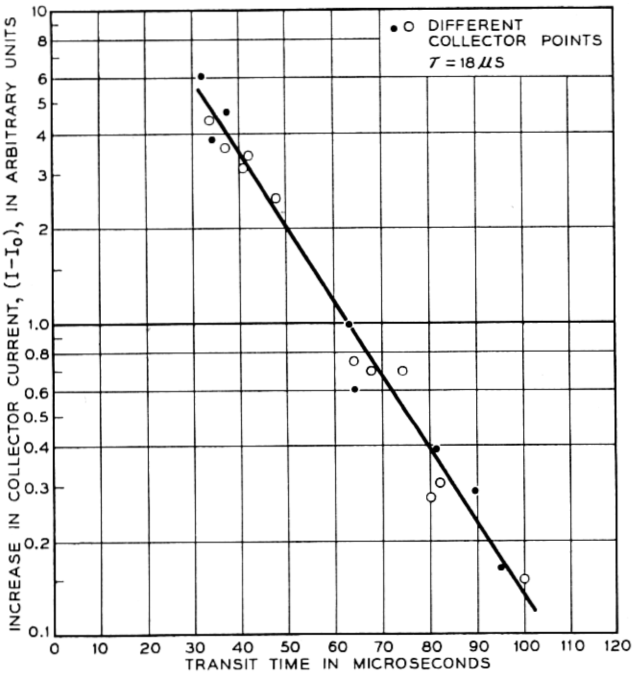

Fig. 7—The decay of injected holes in a sample of *n*-type germanium.

the form of a light quanta (photon) or a quantum of thermal vibration of the crystal lattice (phonon). The recombination time for the photon recombination process can be estimated from the optical constants for germanium and the theory of radiation density using the principle of detailed balancing, which states that under equilibrium conditions the production of hole electron pairs by photon absorption equals the rate of recombination with photon emission; the lifetime obtained in this way is about 1 second at room temperature indicating that the photon process is unimportant.[19] As has been pointed out by A. W. Lawson,[16] the highest energy phonon will have insufficient energy to carry away the "exciton energy" of a hole-electron pair and, therefore, the release of energy will require the cooperation of several phonons with a correspondingly small transition probability.

When a square pulse of holes is injected in an experiment like that of Fig. 1, the leading and trailing edges of the current at the collector point are deformed for several reasons. Due to the high local fields at the emitter point, some of the holes actually start their paths in the wrong direction—i.e. away from the collector; these lines of flow later bend forward so that those holes also pass by the collector point but with a longer transit time than holes which initially started towards the collector. A spread in transit times of this sort is probably largely responsible for the loss of gain at high frequencies in transistors. For the experiments described below, however, this effect is negligible compared to two others which we shall now describe.

On top of the systematic drift of holes in the electric field, there is superimposed a random spreading as a result of their thermal motion. This would cause a sharp pulse of holes to become spread so that after drifting for a time t_d the hole concentration would extend over a distance proportional to $\sqrt{Dt_d}$ where D, the diffusion constant for holes, $= kT\mu_p/q = 45$ cm²/sec. As a result of this effect, the leading and trailing edges of the square wave of emission current become spread out when they arrive at the collector. This is shown in Fig. 8, curve A for the leading edge and B for the trailing edge. The points are 10 microsecond marker intervals traced from an oscilloscope, the time being measured from the instant at which the emitter current starts. For A and B the emitter current was so small compared to the current I_b that the holes produced a negligible modulation of conductivity and each hole moved in essentially the same electric field. It is to be observed that the wave shapes are nearly symmetrical in time about the half rise point and that the A and B waves are identical except for sign. This is just the result to be expected from diffusion. Furthermore, analysis shows that the spread in arrival time is in good quantitative agreement with the theoretical wave shape using the diffusion constant appropriate for holes. For this case the mid-point of the rise, corresponding to the crossing point of the curves, gives the average arrival time and has been used to obtain an accurate measure of the mobility.

Curves C and D correspond to conditions in which the emitter current was relatively large—two thirds of the base current. High impedance sources are used so that I_b is constant and I_e is a good flat topped wave. For the currents used in this experiment, the conductivity is appreciably modulated by the presence of holes. This accounts for the shape of curve C, corresponding to the arrival of holes at the collector. It is seen that this curve is not symmetrical but is much more gradual towards later times. The reason for this is that the first holes to arrive are those which have diffused somewhat ahead of the rest and move in material of low conductivity. The later holes travel in an environment of relatively high conductivity and, consequently, in a lower electric field. (Since the current is the same at all points between emitter and collector, the field is inversely proportional to the conductivity.) The transit time for the later holes is, therefore, longer and the hole density builds up more slowly for the latter part of the incoming pulse of holes. The wave form obtained from the trailing edge of the emitter pulse, curve D, is in striking contrast with the leading edge. The first gradual decay, up to

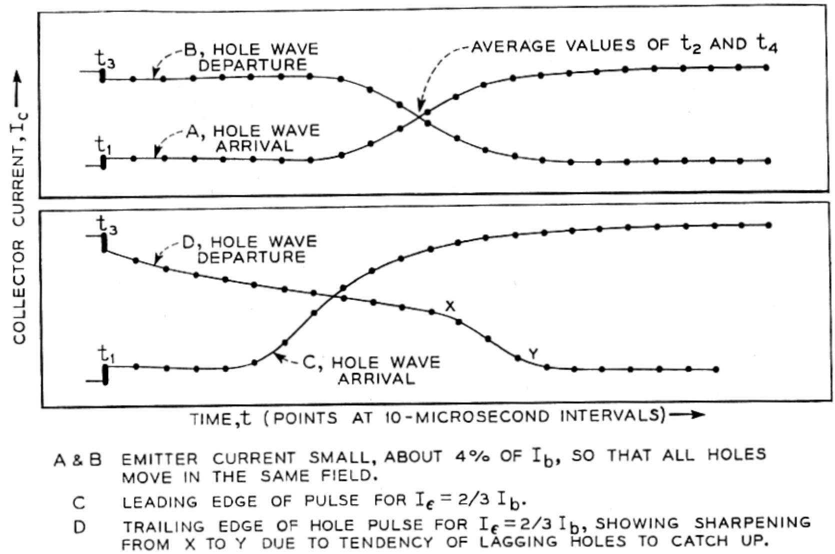

Fig. 8—Collector current characteristics for the circuit shown in Fig. 1.

point X, is due to recombination of holes and electrons; at t_3 the emitter current becomes zero; consequently, the electric field is reduced and the holes arriving at X have taken a longer transit time than the holes arriving at t_3 and a larger fraction of them have recombined with electrons. The true trailing edge, running from X to Y, is appreciably sharper than the leading edge. The reason for this is that holes lagging behind the main body of holes are in a region of relatively low conductivity and high electric field and tend to catch up with the main body. Thus the same effect which lengthens wave C acts to shorten wave D.

C. Herring has been able to obtain mathematical solutions for the appropriate equations bearing on the matters just discussed. His theory is presented elsewhere in this issue.[17]

The delay feature discussed in connection with Figs. 1 and 8 indicates interesting possibilities of using germanium filaments as delay or storage elements.

5. The Theory of the Filamentary Transistor

In Fig. 9 we show a transistor with a filamentary structure.[18] Modulation is achieved in this case by injecting holes at the emitter point which flow to the right and modulate the resistance in the output branch between emitter and collector. Structures of this sort can be produced by the sand-blasting technique discussed in Section 2. The enlarged ends, which give the

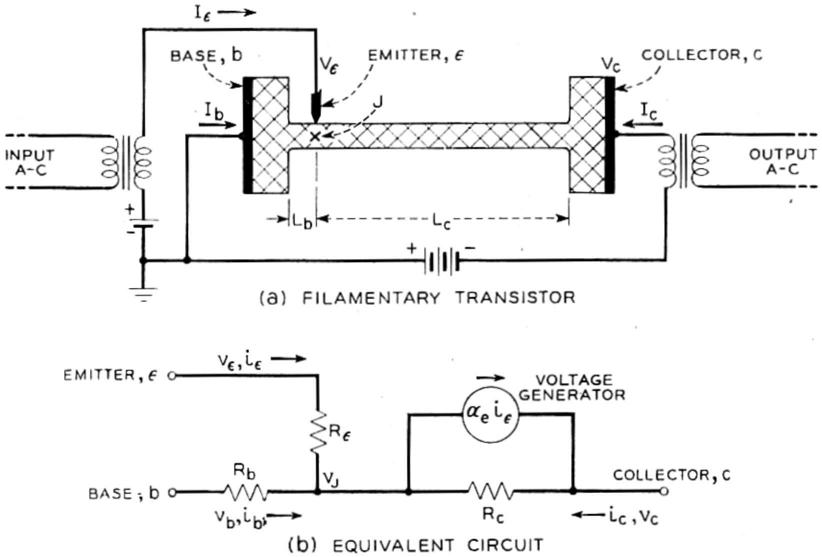

Fig. 9—Filamentary transistor and equivalent circuit.

unit a dumbbell appearance, decrease the problem of making contact to the unit. The large area at the left side serves the additional purpose of reducing unwanted hole emission from the metal electrode and affords an opportunity for any emitted holes to recombine before they enter the narrow part of the unit.

The theory of this transistor is relatively simple and most of the features we shall discuss in connection with it have counterparts in the theory of the type-A transistor. We shall discuss the case for which the injected current is a small fraction of the total current in the filament. Under these conditions we can use a simple linear theory. We shall show that the behavior of the transistor can be given for small a-c signals by the equivalent circuit in

Figure 9(b), which shows the current and voltage relationships in a form equivalent to those used in connection with the type-A transistor.

The point J in Fig. 9 represents a point in the filament near the emitter point. The current from the emitter point will be determined by the difference between its voltage V_ϵ and that of the surrounding semiconductor, namely the voltage at J. Thus we can write

$$I_\epsilon = f_\epsilon(V_\epsilon - V_J). \tag{5.1}$$

For small a-c variations, i_ϵ, v_ϵ and v_J, this equation leads to the relationship

$$i_\epsilon = (v_\epsilon - v_J)f'_\epsilon, \tag{5.2}$$

where f'_ϵ is the derivative of f_ϵ in respect to its argument. Letting $f'_\epsilon = 1/R_\epsilon$ this equation becomes

$$v_\epsilon - v_J = R_\epsilon i_\epsilon. \tag{5.3}$$

This relationship is correctly represented by the R_ϵ branch of the equivalent circuit. The voltage at J, under the assumed operating conditions with I_ϵ positive and much less that I_c, will be $-I_b R_b$ where R_b is the resistance from the base to an imaginary equipotential surface passing through J and $v_b = 0$, corresponding to *grounded base* operation. This leads to

$$v_J = -R_b i_b = +R_b i_\epsilon + R_b i_c, \tag{5.4}$$

since $i_b + i_\epsilon + i_c = 0$. This relationship is obviously satisfied by the R_b branch of the equivalent circuit.

We now come to the collector branch which we have represented as a resistance R_c and a parallel current generator* $\alpha_e i_\epsilon$. (This circuit is equivalent to another in which the parallel current generator is replaced by a series voltage generator $\alpha_e R_c i_\epsilon$.) We must show that this part of the equivalent circuit represents correctly the effect of injecting holes into the right arm of the filament. We shall suppose that there is negligible recombination so that the hole current injected at the emitter point flows through the entire filament. (We consider recombination in the next section.) The current I_c in the collector branch thus contains a component $-\gamma I_\epsilon = I_p$ of hole current (minus because of the algebraic convention that positive $I_c(= -I_b - I_\epsilon)$ flows to the left). The added hole and electron concentrations lower the resistance and R_c changes to $R_c + \delta R_c$, where δR_c is negative. The current voltage relationship for this branch of the filament then becomes

$$V_c - V_J = (R_c + \delta R_c)I_c. \tag{5.5}$$

* α_e in the equivalent circuit differs from $\alpha = -(\partial I_c/\partial I_\epsilon)v_c$ by the relationship $\alpha_e = \alpha + (\alpha - 1)(R_b/R_c)$, equivalent to equation (6.8).

Our problem is to reexpress this relationship in terms of the small a-c components and show that it reduces to the relationship

$$v_c - v_J = R_c(i_c + \alpha_e i_\epsilon) \tag{5.6}$$

corresponding to the equivalent circuit. For small emitter current the analysis is carried out conveniently as follows: The ratio of hole current to the total current is $-\gamma I_\epsilon/I_c$. The ratio $(R_c + \delta R_c)/R_c$ corresponds to G_0/G discussed in connection with Fig. 3. The ratio of hole current to total current is given in (2.6) in terms of G_0/G and may be rewritten as

$$-\frac{\gamma I_\epsilon}{I_c} = \frac{1 - (G_0/G)}{1 + b} = \frac{-\delta R_c}{(1 + b)R_c}, \tag{5.6}$$

giving

$$\delta R_c = R_c(1 + b)\gamma\, I_\epsilon/I_c\,. \tag{5.7}$$

(Since I_c is negative and I_ϵ is positive this equation shows that δR_c is negative, i.e., the conductivity has been increased by the hole current.) Putting this value of $R_c + \delta R_c$ into the equation for $V_c - V_J$ gives

$$\begin{aligned} V_c - V_J &= (R_c + \delta R_c)I_c \\ &= R_c[I_c + (1 + b)\gamma I_\epsilon]. \end{aligned} \tag{5.8}$$

If we consider small a-c variations in the currents and voltages, this reduces to the equation given by the equivalent circuit with

$$\alpha_e = (1 + b)\gamma. \tag{5.9}$$

The data of Section 2 indicate that for holes injected into n-type germanium $\gamma = 1$, and since $b = 1.5$ we obtain $\alpha_e = 2.5$.

The quantity v_J can be eliminated by using $v_J = R_b(i_\epsilon + i_c)$ in equation (5.3) for v_ϵ and the small signal form of (5.8) for v_c leading to the pair of equations

$$v_\epsilon = (R_\epsilon + R_b)i_\epsilon + R_b i_c \tag{5.10}$$

$$v_c = (R_b + \alpha_e R_c)i_\epsilon + (R_c + R_b)i_c\,. \tag{5.11}$$

These equations are formally identical with those for the equivalent circuits of the type-A transistor.

It should be emphasized that although hole injection into n-type germanium plays a role in both the type-A and the particular form of filamentary transistor shown in Fig. 9, there are differences in the principles of operation. One important feature of the type-A is the high impedance of the rectifying collector contact which, however, does not impede hole flow and another important feature is the current amplification occurring at the collector contact. Neither of these features is present in the filamentary type shown. Instead, the high impedance at the collector terminal arises from the small cross-section of the filament. The modulation of the output current takes place through the change in body conductivity due to the presence of the added holes, a change which appears to be unimportant in the type-A transistor. In the filamentary type, current amplification is produced by the extra electrons whose presence is required to neutralize the space charge of the holes. Current amplification in the type-A transistor is, probably, also produced by the space charge of the holes[3] but the details of the mechanism are not as easily understood.

6. Effects Associated with Transit Time

Two important effects arise from the fact that a finite transit time is required for holes to traverse the R_c side of the filament: during this time the holes recombine with electrons and the modulation effect is attenuated for this reason; also the modulation of the conductivity of the filament at any instant is the result of the emitter current over a previous interval and for this reason there will be a loss of modulation when the period of the a-c signal is comparable with the transit time or less.

For the small signal theory, the effect of transit time is readily worked out in analytic terms. We shall give a derivation based on the assumption that the lifetime of a hole before it combines with an electron is τ_p. According to this assumption, the fraction of the holes injected at instant t_1 which are still uncombined at time t_2 is $\exp[-(t_2 - t_1)/\tau_p]$. This means that the effect in the filament at any instant t_2 is the average, weighted by this factor, of all the contributions prior to t_2 back to time $t_2 - \tau_t$ where τ_t is the transit time; holes injected prior to $t_2 - \tau_t$ have passed out of the filament by time t_2. If the emitter current is represented by $i_{\epsilon 0}e^{i\omega t}$, the effective average emitter current is

$$i_{\epsilon\,\mathrm{eff}}(t_2) = i_{\epsilon 0} \int_{t_2-\tau_t}^{t_2} e^{i\omega t_1 - (t_2 - t_1)/\tau_p}\, dt_1/\tau_t\,. \tag{6.1}$$

The term dt_1/τ_t is chosen so that a true average is obtained since the sum of all the dt_1 intervals add up to τ_t. The integral is readily evaluated and gives

$$i_{\epsilon\,\mathrm{eff}}(t_2) = i_{\epsilon 0}\, e^{i\omega t_2}\, \frac{1 - \exp\,[-\,i\omega\tau_t - (\tau_t/\tau_p)]}{i\omega\tau_t + (\tau_t/\tau_p)}\,. \tag{6.2}$$

The result so far as the equivalent circuit is concerned is that obtained by taking α_e as*

$$\alpha_e = \gamma(1 + b)\beta, \tag{6.3}$$

* The derivation of equations (5.10) and (5.11), describing the equivalent circuit, shows that hole injection enters only through the term $\delta R_c I_c$ in (5.8). This term leads only to $\alpha_e R_c i_\epsilon = (1 + b)\gamma R_c i_\epsilon$ in (5.11) and should be replaced by $(1 + b)\gamma R_c i_{\epsilon\,\mathrm{eff}} = (1 + b)\gamma\beta R_c i_\epsilon$ leading to (6.3).

where

$$\beta = \frac{1 - \exp\left[-i\omega\tau_t - (\tau_t/\tau_p)\right]}{i\omega\tau_t + (\tau_t/\tau_p)}. \tag{6.4}$$

β represents the effect of recombination and transit angle, $\omega\tau_t$, in reducing the gain.

We shall consider two limiting cases of this expression. First if $\omega\tau_t$ is very small, the new factor becomes

$$\beta = (\tau_p/\tau_t)(1 - e^{-\tau_t/\tau_p}). \tag{6.5}$$

If τ_t is much larger than τ_p, so that the holes recombine before traversing the filament, then the exponential is negligible and β becomes simply τ_p/τ_t. This means that the effectiveness of the holes is reduced by the ratio of their effective distance of travel to the entire length of the filament, i.e., τ_p/τ_t is the ratio of distance travelled in one lifetime to the entire length of the filament. Essentially the holes modulate only the fraction of the filament which they penetrate. The transit time depends on the field in the filament which is $| V_c - V_J |/L_c$, the absolute value being used since V_c is negative. The transit time is thus

$$\tau_t = L_c/[\mu_p \,| V_c - V_J | /L_c] = L_c^2/\mu_p \,| V_c - V_J |. \tag{6.6}$$

For very small emitter currents $V_c - V_J = R_c V_c/(R_c + R_b)$ so that

$$\tau_t = L_c^2(R_c + R_b)/\mu_p R_c |V_c| \tag{6.7}$$

and τ_t is inversely proportional to V_c. For large values of V_c, τ_t approaches zero and β approaches unity. The dependence of β upon V_c has been investigated by measuring α and plotting it as a function of $|1/V_c|$ as shown in Fig. 10. The value of

$$\alpha = -(\partial I_c/\partial I_\epsilon)_{V_c} \tag{6.8}$$

is readily found from the equivalent circuit, using equation (5.11), to be

$$\alpha = \frac{R_b}{R_b + R_c} + \frac{\alpha_e R_c}{R_b + R_c}. \tag{6.9}$$

For the particular structure investigated, the values of R_b and R_c, obtained at $I_\epsilon = 0$, were in the ratio 1:4. The value of α obtained by extrapolating the data to $|V_c| = \infty$ is 2.2; the value given by the formula for this case with $\beta = 1$, is

$$\alpha = 0.2 + 0.8 \times 2.5 \times \gamma, \tag{6.10}$$

from which we find $\gamma = 1.0$, in agreement with the result of Fig. 4 that substantially all of the emitter current is carried by holes. The theoretical curve shown on the Figure is

$$\alpha = 0.2 + 0.8 \times 2.5 \times |V_c/10|(1 - e^{10/|V_c|}). \tag{6.11}$$

This corresponds to

$$\frac{\tau_t}{\tau_p} = \frac{10}{|V_c|} = \frac{L_c^2(R_c + R_b)}{\tau_p \mu_p R_c |V_c|}, \tag{6.12}$$

from which it was concluded that for the particular bridge studied τ_p was 0.2 microseconds.

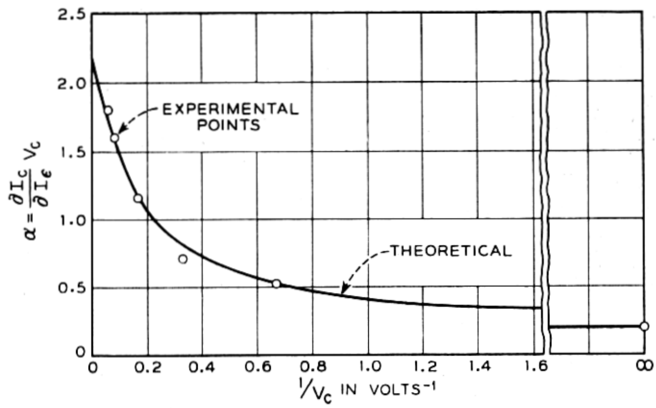

Fig. 10—α versus $1/|V_c|$ showing agreement with the theory for the value of β.

If τ_t is much shorter than τ_p, then the holes penetrate the whole filament and β becomes

$$\beta = \frac{1 - \exp(-i\omega\tau_t)}{i\omega\tau_t} = \frac{e^{-i\omega\tau_t/2} \sin(\omega\tau_t/2)}{(\omega\tau_t/2)}. \tag{6.13}$$

For small values of $\omega\tau_t$, β approaches unity since $(\sin x)/x$ approaches unity as x approaches zero. For $\omega\tau_t/2 = \pi$, the response is zero. This is the condition that $\tau_t = 2\pi/\omega = 1/f$. For this case the filament is just so long that the modulation is averaged over the time of one cycle of the input signal and since this average includes all phases, the modulation vanishes.

Preliminary experiments with filamentary transistors, made in accordance with the principles discussed above, appear to confirm the general aspects of the theory. Power gains of 15 db have been obtained and frequency responses showing a drop of 3 db in α at 10^6 cycles/sec. have been observed. Noise measurements indicate an improvement of 10 to 15 db over the average type-A transistor for comparable conditions of preparation.

Acknowledgments

We have been aided and encouraged in these experiments by many of our colleagues in the Research and Apparatus Development Departments. We are particularly indebted to W. H. Brattain and H. R. Moore for help with experimental problems, to J. Bardeen and W. van Roosbroeck for assistance with the theory and to P. W. Foy and W. C. Westphal for their many contributions in connection with fabricating and measuring the filaments.

References

1. J. Bardeen and W. H. Brattain, *Phys. Rev.*, *74*, 230 (1948).
2. W. H. Brattain and J. Bardeen, *Phys. Rev.*, *74*, 231 (1948).
3. J. Bardeen and W. H. Brattain, *Phys. Rev.*, *75*, 1208 (1949).
4. W. Shockley, *Bell Syst. Tech. J.*, July (1949).
5. J. N. Shive, *Phys. Rev.*, *75*, 689 (1949).
6. W. E. Kock and R. L. Wallace, *Electrical Engineering*, *68*, 222 (1949).
7. E. J. Ryder and W. Shockley, *Phys. Rev.*, *75*, 310 (1949).
8. For reviews of the type-A transistor see reference 3, R. M. Ryder, *Bell Laboratories Record*, Mar. 1949, J. A. Becker and J. N. Shive, *Electrical Engineering*, *68*, 215 (1949) and R. M. Ryder, *Bell Syst. Tech. J.*, July (1949).
9. H. Suhl and W. Shockley, *Phys. Rev.*, *75*, 1617 (1949), *76*, 180 (1949).
10. Experiments of this sort were first reported by J. R. Haynes and W. Shockley, *Phys. Rev.*, *75*, 691 (1949).
11. W. Shockley and G. L. Pearson, *Phys. Rev.*, *74*, 232 (1948).
12. John Bardeen, *Bell Syst. Tech. J.*, July (1949).
13. G. L. Pearson, *Phys. Rev.*, *76*, 179 (1949).
14. Transistors using *p*-type germanium have been described by W. G. Pfann and J. H. Scaff, *Phys. Rev.*, *76*, 459 (1949). Electron injection in *p*-type germanium has also been observed by R. Bray, *Phys. Rev.*, *76*, 458 (1949) and *Phys. Rev.*, *76*, 152 1949.
15. G. Wannier, *Phys. Rev.*, *52*, 191 (1937).
16. Personal communication; a somewhat similar case is treated by B. Goodman, A. W. Lawson and L. I. Schiff, *Phys. Rev.*, *71*, 191 (1947).
17. C. Herring, *Bell Syst. Tech. J.*, July (1949).
18. Transistors of this type, employing *p-n* junctions as well as point contacts as emitters, have been discussed by W. Shockley, G. L. Pearson, M. Sparks, and W. H. Brattain, *Phys. Rev.*, *76*, 459 (1949).
19. Optical constants for germanium have been published by W. H. Brattain and H. B. Briggs, *Phys. Rev.*, *75*, 1705 (1949). The integration over the radiation distribution was carried out by W. van Roosbroeck.