RB-28

BASIC TRANSISTOR DEVICE CONCEPTS

RADIO CORPORATION OF AMERICA
RCA LABORATORIES
INDUSTRY SERVICE LABORATORY

RADIO CORPORATION OF AMERICA

RCA LABORATORIES

INDUSTRY SERVICE LABORATORY

RB-28

Basic Transistor Device Concepts

Basic Transistor Device Concepts

This bulletin will consider some of the elementary device concepts which are essential to the physical understanding of the operation of transistor devices. This discussion will be descriptive in nature with attempts to fortify the consequences of the physical reasoning with some heuristic analysis. While the results thus obtained will have the correct first order functional dependence, multiplying constants may differ from those obtained from a rigorous analysis. From such physical reasoning and simple analysis, one can develop a sense of understanding that many complicated analyses fail to convey.

Because most transistor devices are based on p-n junctions, a study of their properties will constitute the bulk of this bulletin with a brief description of how these properties are utilized in the operation of certain transistor devices.

Physical Concepts of a P-N Junction

Consider a p-n junction to be formed by bringing together a p-type semiconductor and an n-type semiconductor as in Fig. 1. This is not, today, a recommended method of forming a p-n junction. However, as an aid to visualizing the physical processes that go on, this model provides a simple initial state.

The p-type semiconductor contains impurity atoms (acceptors) having one less valence electron each than do the atoms of the host crystal. This robs the normal lattice of one electron for each impurity atom thus creating an electron vacancy or a hole. Since the hole represents an electron deficiency, it is represented as a positive charge. Since the impurity atom (originally electrically neutral) has acquired an additional electron, it is represented as a negative charge. The impurity atom is fixed in the crystal lattice but the hole is free to move and constitute a current.

The converse situation exists in the n-type semiconductor where the impurity atoms (donors) now each have one extra valence electron over those required by the crystal lattice. This extra electron is given up; whereupon the freed electron wanders through the crystal and is free to constitute a current. The fixed impurity atom having given up an electron becomes charged positively. In the separated semiconductors, the mobile charges (electrons and holes) are uniformly distributed throughout the bulk material.

Now, what are the consequences of bringing together these two pieces of semiconductor? It may be expected that the mobile holes and electrons would tend to diffuse throughout the composite crystal. However, restraining forces are set up which counteract this tendency so that the bulk of the holes remain in the p-type material and the bulk of the electrons remain in the n-type material. Thus, the charge densities in regions removed from the junction are essentially unaffected by this juncture.

It should be pointed out that the representation in Fig. 1 is incomplete in that the holes and electrons are considered to be derived only from the ionization of the impurity atoms. In addition to the carriers thus derived, there are, in both types of semiconductor, additional electron-hole pairs formed by the release of electrons from atoms of the host crystal. The energy for ionization of these atoms comes from the thermal energy of the lattice. At normal temperatures, and in material of the type ordinarily used for transistor devices the number of electron-hole pairs thus generated is relatively small in comparison with the number of charge carriers derived from ionization of the impurity atoms.

Now consider the mechanism whereby the holes and electrons are restrained from diffusing throughout the composite crystal. For simplicity consider one type of charge, the holes of the p-type region, for example. The

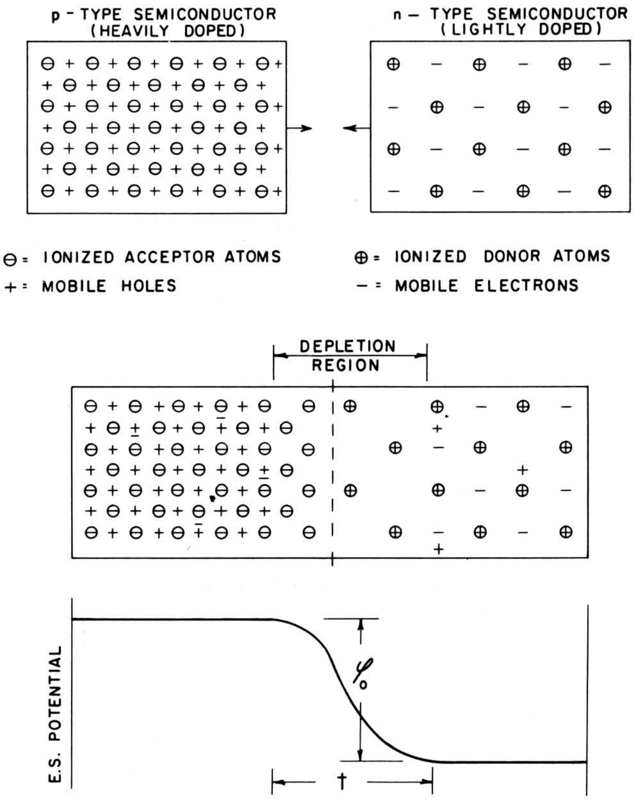


FIG. 1 – Physical picture of p-n junction.

holes tend to diffuse out of the p-type region where their density is high into the n-type region where there are but few holes. The fixed charges (negatively ionized acceptors) are then no longer electrically compensated and give rise to a negative charge density near the transition region between the p and n-type materials. This negative charge being opposite to that of the holes will tend to retard the flow of holes out of the p-type region. Similarly, diffusion of electrons from the n-type region results in an uncompensated positive fixed charge density on the n side of the junction. This further restricts the loss of holes from the p-type region. Thus, under normal equilibrium conditions, the tendency for charges to diffuse across the junction sets up restraining forces in the form of uncompensated fixed charges in the transition region which act to keep the holes in the p-type region and the electrons in the n-type region.

To establish the uncompensated charge densities, the transition region is depleted of mobile charges and this region is often referred to as the depletion region or as the space charge region. The uncompensated charge densities set up an electric dipole resulting in an electrostatic potential difference, ϕ_0, across the junction. Hence, reference is often made to this region as the barrier region, the thickness of the depletion layers being the barrier thickness, *t*, and the electrostatic potential, ϕ_0, the barrier height. As in Fig.1, potential energy diagrams will be drawn, in this bulletin, so that electrons run down hill and holes run up hill.

Current Flow in a P-N Junction

The discussion thus far has referred to a p-n junction under conditions of thermal equilibrium with no applied voltage. Before considering a voltage applied to the junction, let us consider the nature of the currents that might flow.

First we have the picture of holes tending to diffuse across the junction from the p-type region but being hindered by a barrier of height, ϕ_0. The number that do succeed in crossing will depend on the density of holes in the p-type region, $p_p = N_a$, and on the barrier height, ϕ_0. Statistical mechanics tells us in situations like this, the dependency on the barrier height is exponential and so for the hole density on the n-region side of the barrier,

$$p_o = p_p \, e^{-\frac{q}{kT}\phi_o} . \qquad (1)$$

This is referred to as an injected hole density in anticipation of hole injection in transistors. The current will be proportional to the injected charge density so

$$I_i = constant \; e^{-\frac{q}{kT}\phi_o} . \qquad (2)$$

An analogous flow of electrons into the p-region occurs and we may take (2) as being the sum of the hole flow to the right and electron flow to the left. Although the particle flow is conceived of as being in opposite directions, the current flow is in the same direction since the holes and electrons are oppositely charged.

Recall that a junction is being discussed to which no external potentials have been applied. Under equilibrium conditions no net current must flow, so there must be a counter flow to achieve this balance. In the n-region there are normally present small numbers of holes as a result of electron-hole pair generation by thermal processes. There is also an internal potential established across the junction which is in the direction to extract holes from the n-region and sweep them into the p-region. This then constitutes a reverse flow of holes which will balance those diffusing over the barrier into the n-region. How large is this current? The barrier potential can, of course, only extract those holes in the immediate vicinity and in order for additional holes to be extracted they must diffuse to the barrier. Thus the magnitude of this current will be governed by the laws of current flow by diffusion. The "Ohms law," for diffusion current flow is simply (for holes)

$$I_p = -q D_p \frac{\delta p}{\delta x} \qquad (3)$$

i.e. the current is proportional to the charge density gradient. This simply says that particles tend to move from regions of high concentration to regions of low concentration. Thus to determine this current it is necessary to determine the density gradient.

A first-order approximation for the density gradient may be obtained by physical arguments concerning the density at the junction and at some distance away from the junction. Referring now to Fig. 2; at the junction, it may be argued, the hole density is zero since the holes will be immediately swept away by the barrier potential. Now a charge carrier in a semiconductor leads a rather hazardous life since it was derived from an atom which normally would like to get it (or a similar charge) back. Thus, there is always a probability that an electron will drop back into a vacancy represented by a hole, thus ending its life (for the moment) as a free charge capable of carrying current. The barrier potential cannot collect those carriers created so far away that they are lost by recombination before diffusing to the barrier. If L_p is the average distance a hole in the n-type material can diffuse before recombination, then at a distance, L_p, from the junction, the normal hole density, p_n, will be unaffected by the presence of the junction. The density gradient then is p_n/L_p and the diffusion current flow is

$$I_{p_s} = -q D_p \frac{p_n}{L_p} = -\frac{kT}{q} \frac{b}{(1+b)^2} \frac{\sigma_i^2}{(\sigma_n L_p)} \qquad (4)$$

where the latter form may be obtained from simple manipulation using

$$\sigma_i = q(\mu_p + \mu_n)\, n_i;\; b = \frac{\mu_n}{\mu_p};\; p_n n_n = n_i^2;\; \sigma_n = q \mu_n n_n;\; \mu = \frac{q}{kT} D$$

By analogy the electron flow from the p-region to

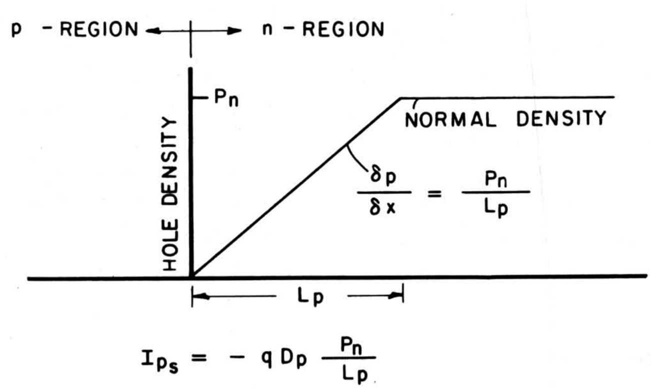


FIG. 2 – Hole saturation current. Current flow is to left.

the n-region is (Fig. 3)

$$I_{n_s} = -q D_n \frac{n_p}{L_n} = -\frac{kT}{q} \frac{b}{(1+b)^2} \frac{\sigma_i^2}{\sigma_p L_n} \qquad (5)$$

The total counter flow of current under equilibrium conditions is then

$$I_s = I_{p_s} + I_{n_s} = -\frac{kT}{q} \frac{b}{(1+b)^2} \sigma_i^2 \left(\frac{1}{\sigma_n L_p} + \frac{1}{\sigma_p L_n}\right) \qquad (6)$$

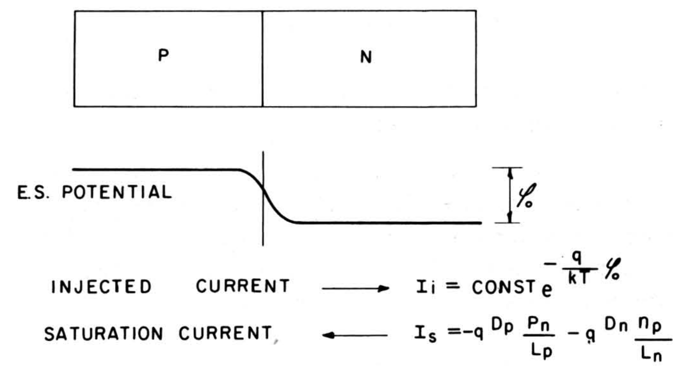


FIG. 3 – Nature of current flow. Under equilibrium conditions $I_i = I_s$.

Note that these currents do not depend on the barrier height. This reverse flow then will *not* depend on the applied voltage and is known as the saturation current. Under equilibrium conditions, the saturation current just balances the injected current ((2) above). The barrier height adjusts itself so that these currents are balanced for both hole and electron flow individually and no net current flows across the junction.

What happens to the injected carriers as in (1) that succeed in overcoming the barrier? The injected holes tend to diffuse away from the barrier into the n-type material. As charge carriers they are subject to the same hazardous life as the holes normally present and on the average diffuse a distance, L_p, before being lost by recombination. Thus, the injected hole density distribution is approximately as shown in Fig. 4. Recalling the form for current flow due to diffusion,

$$I_p = -q D_p \frac{\delta p}{\delta x}$$

The hole current due to the injected charge density is simply:

$$I_p = q D_p \frac{p_o}{L_p}$$

where p_o is related to the barrier height by (1). A simi-

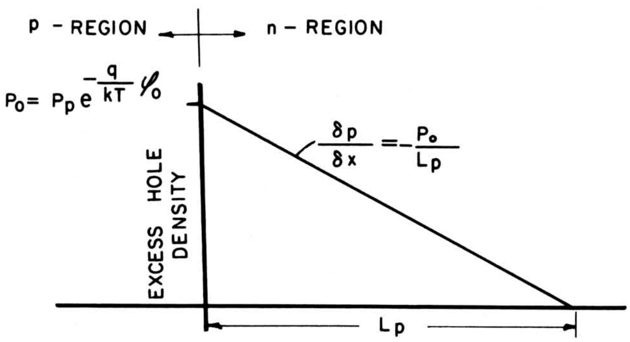


FIG. 4 – Injected hole current.

lar relation holds for the electrons surmounting the barrier in the opposite direction.

If an external potential is applied to the junction, the effect is to alter the height of the barrier. As noted above, this does not affect the saturation component of current flow but will change the injection current flow. The injection current flow now becomes

$$I_i = \text{const}\ e^{\frac{-q}{kT}(\phi_o - V)} \tag{7}$$

where V is taken positive when a positive potential is applied to the p-side of the junction as in Fig. 5.

The constant in (7) can readily be evaluated since it is known that, when $V = 0$, the injection current must equal the saturation current. A little manipulation shows that the total flow is

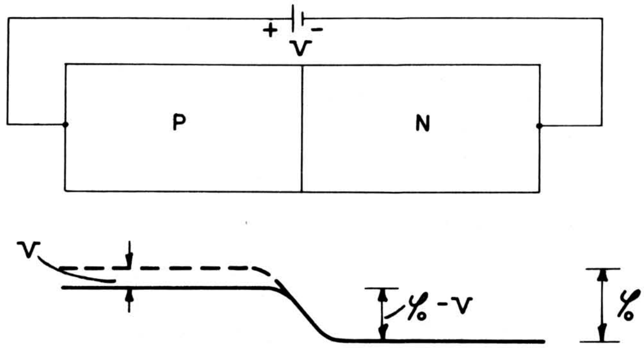
$$I = I_s\,(e^{\frac{q}{kT}V} - 1) \tag{8}$$

and the same form applies individually to both the electron and hole components of the current. When V is positive a very large current will flow because of the exponential relation. When V is negative the current will be extremely small and is the saturation current. It is seen from this relation that the rectification properties of a p-n junction have no direct relation to the proportion of current carried by holes or electrons. The ratio of forward to reverse current being simply.

$$(e^{\frac{q}{kT}V} - 1)$$

On the other hand for many device applications, the property of a p-n junction to inject holes into n-type material or to inject electrons into a p-type material is essential. It will be of interest to discuss these properties further.

First, however, indicate the a-c conductance obtained by differentiation of (8).



$I_i = \text{CONST}\ e^{-\frac{q}{kT}(\phi_0 - V)}$ →

I_s ←

FROM $I_i = I_s$; $V = 0$

$I = I_s\,(e^{\frac{q}{kT}V} - 1)$

FIG. 5 – Diode relation.

$$g = \frac{\delta I}{\delta V} = \frac{q}{kT} I_s\, e^{\frac{q}{kT}V}$$

$$\text{If } V \gg \frac{kT}{q},\quad g = \frac{q}{kT} I \tag{9}$$

$$\text{If } V \ll 0,\quad g \to 0$$

In particular, note the extremely small reverse conductance given by this simple theory. Subsequently it will be shown how another phenomenon in a p-n junction leads to a finite a-c conductance.

Hole and Electron Currents—Injection Efficiency

Note from (8), that the current in a p-n junction is proportional to the saturation current. Hence, an inspection of the saturation currents for holes and electrons will give information on the character of the diode current whether it be in the forward or reverse direction. From (4) and (5) it is seen that the hole current depends only on the properties of the n-type region and the electron current only on the properties of the p-type region. The injection ratio of the hole and electron currents is from (4) and (5),

$$\frac{I_p}{I_n} = \frac{I_{p_s}}{I_{n_s}} = \frac{1/\sigma_n L_p}{1/\sigma_p L_n} = \frac{\sigma_p L_n}{\sigma_n L_p} \tag{10}$$

For diffusion lengths of roughly equal magnitude, the predominant current across the junction corresponds to the majority carrier of the material having the greater conductivity. In this way a p-n junction may be utilized to inject a current of minority carriers into a semiconductor body and to suppress a flow of undesired majority carrier current from that body. This property of the p-n junction forms the basis for the common type of bipolar transistor.

It is common in discussing various aspects of transistor theory to refer to the injection efficiency, γ, rather than the ratio given by (10). The hole injection efficiency is the fraction of the total current across the junction carried by holes and is

$$\gamma = \frac{I_p}{I} = \frac{I_p}{I_p + I_n} = \frac{1}{1+\sigma_n L_p/\sigma_p L_n} \approx 1 - \frac{\sigma_n L_p}{\sigma_p L_n}$$

where the approximation is valid for efficiencies near unity.

The above discussion of various aspects of the p-n junction has been given on the basis that the currents were small. However, as the currents are increased other effects become important. Large currents are necessary in large signal operation and in power transistors. This discussion of the injection efficiency is an opportune place to consider one of these effects, namely, the reduction of injection efficiency with increasing currents. This consideration will also bring to light a basic physical concept which, in the interests of simplicity, has not yet been pointed out.

For clarity consider a specific type of junction, in particular one in which $\sigma_p >> \sigma_n$ so that the current across the junction is carried principally by holes being injected into the n-type region. At low currents, the ratio of hole to electron current is given by (10) above. Now how is this altered at high currents? When a hole is injected into the n-type region, an electron also enters through the external connection to preserve charge neutrality. Although the injection of one type of charge has been considered principally, this injection is accompanied by an equal flow of carriers of the opposite type. Reference is often made to this fact by stating that one should really speak and think of the injection of electron-hole pairs rather than of the injection of one type of carrier. Indeed, a consequence of this is the reduction of emitter efficiency at high currents. If then p_o is the injected hole density corresponding to an injected current, $I_p = qD\, p_o/L_p$, as in Fig. 6, then an equal compensating charge density, $n_o = p_o$ is added to the electron density normally present. Those electrons normally present arise from ionization of the donor atoms, N_d, so that the total electron density is $p_o + N_d$. This then is the total electron density at the junction being held back from entering the p-region by the barrier, $(\phi_o - V)$. The number crossing the barrier is

$$(p_o + N_d)\, e^{\frac{-q}{kT}(\phi_o - V)}.$$

From the earlier calculation of diode current where the increase in electron density was neglected it will be recalled that for zero applied voltage the injected current was equal to the saturation current so that

$$N_d\, e^{\frac{-q}{kT}\phi_o} = I_{n_s}$$

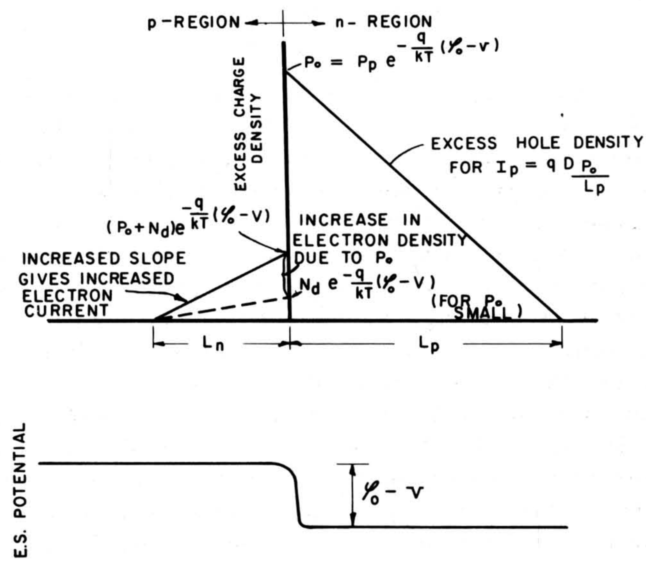


FIG. 6 – Excess charge densities for p-n junction where $\sigma_p >> \sigma_n$ and showing how electron current in p-region is increased by injection of p_o into n-region. The proportion of total current carried by holes is thereby reduced.

From this $e^{\frac{-q}{kT}\phi_o}$ can be evaluated so that I_n becomes approximately

$$I_n = \frac{p_o + N_d}{N_d} I_{n_s} e^{\frac{q}{kT}V} = (1 + \frac{L_p}{qDN_d} I_p)\, I_{n_s} e^{\frac{q}{kT}V} \qquad (11)$$

where, since large currents are assumed the saturation current flow is omitted. The electron current flow across the junction has been increased by the factor, $(1 + \frac{p_o}{N_d})$, which is a function of the hole current. A similar argument may be applied to the hole current but for the case

$\sigma_p >> \sigma_n$ the factor is small and may be neglected for this first order calculation. The injection ratio is then

$$\frac{I_p}{I_n} = \frac{I_{p_s}}{(1+p_o/N_d)I_{n_s}} = \frac{\sigma_p L_n}{\sigma_n L_p}\left(\frac{N_d}{p_o + N_d}\right) \qquad (12)$$

In this case, the injection efficiency is reduced by a disproportionate *increase in the electron current* crossing the junction as the voltage is increased to increase the injected hole current.

Bipolar Transistor

Now consider how the properties of the p-n junction may be applied to form the conventional bipolar transistor. Consider, as in Fig. 7, a p-n-p transistor formed from two p-n junctions placed back to back with a base width, *W*, which is much less than the diffusion length for holes in this region and with voltages applied as indicated. With a positive voltage applied to the left-hand junction a large current will flow into the base. If the conductivity of the emitter is much greater than that of the base, this current will be predominantly a hole flow. The injected holes are minority carriers in the base region and will diffuse through the base. Because the base region is thin, most of these will reach the right hand junction and only a small fraction lost by recombination in the base. The right hand junction is biased negatively or in a direction to collect holes from the base region and transfer them to the right-hand p-region. Thus the hole current injected by the emitter p-n junction diffuses through the base and is collected by the collector junction and the hole current is substantially constant through the device. Electron currents across the junctions are unwanted currents (in a p-n-p transistor) and one of the problems of transistor design concerns the minimization of these currents.

The forward conductance of a p-n junction is large so that little power is required to inject hole current into the base region. On the other hand the conductance of a junction biased in the reverse direction is very small. Now, because the same current flows through the small conductance of the collector junction that was injected at the cost of very little power through the emitter junction, a considerably increased power may be developed in an external load.

Transistor action depends primarily on the diffusion of minority carriers through the base region. An analysis of the base region provides the essential features of transistor performance while the properties of the end regions are principally concerned with the flow of unwanted currents. A rigorous analysis then would solve the diffusion equation in the base, subject to the boundary conditions imposed by the junctions. This becomes somewhat involved analytically so consider what can be done somewhat heuristically on the basis of the current flow in p-n junctions as discussed above.

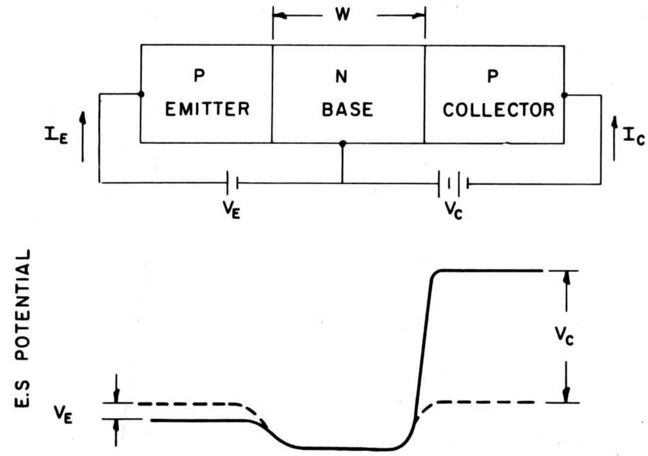


FIG. 7 – P-n-p transistor.

The current flow across the emitter junction can be conceived of as

I_e =(*hole + electron flow due to* V_e) + (*hole + electron flow due to* V_c),
and that across the collector junction,
I_c = (*hole + electron flow due to* V_e) + (*hole + electron flow due to* V_c).

The first term of I_e we can almost write by considering this as a straight forward diode and using (8). However, one modification must be made. Diode relation (8) was computed on the basis that the diffusion length, L_p, in the n-region was smaller than the extent of the region. In the transistor, we have made *W* smaller than L_p. It will be recalled that L_p was the distance in which the injected hole density decreased to zero. In the transistor case, the collecting action of the collector junction reduces the hole density to zero at *W*. It is reasonable to replace L_p by *W* as the factor in determining the gradient giving rise to diffusion flow of current. Similarly, the second term of I_c can immediately be written down with the above modification. This is essentially the reverse current of the collector junction, i.e., the current for a junction biased negatively.

This leaves the transfer terms to be considered. Consider first the transfer term of I_c. From a physical explanation of the operation of the transistor this is simply the hole current injected by the emitter junction and collected at the collector (p-n-p transistor). It is given by the hole component of (8) again with the modification of replacing L_p with W.

Now the structure is quite symmetrical and nothing in the device indicated that the left hand junction should be the emitter. Thus, the transfer terms must have similar coefficients so that, if desired, the voltage polarities can be interchanged and transistor action obtained using the right hand junction as the emitter. Thus the second term in I_e is similar to the first in I_c.

$$I_e = \frac{kT}{q}\frac{b}{(1+b)^2}\sigma_i^2\left\{\left(\frac{1}{\sigma_b W}+\frac{1}{\sigma_e L_e}\right)\left(e^{\frac{q}{kT}V_e}-1\right)-\frac{1}{\sigma_b W}\left(e^{\frac{q}{kT}V_c}-1\right)\right\}$$

$$I_c = \frac{kT}{q}\frac{b}{(1+b)^2}\sigma_i^2\left\{\frac{1}{\sigma_b W}\left(e^{\frac{q}{kT}V_e}-1\right)-\left(\frac{1}{\sigma_b W}+\frac{1}{\sigma_c L_c}\right)\left(e^{\frac{q}{kT}V_c}-1\right)\right\} \quad (14)$$

which is the first order approximation to the solution obtained from a rigorous mathematical development. In the rigorous solution the coefficients of $\left(e^{\frac{q}{kt}V_i}-1\right)$ are given in terms of hyperbolic functions and in practice it is the first order approximation given in (14) that is most often used in calculations.

Capacitive Effects in a P-N Junction

Two phenomena give rise to the flow of capacitive currents in p-n junctions. The first of these, which is termed a diffusion capacitance, is the result of the nature of minority carrier flow in a semiconductor, i.e., a diffusion flow. The second, which is referred to as a transition capacitance, is a consequence of the depletion of mobile charges near the junction as has been discussed earlier.

Consider first the diffusion capacitance arising from hole injection into the n-region of a p-n junction. The hole flow in the n-region corresponds to a charge density gradient as shown in Fig. 8. If the current is altered by changing the applied voltage, the gradient and hence the charge density distribution must change, thus changing the total charge. This change in total charge with the applied voltage corresponds to a capacitance (a diffusion capacitance).

The total hole charge within a diffusion length of the junction is

$$Q = q\frac{p_o}{2}L_p$$

The diffusion current is $I_p = -qD\frac{\delta p}{\delta x} = qD\frac{p_o}{L_p}$

Then

$$C = \frac{\delta Q}{\delta V} = \frac{\delta Q}{\delta p_o}\frac{\delta p_o}{\delta I_p}\frac{\delta I_p}{\delta V} = \left(q\frac{L_p}{2}\right)\left(\frac{L_p}{qD_p}\right)\left(\frac{q}{kT}I_p\right) = \frac{q}{kT}\frac{L_p^2}{2D_p}I_p \quad (15)$$

where the a-c conductance $\frac{\delta I_p}{\delta V} = \frac{q}{kT}I_p$ from (9) was used.

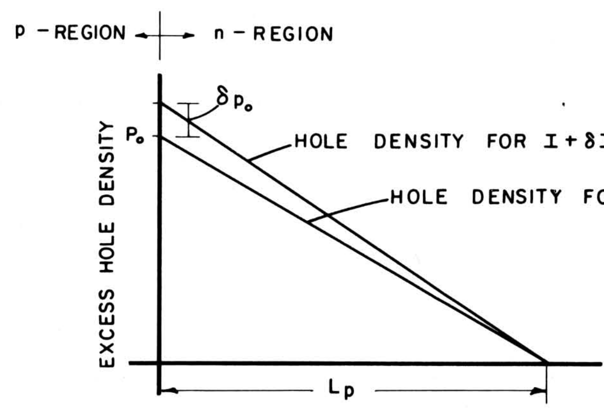


FIG. 8 – Hole density distributions in n-region for calculation of hole diffusion capacitance.

An analogous expression exists for the diffusion capacitance for the electron flow into the p-region. The total diffusion capacitance is ordinarily much larger than the transition capacitance and plays an important part in limiting the frequency response of the transistor. In the transistor, however, remember that the minority carriers are stored in the distance W and not L_p since ordinarily $W \ll L_p$. Thus replace L_p by W in thinking of the diffusion capacitance of the injecting junction of a p-n-p transistor.

Consider next the transition capacitance which, in a transistor, is important for the collector junction. It has been shown earlier how the potential barrier in a p-n junction was a consequence of the depletion of mobile charges in this region leaving uncompensated fixed charges. Further, if the potential drop is varied by an externally applied voltage, these densities must be altered to correspond to the new conditions. This process requires a flow of charge in response to the change in voltage, i.e., a capacitative flow. It is apparent that the exact nature will depend on the distribution of impurity atoms in the transition region. Two cases of practical importance have been much discussed: (1) an abrupt or step transition in which the impurity type changes discontinuously from n to p and (2) a gradual transition in which the net impurity concentration changes linearly from n to p. Consider the first case, a discontinuous transition, and, in particular, a junction in which $\sigma_n >> \sigma_p$. This is shown in Fig. 9, where the depletion layer is considered to exist only in the p-region by virtue of its much lower conductivity. If a pill-box of unit cross sectional area is constructed with one face at $x = o$ at the edge of the depletion layer where the field is zero and the other face at $x = x$, the field at x is given by Gauss' Law as

$$\int_{surface} \mathcal{E}_n \, dA = \frac{4\pi Q}{\kappa}$$

here $\mathcal{E}_n$ is the normal surface component of the field and Q is the charge enclosed in the pill-box. Upon integration

$$\mathcal{E} = \frac{4\pi}{\kappa} q N_a x$$

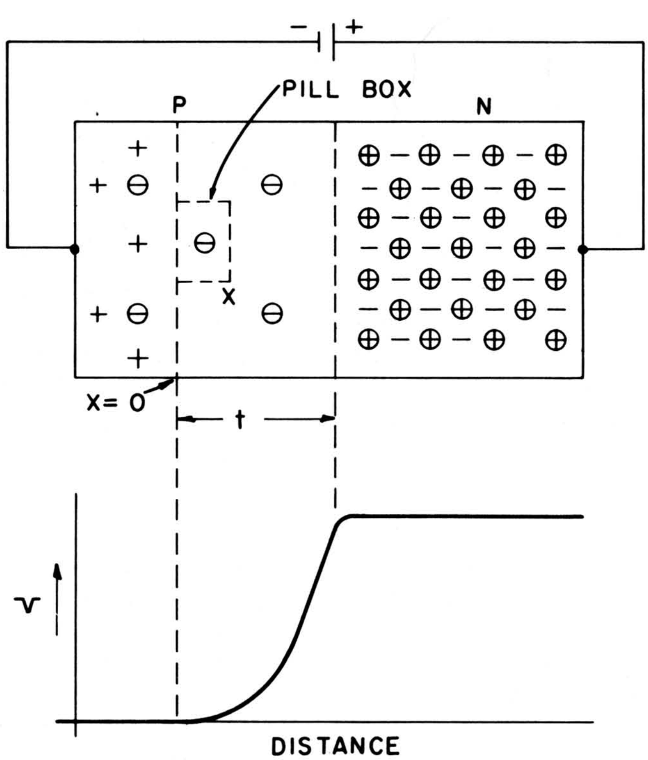


FIG. 9 – Figure for calculation of transition capacitance when $\sigma_n >> \sigma_p$.

The potential can be found by integration over the barrier thickness, t:

$$V = \int \mathcal{E} \, dx = \frac{2\pi q}{\kappa} N_a t^2 \qquad (16)$$

so the barrier thickness is

$$t^2 = \frac{\kappa \mu_p}{2\pi \sigma_p} V \qquad (17)$$

The capacitance can then be determined by taking the incremental change in charge with voltage

$$\text{or} \quad C = \frac{\delta Q}{\delta V} = \frac{q N_a \, \delta t}{(2\pi q/\kappa) N_a 2t \, \delta t} = \frac{\kappa}{4\pi t} \qquad (18)$$

which is the capacitance per unit area of a parallel plane condenser with electrode spacing, t, filled with a material of dielectric constant, κ.

If a similar analysis is performed for the depletion region in the n-region (which was neglected above in the interests of simplicity), the total depletion region thickness is given by:

$$t^2 = \frac{\kappa \mu_p}{2\pi} \left(\frac{b}{\sigma_n} + \frac{1}{\sigma_p} \right) V \qquad (19)$$

where V is the total barrier potential including both the internal electrostatic potential and the applied voltage. In many practical transitions, such as those characteristic of alloyed junctions, one of the conductivities is much greater than the other and one term may be dropped as in the example.

For a linear transition the barrier thickness is given by

$$t^2 = \frac{3\kappa}{4\pi q a} V \qquad (20)$$

where a is the net impurity density gradient.

Some Consequences of Variable Barrier Thickness

Because the barrier thickness depends on the applied voltages, the simple picture of transistor operation we have given earlier must be modified in detail. Consider, for example, the effect of a variable voltage across the collector junction. Such a voltage is present when the transistor is operating into a load across which an a-c voltage is developed. The "electrical" thickness of the collector junction now varies with the instantaneous voltage and in so doing, alters the base width of the transistor (base width modulation). This effect modifies both the apparent conductance and susceptance of the collector junction. Consider a physical picture of these phenomena. Fig. 10 shows the situation in the base region of the transistor as the effective position of the collector junction is varied.

Remembering that the diffusion current is proportional to the density gradient

$$I = -qD\frac{\delta p}{\delta x}$$

$$I_{max} = qDp_o\frac{1}{W - \frac{\delta W}{2}} \qquad I_{min} = qDp_o\frac{1}{W + \frac{\delta W}{2}} \qquad (21)$$

Then

$$g = \frac{\delta I}{\delta V} = \frac{I_{max} - I_{min}}{\delta V} = q\frac{Dp_o}{W}\frac{1}{W}\frac{\delta W}{\delta V} = I\frac{1}{W}\frac{\delta W}{\delta V}$$

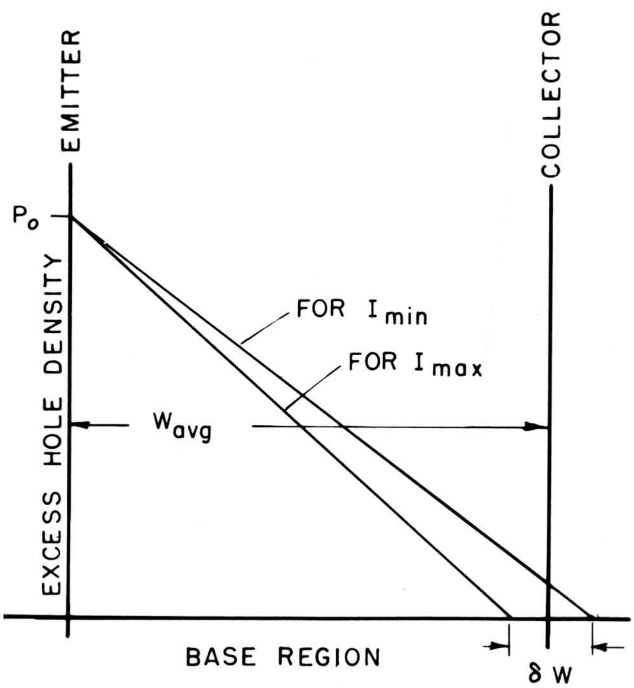


FIG. 10 – Hole distributions in base of p-n-p transistor as effective base width is varied by variation in collector voltage.

giving an approximate expression for the a-c conductance due to modulation of the base width, W, by an applied signal. As we have seen earlier the simple theory gave a conductance for a reversed bias junction of the form $e^{\frac{q}{kt}V}$. For even relatively small negative values of V this conductance is extremely small. The conductance developed above turns out to be much greater and for a practical unit with negligible leakage does express the finite conductances found.

The situation of Fig. 10 shows that the charge in the base region also varies as the effective position of the collector junction moves. This variation of charge with voltage may be interpreted as a capacitance which adds in parallel to the junction transition capacitance discussed above. This capacitance may be found as

$$Q_{mas} = q\frac{p_o}{2}\left(W + \frac{\delta W}{2}\right) \qquad Q_{min} = q\frac{p_o}{2}\left(W - \frac{\delta W}{2}\right)$$

$$C = \frac{\delta Q}{\delta V} = \frac{Q_{max} - Q_{min}}{\delta V} = \frac{qp_o}{2}\frac{\delta W}{\delta V} = I\frac{W^2}{2D}\frac{1}{W}\frac{\delta W}{\delta V} \qquad (22)$$

This is similar in form to the diffusion capacitance associated with the emitter due to the flow of holes into the base. However, in this case the additional factor $\frac{1}{W}\frac{\delta W}{\delta V}$ makes this capacitance small and in practice it is but a fraction of the transition capacitance.

Field-Effect Transistor

The phenomenon of a variable barrier thickness has formed the basis for a different type of transistor device – the field-effect transistor. Such a device is shown in Fig. 11 and consists of a thin piece of semiconductor on the opposite sides of which are two p-n junctions. An ohmic contact is located at either end of the semiconductor. It is seen that a conducting channel exists between the two ohmic contacts. This channel is defined by the two p-n junctions. If a reverse bias is applied to the two junctions, the conducting channel becomes still further limited by the depletion layers of the junctions. Thus the channel conductance may be varied by modulating its cross section through a vari-

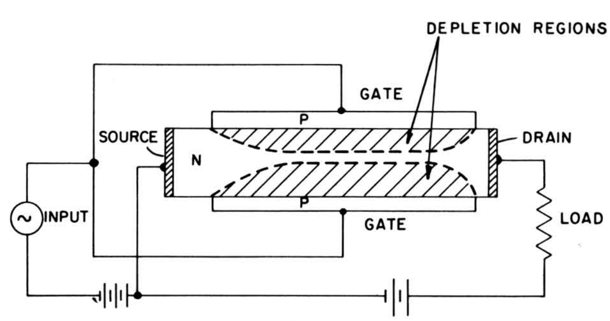


FIG. 11 – Field effect transistor.

able voltage applied to the p-n junction. In this way the signal applied to the p-n junction controls the current flow between the ohmic contacts and through the load.

Origin of Temperature Effects

The most important factor in determining temperature effects in transistors will be briefly examined. The transistor equations previously stated showed that the intrinsic conductivity was the most sensitive factor. Recalling that the diode currents were proportional to the saturation currents, consider the hole saturation current, for example:

$$I_{p_s} = -q\,D\,\frac{p_n}{L_p}$$

where p_n is the density of holes of thermal origin in the n-type material. These holes are created by the loss of an electron from an atom of the host crystal. Writing this in another fashion, using the relation $n_i^{\,2} = n_n\,p_n$, gives

$$I_{p_s} = -q\frac{D}{L_p}\,\frac{n_i^2}{n_n}$$

where n_i is the density of electrons (or holes) in intrinsic material and n_n is the density of electrons in the n-type material which come principally from the donor atoms.

In Fig. 12 are shown, relative to thermal energy at room temperature (solid) and 100 degrees C (dotted),

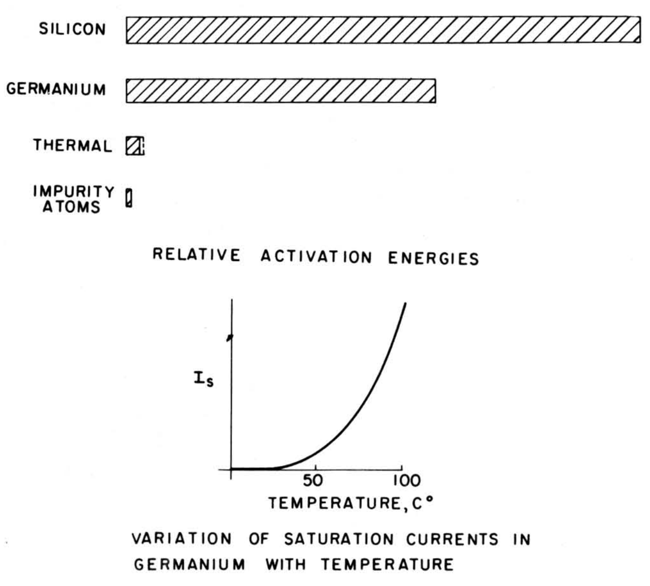


FIG. 12 – Temperature effects.

the energies required to release an electron from an impurity atom normally used in doping germanium and from a germanium atom and from a silicon atom. It is seen that thermal energy is somewhat larger than that required to release an electron from an impurity atom. It is for this reason that at normal temperatures it can be assumed that all of the impurity atoms are ionized and consequently that the normal electron density (in n-type material), n_n, is equal to the donor density, N_d. Furthermore, having ionized all of the donor atoms at normal temperatures, further increases in temperature do not change the number of electrons obtainable from this source. Then in I_{p_s} above, n_n does not vary rapidly with with temperature.

On the other hand, normal thermal energies are much smaller than the energy required to extract an electron from either a germanium or silicon atom. The number of electrons that will surmount a barrier as has been discussed in connection with the barrier in a p-n junction, is an exponential function of the barrier height in terms of thermal energy. Indeed,

$$n_i^2 \propto e^{\frac{-q}{kT}E_G}$$

where E_G is the "band gap" or activation energy shown relatively in Fig. 12. Although the numbers are relatively small in comparison with n_n in practical transistors, the variation with temperature, due principally to the exponential dependence, is extremely rapid. The graph of Fig. 12 illustrates the temperature dependence of the saturation current in germanium between zero and 100 degrees C. This corresponds to the reverse collector current, I_{co} or COI_c. Because the activation energy in silicon is greater than that in germanium, the density of intrinsic electrons (and holes) at the same temperature is much less (again because of the exponential dependence on barrier height). However, the variation is similar in form and percentage wise about the same, but because the density at normal temperatures is so small, the currents are unobjectionable until the temperature is increased considerably.

Other Transistor Devices

Now consider briefly the mode of operation of a few other transistor devices made of an assemblage of p-n junctions.

The hook transistor shown in Fig. 13 will be used to illustrate a different principle that provides a possible

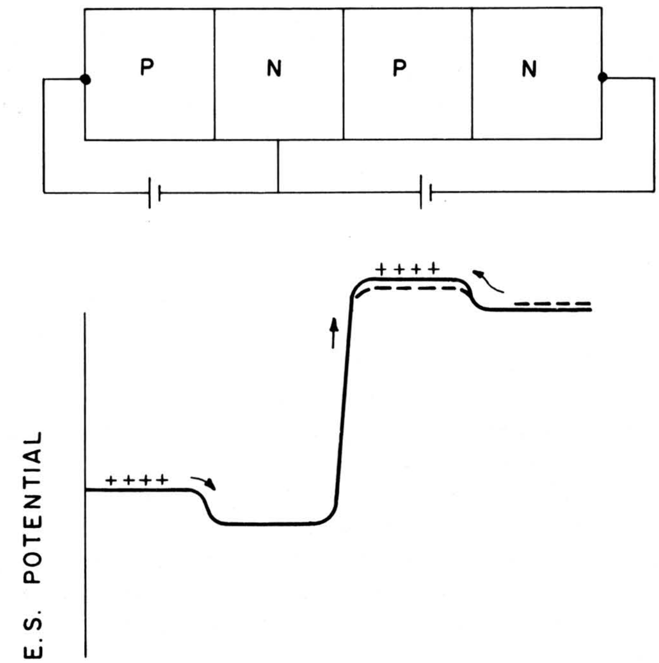


FIG. 13 – Hook transistor.

explanation for a certain behavior of point contact transistors. The hook transistor shown is a p-n-p-n structure with no external connection made to the internal p-region. The left hand junction is biased in the forward direction as an emitter. A negative potential is applied to the right-hand n-region. This biases the internal p-region negatively with respect to the base – in the direction to collect the holes emitted by the emitter. Under static conditions the internal p-region floats at a potential such that the currents entering and leaving it are equal. If holes are injected at the emitter and collected by the internal p-region, they are essentially trapped there. The accumulation of holes lowers the potential of the internal p-region biasing it in a forward direction with respect to the right hand n-region. This permits an even greater number of electrons to be injected from the right which diffuse through the p-region and are collected by the n-type base. In this fashion the current injected by the left hand junction can control an even greater current flowing across the right-hand junction. Thus, a current gain factor, alpha, greater than unity may be obtained. This is in contrast to the p-n-p junction transistor in which the collector current could at most be equal to the injected current, i.e., an alpha of unity.

Fig. 14 shows a point contact transistor. Its operation can be at least qualitatively explained in terms of p-n junction theory. However, the details of its operation are not developed to the relatively refined state of the junction transistor. Here the emitter is a metal-to-semiconductor contact which, with the aid of surface states on the crystal, is able to create an electron-deficient or p-type region in the vicinity of the contact. Such a contact, similar to the p-n junction, can inject holes into the body of the germanium. The collector junction of a point contact is normally "formed" by pulsing it with an electric current. This forms a p-type region somewhat under the surface of the crystal. Biased negatively this acts as a collector for the holes injected by the emitter. Thus far this is not much different than a p-n-p junction transistor except for geometry. However, point contact transistors, as it is well known, have current-gain-factors (alpha) greater than unity. Just how this is achieved has not yet been completely settled. If one considers that between the formed p-region and the contact there exists an n-type region, then the structure is serially similar to the junction "hook" transistor discussed above. The reasoning that gave the junction "hook" transistor a current-gain-factor greater than unity can be invoked to "explain" this behavior of the point contact transistor. This has not been entirely successful.

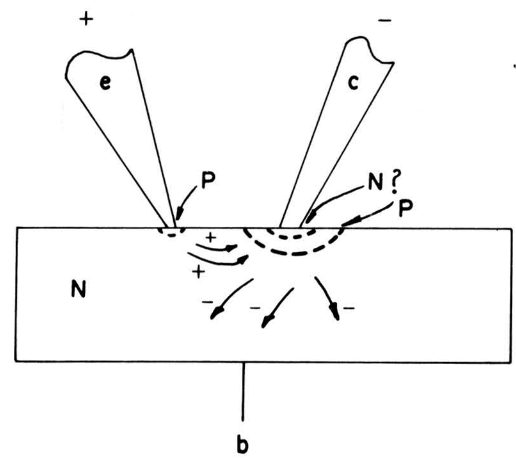


FIG. 14 – Point contact transistor.

An alternative proposal has been made that the collector forming process introduces traps in the region near the collector point. These traps are temporarily filled by holes injected by the emitter and in this state represent an accumulation of positive charge directly in front of the metallic collector point. The metallic collector point can supply an abundance of electrons and these are extracted from the metal by the collection of positive holes in the traps. This electron current flow may be much larger than the hole current injected by the emitter to give current gain factors greater than unity.

A p-n-i-p transistor is shown in Fig. 15. This may be considered much like a usual p-n-p junction transistor with a modified collector region. The modification in this case consists of the intrinsic region interposed between the n-type base and the p-type collector. In our discussion of depletion layers, it was shown how the depletion layer extended principally into the low con-

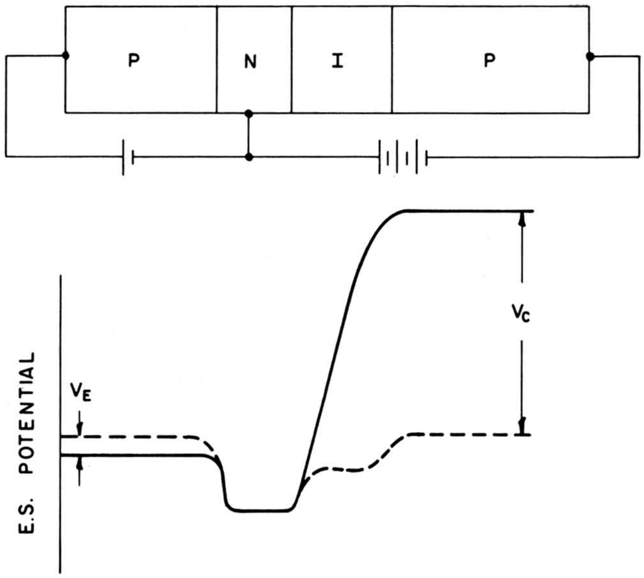


FIG. 15 – P-n-i-p transistor.

ductivity material. The p-n-i-p transistor may be viewed as a p-n-p transistor in which the conductivity of the side of the base region near the collector has been made extremely low. When a potential is applied to the collector, the field extends through the intrinsic layer to the base. This provides a collector junction whose barrier thickness is very wide – the thickness of the intrinsic layer. Because the applied potential is distributed over a larger region, this type of junction can withstand a high reverse voltage and has a lower capacitance. The low collector capacitance improves the high frequency operation. The high frequency operation of this type also depends, as in the conventional p-n-p transistor, upon making the n-type base region very thin. High frequency operation also demands that the n-type region have a high conductivity which may seriously limit the voltage that may be applied to the collector junction of a conventional unit. In the p-n-i-p structure higher conductivity may be used in the n-type base with improved high frequency performance without compromising the collector breakdown voltage.

Harwick Johnson

Symbols

I = *current density*

V = *voltage*

p = *hole density*

n = *electron density*

N_a = *acceptor impurity density*

N_d = *donor impurity density*

k = *Boltzmanns constant*

T = *Kelvin temperature*

ϕ = *barrier potential*

q = *magnitude of electronic charge*

D = *diffusion constant*

μ = *mobility*

L = *diffusion length*

σ = *conductivity*

b = μ_n/μ_p = *ratio of electron to hole mobility*

g = *a.c. conductance*

C = *capacitance per unit area*

Q = *charge*

t = *barrier thickness*

W = *base region width*

κ = *dielectric constant*

$\mathcal{E}$ = *electric field intensity*