10.1109/JRPROC.1957.278485

The High Current Limit for Semiconductor Junction Devices*

NEVILLE H. FLETCHER†

Summary—**At very high operating levels the density of carriers injected into the body of a semiconductor junction device is comparable with the carrier density in the emitter regions of the device. The effect of these high densities on the lifetime and mobility of carriers is considered, and new equations are derived relating the carrier densities on either side of a forward biased junction. These equations are applied to derive the forward characteristics of two diode types, and to consider the dependence of emitter efficiency on current density for alloy junction transistors.**

For a PIN diode, over a considerable current range, the forward current varies approximately as exp ($\lambda qV/kT$), where $\frac{1}{2} \leqq \lambda \leqq 1$; $\lambda = 1$ for very thin diodes, and $\lambda = \frac{1}{2}$ for thick diodes. For PIR diodes, the forward current varies as exp ($qV/2kT$). Both types show additional voltage drop at high currents. Transistor emitter efficiency decreases with increasing current, but saturates to a finite value at high currents. These predictions are in accord with experiment and suggest design considerations for optimum performance. A brief discussion is also given of the usefulness of these new results in device applications.

Introduction

THEORIES of the operation of semiconductor junction devices are conveniently classified on the basis of carrier injection level. Junction devices typically have regions of relatively high resistivity (a few ohm cm) in conjunction with regions of very low resistivity (hundredths or thousandths of an ohm cm) and these regions provide a yardstick for measuring carrier densities.

Low level theories, such as that of Shockley *et al.*,[1,2] are characterized by the fact that injected carrier densities are assumed to be very much less than the equilibrium majority carrier density in the high resistivity regions. Such theories give small signal parameters which are constants and the theory is essentially linear.

In transition theories, such as those of Webster,[3] Rittner,[4] and Misawa,[5,6] injected carrier densities are considered to be comparable with majority carrier densities in the high resistivity regions. These theories are relatively complicated and the small signal parameters are no longer constants. For very high injection levels, these theories merge with the high level theories and their form becomes simpler.

High level theories are characterized by the assumption that injected densities are very much greater than the equilibrium densities in the high resistivity regions, but still very much less than the density of carriers in the low resistivity regions. Theories in this group have been restricted to diodes, as transistors have been described by the more elaborate transition theories which include the high level case as a limit. Hall[7] has considered a two-junction diode which we can denote as PIN, while Kinman, *et al.*,[8] and Spenke[9] have treated a junction and recombination-surface diode which we shall call PIR. The interesting result of these treatments is that the diode forward current is proportional to exp ($qV/2kT$) in both cases. A conflicting theory for the PIN diode has been proposed by Kleinman,[10] who obtains an exp (qV/kT) characteristic at variance with experiment. We believe this discrepancy to be due to Kleinman's inadequate treatment of recombination in the I region. His paper does, however, include some treatment of higher level effects.

Limiting theories deal with the case of still higher injection levels, when the injected carrier density is comparable with the majority carrier density in the low resistivity regions. It is to the development of a theory of this kind that the present paper is devoted.

Phenomena at Very High Levels

Before proceeding to develop our theory, we shall examine some of the assumptions usually made and see to what extent they must be modified at the very high carrier densities which we shall encounter.

Lifetime

The recombination of excess carriers in germanium and silicon is primarily caused through the agency of trapping centers which have energy levels lying near the middle of the forbidden band. The theory of recombination through these centers has been given by Hall[11]

* Original manuscript received by the IRE, September 24, 1956; revised manuscript received February 25, 1957.

† Division of Radio Physics, Commonwealth Scientific and Industrial Research Organization, Sydney, Australia.

[1] W. Shockley, "The theory of p-n junctions in semiconductors and p-n junction transistors," *Bell Sys. Tech. J.*, vol. 28, pp. 435–489; July, 1949.

[2] W. Shockley, M. Sparks, and G. K. Teal, "The p-n junction transistor," *Phys. Rev.*, vol. 83, pp. 151–162; July, 1951.

[3] W. M. Webster, "On the variation of junction-transistor current-amplification factor with emitter current," Proc. IRE, vol. 42, pp. 914–920; June, 1954.

[4] E. S. Rittner, "Extension of the theory of the junction transistor," *Phys. Rev.*, vol. 94, pp. 1161–1171; June, 1954.

[5] T. Misawa, "Emitter efficiency of junction transistor," *J. Phys. Soc. Japan*, vol. 10, pp. 362–367; May, 1955.

[6] T. Misawa, "A note on the extended theory of the junction transistor," *J. Phys. Soc. Japan*, vol. 11, pp. 728–739; July, 1956.

[7] R. N. Hall, "Power rectifiers and transistors," Proc. IRE, vol. 40, pp. 1512–1518; November, 1952.

[8] T. H. Kinman, G. A. Carrick, R. G. Hibberd, and A. J. Blundell, "Germanium and silicon power rectifiers," *Proc. IEE*, vol. 103, Part A, pp. 89–111; April, 1956.

[9] E. Spenke, "Durchlass- und sperreigenschaften eines p-i-metall-gleichrichters," *Z. Naturf.*, vol. 11a, pp. 440–456; June, 1956.

[10] D. A. Kleinman, "The forward characteristic of the PIN diode," *Bell Sys. Tech. J.*, vol. 35, pp. 685–706; May, 1956.

[11] R. N. Hall, "Electron-hole recombination in germanium," *Phys. Rev.*, vol. 87, p. 387; July, 1952.

and by Shockley and Read,[12] who find that the lifetime τ depends on the injected carrier density δp according to

$$\tau = \frac{\tau_0(1 + a\delta p)}{1 + b\delta p} \tag{1}$$

where τ_0 is the low-level lifetime and a and b are constants depending on the characteristics of the recombination centers. This relation shows that τ initially rises or falls as δp is increased, but attains a constant value at high injection levels. Experiments on germanium[13] and silicon[14] indicate that, at any rate for the samples studied, τ rises initially before attaining its high level value. The result (1) is derived on the assumption that recombination at a trap can take place instantaneously, no excited levels being involved. If an excited level is involved in the trapping process, then the recombination process takes an appreciable time, and this could conceivably limit the recombination rate and lead to an increase in τ at extremely high levels. Burton, *et al.*,[15] have shown that a concentration of 10^{12} atoms/cm^3 of nickel or 10^{13} atoms/cm^3 of copper could explain the observed lifetimes of "good" germanium crystals ($\tau \sim 10^{-4}$ sec). If this were the case, and the recombination process occupied more than 10^{-10} sec, recombination rate limiting should be appreciable at carrier densities of the order of 10^{19} pairs/cm^3. There may be additional complications, if there is more than one type of recombination center, but we shall not consider this possibility here.

Recombination by way of traps is, however, only one possible recombination mechanism. The most important remaining mechanism is photon-radiative recombination. This has been treated by van Roosbroeck and Shockley[16] who show that the recombination rate is

$$R_c = \frac{np}{n_i^2} R \tag{2}$$

where, from optical data, for germanium

$$R = 1.57 \times 10^{13} \text{ cm}^{-3} \text{ sec}^{-1} \tag{3}$$

corresponding to a lifetime of 0.75 sec, in intrinsic material. At high levels $n = p$ so that (2) shows that $\tau = 100$ μsec for $n = 4 \times 10^{17}$ cm^{-3}. Since high level lifetimes are $\sim$100 μsec in some cases,[13] this additional mechanism may reduce the lifetime for carrier densities above 10^{18} cm^{-3}.

In our discussion we shall treat high level lifetimes as constant, but recognize the possibility of variations at very high carrier densities. The effect of such variations will be negligible in most devices.

Mobility

It is well known that charged impurities reduce the mobility of holes and electrons by scattering, and this effect becomes noticeable in germanium for impurity densities greater than about 10^{16} cm^{-3}. This effect has been treated on a classical approximation by Conwell and Weisskopf[17] who derive the expression (in terms now of a diffusion coefficient)

$$D_I = 2^{7/2}\pi^{-3/2}N^{-1}\kappa^2 q^{-4} m^{-1/2}(kT)^{5/2} \cdot [\log_e(1 + 9\kappa^2 N^{-2/3} q^{-4} k^2 T^2)]^{-1} \tag{4}$$

where N is the density of singly charged impurity centers, m is the effective mass of the charge carrier, and κ is the dielectric constant of the semiconductor. The logarithmic term is an approximation taking into account the finite range of interaction with the field of each ion. A more rigorous quantum mechanical treatment by Brooks[18,19] gives a different form for this cutoff term, but agreement between the two results is reasonably good in most cases.

The result (4) has also been given in the theory of gas diffusion. In addition, Chapman and Cowling[20] discuss the case of the mutual diffusion of two groups of charged particles. This case is applicable to the interaction of holes and electrons on a classical approximation. The first approximation to the result is, for $n = p$

$$D_{12} = 3\pi^{-1/2}2^{-3/2}n^{-1}\kappa^2 q^{-4}\left(\frac{m_1 + m_2}{m_1 m_2}\right)^{1/2}(kT)^{5/2} \cdot [\log_e(1 + 16\kappa^2 n^{-2/3} q^{-4} k^2 T^2)]^{-1} \tag{5}$$

where m_1 and m_2 are the effective masses of electrons and holes, respectively. The difference in the cutoff term is not significant, as we expect a quantum correction in any case. The result (5) differs from (4) by a factor 0.294, apart from the mass term. This difference is due to the fact that Chapman's analysis, resulting in (5), includes the effects of like carrier collisions, which were neglected in the Conwell-Weisskopf result (the Lorentz approximation). Debye and Conwell[19] discuss like carrier collisions for the impurity scattering case and suggest that (4) should be multiplied by a factor of about 0.6 to allow for this effect.

[12] W. Shockley and W. T. Read, "Statistics of the recombination of holes and electrons," *Phys. Rev.*, vol. 87, pp. 835–842; September, 1952.

[13] L. D. Armstrong, C. L. Carlson, and M. Bentivegna, "PNP transistors using high-emitter-efficiency alloy materials," *RCA Rev.*, vol. 17, pp. 37–45; March, 1956.

[14] G. Bemski, "Lifetime of electrons in p-type silicon," *Phys. Rev.*, vol. 100, pp. 523–524; October, 1955.

[15] J. A. Burton, G. W. Hull, F. J. Morin, and J. C. Severiens, "Effect of nickel and copper impurities on the recombination of holes and electrons in germanium," *J. Phys. Chem.*, vol. 57, pp. 853–859; November, 1953.

[16] W. van Roosbroeck and W. Shockley, "Photon-radiative recombination of electrons and holes in germanium," *Phys. Rev.*, vol. 94, pp. 1558–1560; June, 1954.

[17] E. Conwell and V. F. Weisskopf, "Theory of impurity scattering in semiconductors," *Phys. Rev.*, vol. 77, pp. 388–390; February, 1950.

[18] H. Brooks, "Scattering by ionized impurities in semi-conductors," *Phys. Rev.*, vol. 83, p. 879; August, 1951.

[19] P. P. Debye and E. Conwell, "Electrical properties of N-type germanium," *Phys. Rev.*, vol. 93, pp. 693–705; February, 1954.

[20] S. Chapman and T. G. Cowling, "The Mathematical Theory of Non-Uniform Gases," Cambridge University Press, Cambridge, Eng., pp. 177–179; 1952.

In Fig. 1 we have plotted the diffusion coefficients of electrons and holes in germanium as a function of injection level. Approximate effective mass values were taken as $\frac{1}{5}$ and $\frac{1}{4}$ respectively, and the diffusion coefficient given by (5) was combined with the lattice diffusion coefficient by addition of reciprocals.

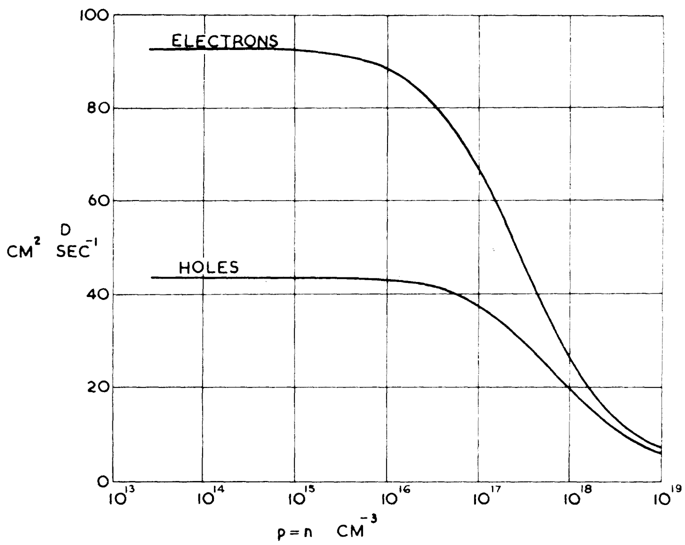

Fig. 1—Diffusion coefficients for electrons and holes in intrinsic germanium at 300°K, as functions of injected carrier density.

From this figure it is evident that hole-electron scattering is important at densities greater than 10^{17} cm^{-3}, and that at densities greater than 10^{18} cm^{-3} the electron and hole diffusion constants are essentially equal.

Of course, this whole approach represents a simple approximation, since the band structure of germanium is quite complex.[21] The assumption of an isotropic effective mass for electrons and of a single isotropic effective mass for holes is far from reality, but the effect of departures from these idealizations on the scattering process as calculated should not be large.

Degeneracy

All theories mentioned in the Introduction use Boltzmann statistics, and this is, in fact, necessary, if the results are to be obtained in simple form. This approximation leads to an error at high injection levels when the difference between Fermi and Boltzmann statistics is appreciable. This has been discussed in some detail by Shockley.[22]

From the formulas given by Herman[21] for the band structure of germanium, the effective numbers of states in the conduction and valence bands are

$$N_c \approx 1.2 \times 10^{19} \text{ cm}^{-3}$$

$$N_v \approx 3.8 \times 10^{18} \text{ cm}^{-3}.$$

[21] F. Herman, "The electronic energy band structure of silicon and germanium," Proc. IRE, vol. 43, pp. 1703–1732; December, 1956.

[22] W. Shockley, "Electrons and Holes in Semiconductors," D. Van Nostrand Co., Inc., New York, N. Y., pp. 230–249; 1950.

It is thus evident that degeneracy will first become appreciable in the valence band. From Shockley's discussion, the error in using Boltzmann statistics is less than a factor of two for predicted carrier densities less than about $4N$—a hole density of 1.5×10^{19} cm^{-3} for germanium.

In all cases, the classical result overestimates the carrier density in the band. This may have an appreciable effect at high carrier densities, but we shall use Boltzmann statistics in our discussion to follow, in order that our results may be presented in reasonably simple form.

Junction Relations

In all other theories, modulation of the conductivity of heavily doped regions has been ignored, but this is no longer valid in the high level limit. We shall now derive the correct relations between the carrier densities on either side of a PI junction. These results can be immediately extended to general PN junctions, but, since we only require the PI case, we shall restrict our discussion to this. The relations derived in this section form the basis of our treatment of the high level limit.

Boltzmann statistics show that if we have two identical sets of energy levels in thermodynamic equilibrium, but differing in energy by an amount E, then the populations of the two sets of levels will be related by

$$n_1 = n_2 e^{E/kT}. \tag{6}$$

Applying this to an abrupt PI junction, in which the equilibrium hole density in the P region is p_0 and the equilibrium hole or electron density in the I region is n_i, we find that

$$p_0 = n_i e^{q\phi/kT} = n_i e^{\theta\phi} \tag{7}$$

where ϕ is the "built in" potential difference across the junction. For brevity, we shall in future write θ for q/kT.

Now we are interested in the quasi-equilibrium or steady state case in which a current is flowing across the junction. It seems legitimate to extend the equilibrium relation (6) to cover this case, provided that the carrier density and potential gradients in the homogeneous regions are sufficiently small, for then there is no appreciable change in these quantities in one mean free path of a carrier. When this condition is not satisfied to a sufficient approximation, the relation (6) will be inaccurate, but may still be retained as a first approximation.

There is one further complication in that, when the equilibrium is disturbed in the way we are contemplating, the carrier densities tend to return to their equilibrium values with a characteristic lifetime. This will not affect our junction relations, provided the width of the junction transition region is much less than a carrier diffusion length. This will always be true for the type of junctions we shall consider. If we are to be able to treat the junction as negligibly thin, it is also necessary that the Debye length in the lightly doped regions be very much less than the minimum dimensions of these regions. Shockley[1] discusses this at some length and shows that the Debye length in intrinsic germanium at room temperature is $\sim 7\times 10^{-5}$ cm. Since this length varies as the inverse square root of the carrier density, it is truly negligible in the cases with which we shall deal.

Suppose that these conditions are fulfilled and that we apply a voltage V to the junction in the forward direction (P region positive). Then the total junction potential is reduced from ϕ to $(\phi - V)$ and (6) gives

$$p(P) = p(I)e^{\theta(\phi-V)} \tag{8}$$

for the relation between the hole populations on either side of the junction, and

$$n(P) = n(I)e^{-\theta(\phi-V)} \tag{9}$$

for the electron populations.

In place of the relation

$$np = n_i^2, \tag{10}$$

which holds for the true equilibrium case, we now have electrical neutrality relations for the I and P regions

$$\left.\begin{aligned} p(I) &= n(I) \\ p(P) - p_0 &= n(P) - \frac{n_i^2}{p_0} \end{aligned}\right\}. \tag{11}$$

If we neglect n_i^2/p_0 in comparison with p_0, then (7)–(9) and (11) can be combined to give

$$\left.\begin{aligned} p(P) &= p_0 A \\ n(P) &= p_0(A-1) \\ p(I) &= n(I) = n_i A e^{\theta V} \end{aligned}\right\} \tag{12}$$

where

$$A = \left[1 - \frac{n_i^2}{p_0^2}e^{2\theta V}\right]^{-1} = \left[1 - e^{2\theta(V-\phi)}\right]^{-1}. \tag{13}$$

These expressions will be the basis of all that follows.

We note that the relations (12) reduce to the usual low level relations[1] for $V \ll \phi$, but as $V \to \phi$ all the carrier densities approach infinity. This actually causes no trouble, as it proves impossible to apply a voltage equal to ϕ to the junction without applying an infinite voltage to the device as a whole.

It should also be pointed out that the relations (12) are only approximate in that they employ classical statistics and are based on the assumption of quasi-equilibrium. To gain some idea of the validity of this second assumption, we observe that a mean free path in germanium or silicon at room temperature is about 10^{-5} cm.[23] Since we customarily consider regions greater than 10^{-3} cm in width, we can thus tolerate quite large carrier density variations without invalidating our assumptions.

[23] W. Shockley, "Electrons and holes in semiconductors," *op. cit.*, p. 215.

The RPINR Diode

As our first application of the theory we shall consider the case of the RPINR diode shown in Fig. 2. The two combining surfaces bounding the low resistivity regions have been introduced, both to simplify the theory and to represent a closer approach to reality, for alloy junctions, than the assumption of semi-infinite end regions. We shall define a recombining surface to be such as to maintain carrier densities at their equilibrium values at this surface, and to offer no barrier to carrier flow. In practice this could be attained by a thin semiconductor region of very low lifetime backed by metal.

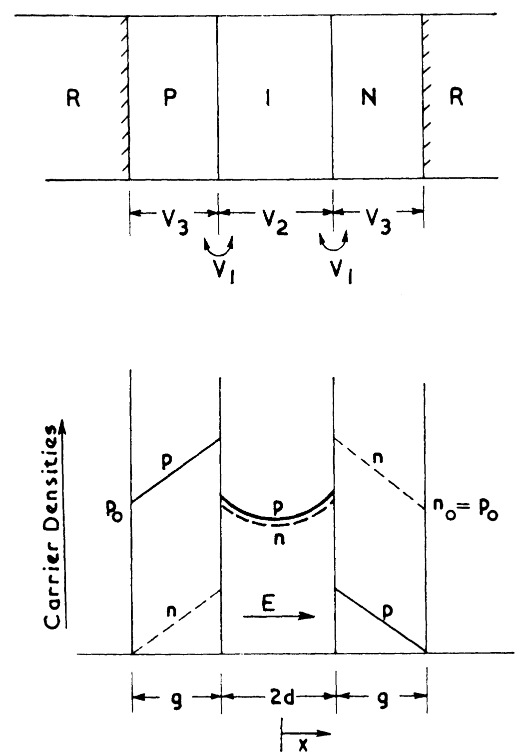

Fig. 2—The RPINR diode.

We make the simplification of neglecting volume recombination in the P and N regions, but suppose that carriers in the I region recombine with a constant high level lifetime τ. Neglect of recombination in the P and N regions makes no fundamental difference to the results. Current flow will be assumed one dimensional throughout.

With these assumptions it is possible to carry through the calculation quite generally, but the arithmetic is forbiddingly complicated. In order to elucidate the behavior of the diode, we make the simplifying assumption that electron and hole mobilities are equal and that the conductivities of the N and P regions are also equal. The first of these assumptions is not as restrictive as it first appears, since Fig. 1 shows it is reasonably true for intrinsic germanium at injection levels of 10^{17} cm^{-3} or greater, and, in any case, it will introduce only small quantitative variations from reality. We allow for impurity and carrier scattering in the N and P regions by writing the diffusion coefficient in these regions as βD, and for hole-electron scattering in the I region by writing αD. The approximate reciprocal rule for addition of diffusion coefficients then gives

$$\left.\begin{aligned}\frac{1}{\alpha D} &= \frac{1}{D} + \frac{1}{D_{12}} \\ \frac{1}{\beta D} &= \frac{1}{D} + \frac{1}{D_{12}} + \frac{1}{D_I}\end{aligned}\right\} \tag{14}$$

where D is the diffusion coefficient due to lattice scattering, D_{12} accounts for hole-electron scattering and is given by (5), and D_I accounts for impurity scattering and is given by (4). Since holes and electrons have different scattering environments in the N and P regions, they will have slightly different β values, but we shall use a single average value in accord with our proposed approximation.

The problem is now essentially symmetric about the center of the I region, and the dimensions and carrier densities are as shown in Fig. 2. The applied voltage, V, may be conveniently divided into components as shown in the upper part of the figure. Within the I region $n = p$, so that the carrier density satisfies

$$\frac{d^2 n}{dx^2} = \frac{n}{L^2} \tag{15}$$

where

$$L^2 = \alpha D \tau. \tag{16}$$

Using (12) we then find

$$n = n_i A e^{\theta V_1} \operatorname{sech} \frac{d}{L} \cosh \frac{x}{L}, \tag{17}$$

where A is defined by (13) with V_1 substituted for V. Within the I region the hole and electron current densities are given by

$$J_p = qp\alpha\mu E - q\alpha D \frac{dp}{dx} \tag{18}$$

$$J_n = qn\alpha\mu E + q\alpha D \frac{dn}{dx} \tag{19}$$

where currents are positive flowing to the right in Fig. 2. The same equations apply to the P and N regions with α replaced by β. The requirement that hole current and electron current be continuous across the junctions determines E on each side of the junctions, and with the aid of (17) the total diode current density is found to be

$$J = 2qn_i A \alpha D e^{\theta V_1} Q \tag{20}$$

where

$$Q = \frac{2A - 1}{L} \tanh \frac{d}{L} + \frac{\beta}{\alpha} \frac{n_0}{n_i} \frac{1}{g} \frac{(A - 1)}{A} e^{-\theta V_1}. \tag{21}$$

We should note that for the case of small currents ($V_1 \ll \phi$) this expression reduces exactly to Hall's expression,[7] if we put $b = 1$, as required by our assumptions, and neglect scattering.

By adding (18) and (19), E is found as a function of J in the I region, and this can be immediately integrated to give

$$V_2 = 4\theta L \left[(\tan^{-1} e^{d/L}) - \frac{\pi}{4} \right] \cosh \frac{d}{L} \cdot Q \tag{22}$$

and this is again seen to reduce to Hall's expression[7] for V_1 small.

The field in the P and N regions can be found similarly, if we observe that the carrier distribution in these regions is linear to a good approximation. Integration then yields

$$V_3 = \theta \frac{\alpha}{\beta} \frac{n_0}{n_i} g e^{-\theta V_1} \log_e [2A - 1] Q \tag{23}$$

which, for small V_1, is found to reduce to the simple ohmic voltage drop over the region concerned. The total voltage drop across the diode is

$$V = 2V_1 + V_2 + 2V_3. \tag{24}$$

We may also note in passing that if we set $d = 0$, then (20) reduces to

$$J = \frac{2qn_0\beta D}{g}(A - 1) = \frac{2qn_0\beta D}{g}\left[e^{\theta(2V_1 - 2\phi)} - 1\right]A \tag{25}$$

which is the correct result for a symmetrical RPNR diode with total built in potential 2ϕ and total applied junction voltage $2V_1$.

Diode Characteristics

The calculated forward characteristics for typical diodes with I regions of various widths are shown in Fig. 3. Values of the other parameters used are

$$g = 0.01 \text{ cm}, \quad \tau = 50\ \mu\text{sec}, \quad n_0 = p_0 = 10^{18}\ \text{cm}^{-3},$$
$$T = 300^\circ \text{K}, \quad D = 50\ \text{cm}^2/\text{sec}$$

other parameters being approximately those for germanium. It will be observed that for thick diodes the current commences as $\exp(\frac{1}{2}\theta V)$ and at high levels falls off well below this value. For diodes of intermediate thickness, the characteristic is initially $\exp(\frac{1}{2}\theta V)$, but rises above this value for moderate currents and over quite a large range is represented fairly well by $\exp(\lambda\theta V)$ where $\frac{1}{2} \leq \lambda \leq 1$. Again the current falls off at very high levels. When $d = 0$, the characteristic is $\exp(\theta V)$ at all moderate levels until it finally drops off at very high levels.

Qualitatively the behavior may be expressed as follows. At low levels Hall's analysis is accurate and substantially all current flow is due to recombination in the I region. As more voltage is applied, the carrier concentration in the I region rises and the injection efficiency falls. The current flowing over both junctions varies approximately as $\exp(\theta V)$ and increases the slope of the curve at higher levels. At very high currents the voltage

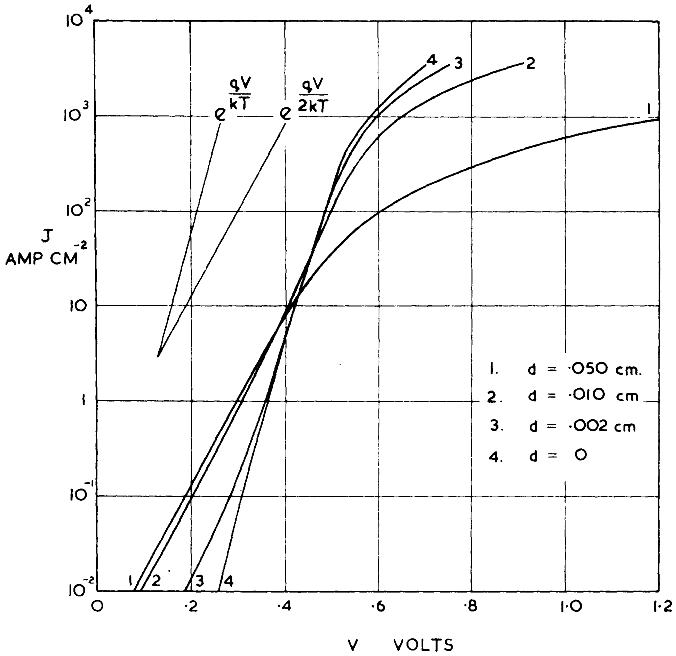

Fig 3—Theoretical RPINR diode characteristics.

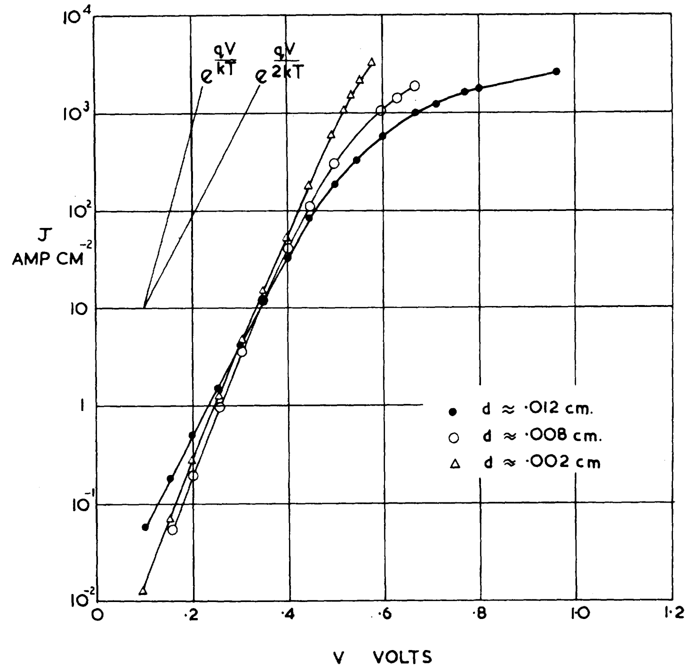

Fig. 4—Experimental characteristics for In-Ge-(SnPbSb) diodes.

drop across the I region increases greatly, voltage drops across the P and N regions become appreciable, and the curve flattens out. Hole-electron scattering magnifies this flattening by reducing the current flow for a given carrier density.

We should remark further that if the "intrinsic" region has an appreciable donor or acceptor concentration, then the low level characteristic will have the typical $[\exp(\theta V)-1]$ shape, so that there may be no $\exp(\frac{1}{2}\theta V)$ region for rather thin diodes made with nonintrinsic material. We should also point out that our analysis is not valid, even for truly intrinsic material, at currents comparable with the reverse saturation current of the device. This is not an inherent limitation of the theory, but arises because, in the interests of simplicity, we have neglected certain small terms.

Similar behavior is shown in the experimentally determined curves of Fig. 4. Diodes were made from germanium wafers of about 3 ohm cm resistivity by alloying indium to one face and a mixture of lead and tin in eutectic ratio with 5 per cent of added antimony to the other. Precautions were taken to reduce the ohmic resistance of contacts to the diode to a negligible value, and measurements were taken in a liquid bath to minimize heating effects. Since no attempt has been made to relate accurately the parameters used for the theoretical curves to those of the measured diodes, Figs. 3 and 4 should only be compared qualitatively.

Design Considerations

The results of this analysis can be applied with advantage to optimizing the design of high current diodes. Here, two things must be taken into account: the forward voltage drop at a given large current must be as small as possible, and the reverse characteristics must be adequate from both leakage and breakdown points of view.

First consider forward currents. From Fig. 3 there is clearly an optimum value of d for a general purpose diode (in this case $d \approx 0.01$ cm). For d values much larger than this, operation becomes very inefficient at high levels, while for d much smaller, operation becomes inefficient at low currents, though this is much less important and can usually be neglected.

The optimum values of g and p_0 are much more difficult to determine and depend on all the other parameters, as well as the current level at which the diode is required to work. Too large a value of g obviously makes V_3 excessive, whilst too small a value can reduce the injection efficiency of the junctions to such an extent that V_2 becomes excessive. Similarly, a low value of p_0 reduces emitter efficiency and increases V_2 and V_3, so that the curve flattens off at a relatively low current. For a very large value of p_0, the current will show little rise above $\exp(\frac{1}{2}\theta V)$, but will maintain this slope to very high current densities.

Actually, values of p_0 and g are almost completely determined by the alloying process and they cannot be varied over very large ranges without considerable inconvenience. The worth of a particular change can be determined by explicit calculation or perhaps more easily by experiment.

Now consider the reverse characteristic. Saturation and leakage currents are determined primarily by the semiconductor surface, so that changes in d have relatively little effect, provided both junctions are of good quality. The contribution of the volume to the saturation current will in any case be approximately proportional to d.

Breakdown of junctions, whether by avalanche or by Zener mechanism, is characterized by a critical field E_c, dependent in magnitude generally upon the exact impurity distribution in the junction. For the case of abrupt junctions such as we expect from the alloying process, the field distribution under reverse bias conditions is as shown in Fig. 5.

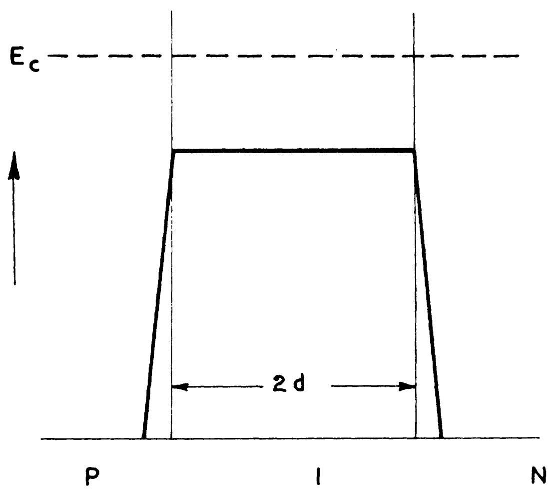

Fig. 5—Field distribution in reverse-biased PIN diode.

The voltage across the diode is given by the area under the curve, and the breakdown voltage is clearly

$$V_c = 2dE_c. \tag{26}$$

If the I region is not truly intrinsic, but lightly doped, then the top of the curve will be sloped rather than flat, and the breakdown voltage will be reduced somewhat.

For an abrupt junction on germanium, the critical field is of the order of 10^5 volts/cm[24] so that from (26) $2d \sim 10^{-5}\ V_c$. For a diode with 300-volt breakdown, we thus require $2d > 0.003$ cm and from Fig. 3 this value is small enough to give a characteristic with slope substantially θ.

The RPIR Diode

We now consider the high level theory of the RPIR diode. The analysis in this case can be carried through quite generally, though it is desirable to neglect all volume recombination if the results are to be reasonably manageable. This will have little effect on the results, provided the width of the regions considered is substantially less than a diffusion length. In the analysis whose results are presented below, we have also omitted the effects of impurity and collision scattering, since this complicates the result without affecting it very much.

The geometry of the diode is shown in Fig. 6, together with applied voltages and carrier densities.

We proceed very much as we did in the preceding section. The carrier densities in the two regions are linear to a good approximation and we can write down the equations of current flow. The requirement of continuity of hole and electron current across the junction gives the fields in the P and I regions at this point and, from the equations of current flow, the fields everywhere can be determined and integrated to give the voltage drops.

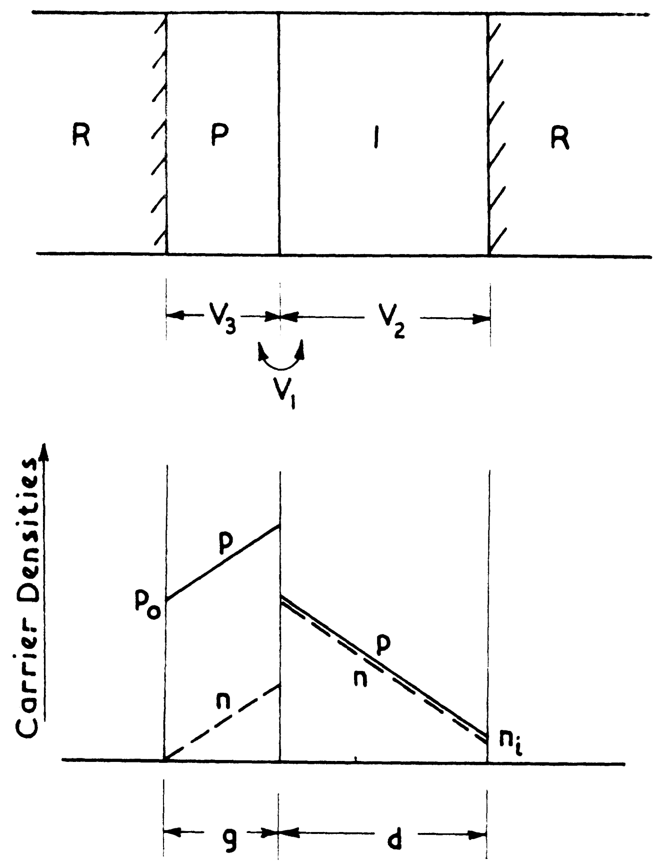

Fig. 6—The RPIR diode.

Summarizing the results we find

$$J_p = qAn_iD_pe^{\theta V_1}\left\{\frac{2A}{d} + \frac{(2A-1)}{g}\sqrt{\frac{A-1}{A}}\right\} \tag{27}$$

$$J_n = qAn_iD_ne^{\theta V_1}\left\{\frac{2(A-1)}{d} + \frac{(2A-1)}{g}\sqrt{\frac{A-1}{A}}\right\} \tag{28}$$

$$V_2 = \theta d\left\{\frac{2A-1}{d} + \frac{(2A-1)}{g}\sqrt{\frac{A-1}{A}}\right\} \cdot \left\{\theta V_1 + \log_e A\right\} \tag{29}$$

$$V_3 = \theta g\left\{\frac{2A-1}{g} + \frac{(2A-1)}{d}\sqrt{\frac{A}{A-1}}\right\} \cdot \log_e (2A-1) \tag{30}$$

$$V = V_1 + V_2 + V_3, \tag{31}$$

[24] R. D. Knott, I. D. Colson, and M. R. P. Young, "Breakdown effect in p-n alloy germanium junctions," *Proc. Phys. Soc. B (London)*, vol. 68, pp. 182–185; March, 1955.

where again A is a function of V_1. These results all reduce to the usual simple forms[8,9] for $V_1 \ll \phi$.

Diode Characteristics

In Fig. 7 we have plotted the J-V characteristic of an RPIR diode for typical parameter values, neglecting the effects of scattering. Values used were

$$g = 0.01 \text{ cm}, \quad d = 0.02 \text{ cm}, \quad p_0 = 10^{18} \text{ cm}^{-3}, \quad T = 300°\text{K}$$

other constants being appropriate to germanium.

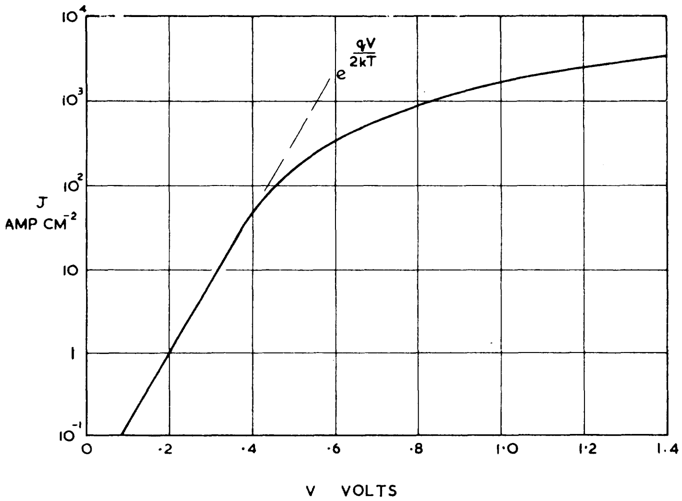

Fig. 7—Theoretical characteristic for RPIR diode (neglecting scattering).

If we had included scattering effects, region I mobilities would have been reduced about a factor of two at the top of the curve, so that the over-all current scale would be compressed by about this same factor, the curve shape being little altered.

It is seen that the curve, plotted on a logarithmic scale, is initially a straight line of slope $\frac{1}{2}\theta$, but soon falls well below this value. In this case, there is no rise above the exp $(\frac{1}{2}\theta V)$ line at any stage. The shape of this characteristic is simply explained as follows. At relatively low levels all the current is carried by holes. Half of the applied voltage appears as V_1 across the junction and half as V_2 across the I region. This gives an exp $(\frac{1}{2}\theta V)$ characteristic. At higher levels the emitter efficiency falls, and appreciable current is carried by electrons. This leads to a large increase in the voltage drop in the part of the intrinsic region near the recombination electrode, since carrier density is quite low here. V_2 thus increases and the curve flattens out.

Several diodes have been made by alloying indium to one side of a germanium wafer and pure lead or tin to the opposite side. The characteristics all have the general shape predicted above. In practice, for non-intrinsic material, the straight line portion of the curve is not very evident due to transition to the [exp $(\theta V)-1$] characteristic which applies at low levels. A typical curve, in this case for a diode with a tin recombination electrode, is shown in Fig. 8. Again the parameters are different from those of the theoretical calculation and only qualitative comparison should be made.

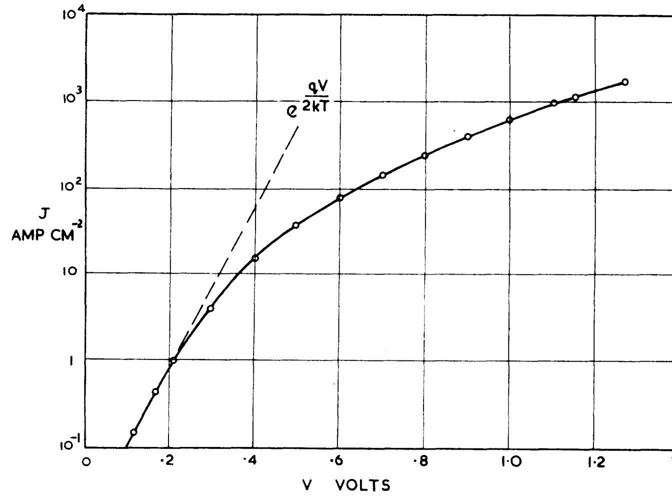

Fig. 8—Experimental curve for In-Ge-Sn diode with $d \approx 0.015$ cm.

Design Considerations

From the formulas derived above and the curves of Figs. 3, 4, 7, and 8, it is clear that the RPIR diode is somewhat inferior to the RPINR diode as a high current device. The reverse characteristic is also inferior,[9] in that a surface of high recombination velocity exists close to the reverse biased junction, and is not shielded from it by low resistivity material as in the case of the RPINR diode.

For optimum operation the intrinsic region should be as thin as allowed by breakdown considerations, and the P region doping should be as heavy as possible to maintain injection efficiency. The best attainable characteristic will not rise above the exp $(\frac{1}{2}\theta V)$ level.

Transistor Emitter Efficiency

Current flow in a junction transistor is in many ways analogous to current flow in an RPIR diode, but there is an important difference, in that the transistor is a three-terminal device and an appreciable amount of current may flow to the base contact.

Fig. 9 shows the carrier distributions in a PNP transistor under normal operation conditions. For simplicity we shall assume that the base region is intrinsic, since this assumption has no effect on the high level performance. If the collector junction is at a potential $-V_c$ with respect to the base, then the carrier density in the base at the collector junction is n_i exp $(-\theta V_c)$, which we may neglect in comparison with injected densities. We have again assumed that a diffusion length in the emitter is long compared with its width and have assumed a recombination surface bounding the emitter region. We shall note later that an identical result is obtained if we assume a long emitter with minority carrier diffusion length g, and, of course, the two cases can be combined, if g is taken to be an appropriately weighted length.

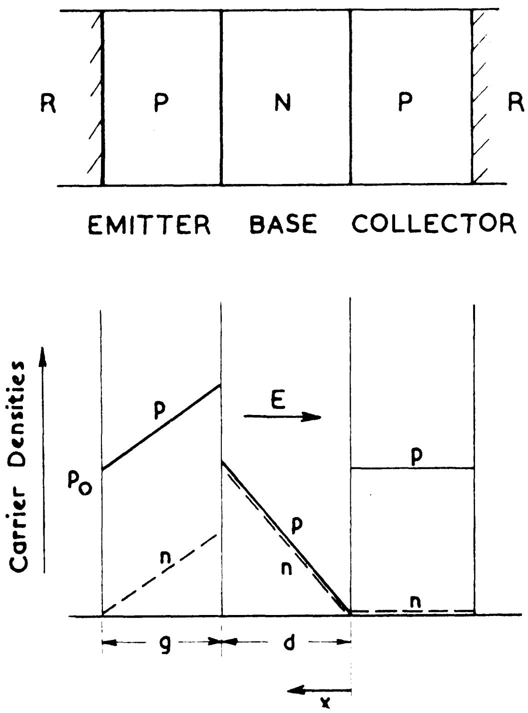

Fig. 9—RPNP transistor.

We shall treat this case quite generally, writing δD_p and ϵD_n for diffusion coefficients in the base region, and ζD_p, ηD_n for diffusion coefficients in the emitter; δ, ϵ, ζ, η are determined from scattering theory. This is not strictly accurate, since the diffusion coefficients are not constant across either base or emitter regions, but it is a sufficient approximation for our purpose. We shall take conventional currents as positive when flowing from emitter to collector and measure x as shown. Hole currents will be taken to be one dimensional, and the self-bias effect[25] of the transverse electron current will be neglected.

The electron current in the base region is not independent of x, since none of it flows through the collector junction. All of the electron current, in fact, flows transversely through the base region to the base lead. The conductivity of the various parts of the base layer depends on the injected carrier density, and if we assume this to vary approximately linearly as shown, then

$$J_n(x) = J_n(d)\,\frac{x^2}{d^2}\,. \tag{32}$$

Small departures from this assumed linearity have little effect on the final result.

Now we proceed as usual to find the fields in the base region at the emitter junction by requiring continuity of hole and electron currents across this junction. These field expressions both contain the term $dp/dx\big|_{x=d}$. To eliminate this, we observe that, from the equations of current flow, since $n=p$ in the base region,

$$J_p(d) = qE\delta\mu_p p + q\delta D_p \frac{dp}{dx} \tag{33}$$

$$\frac{x^2}{d^2} J_n(d) = qE\epsilon\mu_n p - q\epsilon D_n \frac{dp}{dx} \tag{34}$$

so that

$$\frac{dp}{dx} = \frac{1}{2qD}\left[\frac{1}{\delta} J_p(d) - \frac{1}{b\epsilon}\cdot\frac{x^2}{d^2}\cdot J_n(d)\right] \tag{35}$$

where D is D_p and $b = D_n/D_p$. Now integrating (35) from 0 to d we get, using (12) for a forward voltage V applied to the emitter junction,

$$An_ie^{\theta V} = \frac{1}{2qD}\left[\frac{d}{\delta} J_p(d) - \frac{1}{3}\,\frac{d}{b\epsilon} J_n(d)\right] \tag{36}$$

which may be used in conjunction with (33) and (34) for $x=d$ to eliminate the unknown gradient. The final results are

$$J_p = \frac{\dfrac{6q\delta D_p}{d}\,\dfrac{\epsilon}{\eta} A^2 n_i e^{\theta V} + \dfrac{q\delta D_p p_0}{g}(2A-1)(A-1)}{3\dfrac{\epsilon}{\eta}A - \dfrac{\delta}{\zeta}(A-1)} \tag{37}$$

$$J_n = \frac{\dfrac{6q\epsilon D_n}{d}\,\dfrac{\delta}{\zeta} A(A-1)n_i e^{\theta V} + \dfrac{3q\epsilon D_n p_0}{g}(2A-1)(A-1)}{3\dfrac{\epsilon}{\eta}A - \dfrac{\delta}{\zeta}(A-1)} \tag{38}$$

where A is given by (13) and the junction voltage V appears as a variable parameter.

Since the only place the emitter length g enters the problem is in the expression for the electron density gradient in the emitter, we see that if we assume instead a long emitter with electron diffusion length L_e, then the hole and electron currents are again given by (37) and (38) with g replaced by L_e.

We observe that, for $V \ll \phi$, (37) and (38) reduce to their low level form. If we define a dc or large signal emitter efficiency by

$$\Gamma = \frac{J_p}{J_p + J_n}, \tag{39}$$

then, for $V \ll \phi$, Γ reduces to the large signal value derived from Webster's analysis.[3,26]

For large currents the expressions (37)–(39) give a different result from that obtained by Webster. This is not surprising, since Webster's treatment is inapplicable when Γ is appreciably less than unity. In fact, his treatment, invalidly extended to large currents, predicts a

[25] N. H. Fletcher, "Self-bias cutoff effect in power transistors," Proc. IRE vol. 43, p. 1669; November, 1955.

[26] N. H. Fletcher, "Note on the variation of junction transistor current amplification factor with emitter current," Proc. IRE, vol. 44, pp. 1475–1476; October, 1956. This note points out the origin of an error of a factor $\frac{1}{2}$ in Webster's analysis.

limiting Γ of zero which is clearly incorrect. From the present treatment we find that for large currents

$$\Gamma \to \frac{3g + d}{6g + 4d}. \tag{40}$$

This limit is independent of whether the transistor is PNP of NPN, since, at very high levels, scattering makes electron and hole mobilities equal.

In Fig. 10 we have plotted Γ as a function of $J = J_n + J_p$ for the case of a typical PNP germanium transistor. Parameters used were

$$p_0 = 10^{18} \text{ cm}^{-3}, d = 5 \times 10^{-3} \text{ cm}, L_e(=g) = 10^{-3} \text{ cm}$$

which are approximately the values for the transistor discussed by Webster.[3] We observe that Webster's result is a surprisingly good approximation up to current densities of 10^4 amp/cm² in this case. This good agreement is fortuitous since Webster neglected both J_n and all scattering effects. Agreement remains reasonably good for other values of the base width so that Webster's result is a useful approximation for "ordinary" transistors, even under conditions where his assumptions are far from valid.

We should point out that, since both surface and volume lifetime effects saturate to small values at relatively small current densities, the dc high level value of I_c/I_e is essentially just Γ.

Experimental verification of the curves shown in Fig. 10 is difficult, since the base current causes a self-bias effect[25] which considerably disturbs the injected current density distribution at high levels. However, pulse measurements at very high currents show that Γ tends to a finite value, which is approximately that predicted by the theory.

Fig. 11 shows a typical curve of J_p vs J_n (*i.e.*, collector current vs base current) found by pulse measurement of a small PNP power transistor. The straight line portion for various transistors gives values of dI_c/dI_b ranging from 0.38 to 0.44 in excellent agreement with the value near 0.4, predicted by our theory and an approximate knowledge of L_e and d for these particular transistors. The transistors used were selected because of their rather wide base width (>0.01 cm), this making the currents more easily manageable. Because of the effect referred to above, true current densities will be much higher than the nominal values shown.

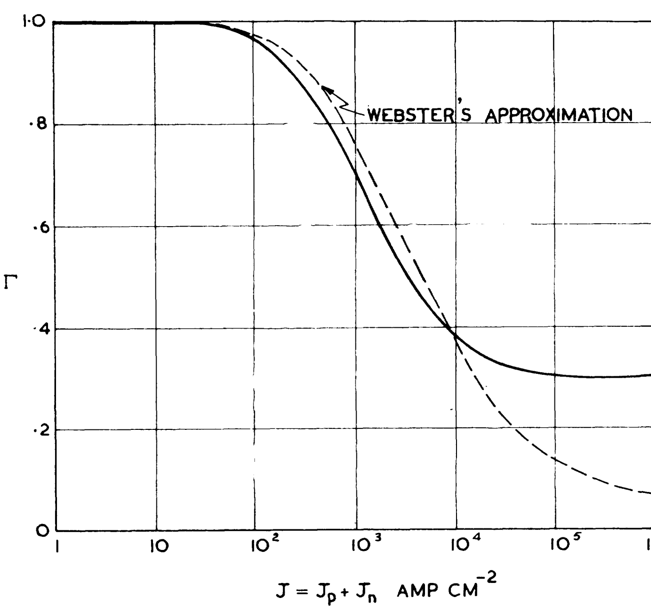

Fig. 10—Theoretical variation of dc emitter efficiency with emitter current for typical PNP transistor.

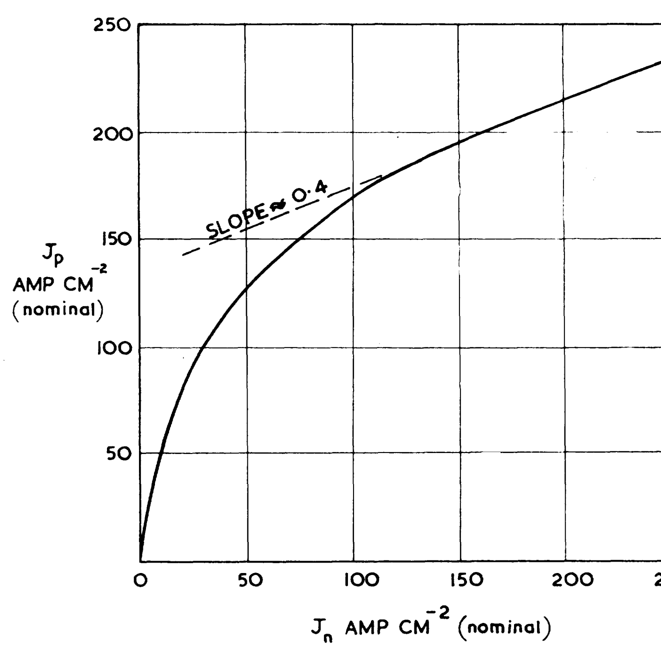

Fig. 11—Experimental relation between hole and electron current densities (derived from collector and base currents) for typical small PNP power transistor with base width ~0.01 cm.

Practical Applications

Most of our discussion so far has been concerned with device design parameters and their effect on the electrical performance of the device from the point of view of the designer. It seems appropriate to conclude with some remarks on the use of semiconductor devices at very high levels in the light of our new results.

We should first emphasize that all our results and calculations have been on the essential semiconductor structure alone and held at a fixed temperature. In practice, allowance has to be made for lead resistances and for heating effects. We have also restricted ourselves to dc considerations so that further analysis is necessary, if transient conditions are to be adequately treated.

To deal first with diodes, we have shown that, for general use, a diode with a PIN structure is usually preferable to one with a PIR structure, and that PIN diodes can be made whose forward current is fairly accurately exponential over large ranges. For example, the thinnest diode shown in Fig. 4 has an exponential characteristic over a current range of five decades, and this behavior is well supported by theory.

The usefulness of this characteristic is limited in practice by heating effects. On the other hand, for pulsed or low duty cycle circuits with time constants sufficiently long that transient effects can be neglected, the large extent of the exponential curve may be of value. With the advent of suitably designed alloyed silicon diodes it should be possible to extend the exponential characteristic several decades, because of the smaller reverse currents.

It may be worth noting in passing that most PIR diode characteristics, for example that of Fig. 8, are approximated quite well by power law curves of the form $J=AV^n$ where n lies between about 3 and 4. The fit is in some cases quite good and extends over as much as four decades of current. These may provide useful approximate power law elements in some applications.

When simple current rectification is required, diode performance is usually limited by cooling difficulties rather than by electrical limitations. For example, from Figs. 3 and 4 we can make a PIN rectifier which maintains its exponential characteristic to 1000 amps/sq cm. However, at this level the dissipation is 500 watts/sq cm for our best rectifier. This requires considerable cooling and in practice it is often convenient to spread the dissipation over a larger area which then operates at a lower current density. Typical rectifiers are not usually run at more than a few hundred amps/sq cm and often very much less if high reverse voltages are encountered.

While it is not usual to operate transistors in current ranges for which α is less than about 0.9, it is useful to know the variation of α with current at very high levels. Our analysis has shown that a naive application of Webster's formula is valid for α greater than about 0.5 for a PNP transistor, but that the ultimate α is about 0.3. Almost the same thing is true for NPN transistors on an extended current scale. The self-bias effect which we have discussed before causes further apparent variations from Webster's result for large transistors and may considerably limit the maximum current attainable. Again heating effects are very important and usually limit the allowable current before electrical restrictions become important. Alloys for improved emitter efficiency should make low α operation relatively rare in the future.

Conclusion

At very high current densities new formulas must be derived considering previously neglected effects. Of these the most important is a more accurate consideration of carrier density relationships across a junction biased in the forward direction. Formulas derived using these relations give a good description of the high current behavior of junction devices and explain the deviations from previous high level theories.

The effect of hole-electron collisions has been shown to be considerable at high carrier densities. Its effects are only secondary in determining the device characteristics, but it must be considered in any accurate calculation.

Possible further complications arising from the high carrier concentrations are departures from classical statistics (degeneracy) and the possibility of radiative recombination reducing carrier lifetime. These are found to be of minor importance at the carrier levels usually encountered in practice.

The predictions of the theory for particular devices are in good agreement with experiment and show in what manner the available parameters should be varied to achieve specific results. This is true even for practical diodes which are only approximations to the ideal diodes of the theory. In particular, we have shown both theoretically and experimentally that it is easily possible to construct diodes with exponential characteristics over as much as five decades of current and up to current densities as high as 3000 amps/sq cm, without using special alloying materials.

While in many practical applications thermal considerations forbid operation at these extreme current densities, application of the theory allows most efficient design at the current level chosen.

Acknowledgment

It is a pleasure to record my thanks to the people who have helped with this study, particularly to H. Flood, who made all the diodes used as well as the special alloying and measuring jigs, to F. Tonking, who performed the pulse measurements on transistors, and to Dr. L. W. Davies for helpful discussions.