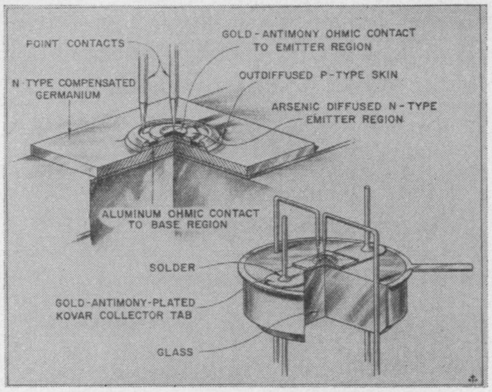


Fig. 10—Artist's representation of a typical *n-p-n* double diffused transistor.

Conclusions

The diffusion of antimony out of germanium has been studied and the use of this process has been shown to be an easy way of producing *p-n* junctions in compensated *n*-type germanium. Outdiffusion has been used to produce high-speed narrow base diodes and high-speed *n-p-n* germanium transistors. While diodes and transistors which have been fabricated using this technique have many desirable and presently unavailable electrical characteristics, the potentialities of the process have been by no means exhausted, either for the production of these or other types of semiconductor devices.

Acknowledgment

The authors wish to express their appreciation to F. M. Sullivan and C. R. Grant for help in fabricating the diodes and transistors and to P. L. Moody and S. A. Kulin for growing the compensated germanium. They are also indebted to E. Chatterton for designing the circuits used to determine the frequency response of the transistors and for taking these measurements. The advice and encouragement of D. T. Stevenson is gratefully acknowledged.

The Evolution of the Theory for the Voltage-Current Characteristic of *P-N* Junctions*

J. L. MOLL, MEMBER, IRE†

Summary—**The rectifying action in semiconductor *p-n* junctions is controlled essentially by the equilibrium densities, diffusion constants, and recombination times of minority carriers. The low-level behavior of germanium junctions at room temperature is adequately described by a theory which is based on the rate of diffusion and recombination of minority carriers on either side of the barrier region. It is necessary to include the effects of carrier recombination and generation in the barrier region to explain the low-level behavior of silicon.**

At high current densities, the junctions depart from the ideal low-level rectifier law because of effects associated with majority carrier modulation. As a consequence of recombination current to the barrier region at low levels and conductivity modulation effects at high levels, the simple $I_s(\exp qV/kT-1)$ behavior is rarely observed in silicon junctions at room temperature.

EVER since the observation of nonohmic behavior in metal-semiconductor contacts, the theoretical explanation of rectification has received a great deal of attention. Most of the theories that were proposed prior to Shockley's definitive article[1] in 1949 contained some features in common with modern theory. However, the theories generally omitted the important role of minority carriers and, for this reason, can be considered of value only to the extent that they helped workers in the field to reach present-day concepts. There is one very notable exception to the preceding statement, *viz.*, a theoretical paper by Davydov in 1938.[2] In this paper Davydov considered most of the essential features of modern low-level rectification theory and specifically pointed out the importance of minority carriers in determining the rectifying action. He also included the concepts of nonequilibrium density and finite lifetime, with perhaps the most notable lapse being the assumption that equilibrium minority and majority carrier density was related to lifetime.

An extremely abbreviated list of pre-1949 contributions would include such names as Wilson and Nordheim who independently proposed the tunnel effect as an explanation of rectification.[3,4] This theory was quickly discarded since it gave the wrong direction of

* Original manuscript received by the IRE, April 4, 1958.
† Bell Telephone Labs., Inc., Murray Hill, N. J.
[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] B. Davydov, "The rectifying action of semiconductors," *Tech. Phys.*, (*USRR*), vol. 5, pp. 87–95; February, 1938.

[3] A. H. Wilson, "A note on the theory of rectification," *Proc. Roy. Soc.*, vol. 136, p. 487; May, 1932.

[4] L. W. Nordheim, "Xur Theorie der Detector wirkung," *Zeits. f. Phys.*, vol. 75, p. 434; April, 1932.

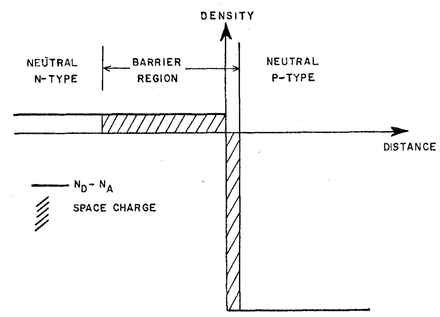
rectification. However, Esaki has noted rectification by tunneling in extremely thin (*i.e.*, highly doped) junctions.[5]

The possibility of relatively thick barriers between the p and n-type regions was also considered at an early date, and the role of impurity density in determining barrier-layer thickness was recognized.[6,7] If Davydov's paper is disregarded (as was generally done at the time even though it was published in English in 1938), the general view of rectification until 1949 was that n-type and p-type semiconductors were "boxes" of Maxwell-Boltzmann gases of electrons and holes, respectively, with a potential barrier at the p-n boundary. The potential barrier keeps the two gases from spilling across the boundary, and the thickness of the potential barrier is determined from the height of the barrier and the density of ionized impurities. As the applied voltage is changed, escape of particles from the two gases becomes more or less difficult depending on the direction of the applied voltage.[8] Now this view of the rectifying action gives the right qualitative result of an exponential law for rectification and is, in fact, consistent with present theory. However, the theory which has been built up with Shockley's 1949 paper[1] as a basis is more useful for understanding p-n junctions. In Shockley's theory the elementary processes of the escape of the Maxwell-Boltzmann gas have been separated, with the result that the various modifications of the theory which are necessary to fit specific situations are easier for the intuition to grasp. With Shockley's theory it becomes possible, for the first time, to make a quantitative check of the theory in terms of the basic semiconducting properties of the component parts of the junction.[9]

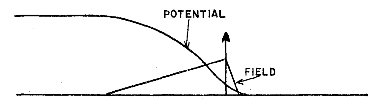
The escape of an electron from the n side to the p side of a junction involves:

1) Surmounting the potential barrier at the junction,
2) Disappearance on the p side by recombination (or removal across a reverse junction in the case of a transistor).

In addition to the process of escape of carriers with recombination, there is, of course, the inverse process of generation of minority carriers which diffuse to the junction where they are swept down the potential barrier. Where the generation is thermal, the process of generation appears as simply a negative recombination or vice versa. If the generation is by means of photons or other particles, it is usually easiest to treat the generation as a separate process. It is not our object here to treat any situations other than the rectification of p-n junctions.



(a)



(b)

Fig. 1—Idealized planar step junction. (a) Assumed distribution of impurity distribution. It is assumed for purposes of analysis that all of the mobile carriers have been swept out of the barrier region resulting in a space-charge density which is simply the ionized donor or acceptor density. (b) Field distribution and potential in a step junction. The width of the barrier region adjusts itself so that the resulting dipole will sustain the contact potential plus the applied voltage.

Following Shockley's treatment[1] the mathematical model which is used to represent the p-n junction is illustrated with the aid of Fig. 1. The junction and infinitely long pieces of p-type and n-type semiconductor are divided into three separate regions: 1) the neutral n-type region, 2) the neutral p-type region, and 3) the barrier region, or depletion region on both sides of the p-n boundary. It is assumed that the majority carriers have been essentially all swept out of the barrier region, and that a space charge of the ionized donors and acceptors exists. The essentially field-free nature of the neutral region requires that the positive charge per unit area of the donors on the n side of the boundary be balanced exactly by the negative charge per unit area of the acceptors on the p side. Thus, the barrier region extends far enough on either side of the junction to sustain the barrier potential, with the relative extent of the region on the two sides being governed by the requirement of net neutrality. The validity of the assumption of an abrupt transition from neutral material to swept-out material is considered in some detail in Shockley's paper[1] where the assumption is shown to be valid for junctions with steep concentration gradients. For junctions with sufficiently shallow concentration gradients, there is no space-charge region or rectification. For the case of the very shallow gradient, the change in equilibrium carrier concentration with distance is so slow that it is not possible to disturb the equilibrium with applied field, and there is no rectification. The change in equilibrium conditions at the p-n boundary is

[5] L. Esaki, "A new phenomenon in narrow germanium p-n junctions," *Phys. Rev.*, vol. 109, pp. 603–604; January, 1958.
[6] N. E. Mott, "The theory of crystal rectifiers," *Proc. Roy. Soc.*, vol. A171, p. 27; May, 1939.
[7] W. Schottky and E. Spenke, "Zur Quantitativen Durch führung der Rauinladungs-und Randschichttheorie der Kristallgleichrichter," *Wiss. Veroff. Siemenswerke*, vol. 18, p. 225; October, 1939.
[8] See for example, H. C. Torrey and C. A. Whitmer, "Crystal Rectifiers," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 78–79; 1948.
[9] F. S. Goucher, G. L. Pearson, M. Sparks, G. K. Teal, and W. Shockley, "Theory and experiment for a germanium p-n junction," *Phys. Rev.*, vol. 81, pp. 637–638; February, 1951.

essential to the rectifying action. The junctions that are of interest in semiconductor technology generally fall well within the required range of steepness to justify the assumption that the transition between neutral material and material which has had all of the majority carriers swept out is infinitely sharp.

The part of the conduction process which involves surmounting the potential barrier at the junction is most easily understood as a deviation from the thermal equilibrium condition. At thermal equilibrium, the product of electron and hole densities throughout the sample is constant. Thus, we may write[1]

$$pn = n_i^2 \tag{1}$$

where

p = hole density
n = electron density
n_i = intrinsic carrier density.

The intrinsic carrier density is a characteristic of the semiconductor and is a strong function of temperature. The value of n_i^2 for germanium at room temperature is approximately 10^{26}, and doubles with each increase in temperature of 8°C. On the n-type side of the junction, the number of electrons is very nearly equal to the number of excess donors, so that electrons outnumber holes by many orders of magnitude for reasonably large excess donor density. Reciprocal considerations hold for the p-type side. The solid curves in Fig. 2 show the variation of electron and hole density through the junction at equilibrium. The sum of the logarithms of the equilibrium densities is a constant, independent of position.

When a bias is applied to the junction, the equilibrium is disturbed. The p-n product changes in the barrier region by a Boltzmann factor,[1] and becomes

$$pn = n_i^2 \exp(qV/kT) \tag{2}$$

where

V is the applied voltage (p positive with respect to n)
q = electronic charge
k = Boltzmann constant
T = absolute temperature.

The dotted curves in Fig. 2 show the disturbed carrier densities for a forward applied bias. The p-n product is constant through the barrier region and approaches the thermal equilibrium value in the neutral material as the excess carriers diffuse away from the junction and recombine.

The effect of (2) is that when the injected minority carrier density is small compared to the equilibrium majority carrier density, the majority carrier density on either side of the barrier region remains unchanged. However, the minority carrier density at the boundary of the neutral regions and the barrier is increased by the exp (qV/kT) factor. We see then that the effect of a given voltage on a junction is to disturb the carrier

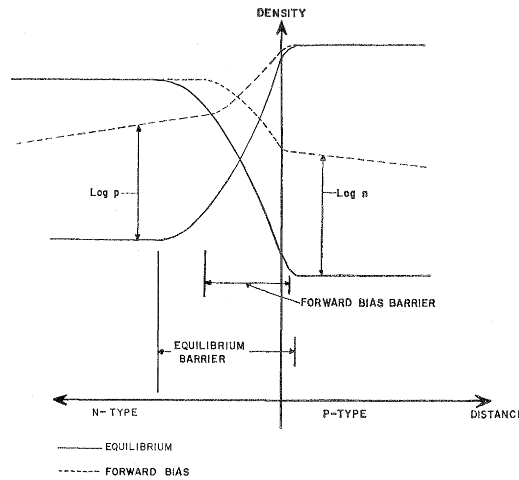


Fig. 2—Carrier density at and near a p-n junction. The solid curves are log n and log p, respectively, for the thermal equilibrium conditions. The sum of the logarithms is constant throughout the specimen. The dashed curves show the densities with a moderate forward bias. The sum of the logarithms is essentially constant in the barrier region implying a constant n-p product. The n-p product decreases in the neutral material from its value in the barrier to the thermal value far from the junction.

densities. The actual current through the junction is limited by the rate of disappearance (or appearance) of the excess (or deficit) minority electrons and holes.

The rate of disappearance of the minority carriers is governed by their rate of diffusion and recombination in the neutral regions. The problem of determining the junction current now becomes the problem of calculating hole and electron densities (and flow) in each of the two neutral regions subject to the boundary conditions of (2) at the boundary of the barrier, and (1) at $x \to \pm\infty$.

The transport equations for minority carriers are particularly simple for small densities (linear equations) and uniform semiconducting regions (no build-in fields). The result of the calculation of carrier densities for this simple case, interpreted in terms of current, is[1]

$$I = I_s[\exp(qV/kT) - 1)], \tag{3}$$

where

I = junction current amp/cm²
V = junction voltage—volts
I_s = saturation current $= q[(n_{p0}/\tau_n)L_n + (p_{n0}/\tau_p)L_p]$
τ_n = electron lifetime in p-type semiconductor
τ_p = hole lifetime in n-type semiconductor
n_{p0} = thermal electron density in p-type semiconductor
p_{n0} = thermal hole density in n-type semiconductor
$L_n = \sqrt{D_n \tau_n}$ = electron diffusion length
$L_p = \sqrt{D_p \tau_p}$ = hole diffusion length
D_p = electron diffusion constant
D_p = hole diffusion constant.

We will refer to this result (3) as the ideal rectifier equation since it was derived for the p-n junction with quite drastic simplifying assumptions on the nature of the p-n boundary and the conduction process. In addition, the resulting rectification ratio is the best that is possible in a simple electronic process. For a forward bias ($V>0$) greater than a few kT/q, the current increases exponentially with voltage. For reverse bias ($V<0$), I saturates at $-I_s$ which corresponds to the rate of generation of minority carriers in the slice which is L_p thick on the n side and L_n thick on the p side of the junction. The solid curve in Fig. 3 is a plot of the ideal rectifier (3). In the case of a reverse bias greater than several kT/q, the boundary condition (2) does not give the correct minority carrier densities.[10] For this case, practically no carriers surmount the barrier, and the density is determined by the rate of flow of holes from the n side and of electrons from the p side across the junction. However, the minority carrier density has become so small that (2) continues to give the correct answer in most practical problems. Eq. (3), which predicts a saturation of current with reverse voltage, continues to hold.

One of the high points in the evolution of junction theory was the development of (3) which predicts the current-voltage characteristic in terms of basic semiconductor properties and its quantitative verification in germanium p-n junctions.[1,9] The simple theory of Shockley[1] has served as a basis for many extensions and modifications. In particular, it was found that the range of current and voltage for which junctions satisfied (3) was limited in the case of germanium to low-level operation and in the case of silicon was nonexistent at room temperature. The departures from (3) are always in the direction of poorer rectification, where rectification is taken as the ratio of forward to reverse current at some low voltage. As based on the semiconducting properties of the two neutral regions and the simple theory (3), more current is conducted in the reverse or blocking direction than I_s, and in the forward or easy conducting direction more voltage than the amount specified in (3) is required. This departure is illustrated in Fig. 3 where voltage is plotted on a linear scale, and current on a logarithmic scale. The slope of the ideal forward characteristic is kT/q volts per e (naperian log base) of current or 2.3 kT/q per decade of current. We will refer to the ideal case as a kT/q slope or exp qV/kT behavior in discussing departures from the simple case. The degraded rectification characteristic in Fig. 3 has a slope steeper than kT/q.

The reasons for the departure from the ideal case are seen most easily as modifications of the simple theory. Our approach will be to consider departures first in the reverse characteristic, then at low forward bias, and finally at high forward bias since this is generally in the direction of increasing complexity. It is convenient to divide the diode V-I characteristic into four distinct regions for the purposes of analysis. These are: 1) high reverse bias, 2) low and medium reverse bias, 3) low forward bias, 4) high forward bias. The boundaries of

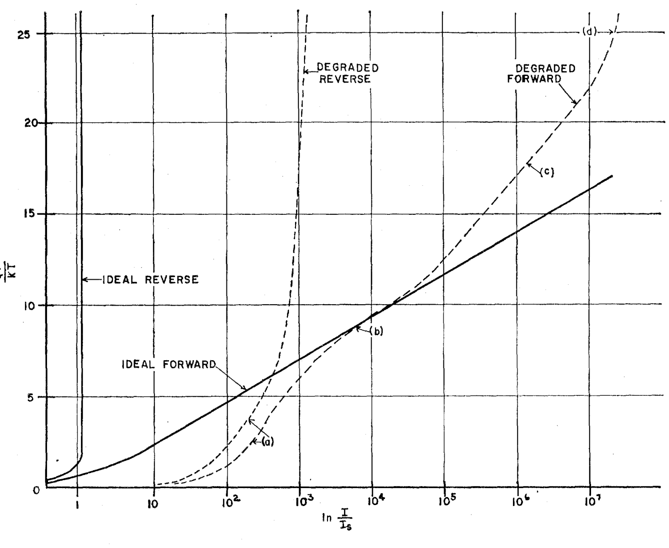


Fig. 3—Rectifier characteristics. Modifications of the ideal theory lead to degradation of the rectifier characteristic. In (a) the conduction mechanism is dominated by the space-charge region carrier recombination and generation. Note the lack of saturation in the reverse characteristic and extra conduction at low forward voltages. As the forward voltage is increased the diffusion current becomes dominant and there may be a tendency towards ideal behavior (b). At high-level injection (c) the diffusion current departs from the ideal law, with an increase in the semilogarithmic slope. Finally the current is limited by series ohmic resistance (d).

these regions are not very sharp but general distinctions can be made.

The problem of high reverse bias is the problem of the mechanism of reverse breakdown and is a subject unto itself. It is not within the scope of this article to consider the breakdown mechanism in detail; however, a short discussion is warranted. In 1953 McKay and McAfee showed that the breakdown in moderately wide junctions occurs through a mechanism of carrier multiplication or avalanche in a manner almost completely described by the Townsend gaseous discharge mechanism.[11–13] In the avalanche process, electrons and holes acquire enough energy from the applied field to cause secondary production of electron-hole pairs. The secondary pairs, in turn, can cause further ionizations. In very thin junctions, the breakdown is believed to be by the Zener mechanism whereby electrons are pulled out of the valence bond by the excessively high fields that occur.[14] In silicon junctions the breakdown which occurs at voltages greater than about 6 v is avalanche breakdown. However, through no fault of Zener, the breakdown has been called "Zener breakdown." A name more appropriate than the erroneous "Zener Diode," as applied to the usual voltage-regulator diode, might be

[10] C. T. Sah, R. N. Noyce, and W. Shockley, "Carrier generation and recombination in p-n junction characteristics," Proc. IRE, vol. 45, pp. 1228–1243; September, 1957.

[11] K. G. MaKay and K. B. McAfee, "Electron Multiplication in silicon and germanium," *Phys. Rev.*, vol. 91, p. 1079; September, 1953.

[12] K. G. McKay, "Avalanche breakdown in silicon, "*Phys. Rev.*, vol. 94, pp. 877–884; May, 1954.

[13] S. Miller, "Avalanche breakdown in germanium," *Phys. Rev.*, vol. 99, pp. 1234–1241; August, 1955.

[14] A. G. Chynoweth and K. G. McKay, "Internal field emission in silicon p-n junctions," *Phys. Rev.*, vol. 106, pp. 418–426; May, 1957.

simply, "Avalanche Diode" or perhaps "*V-R* Diode" (Voltage Regulator Diode), in analogy to the *V-R* tube which performs the same function in circuits and actually uses the avalanche mechanism. We will take medium reverse bias as being reverse bias sufficiently small that avalanche effects are negligible. This requirement is satisfied for silicon and germanium diodes at voltages roughly less than half the breakdown voltage. (Since the multiplication increases the alpha of a transistor above unity, this restriction to half the breakdown voltage is generally not nearly strict enough for the case of the transistor collector.)

For medium reverse bias, measurements on actual diodes generally result in more current than I_s as given in (3). Part of this current may be a result of surface leakage, inversion layers,[15,16] or simply gross body defects in the junction. The surface leakage as well as inversion layers can be eliminated by proper treatment of the surface while body defects can be eliminated by making the junction so small that there is a statistical chance that all gross defects have been excluded. After all of these precautions have been taken, excess current is still observed in silicon junctions at room temperature and in germanium junctions at lower temperatures. This extra (over the simple theory) current has been attributed to charge generation in the barrier region.[10,17–19] The rate of carrier generation by recombination-generation centers[20] in the barrier region, where almost all of the mobile carriers have been swept away, is approximately

$$G_{\text{barrier}} = n_i/\tau \tag{4}$$

whereas the generation rate in a region where only the minority carriers have been removed is

$$G_{\text{neutral}} = n_{\text{min}}/\tau \tag{5}$$

where n_{min} is the equilibrium density of minority carriers, and τ = lifetime.

The equilibrium minority carrier density is much less than the intrinsic density so that the generation rate in the barrier region is much greater than in the neutral material. The saturation current in (3) varies as the second power of intrinsic carrier density, but the generation current that arises from the generation rate in (4) varies as the first power of intrinsic carrier density. As a result, this generation current is dominant when the intrinsic carrier density is small as is the case for silicon at room temperature or germanium at lower temperatures. The reverse space-charge generation current is proportional to the volume in which generation occurs and is thus proportional to the barrier width. The barrier width generally increases with reverse voltage so that there is no true reverse saturation of current for the case where space-charge generation dominates reverse current. For germanium at room temperature, the intrinsic carrier density is large enough for the saturation current of (3) to be dominant, so that germanium *p-n* junctions obey the simple theory for moderate reverse bias.

If the barrier region generation term is dominant in the case of small reverse bias, this effect will also dominate the conduction mechanism of (3) for a small forward bias.[10] As a result, more current will be conducted at a small forward voltage than is predicted by (3). This current[10] will have a semilogarithmic slope greater than kT/q at biases of several kT/q and explains the departure of silicon *p-n* junctions from the simple theory even at small biases (Fig. 3).

As the forward bias is increased, the diffusion current into the neutral region increases faster than the recombination current in the barrier layer until, finally, the diffusion and recombination of carriers in the neutral region is the essential conduction mechanism. At this point we might expect the diffusion theory that results in (3) to give the correct diode behavior, *i.e.*, a kT/q slope and, in fact, a small region where the *V-I* tends towards exp (qV/kT) behavior has been observed in silicon rectifier diodes.[19] However, at high forward bias so many minority carriers are injected across the junction that the minority carrier density becomes of the same magnitude as the majority carrier density, and (3) must be modified. For this case of high injection, the assumption of charge neutrality outside the barrier region is still satisfied. However, the electric field in the neutral region is no longer negligible. Majority carriers move to the junction to neutralize the minority carriers —thus establishing a gradient of majority carriers. Fig. 4 shows the near equality of majority and minority carriers under conditions of high injection. In this case the *p*-type side is more heavily doped (more extrinsic) than the *n*-type side so that modulation occurs essentially on the *n*-type side only. The gradient of electrons results in a diffusion current of electrons; however, a small unbalance of charge results in a field that is just strong enough to result in an equal and opposite drift current. Exact analysis is difficult since the transport equations for carriers become nonlinear.[21,22] The result-

[15] M. Cutler and H. M. Bath, "Surface leakage current in silicon fused junction diodes," Proc. IRE, vol. 45, pp. 39–43; January, 1957.

[16] W. T. Eriksen, H. Statz, and G. A. DeMars, "Excess surface currents on germanium and silicon diodes," *J. Appl. Phys.*, vol. 28, pp. 133–139; January, 1957.

[17] H. Kleinknecht and K. Seiler, "Einkristalle und ph Schichtkristalle aus Silizium," *Zeits. f. Phyisk*, vol. 139, pp. 599–618; December, 1957.

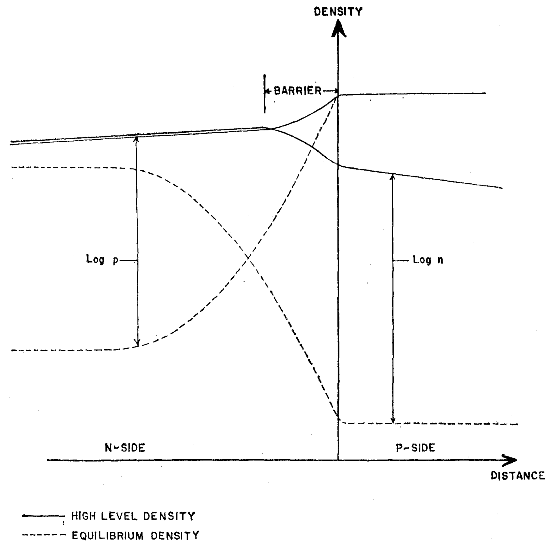
[18] E. M. Pell and G. M. Roe, "Reverse current and carrier lifetime as a function of temperature in germanium junction diodes," *J. Appl. Phys.*, vol. 26, pp. 658–665; June, 1955.

[19] H. S. Veloric and M. B. Prince, "High voltage conductivity-modulated silicon rectifier, "*Bell Sys. Tech. J.*, vol. 36, pp. 975–1004; July, 1957.

[20] W. Shockley and W. T. Read, Jr., "Statistics of the recombinations of holes and electrons," *Phys. Rev.*, vol. 87, pp. 835–842; September, 1953.

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

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



ig. 4—The solid curves show carrier density near a *p-n* junction under conditions of high-level injection. The *p-n* product is constant in the barrier region, and $p \approx n$ in the neutral *n*-type side near the junction. For reference, equilibrium densities are shown in the dashed lines.

ing minority carrier drift current is in the same direction as the minority carrier diffusion current, thereby increasing minority carrier current.

We must re-examine the boundary condition in the barrier as expressed in (2) for application to the high-level case. The solid curves in Fig. 4 show the carrier densities under high-level injection conditions on one side (*n* side). The *p-n* product is constant, as before, through the barrier region. However, the majority carrier density (electrons) at the edge of the barrier on the *n* side has increased to several times the equilibrium number (conductivity modulated). Eq. (2) can still be used to obtain the *p-n* product in the barrier if the voltage V is interpreted to include the voltage associated with the field that holds the majority carriers in place as well as the voltage that appears across the barrier region, but excluding ohmic drops.[10] The near equality of majority and minority carriers under high-level injection conditions results in an exp $qV/2kT$ rate of increase of minority carriers. Thus the high-level rate of increase of minority carriers is slower than the low-level rate, and is in the direction of less current than (3). An additional complication analyzing the junction characteristic is the fact that carrier lifetime is a function of carrier density with lifetime increasing with minority carrier density in silicon.

The general approach to the problem of calculating the voltage-current characteristic for the *p-n* junction with high-level forward bias has been to simplify the problem in some respect until the differential equations can be handled. A junction of particular interest is the *PIN* structure. In order to reduce series ohmic resistance to a minimum (and still maintain high reverse breakdown), high conductivity contacts are placed near enough to a junction where one side is only weakly extrinsic so that conductivity modulation extends from one contact to the other at high injection. The density of impurities in the weakly extrinsic part of the junction, and even the question of whether the conductivity is *p* or *n*-type becomes immaterial in the high injection condition. Thus the three-layer junction which consists of 1) heavily doped *p*-type, 2) a central layer of lightly doped *p* or *n*-type, and 3) heavily doped *n*-type semiconductor is called the *PIN* diode. At low biases the *PIN* diode (as usually constructed) behaves like a junction between heavily doped and lightly doped semiconductors of opposite conductivity types. In 1952 Hall[23] set up a model for the *PIN* structure in which he assumed that all of the recombination occurs in the central region, with the result that for almost all forward biases, the junction follows exp $qV/2kT$. In this case, equal voltages must be supplied at the $p+I$ and $IN+$ junctions to get carriers into the central intrinsic region.

In 1956 Kleinman[24] considered the same problem with the assumption that recombination in the central $I-$ region was negligible, with the result that the voltage drop across the central region is negligible, and the high-level, forward current-voltage characteristic is determined by the $P+$ and $N+$ end regions. Kleinman concluded that a nonlinear recombination law was necessary to explain the departure of the diode from the ideal theory.

The actual *PIN* diode has a finite amount of recombination in each of the three regions, and is undoubtedly a compromise between the two idealized models. However, the analytic problem of the *PIN* diode with finite recombination rates that are functions of density in all three regions is extremely difficult. At the present time we can take the results of Hall[23] and Kleinman[24] to give the qualitative result that the semilogarithmic slope of the forward characteristic should be greater than kT/q.

The *p-n* step junction has also received considerable attention in the high forward bias range. In this case the high injection range begins when the minority density on one side becomes comparable with the majority carrier density.[10,25] High injection begins on the high resistivity side since the minority carrier density is always greater on this side as compared to the heavily doped side. The calculation of the diffusion current again involves solution of nonlinear differential equations, with the result of various approximations again indicating exp $qV/2kT$ behavior.

A *p-n* junction that is simple enough for the separate high-level effects to be separately considered is the emitter of a transistor. In this case the carriers that are injected into the base layer disappear in an average time,

[23] R. N. Hall, "Power rectifiers and transistors," PROC. IRE, vol. 40, pp. 1512–1519; November, 1952.
[24] D. A. Kleinman, "The forward characteristic of the *PIN* diode," *Bell Sys. Tech. J.*, vol. 35, pp. 685–706; May, 1956.
[25] J. S. Saby, "Junction Rectifier Theory," Proc. Rugby Conference, Rugby, Eng.; 1956.

which is the diffusion time across the base layer, and the injected current is given by[26]

$$I = \frac{qpnD_{\min}}{\int_0^w ndx} \tag{6}$$

where

p and n are the carrier densities just inside the base layer at the emitter junction,

$D_{\min}$ is the average diffusion constant for the minority carrier in the base, and

$\int_0^w ndx$ is the integral of majority carrier density the base layer and is the total number of majority carriers per cm^2 in the base layer.

From (2) and (6) we see that for the one-dimensional transistor case the emitted current is not reduced until the average number of majority carriers in the base layer is increased. It is possible, for example, in diffused emitter transistors for the emitter to be in a high injection state but for the total number of carriers to remain essentially constant. In this case, the high-level injection condition holds, but the emitter diode follows the ideal diode formula. When high injection ($p \approx n$) extends across most of the base layer, the number of majority carriers in the base is proportional to the density at the emitter and the current-voltage characteristic follows an exp $(qV/2kT)$ law where V must be interpreted as the voltage across the emitter and includes the contribution to voltage of the field in the neutral region which arises as a result of majority carrier modulation but does not include resistive drops.

A modification of the emitter junction, which is of some interest, is the emitter junction of the diffused base, diffused emitter transistor where the contact is made to the base by alloying through the emitter layer.[27] In this case a relatively small contact is obtained and the problem must be considered as a two-dimensional one. Lateral flow of base current in the relatively high resistivity base causes the emission to be concentrated around the base contact so that the effective area is reduced as the current increases. The result for some cases of idealized geometry is exp $(qV/2kT)$ behavior.[28] Emission concentration can occur simultaneously with high-level injection with the result that behavior ranging from exp $(qV/1.5kT)$ to exp $(qV/3kT)$ has been observed. In silicon alloy transistors made on high resistivity base material, the lateral concentration effect crowds emission to the edge of the emitter region resulting in a semilogarithmic slope greater than kT/q.

A fairly wide range of exp (qV/kT) behavior has been observed in the emitters of diffused base silicon transistors with alloyed emitters. The minority carriers were removed by the collector so that the effective lifetime was the essentially constant diffusion time across the base layer. In addition, the conductivity of the base was high enough that there was no appreciable emission crowding. This type of junction did exhibit exp $(qV/2kT)$ behavior at low currents due to space-charge generation of carriers.

[26] J. L. Moll and I. M. Ross, "The dependence of transistor parameters in the distribution of bass layer resistivity," PROC. IRE, vol. 44, pp. 72–78; January, 1956.

[27] M. Tanenbaum and D. E. Thomas, "Diffused emitter and base silicon transistors," *Bell Sys. Tech. J.*, vol. 35, pp. 1–22; January, 1956.

[28] F. M. Smits (private communication).

Conclusions

The low level p-n junction theory of Shockley, which is based on the statistical escape of particles across a barrier and diffusion into the region beyond the barrier predicts the ideal diode formula, (3), for rectification. This theory adequately explains the low-level operation of p-n junctions that have fairly large saturation currents such as germanium junctions at room temperature.

The low-level operation of junctions made from materials with low saturation currents (such as silicon at room temperature) requires the inclusion of the effects of generation and recombination of electron-hole pairs in the barrier region. In silicon at room temperature and germanium at low temperatures, the generation-recombination dominates the low-level operation with a resulting departure from ideal rectification.

The high-level forward biased junction has a greater voltage drop than predicted by the ideal rectifier formula. This extra voltage arises because of the phenomena associated with conductivity modulation in part of the diode. Exact theory for the conductivity modulated case is difficult to obtain, but solutions for various limiting cases result in a qualitative explanation of the diode characteristics.

As a result of the departure from the ideal law at low biases due to space-charge recombination and generation, ideal rectifier behavior is never observed in silicon junctions at room temperature. The departures from the simple theory at high biases limits the possible range of kT/q behavior to moderate forward bias, and in most cases there is no range of kT/q behavior at all. However, germanium does show ideal behavior at low biases.

The exact formulation of the problem of calculating the V-I of the p-n junction is very difficult when either the low-level effect of barrier region recombination and generation or the high-level effect of conductivity modulation is included. In addition, it is impractical to obtain better than order-of-magnitude guesses as to the magnitudes of some of the semiconductor parameters. As a result, close agreement between theory and experiment is difficult. However, effects that have been discussed are in good qualitative agreement with experiment in both silicon and germanium p-n junctions.

Acknowledgment

The author wishes to acknowledge the helpful discussion and criticism of many of his colleagues.