THE UNIVERSITY OF MICHIGAN

INDUSTRY PROGRAM OF THE COLLEGE OF ENGINEERING

PHYSICAL ELECTRONICS OF JUNCTION TRANSISTOR CHARACTERISTICS

William G. Dow

September, 1960

IP-462

ensn

UMR8894

ACKNOWLEDGMENTS

The author wishes to acknowledge his indebtedness to Dr. Earle Thomas for his clear presentation, at a series of lectures given at The University of Michigan in the summer of 1955, of many of the concepts underlying this paper, and also to many associates at The University of Michigan, especially Gunnar Hok, J. E. Rowe, M. H. Miller, and D. C. Ray.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

The major aspects of voltage-current relations of the usual junction transistors, that employ semiconductors having relatively low number densities of donor and acceptor impurities, can be described adequately in terms of a classical kinetic theory model of charge carrier flow. Use of this model is justified because the charge carrier densities lie in the range of non-degeneracy of the ideal gas model, for which classical Maxwell-Boltzmann energy distribution statistics apply. This is in sharp contrast to the necessity for use of a quantum-mechanical model and the Fermi statistics of a degenerate gas to describe charge carrier behavior in metallic conductors and in the Esaki or tunnel diode, because of their much higher ranges of carrier densities. Simple classical statistics applied to charge carriers in junction transistors provide this paper's basis for derivation of the Ebers and Moll transistor equations.(1)

The analytical study will have the following steps: (a) Determination of the contact potentials across the emitter-base and collector-base interfaces, (b) expression of the emitter and collector currents as resulting from net random-current flow across the interfaces (c) determination of the spatial density distribution of active carriers within the base, in terms of diffusion and carrier lifetime properties, (d) expression of the emitter and collector currents in terms of concentration gradients in the base at the two interfaces, (e) elimination of interface boundary properties between (b) and (d) to give the Ebers and Moll equations. The kinetic-theory mks notation used will be that in Chapter XII of the author's book "Fundamentals of Engineering Electronics," 2nd edition.(2) The model used in the derivation will be an n-p-n transistor.

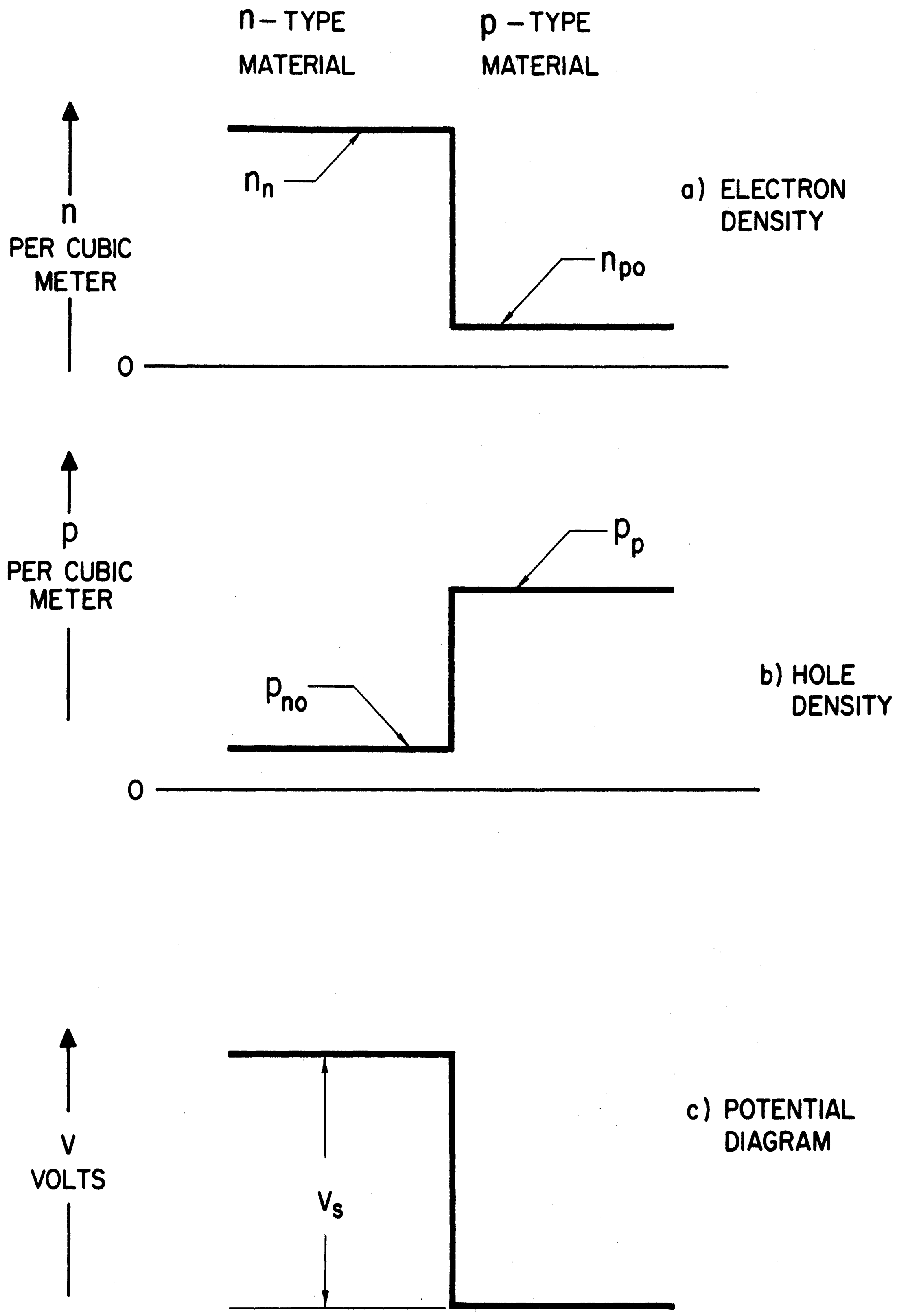

Figure 1. Contact potential difference V_s and step-function density distribution of electrons and of holes at the interface of an n-p semiconductor diode in the disconnected state.

II. CONTACT POTENTIAL DIFFERENCE ACROSS A DISCONNECTED n-p INTERFACE

Figures 1a and 1b (not to scale) represent respectively the electron and hole number density distribution, and Figure 1c the voltage distribution, in an electrically isolated n-p junction, that is, one that is disconnected from any other circuit elements, and is as a whole electrically uncharged. Electron and hole number densities, in particles per cubic meter, will be symbolized by n and p, with subscripts as follows:

In the n-type material,

n_n for majority carriers (electrons)

p_n for minority carriers (holes)

p_{no} for the disconnected-state hole density:

In the p-type material

p_p for majority carriers (holes)

n_p for minority carriers (electrons)

n_{po} for the disconnected-state hole density

Because the random thermal motions of the n_n electrons and the p_p holes tend to move them across the interface into regions having the sparse n_{po} and p_{no} populations, equilibrium flow exists only when there exists a contact difference of potential V_s opposing these majority-carrier movements out of their home environments. The average or "drift" velocity of the charge carriers due to the usual conduction-current electric fields is negligibly small relative to the characteristic random velocity, so that the rate of arrival of the carriers at an interface is governed by the random motions. For the Maxwellian distribution of random velocities, in general terms[2]

$$\left.\begin{array}{l}\text{Rate of arrival of electrons at an in-}\\ \text{terface, per square meter per second}\end{array}\right] = \frac{n\zeta_n}{2\sqrt{\pi}} \qquad (1a)$$

$$\left.\begin{array}{l}\text{Rate of arrival of holes at any}\\ \text{interface}\end{array}\right] = \frac{p\zeta_p}{2\sqrt{\pi}} \qquad (1b)$$

where n, p, are the respective carrier densities adjacent to the approach side of the interface, and

ζ_n, ζ_p, are the characteristic random velocities for the electrons and holes respectively, being properties of the temperature and of the semiconductor substance used, but not of the majority carrier density; the defining equation is

$$kT = q_e V_T = 1/2\, m_n \zeta_n^2 = 1/2\, m_p \zeta_p^2 \qquad (2)$$

Symbolism appearing here and later is:

T = temperature of the device in degrees Kelvin,

k = Boltzmann's constant, 1.38×10^{-23} joule per degree Kelvin,

q_e = the absolute value of the charge carried by an electron, 1.60×10^{-19} coulomb,

m_e = the mass of an electron, 9×10^{-31} kilogram,

V_T = the kinetic temperature, expressed in volts (or electron volts), and defined by (2),

V_s = the contact potential difference at the interface, being qualitatively like the contact potential between metals, in that it is measurable only by electrostatic means,

m_n, m_p, are the effective masses respectively of electrons and holes within the semiconductor, determinable by experimental means; the order of magnitude is the same as for m_e; m_p exceeds m_n.

The quantity $kT = q_e V_T$ is the kinetic energy per particle characteristic[2] of the temperature T, being the energy per particle that occurs in the fundamental equations of statistical mechanics; ζ_n, ζ_p, are the velocities corresponding to the characteristic energy. Numerically, (2) becomes $V_T = T/11600$; $V_T = 0.0258$ volt when $T = 300°K = 27°C$. Also, $\zeta_n = 5.93 \times 10^5 \sqrt{V_T} \,/\, \sqrt{m_n/m_e}$, and similarly for ζ_p.

If the electrons or holes that approach an interface at the (1) rate experience there a potential barrier of V volts, the rate of penetration ("climbing") against this barrier is less[2] than the (1a,b) flow rate by the factor $\exp(-V/V_T)$. If the interface presents no barrier, the carriers being able to pass freely ("fall"), the flow is as in (1a,b).

For the Figure 1 diode in the disconnected state, the contact potential V_s is a barrier to electron flow from the n-type to p-type material, and to hole flow in the reverse direction. This permits the following book-keeping:

Electron flow from left to right; electrons must climb down the potential hill.	=	$\frac{n_n \zeta_n}{2\sqrt{\pi}} \exp \frac{-V_s}{V_T}$;	(3a)
Electron flow from right to left; electrons fall up the potential hill.	=	$\frac{n_{po} \zeta_n}{2\sqrt{\pi}}$	(3b)
Hole flow right to left; holes must climb up the hill.	=	$\frac{p_p \zeta_p}{2\sqrt{\pi}} \exp \frac{-V_s}{V_T}$;	(4a)
Hole flow left to right, holes fall down the hill.	=	$\frac{p_{no} \zeta_p}{2\sqrt{\pi}}$.	(4b)

In the disconnected state there must of course be zero net charge transfer across the interface. Obviously a sufficient condition for this is zero net transfer of electrons and zero net transfer of holes. To show that this is also a necessary condition, imagine zero net charge transfer occurring by means of a net flow of electrons and an equal net flow of holes from left to right across the Figure 1 interface. This requires a net volume recombination in the p-type material, introducing heat there, and an equal net volume generation of electron and hole pairs, with an associated equal heat removal, in the n-type material. But the second law of thermodynamics prohibits the temperature differential that would build up to provide the return-flow heat transfer. Mechanistically, use of bulk carrier flow principles described below show that the postulated flow would call for a less potential barrier to the climbing electrons than to the climbing holes -- an obvious impossibility. Thus both mechanistically, and from heuristic energy considerations, zero net interface transfer of each kind of carrier is a necessary condition for zero net charge transfer.

For such zero net transfer of each kind of carrier, (3a,b) and (4a,b) reduce to

$$\frac{n_{po}}{n_n} = \exp\frac{-V_s}{V_T} \quad ; \qquad \frac{p_{no}}{p_p} = \exp\frac{-V_s}{V_T} \quad . \qquad (5)$$

Comparison shows that $n_n p_{no} = p_p n_{po}$. Thus, from classical kinetic theory considerations, the product of minority and majority carrier densities is a constant of the material that is invariant with donor or acceptor impurity density. This constant's magnitude, governed by quantum mechanical considerations, is expressible as follows:

$$n_n p_{no} = p_p n_{po} = n_i^2 = p_i^2 = A_i T^3 \exp \frac{-V_i}{V_T} \qquad (6)$$

The new symbols have meanings as follows:

n_i, p_i, are the respective electron and hole numbers densities in the intrinsic (i.e., undoped) semiconductor; $n_i = p_i$;

A_i, V_i, are constants of the material, thus:[3]

For germanium, $A_i = 3.1 \times 10^{44}$, $V_i = 0.785$ electron volt; (7a)

For silicon, $A_i = 1.5 \times 10^{45}$, $V_i = 1.21$ electron volts. (7b)

The dependence of contact potential V_s on temperature and on properties of materials is in accordance with (5).

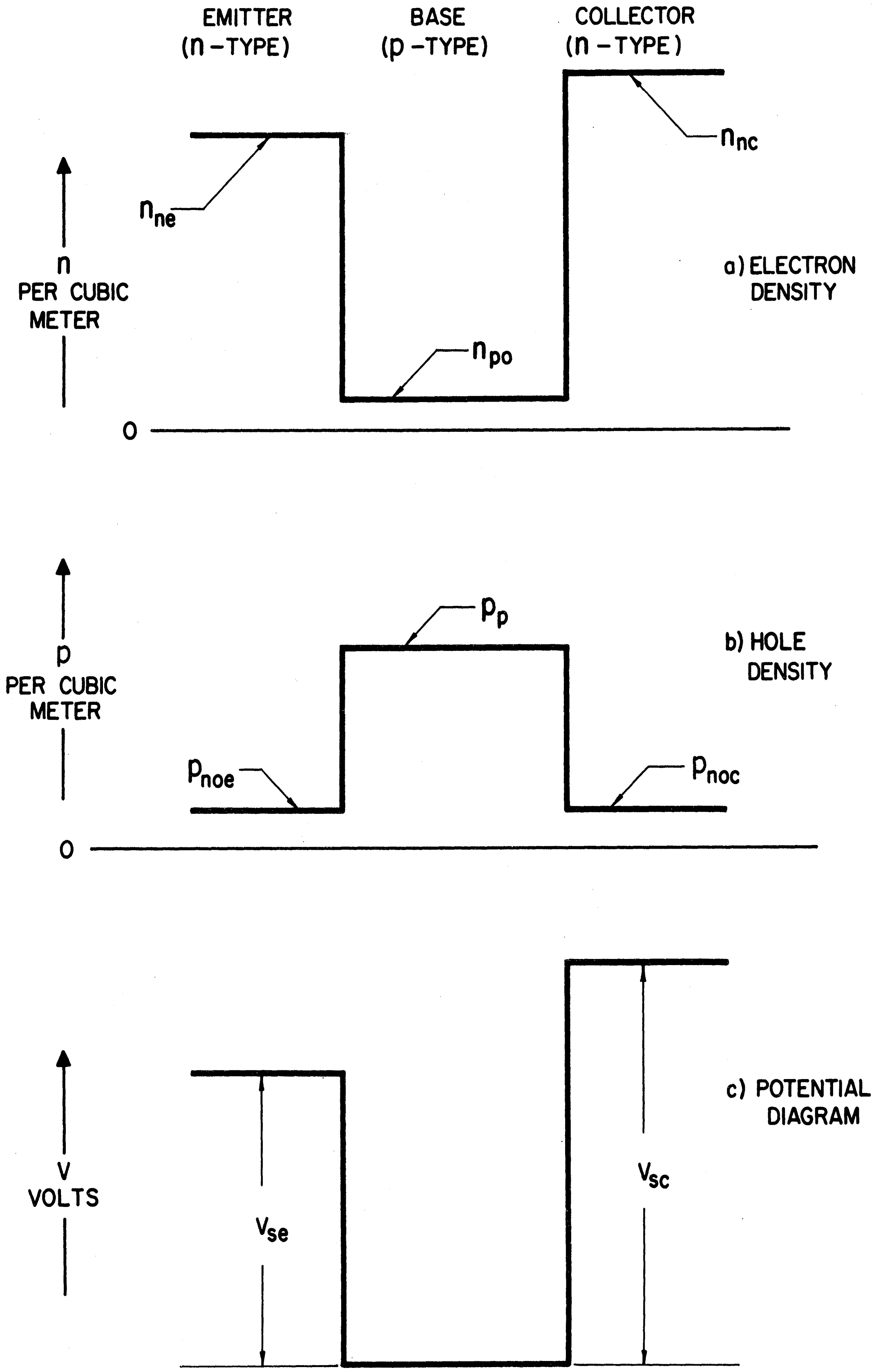

Figure 2. The two contact potential difference V_{se} and V_{sc}, and the number density distributions for holes and electrons for an n-p-n junction transistor in the disconnected state.

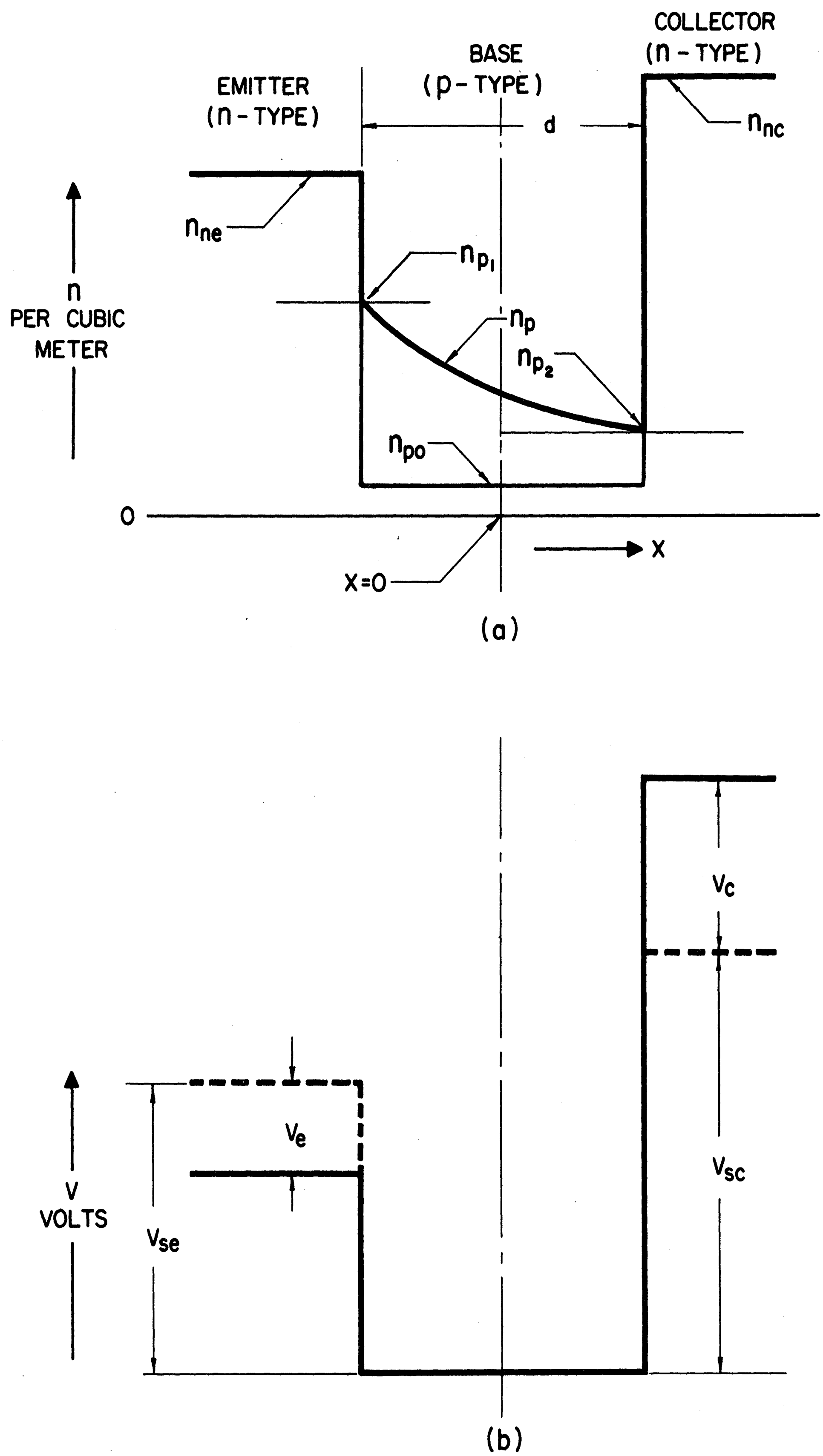

Figure 3. The physical concept model of an operating n-p-n junction transistor, showing the electron number density distribution and the potential distribution, when connected into a typical transistor amplifier circuit.

III. CHARGE TRANSPORT ACROSS TRANSISTOR INTERFACES

A junction transistor has two inteefaces, with in general two different contact potentials, obeying two sets of equations like (5); see Figure 2 for the disconnected state. For studying transport when connected into a circuit, the model illustrated by Figure 3 will be used. In Figures 2 and 3, or for later use,

V_e = emitter voltage, numerically negative for the Figure 3 model;

V_c = collector voltage, numerically positive for Figure 3;

I_e = emitter current, numerically negative for Figure 3;

I_{en}, I_{ep}, are the portions of I_e carried across the emitter-base interface by electrons and holes respectively;

I_c = collector current, numerically positive for Figure 3;

I_{cn}, I_{cp}, are the portions of I_c carried across the base-collector interface by electrons and holes respectively; obviously

$$I_e = I_{en} + I_{ep} \quad ; \quad I_c = I_{cn} + I_{cp} \quad ; \qquad (8a)$$

I_b = base current, numerically positive in Figure 3; obviously

$$I_e + I_c + I_b = 0 \qquad (8b)$$

because each of these currents is considered numerically positive for conventional positive-charge current flow into the respective emitter, collector, and base terminals. Also

n_{p1},n_{p2}, are minority-carrier number densities in the base at locations respectively adjacent to the emitter and collector interfaces;

p_{n1},p_{n2}, are minority-carrier densities respectively in the emitter and the collector, adjacent to the interfaces with the base (not shown in Figure 3);

n_{ne}, n_{nc}, are majority carrier densities in the emitter and in the base,

p_{noe}, p_{noc}, are the disconnected-state minority carrier densities in the emitter and in the base,

V_{se}, V_{sc}, are contact potential differences respectively for the two interfaces, each as in expressions like (5),

S = cross-sectional area of the device, the same at the two interfaces,

d = the distance between the interfaces (base thickness).

At the emitter-base interface, the new forms of (3), (4) for Figure 3 are:

$$\text{Electron flow left to right} = \frac{n_{ne}\zeta_n}{2\sqrt{\pi}} \exp \frac{-(V_s+V_e)}{V_T}, \qquad (9a)$$

$$\text{Electron flow right to left} = \frac{n_{p1}\zeta_n}{2\sqrt{\pi}}, \qquad (9b)$$

$$\text{Hole flow right to left} = \frac{p_p\zeta_p}{2\sqrt{\pi}} \exp \frac{-(V_s+V_e)}{V_T}, \qquad (9c)$$

$$\text{Hole flow left to right} = \frac{p_{n1}\zeta_p}{2\sqrt{\pi}} \qquad (9d)$$

At this interface, the electron-borne portion of the emitter current is the net charge passing through the area S due to (9a,b), being

$$I_{en} = -q_e S \frac{n_{ne}\zeta_n}{2\sqrt{\pi}} \exp \frac{-(V_{se}+V_e)}{V_T} + q_e S \frac{n_{p1}\zeta_n}{2\sqrt{\pi}}, \qquad (10a)$$

whereas the hole-borne portion is

$$I_{ep} = -q_e S \frac{p_p\zeta_p}{2\sqrt{\pi}} \exp \frac{-(V_{se}+V_e)}{V_T} + q_e S \frac{p_{n1}\zeta_p}{2\sqrt{\pi}}. \qquad (10b)$$

Correspondingly, at the base-collector interface the electronborne portion of the collector current is

$$I_{cn} = - q_e S \frac{n_{nc}\zeta_n}{2\sqrt{\pi}} \exp \frac{-(V_{sc}+V_c)}{V_T} + q_e S \frac{n_{p2}\zeta_n}{2\sqrt{\pi}} \quad , \qquad (11a)$$

whereas the hole-borne portion is

$$I_{cp} = - q_e S \frac{p_p\zeta_p}{2\sqrt{\pi}} \exp \frac{-(V_{sc}+V_c)}{V_T} + q_e S \frac{p_{n2}\zeta_p}{2\sqrt{\pi}} \qquad (11b)$$

In (10a) the electron-borne portion I_{en} of the emitter current is, for the Figure 3 model, the very small difference between the two very large terms on the right. If I_{en} is considered to be vanishingly small relative to each of these terms, (10a) rearranges into an expression of the Boltzmann-relation type[2] for the emitter-base interface, in that then, nearly enough for this <u>emitter</u> interface in the Figure 3 model,

$$\frac{n_{p1}}{n_{ne}} = \exp \frac{-(V_{se} + V_e)}{V_T} \quad , \qquad (12)$$

From (10b), an identical equation relates p_{n1} to p_p, for this model. Thus because of the smallness of I_{en} relative to each of the two oppositely-directed contributions comprising it, this Boltzmann-type relation is acceptably valid at the emitter-base for the Figure 3 model, although it is applied to not quite an equilibrium state -- there is for this mode a quasi-equilibrium state at this interface, both as to I_{en} and I_{ep}.

In sharp contrast, at the Figure 3 base-collector interface V_c is significantly positive, making the exponential terms on the right sides of (11a,b) relatively very small, so that, nearly enough for this <u>collector</u> interface in the Figure 3 model,

$$I_{cn} = q_e S \frac{n_{p2}\zeta_n}{2\sqrt{\pi}} \quad , \qquad (13a)$$

$$I_{cp} = q_e S \frac{p_{n2}\zeta_p}{2\sqrt{\pi}} \quad . \tag{13b}$$

Here the charge transport is just that due to the random current densities, the condition being totally non-equilibrium in nature. Use of the Boltzmann relation at the base-collector interface would for the Figure 3 model be violently wrong. This is a consequence of choosing, in this model as in the usual n-p-n transistor amplifier circuit, a significantly positive value for V_c, in contrast to the negative value for V_e.

The need is for equations valid for any set of applied voltages V_e and V_c, not merely such a set as Figure 3 illustrates. Thus neither (12) nor (13) are adequate, although both are conceptually useful in understanding the behavior of the dominant carriers in operating circuits. It is convenient to eliminate V_{se} and V_{sc} from (10) and (11) by using applicable forms like (5), then to add and subtract either n_{po} or p_{no} terms, to obtain, for the emitter-base interface,

$$I_{en} = - q_e S \frac{\zeta_n n_{po}}{2\sqrt{\pi}} \left(\exp \frac{-V_e}{V_T} - 1 \right) + q_e S \frac{\zeta_n (n_{p1} - n_{po})}{2\sqrt{\pi}} , \tag{14a}$$

$$I_{ep} = - q_e S \frac{\zeta_p p_{noe}}{2\sqrt{\pi}} \left(\exp \frac{-V_e}{V_T} - 1 \right) + q_e S \frac{\zeta_p (p_{n1} - p_{noe})}{2\sqrt{\pi}} . \tag{14b}$$

Similarly, at the base-collector interface,

$$I_{cn} = - q_e S \frac{\zeta_n n_{po}}{2\sqrt{\pi}} \left(\exp \frac{-V_c}{V_T} - 1 \right) + q_e S \frac{\zeta_n (n_{p2} - n_{po})}{2\sqrt{\pi}} , \tag{15a}$$

$$I_{cp} = - q_e S \frac{\zeta_p p_{noc}}{2\sqrt{\pi}} \left(\exp \frac{-V_c}{V_T} - 1 \right) + q_e S \frac{\zeta_p (p_{n2} - p_{noc})}{2\sqrt{\pi}} . \tag{15b}$$

Note the occurrence of the voltages in the exponential-minus-one types of terms that characterize the Ebers and Moll equations.[1]

These (14, 15) equations retain generality in being applicable for any applied voltages short of "breakdown," also for any base thickness, and for any design selections of majority carrier densities, as long as all remain small enough to permit use of classical kinetic theory. Geometry and materials choices employed for an n-p-n device often make $n_{ne} \ll p_p$; then the (14a) electron flow dominates charge transport at the emitter-base interface, and I_{ep} of (14b) can be neglected.

However, the designer may still choose the majority-carrier density n_{nc} in the collector to be much larger than, or comparable with, or much smaller than, the majority-carrier density p_p in the base. If he makes it small, ($n_{nc} \ll p_p$), charge transport across the base-collector interface is primarily due to passage of holes as in (15b), the (15a) electron-borne current being trivial. If he makes $n_{nc} \gg p_p$, this charge transport will be primarily as in (15a), (15b) being trivial. The complete (15a,b) equations can be used to describe the behavior for any choice of n_{nc} relative to p_p, as long as both are small enough to permit use of the kinetic-theory equations.

Completely parallel but inversely arranged equations apply for a p-n-p device.

IV. CHARGE TRANSPORT BY DIFFUSION AND BY ELECTRICAL CONDUCTION; THE EINSTEIN RELATION

In this problem, as in most plasma and semiconductor studies, it is found that in the bulk material the net current density (sometimes called the "drift current") is small relative to the "random current density,"[2] $n\zeta_n q_e/2\sqrt{\pi}$ or $p\zeta_p q_e/2\sqrt{\pi}$. To initiate a study of bulk-material current flow, let

D_n, D_p, symbolize the coefficients of diffusion respectively for electrons and holes, and

G_n, G_p, symbolize correspondingly the respective mobilities;[3] thus in the bulk material:

$$\left.\begin{array}{l}\text{Flow of electrons per}\\ \text{unit area per second}\end{array}\right] = -D_n \text{grad } n + nG_n \text{grad } v, \tag{16a}$$

$$\left.\begin{array}{l}\text{Flow of holes per unit}\\ \text{area per second}\end{array}\right] = -D_p \text{grad } p - pG_p \text{grad } v. \tag{16b}$$

These are vector equations, the signs indicating that the flow in response to grad v takes place toward a higher potential for electrons, and toward a lower potential for holes, whereas the diffusion flow for both electrons and holes take place in the direction of decreasing number density of the kind of particle flowing. The Einstein relation, familiar in electrical plasma studies as well as in semi-conductor problems, is that

$$D_n = V_T G_n \quad ; \quad D_p = V_T G_p \tag{17}$$

This permits restating (16) as

$$\left.\begin{array}{l}\text{Flow of electrons per}\\ \text{unit area per second}\end{array}\right] = -D_n\left[\text{grad } n + n \text{ grad}(v/V_T)\right] \tag{18a}$$

$$\left.\begin{array}{l}\text{Flow of holes per unit}\\ \text{area per second}\end{array}\right] = -D_p\left[\text{grad } p - p \text{ grad}(v/V_T)\right] \tag{18b}$$

Study of bulk-material charge transport in the light of (18), and of Poisson's equation relating space-charge density to electric field variance, shows that in the types of semiconductors here being studied, any current due to majority carrier flow occurs almost wholly in response to electric field forces, whereas current due to minority carriers results from their diffusion flow. Also, for these materials mobilities and majority carrier densities are such that for all ordinary circuit currents the bulk-material voltage gradients are very small.

Thus, for semiconductors having donor and acceptor impurity densities usually employed in junction transistors

(a) Majority-carrier transport and its associated current flow is governed by the grad (v/V_T) terms in (18a,b),

(b) Minority-carrier transport and its associated current flow is governed by the density-gradient terms in (18a,b).

In each case the non-governing terms can be neglected.

As to gross behavior in the junction transistor, in the very thin base the x-directed charge flow, due in the Figure 3 n-p-n model to passage of electrons from the emitter-base interface to the base-collector interface, takes place by diffusion of minority carriers, and not by electrical conduction of majority carriers. This diffusion flow is accompanied by recombination of holes with electrons, annihilating equal numbers of each, causing fewer electrons to leave via the collector than enter via the emitter. The holes that participate in the recombination are majority carriers, so that their lateral motion (perpendicular to the paper in Figure 3), from the base circuit connection to the interior, occurs by electrical conduction, constituting the base current. There is of course, an ohmic "base resistance." In summary in the base the major (x-directed) current occurs by diffusion of minority carriers, and the much smaller lateral base current occurs by electrical conduction.

V. RECOMBINATION AND THE DIFFERENTIAL EQUATION FOR THE MINORITY-CARRIER DENSITY DISTRIBUTION

There can occur in the bulk material either net volume generation or net volume loss of both electrons and holes. Recombination occurs in "trapping centers" in which minority carriers are "trapped" in a way that favors recombination with one of the many majority carriers present.

Let ℓ_t = mean free path of a minority carrier to a trapping condition, which represents its "end of life,"

τ_n, τ_p, symbolize mean free time to trapping, that is, the "lifetime," for the minority-carrier electrons and holes respectively,

$\bar{c}$ = average random velocity of these minority carriers.

As an example of the meanings of these definitions, in the base of the Figure 3 model:

$$\left.\begin{array}{l}\text{Trapping collisions per second,}\\ \text{for all electrons in a unit volume}\end{array}\right] = \frac{n_p c}{\ell_t} \quad , \tag{19}$$

and for any one electron

$$\tau_n = \ell_t / \bar{c} \quad , \tag{20}$$

so that

$$\left.\begin{array}{l}\text{Trapping collisions per second per}\\ \text{unit volume, for all the electrons}\end{array}\right] = n_p/\tau_n \tag{21}$$

For a semiconductor in the disconnected state, the rate of generation of electron-hole pairs must in each local region equal the rate of recombination, which is obtained by adapting (21) to the form n_{po}/τ_n. But the rate of generation is a spontaneous thermal property of the material, and is therefore the same in the connected as in the disconnected state. Net recombination is of course recombination minus generation, so that:

In a p-type material (the base in Figure 3):

$$\left.\begin{array}{l}\text{Volume-rate net recombination}\\ \text{loss of minority carriers}\end{array}\right] = \frac{n_p - n_{po}}{\tau_n} \tag{22}$$

In an n-type material:

$$\left.\begin{array}{l}\text{Volume-rate net recombination}\\ \text{loss of minority carriers}\end{array}\right] = \frac{p_p - p_{no}}{\tau_p} \tag{23}$$

The minority-carrier loss rate is important, because the minority carriers must be introduced from some external source, as across an interface; the loss rate is insignificant in relation to majority carriers.

Recombination represents a "sink" for both electrons and holes. Therefore in relation to (18a,b), considering only the diffusion terms for the minority carriers, the density distributions of minority carriers are governed by the following differential equations, for p-type and n-type regions respectively:

$$D_n \text{ div grad } (n_p - n_{po}) = (n_p - n_{po})/\tau_n \quad , \tag{24a}$$

$$D_p \text{ div grad } (p_n - p_{no}) = (p_n - p_{no})/\tau_p \quad . \tag{24b}$$

The inclusion of grad $(n_p - n_{po})$ and grad $(p_n - p_{no})$ rather than grad n_p and grad p_n is permissible and adds convenience. The present study will deal with (24) for planar flow, becoming

$$D_n \frac{d^2(n_p - n_{po})}{dx^2} = \frac{n_p - n_{po}}{\tau_n} \quad , \tag{25a}$$

$$D_p \frac{d^2(p_n - p_{no})}{dx^2} = \frac{p_n - p_{no}}{\tau_p} \quad . \tag{25b}$$

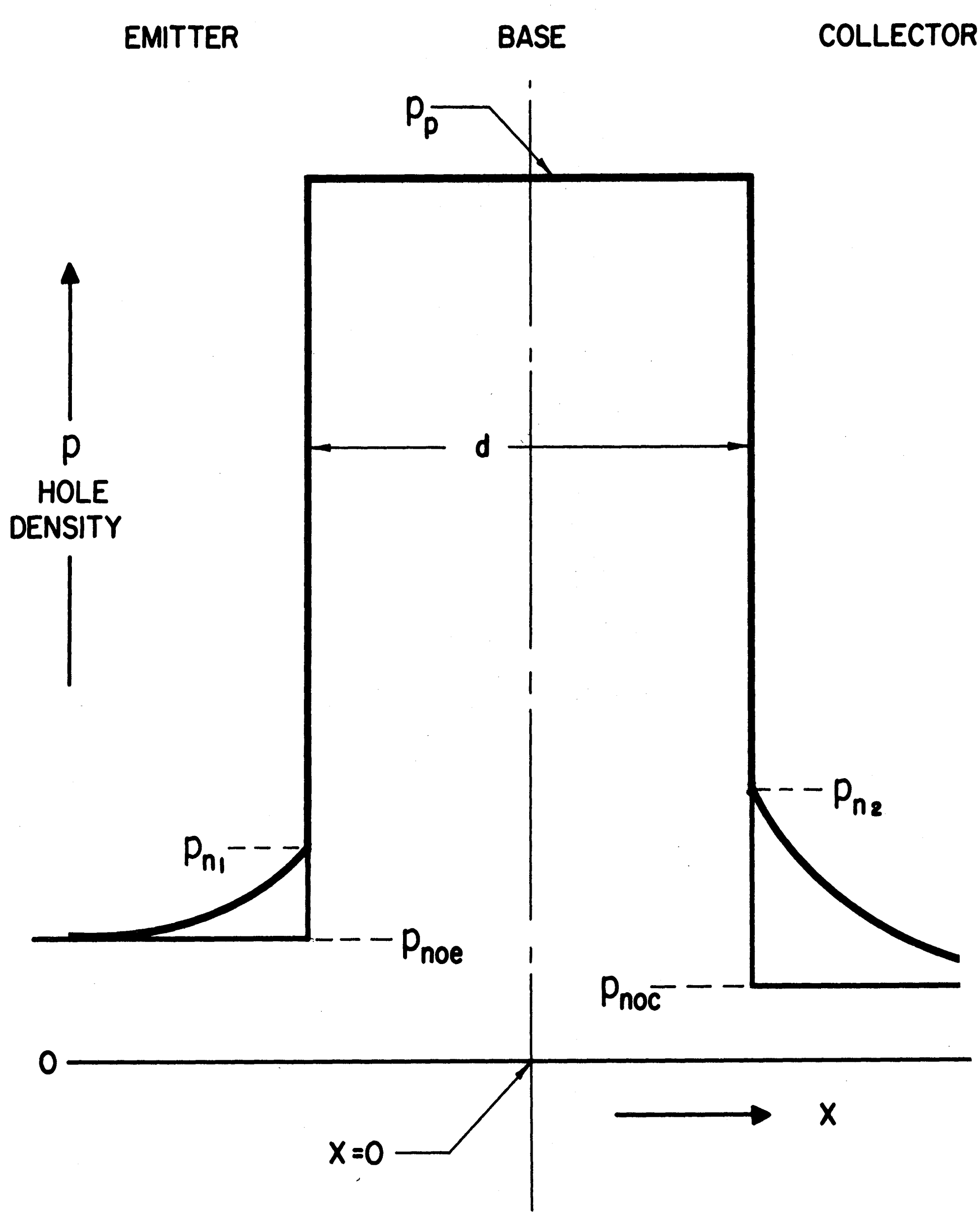

Figure 4. For a condition in which both collector and emitter currents are negative, causing electrons to enter the base, and holes to leave the base, across both interfaces, illustrative hole density distribution diagram (n-p-n transistor).

VI. DENSITY DISTRIBUTION OF MINORITY CARRIERS

The initial form of the solution of (25a) as applied to the base in the Figure 3 model is

$$(n_p - n_{po}) = A\cosh(x/\sqrt{D_n\tau_n}) + B\sinh(x/\sqrt{D_n\tau_n}), \qquad (26)$$

with A, B, governed by boundary conditions at the two interfaces. $x = 0$ is chosen at the midplane of the base, of thickness d. Then, from Figure 3a,

$$\begin{aligned} &\text{When } x = -d/2 \quad , \quad n_p = n_{p1} \quad ; \\ &\text{When } x = +d/2 \quad , \quad n_p = n_{p2} \quad . \end{aligned} \qquad (27)$$

Use of these gives, as the density distribution of minority carriers in the base:

$$(n_p - n_{po}) = (n_{p1} - n_{po}) \frac{\sinh[(\frac{1}{2}d - x)/\sqrt{D_n\tau_n}]}{\sinh(d/\sqrt{D_n\tau_n})} + (n_{p2} - n_{po}) \frac{\sinh[(\frac{1}{2}d + x)/\sqrt{D_n\tau_n}]}{\sinh(d/\sqrt{D_n\tau_n})} \quad . \qquad (28)$$

In the emitter and collector the boundary conditions are that

$$\begin{aligned} &\text{When } x = -d/2 \quad , \quad p_n = p_{n1} \\ &\text{When } x = -\infty \quad , \quad p_n = p_{noe} \end{aligned} \qquad (29a)$$

$$\begin{aligned} &\text{When } x = +d/2 \quad , \quad p_n = p_{n2} \\ &\text{When } x = +\infty \quad , \quad p_n = p_{noc} \end{aligned} \qquad (29b)$$

The $x = \pm\infty$ conditions appear because the hole lifetimes in the emitter and collector are such that all the holes entering from the base recombine in a distance short relative to the total emitter or collector length.

Application of these boundary conditions to the solution of (25b) gives the following for the minority carrier densities in the emitter and collector respectively, illustrated in Figure 4 (for circuit conditions very different from those for Figure 3):

$$(p_n - p_{noe}) = (p_{n1} - p_{noe}) \exp[(\tfrac{1}{2} d + x)/\sqrt{D_p \tau_p}] \quad \text{, and} \qquad (30a)$$

$$(p_n - p_{noc}) = (p_{n2} - p_{noc}) \exp[(\tfrac{1}{2} d - x)/\sqrt{D_p \tau_p}] \quad . \qquad (30b)$$

VII. VOLT-AMPERE EQUATIONS FOR ELECTRON-BORNE INTERFACE CURRENTS

The minority-carrier diffusion current at any point in the Figure 3 base is expressible by means of the following one-dimensional adaptation of the minority-carrier form of (18a), S being cross-sectional area:

$$I_n = q_e S D_n \frac{d(n_p - n_{po})}{dx} \qquad (31)$$

For Figure 3 this is numerically negative in that it describes a flow of negative charge in the +x direction, because in the Figure 3 base the density gradient is everywhere negative. Use of this in relation to (28), with specialization to give I_{en} at $x = -d/2$, I_{cn} at $x = +d/2$, gives

$$I_{en} = \frac{q_e S \sqrt{D_n/\tau_n}}{\sinh(d/\sqrt{D_n \tau_n})} \left[(n_{p1} - n_{po}) \cosh \frac{d}{\sqrt{D_n \tau_n}} - (n_{p2} - n_{po}) \right], \qquad (32a)$$

$$I_{cn} = \frac{q_e S \sqrt{D_n/\tau_n}}{\sinh(d/\sqrt{D_n \tau_n})} \left[(n_{p2} - n_{po}) \cosh \frac{d}{\sqrt{D_n \tau_n}} - (n_{p1} - n_{po}) \right]. \qquad (32b)$$

Evidently $(n_{p1} - n_{po})$ and $(n_{p2} - n_{po})$ can be eliminated between (32),(14a and 15a), giving a pair of equations relating the electron-borne interface currents I_{en} and I_{cn} to the voltage terms. The procedure is straightforward, and yields the following

$$I_{en} = -\frac{I_o}{K_s} \left[K_n \left(\exp \frac{-V_e}{V_T} - 1 \right) - \left(\exp \frac{-V_c}{V_T} - 1 \right) \right], \qquad (33a)$$

$$I_{cn} = -\frac{I_o}{K_s} \left[-\left(\exp \frac{-V_e}{V_T} - 1 \right) + K_n \left(\exp \frac{-V_c}{V_T} - 1 \right) \right]. \qquad (33b)$$

Here and for later use:

I_o = the random current across the area S due to minority carriers of the base in the disconnected state; for the Figure 2, Figure 3 model,

$$\left.\begin{array}{l}\text{Minority-carrier base random current}\\ \text{density in the disconnected state}\end{array}\right] = \frac{q_e \zeta_n n_{po}}{2\sqrt{\pi}} \; ; \qquad (34a)$$

$$I_o = \frac{q_e S \zeta_n n_{po}}{2\sqrt{\pi}} \; ; \qquad (34b)$$

$$K_s \equiv 2 \cosh \frac{d}{\sqrt{D_n \tau_n}} + \left(R_n + \frac{1}{R_n}\right) \sinh \frac{d}{\sqrt{D_n \tau_n}} \; ; \qquad (34c)$$

$$K_n \equiv \cosh \frac{d}{\sqrt{D_n \tau_n}} + \frac{1}{R_n} \sinh \frac{d}{\sqrt{D_n \tau_n}} \; ; \qquad (34d)$$

$$K_p = K_s \frac{\zeta_p/\zeta_n}{1 + R_p} \; ; \qquad (34e)$$

$\sqrt{D_n \tau_n}$ = [a "scale distance," comparable in concept to the time constant in an electrical transient, also called diffusion length] (35a)

$\sqrt{D_n/\tau_n}$ = [a velocity type of quantity used to characterize the (31,32) minority-carrier flow in the presence of trapping recombination] (35b)

$$R_n = \frac{\zeta_n/2\sqrt{\pi}}{\sqrt{D_n/\tau_n}} = \frac{\text{Electron random-current-density velocity}}{\text{Velocity-dimensioned quantity } \sqrt{D_n/\tau_n}} \qquad (36a)$$

$$R_p = \frac{\zeta_p/2\sqrt{\pi}}{\sqrt{D_p/\tau_p}} = \frac{\text{Hole random-current-density velocity}}{\text{Velocity-dimensioned quantity } \sqrt{D_p/\tau_p}} \qquad (36b)$$

To evaluate R_n and R_p, (17) is used to eliminate D_n in favor of the mobility G_n, and $2q_e/m_e$ is numerically stated and used in (2) to express ζ, thus obtaining

$$R_n = \frac{5.93 \times 10^5}{2\sqrt{\pi}\sqrt{m_n/m_e}} \sqrt{\frac{\tau_n}{G_n}} \quad , \text{ (in the base) ,} \tag{37a}$$

$$R_p = \frac{5.93 \times 10^5}{2\sqrt{\pi}\sqrt{m_p/m_e}} \sqrt{\frac{\tau_p}{G_p}} \quad , \text{ (in emitter and collector)} \tag{37b}$$

Use here of typical measured values of the mass ratios and the τ's and G's, show that almost universally,

$$R_n \gg 1 \quad ; \quad R_p \gg 1. \tag{37c}$$

The (33) equations have the Ebers and Moll forms, but express only the electron-borne interface currents, not the total emitter and collector currents.

VIII. VOLT-AMPERE EQUATIONS FOR HOLE-BORNE INTERFACE CURRENTS

For minority-carrier diffusion currents in the emitter or the collector, the counterpart of (31) is

$$I_p = - q_e S D_p \frac{d(p_n - p_{no})}{dx} \quad . \tag{38}$$

Use of this in relation to (30a,b), employing x = -d/2 and x = +d/2 to give I_{ep} and I_{cp}, permits obtaining

$$I_{ep} = - q_e S \sqrt{D_p/\tau_p}(p_{n1} - p_{noe}), \tag{39a}$$

$$I_{cp} = - q_e S \sqrt{D_p/\tau_p}(p_{n2} - p_{noc}), \tag{39b}$$

Use of (14b) and (15b) eliminates p_{n1} and p_{n2}; then rearrangement and expression in terms of I_o of (34b), with use of (34e), leads to hole-borne interface current expressions in forms convenient for combination with (33):

$$I_{ep} = - I_o \frac{p_{noe}}{n_{po}} \frac{K_p}{K_s} \left(\exp \frac{-V_e}{V_T} - 1 \right) \quad , \tag{40a}$$

$$I_{cp} = - I_o \frac{p_{noc}}{n_{po}} \frac{K_p}{K_s} \left(\exp \frac{-V_c}{V_T} - 1 \right) \quad . \tag{40b}$$

IX. THE EBERS AND MOLL VOLT-AMPERE EQUATIONS: CONTINUITY

Use with (33) and (40) of $I_e = I_{en} + I_{ep}$ and $I_c = I_{cn} + I_{cp}$, from (8a), gives the following volt-ampere equations, including contributions of both electrons and holes:

$$I_e = -\frac{I_o}{K_s}\left(K_n + K_p\frac{p_{noe}}{n_{po}}\right)\left(\exp\frac{-V_e}{V_T} - 1\right) + \frac{I_o}{K_s}\left(\exp\frac{-V_c}{V_T} - 1\right) \quad (41a)$$

$$I_c = +\frac{I_o}{K_s}\left(\exp\frac{-V_e}{V_T} - 1\right) - \frac{I_o}{K_s}\left(K_n + K_p\frac{p_{noc}}{n_{po}}\right)\left(\exp\frac{-V_c}{V_T} - 1\right). \quad (41b)$$

These have the Ebers and Moll form

$$I_e = A_{11}\left(\exp\frac{-V_e}{V_T} - 1\right) + A_{12}\left(\exp\frac{-V_c}{V_T} - 1\right), \quad (42a)$$

$$I_c = A_{21}\left(\exp\frac{-V_e}{V_T} - 1\right) + A_{22}\left(\exp\frac{-V_c}{V_T} - 1\right), \quad (42b)$$

in which $A_{12} = A_{21}$, and A_{11} differs from A_{22} only because of a difference between the impurity densities in the emitter and the base, causing p_{noe} to differ from p_{noc}. There is complete mathematical symmetry but not materials symmetry, as between the emitter and collector.

Section VII and its origins deal with a river of electrons flowing into the Figure 3 n-p-n transistor at its emitter terminal, passing through the base to the collector, then out through the collector terminal. Electrons are lost from this river by recombination in any region where the minority-carrier density in the circuit-connected conditions exceeds that for the the disconnected state (p_{no} or n_{po}). This occurs first - and usually only in small degrees - in the emitter in the region of approach to the emitter-base interface. It occurs quite substantially in the base; there may be loss of electrons by recombination in the collector. There applies to this river of electrons, at every point, the concept of "continuity"; the flow into any local region must equal the flow out, plus the loss by recombination within the region. Within the emitter and collector, the electron river flow constitutes a majority-carrier current, carried by electrical conduction. Within the base, it constitutes a minority-carrier current, carried by diffusion. In (41a,b) all terms except those containing K_p relate to the electron flow at the interfaces.

Section VIII and its origins deal with a river of holes flowing into the Figure 3 n-p-n transistor at its base terminal, moving laterally across the base, being largely lost by recombination in the base, as a desert river may disappear into the ground. There may be some overflow of holes across the interfaces into the emitter and collector, where the recombination is completed. Thus all the holes that enter at the base terminal are lost by recombination, in one or another of the three portions of the transistor. Just as for the electrons, the concept of "continuity" applies to holes in any local region. Within the base, the hole river flow constitutes a majority-carrier current - primarily but not wholly lateral - carried by electrical conduction. This river's overflows into the emitter and collector constitute minority-carrier currents, carried by diffusion. In (41a,b), only the terms containing K_p relate to the hole flow at the interfaces.

X. JUNCTION TRANSISTOR CHARACTERISTIC EQUATIONS

Transistor characteristic curves usually chart the dependence of collector current on collector voltage, with emitter current a parameter. This suggests inversion of (41) and (42) into the form

$$I_c = -\alpha_N I_e - I_{co}\left(\exp\frac{-V_c}{V_T} - 1\right) \quad , \tag{43a}$$

$$I_e = -\alpha_I I_c - I_{eo}\left(\exp\frac{-V_e}{V_T} - 1\right) \quad . \tag{43b}$$

Here, after using in (41) the (6) relations $p_{noe}/n_{po} = p_p/n_{ne}$, and $p_{noc}/n_{po} = p_p/n_{nc}$,

$$1 - \alpha_N = \frac{A_{11} - A_{21}}{A_{11}} = \frac{K_n - 1 + K_p(p_p/n_{ne})}{K_n + K_p(p_p/n_{ne})} \quad , \tag{44a}$$

$$1 - \alpha_I = \frac{A_{22} - A_{12}}{A_{22}} = \frac{K_n - 1 + K_p(p_p/n_{nc})}{K_n + K_p(p_p/n_{ne})} \tag{44b}$$

$$I_{co} = (A_{12}A_{21} - A_{11}A_{22})/A_{11}, \tag{45a}$$

$$I_{eo} = (A_{21}A_{12} - A_{22}A_{11})/A_{22}. \tag{45b}$$

Values given K_n in engineering devices will be governed by the requirement that $1 - \alpha_N$, and sometimes $1 - \alpha_I$, must be as small as possible consistent with reasonable reproducibility from unit to unit. This requires from (44) that $K_n - 1$ be small, which in turn requires that $d/\sqrt{D_n\tau_n}$ be small, that is, the base thickness must be a moderately small fraction of the diffusion length in the base. Use of this, and of $R_n \gg 1$ from (37c) means that, nearly enough in most designs,

$$K_n = 1 + (d^2/2D_n\tau_n), \text{ and } K_n^2 - 1 = d^2/D_n\tau_n \quad . \tag{46a}$$

With $d/\sqrt{D_n\tau_n}$ small, $R_n \gg 1$, $R_p \gg 1$, and use of (36a) for R_n, (34c,e) for K_s and K_p reduce to

$$K_s = R_n \frac{d}{\sqrt{D_n\tau_n}} = \frac{\zeta_n}{2\sqrt{\pi}} \frac{d}{D_n} \quad , \tag{46b}$$

$$K_p = \frac{\sqrt{D_p/\tau_p}}{\sqrt{D_n/\tau_n}} \frac{d}{\sqrt{D_n\tau_n}} = \frac{d}{\sqrt{D_p\tau_p}} \frac{D_p}{D_n} \tag{46c}$$

Note that $\sqrt{D_p\tau_p}$ is the scale distance (diffusion length) in the emitter and collector.

With these approximations introduced, because of the requirement that $1 - \alpha_N$ must be small, so making $d/\sqrt{D_n\tau_n}$ small, the α_N, α_I, transmission coefficients and the I_{co} residue current in (43a,b) become, nearly enough in most devices,

$$1 - \alpha_N = \frac{\dfrac{d^2}{2D_n\tau_n} + \dfrac{p_p}{n_{ne}} \dfrac{d}{\sqrt{D_p\tau_p}} \dfrac{D_p}{D_n}}{1 + \dfrac{d^2}{2D_n\tau_n} + \dfrac{p_p}{n_{ne}} \dfrac{d}{\sqrt{D_p\tau_p}} \dfrac{D_p}{D_n}} \quad ; \tag{47a}$$

$$1 - \alpha_I = \frac{\dfrac{d^2}{2D_n\tau_n} + \dfrac{p_p}{n_{nc}} \dfrac{d}{\sqrt{D_p\tau_p}} \dfrac{D_p}{D_n}}{1 + \dfrac{d^2}{2D_n\tau_n} + \dfrac{p_p}{n_{nc}} \dfrac{d}{\sqrt{D_p\tau_p}} \dfrac{D_p}{D_n}} \quad ; \tag{47b}$$

$$I_{co} = q_e S n_{po} \sqrt{\frac{D_n}{\tau_n}} \; \frac{\dfrac{d}{\sqrt{D_n\tau_n}}\left(1 + \dfrac{p_p^2}{n_{ne}n_{nc}} \dfrac{D_p\tau_n}{D_n\tau_p}\right) + \left(\dfrac{p_p}{n_{ne}} + \dfrac{p_p}{n_{nc}}\right)\left(1 + \dfrac{d^2}{2D_n\tau_n}\right)\sqrt{\dfrac{D_p\tau_n}{D_n\tau_p}}}{1 + \dfrac{d^2}{2D_n\tau_n} + \dfrac{p_p}{n_{ne}} \dfrac{d}{\sqrt{D_p\tau_p}} \dfrac{D_p}{D_n}} \tag{47c}$$

The expression for I_{eo} is obtained by using n_{nc} rather than n_{ne} in the third denominator term in this last equation. Equations (43), (47) describe volt-ampere properties that correspond as to general form with those observed experimentally for typical junction transistors.

From (47a), it is clear that to make $1 - \alpha_N$ as small as possible, the majority carrier density in the base should be small relative to that in the emitter ($p_p \ll n_{ne}$), and this is common design practice. With this condition met, but assuming the majority carrier density n_{nc} in the collector to be comparable with the value p_p in the base, the two significant equations for use in (43a) to chart characteristics become, after rearrangement using (6),

$$1 - \alpha_N = \frac{d^2/2D_n\tau_n}{1 + (d^2/2D_n\tau_n)} \quad ; \qquad (48a)$$

$$I_{co} = q_e \frac{(Sd)}{1 + (d^2/2D_n\tau_n)} \frac{n_{po}}{\tau_n} + q_e(S\sqrt{D_p\tau_p}\,) \frac{p_{noc}}{\tau_p} \qquad (48b)$$

The second term on the right of this last equation describes an effect of hole flow across the base-collector interface; there is no significant hole flow across the emitter-base junction in this design. Study of these forms, applying when $n_{ne} \gg p_p$, but n_{nc} is comparable with p_p, leads to the following comments:

1. The forward carrier transmission coefficient α_N is not affected by the occurrence of the substantial hole flow across the base-collector interface;

2. The variation of $1 - \alpha_N$ as d^2 indicates a very great sensitivity to base thickness. Extremely thin dimensions are difficult to

control to close tolerances; therefore serious engineering difficulties appear in efforts to make $1 - \alpha_N$ (or $1 - \alpha_I$) at the same time very small and highly reproducible from unit to unit. The inverse variation as the square of the diffusion length $\sqrt{D_n \tau_n}$ indicates that to retain reproducibility, from batch to batch of units, very close quality control must be maintained on this property of the bulk material of the base.

3. The residue current I_{co} may in this design consist primarily of hole flow across the base-collector interface; as d is presumably a rather small fraction of $\sqrt{D_p \tau_p}$, the second term on the right in (48b) may be expected to be much larger than the first.

4. The form of (48b) shows that the residue current is a measure of the rate of generation of electron and hole pairs, as discussed in the following section.

XI. THE RESIDUE CURRENT MEASURES PAIR GENERATION

The form of (43a) shows that I_{co} is the lower limit approached by the collector current as V_c is indefinitely increased, the emitter terminal being disconnected from any external circuit, that is, $I_e = 0$. I_{eo} is similarly related to V_e and a zero-value I_c. To aid clarity, the physical nature of I_{co} will be discussed for the model underlying (48a,b) in which $n_{ne} \gg p_p$, but n_{nc} is comparable with p_p.

With V_c strongly positive, and $I_e = 0$, the collector current, now I_{co}, is numerically positive, comprising a flow of electrons into the collector terminal, and the base current, now $-I_{co}$, is numerically negative, comprising a flow of holes to the base terminal. With no carriers of either kind entering, the I_{co} current must result from a net generation of electron and hole pairs within the transistor. From (22), (23), and their origins, pair generation appears mathematically as negative recombination, and as such can occur in either an n-type or p-type material when the existing minority-carrier density is less than the disconnected-state value n_{po} or p_{no}. The maximum net generation density rate is n_{po}/τ_n or p_{no}/τ_p. Therefore, in (48b):

1. The first term describes a current that is very slightly less than that due to generation at the maximum volume rate n_{po}/τ_n throughout the base of volume (Sd). For a vanishingly small $(d^2/2D_n\tau_n)$, I_{co} becomes just a measure of the volume generation of electron and hole pairs in the base.

2. The second term describes a current corresponding to generation at the volume rate p_{noc}/τ_p in the collector-region volume $(S\sqrt{D_p\tau_p})$, that is, the portion of the base extending one diffusion length away from the base-collector interface. This is presumably much larger than the first term, for the reason that d is much smaller than either diffusion length, while p_{noc} and n_{po} will be of comparable magnitudes, because p_p and n_{nc} are presumed so; τ_n and τ_p will be comparable for the two opposite-kind but comparable impurity densities.

Thus the form of the first term in (48b) implies that when V_c is moderately large and $I_e = 0$, $n_p \ll n_{po}$ throughout the base.

This can be verified by solving (25a) for the appropriate boundary conditions, which are that: (a) because $n_{ne} \gg p_p$, hole current is negligible at the emitter-base interface, so that with $I_e = 0$, also $I_{en} = 0$, thus calling for zero slope of the n_p vs. x curve at the left interface; (b) because V_c is large, electrons pass only from the left to right across the base-collector interface; therefore also, (c) the electron diffusion in the base is toward the collector, requiring the density distribution to have an increasingly negative slope toward the right. With $R_n \gg 1$, and $d/\sqrt{D_n\tau_n}$ a moderately small fraction of unity, this solution leads to the observation that $n_p \ll n_{po}$ throughout the base, and also to the first term on the left of (48b).

The corresponding solution for (25b) in the collector employs boundary conditions that: (d) the curve of p_n vs. x must be asymptotic to p_{noc} at $x = \infty$ (this implies that the collector extends much farther than $d/\sqrt{D_p\tau_p}$); and (e), the holes generated in the collector must flow toward the base-collector interface, thus requiring in the collector a decreasing positive slope of hole distribution. The curve so obtained is exponential, with $p_n \ll p_{noc}$ everywhere. The average generation described by then using (23) is equal to that in the diffusion length $d/\sqrt{D_p\tau_p}$; this verifies the second term on the right in (48b).

It is found by measurement that in transistors designed for engineering utility, the residue current I_{co} is much larger than as given by (48b), presumably for the reason that there is a large contribution from the generation of pairs at or near the external surfaces (interfaces with vacuous or gas-filled regions) rather than by generation of pairs in the bulk material. This deserves comment.

An n-type semiconductor material, a metallic conductor and a conducting plasma all contain large numbers of freely-roving electrons whose random velocities greatly exceed any drift velocity. All three also have positively-charged particles that are either stationary or nearly so. In the neighborhood of an external surface, the electrons' random motions should presumably tend to make them roam out across this surface into the outer region, perhaps remaining there permanently. This does not happen, as evidenced by the survival of the materials and of the plasma. The observed retention of the electrons can happen only because of the existence of a near-the-surface region in which the potential drops away toward the surface -- sufficiently so to prevent significant electron escape. A positive space-charge density must exist in the region just inside the surface to permit an adequate drop in potential to occur there.

In an n-type semiconductor, this positive space charge is due to an excess of the density of positively-charged donor-type impurity centers over the majority carrier density, for this near-the-surface region. It is not difficult to estimate the depth to which this space-charge region must extend in order to create a potential barrier sufficient to oppose the majority-carrier random-motion travel outward toward the surface. The positive space-charge density of course tapers off gradually toward the interior (rather than abruptly). This space-charge region is the analog of the positive ion sheath at an inactive boundary of a conducting plasma.

Evan for a single small, disconnected, electrically uncharged sample of an n-type material this positive surface sheath cannot comprise the total electrical surface phenomena, because if it did the actually uncharged object would appear to have a net positive charge. The positive space-charge sheath is in fact the positive half of a surface-region electrical "double layer" of charge. In a low-density confined plasma, the negative half of the corresponding double layer is the surface charge on the confining envelope. In a metallic conductor, the negative half consists of the electrons that succeed in moving a small distance -- of the order of the lattice spacing -- out away from the last layer of atoms. In here discussing the negative half for a semiconductor, a completely clean external surface is postulated.

It will now be shown that, in an n-type semiconductor, the negative half of the double layer exists as a surface charge associated with the valence-bond electrons of the last layer of atoms. As to valence properties, an atom of germanium may equally well be thought of as being hungry to gain four more valence electrons, or desirous of getting rid of the four that it has. In the bulk material it accomplishes both simultaneously, by sharing two valence bonds with each of four adjacent atoms. One may say that its hunger has been satisfied by adding four electrons, one from each of its four near neighbors, or one may say it has achieved solitude by giving away its own four electrons, one to each neighbor. Either way, the pattern achieved is one of completeness of satisfaction of each atom's needs through the assistance of its four neightbors -- in the bulk material.

For an atom at the external surface, not all the needs can be satisfied in this way, because at least one of its near neighbors is missing. With one neighbor missing, the point of view may be that this atom has only succeeded in giving away three of its four electrons, so can achieve a complete pattern by giving away one more, thus acquiring a positive charge. The released electron can become a freely-moving bulk-material conduction electron. Or, the point of view may be that this atom has gained only three electrons from its neighbors, and can achieve completeness by stealing one more from the interior region, thus acquiring a negative charge. The stolen electron leaves behind a hole which can move in to provide bulk-material conductivity. After either type of change, the affected atom is said to occupy a "surface energy state." It is evident that any originally unsatisfied atom may thus act either as a donor or acceptor, with essentially equal electron-volt energies for the two types of exchange. Each atom that acts as a donor has become an element of positive surface charge, and each acting as an acceptor an element of negative surface charge. For an intrinsic semiconductor, these will occur equally, and there will be no surface charge. For an n-type sample, the negative surface charge elements will predominate to the degree necessary to provide the negative half of the surface-region double layer that prevents near approach of conduction electrons to the surface, and for a p-type sample the positive surface charges similarly predominates.

Some surface atoms may remain unsatisfied, and there may be trading around of the charge-holding property among the satisfied and unsatisfied surface atoms, representing a capability for diffusion and conduction along the surface. If by means of a tangential surface field, or by diffusion, there is a steady drift of these surface charges away from any given region, there must occur a continuous generation of them, with an associated continuous release of bulk-material electrons and holes to the interior.

Within the interior, whether for an n-type or p-type material, the respective densities of majority and minority carriers, that appear in (6), are determined by means of quantum-mechanical effects that include as a criterion of self-consistency that the bulk material must remain electrically uncharged. In the near-the-surface space-charge sheath this criterion does not exist -- there will not be electrical neutrality. In general, for a p-type material, the majority carrier density will in this space-charge region become less than the bulk p_p, and the minority carrier density greater than the bulk n_{po}. With this change will come an increase-perhaps a very marked one--in the volume generation rate above its bulk-material value n_{po}/τ_n.

Thus it appears that, because of boundary-value external-surface requirements, there will be substantial pair generation capabilities near and perhaps at such surfaces quite above those existing in the bulk material. Yet these surface regions are subject, as is the bulk material, to the n-type and p-type interface drain of electrons toward the high-potential side, and of holes to the low-potential side, of a base-collector interface operating at a high collector voltage. This tends to drain away the surface-region generation products, just as for the bulk-material generation products, and a detailed study of the physical electronics again brings in the Ebers and Moll voltage factor $[\exp(-V_c/V_T) - 1]$.

The surface-region regeneration occurs of course along the external surface of the base adjacent to the junction with the collector, and to some degree also along the external surface of the collector adjacent to the same junction. Although transistor design usually calls for a very thin base, the base external surface may be substantial, so that the generation region contributing to I_{co} may extend a distance away from the base-collector junction that is many times the thickness of the base. The surface-region flow of current may take place by diffusion, leading to the view that the distance away from the base-collector junction within which generation contributing to I_{co} takes place may be roughly the scale distance $\sqrt{D_n \tau_s}$, where τ_s is a "lifetime" property of the surface region, perhaps very different from τ_n of the bulk material.

The surface-region contributions to the residue currents I_{co} and I_{eo} may be from 10 to 1000 times greater than the bulk generation contribution described by (47c) and (48b), but the surface region behavior has relatively little effect on the transport coefficients α_N and α_I. Surface contamination may strongly affect surface generation.

There is often observed experimentally a contribution to the residue currents I_{co} and I_{eo} which is obviously of a conduction-current nature, in that this current contribution varies in proportion to the voltage. This is attributable to electrical conduction along the external surface of the base, as caused by the voltage difference between collector and emitter. This surface conduction may be though of as due partly to conduction in very thin layers of contaminants overlaying the semiconductor surface, or as strictly surface-charge conduction existing as part of the double-layer structure at the boundary between a p-type or n-type semiconductor and its external vacuous or gas-filled environment. Neither of these effects involves the factor $[\exp(-V_c/V_T) - 1]$, and must therefore be accounted for by including, on the right sides of (43a,b), an additional term directly proportional to the appropriate voltage, in one case V_c-V_e, in the other V_e-V_c. In all ordinary cases of transistor usage, either $V_e \ll V_c$, making the added residue current in fact proportional to V_c, or, the conduction current residue is negligible relative to that called for by the exponential term. The effect on the collector volt-ampere characteristic curves is to give them all a less than infinite slope in the region of the chart useful for amplifier design.

XII. ALTERNATIVE APPROACHES TO THE SOLUTION

There were described early in Section XI means to evaluate I_{co} by study of the $I_e = 0$, large V_c condition, and similarly to evaluate I_{eo} by study of the $I_c = 0$, large V_e condition. A solution for the small V_c and small V_e conditions is necessary to establish the exponential-minus-one multiplier to the I_{co} and I_{eo} terms in (43). This paper has presented one method of making this solution.

An alternative method is to study the dependence on V_c, first when $I_c = 0$, but I_e may have any value, second when $I_e = 0$, but I_c may have any value, thus obtaining in turn $(-I_{co}/\alpha_N)$ and I_{co} of (43a). This method uses only two-terminal models, the third terminal being in each case disconnected.

Still another method is to evaluate A_{11} and A_{12} in (42a) by first studying the physical dependence of I_e on V_e, with $V_c = 0$, then the dependence of I_e on V_c with $V_e = 0$. I_e and I_c may each have values when the respective voltages V_e and V_c are zero.

The alternative methods suggested are nq simpler mathematically than the one used in this paper, if hole flow across interfaces is to be considered; note that zero net current does not imply either zero electron current or zero hole current. Both methods achieve substantial simplicity if hole flow is ignored, but this places rather severe limitations on attention to the effects of design choices. Neither alternative method makes it initially clear why the device has the linear dependence on the exponential-minus-one factors; this dependence is usually initially introduced by heuristic considerations, which are sometimes confused by an attempt to apply the Boltzmann relation to grossly non-equilibrium conditions; see the discussion below (12). The method here used has the distinct advantage of using for the analytical model the Figure 3 density distribution configuration which corresponds to the actual behavior in a circuit-connected device, thus giving the reader initially a correct intuitive understanding of the internal behavior in regard to diffusion and recombination.

REFERENCES

1. Ebers, J. J. and Moll, J. L. "Large-Signal Behavior of Junction Transistors," Proc. I.R.E., Vol. 42, (December, 1954) 1761-1772.

2. Dow, W. G. "Fundamentals of Engineering Electronics," 2nd Edition, John Wiley and Sones, 1952: References are primarily to Chapter XII, dealing with the derivation of and equation forms for the time-exposure-over-a-surface and snap-shot-throughout-a-volume Maxwellian distribution of random particle velocities.

3. Conwell, E. M. "Properties of Silicon and Germanium: II": Proc. I.R.E. Vol. 46, (June, 1958) 1281-1300.