DERIVATION AND INTERPRETATION OF A GENERALIZED CHARGE-CONTROL THEORY AND RECIPROCITY FOR A BIPOLAR TRANSISTOR

Thesis by

E. Ted Grinthal

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1969

(Submitted April 22, 1969)

ACKNOWLEDGEMENTS

I am very much indebted and grateful to Professor R. D. Middlebrook for his help and guidance. This work was in part supported by the U. S. Navy through Contract N60530 - 68 - C - 0530, and partially conducted with the aid of National Science Foundation and Radio Corporation of America Fellowships, for which I am grateful.

The patience and skillful typing of Mrs. Carol M. Teeter is very much appreciated. Finally, I would like to thank my colleagues at Caltech for numerous helpful discussions during the course of this research.

ABSTRACT

The charge-control relations are rigorously derived from the carrier transport and continuity equations for a bipolar transistor with an arbitrary three-dimensional geometry, arbitrary base impurity distribution, arbitrary recombination mechanisms (including spatial nonuniformity), and for both low and high injection levels. A one-to-one correspondence is maintained between internal processes and the charge-control parameters, so that conceptual understanding of, and insight into, device operation is enhanced. In the absence of recombination, the average carrier velocity is used to obtain the average carrier transit time across the base region. The current is then obtained as the ratio of injected base charge to the average transit time. In the presence of recombination, the injected carriers are divided into two groups according to whether they recombine or are collected. The collected current is then obtained as the ratio of the collected charge to the average transit time of the collected carriers. The Beaufoy-Sparkes "collector time constant" is related to the carrier transit time and is given a conceptual interpretation as a collection lifetime in analogy with the recombination lifetime. A recombination transit time is introduced in analogy with the collection transit time.

The theory, which is generally valid up to frequencies of the order of the reciprocal transit time, is extended to include high injection levels and some second-order phenomena, such as the Early effect and nonideal base contacts. It is pointed out that the integration of the basic equations over the base region may lead to a loss of detailed information, so that the charge-control theory may accurately describe

only the average behavior of the device; a solution of this difficulty is suggested. The Ebers-Moll circuit relations are derived from charge-control principles so that a charge-control interpretation of the Ebers-Moll parameters and an electrical interpretation of the charge-control parameters is obtained. This leads to the conclusion that the short-circuit saturation currents are of more fundamental significance than are the open-circuit saturation currents.

Finally, the separation of injected carriers into recombining and nonrecombining components is used to obtain a conceptually clear derivation of the principle of reciprocity for a transistor.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

In 1957, Beaufoy and Sparkes [1,2] presented the basic concepts for consideration of the junction transistor as a charge-controlled, rather than a current controlled, device. Their analysis was strictly applicable only to a uniform-base transistor under the conditions of spatially uniform recombination, one-dimensional current flow, and low injection levels. In addition, the excess minority carrier distribution in the base was assumed to be approximately linear with position, even in the presence of recombination. This model was highly mathematical in that each time constant introduced was defined as the ratio of stored (injected) base charge to the relevant current, and no conceptual interpretation was attached to this ratio. Johnson and Rose [3] introduced the ideas of charge in transit and transit time, and defined the current gain in terms of the effective lifetime of the minority carriers (which was undefined) and the transit time. Their analysis is quite general and involves no assumptions as to geometry, recombination, injection level, or base region impurity distribution.

Other workers [4-12] have extended and refined the charge-control concepts, but many questions have been left unanswered. Moll and Ross [13] calculate the minority carrier transit time, but introduce a velocity which is given no conceptual meaning. Varnerin [4] attempts to overcome this objection by referring to the ratio of excess (injected) charge in the base to the emitter current as the average time spent per carrier in the base (in the absence of recombination), so that this ratio can be called a transit time. Baker [7] limits the recombination rate to be spatially uniform in the base region, while Sparkes [8] indicates the limitations of the one-dimensional analysis. Sparkes also extends the analysis to graded-base transistors, states that the collector time-constant is equal to the mean transit time if the depletion layer width variation with voltage [14] can be neglected, and discusses the effect of high level injection on the transit time. Baker and May [9] extend the high frequency usefulness of charge-control analysis by deriving a time-dependent transit time to account for delay effects, but fail to mention the inherent breakdown of the concepts at transit-time frequencies. Schmeltzer [11] extends the analysis to three dimensions, combines surface and bulk recombination in an effective lifetime, and includes time dependence from the beginning, but he defines time constants mathematically with no clear relation to carrier motions or processes occurring within the device. den Brinker, et al. [12] separate the injected current into two components: the current lost by recombination near the injecting junction, and the current which is collected. They then neglect the first component and treat the second component as $I_C = \alpha I_E$.

From the foregoing it can be seen that the charge-control method of transistor analysis is generally limited to low-level injection and one-dimensional current flow in a uniform base transistor. Although some attempts have been made to overcome these restrictions, there is no single, comprehensive, coherent theory which eliminates all of them simultaneously. Further, the velocities and time constants introduced by various workers are mathematical quantities, with little or no conceptual meaning. In addition, those workers who refer to a "transit time" fail to explain of what it is the transit time if there is recombination, when not all of the injected carriers are collected. As is pointed out by Narud, Hamilton, and Lindholm [15,16], and Koehler [17,18], the charge-control concepts lie somewhere between a physical model (such as that of Linvill [19]) and an electrical model (such as that of Ebers and Moll [20]).

As originally proposed [1,2], the charge-control method was primarily employed for large-signal analysis of transistor operation, although it has been extended to small-signal and transient conditions. Beaufoy [5], Schmeltzer [11], den Brinker, et al. [12], and others have given examples of circuit design and analysis using the charge-control method, and techniques for measuring the charge-control parameters have been presented by Beaufoy and Sparkes [1], Sparkes [6,8], and Boothroyd [10]. A comparison of the charge-control method with other large-signal models of the transistor has been given by Narud, et al. [15,16] and Koehler [17,18].

It is the purpose of this work to present a coherent theoretical foundation for the charge-control methods. General charge-control results will be derived for a bipolar transistor with arbitrary three-dimensional geometry, arbitrary impurity distribution, arbitrary recombination mechanisms (including spatial nonuniformity), and for any injection level. The parameters involved in the theory will be clarified by relating them to conceptual processes occurring within the device. This will allow the theory to be used for obtaining understanding of, and insight into, device operation, for DC as well as low frequency AC and slow transient conditions. The parameters will also be given electrical (circuit) interpretations where appropriate, and the usefulness of these electrical and conceptual interpretations will be demonstrated. These results will be used to obtain a derivation of transistor reciprocity which is based on conceptual processes occurring within the device. Some fundamental and practical limitations of the theory will be presented. Primary concern will be with DC and low frequency AC steady-state conditions. The displacement current will be neglected throughout this work.

In Chapter II the charge-control parameters are rigorously and logically derived from the transport and continuity equations. Each parameter is given a clear conceptual significance by relating it to conceptual processes occurring within the transistor. This is accomplished mainly by separation of the injected charge into the charge that recombines and the charge that survives (i. e. does not recombine). This leads to the introduction of two new parameters: δ, <u>the fraction of charge that recombines</u> and t_r, <u>the recombination transit time</u>. Some limitations, fundamental and otherwise, of the charge-control theory are presented in Chapter III and some applications and extensions of the theory are given in Chapter IV. In Chapter V the Ebers-Moll equations are derived from the charge-control theory so far developed, allowing both an electrical (circuit) interpretation to be given to the charge-control parameters and a charge-control interpretation to the Ebers-Moll parameters. Finally, in Chapter VI the separation of injected charge into recombining and nonrecombining (surviving) components is used in a derivation of the principle of reciprocity for a bipolar transistor.

In the appendix the conventional (one-dimensional) derivation of the current relations for a diode under high-injection conditions is presented.

CHAPTER II

BASIC CHARGE-CONTROL CONCEPTS

In their original paper, Beaufoy and Sparkes [1] treat a one-dimensional transistor with a uniformly doped base. They introduce time constants, each of which is defined as the ratio of injected charge to the relevant current, but fail to attribute any conceptual meaning to them. Under the assumptions of one-dimensional current flow and no recombination, but with a nonuniform impurity distribution allowed, Moll and Ross [13] derive an expression for carrier transit time but, as Varnerin [4] points out, they introduce a velocity with no direct conceptual significance. In addition, it is not obvious how to extend their results to three dimensions or to include recombination.

In general (DC or AC, steady-state or transient) the carrier transport current through a surface is given by the time rate of flow of charge through that surface

$$I = \frac{dQ}{dt} \tag{2.1}$$

In the DC steady-state the current is constant, so that Eq. 2.1 may be integrated as

$$I \int_0^{\tau} dt = \int_0^{Q} dQ \tag{2.2}$$

$$I\tau = Q \tag{2.3}$$

$$I = \frac{Q}{\tau} \tag{2.4}$$

where Q is the total amount of charge passing through the surface in time τ. Equation 2.4 is valid for any time τ and its associated charge Q in the DC steady-state. Furthermore, if this same charge passes through another surface some distance removed from the first surface*, Eq. 2.4 gives the current through that surface also.

In the case of time-varying signals, Eq. 2.4 does not give exactly the current through the two different surfaces. The reason for this is that, so far, the nonzero time taken by the charge to travel between the two surfaces has not been accounted for. It is clear however, that if the signal frequency is sufficiently low then this effect can be neglected. This represents a (frequency) limitation of the theory and is discussed further in Chapter III.

Charge-control theory, as developed from the Beaufoy-Sparkes model, starts from Eq. 2.4 and declares that the total excess minority charge in the base is a convenient value for Q and that τ is whatever is required to obtain the proper current value. For the base current, τ is the recombination lifetime if it is uniform throughout the base. For the collector current (for emitter injection) in the absence of recombination, τ is the average time taken by carriers to travel from the emitter to the collector (the transit time). Unfortunately, τ is given no conceptual interpretation at the injecting

---

* For instance, the emitter and collector junctions of a bipolar transistor with no recombination.

junction, or at the collecting junction in the presence of recombination.

In this chapter Eq. 2.4 will be derived from the carrier transport and continuity equations in such a way that not only the convenience, but also the limitations of using the excess minority charge for Q will become obvious. The associated time τ will be obtained simultaneously with the charge from the same equations, so that its significance will be clear. No restrictions will be made on geometry or impurity distribution, and nonuniform, although linear, recombination rates will be allowed. For the sake of definiteness an NPN transistor will be considered, with the positive (hole) current directions as shown in Figure 2.1. Note that electron flow from emitter to collector represents a positive emitter current but a negative collector current.

For conciseness of notation, the following convention will be adopted. If there is no subscript to indicate whether hole (p) or electrons (n) are being considered, then the parameter in question refers to either or both. A current (density) I (j) with a single subscript (E,B,C) indicates the total flow through the emitter, base, or collector due to all mechanisms being considered. The symbol ${}_{\mu}I_{\nu}$ (${}_{\mu}j_{\nu}$) refers to the current (density) through surface ν due to injection at surface μ. Although the discussion will be primarily in terms of electrons injected into the base from the emitter of an NPN transistor, the principles also hold for holes as the current carrier (majority or minority), injection from the collector, or for a PNP transistor.

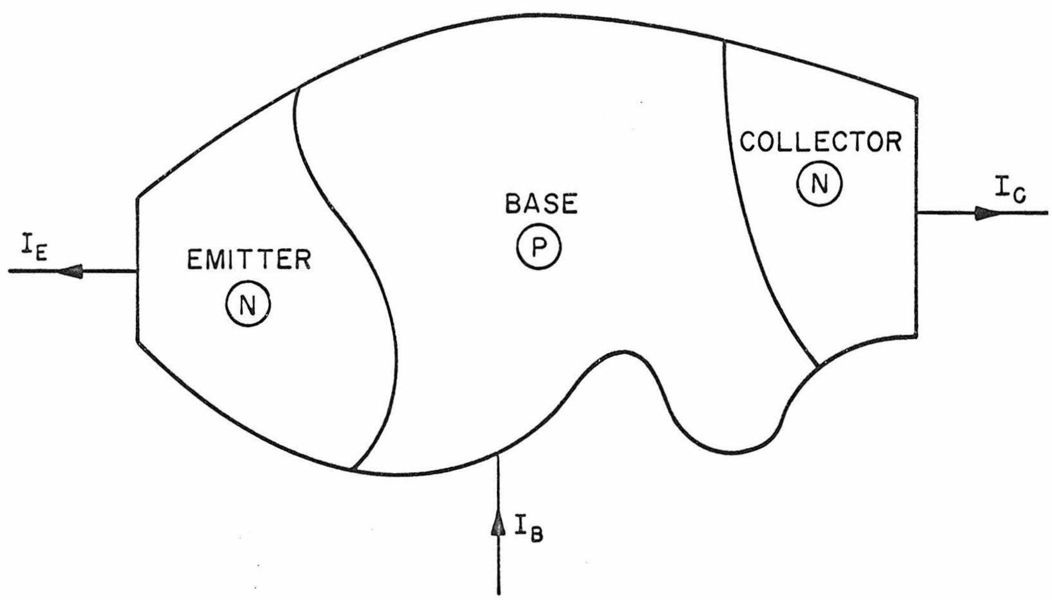

Fig. 2.1 Definition of positive current directions. Holes flow in direction of arrows, electrons flow opposite from arrows for positive currents.

2.1 Three Dimensions, No Recombination

2.1.1 Velocities

The electron transport current density* can be expressed as [21]

$$j = e \int_{u=-\infty}^{u=\infty} u \, dn(u) \tag{2.5}$$

where u is the component of individual electron velocity along the direction of current flow $(\underset{\sim}{j}/j)$ and can be either positive or negative according to whether the velocity is parallel or anti-parallel to the net carrier flow, and $dn(u)$ is the density of electrons with velocities between u and $u + du$. The integration is to be carried out over the total density of electrons at the point at which the current density is to be evaluated. The quantities j, u, and dn may vary with position, but the total current must be independent of position for the DC steady-state since the condition of no recombination has been specified.

The electron density and the velocity may each be expressed as the sum of two terms: that which is due to conditions existing in

---

* The quantity j (as defined by Eq. 2.5) is the negative of the conventionally defined electron current density, so that Eq. 2.5 must be considered as the electron charge flux density. To avoid minus signs, Eq. 2.5, rather than its negative, will be used throughout this work, and j will be referred to as a current density. For holes, a similar relation to Eq. 2.5 will be used, so that both j_p and j_n are positive for positive carrier velocities. Thus the total current density is $j_p - j_n$ rather than the usual $j_p + j_n$.

equilibrium (subscript o), and that which is due to a change in boundary conditions such as an increase in applied voltage (primed). Then Eq. 2.5 becomes

$$j = e \int (u_o + u')(dn_o + dn') \qquad (2.6)$$

$$= e \int_{u_o=-\infty}^{\infty} u_o dn_o(u_o) + e \int_{u'=-\infty}^{\infty} u' dn_o(u') + e \int_{u_o=-\infty}^{\infty} u_o dn'(u_o)$$

$$+ e \int_{u'=-\infty}^{\infty} u' dn'(u') \qquad (2.7)$$

where u_o is that part of the individual electron velocity u which can be attributed to random thermal motion and drift in a built-in electric field. Thus the velocity u_o can be attributed to conditions which exist in thermal equilibrium, and it will be referred to as the equilibrium velocity of individual electrons. The velocity u' is that part of the individual electron velocity u which can be attributed to nonequilibrium conditions, such as an electric field which is caused by the injected carriers (i.e. due to a breakdown of quasi-neutrality at high injection levels); this velocity component will be referred to as the excess velocity of individual electrons. The parameters u_o and u' are algebraic quantities, with a positive or negative value indicating whether the velocity component is parallel or anti-parallel to the direction of net carrier flow.

The first two terms to the right of the equal sign in Eq. 2.7

are to be integrated over the equilibrium electrons with the indicated velocity ranges. The last two terms are to be integrated over the excess electrons with the indicated velocity ranges.

The first term on the right of Eq. 2.7 is the equilibrium current and is of course zero*. The second tern is the current due to a change in velocity of the equilibrium carriers, as in the un-pinched-off region of a junction field effect transistor [22, 23] or in an ordinary resistor. This term is also important for high injection levels in diodes and transistors. The third term is the current due to injected (excess) carriers traveling at equilibrium velocities u_o (positive or negative), as in a junction diode or transistor under low-level injection conditions. The last term is the current due to the injected carriers traveling at excess velocities u' which are determined, in part, by the injected carriers (or external bias conditions). This term is important in a diode or transistor at high injection levels and is the entire current in the pinched-off region of a junction field effect transistor [24, 25]. In this work low-injection conditions in a diode or transistor will be of primary interest, and so attention will be focussed on the third term in Eq. 2.7. However, in Chapter IV, Section 4.2, high injection levels will be explicitly considered, and the second and fourth terms will be included. Although attention will be restricted to PN junction diodes and bipolar transistors, the principles herein developed are also applicable (with some modifications)

---

* In equilibrium there are as many carriers with velocity $+|u_o|$ as there are carriers with velocity $-|u_o|$, so that the first integral is zero without either u_o or n_o being zero.

to other types of devices, as demonstrated by Johnson and Rose [3] and Middlebrook [26].

In order to eliminate the integrations from Eq. 2.7, we make the following definitions:

$$n_o \equiv \int dn_o \tag{2.8}$$

$$n' \equiv \int dn' \tag{2.9}$$

$$n \equiv n_o + n' \tag{2.10}$$

$$v_o \equiv \frac{1}{n'} \int_{u_o=-\infty}^{\infty} u_o dn'(u_o) \tag{2.11}$$

$$v' \equiv \frac{1}{n} \int_{u'=-\infty}^{\infty} u' dn(u') \tag{2.12}$$

where the integrations in Eqs. 2.8 and 2.9 are carried out over all equilibrium or excess electrons respectively, regardless of their individual velocities, the integration in Eq. 2.11 is carried out over all excess electrons, and the integration in Eq. 2.12 is carried out over all electrons.

From Eq. 2.11 it is seen that v_o is the average value of the individual equilibrium velocity for excess electrons at a point*. The velocity v_o will be referred to as the point-average equilibrium velocity of excess electrons and will later be related to the ensemble velocity of a group of electrons (see Eqs. 2.15 and 2.16). Millman and Seely [27] have shown that electrons injected from a metal into a vacuum obey the Maxwell-Boltzmann distribution function in the vacuum if they were Fermi-Dirac distributed in the metal and the barrier potential (work function) is more than a few kT above the Fermi level. This result also holds for injection from the emitter into the base region of a transistor, if the density of carriers in the base is sufficiently small that the Pauli exclusion principle can be ignored in determination of the electron energy distribution. If this condition is satisfied, then the energy distribution of injected electrons is unaffected by the electrons already in the base region, so that the injected electron energy distribution is the same as if the electrons were injected into a vacuum except that the effective mass, rather than the free-electron mass, must be employed. Thus, if the density of injected electrons is sufficiently large that a statistical analysis is

---

* This is an average over all excess electrons at a point in space and is independent of time in the DC steady-state. It may be imagined that we obtain v_o by adding algebraically the velocities of excess electrons crossing a small surface area normal to the current flow, dividing by the total density of electrons involved, and taking the limit as the element of surface area becomes infinitesimal. The quantity v_o is to be distinguished from the time-average velocity of an individual electron, which is zero in the absence of an electric field and is the drift velocity if there is an electric field.

valid*, it may be concluded that the point-average equilibrium velocity v_o of injected electrons is independent of the density n' of injected electrons.

From Eq. 2.12 it is seen that v' is the average value of the individual electron excess velocity u' for all electrons. The quantity v' will be referred to as the average excess velocity of electrons.

With the use of Eqs. 2.8 - 2.12, Eq. 2.7 may be rewritten as (since the first term is zero)

$$j = env' + en'v_o \qquad (2.13)$$

For low injection levels the first term is negligible** so that we have

$$j = en'v_o \qquad (2.14)$$

where j is the transport current density in the direction of current flow, n' is the density of excess or injected electrons, and v_o is the point-average equilibrium velocity of the injected electrons. Thus,

---

* For very low injection levels, the injected density will be too small to yield a meaningful statistical average, so that Eq. 2.11 will not be a particularly useful expression. For the remainder of this work, we will always consider a sufficiently high injection level that statistical averages will be meaningful.

** This can be seen by comparing the two terms with the aid of Eqs. 4.6-4.8 for the velocities and Eq. A.4 for the current density.

since all terms in Eq. 2.14 have a clear conceptual meaning, Varnerin's objections [4] to using Eq. 2.14 as a starting point for the derivation of the charge-control parameters have been overcome.

An alternative interpretation of the velocity v_o can be obtained from the usual expression for electron current density in low injection

$$\underset{\sim}{j} = - e \mu_n \underset{\sim}{E}_o n' - eD_n \underset{\sim}{\nabla} n' \tag{2.15}$$

where $\underset{\sim}{E}_o$ is the electric field due to any impurity gradient. Comparison of Eq. 2.14 (in vector form) and Eq. 2.15 shows that

$$\underset{\sim}{v}_o = - \mu_n \underset{\sim}{E}_o - D_n \frac{\underset{\sim}{\nabla} n'}{n'} \tag{2.16}$$

so that v_o can be interpreted as the component, in the direction of current flow, of the vector sum of drift and diffusion velocities. Under low-injection conditions (for which Eqs. 2.14 - 2.16 are valid), the shape of the spatial distribution of excess electrons is independent of the magnitude of n' (or the externally applied bias). Hence, it may be concluded that v_o as given by Eq. 2.16 is independent of the injected electron density, as it must be if Eqs. 2.14 and 2.15 are to give the same result. The velocity v_o can be considered either as the point-average equilibrium velocity of individual electrons or as an ensemble velocity. The former interpretation was used in the discussion of Eq. 2.11, while the latter description is more appropriate here. Since the excess density gradient is of opposite sign (and may be of different shape and magnitude) for emitter and collector injection, it follows that in general v_o is not the same for emitter and collector injection, or

$$_{E}v_o \neq {}_{C}v_o \tag{2.17}$$

where $_{\mu}v_o$ is the ensemble velocity of carriers that were injected at surface μ.

Equation 2.16 leads to the conclusion that v_o becomes infinite when n' becomes zero if the excess density gradient approaches zero at a slower rate than does n', or remains nonzero. This is, of course, impossible. There are three conditions which may cause the injected electron density to become zero at a certain position: first, the trivial case of zero applied voltage or current: second, the condition of infinite recombination rate, which is excluded here; and finally, the condition of a forward-biased emitter and a reverse-biased collector. If the collector injection is considered to be negative, the net injected density will go through zero in the base region, while the gradient remains nonzero. In this case, the problem of an infinite velocity can be eliminated by retention of the distinction between emitter and collector injection. In this way a v_o can be defined for each injecting junction, neither of which becomes infinite. Alternatively, the distinction between injected and equilibrium carriers could be eliminated and only the total electron density (which is never zero) considered. Unfortunately, this does not lead to easily understandable results. Conventional first-order theory contradicts the above argument by assuming that the density of emitter-injected electrons is zero at the collector. This is impossible and will not be considered further. (See Middlebrook [28,29] for a discussion of this point.)

2.1.2 Characteristic Times and Charges

To simplify the analysis in a general three-dimensional geometry, it will be convenient to introduce a set of orthogonal flow-tube coordinates. In this system, x is defined to be the position coordinate along the axis of a flow tube of current (i.e. along the direction of current flow $\underset{\sim}{j}/j$), and y and z are the position coordinates, orthogonal to each other, lying in the surface normal to the flow tube (i.e. on a constant quasi-Fermi level surface). This coordinate system is illustrated in Figure 2.2. It may be noted that this coordinate system reduces to the familiar Cartesian system if the base region is rectangularly shaped and the impurity gradient is appropriately directed.

With this definition of coordinates, Eq. 2.14 may be written as (for emitter-injection only)

$$j(x,y,z) = e \, {}_E n'(x,y,z) \, {}_E v_o(x,y,z) \tag{2.18}$$

where ${}_E n'$ is the excess electron density due to injection from the emitter, ${}_E v_o$ is the ensemble velocity of electrons injected from the emitter, and $j(x,y,z)$ is the current density along a flow-tube. (Similar notation will be used for collector injection and for holes.) Rearrangement of Eq. 2.18 and integration over the base volume leads to

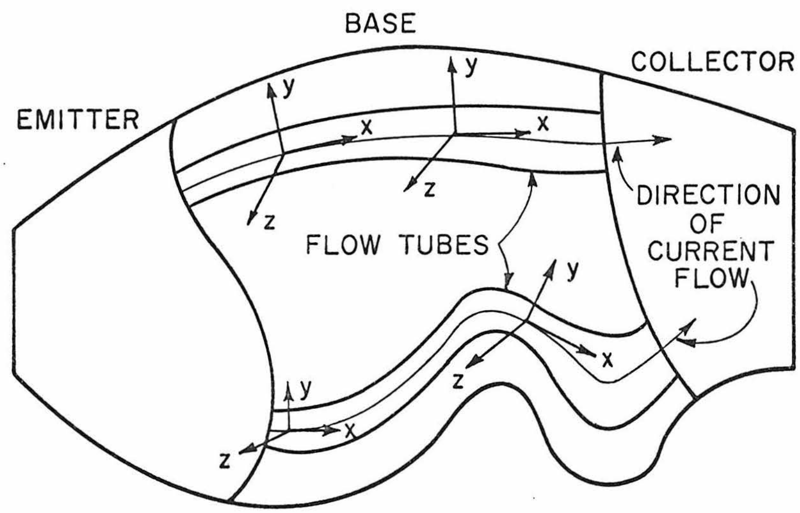

Fig. 2.2 Example of a flow-tube coordinate system.

$$\iiint_{\text{Base}} \frac{j(x,y,z)}{{}_E v_o(x,y,z)}\, dx\,dS = \iiint_{\text{Base}} e_E n'(x,y,z)\,d(\text{vol}) \tag{2.19}$$

where dS is the cross-sectional area of a flow tube and is not necessarily constant along a flow tube. Since recombination has been excluded, the current must be constant along a flow tube, so that $j(x,y,z)dS$ must be independent of position x along the flow tube. Thus, Eq. 2.19 can be rewritten as

$$\iint_S j(x,y,z) \left[\int_0^{W(y,z)} \frac{dx}{{}_E v_o(x,y,z)} \right] dS = \iiint_{\text{Base}} e_E n'(x,y,z)\,d(\text{vol}) \tag{2.20}$$

where $W(y,z)$ is the length of a flow tube from the emitter to the collector, and the surface integration may be carried out over any surface that crosses all of the flow tubes in the base region. The x integral on the left-hand side of Eq. 2.20 yields the emitter-to-collector transit time for a particular flow tube

$${}_e t_c(y,z) = \int_0^{W(y,z)} \frac{dx}{{}_E v_o(x,y,z)} \tag{2.21}$$

This is essentially the same expression for transit time as given by Moll and Ross [13] except that here it has been generalized to an arbitrary three-dimensional geometry and the velocity has been given a clear conceptual interpretation as the point average equilibrium velocity of excess electrons (Eq. 2.11) or the ensemble velocity of excess electrons (Eq. 2.16). Recognizing that the right-hand side of Eq. 2.20 is the total excess charge of electrons in the base due to injection from the emitter ${}_E Q'$, we have

$$\iint_S j(x,y,z)\ {}_e t_c(y,z)\,dS = {}_E Q' \tag{2.22}$$

The average transit time can now be defined as

$${}_E t_C = \langle {}_e t_c \rangle = \frac{\iint_S j(x,y,x)\ {}_e t_c(y,z)\ dS}{\iint_S j(x,y,z)\ dS} \tag{2.23}$$

so that ${}_E t_C$ is the average emitter-to-collector transit time per flow tube, weighted by the current carried by the flow tube. It is important to note that, since v_o is different for emitter and collector injection, the average emitter-to-collector transit time is not equal to the average collector-to-emitter transit time:

$${}_E t_C \neq {}_C t_E \tag{2.24}$$

Since recombination has been excluded, the current through any surface S is the same as the current through any other surface (delay time neglected), so that the denominator of Eq. 2.23 is simply the total current. If we substitute Eq. 2.23 into Eq. 2.22 and rearrange, we arrive at the desired result

$$ {}_E I_E = - {}_E I_C = \frac{{}_E Q'}{{}_E t_C} \tag{2.25} $$

This is the usual charge-control result. However, the reason for using the total excess (injected) electron charge ${}_E Q'$ in the base is now clear: this quantity arises "naturally" from integration of the equation of carrier transport over the entire base volume (a "natural" choice for the integration region). Furthermore, the "time constant" involved in the result has been directly related to internal processes (i.e. carrier transport) rather than being defined as the ratio of charge to current.

The charge ${}_E Q'$ may be given either of two interpretations. The usual description is that ${}_E Q'$ is the total excess electron charge stored in the base; however, this implies a static condition which does not exist (any steady-state condition which may exist is a dynamic steady-state). A dynamic interpretation is that ${}_E Q'$ is the electron charge injected into the base. In the absence of recombination, as here, ${}_E Q'$ is also the collected charge and the charge in transit associated with one transit time. That is, a total charge ${}_E Q'$ crosses the base from the emitter to the collector in one transit time.

2.2 Three Dimensions, Recombination

In this section only Shockley-Read [30] or similar recombination mechanisms will be considered.

2.2.1 Velocities

The presence of recombination in no way affects the discussion of velocity in Section 2.1.1. The point-average equilibrium velocity of the injected electrons (v_o) is unchanged and is to be associated with an electron as long as it exists in the base.

2.2.2 Characteristic Times and Charges

As in Section 2.1.2, the electron current density due to injection at the emitter (for low-injection) may be written

$$ {}_{E}j(x,y,z) = e\,{}_{E}n'(x,y,z)\;{}_{E}v_{o}(x,y,z) \tag{2.26} $$

However, in the presence of recombination, the current is no longer constant along a flow tube*, so that Eq. 2.26 cannot be integrated as was Eq. 2.18. In order to overcome this difficulty the injected charge is separated into two components, 1) the density of emitter-injected (E) electrons at point (x,y,z) that are destined to recombine (r) before reaching the collector ${}_{E}n'_{r}(x,y,z)$, and 2) the density of emitter-injected (E) electrons at point (x,y,z) that are destined to reach the collector (C) ${}_{E}n'_{C}(x,y,z)$. If collection by the base contact were to be accounted for, it would be necessary to add another component ${}_{E}n'_{B}$, but this current will be neglected here. These components are shown schematically in Figure 2.3. Clearly we have

$$ {}_{E}n'(x,y,z) = {}_{E}n'_{r}(x,y,z) + {}_{E}n'_{C}(x,y,z) \tag{2.27} $$

$$ {}_{E}n'_{r}(C,y,z) = 0 \tag{2.28} $$

* We consider the electron flow tubes only. The hole flow from the base contact does not affect the electron flow tubes under low injection conditions.

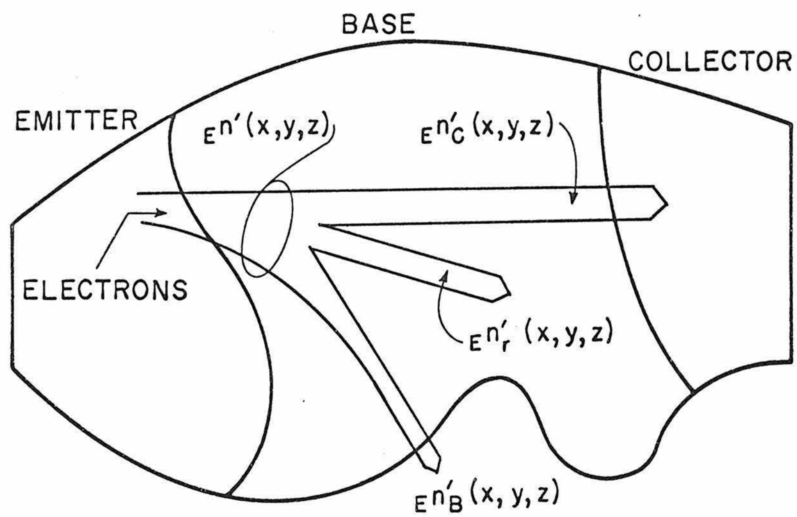

Fig. 2.3 Schematic diagram showing the destinations of emitter-injected electrons in the DC steady-state. Similar diagrams can be made for holes or for injection from other surfaces.

$$_E n' (C,y,z) = {}_E n'_C(C,y,z) \qquad (2.29)$$

if ${}_E n'_B$ is neglected, where C indicates the x coordinate at the collector.

The current density of nonrecombining (collected) electrons may now be written as

$$_E j_C(x,y,z) = e \; {}_E n'_C(x,y,z) \; {}_E v_o(x,y,z) \qquad (2.30)$$

and the current density at the emitter as

$$_E j_E(y,z) = {}_E j(E,y,z) = e \; {}_E n'(x,y,z) \; {}_E v_o(x,y,z)]_E \qquad (2.31)$$

where it has been assumed that the electrons that recombine and those that survive (do not recombine) have the same energy and velocity distribution, so that the same average velocity may be used for both.

The separation of the injected charge into two components has led to an expression for the collected current density which can be integrated in the same way as was Eq. 2.18 with no recombination. This is possible because ${}_E j_C(x,y,z)$ is the current density of the electrons that do not recombine and hence can be treated as though there were no recombination. Thus, Eq. 2.30 becomes

$$_E I_C = \frac{-{}_E Q'_C}{{}_E t_C} \qquad (2.32)$$

where

$$ {}_{E}Q'_{C} = \iiint_{\text{Base}} e \; {}_{E}n'_{C}(x,y,z)\,d(\text{vol}) \tag{2.33} $$

$$ {}_{E}t_{C} = \frac{\iint_{C} {}_{E}j_{C}(x,y,z)\; {}_{e}t_{c}(y,z)\,dS}{\iint_{C} {}_{E}j_{C}(x,y,z)\,dS} \tag{2.34} $$

and

$$ {}_{e}t_{c}(y,z) = \int_{0}^{W(y,z)} \frac{dx}{{}_{E}v_{o}(x,y,z)} \tag{2.35} $$

Equation 2.34 can also be rewritten as

$$ {}_{E}t_{C} = \frac{\iiint_{\text{Base}} e\,{}_{E}n'_{C}(x,y,z)\,d(\text{vol})}{\iint_{C} e\,{}_{E}n'_{C}(x,y,z)\,{}_{E}v_{o}(x,y,z)\,dS} \tag{2.36} $$

It can be seen from Eq. 2.36 that under conditions of low injection level and neglect of the Early effect [14], ${}_{E}t_{C}$ is independent of the applied current or voltage.

Under DC conditions, ${}_{E}Q'_{C}$ is the charge that is collected in one transit time, or is the charge in transit across the base associated with one transit time.

It is seen that ${}_EQ'_C$ is the total electron charge that is removed from the base by collection in one transit time. Thus, it is clear that ${}_EQ'_C$, rather than ${}_EQ'$, is the relevant charge for determination of the collected current.

The quantity ${}_Et_C$ is the weighted average emitter-to-collector transit time of the injected electrons that reach the collector. That is, any electron injected at the emitter either reaches the collector in an average time ${}_Et_C$ or does not reach the collector at all (i.e. recombines).

The emitter current may be obtained by integration of Eq. 2.33 over the emitter junction

$$ {}_EI_E = \iint_E {}_Ej_E(y,z)dS = \iint_E e_E n'(x,y,z) \; {}_Ev_o(x,y,z)dS \qquad (2.37) $$

We now <u>define</u>

$$ {}_E\tau \equiv \frac{\iiint_{Base} e_E n'(x,y,z)d(vol)}{\iint_E e_E n'(x,y,z) {}_Ev_o(x,y,z)dS} \qquad (2.38) $$

so that substitution of Eq. 2.38 into Eq. 2.37 yields the expression

$$ {}_EI_E = \frac{{}_EQ'}{{}_E\tau} \qquad (2.39) $$

where

$$ {}_EQ' = \iiint_{Base} e\,{}_En'(x,y,z)d(vol) \tag{2.40} $$

which is the usual charge-control relation for the emitter current under DC conditions. A conceptual interpretation of ${}_E\tau$ (the emitter injection time) will be given in Section 2.2.5. Again, the reason for use of the total injected charge in the expression for the emitter current is clear. This charge enters "naturally" from integration of the carrier transport equation over the entire base volume (a "natural" choice for the integration region). Also, ${}_EQ'$ is the total charge injected in one emitter injection time.

The total injected electron charge that recombines may also be obtained by integration of the density of injected electrons that are destined to recombine before being collected:

$$ {}_EQ'_r = \iiint_{Base} e\,{}_En'_r(x,y,z)d(vol) \tag{2.41} $$

This result will be used in Section 2.2.4.

2.2.3 Average Recombination Lifetime in the Base

The DC continuity equation for electrons is

$$ \nabla \cdot \mathbf{j} = e\,U \tag{2.42} $$

Integration of Eq. 2.42 over the base volume and use of Gauss' theorem leads to

$$I_B = \iiint_{Base} e\ U\ d(vol) \equiv R \tag{2.43}$$

where R may be considered to be the recombination rate of excess charge for the entire base volume, while U is the recombination rate of excess carriers for an infinitesimal volume. In analogy with the point lifetime, which Shockley and Read [30] define as

$$\tau_{nr}(x,y,z) \equiv \frac{n'(x,y,z)}{U(x,y,z)} \tag{2.44}$$

the recombination lifetime for the entire base (for injection from the emitter) may be defined as

$${}_E\tau_{nr} \equiv \frac{{}_EQ'}{R} \tag{2.45}$$

Use of Eqs. 2.43 and 2.44 in Eq. 2.45 gives

$$\frac{1}{{}_E\tau_{nr}} = \frac{\iiint_{Base} \frac{{}_En'}{\tau_{nr}}\ d(vol)}{\iiint_{Base}\ {}_En'\ d(vol)} \tag{2.46}$$

so that ${}_E\tau_{nr}$ may be interpreted as the average* excess electron lifetime in the base for injection from the emitter, weighted by the density

---

* $1/\tau_{nr}$ is related to the probability of capture [30]; it is really the probability that is averaged.

of electrons subject to a particular recombination rate (lifetime, or probability of capture).

If the recombination rate is spatially uniform and independent of the carrier density (low-injection), then

$$_E\tau_{nr} = \tau_{nr} \qquad (2.47)$$

and

$$_C\tau_{nr} = {}_E\tau_{nr} \qquad (2.48)$$

On the other hand, if the recombination rate is not spatially uniform or if high injection conditions exist, Eq. 2.47 is not valid. In general,

$$_En'(x,y,z) \neq {}_Cn'(x,y,z) \qquad (2.49)$$

so that the reciprocal lifetime is not averaged with the same weighting function for both emitter and collector injection. Hence, in general

$$_E\tau_{nr} \neq {}_C\tau_{nr} \qquad (2.50)$$

2.2.4 Collection Lifetime and Recombination Transit Time

In the previous section the collector current was obtained in terms of the collected charge and its transit time. In contrast, Beaufoy and Sparkes [1,2], Sparkes [8], Baker and May [9], Gray, et al. [31], and others, express the collector current in terms of the total injected charge to define a collector time constant

$$ {}_E\tau_C \equiv \frac{-{}_EQ'}{{}_EI_C} \tag{2.51} $$

to which they give no conceptual interpretation*. In this section, we will give a conceptual interpretation to ${}_E\tau_C$ and show that it is analogous to the recombination lifetime ${}_E\tau_r$.

Ignoring microscopic details, we may say that recombination and collection in a transistor base are similar mechanisms in that they both serve to remove injected carriers from the base region. Thus, since the recombination lifetime is given by (see Section 2.2.3)

$$ {}_E\tau_r = \frac{{}_EQ'}{{}_EI_B} \tag{2.52} $$

we may, be comparing Eqs. 2.51 and 2.52, consider ${}_E\tau_C$ to be a transit or collection lifetime. That is, ${}_E\tau_r$ is the mean time required for a carrier to be removed from the base by recombination alone, and ${}_E\tau_C$ is the mean time for a carrier to be removed from the base by collection alone.

Furthermore, in analogy with Eq. 2.32 for the collected current, we may define a recombination time constant by

---

* Gray, et al. [31] show that ${}_E\tau_C = W^2/2D_n$ in a one-dimensional, uniform-base transistor if recombination is neglected, but fail to identify this term as the carrier transit time.

$$ {}_{E}t_{r} \equiv \frac{{}_{E}Q'_{r}}{{}_{E}I_{B}} \tag{2.53} $$

Since ${}_{E}t_{C}$ is the mean time for a carrier to travel from the injection point to the collection point (where it is removed from the base), ${}_{E}t_{r}$ may be interpreted as the mean time for a carrier to travel from the injection point to the recombination point (where it is removed from the base). Thus ${}_{E}t_{r}$ is the recombination transit time for injection at the emitter. It is obvious that the recombination lifetime is equal to or larger than the recombination transit time, or

$$ {}_{E}\tau_{r} \geq {}_{E}t_{r} \tag{2.54} $$

since the carriers with longer recombination transit times are collected before they can recombine and are not included in the determination of ${}_{E}t_{r}$. A similar argument for the collection times leads to the conclusion that the collection lifetime is larger than or equal to the collection transit time, or

$$ {}_{E}\tau_{C} \geq {}_{E}t_{C} \tag{2.55} $$

If we define

$$ {}_{E}\delta \equiv \frac{{}_{E}Q'_{r}}{{}_{E}Q'} \tag{2.56} $$

as the fraction of injected charge that recombines*, then the integrated form of Eq. 2.27 yields

$$ {}_{E}Q'_{C} = (1 - {}_{E}\delta)\,{}_{E}Q' \tag{2.57} $$

Furthermore, Eqs. 2.32 and 2.51, and Eqs. 2.52 and 2.53, together with Eqs. 2.56 and 2.57, lead to

$$ {}_{E}\tau_{C} = \frac{{}_{E}t_{C}}{1 - {}_{E}\delta} \tag{2.58} $$

and

$$ {}_{E}t_{r} = {}_{E}\delta \; {}_{E}\tau_{r} \tag{2.59} $$

from which Eqs. 2.54 and 2.55 follow immediately, since ${}_{E}\delta \leq 1$.

The foregoing results may be interpreted as follows. The entire injected charge is to be associated with the lifetimes, both recombination and collection, whereas only the charges involved in the particular mechanism are to be associated with the transit times. Since ${}_{E}\tau_{r}$ is related to the probability per unit time of recombination [30], it may be said that ${}_{E}\tau_{C}$ is related to the probability per unit time of collection. Also, ${}_{E}Q'_{r}$ can be interpreted as the charge removed from the base by recombination in one recombination transit time. In the DC

---

* The quantity ${}_{E}\delta$ can also be interpreted as the probability of recombination for minority carriers injected at the emitter.

steady-state, this charge is also the majority charge injected through the base contact in one recombination transit time ${}_E t_r$.

2.2.5 Base Current and Injection Time

To complete the solution of the transistor, the base current must be obtained. If we use Eqs. 2.32 and 2.39 to obtain I_B as the sum of collector and emitter currents, the result will be valid only under DC conditions since we have, in those equations, not accounted for the increase in excess charge in the base region due to the nonzero emitter-to-collector transit time. In the derivations in Section 2.2.2 an instantaneous redistribution of charge after a change in the boundary conditions was assumed, whereas, of course, a nonzero time is required. During this time there is a component of base current consisting of majority carriers being injected at the base contact at a rate equal to the rate of increase in minority carriers due to emitter injection. There are two ways to account for this base current component. The tedious method is to include the time delay when calculating the emitter and collector currents so that they can be used to obtain the total (time-varying) base current. The easier method is to use the charge-control approach to obtain the base current from the continuity equation. Furthermore, this approach will lead, in a straightforward manner, to the inclusion of the recombination lifetime in the result.

For emitter injection, the electron continuity equation is

$$\underset{\sim}{\nabla}\cdot\underset{\sim}{j} = \frac{e\,{}_E n'}{\tau_r} + e\frac{\partial\,{}_E n'}{\partial t} \qquad (2.60)$$

where the recombination lifetime τ_r includes both bulk and surface recombination [32-35] and may vary with position or carrier density. Integration of Eq. 2.60 over the volume of the base, and use of Gauss' theorem in the left-hand side, leads to

$$\oiint_{\text{Base}} \underset{\sim}{j} \cdot d\underset{\sim}{S} = \iiint_{\text{Base}} \frac{e_E n'}{\tau_r} \, d(\text{vol}) + \frac{d}{dt} \iiint_{\text{Base}} e_E n' \, d(\text{vol}) \tag{2.61}$$

On the assumption that electrons do not cross the base contact (i.e. ideal base contact), so that the left hand side of Eq. 2.61 is the difference between the emitter and collector electron currents (which is just the base hole current), this result can be written as

$${}_E I_B = \frac{{}_E Q'}{{}_E \tau_r} + \frac{d\,{}_E Q'}{dt} \tag{2.62}$$

This is the usual charge-control expression for the base current [7,9], except that it has been extended to nonlinear, nonuniform, and surface recombination.

We can now obtain an expression for the (time-varying) emitter current accounting for the increase in excess charge in the base, and thus extend Eq. 2.39, by using Eqs. 2.32 and 2.62. Thus

$${}_E I_E = {}_E I_B - {}_E I_C \tag{2.63}$$

$$= \frac{{}_E Q'}{{}_E \tau_r} + \frac{{}_E Q'_C}{{}_E t_C} + \frac{d\,{}_E Q'}{dt} \tag{2.64}$$

From Eqs. 2.39, 2.57 and 2.64 it is seen that (for DC conditions)

$$\frac{1}{{}_E\tau} = \frac{1}{{}_E\tau_r} + \frac{1 - {}_E\delta}{{}_Et_C} \qquad (2.65)$$

The numerator of the transit time term is explained by the fact that only a fraction $(1 - {}_E\delta)$ of the injected charge is to be associated with the transit time, whereas all of the injected charge is associated with the recombination lifetime and the emitter injection time, as discussed in Section 2.2.4. From Eq. 2.65 it is seen that ${}_E\tau$ is the parallel combination of the recombination and collection lifetimes. This is in accordance with intuition since the two mechanisms for removal of carriers from the base operate independently and simultaneously on all carriers. It is now seen that the emitter injection time ${}_E\tau$, as defined by Eq. 2.38, has the significance that in the time ${}_E\tau$ all of the injected carriers are removed from the base and replaced by new carriers (in the steady-state).

2.3 Conclusions

The basic charge-control parameters have been derived from the carrier transport and continuity equations in a clear and logical fashion. No restrictions have been made as to geometry or impurity distribution, and spatially nonuniform recombination (bulk and surface) has been included. Most of the detailed discussion is for low injection levels, but the concepts are also applicable to high injection levels; this extension will be made explicit in Chapter IV. Throughout the discussion, a one-to-one correspondence has been maintained between charge-control parameters and conceptually clear processes. In this way the original Beaufoy-Sparkes model, which is more mathematical than conceptual (as has been pointed out by Hamilton, et al. [16] and Koehler [18]), has been converted into a model which can be used to obtain an understanding of and insight into device operation for DC as well as low frequency AC and slow transient conditions.

The major obstacle to a clear conceptual interpretation of the charge-control parameters has lain in the meaning of the transit time in the presence of recombination, as has been pointed out by one of the originators of the charge-control model [36]. The separation of the injected charge into that which is collected and that which recombines removes this obstacle and allows a clear conceptual understanding of the transit time and carrier velocity to be obtained, with or without recombination. A conceptual interpretation of the Beaufoy-Sparkes [1] collector time constant, which heretofore was a strictly mathematical parameter, is also obtained. Two new parameters have been introduced: the recombination transit time, and the fraction of charge that recombines.

CHAPTER III

LIMITATIONS OF CHARGE-CONTROL CONCEPTS

In the previous chapter the basic charge-control parameters were defined and derived without an indication of the range of conditions over which they are meaningful. It is generally understood that the charge-control concepts are valid for sufficiently slow variations in the external boundary conditions [31], and that this frequency limitation is in some way connected with the carrier transit time across the base [3]. However, this connection is not always fully explained.

In this chapter some frequency limitations of the charge-control concepts will be presented. Some other limitations of usefulness or convenience will also be discussed which, while not fundamental, are of importance. One of these limitations will be overcome by a modification of the theory.

3.1 Continuity and Transport Equations

All derivations of charge-control relations [1,8,9,16,18], including those in the present work, assume the validity of the carrier transport and continuity equations. This is true also for the Ebers-Moll [20] circuit model and Linvill's [19] lumped model. Thus, all of these results are limited to conditions in which these equations are valid, as has been pointed out by McKelvey, et al. [37], McKelvey [38], and Hamilton, et al. [16].

If the transport and continuity equations are to be valid, then the concepts of drift and diffusion must be meaningful. This requires that the mean time to recombination (lifetime) be large compared to the mean time between collisions, and that the dimensions of the base region be large compared to a mean free path [16,37,38].

For most devices, these restrictions are almost always satisfied, so that this limitation will not be further considered.

3.2 First-Order Approximation

The charge-control relations obtained in Chapter II are valid within, and limited by, the usual first-order approximations made in conventional transistor analysis. In particular, it has been assumed that the device can be separated into fully depleted regions and quasi-neutral regions, with abrupt boundaries between them. It has been assumed that the density of injected minority carriers is zero at the collecting junction; that is, the collecting junction is an infinite sink for minority carriers. Also, carrier generation and recombination within the depletion layers and the emitter and collector regions has been neglected, as have any other phenomena which may occur outside the quasi-neutral base region (such as charge storage in the collector region).

3.3 Delay Time

In Chapter II, in the expressions for the emitter, base, and collector current, it was assumed that all of the currents could be expressed in terms of the same charge* $_EQ'$. This assumption is clearly violated for a time-varying signal. Account must be taken of the fact that the collector current at time t is not determined by the charge being injected into the base at time t, but rather by the charge that was injected approximately one transit time earlier. This is the delay

---

* This is essentially the assumption of instantaneous redistribution of charge that appears in the literature [5,16-18].

time which must be used in the boundary conditions in the solution of the continuity or diffusion equation. For sufficiently low frequencies and signal levels that the injected charge does not change greatly during one transit time, this effect can be neglected. It is thus required that

$$t_{delay} \ll T = \frac{2\pi}{\omega} \tag{3.1}$$

in order that the delay time should be negligible, where ω is the signal frequency, T is the signal period, and t_{delay} is the delay time, which is of the order of magnitude of the transit time.

For a one-dimensional, uniform-base transistor, Baker and May [9] have determined the delay time as

$$t_{delay} = \frac{{}_E t_C}{6} \tag{3.2}$$

However, instead of associating the delay time with the charge in transit, they consider it as a time-dependent transit time. Mathematically this is perfectly valid, since in an equation of the form

$${}_E I_C\ (s) = \frac{-{}_E Q'_C(s)}{{}_E t_C}\ F(s) \tag{3.3}$$

where s is the Laplace transform variable, it is arbitrary as to whether $F(s)$ is associated with ${}_E Q'_C(s)$ or with ${}_E t_C$. However, it is clear that association of $F(s)$ with ${}_E Q'_C(s)$ would more closely represent the process involved.

The delay time limitation can be circumvented by evaluation of the transport current as

$$I = \frac{Q}{\tau} \tag{3.4}$$

where τ is sufficiently small that the current and charge are approximately constant during the time τ. For $T < t_t$ (where t_t is the average transit time) a time interval can always be chosen (e.g. T/100) such that I is approximately constant during τ. If this is done, however, then τ has no direct conceptual significance and is a direct function of the applied signal. In this case charge-control no longer gives simple, understandable results: it is valid, but not useful. This approach will ultimately lead back to

$$I = \frac{dQ}{dt} \tag{3.5}$$

for sufficiently high frequencies and small τ.

3.4 Displacement Current

As usually formulated, the charge-control relations assume a sufficiently low signal frequency that the displacement current can be neglected. This is tantamount to the assumption that the period of the signal T is large compared to the relaxation time t_{relax}. Since the transit or injection time is usually larger than the relaxation time, the effects described in the previous section limit the validity of the results to a lower frequency, so that the displacement current can be neglected. This assumption will be made throughout this work. However, if

$$t_{relax} \geq t_t \qquad (3.6)$$

where t_t is the average transit time, then the displacement current must be accounted for.

3.5 Other Frequency Effects

So far, only phenomena which are fundamentally involved in the operation of a transistor have been considered. However, other effects must sometimes be taken into account. In particular, for sufficiently large applied biases and/or small base widths, the depletion layer widening effect on the base width may be significant [14]. This can introduce an additional time dependent term into the transit time and must therefore be accounted for. Its effect, through the change in base volume, on the injected charge and charge in transit must also be included, as must the time delay between a change in the external signal and the attendant depletion layer change in width.

In general, most second-order effects have some frequency dependence associated with them which must be considered in determination of the region of validity of the charge-control equations. These are generally not fundamental to transistor operation (but may be of practical importance) and can more or less easily be accounted for. These effects will not be considered.

3.6 Loss of Simplicity

The discussion of charge-control in Chapter II was mainly concerned with first-order effects encountered under low-injection conditions, although high-injection was explicitly allowed for in a few places. This was not intended to imply any basic limitation of the theory, but rather to simplify the discussion; in Chapter IV the theory will be extended to high-injection conditions and other phenomena. In this section it will merely be noted that, while the charge-control concepts and methods can be extended to some higher order phenomena (except possibly high frequency effects), such an extension is almost always accompanied by an attendant increase in complexity of the parameters and concepts. This, of course, is also true of other models. For high injection levels or base width variations with voltage, for instance, the transit time is no longer a constant dependent only on the device structure, but becomes a (possibly complex) function of applied bias, injection level, or time. If the restriction to low-injection Shockley-Read recombination is eliminated, then the lifetime also becomes a function of injection level. It is to be emphasized, however, that the charge-control principles are valid and useful under these conditions, although the low-injection simplicity is lost.

3.7 Differential Versus Integral Results

A limitation of the charge-control theory arises in switching applications. This is due to what Gray, et al. [31] refer to as charge storage in the remote regions of the base (see Figure 3.1). In transistor switch-off operation (i.e. removal of all injected charge from the base) the first charge to be removed is that which is physically between the emitter and collector junctions, in the active region. This is because the time for these carriers to reach the emitter junction is less than

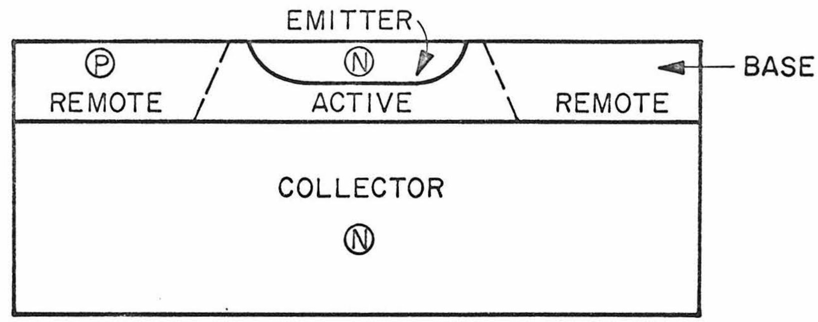

Fig. 3.1 Remote and active regions of the base in a planar transistor.

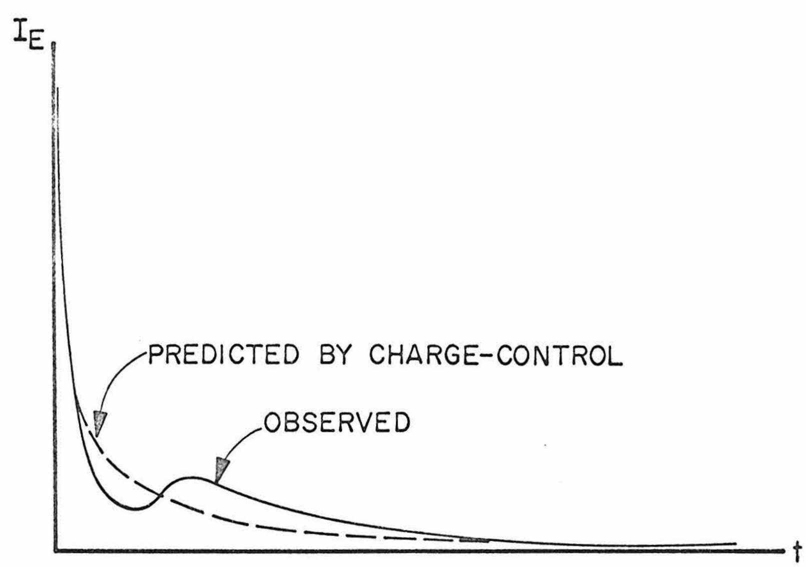

Fig. 3.2 Emitter current during switching from the saturation to the off condition, illustrating the "wiggle" effect. (After Gray, et al. [31])

that for the carriers in the remote regions. After the rapid removal of charge from the active region, the charge from the remote regions diffuses and drifts into the active region, causing an increase in current. Thus, the current waveform is similar to the solid curve shown in Figure 3.2. However, the theory as developed in Chapter II combines the active and remote region transit times into one average transit time, so that this "wiggle" effect is not predicted by the theory; only the average behavior is predicted (dashed curve in Figure 3.2).

The above result is inherent in the charge-control approach, in which only the gross, average, or integrated behavior of the device is represented. Integration of the transport and continuity equations results in the loss of detailed information, but the average behavior is correctly predicted*. This loss of detailed information may be important in certain applications, and thus constitutes a limitation on the range of applicability of the theory.

For the specific situation referred to above, of remote charge storage in the base, it is possible to modify the theory to overcome this limitation. Instead of the total injected charge and one transit time (for each injecting junction) being considered, the injected charge and transit time may be divided into "active" and "remote" components so that the (possibly) very different transit times in the two regions can be accounted for. By this means, the concepts of a lumped

---

* Hamilton, et al. [16] claim that the integrated equations represent an approximation, but this is incorrect; they represent an accurate average.

model [19,39] can be adapted to the concepts of the charge-control model, and a synthesis of the two thereby obtained.

3.8 Conclusions

Limitations to the validity of the charge-control concepts have been discussed. The most fundamental restriction is the assumption of the validity of the transport and continuity equations; without these equations, all else is meaningless, even for DC. Also of great importance is that the theory is only valid within the framework of the first-order approximations of conventional transistor analysis. The most important frequency limitation is that imposed by transit (or delay) time, although relaxation time effects may be important, especially for switching or transient analysis. Second-order frequency effects must also be considered.

Other limitations of the theory are related more to convenience or utility than to validity. In the attempt to include high injection levels and second-order effects, the parameters and concepts lose their simplicity, and the entire theory may lose its attractiveness, although not its validity. This, however, is also true of other models.

By integration of the basic equations and consideration of only average behavior, detailed information is lost, as in the case of remote charge storage in the base. This constitutes only a loss of information, not a loss of accuracy; the average behavior is accurately described. By combination of some of the concepts of the lumped model with the charge-control model, this limitation can be overcome at the cost of added complexity.

CHAPTER IV

APPLICATIONS AND EXTENSIONS OF THE CHARGE-CONTROL THEORY

In this chapter it will be shown how the basic concepts of charge-control, as developed in Chapter II, can be used to determine device capabilities and performance. The theory will also be extended to high injection levels and second-order phenomena, and a relation between the injected charge and the applied voltage will be obtained.

4.1 Common-Emitter Current Gain

In this section an expression for the DC common-emitter current gain β will be obtained, under the assumption that majority carrier injection from the base into the emitter can be neglected (unity emitter efficiency). Then

$$\beta \equiv \left| \frac{{}_E I_C}{{}_E I_B} \right| \tag{4.1}$$

Using Eqs. 2.32 and 2.62 in Eq. 4.1 we obtain

$$\beta = \frac{{}_E Q'_C / {}_E t_C}{{}_E Q' / {}_E \tau_r} \tag{4.2}$$

Equation 2.57 for ${}_E Q'_C$ yields

$$\beta = \frac{(1 - {}_E \delta)\, {}_E \tau_r}{{}_E t_C} \tag{4.3}$$

$$\beta = \frac{{}_E \tau_r}{{}_E \tau_C} \tag{4.4}$$

where Eq. 2.58 for ${}_E\tau_C$ has been used. Gray, et al. [31] have obtained the same result under the restriction of a one-dimensional, uniform-base transistor, without giving a clear interpretation of the parameters. The result obtained here is valid for an arbitrary geometry and impurity distribution.

If ${}_E\tau_C$ is considered as a collection lifetime, Eq. 4.4 indicates that β is the ratio of recombination to collection lifetimes. If the fraction of recombined carriers ${}_E\delta$ is sufficiently small, the current gain approaches the ratio of recombination lifetime to collection transit time.

4.2 High Injection Levels

In this section the results of Chapter II will be extended to the case of high injection levels. It will also be shown that the usual high-injection equations can easily be derived from charge-control principles.

For high-injection, the electron current density can be written as (from Section 2.1.1)

$$j_n = en'v_{nDo} + env'_{nD} + en'v_{nd} \tag{4.5}$$

where

$$v_{nDo} \equiv -(\mu_n \underset{\sim}{E}_o) \cdot \hat{x} \tag{4.6}$$

$$v'_{nD} \equiv -(\mu_n \underset{\sim}{E}') \cdot \hat{x} \tag{4.7}$$

$$v_{nd} \equiv -(D_n \frac{\underset{\sim}{\nabla} n'}{n'}) \cdot \hat{x} \tag{4.8}$$

and $\hat{x}$ is a unit vector in the direction of current flow, so that Eqs. 4.6 - 4.8 represent the component of the velocities along a flow tube.

For the case of no recombination* Eq. 4.5 can be integrated to obtain

$$I_n = \frac{Q_{no} + Q'_n}{t'_{nD}} + Q'_n \left(\frac{1}{t_{nDo}} + \frac{1}{t_{nd}}\right) \tag{4.9}$$

where Q_{no} and Q'_n are the equilibrium and excess electron charge, respectively, and

$$\frac{1}{t_{nDo}} \equiv \frac{\iint_S en'v_{nDo}dS}{Q'_n} \tag{4.10}$$

$$\frac{1}{t_{nd}} \equiv \frac{\iint_S en'v_{nd}dS}{Q'_n} \tag{4.11}$$

$$\frac{1}{t'_{nD}} \equiv \frac{\iint_S env'_{nD}dS}{Q_n} \tag{4.12}$$

$$Q_n = Q_{no} + Q'_n \tag{4.13}$$

---

* The methods of Section 2.2.2 can be used in the presence of recombination.

Since recombination has been neglected, the integrations can be carried out over any surface that crosses all of the flow tubes in the base region. It is clear that Eq. 4.9 could have been written down immediately (without starting with Eq. 4.5) from a knowledge of the various forces and pseudo-forces (diffusion, equilibrium drift, excess drift) acting on the equilibrium and excess carriers, and the charge-control principles.

A comparison of Eqs. 4.10 - 4.12 with Eq. 2.36 indicates that t_{nDo} and t_{nd} are the transit times due to the built-in field and diffusion, respectively, while t'_{nD} can be considered as the transit time due to the injection-caused field. This interpretation of t'_{nD} leads to the conclusion that Q_{no}/t'_{nD} represents a shift in the distribution of equilibrium carriers. This is not to be taken literally since all electrons are indistinguishable, but it may be a useful concept.

For low injection levels, the injection-caused field (or velocity) is negligible, so that

$$I_n = Q'_n \left(\frac{1}{t_{nDo}} + \frac{1}{t_{nd}}\right) \tag{4.14}$$

For high-injection $Q'_n \gg Q_{no}$, leading to

$$I_n = Q'_n \left(\frac{1}{t_{nDo}} + \frac{1}{t_{nd}} + \frac{1}{t'_{nD}}\right) \tag{4.15}$$

Thus it is seen that for both low and high injection levels, the equilibrium minority charge contributes a negligible amount to the total current. For medium injection levels, where the above approximations are not valid, this term cannot be neglected.

The hole current can be similarly expressed as

$$I_p = \frac{Q_p}{t'_{pD}} + Q'_p\left(\frac{1}{t_{pDo}} - \frac{1}{t_{pd}}\right) \tag{4.16}$$

where the terms are analogous to those of Eq. 4.9.

An expression for the electron current (assumed the minority carrier current) that does not contain the transit time due to injected carriers (or the injection-caused field) will now be obtained.

First, it is noted that since the hole and electron drift velocities differ only by the mobility ratio, the transit times of individual holes and electrons will differ only by the mobility ratio. Therefore the average drift transit times must also differ only by the mobility ratio. Thus

$$\frac{t_{pDo}}{t_{nDo}} = \frac{t'_{pD}}{t'_{nD}} = \frac{\mu_n}{\mu_p} \tag{4.17}$$

under all conditions.

Solution of Eq. 4.16 for t'_{pD} and use of Eq. 4.17 leads to

$$\frac{1}{t'_{pD}} = \frac{\mu_p}{\mu_n}\frac{1}{t'_{nD}} = \frac{I_p}{Q_p} + \frac{Q'_p}{Q_p}\left(\frac{1}{t_{pd}} - \frac{\mu_p}{\mu_n}\frac{1}{t_{nDo}}\right) \tag{4.18}$$

Substitution of Eq. 4.18 for t'_{nD} into Eq. 4.9 for the electron current and rearrangement, yields

$$I_n = \frac{Q'_n}{t_{nd}}\left(1 + \frac{t_{nd}}{t_{pd}} \frac{\mu_n}{\mu_p} \frac{Q'_p}{Q'_n} \frac{Q_n}{Q_p}\right) + \frac{Q'_n}{t_{nDo}}\left(1 - \frac{Q'_p}{Q'_n} \frac{Q_n}{Q_p}\right) + \frac{Q_n}{Q_p} \frac{\mu_n}{\mu_p} I_p \quad (4.19)$$

Equation 4.19 is essentially the integrated form of Eq. A.4 (appendix) generalized to arbitrary geometry and a three-dimensional impurity distribution.

The assumption of quasi-neutrality results in

$$n' = p' \quad (4.20)$$

$$Q'_n = Q'_p \quad (4.21)$$

Under these conditions, the diffusion velocities (Eq. 4.8) also will differ only by the mobility ratio, so that

$$\frac{t_{pd}}{t_{nd}} = \frac{\mu_n}{\mu_p} \quad (4.22)$$

With substitution of Eqs. 4.21 and 4.22 into Eq. 4.19, the electron current is obtained as

$$I_n = \frac{Q'_n}{t_{nd}}\left(1 + \frac{Q_n}{Q_p}\right) + \frac{Q'_n}{t_{nDo}}\left(1 - \frac{Q_n}{Q_p}\right) + \frac{\mu_n}{\mu_p} \frac{Q_n}{Q_p} I_p \quad (4.23)$$

which is the integrated, three-dimensional, generalized form of Eq. A.7.

For high injection levels $n \approx p$, so that

$$Q_n \approx Q_p \tag{4.24}$$

and Eq. 4.23 becomes (compare with Eq. A.9)

$$I_n = \frac{2Q'_n}{t_{nd}} + \frac{\mu_n}{\mu_p} I_p \tag{4.25}$$

which shows the usual multiplicative factor of 2 as compared with the low-injection result [40,41] (see the appendix). Since the conventional interpretation is that high-injection causes the diffusion constant (and hence the carrier diffusion velocity) to double, it is seen that the diffusion transit time can be considered as halved.

Equation 4.25 (together with the results of Section 4.6 below) is the usual first-order high-injection diode or transistor solution, except that it is more general. The solution presented here is valid for any geometry or impurity distribution, whereas the usual results are strictly valid only in one dimension with uniform doping, although they are usually assumed valid in three dimensions. The generalization to three dimensions has been obtained at no cost of added complexity as compared to the conventional solution, while the generalization to non-uniform doping is equally simple by either method.

When neutrality breaks down at the diode contact [42] or the collector junction [28,29] for very high injection levels, then for

Eqs. 4.20 - 4.25 to be valid, the Q's must represent the charge in the quasi-neutral region and the t's must refer to the transit times across the quasi-neutral region.

The above derivation is for a diode or transistor with no recombination. However, recombination can be included, and the device treated as either a transistor or diode according to whether the recombination current is considered as a separate (base) current or is combined with the collected current [21,26].

4.3 Nonlinear Recombination Rates

In Chapter II only Shockley-Read [30] recombination mechanisms were considered; however, regardless of the type of recombination, a lifetime can always be defined as

$$\tau_{nr} \equiv \frac{n'}{U_n} \tag{4.26}$$

Further, an average recombination lifetime can always be defined as in Section 2.2.3 (for injection from surface μ)

$$_\mu\tau_{nr} \equiv \frac{_\mu Q'_n}{\iiint\limits_{Base} U_n d(vol)} \tag{4.27}$$

This average lifetime can be used throughout the charge-control equations, but it is only in the case of a linear recombination rate ($U_n \propto n'$) that $_\mu\tau_{nr}$ is independent of the injected density. For high injection levels and nonlinear recombination rates, $_\mu\tau_{nr}$ is a function of the injected charge and the expressions for recombination current become correspondingly complex.

4.4 Early Effect

So far in the analysis, the base volume and the length of a flow tube have been assumed to be independent of the applied bias. Unfortunately for the simplicity of the model, the depletion layer width varies with the voltage across it and hence affects the base width, as Early pointed out in 1952 [14]. As Gray, et al. [31] have noted, this variation of base layer width affects both the transit times and the total charges.

One way of accounting for these effects is to introduce a voltage (or time) dependent base volume (or width) into the integral expressions for the various characteristic times (transit time, average lifetime, injection time) and charges (injected charge, recombined charge, charge in transit). However, this complexity is not always necessary.

The flow tube length can always be represented as

$$W = W_o + W_1 \tag{4.28}$$

where W_o is the initial value and W_1 is the change (positive or negative) due to the depletion layer variation. The transit time per flow tube may also be written

$${}_e t_c = {}_e t_{co} + {}_e t_{c1} \tag{4.29}$$

where ${}_e t_{co}$ is the initial value and ${}_e t_{c1}$ is the change (positive or negative) due to the depletion layer variation. If W_1/W_o is sufficiently small that the ensemble velocity can be considered unaffected, then ${}_e t_{cl}$ is merely the transit time across the distance W_1. In general, however, the change in base width will affect (through the diffusion equation) the injected density gradient and injected density in such a way as to change the ensemble velocity. In this case, ${}_e t_{cl}$ will include both effects. Similarly, the average transit time may be written as

$$ {}_E t_C = {}_E t_{Co} + {}_E t_{Cl} \tag{4.30} $$

As expressed above, W_1 and ${}_E t_{Cl}$ include the effects of both the emitter and collector depletion layer variations; however, these effects can be separated for low injection levels.

The average lifetime can be written as

$$ \frac{1}{{}_E \tau_{nr}} = \frac{1}{{}_E \tau_{nro}} + \frac{1}{{}_E \tau_{nrl}} \tag{4.31} $$

where ${}_E \tau_{nrl}$ is to be attributed to the volume added or removed from the base by the depletion layer variation. It is to be noted that ${}_E \tau_{nrl}$, while conceptually clear, may be mathematically quite complex if the recombination rate is nonuniform. However, for a linear and spatially uniform recombination rate the lifetime is unaffected, so that

$$ {}_E \tau_{nr} = {}_E \tau_{nro} = \tau_{nr} \tag{4.32} $$

The charges can also be separated into components (following Gray, et al. [31]) as

$$Q = Q_o + Q_1 \tag{4.33}$$

where Q_o is the initial charge (in transit, injected or recombined) and Q_1 is the change in charge (positive or negative) due to the base width variation.

Care must be taken in using Eqs. 4.30 - 4.33 that linearity is not assumed. The change in base width W_1 is certainly nonlinear in voltage, except for sufficiently small fluctuations, while Q_1 and ${}_Et_{C1}$ may never be linear in voltage (or e^V) or in injected density. The quantities Q_1 and ${}_Et_{C1}$ are composed of two effects. The first is a change in base volume which enters directly as a change in the region of integration. The second effect is less direct and results from the fact that the base width affects, through the diffusion equation, the shape of the excess carrier density distribution. However, this separation of times and charges into components is justified on the basis that the components have clear conceptual significance and hence this separation can enhance understanding of the effect of base width variation on device behavior, regardless of the mathematical form of the equations.

4.5 Additional Transport and Injection Processes

In Chapter II, attention was focussed on electron injection from the emitter to the base of an NPN transistor with an ideal base contact. In this section those results will be extended to both hole and electron injection and collection by the emitter and collector junctions and the base contact.

Regardless of the injection level, the currents can always be written* as (with neglect of the displacement current)

$$I_E = {}_E I_{nE} + {}_E I_{pE} + {}_C I_{nE} + {}_C I_{pE} + {}_B I_{nE} + {}_B I_{pE} \tag{4.34}$$

$$I_C = {}_C I_{nC} + {}_C I_{pC} + {}_E I_{nC} + {}_E I_{pC} + {}_B I_{nC} + {}_B I_{pC} \tag{4.35}$$

$$I_B = {}_B I_{nB} + {}_B I_{pB} + {}_E I_{nB} + {}_E I_{pB} + {}_C I_{nB} + {}_C I_{pB} \tag{4.36}$$

where I_ν is the total current through surface ν and ${}_\mu I_\nu$ is the current through surface ν due to injection at surface μ**. The subscripts n, p have their usual significance.

The charge-control principle, that a transport current is simply the charge divided by the relevant time, results in

$$ {}_E I_{nE} = \frac{{}_E Q'_n}{{}_E \tau_n} = \frac{{}_E Q'_n}{{}_E \tau_{nr}} + \frac{{}_E Q'_{nB}}{{}_E t_{nB}} + \frac{{}_E Q'_{nC}}{{}_E t_{nC}} \tag{4.37}$$

$$ {}_C I_{nE} = - \frac{{}_C Q'_{nE}}{{}_C t_{nE}} \tag{4.38}$$

---

* For simplicity dQ/dt terms are neglected. These must be inserted where appropriate for each injection mechanism for AC analysis.

** Injection at surface μ may, of course, be affected by the voltage at surface ν or λ.

$$ {}_{B}I_{nE} = - \frac{{}_{B}Q'_{nE}}{{}_{B}t_{nE}} \tag{4.39} $$

and similarly for all the other components, where ${}_{\mu}Q'_{\nu}$ is the charge injected at surface μ which reaches surface ν and ${}_{\mu}t_{\nu}$ is the transit time from surface μ to surface ν. Thus the total emitter current can be expressed as

$$ I_E = \frac{{}_{E}Q'_{n}}{{}_{E}\tau_{nr}} - \frac{{}_{E}Q'_{p}}{{}_{E}\tau_{pr}} + \frac{{}_{E}Q'_{nB}}{{}_{E}t_{nB}} - \frac{{}_{E}Q'_{pB}}{{}_{E}t_{pB}} + \frac{{}_{E}Q'_{nC}}{{}_{E}t_{nC}} - \frac{{}_{E}Q'_{pC}}{{}_{E}t_{pC}} $$

$$ - \frac{{}_{C}Q'_{nE}}{{}_{C}t_{nE}} + \frac{{}_{C}Q'_{pE}}{{}_{C}t_{pE}} - \frac{{}_{B}Q'_{nE}}{{}_{B}t_{nE}} + \frac{{}_{B}Q'_{pE}}{{}_{B}t_{pE}} \tag{4.40} $$

Additional terms can, of course, be added to account for excess (non-equilibrium) carrier generation and devices with more or fewer contacts. The base and collector currents can be expressed by equations analogous to Eq. 4.40.

It is important to emphasize that the separation into components carried out above requires absolutely no assumptions as to injection level or linearity. Superposition has not been used. We have merely used the obvious fact that, for instance, part of the emitter current is contributed by electrons injected by the collector junction that are received by the emitter junction. Equations 4.34 - 4.40 do not preclude the possibility that the density (or total number) of these electrons is partially or wholly determined by the emitter junction voltage, conditions at the base contact, or by anything else.

We have merely enumerated the possible origins and destinies of carriers in the base region. Only for low injection levels may the assumption be made that the τ's and t's are constants and that the charges depend only on conditions at the injecting surface. For high injection levels, the Q's, τ's, and t's may be so interrelated as to render the equations mathematically useless. However, they will always remain conceptually useful.

The utility of this separation into components can be seen in the calculation and understanding of the DC common-emitter current gain. The usual first-order approximation was given in the first section of this chapter. That result corresponds to setting

$$ {}_{E}I_{C} = {}_{E}I_{nC} = \frac{-{}_{E}Q'_{nC}}{{}_{E}t_{nC}} \tag{4.41} $$

and

$$ {}_{E}I_{B} = {}_{E}I_{pB} = \frac{{}_{E}Q'_{n}}{{}_{E}\tau_{nr}} \tag{4.42} $$

Examination of Eqs. 4.35 and 4.36 (rewritten similarly to Eq. 4.40 if necessary) will readily show which current flow mechanisms have been neglected. Further study of these mechanisms can lead to quantitative limitations of the first-order result. A less detailed analysis can lead to qualitative results from which we can determine which mechanism is of secondary, tertiary, etc. importance. Expressions like Eq. 4.40 allow us to enumerate all factors entering into the current gain (or other parameter) and to study them systematically, severally or individually according to whether they are interrelated or not. By including the dQ/dt terms in Eqs. 4.34 - 4.36, this discussion can be extended to slowly varying signals.

4.6 Charge-Voltage Relations

Charge-control, as introduced by Beaufoy and Sparkes [1,2] and extended by others, is restricted to current-charge relationships. However, relations between the injected charge and the applied voltages are also of interest, and will be derived in this section. Shockley, Sparks, and Teal [43] show that, for unity injection efficiency

$$n(E,y,z) = n_P(E,y,z)e^{V_{jE}} \tag{4.43}$$

where $n_P(E,y,z)$ is the equilibrium electron concentration in the base at the emitter junction and V_{jE} is the portion of the applied voltage which appears across the emitter junction (normalized to the thermal potential $V_t = kT/e$).

Van Vliet [44], among others, has obtained relations analogous to Eq. 4.43 which are valid for high injection levels and nonunity injection efficiency. He obtains

$$n(E,y,z) = n_P(E,y,z)[b_E(V_{jE}) + 1] \tag{4.44}$$

where

$$b_E(V_{jE}) + 1 \equiv \frac{n_P(E,y,z)p_N[p_P(E,y,z) - n_P(E,y,z)]e^{2V_{jE}} + n_I^2(n_N-p_N)e^{V_{jE}}}{n_N n_I^2 - [n_P(E,y,z)]^2 p_N e^{2V_{jE}}} \tag{4.45}$$

so that

$$_{E}n'(E,y,z) = n_P(E,y,z)b_E(V_{jE}) \tag{4.46}$$

where n_N and p_N are the equilibrium carrier densities in the emitter at the depletion layer edge. This relation assumes that a quasi-equilibrium condition, obeying the Boltzmann relation, is instantaneously established among all carriers (including injected carriers). For low-injection conditions we have

$$V_{jE} = V_{EB} \tag{4.47}$$

where V_{EB} is the externally applied emitter-to-base voltage (normalized to V_t). Then Eq. 4.43 or 4.44 becomes

$$n(E,y,z) = n_P(E,y,z)e^{V_{EB}} \tag{4.48}$$

and Eq. 4.46 becomes

$$_{E}n'(E,y,z) = n_P(E,y,z)(e^{V_{EB}}-1) \tag{4.49}$$

Equation 4.46 represents the emitter boundary condition to be used in solving the diffusion equation. At the collector, the boundary condition is

$$_{E}n'(C,y,z) = 0 \tag{4.50}$$

Using Eqs. 4.46 and 4.50 the emitter-injected electron density may be written as

$$ {}_{E}n'(x,y,z) = n_P(E,y,z)\; b_E(V_{jE})\; {}_{E}f(x,y,z) \tag{4.51} $$

where it is required that

$$ {}_{E}f(E,y,z) \equiv 1 \tag{4.52} $$

$$ {}_{E}f(C,y,z) \equiv 0 \tag{4.53} $$

This is the result of Ebers and Moll [20] generalized to arbitrary injection levels. For low-injection and negligible Early effect, ${}_{E}f$ is independent of the injection level.

To obtain the total injected charge, Eq. 4.51 is multiplied by the electronic charge and integrated over the base volume, so that

$$ {}_{E}Q' = \iiint_{\text{Base}} e\; n_P(E,y,z)\; {}_{E}f(x,y,z)\; b_E(V_{jE})\; d(\text{vol}) \tag{4.54} $$

Similarly

$$ {}_{C}Q' = \iiint_{\text{Base}} e\; n_P(C,y,z)\; {}_{C}f(x,y,z)\; b_C(V_{jC})\; d(\text{vol}) \tag{4.55} $$

where

$$ {}_{C}f(E,y,z) \equiv 0 \tag{4.56} $$

$${}_{C}f(C,y,z) \equiv 1 \tag{4.57}$$

and b_C is analogous to b_E.

The integral of Eq. 4.54 is clearly a function only of the applied voltages V_{EB} and V_{CB}. Thus, Eq. 4.54 may be written as

$${}_{E}Q'(V_{EB}, V_{CB}) = {}_{E}K\, B_E(V_{EB}, V_{CB}) \tag{4.58}$$

where ${}_{E}K$ is a constant, and $B_E(V_{EB}, V_{CB})$ is a function which depends on the applied voltages. If the device is such that the equilibrium concentrations and V_{jE} do not vary along the depletion layer edges, or the injection level is sufficiently low, b_E will be spatially constant, so that (with the neglect of depletion layer width variation with voltage)

$$B_E = b_E \tag{4.59}$$

and

$${}_{E}K = \iiint_{\text{Base}} e\, n_P(E,y,z)\ {}_{E}f(x,y,z)\ d(\text{vol}) \tag{4.60}$$

For low injection levels, regardless of equilibrium carrier density variation with position along the depletion layer edge, Eq. 4.48 is valid, so that

$$b_E = e^{V_{EB}} - 1 \tag{4.61}$$

If any dependence of ${}_{E}f$ on voltage (i.e. because of the Early effect) is neglected, Eq. 4.59 is valid and

$$B_E = e^{V_{EB}} - 1 \tag{4.62}$$

so that

$${}_{E}Q' = {}_{E}K(e^{V_{EB}} - 1) \tag{4.63}$$

For high injection levels, if the equilibrium carrier density is constant along the depletion layer edge, first-order theory results in [40, 41]

$$B_E = b_E \approx e^{V_{EB}/2} \tag{4.64}$$

so that the total injected charge is

$${}_{E}Q' \approx {}_{E}K\, e^{V_{EB}/2} \tag{4.65}$$

and ${}_{E}K$ is given by Eq. 4.60.

Thus, for low injection levels, a simple relation between applied voltage and injected charge can be obtained regardless of the equilibrium carrier distribution. For high injection levels the same formal relation exists, but it is more complex.

4.7 Conclusions

The charge-control principles have been used to derive the DC common-emitter current gain under the assumptions of low injection level and unity injection efficiency. It has been shown that the usual high-injection solution of a diode or transistor can be obtained as easily for three dimensions by charge-control methods as for one dimension in the conventional manner. The results have been further generalized to arbitrary three-dimensional impurity distributions; this can be done as easily for one dimension by the conventional approach.

The concepts of Chapter II were extended to include nonlinear recombination mechanisms and the results of that chapter were extended to cover the Early effect.

Allowance was made for majority carrier currents, collector injection and an injecting and collecting base contact. It was demonstrated that all of these effects are easy to account for by charge-control principles; however, for high injection levels these results are more useful as an aid to insight and understanding than for mathematical manipulations. The charge-control model was extended to include charge-voltage relationships at all injection levels; however, these relations are more useful for low than for high injection levels.

CHAPTER V

THE EBERS-MOLL CIRCUIT RELATIONS

In 1954 Ebers and Moll [20] derived expressions for the emitter and collector currents in terms of the junction voltages. Their results are valid under the following assumptions:

1) Low injection levels, so that the emitter efficiency is constant and all of the externally applied voltage appears across the junctions. (This permits the use of superposition.)

2) Second-order effects (specifically the Early [14] effect) were neglected.

3) Equation 4.48 for the minority carrier density at the junction is valid.

The Ebers-Moll equations can be written as

$$I_E = a_{11}(e^{V_{EB}}-1) + a_{12}(e^{V_{CB}}-1) \tag{5.1}$$

$$I_C = a_{21}(e^{V_{EB}}-1) + a_{22}(e^{V_{CB}}-1) \tag{5.2}$$

$$I_B = (a_{11} + a_{21})(e^{V_{EB}}-1) + (a_{12} + a_{22})(e^{V_{CB}}-1) \tag{5.3}$$

Shockley, Sparks, and Teal [43] obtained these results earlier.

By employing various circuit operations (i.e. open or short circuiting various terminals) Ebers and Moll related the a_{ij} to open-circuit saturation currents and common-base current gains, to obtain

$$I_E = \frac{I_{EBO}}{1-\alpha_N\alpha_I} (e^{V_{EB}}-1) - \frac{\alpha_I I_{CBO}}{1-\alpha_N\alpha_I} (e^{V_{CB}}-1) \tag{5.4}$$

$$I_C = \frac{-\alpha_N I_{EBO}}{1-\alpha_N\alpha_I} (e^{V_{EB}}-1) + \frac{I_{CBO}}{1-\alpha_N\alpha_I} (e^{V_{CB}}-1) \tag{5.5}$$

$$I_B = \frac{(1-\alpha_N) I_{EBO}}{1-\alpha_N\alpha_I} (e^{V_{EB}}-1) + \frac{(1-\alpha_I) I_{CBO}}{1-\alpha_N\alpha_I} (e^{V_{CB}}-1) \tag{5.6}$$

where α_N, α_I are the normal and inverted common-base current gains

$$\alpha_N \equiv \left| \frac{I_C}{I_E} \right|_{V_{CB} = 0} \tag{5.7}$$

$$\alpha_I \equiv \left| \frac{I_E}{I_C} \right|_{V_{EB} = 0} \tag{5.8}$$

and $I_{\mu\nu O}$ is the reverse saturation current through terminal μ when the third terminal is left open circuited.

In terms of the normal and inverted common-emitter current gains

$$\beta_N \equiv \left| \frac{I_C}{I_B} \right|_{V_{CB} = 0} = \frac{\alpha_N}{1-\alpha_N} \tag{5.9}$$

$$\beta_I \equiv \left| \frac{I_E}{I_B} \right|_{V_{EB} = 0} = \frac{\alpha_I}{1-\alpha_I} \tag{5.10}$$

Eqs. 5.4 - 5.6 can be written as

$$I_E = \frac{(1 + \beta_N)(1 + \beta_I)I_{EBO}}{1 + \beta_N + \beta_I} (e^{V_{EB}}-1) - \frac{\beta_I(1 + \beta_N)I_{CBO}}{1 + \beta_N + \beta_I}(e^{V_{CB}}-1) \tag{5.11}$$

$$I_C = \frac{- \beta_N(1 + \beta_I)I_{EBO}}{1 + \beta_N + \beta_I} (e^{V_{EB}}-1) + \frac{(1 + \beta_N)(1 + \beta_I)I_{CBO}}{1 + \beta_N + \beta_I}(e^{V_{CB}}-1) \tag{5.12}$$

$$I_B = \frac{(1 + \beta_I)I_{EBO}}{1 + \beta_N + \beta_I} (e^{V_{EB}}-1) + \frac{(1 + \beta_N)I_{CBO}}{1 + \beta_N + \beta_I}(e^{V_{CB}}-1) \tag{5.13}$$

As has been pointed out by Narud, Hamilton, and Lindholm [15,16], and by Koehler [17,18], Eqs. 5.4 - 5.6 are predominantly electrical equations. That is, they are more oriented toward circuit operations than toward a conceptual understanding of internal processes. In this chapter these relations will be given a clear conceptual interpretation by deriving them from charge-control principles. This will lead to the conclusion that short-circuit, rather than open-circuit, saturation currents would have more conceptual significance in these equations. It will also be shown how, using charge-control, Eqs. 5.1 - 5.3 can be extended to high injection levels. Finally, Eqs. 5.1 - 5.3 will be inverted to obtain the voltages in terms of the external currents and short-circuit current parameters.

5.1 Derivation From Charge-Control Principles

All of the material needed for this section has already been developed in the previous chapters. Here, the previous results will be restated in the appropriate context. In conformity with convention, minority carrier flow through the base contact will be neglected. For conciseness, only minority carrier current will be explicitly considered; majority carrier current can readily be added. Under these conditions, Eqs. 4.34 - 4.36 become*

$$I_E = {}_E I_E + {}_C I_E \tag{5.14}$$

$$I_C = {}_E I_C + {}_C I_C \tag{5.15}$$

$$I_B = {}_E I_B + {}_C I_B \tag{5.16}$$

Then, use of the appropriate portions of Eq. 4.40 for the emitter current and similar results for the collector and base currents results in

$$I_E = \frac{{}_E Q'}{{}_E \tau_r} + \frac{{}_E Q'_C}{{}_E t_C} - \frac{{}_C Q'_E}{{}_C t_E} \tag{5.17}$$

$$I_C = \frac{-{}_E Q'_C}{{}_E t_C} + \frac{{}_C Q'}{{}_C \tau_r} + \frac{{}_C Q'_E}{{}_C t_E} \tag{5.18}$$

$$I_B = \frac{{}_E Q'}{{}_E \tau_r} + \frac{{}_C Q'}{{}_C \tau_r} \tag{5.19}$$

---

* The base current can be considered as a majority carrier (hole) current due to majority carrier injection at the base contact (${}_B I_{pB}$) or as a minority carrier (electron) current due to minority carrier injection at the junctions (${}_E I_{nB}$ + ${}_C I_{nB}$). Since only minority carriers are being considered here, the latter description will be employed.

Since the injected charge is more easily related to the external voltages than is the collected charge, the explicit reference to the charge in transit may be eliminated, so that

$$I_E = {}_EQ'\left(\frac{1}{{}_E\tau_r} + \frac{1-{}_E\delta}{{}_Et_C}\right) - \frac{{}_CQ'(1-{}_C\delta)}{{}_Ct_E} \tag{5.20}$$

$$I_C = \frac{-{}_EQ'(1-{}_C\delta)}{{}_Et_C} + {}_CQ'\left(\frac{1}{{}_C\tau_r} + \frac{1-{}_C\delta}{{}_Ct_E}\right) \tag{5.21}$$

$$I_B = \frac{{}_EQ'}{{}_E\tau_r} + \frac{{}_CQ'}{{}_C\tau_r} \tag{5.22}$$

where ${}_E\delta$, ${}_C\delta$ are the fractions of injected charge that recombine, as discussed in Chapter II. Using the results of Section 4.6 (Eq. 4.58) we can write Eqs. 5.20 - 5.22 as

$$I_E = B_E(V_{EB}, V_{CB})\ {}_EK\left(\frac{1}{{}_E\tau_r} + \frac{1-{}_E\delta}{{}_Et_C}\right) - B_C(V_{EB}, V_{CB})\ \frac{{}_CK(1-{}_C\delta)}{{}_Ct_E} \tag{5.23}$$

$$I_C = -\ B_E(V_{EB}, V_{CB})\ \frac{{}_EK(1-{}_E\delta)}{{}_Et_C} + B_C(V_{EB}, V_{CB})\,{}_CK\left(\frac{1}{{}_C\tau_r} + \frac{1-{}_C\delta}{{}_Ct_E}\right) \tag{5.24}$$

$$I_B = B_E(V_{EB}, V_{CB})\ \frac{{}_EK}{{}_E\tau_r} + B_C(V_{EB}, V_{CB})\ \frac{{}_CK}{{}_C\tau_r} \tag{5.25}$$

In obtaining Eqs. 5.23 - 5.25, no restrictions as to injection level or second-order effects have been made, thus they can be considered as a generalized form of the Ebers-Moll equations. However, the Ebers-Moll restrictions will now be imposed so that their results may be obtained. Substitution of Eq. 4.62 for B_E and a similar relation for B_C into Eqs. 5.23 - 5.25 results in

$$I_E = \frac{{}_EK(e^{V_{EB}}-1)}{{}_E\tau} - \frac{{}_CK(1-{}_C\delta)(e^{V_{CB}}-1)}{{}_Ct_E} \tag{5.26}$$

$$I_C = \frac{-{}_EK(1-{}_E\delta)(e^{V_{EB}}-1)}{{}_Et_C} + \frac{{}_CK(e^{V_{CB}}-1)}{{}_C\tau} \tag{5.27}$$

$$I_B = \frac{{}_EK(e^{V_{EB}}-1)}{{}_E\tau_r} + \frac{{}_CK(e^{V_{CB}}-1)}{{}_C\tau_r} \tag{5.28}$$

where Eq. 2.65 has been used for the emitter injection time (and a similar relation for the collector injection time).

Equations 5.26 - 5.28 are the Ebers - Moll equations (Eqs. 5.4 - 5.6) except that the coefficients of the voltage terms are expressed differently. They have been derived entirely from the charge-control theory as developed in the previous chapters. It is clear that the basis for charge-control, and hence for the above derivation, is exactly the same as for the derivation given by Ebers and Moll, so that it is not surprising that the results are identical for identical restrictions. The significance of the derivation given here is that the results are given in terms of conceptually, rather than electrically, significant parameters. (This may not be an advantage for the designer of discrete circuits, but it may be important for the design and understanding of integrated circuits.) Thus these two derivations provide a means of relating circuit and conceptual parameters. From Eqs. 5.23 - 5.25 it can be seen how these relations can be extended to high injection levels and second-order effects.

5.2 Conceptual Interpretation of the Ebers-Moll Parameters

Now that the Ebers-Moll equations have been derived from two conceptually different bases, the different parameters will be related to each other. This will be done for both open and short-circuit saturation currents and it will be shown that the short-circuit saturation currents are a more useful indicator of device behavior.

From Section 4.1 the current gains are

$$\beta_N \equiv \left| \frac{E^{I}C}{E^{I}B} \right| = \frac{E^{\tau}r}{E^{\tau}C} \tag{5.29}$$

$$\beta_I \equiv \left| \frac{C^{I}E}{C^{I}B} \right| = \frac{C^{\tau}r}{C^{\tau}E} \tag{5.30}$$

Equations 5.29 and 5.30 can be used to determine the normal and inverted common-base current gains

$$\frac{1}{\alpha_N} = 1 + \frac{1}{\beta_N} = 1 + \frac{E^{\tau}C}{E^{\tau}r} = \frac{E^{\tau}C}{E^{\tau}} \tag{5.31}$$

$$\frac{1}{\alpha_I} = 1 + \frac{1}{\beta_I} = 1 + \frac{{}_C\tau_E}{{}_C\tau_r} = \frac{{}_C\tau_E}{{}_C\tau} \tag{5.32}$$

5.2.1 Open-Circuit Saturation Currents

The reverse-saturation current through the emitter with the collector open-circuited (I_{EBO}) can be obtained in terms of the charge-control parameters by comparing Eqs. 5.4, 5.11 and 5.26. Then

$$I_{EBO} = \frac{{}_EK(1-\alpha_N\alpha_I)}{{}_E\tau} = \frac{{}_EK}{{}_E\tau}\frac{(1 + \beta_N + \beta_I)}{(1 + \beta_N)(1 + \beta_I)} \tag{5.33}$$

Use of Eqs. 5.29-5.32 for β_N, β_I, α_N, α_I and Eq. 2.65 for ${}_E\tau$ yields

$$I_{EBO} = \frac{{}_EK}{{}_E\tau_r}\left(1 + \frac{{}_C\tau_E/{}_E\tau_C}{1 + {}_C\tau_E/{}_C\tau_r}\right) \tag{5.34}$$

or

$$I_{EBO} = \frac{{}_EK}{{}_E\tau_r}\left(1 + \frac{\alpha_N(1-\alpha_I)}{1-\alpha_N}\right) \tag{5.35}$$

or

$$I_{EBO} = \frac{{}_EK}{{}_E\tau_r}\left(1 + \frac{\beta_N}{1+\beta_I}\right) \tag{5.36}$$

Similarly

$$I_{CBO} = \frac{{}_CK}{{}_C\tau_r}\left(1 + \frac{{}_E\tau_C/{}_C\tau_E}{1 + {}_E\tau_C/{}_E\tau_r}\right) \tag{5.37}$$

or

$$I_{CBO} = \frac{{}_{C}K}{{}_{C}\tau_r} \left(1 + \frac{\alpha_I(1-\alpha_N)}{1 - \alpha_I}\right) \tag{5.38}$$

or

$$I_{CBO} = \frac{{}_{C}K}{{}_{C}\tau_r} \left(1 + \frac{\beta_I}{1 + \beta_N}\right) \tag{5.39}$$

The conventional interpretation of I_{EBO} is that it is the emitter-base diode leakage (reverse-saturation) current, modified by a factor which accounts for the effects of the collector junction. It can be seen that the expression in parentheses reduces to unity if the transit times become infinite, which corresponds to the junctions having no effect on each other.

From Eqs. 5.34 and 5.37 it can be seen that in the limit of no recombination $({}_{E}\tau_r, {}_{C}\tau_r \rightarrow \infty)$ these two currents become zero and hence are useless for describing device behavior. Under this condition α_N and α_I become unity and it can be shown (see next section) that the ratio of I_{EBO} or I_{CBO} to $(1 - \alpha_N\alpha_I)$, which appears in Eqs. 5.4 - 5.6, remains finite.

The reason for the loss of utility of these open-circuit saturation currents is not hard to see. They are used because, for instance, a transistor with the collector terminal disconnected is generally assumed to be similar to a diode (the emitter-base diode). That this is not so is clear from Middlebrook's [21,26] discussion of the separation of the control and collection functions. In a diode, the contact is ohmic and combines both the control and collection of the moving carriers. In an ideal transistor, however, these functions are separated. The collector junction is nonohmic and fulfills only the collection function for the minority carriers, while the base contact (ideally) will not pass minority carriers and fulfills only the control function. Thus it is seen that the emitter and base terminals, with the collector floating, do not represent the external connections of a diode, so that the usual interpretation of I_{EBO} as a diode leakage current is incorrect for an ideal transistor.

In most real transistors, however, the open-circuit saturation currents are useful parameters. This is because most real transistors have an ohmic, rather than ideal, base contact which can serve both the control and collection functions. (In normal operation, the collection function is minimized by the physical or geometrical location of the base contact.) Another way of looking at this is to realize that the base contact is part of the base region. In order to have no recombination in the base (which leads to $I_{EBO}, I_{CBO} \rightarrow 0$) there must be no volume recombination and no surface recombination, including at the base contact*. This latter condition is violated in almost all real transistors.

From this discussion it can be seen that the open-circuit saturation currents are not mainly indicators of the junction behavior, as is generally assumed, but are, to a much greater extent, indicators of the recombination characteristics of the base region, particularly the base contact. Thus, far from being fundamental parameters of the device, they are only second-order parameters which are of little

---

* This, apparently, is not universally recognized. See, for instance, Sparkes[45] who considers $\alpha_N, \alpha_I < 1$ and $I_{EBO}, I_{CBO} \neq 0$ in the absence of recombination.

theoretical significance, but of somewhat greater practical importance.

5.2.2 Short-Circuit Saturation Currents

The Ebers-Moll equations can be written in terms of short-circuit currents as

$$I_E = I_{E,BC}(e^{V_{EB}}-1) - I_{E,BE}(e^{V_{CB}}-1) \tag{5.40}$$

$$I_C = -I_{C,BC}(e^{V_{EB}}-1) + I_{C,BE}(e^{V_{CB}}-1) \tag{5.41}$$

$$I_B = I_{B,BC}(e^{V_{EB}}-1) + I_{B,BE}(e^{V_{CB}}-1) \tag{5.42}$$

where the symbol $I_{\mu,\nu\lambda}$ is defined as the reverse-saturation current through terminal μ when terminals ν, λ are shorted $(V_{\nu\lambda} = 0)$. Electrically, then, $I_{E,BC}$ is the emitter reverse-saturation current when the base and collector are shorted together.

In terms of charge-control parameters, the short-circuit saturation currents are (from Eqs. 5.26 - 5.32 and Eqs. 5.40 - 5.42)

$$I_{E,BC} = \frac{{}_E K}{{}_E \tau} \tag{5.43}$$

$$I_{C,BE} = \frac{{}_C K}{{}_C \tau} \tag{5.44}$$

$$I_{C,BC} = \frac{{}_E K}{{}_E \tau_C} = \alpha_N I_{E,BC} \tag{5.45}$$

$$I_{E,BE} = \frac{{}_C K}{{}_C \tau_E} = \alpha_I I_{C,BE} \tag{5.46}$$

$$I_{B,BC} = \frac{{}_E K}{{}_E \tau_r} = \frac{I_{C,BC}}{\beta_N} = (1 - \alpha_N)\, I_{E,BC} \tag{5.47}$$

$$I_{B,BE} = \frac{{}_C K}{{}_C \tau_r} = \frac{I_{E,BE}}{\beta_I} = (1 - \alpha_I) I_{C,BE} \tag{5.48}$$

It is clear that, as indicated in the previous section, the emitter and collector short-circuit saturation currents (Eqs. 5.43 - 5.46) remain nonzero as the recombination rate in the base region approaches zero $({}_E\tau_r,\ {}_C\tau_r \to \infty;\ \ {}_E\tau_C,\ {}_E\tau \to {}_Et_C;\ \ {}_C\tau_E,\ {}_C\tau \to {}_Ct_E)$.

From the discussion of control and collection functions given in the previous section, it can be seen that by shorting the base and collector terminals we are combining (externally) the control and collection functions so that this combined terminal, together with the emitter terminal, does indeed represent a diode. This is true even if the control and collection functions are not completely separated internally, since the shorted terminal completely combines them externally. Thus the currents $I_{E,BC}$ and $I_{C,BE}$ can truly be called diode leakage (or saturation) currents. The currents $I_{C,BC}$ and $I_{E,BE}$ are ideally the collection components of the diode leakage currents. In real devices, however, it must be remembered that these currents do not constitute all of the collection current and that they also contain part of the control current. This is because the control and collection functions are not fully separated. The other two currents, $I_{B,BC}$ and $I_{B,BE}$ are ideally the control components of the two diode leakage currents. In real devices, of course, they do not constitute the entire control current and they contain part of the collection current.

5.3 Voltages as Dependent Variables

So far, the transistor currents have been considered as functions of the external voltages. However, as Ebers and Moll [20] point out, it may sometimes be convenient to consider the voltages as functions of the externally applied currents. In accord with this, they have solved Eqs. 5.4 - 5.6 for the junction voltages in terms of the external currents. Since their results are in terms of the open-circuit saturation currents, which are more sensitive to the base contact surface recombination rate than are the short-circuit saturation currents, they are difficult to interpret in the limit of no recombination and high gain. In this section, Eqs. 5.40 - 5.42 will be solved (with the aid of Eqs. 5.43 - 5.48) for the external voltages in terms of the currents, and the no-recombination (high gain) limit obtained.

Solution of Eqs. 5.41 and 5.42 for V_{EB} and V_{CB}, and rearrangement with the aid of Eqs. 5.43 - 5.48 results in

$$V_{EB} = \ln \left\{\frac{(1 + \beta_N)(1 + \beta_I)I_B - (1 + \beta_N)I_C + (1 + \beta_N + \beta_I)I_{E,BC}}{(1 + \beta_N + \beta_I)I_{E,BC}}\right\} \tag{5.49}$$

$$V_{CB} = \ln \left\{\frac{\beta_N(1 + \beta_I)I_B + (1 + \beta_I)I_C + (1 + \beta_N + \beta_I)I_{C,BE}}{(1 + \beta_N + \beta_I)I_{C,BE}}\right\} \tag{5.50}$$

The collector-emitter voltage V_{CE} can be obtained from Eqs. 5.49 and 5.50 as

$$V_{CE} = V_{CB} - V_{EB} \tag{5.51}$$

so that

$$V_{CE} = \ln \{[\frac{\beta_N(1 + \beta_I)I_B + (1 + \beta_I)I_C + (1 + \beta_N + \beta_I)I_{C,BE}}{(1 + \beta_N)(1 + \beta_I)I_B - (1 + \beta_N)I_C + (1 + \beta_N + \beta_I)I_{E,BC}}]\frac{I_{E,BC}}{I_{C,BE}}\} \tag{5.52}$$

For common-base operation, it is more convenient to consider the emitter and collector currents, rather than the base and collector currents, as the independent variables. Use of Kirchoff's current law (see Figure 2.1)

$$I_B = I_E + I_C \tag{5.53}$$

in Eqs. 5.49, 5.50 and 5.52 results in

$$V_{EB} = \ln \{\frac{(1 + \beta_N)(1 + \beta_I)I_E + (1 + \beta_N)\beta_I I_C + (1 + \beta_N + \beta_I)I_{E,BC}}{(1 + \beta_N + \beta_I)I_{E,BC}}\} \tag{5.54}$$

$$V_{CB} = \ln \{\frac{\beta_N(1 + \beta_I)I_E + (1 + \beta_N)(1 + \beta_I)I_C + (1 + \beta_N + \beta_I)I_{C,BE}}{(1 + \beta_N + \beta_I)I_{E,BC}}\} \tag{5.55}$$

$$V_{CE} = \ln \{[\frac{\beta_N(1 + \beta_I)I_E + (1 + \beta_N)(1 + \beta_I)I_C + (1 + \beta_N + \beta_I)I_{C,BE}}{(1 + \beta_N)(1 + \beta_I)I_E + (1 + \beta_N)\beta_I I_C + (1 + \beta_N + \beta_I)I_{E,BC}}]\frac{I_{E,BC}}{I_{C,BE}}\} \tag{5.56}$$

Ebers and Moll have obtained Eqs. 5.52, 5.54 and 5.55 in terms of the open-circut saturation currents and common-base current gains, except that they neglected the saturation current terms in the expressions for V_{CE}. They have also used the reciprocity relation (see Chapter VI)

$$I_{E,BE} = I_{C,BC} \tag{5.57}$$

or

$$\alpha_N I_{E,BC} = \alpha_I I_{C,BE} \tag{5.58}$$

It is to be noted that the Ebers-Moll results are considerably more concise than the above equations and hence are of more practical value for the analysis of an existing device, for which all the parameters are fixed and known. However, the results obtained here are of more theoretical significance since they are more easily manipulated to yield the effect of varying recombination rates. This is primarily because $I_{\mu,\nu\lambda}$ is less sensitive to the recombination rate than is $I_{\mu\nu\sigma}$ and β is a simple function of the lifetime.

In the limit of no recombination $(\tau_r \to \infty)$ we can show, from Eqs. 2.46, 2.58, 5.29 and 5.30, that

$$\lim_{\tau_r \to \infty} \frac{\beta_N}{\beta_I} = \frac{C^{t_E}}{E^{t_C}} \equiv \gamma \tag{5.59}$$

Also,

$$\lim_{\tau_r \to \infty} \alpha_N = 1 = \lim_{\tau_r \to \infty} \alpha_I \tag{5.60}$$

$$\lim_{\tau_r \to \infty} \beta_N = \infty = \lim_{\tau_r \to \infty} \beta_I \tag{5.61}$$

$$\lim_{\tau_r \to \infty} I_B = 0 \tag{5.62}$$

From Eqs. 5.58 and 5.60 we find

$$\lim_{\tau_r \to \infty} I_{E,BC} = \lim_{\tau_r \to \infty} I_{C,BE} \equiv I_s \tag{5.63}$$

Using Eqs. 5.59 - 5.63 in Eqs. 5.49, 5.50 and 5.52 we obtain

$$V_{EB}(\text{no recombination}) = \ln \left\{\frac{(1+\gamma)I_s - I_C}{(1+\gamma)I_s}\right\} \tag{5.64}$$

$$V_{CB}(\text{no recombination}) = \ln \left\{\frac{(1+\gamma)I_s + I_C}{(1+\gamma)I_s}\right\} \tag{5.65}$$

$$V_{CE}(\text{no recombination}) = \ln \left\{\frac{(1 + \gamma)I_s + I_C}{(1 + \gamma)I_s - \gamma I_C}\right\} \qquad (5.66)$$

These expressions can be obtained with the open-circuit saturation currents only if it is recognized that

$$\lim_{\tau_r \to \infty} \frac{I_{EBO}}{1-\alpha_N\alpha_I} = \lim_{\tau_r \to \infty} \frac{I_{CBO}}{1-\alpha_N\alpha_I} = I_s \qquad (5.67)$$

which is essentially the same as using the short-circuit saturation currents, as above.

Equations 5.64 - 5.66 (and also Eqs. 5.49 - 5.52) show that in a transistor with no recombination (or zero base current), if the external voltages are not held constant, then the maximum current that the device can sustain is limited by the short-circuit saturation currents and the ratio of β_N to β_I. This current limitation can easily be understood by considering a zero recombination transistor to be equivalent to two opposing diodes in series. Then, if a current is passed through the diodes, it will be limited by the reverse-saturation current of the reverse-biased diode (until it breaks down). With neglect of voltage breakdown, the voltage across the reverse-biased (high resistance) junction approaches infinity as the current is (attempted to be) increased, while the forward biased (low resistance) junction sustains only a small voltage drop. In practice, of course, this limiting condition will never be observed. Either the current source will reach its limiting voltage and become a voltage source, or the reverse-biased junction will break down and conduct (and possibly burn out).

5.4 Conclusions

The Ebers-Moll equations have been derived from charge-control principles, thus relating two models of the transistor. It has been shown that the open-circuit saturation currents do not have the significance usually attributed to them since they are second-order, rather than fundamental, diode currents. It was shown that the short-circuit saturation currents are fundamental parameters and thus of more theoretical significance. However, owing to the nonideality of most real transistors, the open-circuit saturation currents are of a useful nature.

The expressions for the currents in terms of the voltages were inverted to obtain the voltages as functions of the external currents, short-circuit saturation currents, and common-emitter current gains. These are of greater theoretical significance than the same expressions in terms of the open-circuit saturation currents which are of practical importance for most real transistors. The high gain or no-recombination limit of these expressions was obtained and it was shown that the short-circuit saturation currents are more easily used in this limit than are the open-circuit saturation currents.

CHAPTER VI

TRANSISTOR RECIPROCITY

In n-port circuit theory, if excitations and responses are linearly related, the system equations can be written as

$$
\begin{aligned}
R_1 &= a_{11}E_1 + a_{12}E_2 + \cdots + a_{1n}E_n \\
R_2 &= a_{21}E_1 + a_{22}E_2 + \cdots + a_{2n}E_n \\
\vdots & \qquad\quad \vdots \\
R_n &= a_{n1}E_1 + a_{n2}E_2 + \cdots + a_{nn}E_n
\end{aligned}
\tag{6.1}
$$

where E_i is the excitation applied at port i, R_i is the response measured at port i, and a_{ij} are constants. Equation 6.1 is of course valid for any linear system, not only electric circuits. If the condition

$$R_i(0,0 \ldots E_j = E,0,0 \ldots) = R_j(0,0 \ldots E_i = E,0,0 \ldots) \quad (\text{all } i,j) \tag{6.2}$$

which is equivalent to

$$a_{ij} = a_{ji} \quad (\text{all } i,j) \tag{6.3}$$

is satisfied, then the system is said to be reciprocal. In words, the principle of reciprocity may be stated as follows: given that an excitation E at port i produces a response R at port j; if the same excitation E applied at port j produces the same response R at port i, then the system is reciprocal with respect to ports i and j. If all sets of two-ports are reciprocal, then it is simply said that the system is reciprocal.

This chapter is concerned only with the bipolar transistor, considered as a two-port device. For this device, the principle of reciprocity is valid for any geometry, any base impurity distribution, linear recombination, low-level injection (linearity), and low frequencies.

Searle, et al. [46] give a plausibility argument for the validity of reciprocity from a circuit point of view. For sufficiently small applied voltages, the Ebers-Moll equations (Eqs. 5.1, 5.2) can be written as

$$I_E = a_{11}V_{EB} + a_{12}V_{CB} \tag{6.4}$$

$$I_C = a_{21}V_{EB} + a_{22}V_{CB} \tag{6.5}$$

They then argue that "On physical grounds, it is plausible that the transistor must now be indistinguishable from a simple block of resistive material with three leads. Thus it must possess under these conditions all the properties of such an element, including the property of reciprocity" [46].

A rigorous proof of reciprocity is given by Shockley, et al. [43] for the most general geometry and impurity distribution in the presence of linear recombination. While this proof is completely rigorous it is strictly mathematical, and no conceptual or physical interpretation of the derivation is presented. In particular, the authors introduce a vector

$$\underset{\sim}{A}(x,y,z) \equiv \frac{1}{n_P(x,y,z)}[{}_E n'(x,y,z)\,{}_C\underset{\sim}{j}(x,y,z) - {}_C n'(x,y,z)\,{}_E\underset{\sim}{j}(x,y,z)] \tag{6.6}$$

which is given no conceptual or physical interpretation. The proof consists of mathematical manipulations of this vector, so that no conceptual understanding is obtained. Ebers and Moll [20] use essentially the same method, except that it is less general.

In this chapter the principle of reciprocity will be derived for a bipolar transistor with arbitrary geometry, arbitrary impurity distribution, and an arbitrary spatial variation of linear recombination, from the principles developed in Chapter II and other conceptual considerations. The only restrictions imposed will be low injection levels and low frequencies. The trivial case of a perfectly symmetric transistor (with respect to emitter and collector) will not be considered. For such a device, it is obvious that reciprocity will be valid for any injection level and any frequency, even if all carriers interact.

6.1 No Recombination

In the absence of recombination, transistor reciprocity can be proven rather simply, directly from the Ebers-Moll equations (Eqs. 5.1, 5.2). In this case, Eqs. 5.26 and 5.27 or Eqs. 5.40 - 5.46 show that (since ${}_E\delta, {}_C\delta = 0$ and $\alpha_I, \alpha_N = 1$)

$$a_{11} = - a_{21} \tag{6.7}$$

$$a_{22} = - a_{12} \tag{6.8}$$

so that Eqs. 5.1 and 5.2 reduce to

$$I_E(V_{EB},V_{CB}) = - a_{21}(e^{V_{EB}}-1) + a_{12}(e^{V_{CB}}-1) \quad (6.9)$$

$$I_C(V_{EB},V_{CB}) = a_{21}(e^{V_{EB}}-1) - a_{12}(e^{V_{CB}}-1) = - I_E(V_{EB},V_{CB}) \quad (6.10)$$

To prove reciprocity, it is required to show that

$$a_{12} = a_{21} \quad (6.11)$$

Substitution of Eq. 6.11 into Eqs. 6.9 and 6.10 leads to the conclusion that with equal applied emitter and collector voltages the net current is zero*

$$I_E(V,V) = - I_C(V,V) = 0 \quad (6.12)$$

Clearly, Eqs. 6.11 and 6.12 are equivalent; that is, they imply each other. Thus, a demonstration of the validity of Eq. 6.12 will constitute an indirect proof of reciprocity. The proof is indirect because it is necessary to prove Eq. 6.11, whereas Eq. 6.12 leads directly (with no assumptions as to recombination) to the conclusion that

$$a_{12} = - a_{11} \quad (6.13)$$

$$a_{21} = - a_{22} \quad (6.14)$$

The condition of no recombination must then be used to obtain Eqs. 6.7

---

* This is a consequence of the infinite driving-point (base-emitter or base-collector) impedance of a zero-recombination transistor.

and 6.8 which, with Eqs. 6.13 and 6.14, will lead to Eq. 6.11.

Eq. 6.12 can be proven by expressing the current in terms of the electron and hole current densities, which can be expressed in terms of the carrier quasi-Fermi levels. By showing that the quasi-Fermi levels must be constant for equal applied emitter and collector voltages, we will show that the current densities, and hence the currents, must be zero.

The total current can be written as

$$I(V,V) = I_n(V,V) + I_p(V,V) \tag{6.15}$$

$$= \iint_S \underset{\sim}{j}_p \cdot d\underset{\sim}{S} + \iint_S \underset{\sim}{j}_n \cdot d\underset{\sim}{S} \tag{6.16}$$

where $d\underset{\sim}{S}$ is the cross-sectional area of a flow tube and is always directed in the same direction along the flow tube (i.e. the positive x direction) as in Figure 6.1, $\underset{\sim}{j}_n$ and $\underset{\sim}{j}_p$ are the electron and hole current densities along their respective flow tubes*, and the integrations may be carried out for any surface crossing all of the flow tubes in the base region. Attention will be focussed on the electron current; arguments for the hole current are analogous.

The electron current density along a flow tube can be expressed as

$$\underset{\sim}{j}_n = -eD_n(n_P + {}_En' + {}_Cn')\underset{\sim}{\nabla}\psi_o + eD_n\,\underset{\sim}{\nabla}(n_P + {}_En' + {}_Cn') \tag{6.17}$$

---

* Electrons and holes do not necessarily travel along the same flow tubes.

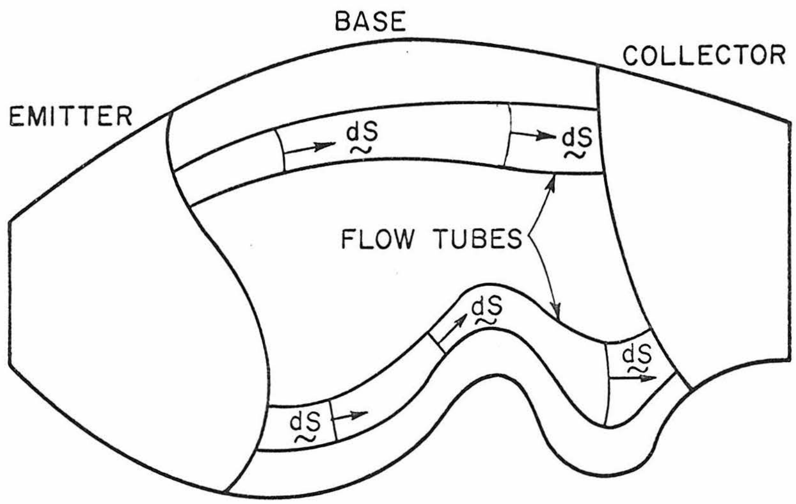

Fig. 6.1 Examples of flow tubes showing the vector cross-sectional area $d\underset{\sim}{S}$, which is always pointed in the same direction along the flow tube.

where ψ_o is the electrostatic potential due to the built-in electric field (normalized to the thermal potential V_t), given by

$$\underset{\sim}{\nabla}\psi_o = \frac{-\underset{\sim}{E}_o}{V_t} = \frac{1}{n_P}\underset{\sim}{\nabla}n_P \tag{6.18}$$

Integration of Eq. 6.18 yields

$$\psi_o = \ln(n_P/n_I) \tag{6.19}$$

where the constant of integration has been taken as the intrinsic carrier density n_I.

The (normalized) electron quasi-Fermi level ϕ_n, defined by [43]

$$n = n_P + {}_E n' + {}_C n' = n_I \, e^{(\psi_o - \phi_n)} \tag{6.20}$$

is now introduced, where n_I is the intrinsic carrier density. Substitution of Eq. 6.20 into Eq. 6.17 leads to

$$\underset{\sim}{j}_n = - e \, D_n n \, \underset{\sim}{\nabla} \, \phi_n \tag{6.21}$$

Since recombination has been excluded, the current $(\underset{\sim}{j}_n \cdot d\underset{\sim}{S})$ cannot vary with position along a flow tube.

Since the direction of $d\underset{\sim}{S}$ is unchanging along a flow tube it follows that $\underset{\sim}{j}_n$, and hence $\underset{\sim}{\nabla}\phi_n$, cannot change sign along a flow tube, so that ϕ_n must be monotonic.

Now consider the electron densities at the emitter and collector depletion layer edges. The first-order approximation gives

$${}_E n'(\text{collector}) = {}_E n'(C,y,z) = 0 \tag{6.22}$$

$${}_C n'(\text{emitter}) = {}_C n'(E,y,z) = 0 \tag{6.23}$$

Use of Eqs. 6.22 and 6.23 in Eq. 6.20 results in

$$n(E,y,z) = n_P(E,y,z) + {}_E n'(E,y,z) = n_I e^{[\psi_o(E,y,z) - \phi_n(E,y,z)]} \tag{6.24}$$

$$n(C,y,z) = n_P(C,y,z) + {}_C n'(C,y,z) = n_I e^{[\psi_o(C,y,z) - \phi_n(C,y,z)]} \tag{6.25}$$

For low injection levels Eq. 4.49 may be used for ${}_E n'(E,y,z)$ and a similar relation for ${}_C n'(C,y,z)$. Then, also using Eq. 6.19 for $\psi_o(E,y,z)$ and $\psi_o(C,y,z)$, Eqs. 6.24 and 6.25 become

$$n(E,y,z) = n_P(E,y,z) + n_P(E,y,z)(e^{V_{EB}}-1) = n_P(E,y,z)e^{-\phi_n(E,y,z)} \tag{6.26}$$

$$n(C,y,z) = n_P(C,y,z) + n_P(C,y,z)(e^{V_{CB}}-1) = n_P(C,y,z)e^{-\phi_n(C,y,z)} \tag{6.27}$$

Comparison of Eqs. 6.26 and 6.27 shows that if the applied voltages are equal, the electron quasi-Fermi level has the same value at the emitter and collector depletion layer edges:

$$\phi_n(E,y,z) = \phi_n(C,y,z) \tag{6.28}$$

Since it has already been shown that ϕ_n is monotonic, it is clear that

$$\phi_n = \text{constant} \tag{6.29}$$

so that Eq. 6.21 results in

$$j_n = 0 \tag{6.30}$$

Since holes are subject to the same injection, collection and transport mechanisms as are electrons, an exactly analogous argument for the hole current density leads to

$$j_p = 0 \tag{6.31}$$

Substitution of Eqs. 6.30 and 6.31 into Eq. 6.16 immediately yields Eq. 6.12.

This result (Eq. 6.12) can also be derived from circuit considerations. Since there is no base current (because there is no recombination), the external voltages can be applied as in Figure 6.2 and the base lead can then be cut without affecting anything else. Then it is obvious that, if

$$V_{CB} = V_{EB} = V \tag{6.32}$$

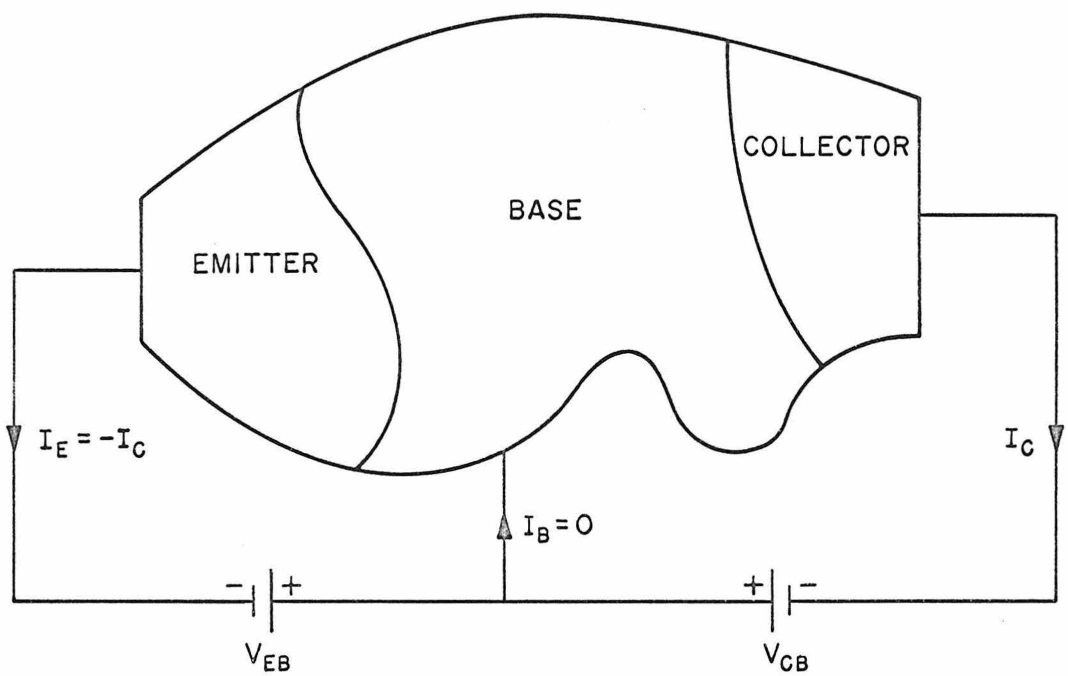

Fig. 6.2 Terminal currents for a transistor with no recombination.

the net voltage across the device is zero, so that there can be no net current flow and Eq. 6.12 is satisfied.

Since Eq. 6.11 is equivalent to Eq. 6.12, the principle of reciprocity has now been proven for a bipolar transistor with no recombination.

6.2 With Linear Recombination

In the presence of linear recombination rates the simple arguments of the previous section cannot be used because

$$a_{11} \neq -a_{21} \tag{6.33}$$

$$a_{22} \neq -a_{12} \tag{6.34}$$

and Eq. 6.12 is not valid.

In this section a direct proof of reciprocity in the presence of linear recombination rates will be presented by deriving Eq. 6.11 directly. For low injection levels, the emitter and collector collection (or transit) currents can be expressed as (with the use of Eqs. 5.1, 5.2, 5.14 and 5.15)

$${}_{E}I_{C}(V_{EB}) = a_{21}(e^{V_{EB}}-1) \tag{6.35}$$

$${}_{C}I_{E}(V_{CB}) = a_{12}(e^{V_{CB}}-1) \tag{6.36}$$

so that the condition of reciprocity $(a_{12} = a_{21})$ is equivalent to

$${}_{E}I_{C}(V) = {}_{C}I_{E}(V) \tag{6.37}$$

Thus, it is necessary to prove that, for equal applied voltages, the emitter and collector collection (or transit) currents are equal. Equivalently, the net current due to nonrecombining (collected) carriers $({}_{E}I_{C}(V) - {}_{C}I_{E}(V))$ must be shown to be zero. This is essentially what was done in the previous section. However, the task was simpler there since in the absence of recombination the transit current and the total current are identical. Here, the total current must first be separated into its transit and recombination components, so that the transit current can be considered alone. Although the details will be more complex owing to the presence of recombination, the derivation in this section will be similar to that in the previous section.

Eq. 6.37 can be proven be expressing the transit current in terms of the nonrecombining excess carriers, which can be expressed in terms of pseudo-Fermi levels (see below). By showing that the pseudo-Fermi levels must be constant for equal applied emitter and collector voltages, it will be shown that the transit current density, and hence the transit current, must be zero, so that Eq. 6.37 is valid.

The transit current $I_t(V,V)$ can be written as

$$I_t(V,V) = {}_{E}I_{C}(V) - {}_{C}I_{E}(V) \tag{6.38}$$

$$= \iint_S \underset{\sim}{j}_p(\text{transit}) \cdot d\underset{\sim}{S} + \iint_S \underset{\sim}{j}_n(\text{transit}) \cdot d\underset{\sim}{S} \tag{6.39}$$

where $d\underset{\sim}{S}$ is the cross-sectional area of a transit (current) flow tube and is always directed in the same direction along the transit flow tube (i.e. the positive x direction) as in Figure 6.1, $\underset{\sim}{j}_n$ (transit) and $\underset{\sim}{j}_p$ (transit) are the electron and hole transit current densities along their respective transit flow tubes, and the integrations may be carried out for any surface crossing all of the transit flow tubes in the base region. Attention will be focussed on the electron transit current; arguments for the hole transit current are analogous.

The electron transit current density along a transit flow tube can be expressed as

$$\underset{\sim}{j}_n(\text{transit}) = -eD_n({}_E n'_C + {}_C n'_E)\,\underset{\sim}{\nabla}\psi_o + eD_n\,\underset{\sim}{\nabla}({}_E n'_C + {}_C n'_E) \qquad (6.40)$$

where ψ_o is the electrostatic potential due to the built-in electric field (see Eq. 6.19). The (normalized) pseudo-Fermi level θ_n of non-recombining excess electrons*, defined by

$$n'_t \equiv {}_E n'_C + {}_C n'_E = n_I\, e^{(\psi_o - \theta_n)} \qquad (6.41)$$

is now introduced, where n'_t is the total density of nonrecombining excess electrons.

Substitution of Eq. 6.41 into Eq. 6.40 results in

$$\underset{\sim}{j}_n(\text{transit}) = -eD_n\, n'_t\, \underset{\sim}{\nabla}\theta_n \qquad (6.42)$$

---

* The term pseudo-Fermi level is used rather than quasi-Fermi level because only nonrecombining excess electrons are considered here, whereas the quasi-Fermi level, as introduced by Shockley, et al. [43] and as commonly used, refers to the total density of electrons.

Since the electrons constituting the transit current do not recombine, the transit current ($\underset{\sim}{j}_n$(transit) $\cdot$ $d\underset{\sim}{S}$) cannot vary with position along a transit flow tube.

Since the direction of $d\underset{\sim}{S}$ is unchanging along a transit flow tube it follows that $\underset{\sim}{j}_n$(transit), and hence $\underset{\sim}{\nabla}\theta_n$, cannot change sign along a transit flow tube. Hence, θ_n must be monotonic along a transit flow tube.

Now consider the densities of nonrecombining excess electrons at the emitter and collector depletion layer edges. The first-order approximation yields

$$ {}_{E}n'_{C}(C,y,z) = 0 \tag{6.43} $$

$$ {}_{C}n'_{E}(E,y,z) = 0 \tag{6.44} $$

Use of Eqs. 6.43 and 6.44 in Eq. 6.41 results in

$$ n'_{t}(E,y,z) = {}_{E}n'_{C}(E,y,z) = n_I e^{[\psi_o(E,y,z) - \theta_n(E,y,z)]} \tag{6.45} $$

$$ n'_{t}(C,y,z) = {}_{C}n'_{E}(C,y,z) = n_I e^{[\psi_o(C,y,z) - \theta_n(C,y,z)]} \tag{6.46} $$

The density of recombining electrons at point (x,y,z) due to injection at the emitter can be written as the total density of electrons at point (x,y,z) due to injection at the emitter mulitplied by the probability of survival (not recombining before collection) of these electrons ${}_{E}\eta(x,y,z)$, and similarly for the electrons injected at the collector. Then

$$ {}_E n'_C(x,y,z) = {}_E\eta(x,y,z)\,{}_E n'(x,y,z) \tag{6.47} $$

$$ {}_C n'_E(x,y,z) = {}_C\eta(x,y,z)\;{}_C n'(x,y,z) \tag{6.48} $$

Substitution of Eqs. 6.47 and 6.48 into Eqs. 6.45 and 6.46 results in

$$ {}_E\eta(E,y,z)\;{}_E n'(E,y,z) = n_I\; e^{[\psi_o(E,y,z)\; -\; \theta_n(E,y,z)]} \tag{6.49} $$

$$ {}_C\eta(C,y,z)\;{}_C n'(C,y,z) = n_I\; e^{[\psi_o(C,y,z)\; -\; \theta_n(C,y,z)]} \tag{6.50} $$

For low injection levels Eq. 4.49 may be used for ${}_E n'(E,y,z)$, and a similar relation for ${}_C n'(C,y,z)$. Then, with use of Eq. 6.19 for $\psi_o(E,y,z)$ and $\psi_o(C,y,z)$, Eqs. 6.49 and 6.50 become

$$ {}_E\eta(E,y,z)\; n_p(E,y,z)(e^{V_{EB}}-1) = n_p(E,y,z)\; e^{-\theta_n(E,y,z)} \tag{6.51} $$

$$ {}_C\eta(C,y,z)\; n_p(C,y,z)\; (e^{V_{CB}}-1) = n_p(C,y,z)\; e^{-\theta_n(C,y,z)} \tag{6.52} $$

which can be rewritten as

$$ {}_E\eta(E,y,z)\; (e^{V_{EB}}-1) = e^{-\theta_n(E,y,z)} \tag{6.53} $$

$$ {}_C\eta(C,y,z)\; (e^{V_{CB}}-1) = e^{-\theta_n(C,y,z)} \tag{6.54} $$

Owing to injection at the emitter, there is a density ${}_{E}n'_{C}(x,y,z)$ of nonrecombining electrons traveling along a transit flow tube. From Eq. 6.47, ${}_{E}n'_{C}(x,y,z)$ can be considered as the density of all emitter-injected electrons traveling along the same transit flow tube, but with a probability ${}_{E}\eta(x,y,z)$ of not recombining before reaching the collector. A similar argument holds for collector-injected electrons. If it is assumed that the probability of recombination of an electron at any point is independent of the past history of the electron, then the probability of recombination or of survival must be independent of the direction of travel of the electrons. Thus the probability that an electron injected at the emitter will not recombine before collection must be the same as the probability that an electron injected at the collector will not recombine before collection, if they are injected at opposite ends of the same transit flow tube so that they will travel along the same transit flow tube. Hence

$$ {}_{E}\eta(E,y,z) = {}_{C}\eta(C,y,z) \qquad (6.55) $$

By comparing Eqs. 6.53 and 6.54, with the aid of Eq. 6.55, it is seen that if the emitter and collector applied voltages are equal, the pseudo-Fermi level at the emitter depletion layer edge is the same as at the collector depletion layer edge:

$$ \theta_n(E,y,z) = \theta_n(C,y,z) \qquad (6.56) $$

Since it has already been shown that θ_n is monotonic along a transit flow tube, it is clear that

$$\theta_n = \text{constant} \tag{6.57}$$

along a transit flow tube, so that Eq. 6.42 becomes

$$j_n(\text{transit}) = 0 \tag{6.58}$$

Since holes are subject to the same injection, collection, transport, and recombination mechanisms as are electrons, an exactly analogous argument for the hole transit current density leads to

$$j_p(\text{transit}) = 0 \tag{6.59}$$

Substitution of Eqs. 6.58 and 6.59 into Eq. 6.39 and then into Eq. 6.38, immediately yields Eq. 6.37, so that reciprocity has now been proven in the presence of recombination. It is seen that the derivation given here reduces to that of the previous section if there is no recombination, so that $\eta = 1$.

6.3 Conclusions

The principle of reciprocity has now been derived for a transistor with an arbitrary geometry, arbitrary impurity distribution, and arbitrary spatial variation of linear recombination, under the restrictions of low injection level and low frequency. The result is, of course, not new. However, the derivation given here is based on conceptual processes occurring within the transistor and is carried out in such a way as to yield a conceptual understanding of the processes leading to the final result.

CHAPTER VII

CONCLUSIONS

In this work the charge-control method has been considered as a means of enhancing conceptual understanding of device operation, and this method has been given a sound theoretical foundation based on a one-to-one correspondence between internal processes and charge-control parameters. All of the results have been obtained for a device with an arbitrary three-dimensional geometry, arbitrary impurity distribution, and arbitrary spatial variation of recombination rate. The results are valid within, and limited by, the usual first-order approximations in which a semiconductor device is separated into completely depleted regions and quasi-neutral regions with abrupt boundaries between them. Some limitations to the theory have been presented and it has been shown how the basic concepts can be used to account for second-order effects and to give a conceptual interpretation to results which heretofore have been given only mathematical or electrical (circuit) significance. Finally, a conceptual derivation of transistor reciprocity has been presented in which the final result was related to internal processes of the device.

In Chapter II the current density was first considered as a flux density of mobile carriers. Separation of the velocity and the carrier density into equilibrium and excess components permitted the several average velocities to be given clear conceptual meanings, directly related to the motion of the individual carriers. In the absence of recombination, the point-average carrier velocity was used to obtain the average carrier transit time across the base of a transistor, and it was shown that for the DC steady-state, the current is simply the ratio of excess or injected base charge to the transit time. In the presence of recombination the derivation could not proceed in exactly the same way because all of the injected carriers do not reach the collecting junction. This difficulty was overcome by separation of the injected carriers into those that recombine and those that survive. This separation permitted the collected current to be treated in the same way as the total current was treated in the absence of recombination. The total injected current was then obtained in a formal mathematical manner by definition of a new parameter, the injection time, which was subsequently shown to be the time required for the removal of the total excess charge from the base by means of recombination and collection. The Beaufoy-Sparkes [1] collection time constant was given a conceptual interpretation by showing that it is the mean time required for an injected carrier to be removed from the base by collection alone, and hence can be considered to be a transit or collection lifetime. In analogy with the collection transit time a new parameter, the recombination transit time, was introduced. This parameter is the mean time for a carrier to travel from the point of injection to the point of recombination. The similarity between recombination and collection as mechanisms for removing carriers from the base was pointed out. In the final section of Chapter II the base current was obtained from the continuity equation and it was shown that the injection time can be obtained as the "parallel" combination of the recombination and collection lifetimes.

In Chapter III, some fundamental and practical limitations to the theory developed in the previous chapter were discussed. An important limitation is the neglect of the nonzero (delay) time required for a variation in the signal applied at the emitting junction to propagate to the collecting junction. Other limitations of the theory are the loss of simplicity when high injection levels and second-order effects are included, and the inherent loss of detailed information owing to the use of gross or average quantities rather than point variable functions (e.g., densities). A synthesis of the concepts of lumped models and charge-control theory was suggested as a possible means of overcoming the latter limitation.

In Chapter IV it was shown that the charge-control concepts can be used very easily to obtain the DC common-emitter current gain under some simplifying assumptions. The results of Chapter II were extended to the case of high injection levels and it was shown that the usual high-injection results can be obtained as easily for three dimensions by charge-control methods as for one dimension in the conventional manner. The result was further generalized to include an arbitrary impurity distribution. It was then shown how the charge-control theory can be extended to account for nonlinear recombination rates, other second-order effects, and additional mechanisms for carrier transport and injection, such as are encountered with a nonideal base contact. Relations between applied voltage and injected charge were also obtained.

In Chapter V the Ebers-Moll circuit equations were derived from and related to the charge-control model. With the tools developed in Chapters II and IV it was a simple matter to derive the Ebers-Moll circuit relations entirely from charge-control principles and to obtain relations among various parameters. It was shown that short-circuit saturation currents would be more appropriate in the circuit equations than the original open-circuit saturation currents, since the former are related to more fundamental processes than are the latter.

In Chapter VI the principle of reciprocity was proven for a bipolar transistor with an arbitrary geometry, arbitrary impurity distribution, and an arbitrary spatial variation of linear recombination. A derivation valid in the absence of recombination was presented, based on the Ebers-Moll equations and the quasi-Fermi level of the carriers. Reciprocity in the presence of recombination was then proven by using the principle of the separation of injected charge into recombining and surviving components, and the concept of the pseudo-Fermi level, to show that the net transit current is zero if the applied emitter and collector voltages are equal. The derivations given here are more closely related to the processes occurring within the device than are previous derivations.

APPENDIX

CONVENTIONAL DERIVATION OF HIGH-INJECTION RELATIONS

In this appendix the carrier transport equations (non-charge-control analysis) will be used to obtain the minority carrier current density in a diode or transistor at all injection levels for which there is a quasi-neutral region.

The P region of an N^+P diode or NPN transistor with an arbitrary one-dimensional impurity distribution will be considered. A one-dimensional geometry and absence of recombination will be assumed.

The transport equations are

$$j_n = e\mu_n nE' + e\mu_n n'E_o + eD_n \frac{dn'}{dx} \quad (A.1)$$

$$j_p = e\mu_p pE' + e\mu_p p'E_o - eD_p \frac{dp'}{dx} \quad (A.2)$$

where n,p are the total carrier densities, n',p' are the excess carrier densities, E_o is the built-in field, and E' is the excess field. Equation A.2 can be solved for the excess field to obtain

$$E' = \frac{1}{e\mu_p p} (j_p - e\mu_p p'E_o + eD_p \frac{dp'}{dx}) \quad (A.3)$$

Use of Eq. A.3 in Eq. A.1 and rearrangement of the result yields the minority carrier current as

$$j_n = eD_n \frac{dn'}{dx} (1 + \frac{n}{p} \frac{dp'/dx}{dn'/dx}) + e\mu_n n'E_o(1 - \frac{n}{p}\frac{p'}{n'}) + \frac{\mu_n}{\mu_p}\frac{n}{p} j_p \quad (A.4)$$

Note that Eq. A.4 contains no approximations; quasi-neutrality has not been assumed.

The assumption of quasi-neutrality results in

$$n' = p' \tag{A.5}$$

and

$$\frac{dn'}{dx} = \frac{dp'}{dx} \tag{A.6}$$

Use of Eqs. A.5 and A.6 in Eq. A.4 leads to an expression for the current density

$$j_n = eD_n \frac{dn'}{dx}\left(1 + \frac{n}{p}\right) + e\mu_n n' E_o\left(1 - \frac{n}{p}\right) + \frac{\mu_n}{\mu_p}\frac{n}{p} j_p \tag{A.7}$$

which is valid at all injection levels in the quasi-neutral region.

The usual derivation seeking to account for the excess field [40,41] is restricted to uniform impurity densities ($E_o = 0$) and generally assumes that the majority carrier current can be neglected*. Thus Eq. A.7 is a generalization of the conventional results, but is still restricted to one dimension.

For high injection levels

$$n \simeq p \tag{A.8}$$

so that Eq. A.7 becomes

$$j_n = 2eD_n \frac{dn'}{dx} + \frac{\mu_n}{\mu_p} j_p \tag{A.9}$$

*Jonscher [47] and Middlebrook [48] are notable exceptions. However, Jonscher makes other unnecessary assumptions, such as $n=n_p+n' \simeq n'$, so that his results are not valid for very low injection levels.

Thus, it is seen that for high injection levels, the built-in field is of no importance. This is, of course, intuitively obvious, since at these injection levels the excess densities completely overcome the equilibrium densities so that the equilibrium gradients (and hence field) are, in a sense, undetectable.

LIST OF PRINCIPAL SYMBOLS

a_{ij} = parameter in Ebers-Moll equations

B = base contact

$B_{\mu}(V_{\mu\nu}, V_{\lambda\nu})$ = function of voltage defined by Eq. 4.58

$b_{\mu}(V_{j\mu})$ = function of voltage defined by Eq. 4.44

C = collector junction

$D_n(D_p)$ = electron (hole) diffusion constant

E = emitter junction; electric field

E_o = electric field due to impurity distribution

E' = electric field due to injected carriers

e = magnitude of electronic charge ($e > 0$)

${}_{\mu}f$ = function giving spatial dependence of carriers injected at surface μ

I = current

I_{ν} = total current at surface ν

${}_{\mu}I_{\nu}$ = current at surface ν due to injection at surface μ

$I_{\mu\nu 0}$ = reverse saturation current at terminal μ when the third terminal is open circuited

$I_{\mu,\nu\lambda}$ = reverse saturation current at terminal μ when terminals ν and λ are shorted together

I_s = $I_{E,BC}$ and $I_{C,BE}$ in absence of recombination

I_t = total current due to nonrecombining carriers

j = current density

${}_{\mu}j$ = total current density due to injection at surface μ

${}_{\mu}j_{\nu}$ = current density at surface ν due to injection at surface μ

j(transit) = current density due to nonrecombining carriers

${}_{\mu}K$	=	constant defined by Eq. 4.58
k	=	Boltzmann's constant
n	=	total electron density
n_o	=	equilibrium electron density
n'	=	total excess (injected) electron density
n'_t	=	total density of nonrecombining injected electrons
${}_{\mu}n'$	=	density of electrons due to injection at surface μ
${}_{\mu}n'_{\nu}$	=	density of electrons due to injection at surface μ that will be collected at surface ν (density of injected electrons in transit from surface μ to surface ν)
${}_{\mu}n'_r$	=	density of electrons due to injection at surface μ that will recombine
n_I	=	intrinsic electron density
$n_P(n_N)$	=	equilibrium electron density in P-type (N-type) material
p	=	total hole density
p_o	=	equilibrium hole density
p'	=	total excess (injected) hole density
$p_P(p_N)$	=	equilibrium hole density in P-type (N-type) material
Q	=	total charge (positive for both holes and electrons)
Q_o	=	equilibrium charge
Q'	=	total injected charge
${}_{\mu}Q'$	=	injected charge due to injection at surface μ
${}_{\mu}Q'_{\nu}$	=	injected charge in transit from surface μ to surface ν
${}_{\mu}Q'_r$	=	charge injected at surface μ that recombines
R	=	recombination rate of charge for the entire base region
s	=	Laplace transform variable
T	=	absolute temperature; period of applied signal

t = time

t_t = general transit time

t_d = transit time due to diffusion

t_{Do} = transit time due to drift in equilibrium field E_o

t'_D = transit time due to drift in excess field E'

${}_{\mu}t_{\nu}$ = average transit time from surface μ to surface ν (per flow tube if subscripts are lower case)

${}_{\mu}t_r$ = average recombination transit time for injection from surface

t_{delay} = delay time

t_{relax} = relaxation time

U = net rate of removal of carriers (recombination rate)

u = velocity of individual carriers in the direction of current flow

u_o = portion of u due to equilibrium conditions: individual electron equilibrium velocity

u' = portion of u due to injected carriers: individual electron excess velocity

V_t = thermal potential = kT/e

$V_{j\mu}$ = voltage across junction μ

$V_{\mu\nu}$ = external voltage across terminals μ and ν ; terminal μ positive with respect to terminal ν for $V_{\mu\nu} > 0$

v_o = point-average equilibrium velocity of carriers (due to equilibrium conditions)

v' = point-average excess velocity of carriers (due to injected carriers)

${}_{\mu}v_o$ = average equilibrium velocity of carriers that were injected at surface μ

v_{Do} = average drift velocity of carriers due to the equilibrium field E_o

v'_D = average drift velocity of carriers due to the excess field E'

v_d	=	average diffusion velocity of carriers
W	=	length of flow tube
x	=	coordinate along flow tube (normal to y,z)
y	=	coordinate, normal to x,z
z	=	coordinate, normal to x,y
$\alpha(\alpha_N,\alpha_I)$	=	general (normal, inverted) common-base current gain
$\beta(\beta_N,\beta_I)$	=	general (normal, inverted) common-emitter current gain
γ	=	ratio of β_N to β_I in absence of recombination
${}_\mu\delta$	=	fraction of charge injected at surface μ that recombines
${}_\mu\eta(x,y,z)$	=	probability that an electron at point (x,y,z) which was injected at junction μ will not recombine before reaching the opposite junction
$\theta_n(x,y,z)$	=	pseudo-Fermi level of nonrecombining electrons
$\mu_n(\mu_p)$	=	electron (hole) mobility
τ	=	time
τ_r	=	recombination lifetime at a point
${}_\mu\tau_r$	=	average recombination lifetime over the entire base region for carriers injected at surface μ
τ_t	=	general collection (transit) lifetime
${}_\mu\tau_\nu$	=	average transit lifetime for carriers injected at surface μ and collected at surface ν
${}_\mu\tau$	=	injection time = average lifetime due to recombination and collection of carriers injected at surface μ
ϕ_n	=	quasi-Fermi level of (all) electrons
ψ_o	=	electrostatic potential due to impurity distribution

ω = signal frequency (radians per second)

Note 1

Subscript n(p) indicates electrons (holes).

Note 2

Subscript o indicates equilibrium or initial conditions.

Note 3

A prime (') indicates the component due to injected carriers.

Note 4

Transit time is denoted by t .

Lifetime is denoted by τ .

Note 5

A pre-subscript indicates the surface of injection of the carriers contributing to the parameter of interest.

Note 6

A post-subscript indicates the destination of carriers, or the surface at which the current (density) is evaluated.

REFERENCES

[1] R. Beaufoy and J. J. Sparkes, "The Junction Transistor as a Charge-Controlled Device", A.T.E. Jour. 13, No. 4 (1957).

[2] J. J. Sparkes and R. Beaufoy, "The Junction Transistor as a Charge-Controlled Device", IRE Proc 45, 1740-1742; December 1957.

[3] E. O. Johnson and A. Rose, "Simple General Analysis of Amplifier Devices with Emitter, Control, and Collector Functions", IRE Proc. 47, 407-418; March 1959.

[4] L. J. Varnerin, "Stored Charge Method of Transistor Base Transit Analysis", IRE Proc. 47, 523-527; April 1959.

[5] R. Beaufoy, "Transistor Switching-Circuit Design Using the Charge-Control Parameters", IEE Proc. 106B, Suppl.17, 1085-1091; May 1959.

[6] J. J. Sparkes, "The Measurement of Transistor Transient Switching Parameters", IEE Proc. 106B, Suppl. 17, 562-567; May 1959.

[7] A. N. Baker, "Charge Analysis of Transistor Operation", IRE Proc. 48, 949-950; May 1960.

[8] J. J. Sparkes, "A Study of the Charge-Control Parameters of Transistors", IRE Proc. 48, 1696-1705; October 1960.

[9] A. N. Baker and W. G. May, "Charge Analysis of Transistor Operation Including Delay Effects", IRE Trans. on Electron Devices ED-8, 152-154; March 1961.

[10] A. R. Boothroyd, "Charge Definition of Transistor Properties", International Solid-State Circuits Conf. 1962, p. 30.

[11] R. A. Schmeltzer, "Transient Characteristics of Alloy Junction Transistors Using a Generalized Charge Storage Model", IEEE Trans. on Electron Devices ED-10, 164-170; May 1963.

[12] C. S. den Brinker, D. Fairbairn, and B. L. Norris, "An Analysis of the Switching Behaviour of Graded Base Transistors", Electronic Engineering 35, 500-505; August 1963.

[13] J. L. Moll and I. M. Ross, "The Dependence of Transistor Parameters on the Distribution of Base Layer Resistivity", IRE Proc. 44, 72-78; January 1956.

[14] J. M. Early, "Effects of Space-Charge Layer Widening in Junction Transistors", IRE Proc. 40, 1401-1406; November 1952.

[15] J. A. Narud, D. J. Hamilton and F. A. Lindholm, "Large Signal Models for Junction Transistors", International Solid-State Circuits Conf. 1963, p. 57.

[16] D. J. Hamilton, F. A. Lindholm, and J. A. Narud, "Comparison of Large Signal Models for Junction Transistors", IEEE Proc. 52, 239-248; March 1964.

[17] D. Koehler, "A New Charge Control Equivalent Circuit for Diodes and Transistors and Its Relation to Other Large Signal Models", International Solid-State Circuits Conf. 1965, p. 38.

[18] D. Koehler, "The Charge-Control Concept in the Form of Equivalent Circuits, Representing a Link between the Classic Large Signal Diode and Transistor Models", Bell Syst. Tech. J. 46, 523-576; March 1967.

[19] J. G. Linvill, "Lumped Models of Transistors and Diodes", IRE Proc. 46, 1141-1152; June 1958.

[20] J. J. Ebers and J. L. Moll, "Large-Signal Behavior of Junction Transistors", IRE Proc 42, 1761-1772; December 1954.

[21] R. D. Middlebrook, to be published.

[22] W. Shockley, "A Unipolar 'Field-Effect' Transistor", IRE Proc. 40, 1365-1376; November 1952.

[23] L. J. Sevin, Jr., Field-Effect Transistors , McGraw-Hill, 1965.

[24] F. N. Trofimenkoff and A. Nordquist, "F.E.T. Operation in the Pinchoff Mode", IEE Proc. 115, 496-502; April 1968.

[25] F. A. Lindholm, "Unipolar Transistors: Some Aspects Related to Their Inclusion in Electronics Curricula", IEEE Trans. on Education E-11, 25-33; March 1968.

[26] R. D. Middlebrook, "A Modern Approach to Semiconductor and Vacuum Device Theory", IEE Proc. 106B, Suppl. 17, 887-902; May 1959.

[27] J. Millman and S. Seely, Electronics, McGraw Hill, 1951.

[28] R. D. Middlebrook, "Conditions at a p-n Junction in the Presence of Collected Current", Solid-State Electronics 6, 555-571; November 1963.

[29] R. D. Middlebrook, "Effects of Modified Collector Boundary Conditions on the Basic Properties of a Transistor", Solid-State Electronics 6, 573-588; November 1963.

[30] W. Shockley and W. T. Read, Jr., "Statistics of the Recombination of Holes and Electrons", Phys. Rev. 87, 835-842; September 1952.

[31] P. E. Gray, D. DeWitt, A. R. Boothroyd, and J. F. Gibbons, "Physical Electronics and Circuit Models of Transistors", SEEC 2, John Wiley and Sons, 1964; Chapter 10.

[32] R. P. Nanavati, An Introduction to Semiconductor Electronics, McGraw Hill, 1963; Chapter 2.19.

[33] G. Bemski, "Recombination in Semiconductors", IRE Proc. 46, 990-1004; June 1958.

[34] C. Kittel, Introduction to Solid State Physics, 2nd Edition, John Wiley and Sons, 1956; Chapter 13, pp. 366-369.

[35] W. Shockley, Electrons and Holes in Semiconductors, D. Van Nostrand, 1956; Chapter 12.6, pp. 318-319.

[36] J. J. Sparkes, Junction Transistors, Pergamon Press, 1966, p.118.

[37] J. P. McKelvey, R. L. Longini, and T. P. Brody, "Alternative Approach to the Solution of Added Carrier Transport Problems in Semiconductors", Phys. Rev. 123, 51-57; July 1961.

[38] J. P. McKelvey, "Analysis of Semiconductor p-n Junctions and Junction Devices by a Flux Method", J. of Appl. Phys. 33, 985-991; March 1962.

[39] P. E. Gray, et al., op. cit., Chapter 6.

[40] R. P. Nanavati, op. cit., Chapter 3.8.

[41] A. B. Phillips, Transistor Engineering, McGraw Hill, 1963; Chapter 6.8.

[42] A. DeMari, "Accurate Numerical Solution of the One-Dimensional p-n Junction in Steady State", Electronics Letters 3, 142-144; April 1967.

[43] W. Shockley, M. Sparks, and G. K. Teal, "p-n Junction Transistors" Phys. Rev. 83, 151-162; July 1951.

[44] K. M. Van Vliet, "High Injection Theories of the p-n Junction in the Charge Neutrality Approximation", Solid-State Electronics 9, 185-201; March 1966.

[45] J. J. Sparkes, op. cit., Appendix D.

[46] C. L. Searle, A. R. Bothroyd, E. J. Angelo, Jr., P. E. Gray, and D. O. Pederson, "Elementary Circuit Properties of Transistors", SEEC 3, John Wiley and Sons, 1964; Chapter 2.1.

[47] A. K. Jonscher, <u>Principles of Semiconductor Device Operation</u>, G. Bell and Sons, 1960; Chapter 6.2, pp. 136-140.

[48] R. D. Middlebrook, unpublished work.