TRANSISTOR SWITCHING ANALYSIS

Thesis by

Carver A. Mead

In Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1960

ACKNOWLEDGEMENTS

I am deeply indebted to my many associates at both the California Institute of Technology and Pacific Semiconductors, Inc., for their assistance and encouragement in this work. Particular appreciation is expressed to Dr. R. D. Middlebrook and Dr. R. V. Langmuir for their constant support, to Mr. Richard Neville for his careful appraisal of the theoretical work, to Mrs. Joan Cartwright for her meticulous preparation of the manuscript, and to my family, without whose patient forbearance this work would never have been completed.

Carver A. Mead

ABSTRACT

With the widespread application of junction transistors in switching applications, the need for a general method of analysis useful in the region of collector voltage saturation has become apparent. Linear equivalent circuits using lumped elements have long been used for small signal calculations of normally biased transistors, but a comparable method for saturated transistors has been lacking. Recently Linvill[3] proposed the method of lumped models which allows the analysis of complex switching problems with the ease of linear circuit calculations. In Part I the method is shown to be equivalent to a well-known linear equivalent circuit under normal bias conditions. Examples of the application of the method and the use of approximations are drawn from practical circuit problems. Emphasis is placed upon the understanding of the physical phenomena involved, a necessary prerequisite to intelligent circuit design. In Part II the operation of transistors at relatively large collector currents is considered. Several effects which contribute to non-linear operation are analyzed. It is shown that at both high and low injection levels, a decrease in current gain is to be expected as a result of the self bias cutoff effect.

Emitter efficiency is shown to be important at high levels because of its contribution on the self bias effect.

The operation of both alloy and diffused base transistors as saturated switches is discussed, with particular reference to the non-linear effects. Theoretical expressions are derived for the saturation voltage as a function of collector and base current for both types.

TABLE OF CONTENTS

PART I. LINEAR OPERATION

PART II. NON-LINEAR THEORY

PART I

LINEAR OPERATION

INTRODUCTION

With the first analysis of a junction transistor triode, it was recognized that such a device was capable of symmetrical operation; that is, either the "emitter" junction or "collector" junction could act as a source of minority carriers in the base region. Thus modes of operation are available in a transistor which have never existed in the vacuum tube. For example, a saturated transistor (both emitter and collector forward biased) will carry signals well in both directions, while if both junctions are reverse biased, essentially no signal is allowed to pass in either direction. The inherently low voltage drop across a saturated transistor makes possible the control of very high powers with low dissipations. For these reasons, transistors find switching service a most important and useful application.

The first detailed analysis of the large signal properties of transistors was done by Ebers and Moll.[1] Later Linvill[3] proposed a technique by which the same results may be obtained, but which has the advantage that a linear model is used and physical insight into the behavior of the device is more readily gained. The purpose of the present article is to extend this method to the general treatment of diodes and transistors in practical circuit applications, to present results in special cases of importance, and to illustrate applications of the analysis in sufficient detail to be generally useful to the design engineer.

Let us examine the requirements on a method of analysis to be used in problems of this nature. Clearly, what is needed is a model, similar to an equivalent circuit such as used for small signal work, yet appropriate for all conditions encountered in transistor operation; i.e., either junction may be either forward or reverse biased, currents may be either small or large, a-c, d-c, or both. Together with a model, the method must include a procedure for evaluating elements in the model and for making approximations where an exact analysis would be too cumbersome.

Furthermore, in order to be of general utility, the method of analysis should possess the following qualities:

(a) The principal variables should be related in a linear manner so linear circuit theory may be used.

(b) Non-linearities in the system should be easily and accurately approximated by a simple piecewise linear idealization.

(c) Elements in the model should be readily obtainable in terms of simple, easily-measured device parameters.

(d) Variables in the model should possess physical significance, and results of the analysis should enhance one's physical insight into the problem.

(e) The model should reduce to familiar form for special cases, i.e., normal bias, small signal.

The variables necessary to solve normal semiconductor problems are:

1. Junction voltage
2. Junction current
3. Minority carrier density

provided we deal only with devices in which the diffusion current predominates. Transistors of this type are typically used in switching service where the method is most generally useful.

1. REVIEW OF BASIC PROCESSES

Before launching into the details of analysis, let us briefly review the basic processes which occur within a semiconductor. If we limit our discussion to one-dimensional diffusion flow, a complete description of the motion of minority carriers within the material consists of:

1. The continuity equation for minority carriers

$$\frac{\partial N}{\partial t} = \frac{N_o - N}{\tau} + D \frac{\partial^2 N}{\partial x^2}$$

where N is the density (number per unit volume) of minority carriers as a function of x and t

N_o is the density of minority carriers at thermal equilibrium

τ is the "lifetime" of minority carriers

x is the distance through the semiconductor

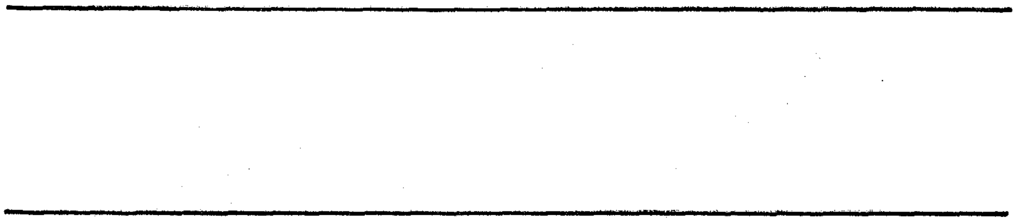
2. The condition that any macroscopic volume element of the material be electrically neutral.

2. INTRODUCTION TO LUMPED MODELS

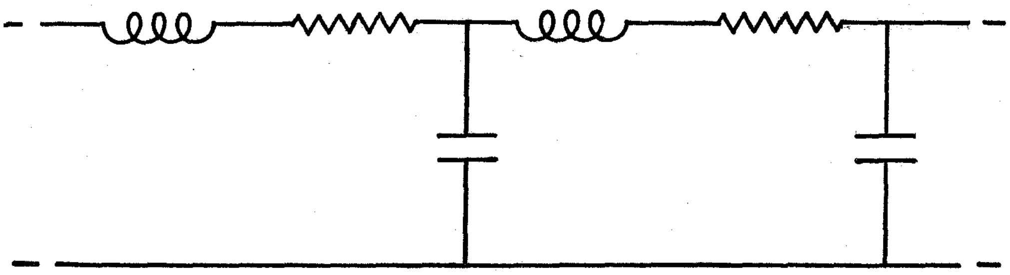
The continuity equation is a partial differential equation, involving both time and space derivatives, and its solutions are in general both difficult and messy, resulting in carrier densities at all points as a function of time.

One is reminded of the transmission line problem where again a partial differential equation must be solved and the results are voltage (or current) at any point on the line as a function of time. Pursuing the analog still further, the analysis of a transmission line is greatly simplified if we are content to find approximate voltages and currents at certain specified points along the line instead of exact voltages and currents at all points along the line. Since normally only the ends of the line are of interest, such a procedure seems highly desirable. To this end we approximate the line by a ladder network as shown in Fig. 1.

The line possesses a series inductance and resistance per unit length, and a shunt capacitance per unit length. These distributed parameters are represented in the ladder network by the lumped elements L, R and C. Thus we have created a lumped model as an approximation to the real transmission line. The voltage at any node, or the current through any element may be obtained by standard circuit analysis techniques. Hence, we have transformed a problem in partial differential equations into a problem in simple circuit theory.



Transmission Line



Lumped Model

Fig. 1. - Transmission line and equivalent lumped element representation.

Each section of the lumped model corresponds to a given length of transmission line. As the number of sections is increased, the length of line to which each section corresponds is decreased and the accuracy of the approximation is improved. In the limit, as the number of sections becomes infinite, each section represents an infinitesimal length of line and we are again faced with the solution of a partial differential equation. Perhaps the most significant feature of the lumped model is that the approximations have been made before any equations were written, and each element in the model has definite physical significance.

3. SEMICONDUCTOR LUMPED MODEL

In order to simplify our expressions for minority carrier flow, it is convenient to write the continuity equation in terms of a new variable, the excess density, $N - N_o$.

$$\frac{\partial (N - N_o)}{\partial t} = - \frac{(N - N_o)}{\tau} + D \frac{\partial^2 (N - N_o)}{\partial x^2}$$

For a p-type semiconductor: $N = n$, $N_o = n_p$

Therefore,

$$\frac{\partial \eta}{\partial t} = \frac{\eta}{\tau} + D \frac{\partial^2 \eta}{\partial x^2} \qquad \eta = n - n_p$$

For an n-type semiconductor: $N = p$, $N_o = p_n$

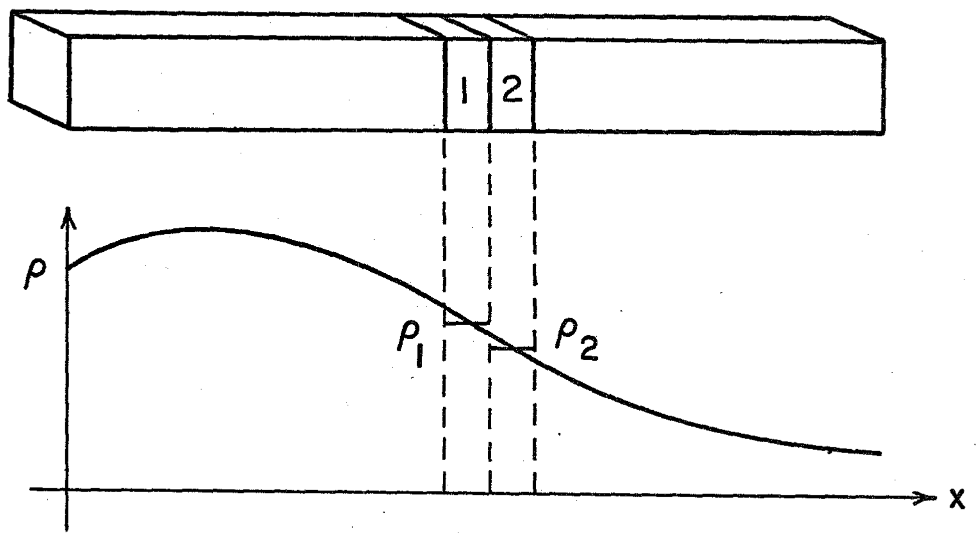


Fig. 2. - Semiconductor bar with excess minority carrier distribution.

Therefore,

$$\frac{\partial \rho}{\partial t} = -\frac{\rho}{\tau} + D\frac{\partial^2 \rho}{\partial x^2} \qquad \rho = p - p_n$$

Now let us examine the physical significance of each term in the continuity equation. Consider a long bar of n-type semiconductor with unit cross-sectional area containing volume elements 1 and 2.

Suppose within the bar there exists a distribution of excess holes $\rho(x)$ as shown in Fig. 2, which is a function of x but independent of y and z. The volume elements will possess average excess densities ρ_1 and ρ_2 respectively.

<u>Storage.</u> Due to the charge neutrality requirement, an excess density of minority carriers implies an equal excess density of majority carriers. Thus a <u>change in minority excess density with time produces a majority carrier current into the volume element.</u> Such a change in density with time is represented by the first term in the continuity equation, and the resulting current is the <u>rate of change of stored charge</u> in the volume element.

<u>Recombination.</u> According to the simple linear recombination law, the recombination rate is proportional to the excess density. Again the charge neutrality condition requires that if the minority carrier density remains fixed, for each recombination a new majority carrier must enter the volume element. However, in order for the minority carrier density to remain fixed, a new minority carrier must also enter the volume element. Thus, since

the carriers possess opposite charges, the net effect of recombination is to bring minority carrier current into the volume element and force an equal majority current out of the volume element. This effect is represented by the second term in the continuity equation.

Diffusion. If the slope of the carrier excess density curve is greater at the left boundary of the volume element than at the right boundary, there will be a net minority carrier diffusion current into the volume element. This effect is represented by the last term in the continuity equation. Since the diffusion current is proportional to the gradient of the excess density, the current flowing from element 1 into element 2 will be approximately proportional to $(\rho_1 - \rho_2)$.

We may state the continuity equation, then, as follows:

> The sum of currents flowing out of a given volume element due to
>
> a) change in stored charge
> b) recombination of carriers
> c) minority carrier diffusion
>
> is equal to zero.

When so stated it is clear that the continuity equation is merely Kerchhoff's law for a continuous system. However, we have seen that such a system may be approximated by a lumped model in which the network node equations replace the continuity equation. We may therefore construct such a model for our bar of semiconductor. Lumped models of this type were first proposed by Linvill.[3]

Lumped Model Elements. The elements required for such a lumped model are clearly of a different nature than those of electric circuit theory. The variables are excess density and current rather than voltage and current. It should always be borne in mind that the semiconductor lumped model is analogous, not identical, to its corresponding electric circuit. However, the same methods are available for the solution of either type of problem.

The types of elements required may be summarized as in Table I. Clearly the excess densities are the direct analogies of voltages in an electrical circuit.

The lumped model for our bar of semiconductor may be constructed in exactly the same manner as for the transmission line, as shown in Fig. 3. Here ρ_1 and ρ_2 are the average excess hole densities in volume elements 1 and 2. G_1 and C_1 represent recombination and storage in volume element 1, while G_2 and C_2 represent recombination and storage in volume element 2. G_{12} represents the diffusion of holes from volume element 1 into volume element 2. It will be noticed that the lines are not wires but are defined as regions of constant excess density. The bottom line is defined as the zero excess density, and current flowing in this line is majority carrier current. This convention follows from the definitions of the elements and the effects which they represent. Although the lumped model elements are linear, it should be borne in mind that the excess density ρ is constrained to be greater than $-p_n$ at all times, since the actual density cannot become negative. This restriction will introduce nonlinearities later when specific problems are considered.

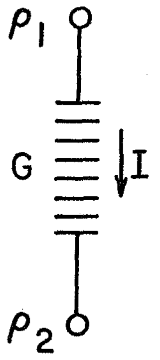
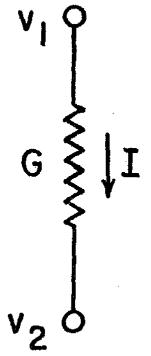
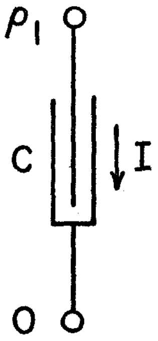
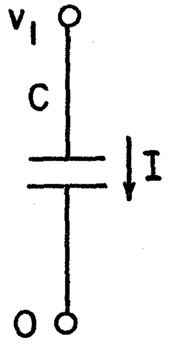
Symbol for Element	Electrical Analogy
	
$I = G(\rho_1 - \rho_2)$	$I = G(v_1 - v_2)$
	
$I = C \frac{d\rho_1}{dt}$	$I = C \frac{dv_1}{dt}$

TABLE I. - Semiconductor lumped elements with electrical analogies.

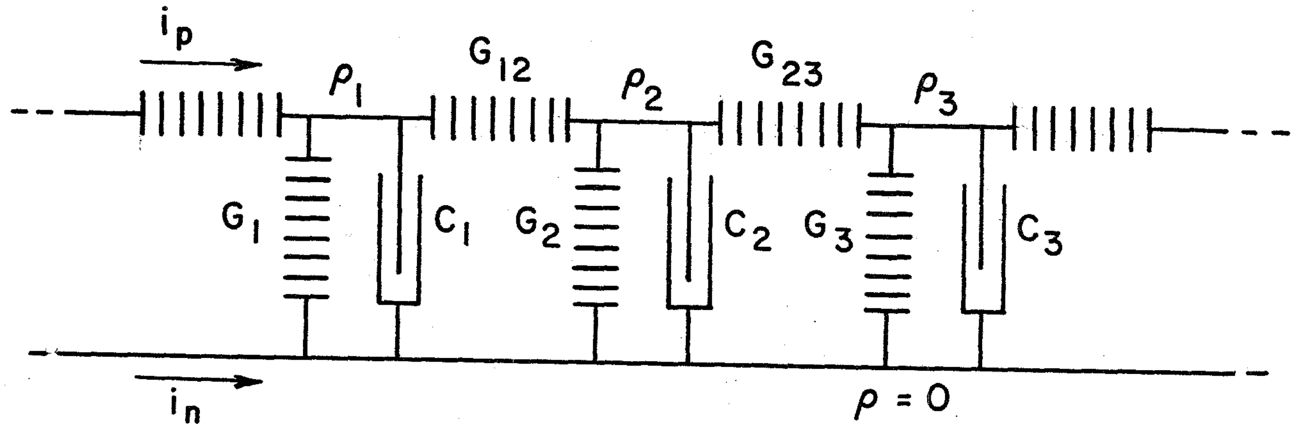


Fig. 3. - Lumped model of semiconductor bar.

As a check on the lumped model, let us assume a uniform excess minority carrier density in the semiconductor bar. Under these conditions the continuity equation becomes

$$\frac{\partial \rho}{\partial t} = -\frac{\rho}{\tau}$$

which has the solution

$$\rho = \rho_o e^{-t/\tau}$$

From the lumped model

$$\rho_1 = \rho_2 = \rho_3 \quad \text{etc.}$$

Thus no current flows in the diffusion conductances and we may write Kerchhoff's law as, for example

$$\rho_2 G_2 + C_2 \frac{\partial \rho_2}{\partial t} = 0$$

or

$$\frac{\partial \rho_2}{\partial t} = -\frac{G_2}{C_2} \rho_2$$

which has the solution

$$\rho_2 = \rho_o e^{-\frac{G_2}{C_2} t}$$

and similarly for the other junctions. Thus the lifetime τ may be identified with the time constants of the lumped model parallel elements

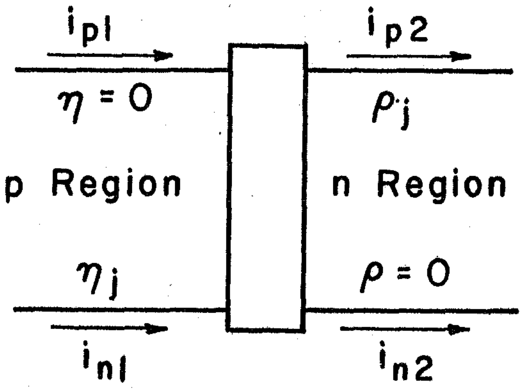
$$\frac{1}{\tau} = \frac{G_1}{C_1} = \frac{G_2}{C_2} = \frac{G_3}{C_3} \quad \text{etc.}$$

It is now clear that the basic diffusion process within the semiconductor is extremely simple, analogous to an RC ladder network where all capacitors are connected to ground, and all shunt elements have time constant τ. No inductive elements are present, since the diffusion process cannot support a propagating wave. Our lumped approximation is complete and we are in a position to solve any problem where the initial conditions or boundary conditions on the excess density are known, by the application of simple circuit theory.

Boundary Conditions, the p-n Junction. In a semiconduotor device such as the transistor, the boundary conditions on minority carrier excess density are generally determined by the voltage across certain p-n junctions in the structure. Depending upon the type of semiconductor material, temperature, operating current density and other factors, the value of excess density determined by a junction with a certain applied potential may vary. However, in most instances the recombination within the junction is small and the following expressions are sufficiently accurate for practical application:

$$\rho_j = p_n \left(e^{qv/kT} - 1\right) \tag{1a}$$

$$\eta_j = n_p \left(e^{qv/kT} - 1\right) \tag{1b}$$



$$i_{p1} = i_{p2}$$

$$i_{n1} = i_{n2}$$

$$\rho_j = p_n(e^{qv/kT} - 1)$$

$$\eta_j = n_p(e^{qv/kT} - 1)$$

Fig. 4. - Model of p-n junction.

where

ρ_j	is the excess hole density at the n side of the junction
η_j	is the excess electron density at the p side of the junction
k	is Boltzmann's constant
T	is the absolute temperature
v	is the voltage across the junction, being taken as positive when the junction is forward biased

In this approximation the hole and electron currents entering the junction from one region emerge undiminished at the opposite side of the junction. Thus the junction provides an excess density essentially independent of current, or acts as an excess density generator controlled by the junction voltage. The analogy in electric circuit theory would be a voltage generator controlled by some other variable in the system, as frequently used in small signal equivalent circuits for active devices. The symbol for such a junction is shown in Fig. 4. Here again the top lines carry hole current, and the bottom lines carry electron current.

4. P-N DIODE

A p-n junction diode consists not only of the p-n junction itself but also of the adjoining p and n regions. Thus the diode model must include lumped element network sections for both regions as well as the junction "excess density generator". The number of sections used for each region is determined by the accuracy required and the fortitude of the analyst. An appropriate compromise is usually obtained at one or two sections per region. A two-section model is shown in Fig. 5a. As before, the junction boundary densities ρ_1 and η_1 are determined by the junction voltage through the exponential junction law.

In cases where the conductivity of one region is much greater than that of the other, carrier injection into the high conductivity region is small and may be neglected. For example, if the p region is of very high conductivity compared with the n region, n_p is very small compared to p_n, (since $p_p n_p = n_i^2$, a constant for given temperature). Thus η_1 will be very small compared to ρ_1, and only hole current need be considered. Under these conditions, considerable simplification in the lumped model is possible, i.e., the elements corresponding to the p region may be deleted, as shown in Fig. 5b.

For many practical applications, sufficient accuracy is obtained by the use of only one lumped section, as shown in Fig. 5c. In this model the capacitance C represents minority carriers stored near the junction and the conductance G represents the combined effect of recombination near the junction and diffusion away from the junction and subsequent recombination. As will be seen, the great virtue of the single section model is the ease with which the element values are determined.

<u>Diode Small Signal Response.</u> If the diode is forward biased and a small a-c signal $v_1 e^{j\omega t}$ is superimposed on the d-c bias voltage v_o, we may write the injected excess density

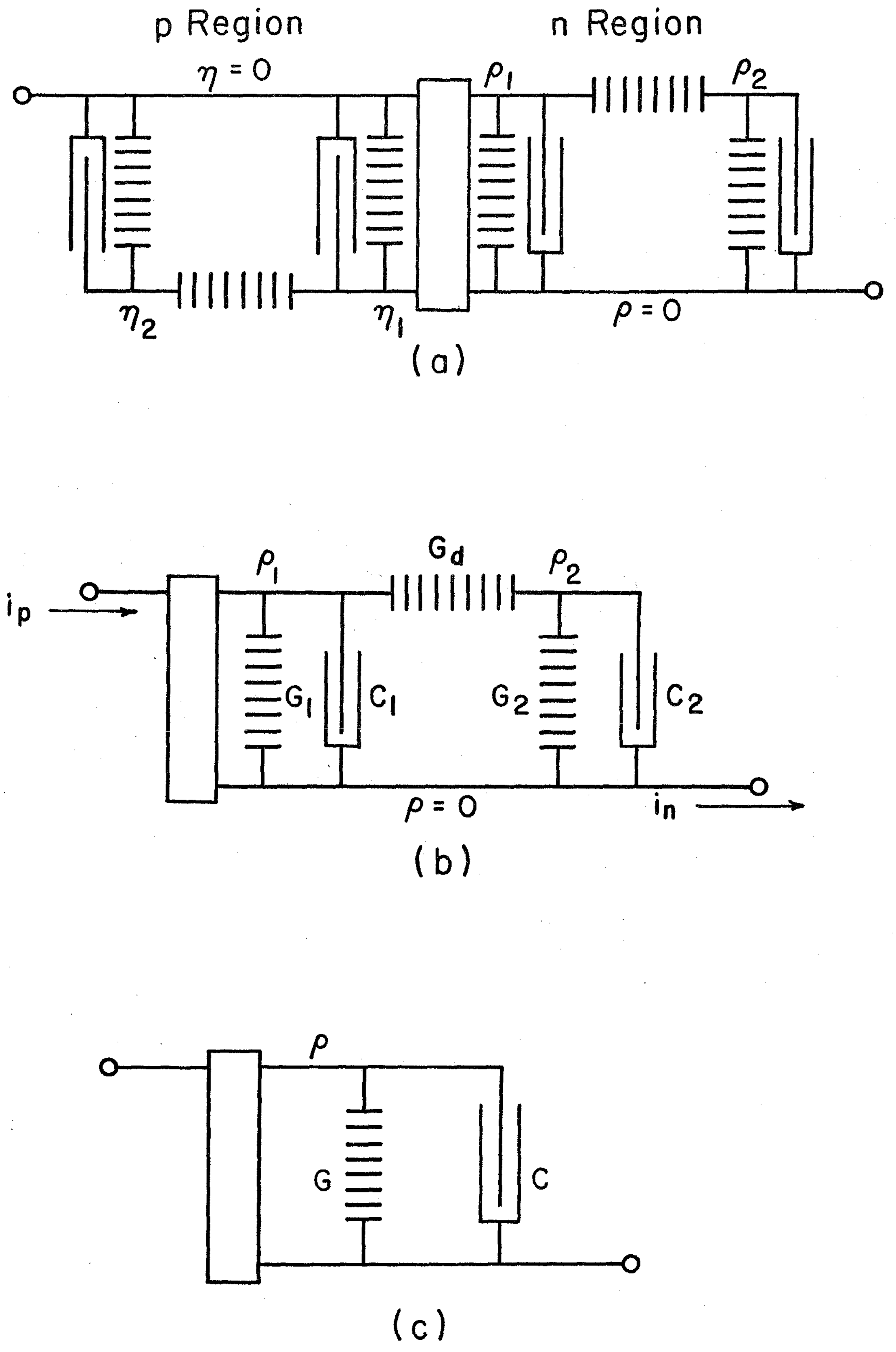


Fig. 5. - Development of diode model.

$$\rho = p_n(e^{qv/kT} - 1)$$

where $v = v_o + v_1 e^{j\omega t}$.

Now if $v_1 << kT/q$ we may expand the exponential and retain only the first two terms

$$\rho \approx p_n \left[e^{qv_o/kT} (1 + \frac{q}{kT} v_1 e^{j\omega t}) - 1 \right]$$

Thus the a-c component of excess density is

$$\rho_1 e^{j\omega t} \approx \frac{q}{kT} v_1 p_n e^{qv_o/kT} e^{j\omega t}$$

if the d-c current through the diode is much greater than its saturation current, the excess density will be much greater than p_n, and the exponential term in the expression for excess density will be large compared to unity. Thus the d-c component of excess density becomes

$$\rho_o \approx p_n e^{qv_o/kT} \tag{2}$$

Combining these two expressions, the amplitude of the a-c excess density is

$$\rho_1 \approx \rho_o \frac{q}{kT} v_1 \tag{3}$$

From the lumped model of Fig. 5c, the current through the diode is composed of a d-c bias component

$$I = \rho_o G$$

and an a-c component

$$i_1 = \rho_1 (G + j\omega C)$$

But from equations 2 and 3

$$i_1 = \frac{qI}{kT} (1 + j\omega \frac{C}{G}) v_1$$

Thus the time constant C/G may be determined by one simple small signal measurement on the diode, i.e., the impedance rolloff frequency

$$\omega_h = \frac{G}{C}$$

To determine the magnitude of G we may measure the d-c junction voltage required to produce a certain current. If the current is much larger than the saturation current of the diode

$$I = \rho G \approx p_n G\, e^{qv/kT}$$

or

$$p_n G = I\, e^{-qv/kT}$$

It will be noticed that we are not able to determine either p_n or G individually from external measurements on the diode, but only their product. However, for this reason we need not know the individual values if we are interested only in voltages and currents, since expressions for these external variables involve only the product. Hence our two simple measurements on the diode, small signal impedance rolloff frequency and d-c voltage drop for a given d-c current, serve to completely determine the lumped model which is now applicable to all operating conditions.

Diode Recovery. Suppose we connect a junction diode in the circuit shown in Fig. 6. If the applied voltage V remains at $+V_1$ for a sufficient time, the current i will reach a steady state value very nearly V_1/R, assuming the forward voltage drop across the diode is small compared with circuit voltages. If now the applied voltage is abruptly changed to $-V_2$, the current is observed to assume a nearly constant value $-V_2/R$ for a storage time t_s, after which it decays rapidly to its small steady state reverse value i_s. The explanation for this action is as follows:[5]

At $t = 0$, the excess density of minority carriers near the junction is as shown by the top curve of Fig. 7. At a slightly later time, the applied voltage has reversed, but the stored carriers have not had time to recombine or diffuse away. The rate at which they may cross the junction is limited to a maximum reverse current of V_2/R, since a higher value would appreciably forward bias the junction and hence be self-annihilating. However, since the excess density of minority carriers at the junction is greater than zero, the junction must remain slightly forward biased, and the voltage across the junction remains small. As long as the junction remains forward biased, the current is determined by the external circuit. In this case, the current during the storage period is very nearly $-V_2/R$. This condition determines the slope of minority carrier density at

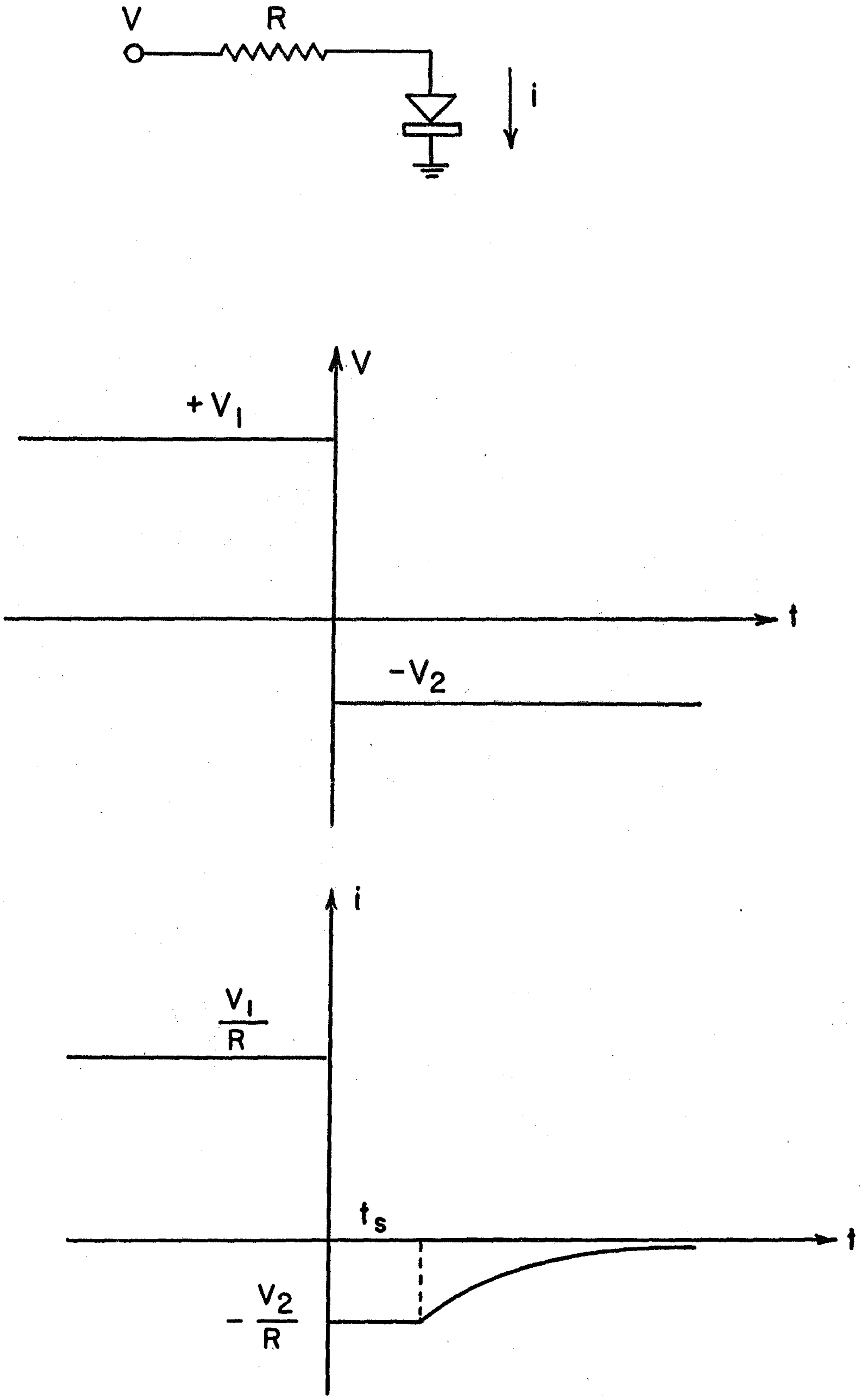


Fig. 6. - Recovery of p-n diode from step input.

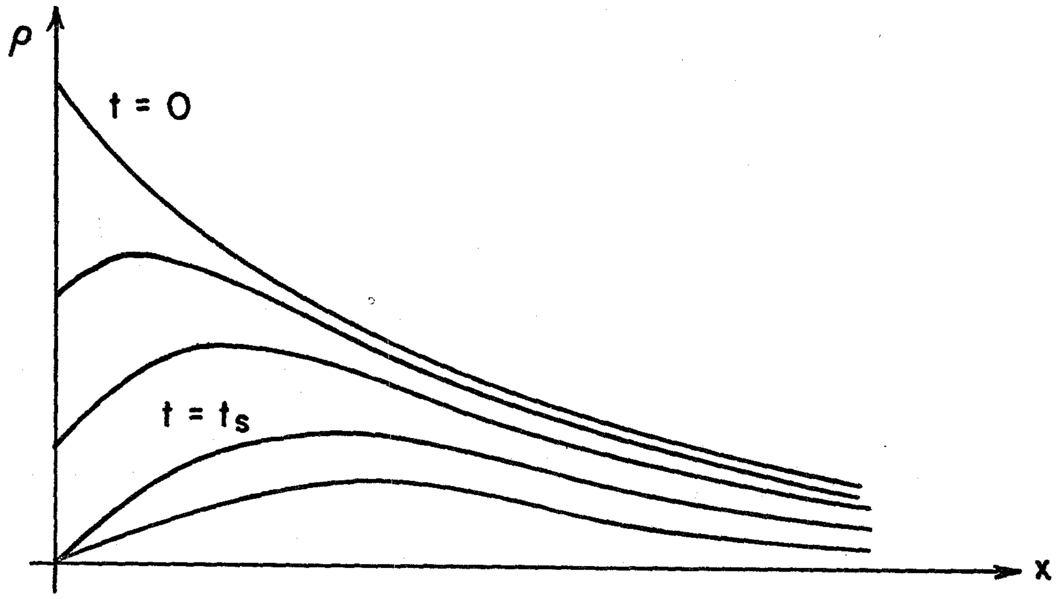


Fig. 7. - Excess density distribution in p-n diode during recovery.

the junction, which is proportional to the junction current. At the end of the storage time t_s the junction excess density has just reached zero and approaches $-p_n$ as the junction becomes reverse biased. Since the junction excess density is essentially constant (nearly zero) for any reverse bias voltage, the current is no longer affected by the junction voltage but is determined only by the conditions within the semiconductor. Therefore, the decay time is characterized by the minority carrier lifetime.

Diode Storage Time Calculation. We may obtain a good estimate of the recovery time from a simple calculation using the one section lumped model of Fig. 5c. Although the results are not nearly as accurate as those which will shortly be obtained for the transistor, they are significant and the method used is typical of all lumped model calculations.

(a) Steady State

Since the diode is forward biased, the voltage across it is small compared with V_1 and we may assume

$$\frac{V_1}{R} \approx I = G\rho_{ss}$$

The voltage across the diode may be found by using the value of ρ_{ss} obtained from this equation and the exponential junction law.

(b) Storage Period

When the applied voltage changes sign, the charge on C cannot change instantaneously, and hence ρ must be continuous. As long as ρ is greater than zero, the diode is forward biased and the

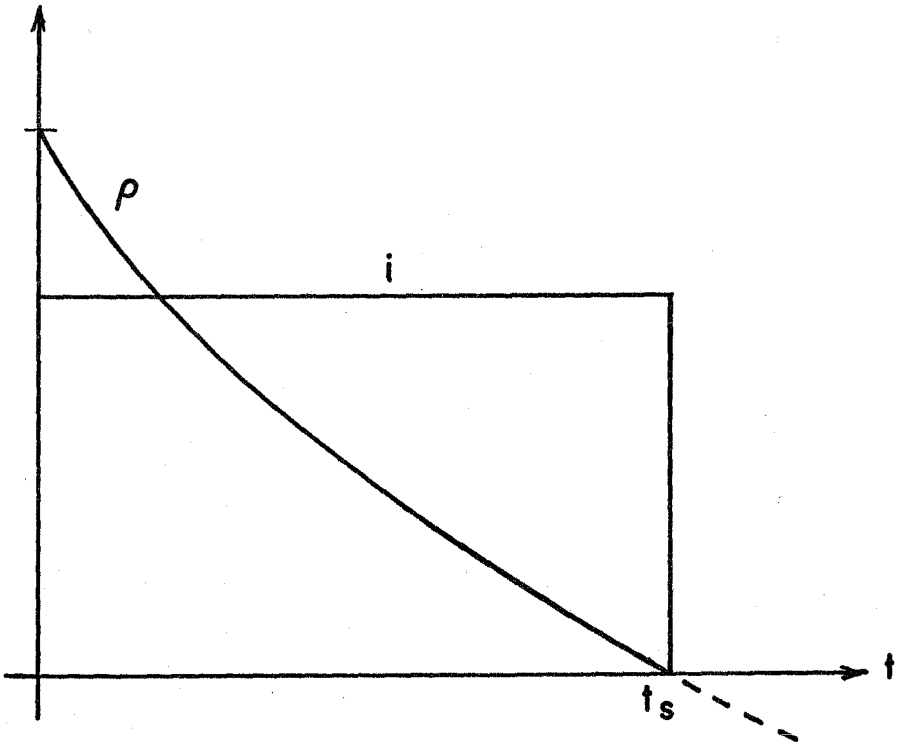


Fig. 8. - Recovery characteristic predicted by simple single section model.

voltage across it is small. Hence the current is very nearly $-V_2/R$. To determine the storage time t_s, we find ρ as a function of time assuming the current remains at $-V_2/R$. Then, when ρ becomes very nearly zero, the junction voltage may assume any negative value, and the current approaches zero. To obtain the complete solution for ρ as a function of time, we superpose the steady state value upon the response to a negative current step of magnitude $(V_1 + V_2)/R$, with the result

$$G = \frac{V_1}{R} - \frac{V_1 + V_2}{R}\left(1 - e^{-\omega_h t}\right)$$

At $t = t_s$, the excess density has just reached zero.

$$t_s = \frac{1}{\omega_h} \operatorname{Ln}\left(\frac{V_1 + V_2}{V_2}\right) \tag{4}$$

Since there is no longer any charge on C, the current will abruptly stop at $t = t_s$, according to the lumped model. In reality, we know it dies away smoothly. Hence, in this particular case, the lumped model has predicted zero decay time, which may be explained as follows: Since we have used only one lumped section, we have not included the effect of carriers at some distance from the junction, and it is just these carriers which diffuse back to the junction and cause the exponential type decay. With any finite number of lumped sections, the calculated current will exhibit a discontinuity since ρ at the junction approaches zero with a finite $d\rho/dt$, and causes current through the capacitor nearest the junction which must abruptly cease as ρ is clamped at $-p_n$. This lack of accuracy was introduced because of our attempt to use one lumped section to approximate too large a region of the semiconductor. The problem is especially bad in fast diodes since the lifetime is made very short and consequently the region to be represented by the lumped model is normally many diffusion lengths. On the other hand, in a transistor the base region is only a very small fraction of a diffusion length and the accuracy obtained is very much better. The justification for using the single section diode model is the ease of determining the element values. In practical design work the choice is usually not between a simple or elegant analysis but rather between a simple analysis or none at all.

5. TRANSISTOR LUMPED MODEL

If the transistor emitter and collector conductivities are high compared to the base conductivity, we may neglect any carrier injection into the emitter and collector and consider only minority carriers in the base region. A two-section model results as shown in Fig. 9. In this figure ρ_1 represents the excess density near the emitter, given by

$$\rho_1 = p_n \left(e^{qv_{eb}/kT} - 1 \right)$$

where v_{eb} is the emitter-base junction voltage; C_1 and G_1 represent storage and recombination near the emitter, G_d represents diffusion from emitter to collector, ρ_2 is the excess density near the collector, given by

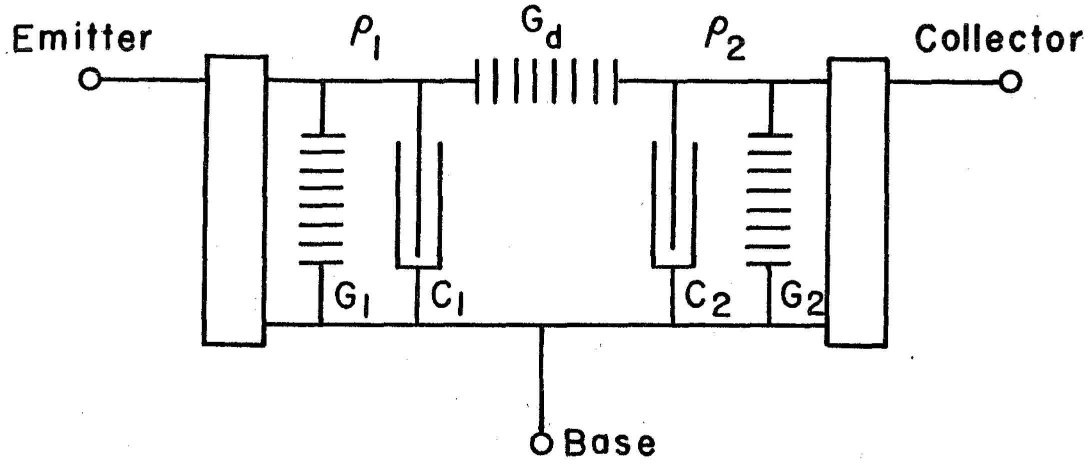


Fig. 9. - Complete two-section lumped model of junction transistor with high conductivity collector and emitter regions.

$$\rho_2 = p_n \left(e^{qv_{cb}/kT} - 1 \right)$$

where v_{cb} is the collector-base junction voltage; C_2 and G_2 represent storage and recombination near the collector. Both voltages are taken positive when the junction is forward biased.

When the transistor is normally biased as shown in Fig. 10a, $\rho_2 = - p_n$ = constant and thus no a-c current flows through G_2 or C_2. Therefore, the a-c collector current

$$i_c = \rho_1 G_d$$

and the a-c base current

$$i_b = \rho_1 (G_1 + sC_1)$$

using the Laplace transform notation. Hence

$$\frac{i_c}{i_b} = \frac{G_d}{G_1 + sC_1} = \frac{G_d/G_1}{1 + s\frac{C_1}{G_1}} = \frac{\beta}{1 + \frac{s}{\omega_\beta}}$$

where β is the low frequency, short circuit, common emitter, current gain and ω_β is the short circuit, common emitter, current gain, cutoff frequency. Thus

$$\beta = \frac{G_d}{G_1}$$

$$\omega_\beta = \frac{G_1}{C_1}$$

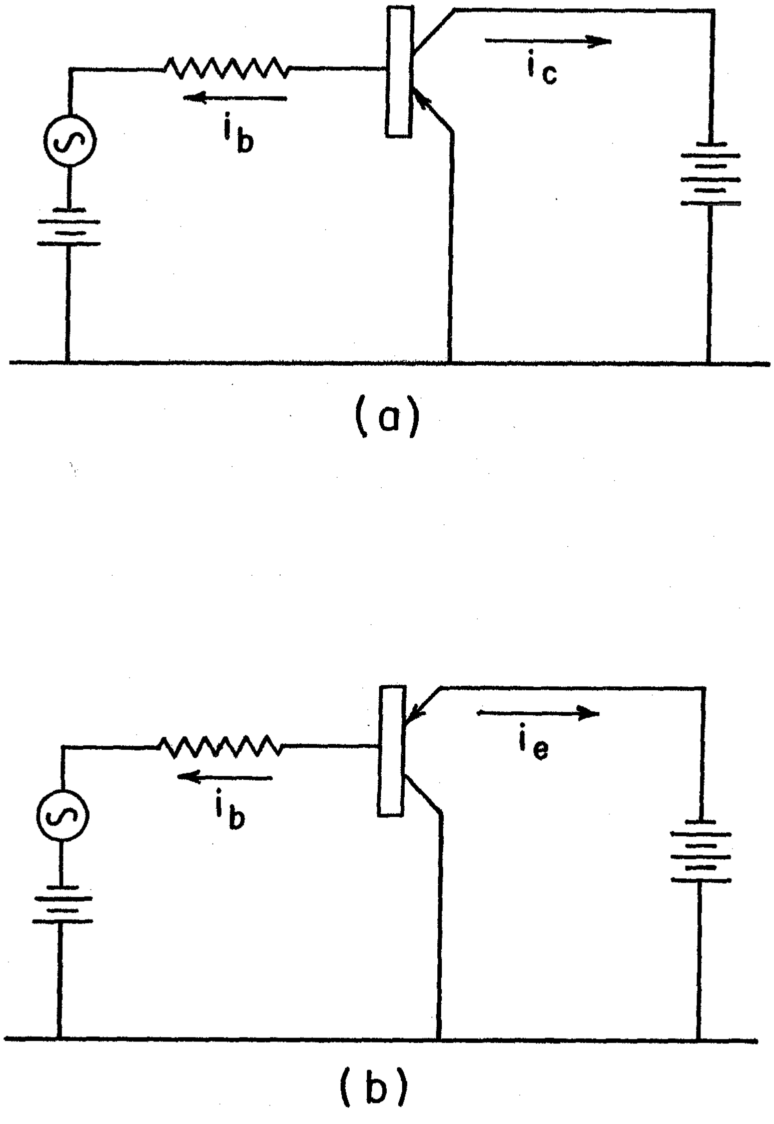


Fig. 10. - p-n-p transistor in (a) normal and (b) inverse connection.

Since for all reasonable transistors $\beta >> 1$, in all cases of interest $G_d >> G_1$. If the transistor is inverted, i.e., the collector forward biased (acting as an emitter) and the emitter reverse biased (acting as a collector) as shown in Fig. 10b, $\rho_1 = -p_n =$ constant and no a-c current flows through G_1 or C_1.

The a-c emitter current

$$i_e = \rho_2 G_d$$

and the base current

$$i_b = \rho_2 (G_2 + C_2 s)$$

Hence

$$\frac{i_e}{i_b} = \frac{G_d}{G_2 + C_2 s} = \frac{G_d/G_2}{1 + s\,\dfrac{C_2}{G_2}} = \frac{\beta_i}{1 + \dfrac{s}{\omega_{\beta i}}}$$

where β_i is the inverted current gain and $\omega_{\beta i}$ is the inverted cutoff frequency. Thus

$$\beta_i = \frac{G_d}{G_2}$$

$$\omega_{\beta i} = \frac{G_2}{C_2}$$

Since β_i is often quite small, we may make no statement with regard to the relative magnitude of G_2 and G_d.

By the four simple measurements of β, β_i, ω_β and $\omega_{\beta i}$, we are able to determine all of the lumped elements in terms of one (preferably G_d). For many calculations we need not proceed further. However, if we are interested in the voltage across a forward biased junction, we need to determine $p_n G_d$. As in the case of the diode, we cannot determine either p_n or G_d separately by external measurements. Perhaps the best method of determining $p_n G_d$ is to measure the d-c emitter-base voltage v_{eb} and the d-c collector current I_c in the normal bias connection.

$$I_c = (\rho_1 + \rho_n) G_d = p_n G_d \, e^{qv_{eb}/kT}$$

$$p_n G_d = I_c \, e^{-qv_{eb}/kT} \qquad (5)$$

For germanium transistors it is also possible to obtain an approximate value of $p_n G_d$ from a measurement of i_{co}, the collector cutoff current when the emitter is open circuited.

$$i_{co} = p_n \left(G_2 + \frac{G_1 G_d}{G_1 + G_d} \right) = p_n G_d \left(\frac{1}{\beta_i} + \frac{1}{1+\beta} \right) \approx p_n G_d \left(\frac{1}{\beta_i} + \frac{1}{\beta} \right)$$

hence

$$p_n G_d \approx \frac{i_{co} \beta \beta_i}{\beta + \beta_i} \qquad (6)$$

However, i_{co} is normally composed of a certain amount of surface leakage current and junction depletion layer generation current. Hence the value of $p_n G_d$ obtained in this way may be considerably in error. A measurement of v_{eb} and i_c at a bias current large compared to i_{co} is much to be preferred. In silicon units the i_{co} is largely determined by carrier generation within the junction depletion region and therefore should never be used in the determination of $p_n G_d$.

Transistor Small Signal Performance. We have already used the common emitter current gain characteristic of the transistor in order to determine the values of the lumped model elements. It is of interest to investigate the other aspects of small signal performance as predicted by the lumped model. If the transistor is used in the common base connection and normally biased, the a-c collector current

$$i_c = \rho_1 G_d$$

and the a-c emitter current

$$i_e = \rho_1 (G_d + G_1 + C_1 s)$$

Therefore

$$\frac{i_c}{i_e} = \frac{G_d}{G_d + G_1 + C_1 s} = \frac{\dfrac{G_d}{G_d + G_1}}{1 + \dfrac{C_1}{G_d + G_1}} = \frac{\alpha}{1 + \dfrac{s}{\omega_\alpha}}$$

where

$$\alpha = \frac{G_d}{G_d + G_1} = \frac{\beta}{1 + \beta}$$

$$\omega_\alpha = \frac{G_d + G_1}{C_1} = (1 + \beta)\, \omega_\beta$$

Thus the lumped model gives the single time constant approximation for the current gain which is quite accurate for operation well below the alpha cutoff frequency and is commonly used for high frequency calculations. If $\beta >> 1$, we may simplify the last expression as follows:

$$\omega_\alpha \approx \frac{G_d}{C_1} = \beta\omega_\beta$$

In the inverse common base connection similar expressions apply.

$$i_e = \rho_2 G_d$$

$$i_c = \rho_2 (G_d + G_2 + C_2 s)$$

$$\frac{i_e}{i_c} = \frac{G_d}{G_d + G_2 + C_2 s} = \frac{\frac{G_d}{G_d + G_2}}{1 + \frac{C_2}{G_d + G_2} s} = \frac{\alpha_i}{1 + \frac{s}{\omega_{\alpha i}}}$$

$$\alpha_i = \frac{G_d}{G_d + G_2} = \frac{\beta_i}{1 + \beta_i}$$

$$\omega_{\alpha i} = \frac{G_d + G_2}{C_2} = (1 + \beta_i)\,\omega_{\beta i}$$

where β_i is not necessarily large compared to unity.

We may now ask what type of complete common emitter, small signal equivalent circuit results from the lumped model. For small signals, the emitter a-c minority carrier density is proportional to the a-c emitter-base voltage as shown in equation 3.

$$\rho_1 = \frac{q}{kT} \rho_o v_{be} = K_1 v_{be}$$

The effective a-c density at the collector is also proportional to the a-c collector-base voltage, due to collector depletion layer widening or Early effect.[2, 6]

$$\rho_2 = K_2 v_{cb} \qquad K_2 << K_1$$

where all quantities of interest are now a-c components.

In the common emitter connection, expressions for the base and collector currents become

$$i_c = \rho_1 G_d + \rho_2 (G_d + G_2 + C_2 s)$$

$$i_b = \rho_1 (G_1 + C_1 s) + \rho_2 (G_2 + C_2 s)$$

Since the emitter voltage is taken as zero, we may rewrite the currents in terms of collector and base voltages.

$$i_c = v_{be} \left[- K_1 G_d + K_2 (G_d + G_2 + C_2 s) \right] - K_2 (G_d + G_2 + C_2 s) v_c$$

$$i_b = - v_{be} \left[K_1 (G_1 + C_1 s) - K_2 (G_2 + C_2 s) \right] + K_2 (G_2 + C_2 s) v_c$$

These equations correspond to the circuit shown in Fig. 11.

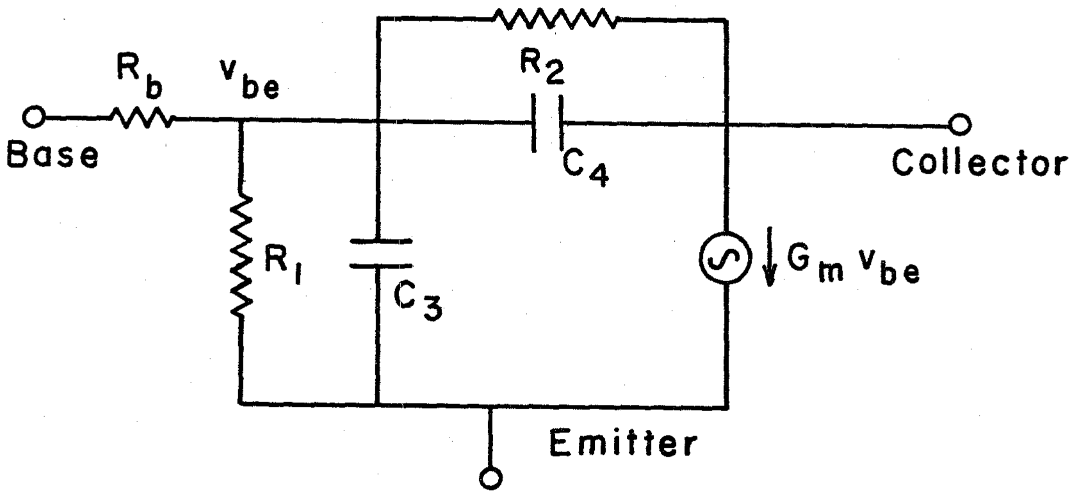


Fig. 11. - Transistor small signal equivalent circuit as developed from lumped model.

where

$$R_1 = \frac{1}{K_1 G_1} \qquad R_2 = \frac{1}{K_2 G_2}$$

$$C_3 = \frac{K_1}{C_1} \qquad C_4 = \frac{K_2}{C_2}$$

$$g_m = K_1 G_d$$

The extrinsic base resistance R_b must be added in series with the base terminal as shown. This circuit is similar to that proposed by Giacoletto[7] and widely used for high frequency work. Thus, the lumped model reduces simply to a quite accurate representation in this special case, yet is much more general in that it is useful for all transistor operating conditions.

6. TRANSISTOR SWITCHING PERFORMANCE

It was recognized early in the development of the junction transistor that the symmetry of the device implied the unique ability to saturate. Since both the collector and emitter junctions are capable of emitting minority carriers into the base region when forward biased, and since the diffusion current across the base is due to the difference in minority carrier density between the two junctions, it is clear that when the transistor is saturated current may flow in either direction, depending upon which density is the larger. In the saturated condition the transistor closely resembles a closed switch, and dynamic resistances of a few tenths of an ohm are easily obtained with small units. If, on the other hand, both junctions are reverse biased, the transistor is cut off and only the very small junction reverse currents flow. Hence the transistor resembles an open circuit, impedances of several megohms being common. When a transistor spends the majority of its time in one of two states, fully saturated or fully cut off, passing through the normally biased region only to get from one state to the other, it is said to be operating as a switch. A distinction should be drawn at this point between true switching service and non-saturating service. A so-called non-saturating switch is one where the transistor may become either cut off or normally biased (usually with a rather low collector voltage), but not saturated. Such operation is more properly termed Class C pulse amplifier service. When the transistor is not caused to saturate, its operation may be analyzed with adequate accuracy by the use of small signal equivalent circuits. However, when the transistor is caused to saturate, the analysis becomes quite complicated and has traditionally been avoided in circuit work. Here the lumped model comes into its own, since it transforms the difficult problem of transistor saturation and storage into one of simple R-C circuit analysis. The advantage of this approach for the circuit engineer is obvious.

Quasi Static Performance. Consider a transistor connected in the circuit shown in Fig. 12. For this analysis, we shall use the lumped model of Fig. 9. However, we may ignore C_1 and C_2 since we are only interested in slowly varying d-c quantities. The three regions of transistor operation will now be considered:

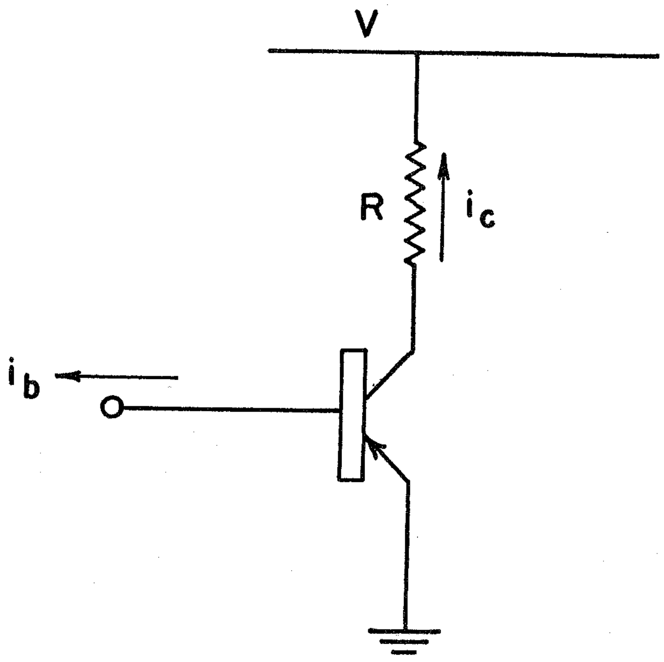


Fig. 12. - Elementary transistor switch.

(a) Cut-Off. In the cut-off state, $\rho_1 = \rho_2 = -p_n$. Therefore

$$i_c = p_n G_2 = \frac{i_{co}(1+\beta)}{1+\beta+\beta_i}$$

which becomes

$$i_c \approx \frac{i_{co}}{1+\beta_i/\beta} \qquad \text{if} \qquad \beta >> 1 \tag{7}$$

In the usual case, i_{co} is very small and hence the collector voltage is approximately equal to the supply voltage

$$v_c \approx V$$

The base current

$$i_b = -p_n(G_1 + G_2) \approx -i_{co} \qquad \text{if} \qquad \beta >> 1 \tag{8}$$

In this state the only significant contribution to the power dissipation of the transistor is from the collector.

$$P \approx i_c V = \frac{i_{co} V}{1+\beta_i/\beta} \tag{9}$$

(b) Active Region. In the active region

$$\rho_2 = -p_n \approx 0 \qquad \text{if} \qquad i_c >> i_{co}$$

The collector excess density may be considered zero provided the collector current is large compared with i_{co}. Thus

$$i_c = \beta i_b$$

$$v_c = V - \beta i_b R$$

Since $\beta >> 1$ and the collector voltage is larger than the base voltage (since the transistor is not yet saturated), the power dissipated due to base current is negligible compared with that due to the collector current. Thus

$$P \approx \beta V i_b - \beta^2 R i_b^2 \tag{10}$$

which reaches a maximum when $v_c \approx V/2$.

(c) Saturation. When the base current is increased to the point where $v_c = v_b$, the transistor saturates and the collector current becomes substantially independent of further increases in base current. Since both junctions are forward biased, the base and collector voltages will be neglected in comparison with V. Hence

$$i_c \approx V/R$$

$$i_b > V/\beta R$$

We may now solve for the excess densities ρ_1 and ρ_2 in order to obtain the junction voltages v_{be} and v_{bc}. We may write the equations for base and collector current as

$$i_c = (\rho_1 - \rho_2) G_d - \rho_2 G_2$$

$$i_b = \rho_1 G_1 + \rho_2 G_2$$

which may be solved for ρ_1 and ρ_2 writing β for G_d/G_1 and β_i for G_d/G_2

$$\rho_1 G_d \approx \frac{(1+\beta_i)\, i_b + i_c}{1 + \frac{\beta_i}{\beta} + \frac{1}{\beta}} \tag{11}$$

$$\rho_2 G_d \approx \frac{\beta_{ib} - i_c}{1 + \frac{\beta}{\beta_i} + \frac{1}{\beta_i}} \tag{12}$$

The combination ρG_d is a convenient quantity with which to deal since it has the dimensions of a current. When $\beta i_b = i_c$, the transistor has just become saturated and

$$i_c = \rho_1 G_d = V/R$$

At higher base currents the collector current remains essentially constant, but $\rho_2 G_d$ increases. Hence $\rho_2 G_d$ is a significant measure of the amount by which the transistor has been driven into saturation. We may now determine the junction voltages since

$$\rho_1 G_d = p_n G_d \left(e^{qv_{be}/kT} - 1 \right)$$

$$\rho_2 G_d = p_n G_d \left(e^{qv_{bc}/kT} - 1 \right)$$

Therefore

$$v_{be} = \frac{kT}{q} \operatorname{Ln} \left(\frac{\rho_1 G_d}{p_n G_d} + 1 \right)$$

$$v_{bc} = \frac{kT}{q} \operatorname{Ln} \left(\frac{\rho_2 G_d}{p_n G_d} + 1 \right)$$

where $p_n G_d$ may be determined from equations 5 or 6. Although the i_{co} expression is not as accurate as a measurement of junction voltage and current, it is often convenient for germanium units.

$$v_{be} \approx \frac{kT}{q} \operatorname{Ln} \left[\frac{i_c + (1 + \beta_i)\, i_b}{\beta_i\, i_{co}} + 1 \right] \tag{13}$$

$$v_{bc} \approx \frac{kT}{q} \operatorname{Ln} \left[\frac{(1 + \beta)\, i_b - i_c}{\beta i_{co}} + 1 \right] \tag{14}$$

The collector saturation voltage is just the difference between the two junction voltages. In most cases the drive current is sufficiently large that ρ_1 and ρ_2 are both much greater than p_n. Thus from equations 13 and 14

$$v_c \approx \frac{kT}{q} \operatorname{Ln} \frac{\rho_1 G_d}{\rho_2 G_d}$$

which from equations 11 and 12 becomes

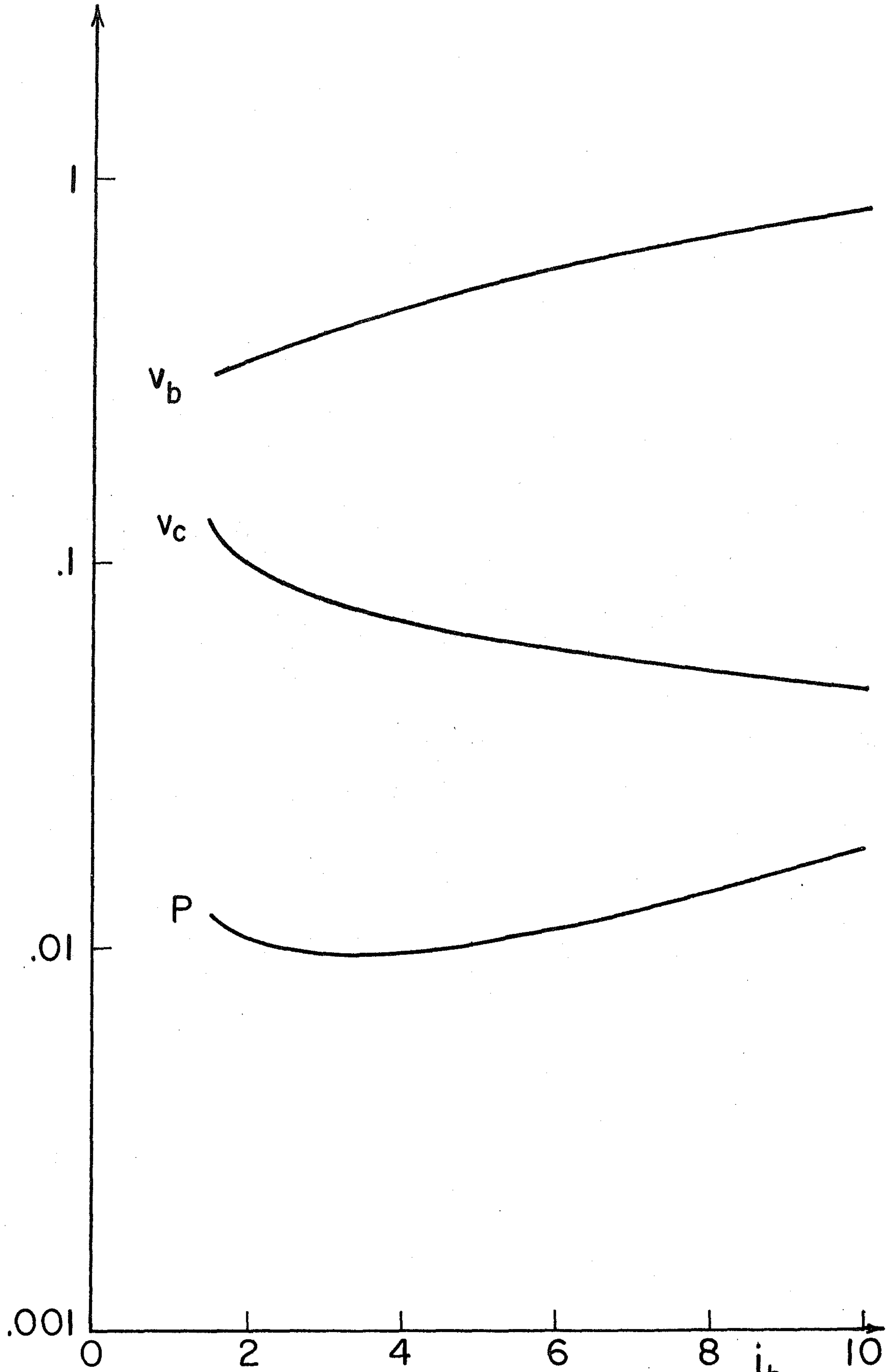


Fig. 13. - Voltage and power variation for normally connected transistor.

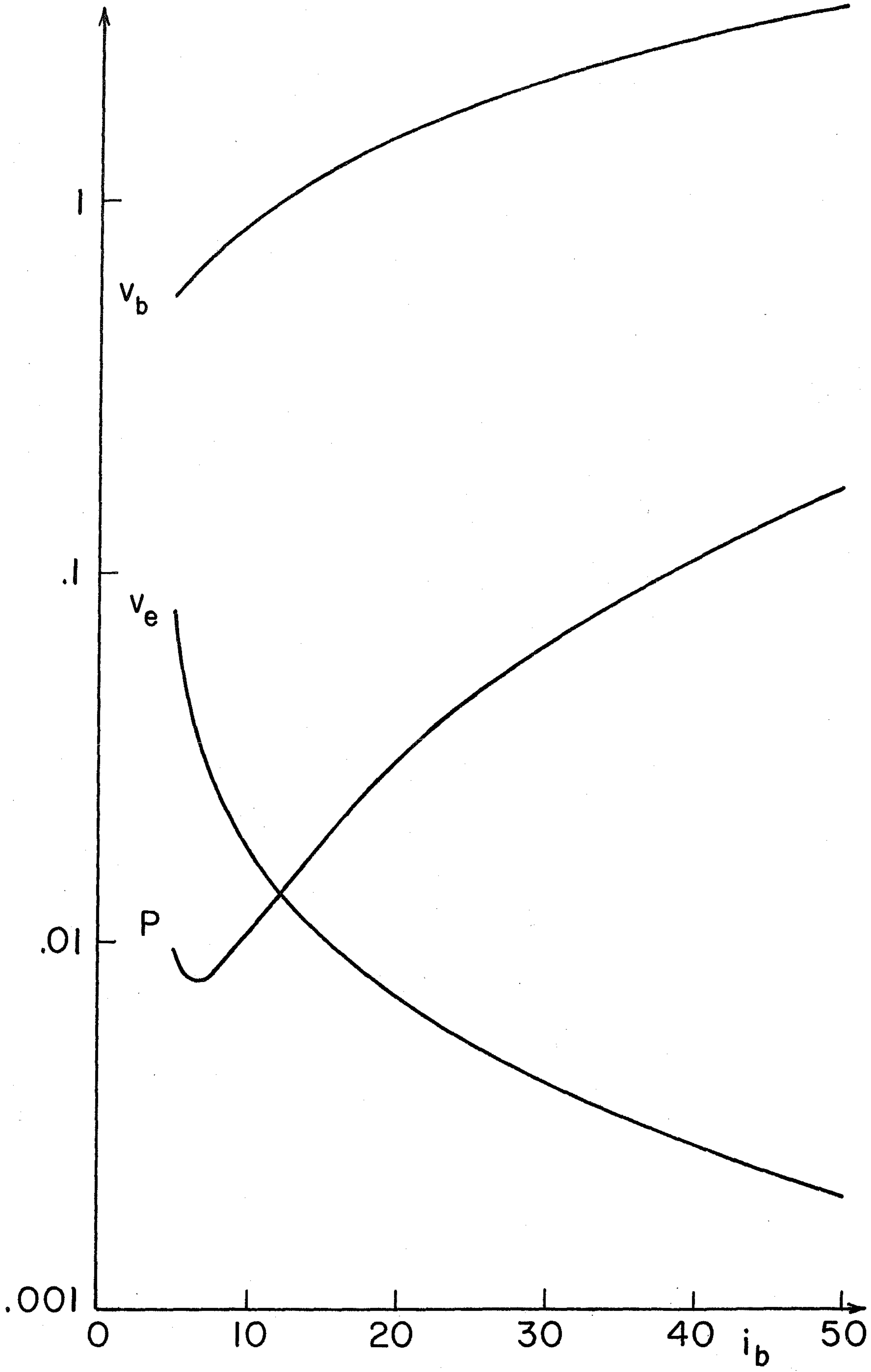


Fig. 14. - Voltage and power variation for inversely connected transistor.

$$v_c = \frac{kT}{q} \operatorname{Ln} \left(\frac{1 + \dfrac{1 + i_c / i_b}{\beta_i}}{1 - i_c / \beta i_b} \right) \tag{15}$$

In the inverted connection the normal and inverse quantities merely exchange places and the junction voltages become

$$v_{bc} = \frac{kT}{q} \operatorname{Ln} \left[\frac{i_e + (1 + \beta)\, i_b}{\beta i_{co}} + 1 \right]. \tag{16}$$

$$v_{be} = \frac{kT}{q} \operatorname{Ln} \left[\frac{(1 + \beta_i)\, i_b - i_e}{\beta_i\, i_{co}} + 1 \right] \tag{17}$$

Also, the emitter saturation voltage becomes

$$v_e = \frac{kT}{q} \operatorname{Ln} \left(\frac{1 + \dfrac{1 + i_e / i_b}{\beta}}{1 - i_e / \beta_i\, i_b} \right) \tag{18}$$

Plots of the base input voltage, saturation voltage, and power dissipation for a typical switching transistor operating in the normal and inverse connections are shown in Figs. 13 and 14. It is of interest to note the minimum in power dissipation which occurs at some base drive current. Clearly this drive current represents an optimum operating point for the particular transistor and collector current involved. The serious nature of large overdrive currents is quite apparent.

Dynamic Resistance. The resistance of a saturated transistor to small a-c signals is often of interest. This dynamic resistance is given by

$$R_s = \frac{\partial v_c}{\partial i_c}$$

for the normal connection and may be obtained from equation 15.

$$R_s = \frac{kT}{q}\left[\frac{1}{(1+\beta)\, i_b - i_c} + \frac{1}{(1+\beta_i)\, i_b + i_c}\right] \qquad (19)$$

A similar expression is obtained for the inverted connection, only the roles of the collector and emitter are interchanged and β is interchanged with β_i. For sufficiently large drive currents

$$R_s \approx \frac{kT}{q}\left[\frac{\beta+\beta_i}{\beta(1+\beta_i)\, i_b}\right] \qquad \beta >> 1 \qquad (20)$$

which shows the saturation resistance inversely dependent on β, β_i and the drive current. The importance of both β and β_i is dramatically illustrated by this formula.

All transistors have certain ohmic resistances associated with their collector and base circuits due to the semiconductor material between the active region of the device and its contact to the outside world. For the first approximation we may assume these resistances constant, and if the voltage drops across them are not negligible, we must add them to the appropriate voltages already calculated. For switching transistors made by the alloying process, and many others, the collector series resistance is negligible, but the base resistance R_b should always be taken into account. The total base voltage thus becomes

$$v_b = v_{be} + i_b R_b$$

Since we are considering base currents up to very high values, we may no longer neglect the power dissipation in the base circuit.

$$P = v_c i_c + v_{be} i_b + i_b^2 R_b \tag{21}$$

As we have noted, a number of simplifying assumptions have been made which under many conditions may be very questionable. However, for practical circuit design one often uses models which are greatly oversimplified, not because of their extreme accuracy but because they provide approximate answers and still allow a qualitative understanding of the problem. For example, the use of small signal equivalent circuits without regard to the magnitude of the signal level is an accepted engineering procedure. The nonlinearities are taken into account qualitatively after the main circuit behavior has been determined from the linear analysis. The lumped model serves in the same capacity for switching problems as a small signal equivalent circuit does for problems where the transistor is normally biased. It provides a straightforward method of obtaining results with reasonable accuracy in the majority of cases and hence may claim a certain engineering importance. The physical insight gained by the lumped model analysis is often much more valuable than a slight improvement in accuracy, since it enables the analyst to make qualitative statements concerning changes in circuit parameters, a very important step in the design procedure.

Transient Response.[8] Again we shall consider a junction transistor connected as shown in Fig. 12.

(a) Turn On. In the cut-off condition $i_b = -i_{co}$ as before. Now let us apply a positive current step of magnitude I_1 (large compared to i_{co}) and calculate the collector current response. As long as the transistor remains normally biased, the collector current may be computed from small signal formulae

$$i_c(s) = \frac{\beta i_b(s)}{1 + \frac{s}{\omega_\beta}} = \frac{\beta \omega_\beta I_1}{s\,(s + \omega_\beta)}$$

$$i_c = \beta I_1 \left(1 - e^{-\omega_\beta t} \right)$$

However, the collector remains reverse biased only so long as

$$i_c < V/R$$

Thus the collector current rises toward the asymptote βI_1 with time constant $1/\omega_\beta$. If $\beta I_1 > V/R$, the transistor will saturate when the collector current reaches V/R . The "rise time" required is thus

$$t_r = -\frac{1}{\omega_\beta} \operatorname{Ln} \left(1 - \frac{V}{\beta I_1 R} \right) \tag{22}$$

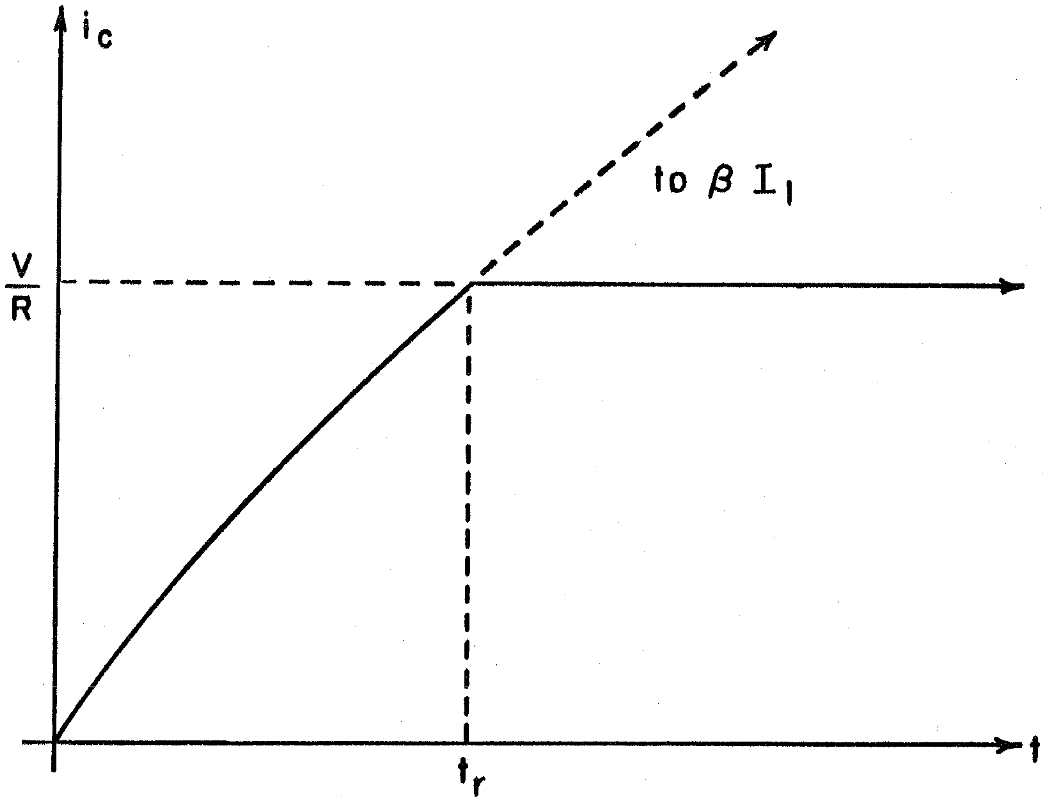


Fig. 15. - Rise transient as predicted by transistor lumped model.

Usually the transistor is driven quite hard in order to minimize the rise time. Under these conditions $\beta I_1 >> V/R$ and we may expand the log, retaining only the first term

$$t_r \approx \frac{V}{\beta\omega_\beta RI_1}$$

Since normally $\beta >> 1$, $\omega_\alpha \approx \beta\omega_\beta$. The rise time may be written

$$t_r \approx \frac{V}{\omega_\alpha RI_1} \tag{23}$$

It is thus clear that the alpha cut-off frequency is the determining factor in turn-on time and not β or ω_β alone. The conditions during turn-on are illustrated in Fig. 15.

(b) Storage. After the transistor has reached the steady state with $i_b = I_2$, let us suddenly reverse the base current to $i_b = -I_3$. As in the case of the diode, the collector junction remains forward biased since ρ_2 cannot change instantaneously. Hence the collector current remains $I_c = V/R$. After a "storage time" t_s, ρ_2 has reached zero and the transistor becomes normally biased.[10] In order to determine ρ_1 and ρ_2 during the storage period, we may determine their initial values from the steady state conditions given by equation 12, assuming $\beta >> 1$

$$\rho_2(0) = \frac{\beta I_2 - I_c}{\frac{\beta}{\beta_i} + 1} \qquad I_c = V/R \tag{24}$$

The transient densities may be found by superposing the steady state solution above (for $I_c = V/R$, $i_b = I_2$) upon the solution for a negative base current step of magnitude $I_2 + I_3$ and constant collector current $i_c = 0$.

The step solution proceeds as follows: Since $i_c = 0$, G_d and the parallel combination of G_2 and C_2 merely act as an "excess density divider". Hence

$$\rho_2 = \rho_1 \frac{G_d}{G_d + G_2 + C_2 s}$$

Also

$$i_b(s) = \rho_1 (G_1 + C_1 s) + \rho_2 (G_2 + C_2 s)$$

From these two equations we may solve for $\rho_2(s)$

$$\rho_2 = \frac{G_d\, i_b(s)}{C_1 C_2 s^2 + (G_d C_2 + G_d C_1 + G_2 C_1 + G_1 C_2) s + G_d G_2 + G_d G_1 + G_1 G_2}$$

We may now identify terms in the denominator D with the transistor parameters

$$D = C_1 C_2 \left[s^2 + (\omega_a + \omega_{ai}) s + (1 + \beta_i + \beta)\, \omega_\beta \omega_{\beta i} \right]$$

which may be factored into two real roots

$$D = C_1 C_2 (s + a)(s + b)$$

where

$$(a, b) = \frac{(\omega_a + \omega_{ai})}{2} \quad 1 = \sqrt{1 - \frac{4\omega_\beta \omega_{\beta i}(1 + \beta_i + \beta)}{(\omega_a + \omega_{ai})^2}}$$

Examination of the term under the radical shows that it is small compared to unity so long as $\beta >> 1$. Therefore the radical may be expanded and only the first terms retained. The roots then become

$$a \approx \omega_a + \omega_{ai} \tag{25}$$

$$b \approx \frac{\omega_a \omega_{\beta i} + \omega_\beta \omega_{ai}}{\omega_a + \omega_{ai}} \tag{26}$$

and the densities may be written

$$\rho_2 \approx \frac{\beta_i \omega_{\beta i} \omega_a}{(s + a)(s + b)} \; i_b(s)$$

Taking the inverse transform and including the initial conditions for the case

$$i_b(s) = - \frac{I_2 + I_3}{s}$$

The collector density may be found as a function of time

$$\rho_2 G_d = \rho_2(0)\, G_d - \frac{(I_2 + I_3)\, \beta_i \omega_{\beta i} \omega_a}{ab} \left(1 + \frac{be^{-at} - ae^{bt}}{a - b} \right)$$

If $\beta >> 1$ then $\omega_\alpha >> \omega_\beta$ and $a >> b$ the exp (- at) term is negligible.

$$\rho_2 G_d \approx \rho_2(0) G_d - (I_2 + I_3) \frac{\beta\beta_i}{\beta + \beta_i} (1 - e^{-bt})$$

When the value of $\rho_2(0)$ is inserted from equation 24, this expression simplifies to

$$\rho_2 G_d \approx (I_2 + I_3) \frac{\beta\beta_i}{\beta + \beta_i} e^{-bt} - \frac{\beta_i I_c + \beta\beta_i I_3}{\beta + \beta_i} \tag{27}$$

where $I_c = V/R$. The "storage time" t_s ends when ρ_2 reaches zero. Hence

$$e^{-bt_s} = \frac{\beta_i I_c + \beta\beta_i I_3}{\beta\beta_i (I_2 + I_3)}$$

or

$$t_s = \frac{1}{b} \mathrm{Ln} \frac{I_3 + I_3}{I_3 + \frac{I_c}{\beta}} \tag{28}$$

from which it is clear that the storage time may be reduced by using large turn-off currents. I_2 must be greater than I_c/β for the transistor to be saturated, but the overdrive may be reduced to decrease the storage time. The controlling time constant for the storage period is b, hence for small storage times both $\underline{\omega_\beta}$ and $\underline{\omega_{\beta i}}$ should be large.

If these frequencies are nearly equal

$$b \approx \omega_\beta \approx \omega_{\beta i}$$

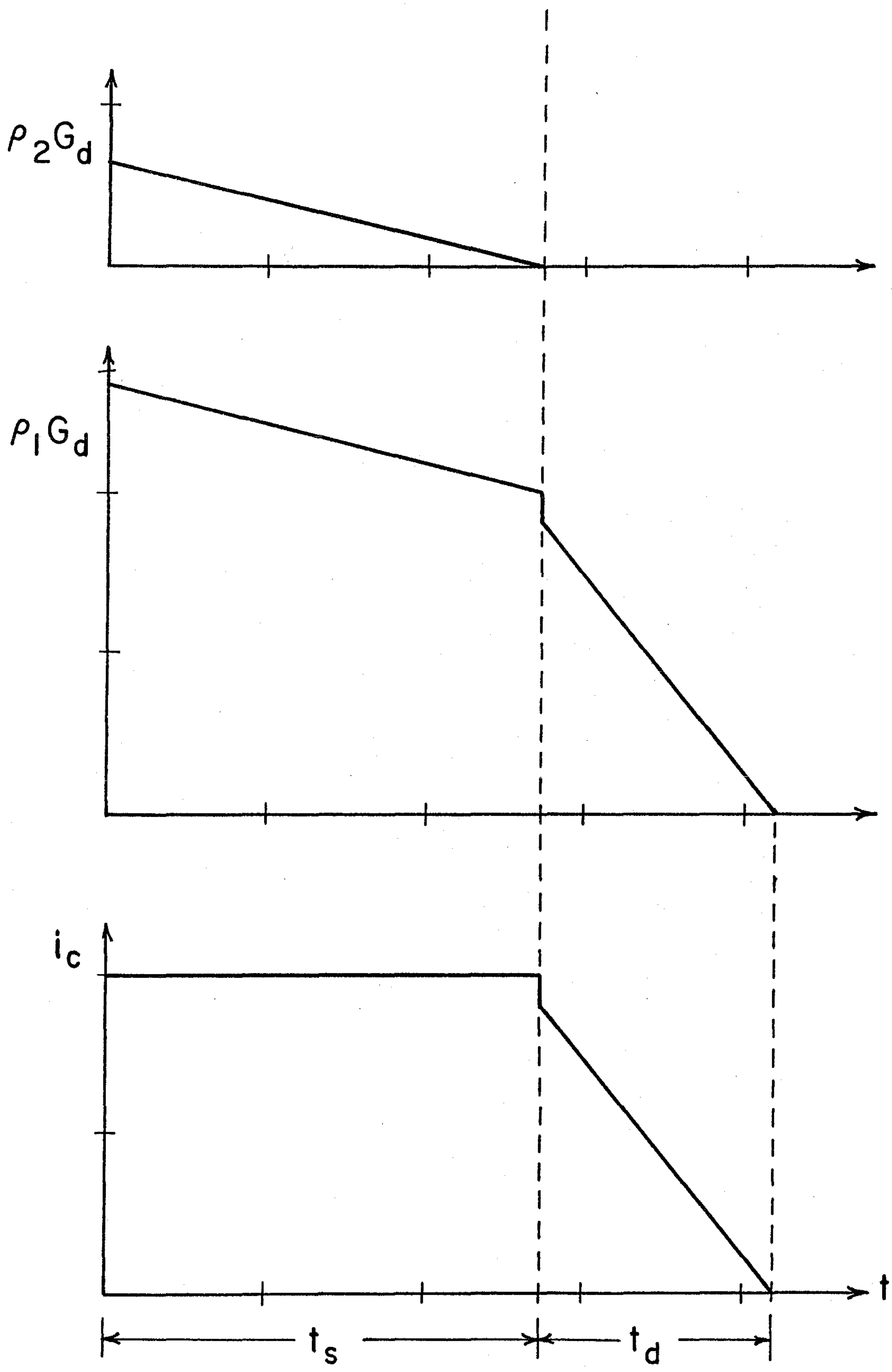


Fig. 16. - Storage and decay transients as predicted by transistor lumped model.

Plots of ρ_1, ρ_2 and i_c during the storage period are shown in Fig. 16.

(c) Turn Off. When ρ_2 approaches zero, the collector current is made up of two components:

$$I_c = \rho_1 G_d + C_2 \frac{d\rho_2}{dt}$$

However, as the collector junction becomes reverse biased, $d\rho_2/dt$ abruptly becomes zero since ρ_2 cannot become less than $-p_n$ and the lumped model predicts a slight discontinuity in collector current. In reality the actual current changes smoothly, and this is another case where the lumped nature of the model fails to give the completely correct physical picture. However, for purposes of calculating the "decay time" or time required for i_c to reach zero, the lumped model expressions including the discontinuity will be more accurate than the corresponding expressions assuming no discontinuity. The reason is that the true collector current very quickly approaches the predicted value as an asymptote and by the time i_c approaches zero, the lumped model expression is quite accurate. Let us therefore calculate the magnitude of the discontinuity.

$$\Delta i = C_2 \frac{d\rho_2}{dt}\bigg|_{t = t_s} = -\frac{C_2}{G_d}\, b\, \frac{(\beta_i I_c + \beta\beta_i I_3)}{\beta + \beta_i}$$

Substituting for b and assuming $\beta >> 1$

$$\Delta i = -\left(I_c + \beta I_3\right)\left(\frac{\omega_\beta}{\omega_a + \omega_{ai}}\right)$$

Thus under normal circumstances the relative magnitude of the discontinuity is quite small. However, for very large overdrive currents it may become important. Under these conditions

$$\Delta i \approx -\beta I_3 \frac{\omega_\beta}{\omega_a + \omega_{ai}} \tag{29}$$

After ρ_2 reaches zero, the transistor is again normally biased and we may use the small signal approach, as with the turn-on period. The collector current approaches the asymptote $-\beta I_3$ with a time constant $1/\omega_\beta$. During this period $i_c = \rho_1 G_d$

$$i_c = (I_c - \Delta i)\, e^{-\omega_\beta t} - \beta I_3 \left(1 - e^{-\omega_\beta t}\right)$$

The "decay time" t_d is the time required for i_c to reach zero

$$t_d = \frac{1}{\omega_\beta} \operatorname{Ln}\left(1 + \frac{I_c + \Delta i}{\beta I_3}\right)$$

For large turn-off current $I_3 >> I_c/\beta$ and we may approximate the logarithm. The decay time therefore becomes

$$t_d \approx \frac{1}{\omega_a}\left(\frac{I_c}{I_3} - \frac{1}{1 + \omega_{ai}/\omega_a}\right) \quad \text{if} \quad \beta >> 1 \tag{30}$$

Thus, as with the rise time, the decay time is determined by the magnitude of the drive current and the α cut-off frequency, and not by β or ω_β alone.

After the turn-off period, the transistor is cut off and the collector current resumes its small steady state value as given by equation 7. Comparing the expression for rise time as given in equation 23, we see that the decay time is always less than the rise time for a given base drive current. This is true because recombination is helpful during the decay period but harmful during the rise period.

PART II

NON-LINEAR THEORY

In part I a rather detailed treatment of the operation of junction transistors has been presented assuming the lumped model elements involved were constant. It has been pointed out that the lumped model approach provides a deep understanding of the physical principles underlying transistor operation. One is naturally led to ask if such an approach would also provide a similar understanding of the non-linear behavior of junction transistors which predominates as current densities are pushed higher and higher. In the sections which follow, the non-linear theory will be investigated in some detail, and modifications of the linear theory which are necessary under a number of conditions encountered in practice will be established. First, however, let us consider the two basic assumptions implicit in the linear theory.

(a) All lumped model elements are not affected by the excess carrier density (i.e. the theory is linear).

(b) The emitter and collector injection efficiencies are nearly unity.

In many practical situations the important problem is what maximum current may be controlled by a given transistor, or what type of transistor may be used to control a given current in the most effective manner under a given set of circumstances. The question of the fundamental limitations of switching transistors has been discussed by a number of authors, but a satisfactory consistent treatment is still lacking. Perhaps one of the basic obstacles has been the tendency to treat certain specific effects in great detail and omit other entire areas of as great or greater importance. Since a complete exact description of the high level effects in any useful transistor is manifestly impossible, and since even the first order linear theory is not entirely free of approximations, it is felt that the most fruitful approach is first to investigate semi-quantitatively the effects to be expected in general, then to estimate the relative magnitudes of these effects in given specific cases, and finally to evaluate the operation of a given unit by including only the large effects and ignoring all others. This method has the advantage that quite general statements may be made concerning the performance and inherent limitations of commonly used devices. Also a deep qualitative understanding of all processes involved is much more readily gained.

Throughout the discussion a pnp transistor is assumed in the interest of clarity and simplicity of notation.

1. SUMMARY OF NON-LINEAR EFFECTS

Let us now investigate in general the effects which would be expected to contribute to the non-linear behavior of a junction transistor. Although a complete list would necessarily be of much greater length than utility, the following effects are of some importance in present devices and merit careful consideration.

Depletion layer recombination. At low injection levels in silicon devices, junction operation is dominated by carrier recombination within the depletion layer. In germanium devices and in silicon devices operating at higher levels, this effect is negligible.[4, 15, 16] Since we are specifically interested in the high level case, we shall assume the effects of depletion layer recombination are small. However, some discrepancies will be found to exist between the high level theoretical expressions and data taken at relatively low level, which may be explained in terms of this omission, but which are not important for the purposes of this discussion. The effect of such recombination on the lumped model is to introduce conductances in parallel with G_1 and G_2 which decrease rapidly as ρ_1 and ρ_2 become appreciable.

Drift-enhanced diffusion flow. When minority carriers are injected into a region of semiconductor, the condition of near charge neutrality requires an equal number of additional majority carriers.

A gradient in excess minority carrier density therefore implies a gradient in majority carrier density. If the majority carrier current is constrained by means of junctions, the diffusion current tendency implied by this density gradient must be canceled by an equal and opposite drift current tendency; hence an electric field is developed. As long as the injected density is small compared to the equilibrium majority density, the drift component of minority carrier resulting from this electric field is small compared to the diffusion component. However, at high injection levels, the drift component becomes appreciable, resulting in an effective diffusion constant larger than at low densities. At very high levels, the effective diffusion constant approaches twice its low level value.[11] As we will later separate the high and low level cases, it will not be necessary to consider the transition in detail. However, since the low level limiting form of the solution is of interest, it must be borne in mind that the high level value of G_d will be twice the low level value due to this effect.

Recombination rate. For most purposes the rate of recombination of minority carriers may be considered proportional to the product of majority and minority carrier densities. However, in semiconductor material as commonly used, recombination occurs principally at traps or recombination centers, and the recombination statistics are altered.[15, 16, 23] A detailed discussion of the recombination statistics as applied to commonly encountered cases is given in Appendix A, with the result that the recombination rate at high levels may still be considered proportional to the excess density.

Conductivity modulation. The conductivity of a semiconductor is usually taken as proportional to the equilibrium density of majority carriers which, under normal conditions, is equal to the density of donor or acceptor atoms. However, at high levels the density of majority carriers is increased as the density of minority carriers becomes appreciable. Hence the number of carriers available for current flow is increased and the conductivity of the material is increased. Ohmic resistance effects (for example, the base resistance and series collector resistance of a transistor) are altered to the extent that these effects arise in the active region of the transistor. As the excess density of injected carriers becomes appreciable compared with the majority carrier density, the increase in majority density must be taken into account in the charge neutrality condition. For example, in an n type semiconductor:

$$n = n_n + \rho$$

$$p \approx \rho$$

The conductivity σ of the material then becomes

$$\sigma = q(\mu_n n + \mu_p p) \approx q[\mu_n n_n(1 + \frac{\rho}{n_n}) + \mu_p \rho]$$

or

$$\sigma = \sigma_o \left[1 + (1 + b) \frac{\rho}{n_n} \right] \tag{31}$$

where σ_o is the low-level conductivity and b is the ratio of minority to majority carrier mobilities.

Junction injection efficiencies. When a p-n junction is forward biased, in addition to the holes injected from the p-region into the n-region, electrons are injected from the n-region into the p-region. The ratio of the injected densities in the two regions is inversely proportional to the ratio of the majority carrier densities. If the equilibrium majority density of one region is much greater than that of the other, the injected density will become comparable to the majority density at much lower levels in the low conductivity region. Hence the ratio of injected densities will vary with the injection level. This effect becomes significant, for example, at a forward biased junction of a transistor where the junction injection efficiency is caused to decrease as the injected density is increased.

Although the contribution to the total base current in a junction transistor due to non-unity injection efficiency is normally small compared with that due to surface recombination, the lateral base current within the active area of the device is often principally due to injection effects and hence of great importance in connection with the self-bias cutoff effect. The dependence of injected current density upon junction voltage and minority carrier excess density is treated in Section 2 of Appendix A. Two cases are considered:

a. Asymmetrical Junction - In many transistor junctions the conductivity of the transistor base region is relatively low compared with that of the other region in question. In this case the current density J injected from the base region into the other region is shown to be related to the excess density ρ in the base region as follows:

$$J \sim \rho \left(1 + \frac{\rho}{n_n}\right) = p_n\left(e^{qv/kT} - 1\right) \qquad (32)$$

b. Graded Junction - The collector junction of a diffused-base transistor is normally a very gradual transition between regions of relatively low conductivity in order to achieve a high breakdown voltage. Hence when such a junction is forward biased, even at rather low injected densities, the minority carrier density is large compared with the equilibrium majority carrier density. Under these conditions the injected densities ρ and η in the two regions are equal, and it is shown that the junction injection law may be written

$$\rho = \eta = n_i e^{qv/2kT} \tag{33}$$

Density crowding or self-bias effect. Because of volume recombination and non-unity junction efficiencies in a junction transistor, majority carrier current is caused to flow laterally in the base region. This current must flow through the ohmic resistance of the base region and hence causes a lateral potential gradient in the base region. This gradient is in such a direction as to decrease the junction forward bias in the center of the base region. Hence at high levels the injected carriers are essentially confined to the outer edge of the base region.[12,20,24] The major consequence of this effect is a reduction in the effective area of the active region of the transistor. In many transistors the high level behavior is dominated by this effect. In this work we will consider the results of density crowding in considerable detail.

Thermal effects. When a junction transistor is operated at high current densities, heat is evolved at junctions where the applied reverse bias is appreciable and also in the ohmic resistances associated with the various regions. Changes in temperature of the device caused by heat generated in this manner cause changes in the electrical properties of the unit which are often very difficult to distinguish from the high level phenomena previously discussed.

Junctions in a transistor or diode are often assigned a certain thermal time constant and thermal resistance to the mounting base of the device, and it is thus implied that the junction itself possesses a certain thermal capacitance, which may be fairly large. However, the flow of heat in a solid is basically a diffusion process and it should be recognized that the lumped approximation implied by the thermal time constant and thermal resistance is valid only under conditions where the thermal gradients are not too severe. Under conditions of very high power input for very short periods, such as during pulse measurements, the lumped approximation would predict a very small change in temperature. In reality, since the power is dissipated only within the junction and hence is localized to a very small volume, the actual local temperature may rise to a very high value. Such a temperature increase causes violent changes in the equilibrium minority carrier density as well as less drastic changes in lifetime, mobility, and surface recombination properties. Although we shall not treat these effects in detail, we must always be aware that experimental results may be colored by their presence.

Hole-electron scattering. As the injected carrier density reaches very high levels, a decrease in the mobility of minority carriers is to be expected due to hole-electron scattering. This effect is treated in considerable detail in a paper by Fletcher,[22] where the mobility of holes and electrons is plotted as a function of minority carrier excess density. Although these results have not been explicitly incorporated into the present work, it should be noted that since the lumped model approach has been used throughout, the determination of injected densities is a necessary step in the solution of any high level problem. Like Fletcher[12] and others,[20,24] we have assumed the mobility to be constant. However, a good estimate of the effects of hole-electron scattering may be obtained using the computed values of injected density in connection with Fletcher's data.

Outline of analysis. In order to gain a basic understanding of transistor high current operation we have briefly considered a number of the more important effects which contribute to its non-linear behavior. We shall now restrict our attention to a number of specific cases of practical interest. Of the bewildering array of transistors now available, we shall consider two major types: the alloy transistor, and the diffused or graded-base transistor. These two general classifications together with their near relatives cover the vast majority of transistors used at the present time.

Due to the self bias effect we expect non-linear performance of both types even at injection levels low compared with the majority carrier density. In this case we shall see that the general formulation for a normally biased transistor may be applied with equal validity to both types. Because of the symmetry of the structure, an inversely connected alloy unit is also properly treated by the same analysis. However, because of the high collector resistivity of graded-base types, we would expect the decrease in collector injection efficiency to become important at much lower injection levels as the collector is forward biased. For this reason a separate analysis is presented for the inverse connection of such transistors.

For many alloy units, much of the operation is performed under very high level conditions. The corresponding analysis must include the effects of decreased injection efficiency and base conductivity modulation. Both of these effects considerably modify the form of the self bias effect, and are treated in some detail.

Finally, the operation of both types of transistor in the saturation region is considered. The injection efficiency of both emitter and collector is seen to be important in alloy units, while in diffused transistors the collector injection efficiency is much lower and the contribution of the emitter may be neglected. In both types, the electric field across the base region due to majority carrier density gradient may make important contributions to the saturation voltage at high levels.

2. NON-LINEAR OPERATION; BASIC EQUATIONS

The basic equations describing the operation of a junction transistor under conditions where carrier density crowding must be taken into account will now be derived. The analysis is presented for a normally connected alloy transistor but is equally valid for the inversely connected alloy units and normally connected diffused base structures.

In an alloy (fused) junction transistor, both emitter and collector regions are highly doped and both junctions are highly asymmetrical. A typical geometry is shown in Fig. 17. The structure is assumed two-dimensional, i.e., independent of the coordinate normal to the page. Thus, the analysis will be accurate for stripes and ring configurations and interleaved finger power transistors. It will also be approximately correct for dot configurations at high levels where the current is confined principally to the edge of the emitter. Initially we shall solve for the distribution of minority carrier density in the base region caused by the self-bias effect as a function of x. This information will allow us to calculate the base current and collector current if the collector is reverse biased, and hence the behavior of current transfer ratio as a function of current level. Another way of expressing the solution is in terms of the variation of the lumped model parameters as a function of the excess density.

Low level approximation. As noted before, if $\rho \ll n_n$ both the electron current injected into the emitter and the bulk

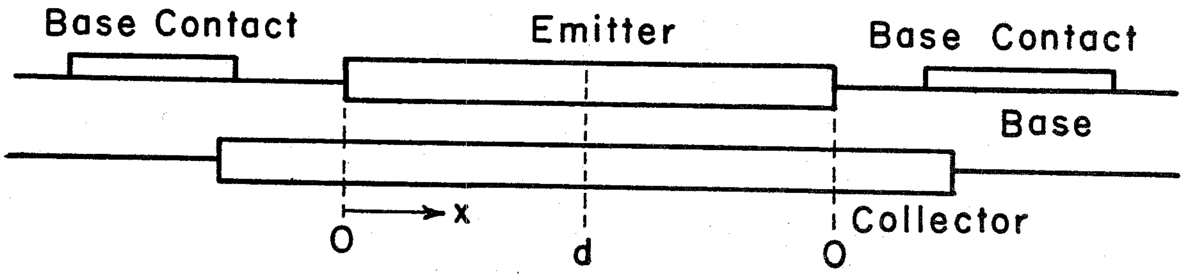


Fig. 17. - Structure of alloy transistor.

recombination may be considered proportional to the excess density. Hence the lateral (x-directed) base current flowing outward from the interior base region is also proportional to the excess density. By the symmetry of the structure the lateral base current is zero at the center of the emitter ($x = d$) and increases as the edge of the emitter ($x = 0$) is approached. Since both the current injected into the emitter and the bulk recombination rate are proportional to ρ_1 the number of carriers lost from a volume element of unit length, width equal to the base width, and depth dx located at $x = x'$ is proportional to $\rho(x')dx'$. The lateral base current per unit length $i(x)$ is equal to the sum of such contributions from all volume elements between x and the center of the emitter. In the limit dx approaches zero and the sum becomes an integral:

$$i(x) = A \int_{d}^{x} \rho(x')\, dx' \tag{34}$$

Here A is the sum of two terms, the first being the total charge lost in a unit volume of semiconductor base material per second due to volume recombination, and the second being the total current flowing from base into emitter per unit area, both evaluated at unit excess density.

The lateral base current flowing through the sheet resistance of the base creates a lateral voltage drop within the base region.

This self bias voltage drop is in such a direction as to decrease the forward bias voltage applied to interior portions of the emitter-base diode.[12] The voltage $v(x)$ between the base at the outer edge of the emitter $(x = 0)$ and at some interior point x may be written as the sum of the voltage drops across the resistances of each volume element, each individual voltage drop is equal to the product of the lateral base current and the base sheet resistance R.

$$dv = R\, i(x')\, dx'$$

Passing to the limit, the total voltage drop becomes

$$-v(x) = R \int_0^x i(x')\, dx' \qquad (35)$$

(Note that A and R may be considered constant only when $\rho \ll n_n$; conductivity modulation of the base region and the decrease in emitter efficiency create important variations which will be treated later.)

The effect of the self-bias voltage $v(x)$ is to decrease the excess hole density injected into the base region from its value at $x = 0$. As we have restricted our attention to the case $p_n \ll \rho \ll n_n$, we may write the injected density in terms of the self bias voltage $v(x)$ from equation 32

$$\rho(x) = \rho(0)\, e^{-qv(x)/kT} \qquad (36)$$

Both the mobility and diffusion constant of minority carriers have been assumed independent of excess density as noted earlier. Hence each volume element may be considered as a separate transistor, and the total diffusion current through each element is proportional to the density gradient, or in the special case where the collector is reversed biased, to the emitter excess density at the element in question. Integrating the contributions to the diffusion current over all elements, the total collector diffusion current i_c may be written,

$$i_c = F \int_0^d \rho(x')\, dx' \tag{37}$$

Where F is directly proportional to the carrier diffusion constant and inversely proportional to the base width.

The formulation of the basic equations necessary for determining the transistor high level operating characteristics is now complete. These equations with appropriate modifications will also be used for cases treated in later sections. It is clear that even in the present simple case the relations are quite non-linear.

3. NON-LINEAR OPERATION: ALLOY TRANSISTOR

Low level approximation. As noted, the basic equations derived in section 2 may be applied directly to the alloy transistor under conditions where the injected density is small compared with the majority carrier density. Even under these conditions a general solution is quite difficult, and the results do not present a closed form relation between i_c and ρ. The matter is treated in some detail in Appendix B with the result that normalized values, ρ^* and i_c^*, of the density and collector current are related as shown in Fig. 18. The normalized parameters are defined as follows:

$$\rho^* = \frac{qRAd^2}{2kT} \quad \rho_1 = \frac{3}{2} K \rho_1 G_{do} \tag{38a}$$

$$i_c^* = \frac{qRAd}{2kTF} i_c \quad = \frac{3}{2} K i_c \tag{38b}$$

where $$K = \frac{qRAd}{3kTF} \quad \text{and} \quad G_{do} = Fd \tag{38c}$$

An important implication of these results is the ability to present a unique form for the dependence of i_c on ρ_1 with only one normalizing constant to be determined. As in the linear case we find the combination ρG_{do} a convenient one with which to deal.

The relation between i_c^* and ρ^* is seen to be linear at values of i_c^* much less than unity. At higher levels i_c^* increases only as the square root of ρ^*. The variation of the effective diffusion conductance with excess density and also with collector current as shown in Fig. 19. In this figure the normalization

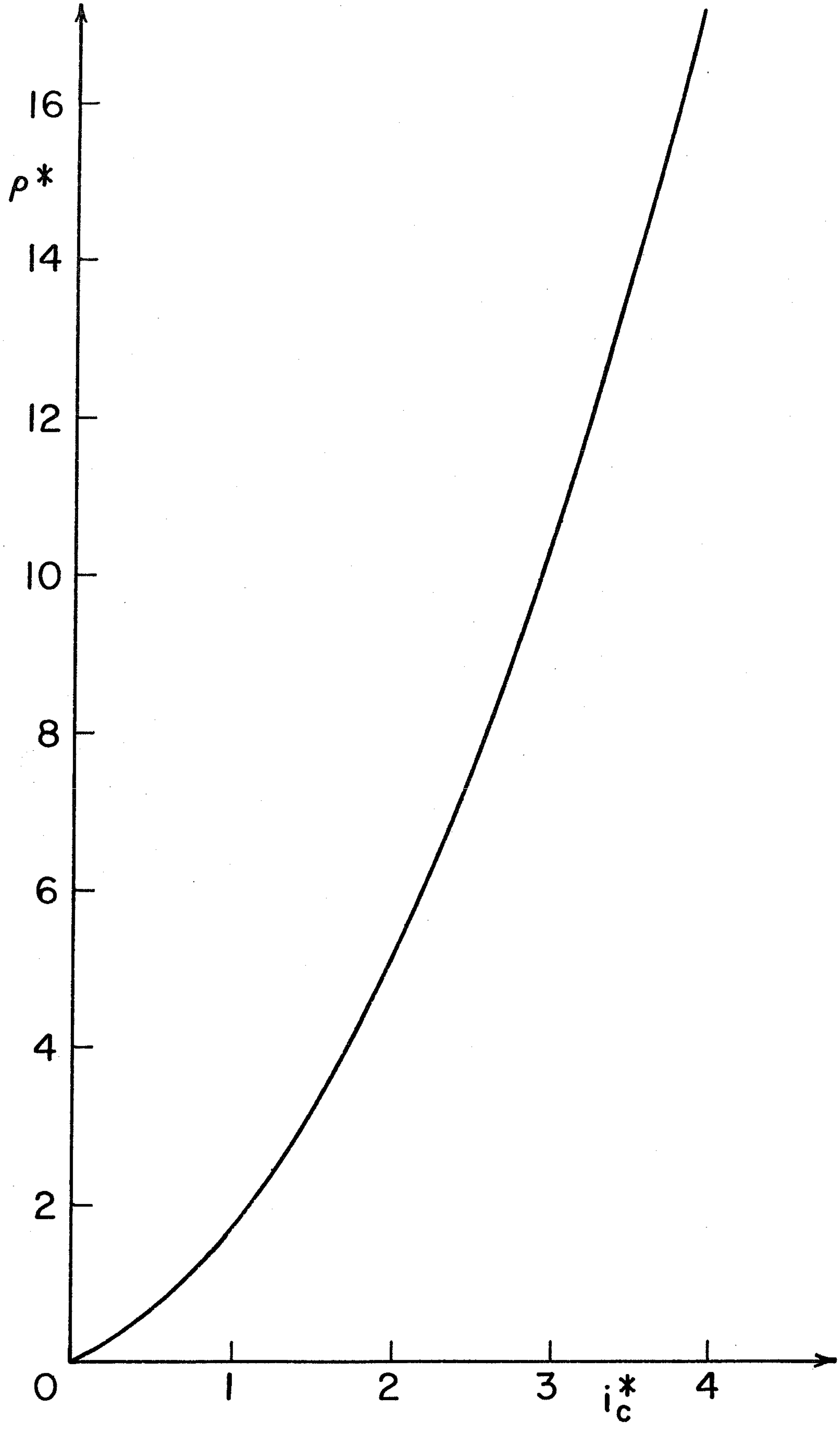


Fig. 18. - Dependence of collector current upon excess density.

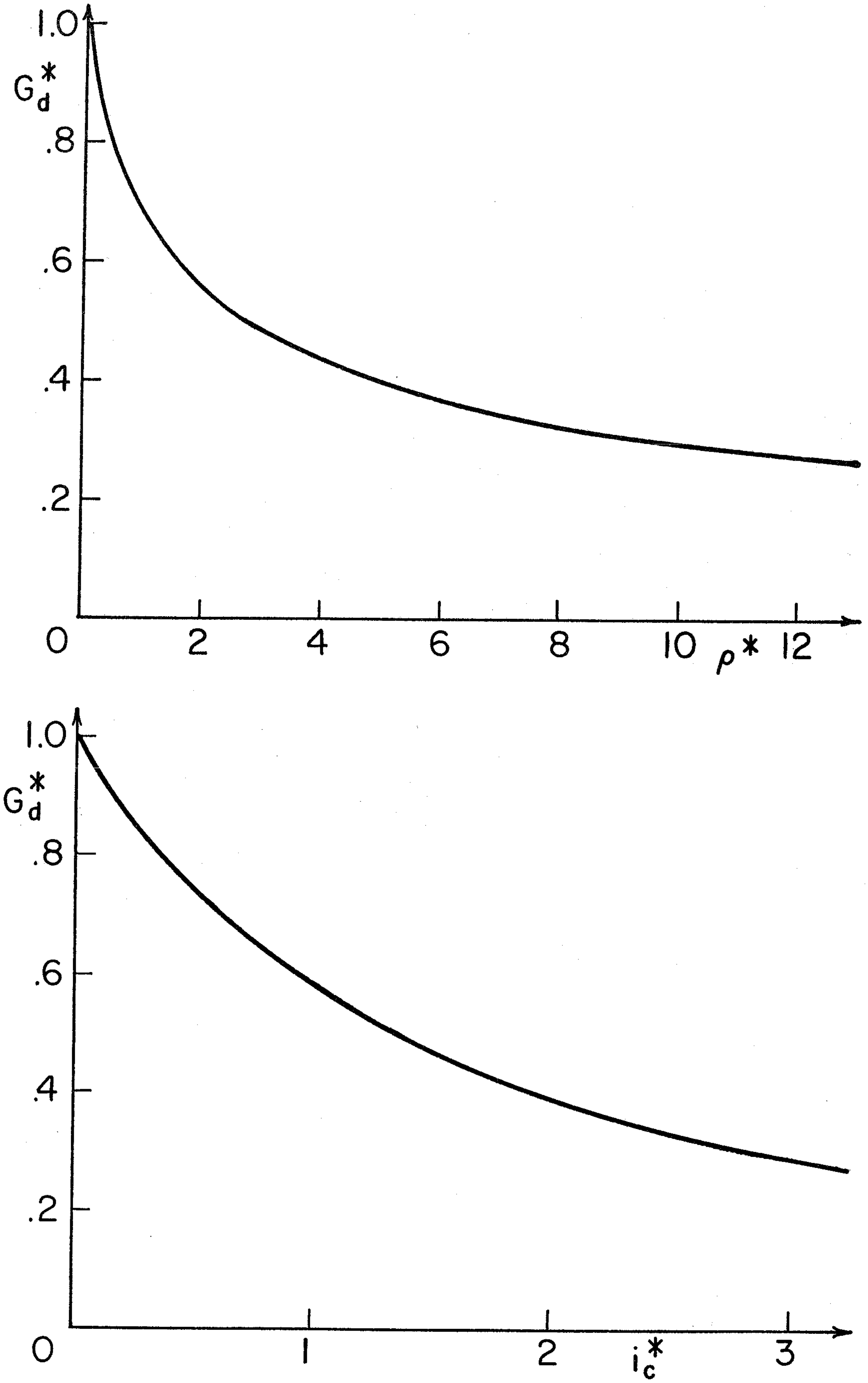


Fig. 19. - Theoretical variation of effective diffusion conductance.

$$G_d^* = \frac{G_d}{G_{do}}$$

has been used. Again only the constant K (in addition to the linear lumped model element values) is needed to completely characterize the transistor. The inverse of this constant, given in units of current, is a valuable figure of merit for transistors with which the present approximations are valid. From a circuit standpoint this value represents the current at which the transistor has begun to fail, and therefore provides a reliable comparison of the current handling capacity of various units. As noted in Appendix B, although the lateral base current within the active area of the transistor changes as G_d changes, in the vast majority of transistors the total base current is nearly all due to surface recombination and may be considered proportional to ρ_1. Since the lateral base current is but a small fraction of the surface recombination current, its contribution to the total base current may be neglected. Thus G_1 may be assumed constant and the variation of G_d with current becomes the variation of β with current, following the $1/i_c$ dependence at higher currents.

From a device design point of view, we are interested in the requirements on the physical parameters of a transistor to make the figure of merit as large as possible. Since the figure of merit is inversely proportional to K we should strive for large values of F and small values of R, A, and d.

The decrease in G_d at higher levels is due to a decrease in the effective area of the transistor due to the density crowding or self-bias effect. A very large reduction in G_d may be observed while the injected density is still small compared with the equilibrium majority density, n_n. In large area power transistors this condition is very prevalent.

Experimental data taken on an alloy power transistor of essentially two dimensional construction are shown in Fig. 20, together with the theoretical relation as given in Fig. 18. The value of K has been determined to give the best fit. It can be seen that the results agree within experimental error over a very wide current range.

a) Near uniform approximation. Closed form solutions for the relation between i_c and ρ_1 may be obtained for the two limiting cases of very slight and very severe crowding. As shown in Appendix B, for the slightly crowded case the diffusion conductance may be written

$$G_d \approx \frac{G_{do}}{1 + K\,\rho_1 G_{do}} \tag{39}$$

Since the condition for validity of this approximation is that the correction term involving K be small compared to unity, we may obtain an approximate expression for the current gain by substituting

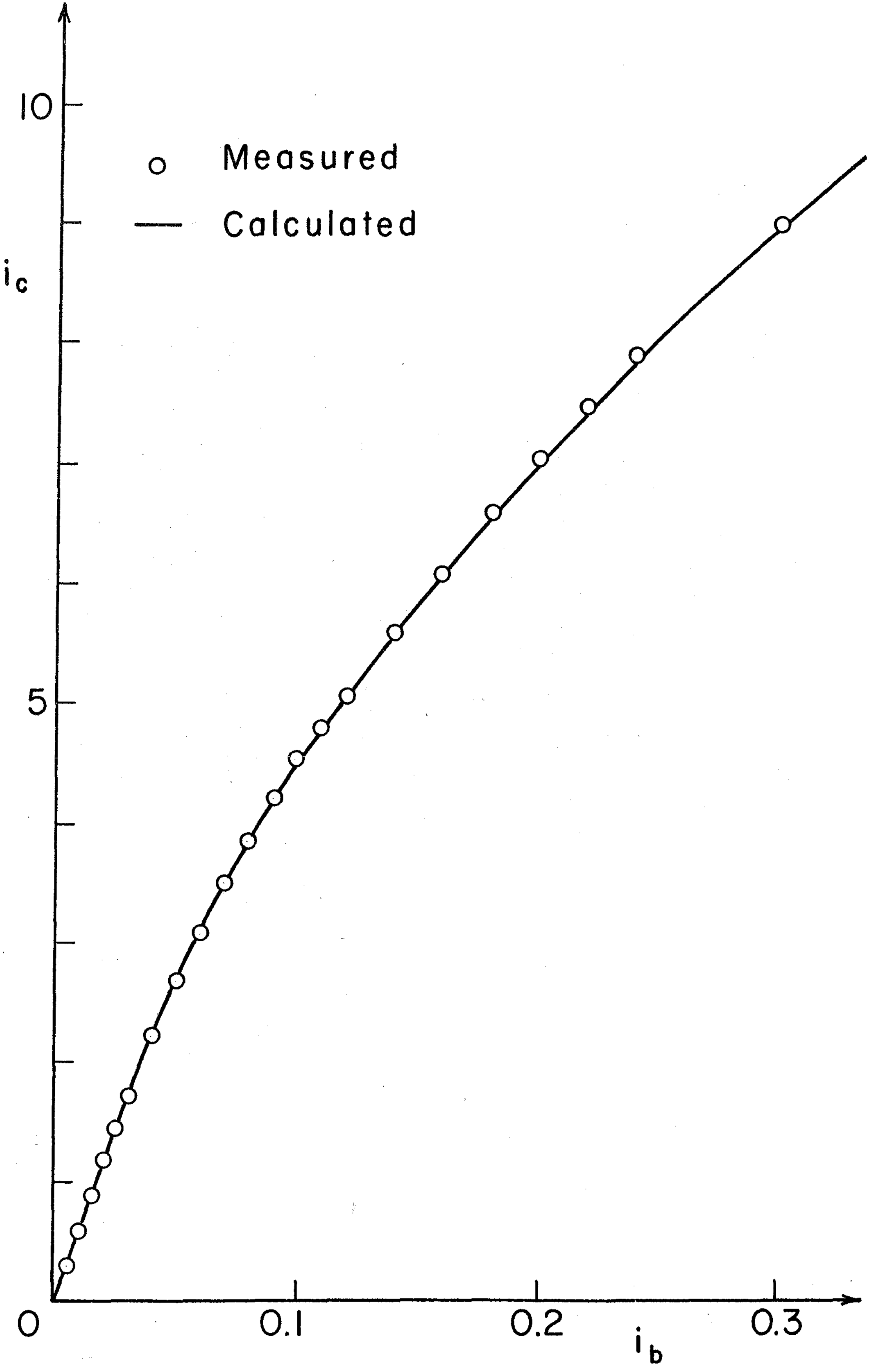


Fig. 20. - Experimental verification of Fig. 18.

$$i_c \approx \rho_1 G_{do}$$

in the correction term. Again the value of G may be assumed constant and the current gain may be written

$$\beta \approx \frac{\beta_o}{1 + K i_c} \tag{40}$$

This expression is very significant since it gives the form of the falloff of β in the most useful range of operation.

b) Severely crowded approximation. As shown in Appendix B, under conditions of severe crowding the injected density and collector current are related by the following expression:

$$\rho_1^* \approx i_c^* + \frac{\pi^2}{4} \qquad \rho^* >> 1$$

In the range where this approximation is valid, the constant term may be neglected without appreciably affecting the accuracy. Hence the diffusion conductance may be written:

$$G_d^* \approx \frac{1}{i_c^*}$$

or

$$G_d \approx \frac{2G_{do}}{3K i_c} \tag{41a}$$

which implies:

$$\beta \approx \frac{2\beta_o}{3K i_c} \tag{41b}$$

It is interesting that this expression agrees with the high current limit of equation 40 within the factor of 3/2. Thus the conclusions to be drawn from either expression are the same.

A plot of $\rho^*(0)$ as a function of i_c^* is shown in Fig. 21 together with the asymptotic forms at low and high currents. It may be seen that results obtained by means of the asymptote are never in error more than about 10%, the worst case occurring when $\rho^* \approx 10$ and convergence is quite rapid at lower or higher values.

High level approximation. The foregoing expression for the figure of merit indicates that the self-bias cutoff effect or emitter density crowding as discussed would be most noticeable in transistors with large emitter dimensions, namely power transistors. Indeed, the effect was first treated in detail by Fletcher[12] in connection with power transistors. However, his analysis assumed infinite emitter width and hence did not treat the transition from uniform flow to the crowded condition It is clear from equation 38 that for a given emitter area the current-carrying capability varies inversely with the emitter width. Hence the most efficient design for such units is an interleaved system of narrow base and emitter stripes or rings.[18,12,19] As the emitter width is decreased, the self-bias effect becomes less noticeable and current densities are pushed higher and higher. In many transistors now available, the assumption of $\rho \ll n_n$ is very

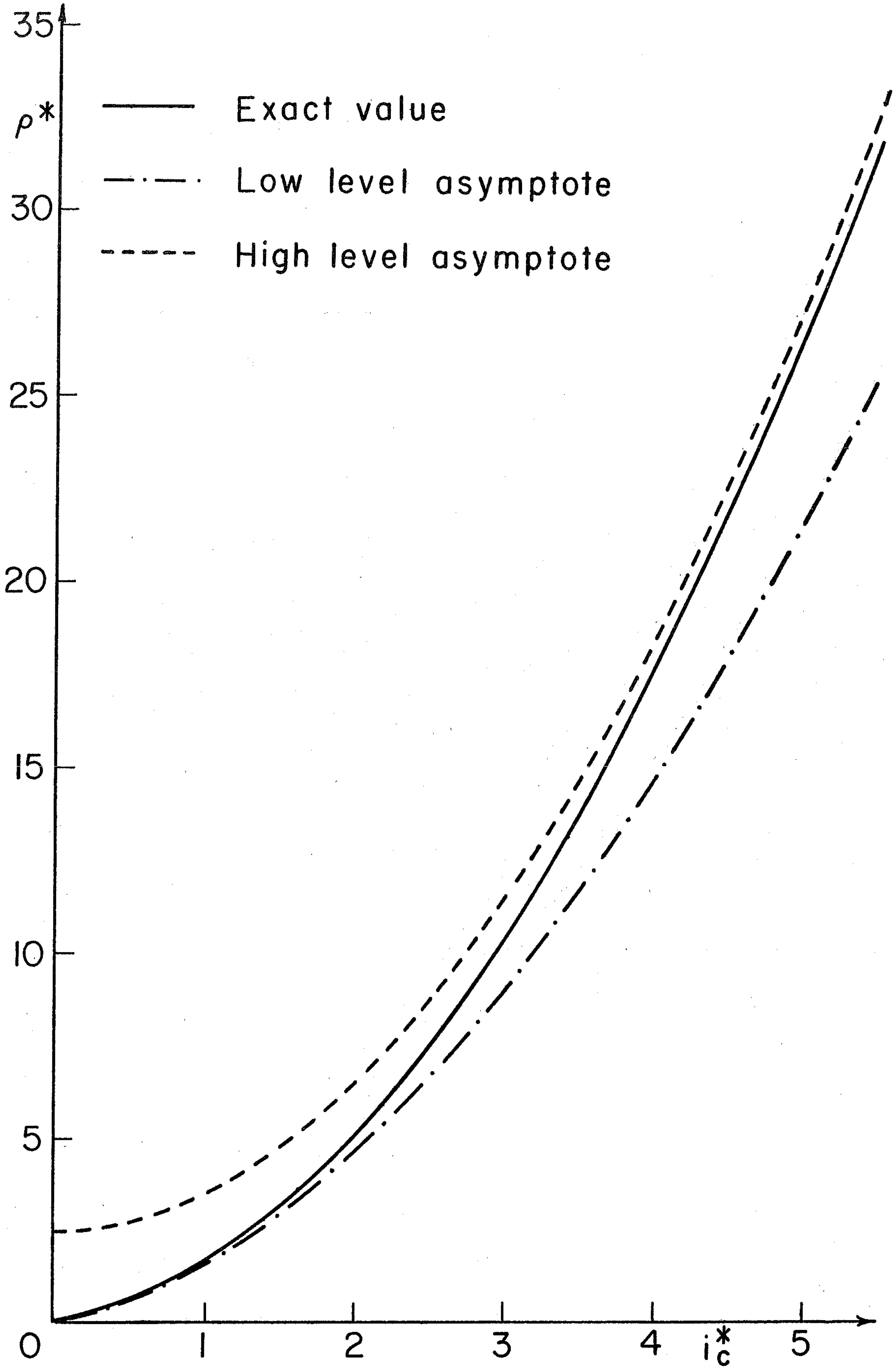


Fig. 21. - Asymptotic to the relation of Fig. 18.

poor, in fact the other extreme $\rho \gg n_n$ is often valid over the greater part of the range of operation. Thus, we must re-examine the analysis including the effects of large injected density.

At high injection levels two major changes in the physical operation occur:

i) The emitter efficiency decreases as shown in equation 32. This decrease may be quite large, and in general at high levels the current injected into the emitter region will be the dominant source of lateral base current inside the area covered by the emitter.

ii) The resistivity of the base region decreases due to conductivity modulation caused by the injected minority carriers, as shown in equation 31. Hence R may not be considered a constant, and the variation with ρ must be shown in the integral.

If we assume that all of the lateral base current is due to the decrease in emitter injection efficiency and neglect bulk recombinations, we may directly apply the expression for current density given in equation 32. The expression for lateral base current then becomes

$$i(x) = A \int_d^x \rho(x') \left[1 + \frac{\rho(x')}{n_n}\right] dx' \qquad (42)$$

The lateral voltage drop caused by this current is also affected by the conductivity modulation of the base region. Since the effective resistivity of each element of base region is reduced as given by equation

31, we may write

$$-v(x) = R \int_0^x \frac{i(x')}{1 + \frac{(1+b)\,\rho\,(x')}{n_n}}\, dx' \tag{43}$$

The junction injection law is also modified at high levels in accordance with equation 32. We may assume as before $\rho \gg p_n$

$$\rho\,(x) \left[1 + \frac{\rho(x)}{n_n} \right] = \rho(0) \left[1 + \frac{\rho(0)}{n_n} \right] e^{-qv(x)/kT} \tag{44}$$

Further analysis becomes very difficult, unless another simplification is achieved. Since we are primarily interested in the high current or high injection level limit on the devices, and since ρ becomes equal to n_n at quite low current levels in many devices,[15, 16] we will make the assumption of high injection level ($\rho \gg n_n$), realizing that the results will break down at the lower levels, where the analysis of Section 2 may be applied. As is often the case with engineering approximations such as these, the regions of mathematical validity for the two approximations are widely separated. However, in practice, the approximations are reasonably good into the intervening range and quite a good representation of the overall performance may be made by the use of these relatively simple cases.

Two approximate solutions of these equations are given in Appendix C.

a) Near uniform approximation. The first solution is derived for conditions where the crowding is not severe. In this case it is shown that the diffusion conductance may be written

$$G_d = \frac{G_{do}}{1 + K \rho_1 G_{do}} \tag{45}$$

where in this case the current falloff factor K becomes

$$K = \frac{qRAd}{3(1 + b)\, kTF}$$

which is seen to differ from the low-level case of equation 40 only by the constant (1 + b). The fact that a unique form is available for the two cases is quite significant for the circuit engineer. It permits the non-linear characterization of a transistor to first order without a knowledge of whether the unit is operating at high or low levels, simply by a measurement of the constant K.

In the case of transistors with wide emitters, severe crowding already exists as ρ becomes comparable with n_n. If we may assume severe crowding, an exact solution to the equations is possible. As shown in Appendix C, for this case the collector current is given by

$$i_c = 2F \sqrt{\frac{6(1 + b)\, kT}{qAR} \rho_1}$$

which indicates a value of G_d which is inversely proportional to the collector current. The base current may be written in the form

$$i_b = \rho_1 \left(G_s + c\rho_1^{3/2} \right) \tag{46}$$

where G_s is the effective surface recombination conductance and

$$c = \frac{A}{5n_n} \sqrt{\frac{qRA}{6(1 + b)\, kT}}$$

As noted before, in most cases the surface recombination term will dominate until very high levels are reached. Under these conditions, β is again proportional to $1/i_c$ as in the other cases.

It should be pointed out that in transistors, notably modern alloy units, where the base resistivity is quite high and the emitter is very heavily doped, the approximations of $\rho \gg n_n$ in the base, and $\eta \ll p_p$ in the emitter are valid over a very wide current range, and hence the analysis assumes a greater importance. However, in other cases where the ratio of the conductivities in the two regions is not nearly so large, the operation may vary smoothly from the low level case to the condition of high injection in both regions.

It seems unnecessary to deal with the correlation between emitter efficiency and current gain fall off as widely treated in the early literature. However, because of the widespread misconception of the principles involved, perhaps a word here is in order. From the results of the foregoing analysis it is clear that the $1/i_c$ dependence of β which was observed is to be expected as a result of the self-bias cutoff effect when the base current is completely dominated by surface recombination. Since the surface recombination is known to dominate in a great many transistors, especially earlier units, it is inconceivable that the base current due to decreased emitter efficiency could have been responsible for the rapid falloff in β which was observed. The great improvement which was effected by the use of highly doped emitter materials was due to the decrease in lateral base current and hence a decrease in the self-bias cutoff effect which resulted in a more uniform distribution of carriers across the emitter surface. This improvement resulted in a marked increase in collector current for a given value of injected density. In other words, the observed variation in β was primarily the result of a decrease in G_d due to emitter crowding, not an increase in G_1. The necessity for an extremely large value of emitter efficiency is now apparent, since the small internal lateral base current is much more important than previously recognized.

As mentioned before, the results of this section are directly applicable to alloy transistors operating in either the normal or inverse connection, the low level expressions are also applicable to normally biased diffused base units. However, it should be noted that many alloy structures utilize a different doping material for the collector junction than for the emitter, and therefore the constant A will not in general be the same for the two connections. Also, the collector is normally considerably larger than the emitter, making the surface recombination conductance G_2 larger than G_1.

4. NON-LINEAR OPERATION; DIFFUSED BASE TRANSISTOR

Diffused base transistors are becoming increasingly common in modern transistor circuit work because of their superior voltage and frequency capabilities. Perhaps the most significant feature of such a transistor is the density distribution of donors in the base region. Typically the doping level is very high near the emitter and decreases more or less exponentially to a very gradual junction with a near-intrinsic collector. This construction provides a rather wide collector space charge region with attendant high voltage breakdown and low capacitance. The gradient of majority carriers supplied by the impurity distribution also produces a built-in electric field which aids minority carrier flow across the base region. This effect increases the conductance G_d and decreases the carrier transit time, improving the high frequency performance of the device.

Normal connection. The emitter junction of the transistor may be fabricated by alloying or by a subsequent p-type diffusion. In any case the conductivity of the emitter region is large compared with that of the base near the emitter, and the characteristics of the junction approach those of an abrupt junction. For this reason, the non-linear performance of such a transistor for the normal connection may be analyzed in exactly the same way as that of an alloy unit. Since the resistivity of the base region is much smaller near the emitter than elsewhere, most of the lateral base current will flow near the emitter. Therefore the approximation $\rho \ll n_n$ will be valid over a much wider range than in normal alloy transistors, n_n being the value of base region equilibrium majority density near the emitter. However, since the value of n_n near the emitter is much closer to the value of p_p in the emitter, the range of currents for which $\rho \gg n_n$ and $\eta \ll p_p$ is likely to be very small, even nonexistent. Hence this case is usually not useful and we shall restrict our attention to the "low level" case as considered in Section 3. The results of that section may be applied directly to normally biased diffused base transistors.

Inverted connection. Since the collector junction of a diffused base unit is normally a gradual transition between two rather high resistivity regions, at even rather low injected densities the minority carrier density will exceed the equilibrium majority density.

a) Moderate level. At moderate injection levels we may assume the injected minority carrier density large compared with the equilibrium majority density in both regions. Under these conditions $p \approx n$ in both regions and we may apply the results of Section 2. From equation 33

$$\rho \approx n_i e^{qv/2kT} \quad \text{for the n-region} \tag{47a}$$

$$\eta \approx n_i e^{qv/2kT} \quad \text{for the p-region} \tag{47b}$$

As noted earlier, the conductivity of the base region is very small near the collector and very large near the emitter. At moderate levels we may assume all of the conduction in the base region to occur in the high conductivity area near the emitter and very little in the lightly doped area near the collector. The conditions for the validity of this assumption are that ρ near the collector be much less than n_n near the emitter. This approximation is seen to be identical to the low-level alloy case if it is recognized that the lateral base current is caused by electron current injected into the collector, which is proportional to η which is equal to ρ . In this case A is much larger than in the normally biased case. Also kT must be replaced by 2kT from equation 33.

The total base current is composed of two components, one due to the lateral base current i(0) and one due to a combination of hole and electron current flowing across the collector junction outside the area covered by the emitter junction.

A word here seems in order concerning the physical operation of a diffused transistor in the inverted connection. Typically such transistors are made by diffusing for example a donor element from a very highly doped surface layer into a rather high resistivity p type slab. Subsequently all of the surface is masked except the area destined to become the emitter, which is subjected to a thin, very high concentration acceptor diffusion. Base contacts are then attached to the surface on each side of the emitter. The resulting geometry is shown in Fig. 22. An interesting variation of such a transistor is the triple diffused $pn\,pp^+$ in which an additional very deep acceptor diffusion is introduced into the collector side of the slab. The effect of such a diffusion is to greatly reduce the collector ohmic resistance, and also to produce an electric field in the collector region due to the impurity density gradient. Cross sections of such a structure both within the area of the emitter and outside showing the net impurity concentration as a function of the distance y through the slab are shown in Fig. 23.

Both the base and collector regions contain electric fields which tend to prevent diffusion flow away from the collector junction when it becomes forward biased and hence act as tanks or reservoirs for injected minority carriers.[21] Under these conditions the characteristics of the base surface facing the collector may not be important

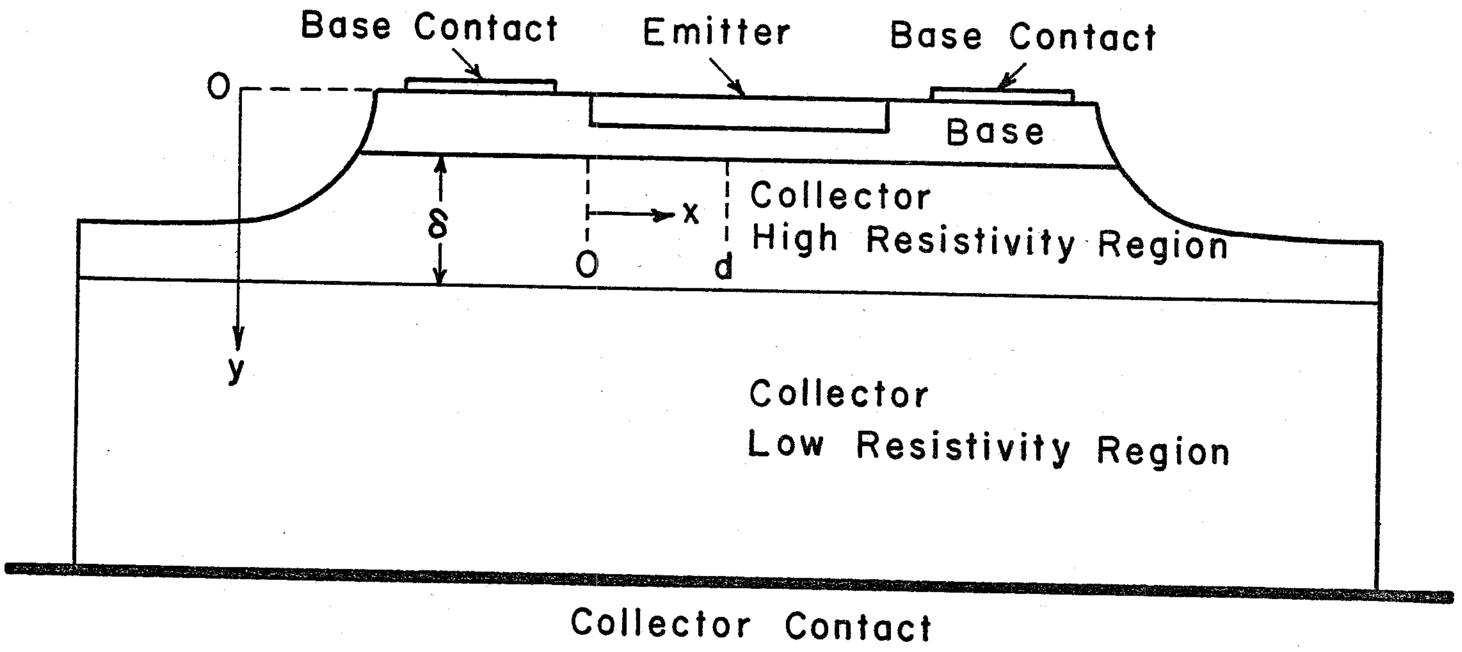


Fig. 22. - Structure of diffused base transistor with diffused collector contact.

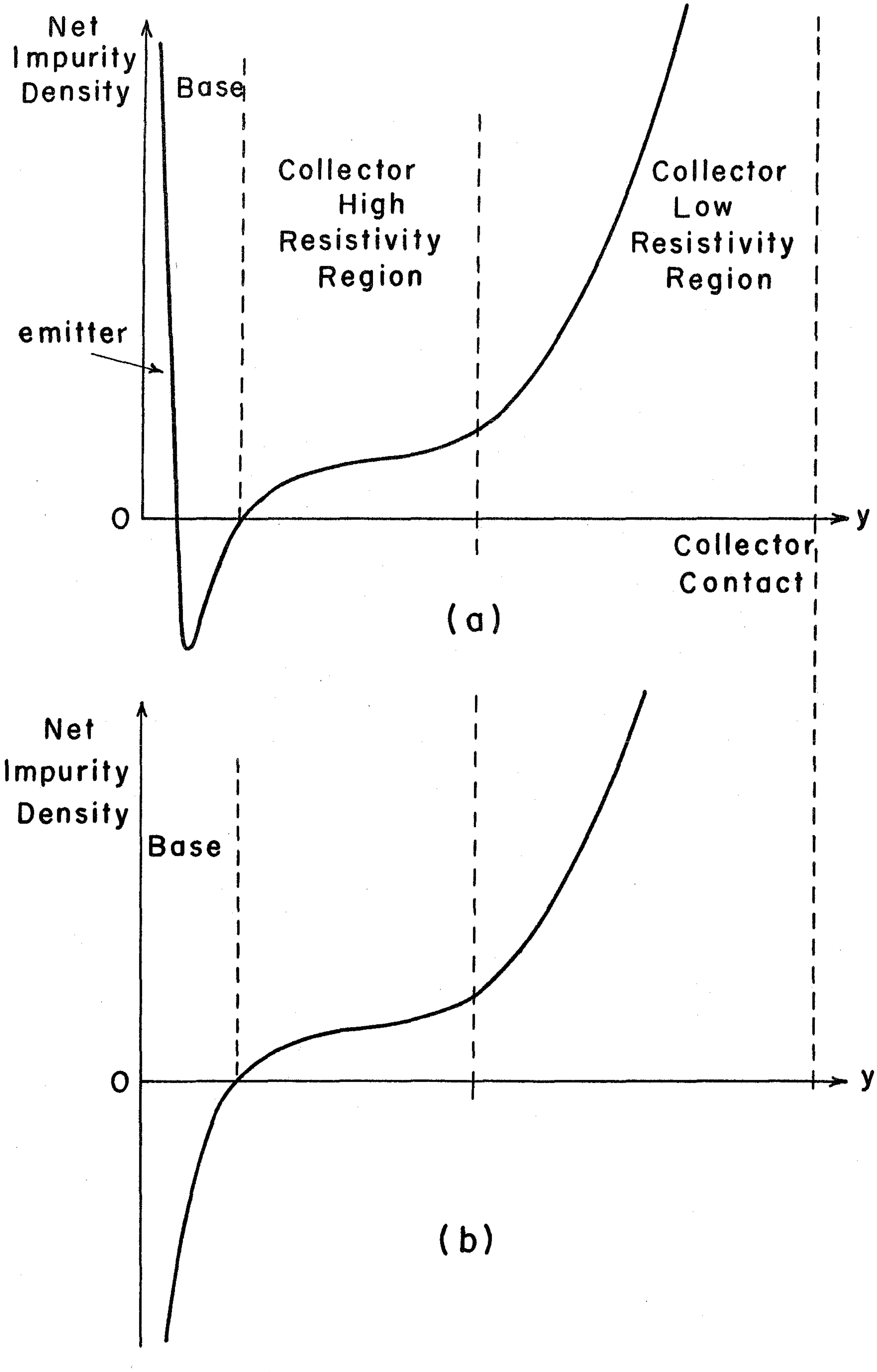


Fig. 23. -Impurity profiles (a) within and (b) outside the emitter region.

since minority carriers are effectively prevented from diffusing this far into the base. The total base current thus observed may be many times smaller than that calculated assuming uniform diffusion to the surface. In the collector region, this storage of carriers tends to prevent electron diffusion and hence may greatly reduce the lateral base current and self-bias cutoff effects. In such cases the bulk recombination rate in the collector is chiefly responsible for lateral base current and should be made as small as possible. In transistors with very low bulk recombination a large fraction of the base current is due to surface recombination at the edge of the collector. In other cases the diffusion of carriers into the body of the collector is quite important. In any case the portion of base current originating outside the emitter area may be assumed proportional to ρ_2 as before.

Since both the total emitter current i_e and the lateral base current $i(0)$ are proportional to the area under the injected density curve,

$$i(0) \sim i_e$$

Therefore the total base current becomes

$$i_b = c_1 i_e + c_2 \rho_2$$

In most cases the first term is quite small compared to the second at low levels and decreases rapidly as the crowding is increased.

Therefore again we may assume i_b proportional to ρ_2 and the equations and figures derived in connection with the low-level alloy case may again be applied directly.

Data taken on on inverted transistor with a graded base structure and near intrinsic collector region are shown in Fig. 24. The theoretical curve was calculated from equation 39. Very good agreement is noted over a wide range of emitter and base currents.

5. SATURATED CONDITIONS; ALLOY TRANSISTOR*

As has been shown in connection with the linear lumped model analysis, one of the most outstanding and useful properties of a junction transistor is its unique ability to saturate. Under saturated conditions the minority carrier density at the collector builds up to the

---

*The saturation voltage of alloy transistors has been analyzed in great mathematical detail by Huang.[26] Unfortunately the two most important features of saturation operation were treated in a self-inconsistent manner. In the determination of the internal lateral base current, the effects of non-unity emitter efficiency were neglected, yet were included later in the calculation of the total base current. Clearly if this contribution were at all important in the total base current, it would be the dominant source of lateral base current. The second inconsistency occurs in the solution for the radial distribution of carrier density assuming the collector and base radii to be equal, (a case which is in general uninteresting). The results obtained under these assumptions are then extended to the case where the radii are not equal. For any case where the crowding is at all severe this extension is grossly in error. In the light of these errors the value of Huang's analysis is seriously questionable.

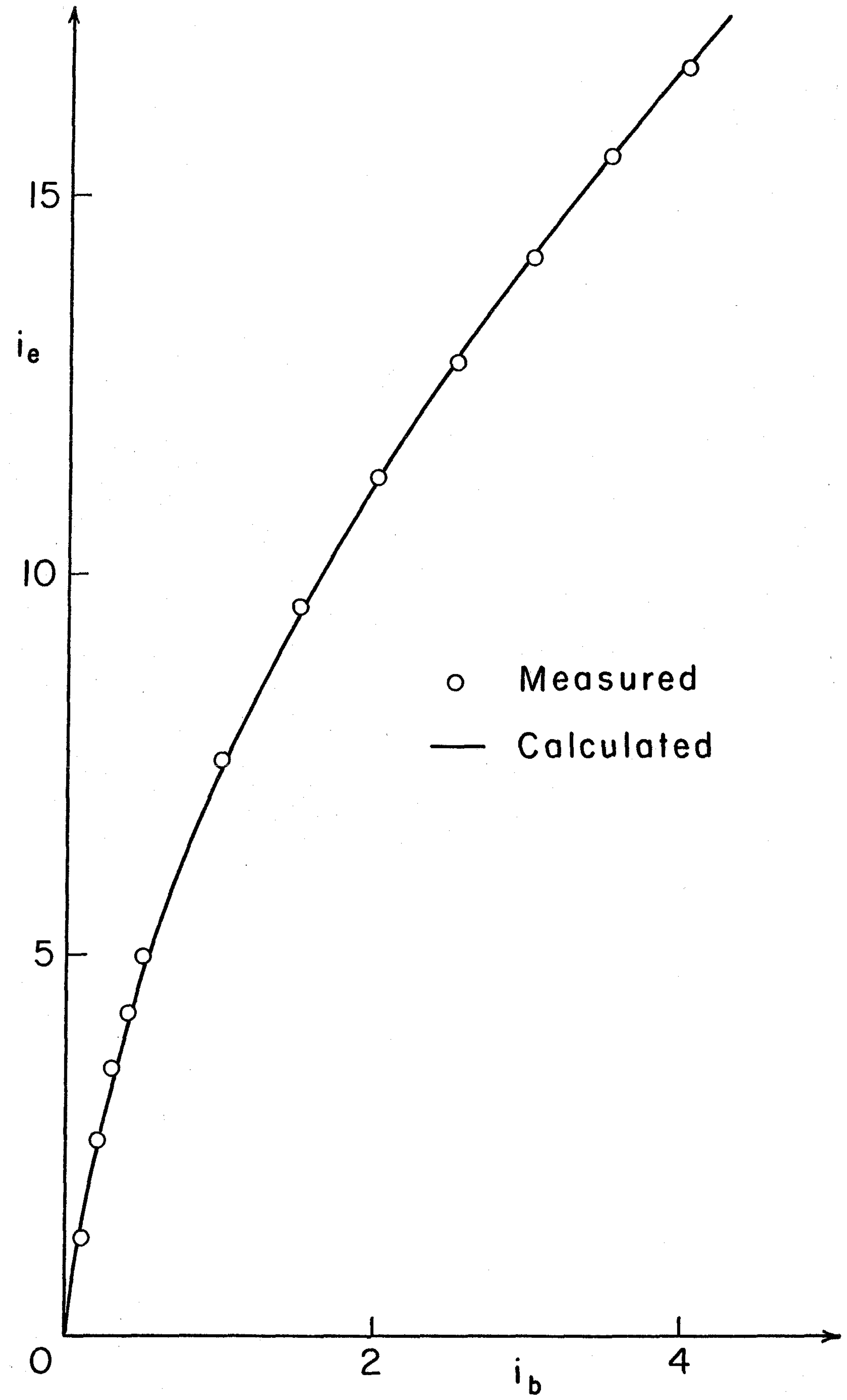


Fig. 24. - Experimental verification of equation 39.

point where the gradient of minority carriers is just sufficient to supply the necessary collector current. As the density curve rises higher, the saturation voltage becomes lower. The limit is reached when the total recombination in the base and injection into the emitter and collector becomes equal to the base driving current. In the linear theory the exact source of the base current was not important. However, as high level conditions are reached, the self-bias effect crowds the carriers toward the outer edge of the collector and hence a large part of the collector junction may be essentially cut off. A rapid increase in base current for a given degree of saturation is caused by the large exposed collector area which has no facing emitter junction and hence carriers are injected only to be lost in the base region.[13]

With the alloying techniques in common use, the base region outside the emitter is considerably wider than inside the active area of the device, and crowding effects may be neglected in this region. Suppose both emitter and collector junctions are forward biased such that at $x = 0$ their respective injected densities are ρ_1 and ρ_2. The analysis of this case is much the same as that used in connection with the normally biased transistor, and is carried out in detail in Appendix D. Under the assumption that the value of A is the same at emitter and collector, the results obtained indicate a decrease in G_d which is identical with that given in Section 3, provided ρ is replaced by $\rho_1 + \rho_2$. As before two special cases are considered: the near uniform and severely crowded approximations. In the former case a closed form solution for the saturation voltage as a function of i_c and i_b may be obtained as shown in Appendix D.

$$v_{sat} \approx \frac{kT}{q} \, \mathrm{Ln} \left[1 + \frac{i_c(\beta + \beta_i) + \beta i_b}{\beta_i(\beta i_b - i_c)} \quad (1 + c) \right] \tag{48}$$

where the correction term

$$c = K \, \frac{i_c(\beta - \beta_i) + \beta i_b(2\beta_i - 1)}{\beta + \beta_i}$$

has been assumed small, and β has been assumed large compared with unity. In this expression K is the current amplification falloff factor as given by equation 38 in connection with the normally biased case. The equations in Appendix D also provide a solution for ρ_1 and ρ_2 and hence the individual junction voltages if such information is needed. It should be noted that ohmic resistances of the collector and emitter are usually negligible and hence the saturation voltage given by equation 48 should be quite accurate. However, in calculating the individual junction voltages the $i_b R_b$ voltage may be dominant, and should always be taken into account.

Low-level switching operation. One of the most important uses of transistors under the conditions of the present approximation is that of a low-level signal switch. For this purpose a transistor is universally used in the inverted connection because of the lower saturation voltage obtainable. In this application the emitter current is very small and the transistor is used merely to transfer voltage information from collector to emitter. If the transistor is to operate with any reasonable success in such a connection ρ_1 and ρ_2 must be very nearly equal, and we may greatly simplify the analysis. From the linear lumped model the saturation voltage under these conditions is predicted to be

$$v_{sat} \approx \frac{kT}{q\beta}$$

However, such operation is quite sensitive to the crowding or self-bias cutoff effect. If we assume $\rho_1 \approx \rho_2 \approx \rho$ and $i_e \ll i_b$ the base current may be written

$$i_b \approx \rho G_{do} \left(\frac{1}{\beta} + \frac{1}{\beta_i}\right)$$

The saturation voltage may be evaluated from equation 48, interchanging the inverse and normal parameters, neglecting the i_e terms and assuming $\beta \gg 1$.

$$v_{sat} \approx \frac{kT}{q} \operatorname{Ln}\left(1 + \frac{1+c}{\beta}\right) \tag{49}$$

where $$c \approx \frac{2K\beta\beta_i \, i_b}{\beta + \beta_i}$$

Since it is desired to obtain the lowest possible saturation voltage, operation at low base currents is seen to be desirable. However, a lower limit is imposed by the junction depletion layer recombination which causes G_1 to increase at low current levels and therefore results in an increased saturation voltage.

Another limitation imposed on the drive current is that for any application, the transistor is required to supply <u>some</u> signal current, even if this current is very small. The drive current must not be less than $1/\beta_i$ of this required minimum value of signal current. It is true that in many cases the transistor signal switch may be used in connection with a self-balancing circuit so at balance no current is required. However, when the circuit is off balance, signal current is required to operate the self-balancing mechanism. Therefore the drive limitation may never be completely ignored. Operation in the region where depletion layer recombination is important is not recommended for another very important reason. The effective G_1 due to this effect and hence the effective β is very dependent upon temperature. Hence the saturation voltage stability with temperature at small drive currents may be much worse than at higher drive currents. Transistors made of material free from dislocations and extraneous impurities are much superior in this respect.

In any transistor which is at all suitable for this type of operation and which is driven with reasonably small base current, the second term in equation 49 will be small compared to unity. Thus the logarithm may be expanded in terms of its argument and equation 49 becomes

$$v_{sat} \approx \frac{kT}{q\beta}\left(1 + \frac{2K\beta\beta_i i_b}{\beta + \beta_i}\right) \tag{50}$$

where again the correction term has been assumed small compared with unity.

Experimental values of v_{sat} for an alloy transistor operating at low level are shown in Fig. 25. The theoretical values fall somewhere within the limits shown depending upon the exact values of β and β_i used in the calculation. The analysis is especially critical to the precise value of β_i used. Perhaps the most satisfactory method of obtaining β_i and K is to fit the i_e vs i_b data to Fig. 18 and use the normalization constants to determine these parameters. The true low level value of β_i is often masked by depletion layer recombination at low operating currents. This effect is quite noticeable in the experimental data of Fig. 25 where the saturation voltage increases slightly at low base currents.

The general requirements on a transistor to be used in low level switching service are that the saturation voltage should be as independent of time, temperature, and variations in drive current as possible. Thus, for efficient operation in this type of application,

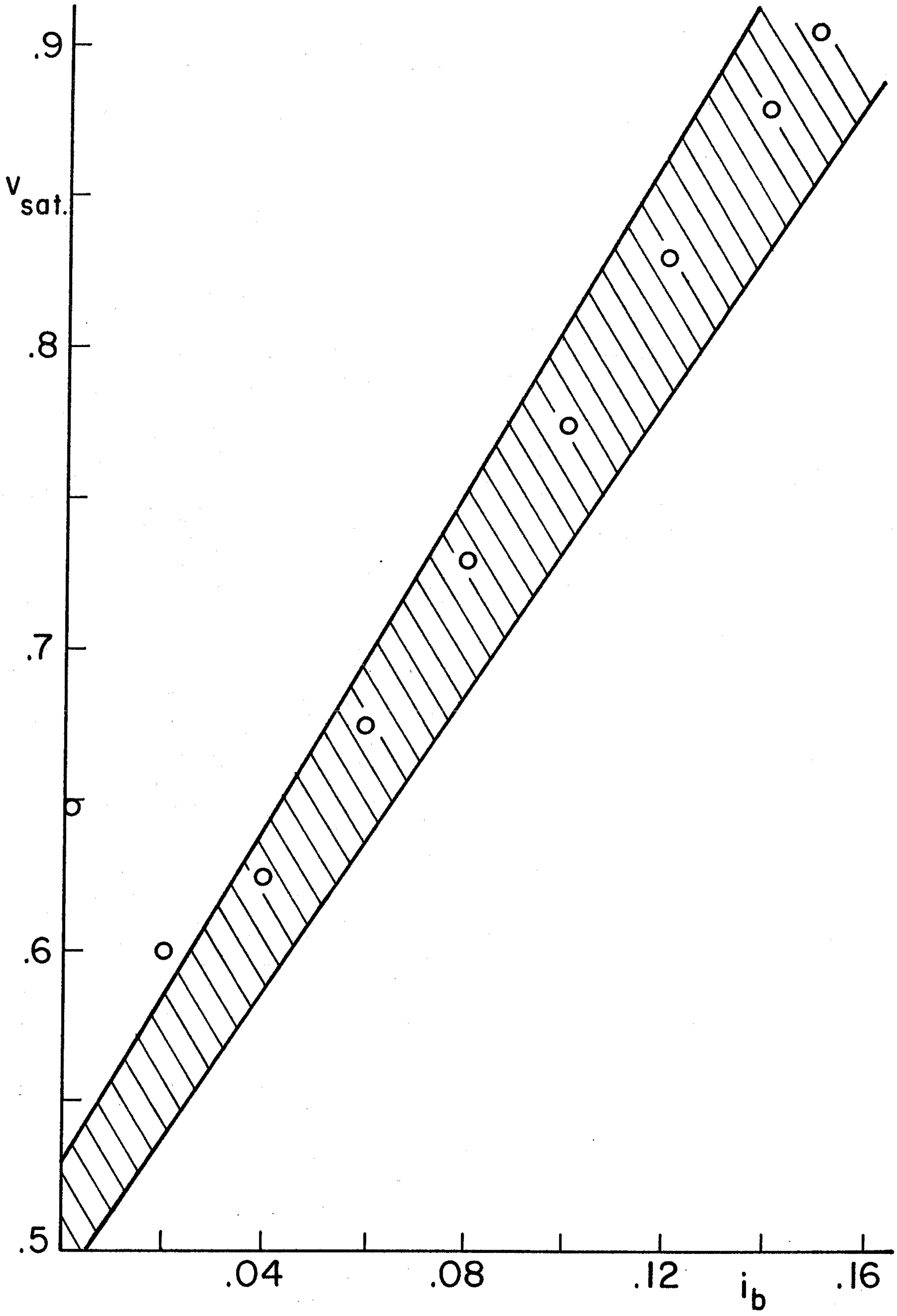


Fig. 25. - Experimental verification of equation 50.

the transistor should have a very high β and low value of K. Also the depletion layer recombination should not become important until very low levels are reached.

An obvious extension of the transistor signal switch is to operate two such switches back-to-back so that their saturation voltages tend to cancel. This procedure has often been used with some success. The problems involved are primarily those of matching β between units, although the variation of v_{sat} with drive current is often objectionable and a matching of the drive correction factor may also be attempted, with the obvious increase in labor involved. Again it should be noted that the depletion layer recombination rate is extremely variable between units and hence operation in the region where this condition is important is even more to be avoided.

The potential applications of transistor low level signal switches are very numerous and exciting. In nearly all such applications the requirements are sufficiently critical that matched back-to-back pairs are desirable. It would seem highly desirable that optimally designed matched units in the same package, mounted on the same mounting base, should be made available to the circuit designer in the same way that avalanche reference elements have been in the past. Perhaps one reason for the overlong delay is the general lack of understanding of the principles underlying the operation of devices used in this type of service. It is hoped that this work will help to clarify such operation and hasten the day when transistors will replace mechanical elements in many low level signal switching applications.

Well-crowded approximation. In large area devices with thin base regions it is often possible to reach the condition of large crowding well before the injected density becomes large compared with the majority carrier density. In such cases the analysis may again be simplified. The results of these calculations give ρ_1 and ρ_2 as functions of i_c and i_b and are presented in Appendix D.

Very high level. If we attempt to solve for the saturated density distribution under very high level conditions by returning to the initial integral equations, we are faced with an extremely difficult problem. The introduction of the second variable ρ_2 introduces tremendous complications into the mathematical manipulation and renders a closed form solution impossible. However, before giving up all hope and resorting to numerical methods, let us consider the real problem with which we are faced. We have derived expressions for the saturation voltage under the assumption that the lumped model elements were linear, and have seen how correction terms were added even at low levels due to the self-bias effect. However, as we have noted, in transistors designed to eliminate the density crowding, (i.e., those with thin emitter stripes) the injected density may well become comparable with n_n at current levels where G_d has not decreased appreciably below its low level value. In switching service such a unit will nearly always be used in the very high level region and we are interested in determining the first order correction to the linear theory under these conditions. As an approximation we will use the variation of G_d with ρ as given in equation 45. Since this variation of G_d with ρ is of the same form as in the low level case, we will assume the previous analysis applies and may use the results with the following modifications.

a) In units where ρ becomes comparable with n_n at reasonably low currents, the value of β as measured or given by the manufacturer is normally twice the low level value due to the drift enhancement of G_d mentioned earlier. Since at higher levels this effect is still present we do not need to add any additional factors to the analysis, but need only be sure that the value of β has been measured at a high enough current that the drift enhancement has become effective.

b) An additional term must be added to the saturation voltage due to the electric field in the base region caused by the gradient of majority carrier density. Under high level conditions the total potential drop across the base region is given by the condition that the drift and diffusion tendencies of majority carrier current are equal

$$\mu n E = D \frac{dn}{dy}$$

Since $n \approx \rho = \rho_1 - \left(\frac{\rho_1 - \rho_2}{w}\right) y$

where w is the base width; the electric field E is

$$E = \frac{D\left(\frac{\rho_1 - \rho_2}{w}\right)}{\mu\left[\rho_1 - \left(\frac{\rho_1 - \rho_2}{w}\right) y\right]}$$

which upon integration yields

$$V = \frac{kT}{q} \; Ln \; \frac{\rho_1}{\rho_2} \tag{51}$$

where we have used the Einstein relation

$$\frac{D}{\mu} = \frac{kT}{q}$$

Since this voltage is identical to the difference in forward bias voltages of the junctions and of the same polarity, the effect of high injection level is to double the observed saturation voltage over that given in the low level case.

c) It should again be emphasized that the present analysis has not included the effects of electron-hole scattering. The decrease in G_d due to this effect will become appreciable at extremely high levels and additional saturation voltage is to be expected.

The results of this section are quite significant to the circuit engineer in that they permit the first order characterization of transistor switches by the use of only one additional parameter, i.e., K the current amplification falloff factor. This modification to the basic theory permits calculations of many switching problems of interest to engineering accuracy. In addition it presents the clear physical insight into the fundamental limitations of the device so necessary for complex circuit work.

6. SATURATED CONDITIONS; DIFFUSED BASE TRANSISTOR

The case of a diffused base transistor in the saturated condition is of great interest since units of this type have excellent frequency response and high collector voltage breakdown, making them very attractive as power switches if reasonably low saturation voltages may be obtained. As pointed out earlier, such transistors are usually fabricated with a rather high resistivity collector region in order to achieve low collector capacitance and high voltage breakdown. For this reason when the transistor is saturated, the collector injection efficiency is quite poor and the current injected into the collector normally dominates the operation. Therefore, the analysis developed in connection with the current dependence of β_i may be used directly in the calculation of the saturation voltage. However, we must include the effects of the built-in field in the base region due to the gradient of majority carriers. As before, we shall consider the case of moderate injection level.

If we assume ρ in the base region to be large compared to n_n near the collector, but small compared to n_n near the emitter, we may obtain the form of the falloff in G_d with ρ_2^* from Fig. 18 making the further assumption that ρ near the emitter has very little effect upon the density crowding compared with that due to the collector injection efficiency.

Here

$$\rho_2^* = 3/2K \rho_2 G_{do} \tag{52}$$

where in this case

$$K = \frac{qRAd}{6FkT}$$

It now remains for us to compute the collector saturation voltage under these conditions of moderate injection level. This saturation voltage may be written as the sum of three terms

$$v_c = v_1 + v_2 + v_3 \tag{53}$$

where v_1 is the voltage due to the difference in forward bias voltages of the two junctions. From the discussion given in connection with the linear theory

$$v_1 \approx \frac{kT}{q} \text{Ln} \frac{\rho_1}{\rho_2} \tag{54}$$

v_2 is the voltage developed across the base region due to the gradient of majority carriers. This voltage is given by the condition that the diffusion and drift tendencies of majority carriers just cancel across the base region

$$v_2 = \frac{kT}{q} \, \mathrm{Ln} \, \frac{n_n}{\rho_2} \tag{55}$$

Here n_n is evaluated near the emitter

v_3 is the voltage drop across the series resistance of the collector region which is modified due to the conductivity modulation provided by carriers injected into the collector. Hall[22] has derived approximate expressions for the voltage drop across such a conductivity modulated section of high resistivity semiconductor as follows:

$$v_3 \approx \frac{2kT}{q} (\delta/L)^2 \qquad \delta \ll L \tag{56}$$

$$v_3 \approx \frac{kT}{2q} \, e^{\delta/L} \qquad \delta \gg L \tag{57}$$

where δ is the thickness of the high resistivity region and L is the diffusion length of minority carriers.

The importance of conductivity modulation of the collector region is clear from these equations. In many cases the major portion of the collector saturation voltage may originate in the collector region.

From these expressions it is possible to determine the saturation voltage of a diffused base transistor for any combination of collector and base currents. The first step is to solve for ρ_2 using the linear analysis presented in Part 1. The decrease in G_d is then evaluated from Fig. 18 or equation 39, normalization constants having been determined during the characterization of the transistor. This new value of G_d is then used to re-evaluate the operating point. In most cases, after one such iteration a quite good value for the saturation voltage may be obtained.

In many cases β_i is small compared with β and if the transistor is well saturated, nearly all the base current may be assumed due to ρ_2.

$$i_b \approx \rho_2 G_2$$

As before the collector current is given by

$$i_c = (\rho_1 - \rho_2) G_d - \rho_2 G_2$$

Therefore we may evaluate the ratio ρ_1 / ρ_2

$$\frac{\rho_1}{\rho_2} \approx 1 + \frac{1}{\beta_i} (i_c / i_b + 1) \tag{58}$$

where β_i is evaluated at the base current in question.

The condition for the validity of this approximation is

$$\frac{\rho_1}{\rho_2} << \frac{\beta}{\beta_i}$$

In any case, the above equation is a valuable starting point for the iterative procedure mentioned earlier.

Under widely varying conditions, different approximations may be in order. The results of Sections 3 and 4 may be applied directly to cases where slight crowding or severe crowding exists.

CONCLUSIONS

In Part I, a large signal model has been developed for the junction transistor which makes possible the analysis of commonly encountered switching problems with a great degree of facility. Several formulae have been derived which apply directly to cases of interest.

In Part II, this method of analysis has been extended to current ranges where the transistor operation may no longer be considered linear. A number of effects were considered which contribute to the non-linear behavior. In all cases it has been shown that the self bias cutoff effect plays an important part in the deterioration of performance at high currents.

In alloy transistors, both normally and inversely biased, an exact solution for the form of the current gain as a function of collector current has been given for the case where the injected density was small compared with the equilibrium base majority carrier density. This solution was shown to also be applicable to normally biased diffused base units. In all cases the decrease in current gain was due to the self bias effect. The sources of lateral base current were bulk recombination and injection into the emitter, the relative amount being immaterial since both were proportional to the injected density. In the small area alloy devices, it was seen that the injected density became large compared with the base majority carrier density before the transistor performance was appreciably degraded. In such cases the lateral base current was seen to normally be dominated by non-unity emitter efficiency.

In all cases, surface recombination was assumed to be the dominant source of total base current, and the first order correction to the linear theory was shown to exhibit a $1/(1 + Ki_c)$ dependence. At higher currents all cases followed the observed $1/i_c$ proportionality. Similar results were obtained for the inversely biased diffused base transistor.

The effect of non-linear phenomena on saturation voltage was considered in some detail. The saturation voltage was shown to increase with drive current, again emphasizing the importance of avoiding large overdrive currents. The use of alloy transistors as low level signal switches was discussed, indicating the necessity for high values of β and low values of K. Expressions for the saturation voltage of diffused base transistors were given. In many cases the major part of this voltage was found to be coursed by the ohmic resistance of the collector region. For low saturation voltage, the necessity for thin collector regions with long lifetimes was demonstrated.

APPENDIX A

1. RECOMBINATION RATE

It has been shown by Shockley and Read[16] and also Hall[23] that many of the observed recombination phenomena may be explained on the basis of an analysis which assumes the recombinations to occur at recombination centers or *traps* which possess a discrete energy level. The results of this analysis have recently[15] been applied to experiments involving silicon diodes and it was found that in typical cases the energy levels of the traps were a few kT from the intrinsic Fermi level. Using the notation of Sah et al.,[15] the recombination rate U is shown to be of the following form:

$$U = \frac{p_n - n_i^2}{(n + n_1)\ \tau_{po} + (p + p_1)\tau_{no}}$$

where n_1 and p_1 are the hole and electron densities to be expected if the Fermi level were located at the trap energy level, and τ_{po} and τ_{no} are respectively the lifetimes of holes and electrons as minority carriers in highly doped samples. If the hole and electron lifetimes are assumed equal, the expression simplifies to

$$U = \frac{1}{\tau_o} \frac{p_n - n_i^2}{n + p + n_1 + p_1} \qquad (A.1)$$

At low levels with traps near the intrinsic Fermi level, in an n-type semiconductor, for example $n \approx n_n$ and n_1, p_1 and p are small compared to n_n, the recombination rate becomes

$$U = \frac{n_n p - n_i^2}{n_n \tau_o}$$

which may be written

$$U = \frac{p - p_n}{\tau_o} = \rho / \tau_o \qquad (A.2)$$

If the traps are still assumed near the intrinsic Fermi level, we may investigate the change in recombination rate as p is increased, neglecting n_1 and p_1.

$$U \approx \frac{1}{\tau_o} \frac{p_n - n_i^2}{n + p}$$

The charge neutrality condition requires

$$n = n_n + \rho$$

therefore

$$U = \frac{1}{\tau_o} \frac{(n_n + \rho)(p_n + \rho) - p_n n_n}{n_n + p_n + 2\rho}$$

For reasonable injection levels and/or not near-intrinsic material, p_n may be neglected in comparison with ρ and/or n_n

$$U \approx \frac{1}{\tau_o} \frac{\rho n_n + \rho^2}{n_n + 2\rho} = \frac{\rho}{\tau_o} \frac{1 + \rho/n_n}{1 + 2\rho/n_n} \qquad (A.3)$$

Hence it can be seen that the bulk recombination rate and hence the value of the recombination elements associated with bulk recombinations decreases smoothly to one-half of their low level value. Later, we shall consider two special cases, where the injected minority carrier density is large and small compared to the equilibrium majority carrier density. In each case, we may assume the recombination rate proportional to the excess density, with the high level rate one-half of the low level rate. Often we may ignore the factor of two since other effects to be discussed make much larger contributions to the device operation.

2. JUNCTION INJECTION EFFICIENCIES

In the study of the effect of high level upon the injection efficiencies of the junctions, we must consider the deviation of the junction boundary value from the low level exp(qv/kT). Under high level conditions the majority carrier density may no longer be considered constant and the boundary condition for either region becomes

$$pn = n_i^2 \, e^{qv/kT} \tag{A.4}$$

where n_i is the equilibrium density of either holes or electrons in an intrinsic specimen at a given temperature. The charge neutrality condition requires the increase in majority electron density to be equal to the increase in hole density

$$p = p_n + \rho$$

$$n = n_n + \rho$$

Thus the junction law may be written

$$p_n n_n + \rho(p_n + n_n) + \rho^2 = n_i^2 \, e^{qv/kT}$$

If we recognize that $p_n n_n = n_i^2$

$$p_n (e^{qv/kT} - 1) = \frac{\rho^2}{n_n} + \rho\left(1 + \frac{p_n}{n_n}\right)$$

We shall now consider two special cases of practical importance.

a) Injected density in p-region small compared with equilibrium hole density (p_p) and n-region not intrinsic $(p_n/n_n \ll 1)$. In this case the injected excess electron density (η) becomes

$$\eta = n_p (e^{qv/kT} - 1)$$

However, the injected hole density is given by

$$p_n(e^{qv/kT} - 1) = \rho\ (1 + \rho/n_n) \qquad (A.5)$$

The quantity of interest is the total lateral electron current flowing in the base for a given value of ρ. The electron current density due to diffusion and recombination in the p-region is proportional to η,

$$J_n \sim (e^{qv/kT} - 1)$$

Hence

$$J_n \sim \rho(1 + \frac{\rho}{n_n}). \qquad (A.6)$$

This result is quite significant for the emitter junction of a normally biased transistor.

b) Injected density in both regions large compared with equilibrium majority density. In this case

$$p \approx n \gg n_n, p_p$$

in both regions and the junction law may be written

$$\rho \approx n_i\ e^{qv/2kT} \quad \text{for the n-region} \qquad (A.7a)$$

$$\eta \approx n_i \, e^{qv/2kT} \qquad \text{for the p-region} \qquad \textbf{(A.7b)}$$

The recombination rate from equation A. 3 is proportional to the excess density

$$U = \frac{\rho}{2\,\tau_o}$$

This case is of interest in connection with the collector junction of diffused-base transistors operated in the saturated or inverse biased conditions.

APPENDIX B

LOW LEVEL ALLOY TRANSISTOR

Under the condition $\rho \gg n_n$ the governing equations are

$$-i(x) = A \int_d^x \rho(x)\, dx$$

$$-v(x) = R \int_o^x i(x)\, dx$$

$$\rho(x) = \rho(0)\, e^{-qv(x)/kT}$$

It may be verified by direct substitution that the solution of this set of simultaneous, non-linear, integral equations is[20]

$$i(x) = K_1 \operatorname{Tan} K_2(d - x) \qquad \text{(B. 1a)}$$

$$\rho(x) = \rho(d) \operatorname{Sec}^2 K_2(d - x) \qquad \text{(B. 1b)}$$

where

$$K_1^{\,2} = \frac{2kT\ \rho(d)\, A}{qR}$$

$$K_2^{\,2} = \frac{\rho(d)\ ARq}{2kT}$$

It should be noted that both expressions contain the value of at the center of the emitter. The value of the injected density at the edge of the emitter is

$$\rho_1 = \rho(0) = \rho(d)\ \mathrm{Sec}^2\, K_2 d \qquad \text{(B.2)}$$

Since the value of ρ_1 does not determine the distribution of ρ as given by equation B.1, we must regard these expressions not as complete solutions but only as indications of the form of the density and current as a function of x.

Since the lateral base current evaluated at the edge of the emitter i(0) is proportional to the area under the density curve, and since the total collector diffusion current (i_c) is also proportional to the area under the density curve, $i_c = F \int_0^x \rho(x)\, dx$. The ratio of the two is a constant

$$i_c = \frac{F}{A}\ i(0) \qquad \text{(B.3)}$$

Thus one might expect the forward current gain β of the transistor to be constant. However, it should be recalled that the effects of bulk recombination and emitter efficiency on β are normally not appreciable compared with the effect of surface recombination.[2] Even an approximate analysis of the surface recombination current is extremely involved and becomes so entangled with the exact geometry of the specific transistor as to be useless in general. However, it has been shown that for a wide variety of practical cases, the surface recombination may be considered nearly proportional to the excess density at the edge of the emitter.[17] This amounts to assuming G_1 independent of ρ . Hence, if for the purpose of determining the total base current we neglect bulk recombination and the effects of non-unity emitter efficiency, i_b is proportional to ρ_1 . We will now investigate the dependence of G_d on the injection level. From equation B.1 we may write

$$i_c = \frac{F}{A} K_1 \operatorname{Tan} K_2 d \qquad \text{(B.4a)}$$

and also

$$\rho_1 = \rho(d) \operatorname{Sec}^2 K_2 d \qquad \text{(B.4b)}$$

thus

$$\left(\frac{i_c A}{F K_1}\right)^2 + 1 = \operatorname{Sec}^2 K_2 d$$

hence

$$\rho_1 = \rho(d) + \frac{3}{2} i_c^{\ 2} \ K/G_{do} \qquad \text{(B.5)}$$

where

$$K = \frac{qRAd}{3kTF}$$

Thus we may solve for $\rho(d)$

$$\rho(d) = \rho_1 - \frac{3}{2}\frac{K}{G_{do}} i_c^2 \qquad (B.6)$$

Let us now investigate the special limiting case where $i_c^2 K$ is very small compared with $\rho(d)\,G_{do}$ and the excess density is uniform across the base region. Under these conditions $K_2 d$ is small and we may approximate the tangent by its argument

$$i_c \approx \frac{F}{A} K_1 K_2 d = Fd\,\rho(d) \approx Fd\,\rho(d)$$

Hence the value of G_d at low levels (G_{do}) is seen to be

$$G_{do} = Fd$$

Returning to the more general case, equations B.4 and B.6 may be written

$$i_c = \sqrt{\frac{2\rho_1 G_{do}}{3K} - i_c^2}\ \mathrm{Tan}\ \frac{RdqA}{2kTF}\sqrt{\frac{2\rho_1 G_{do}}{3K} - i_c^2} \qquad (B.7)$$

In order to generalize the result of this analysis, we will express the quantities of interest in dimensionless normalized form

$$\rho^* = \frac{3}{2}\rho G_{do} K$$

$$i_c^* = \frac{3}{2} i_c K$$

$$G_d^* = \frac{G_d}{G_{do}}$$

using this notation equation B.7 becomes

$$i_c^* = \sqrt{\rho_1^* - i_c^{*2}} \;\text{Tan}\; \sqrt{\rho_1^* - i_c^{*2}} \tag{B.8}$$

The radical may be considered as a parameter by which the values of ρ_1^* and i_c^* are connected. It is also a significant measure of the extent to which the carriers are crowded to the edge of the emitter. At low level, $i_c^* \ll 1$ and equation B.8 reduces to

$$i_c^* \approx \rho_1^*$$

or

$$i_c \approx G_{do}\, \rho_1$$

as before. However, as the injection level increases, i_c^{*2} becomes appreciable and G_d decreases. If we let the parameter

$$\theta = \rho_1^* - i_c^{*2} \tag{B.9}$$

then from equation B.7

$$i_c^* = \theta \;\text{Tan}\; \theta \tag{B.10a}$$

and from equation B.4

$$\rho_1^* = \theta \;\text{Sec}^2 \theta \tag{B.10b}$$

Near-uniform approximation. At low levels, where $\theta \ll 1$ we may obtain the limiting form of the solution by expanding the trigonometric functions in equation B.10 and retaining only the first two terms.

$$i_c^* \approx \frac{\theta^2(1 - \theta^2/6)}{1 - \theta^2/2} \approx \theta^2(1 + \theta^2/3)$$

$$\rho_1^* \approx \frac{\theta^2}{(1 - \theta^2/2)^2} \approx \theta^2(1 + \theta^2)$$

Substituting

$$\rho_1^* \approx i_c^*(1 + 2/3\, i_c^*) \qquad \text{(B.11)}$$

hence the low-level limiting form of G_d becomes

$$G_d \approx \frac{G_{do}}{1 + Ki_c} \qquad \text{(B.12)}$$

Since the correction term has been assumed small, we may find the variation of G_d with ρ to first order by using the low-level value of G_d to replace i_c

$$G_d \approx \frac{G_{do}}{1 + K\rho_1 G_{do}} \qquad \text{(B.13)}$$

where again

$$K = \frac{qRAd}{3FkT}$$

If we assume surface recombination dominant G_1 is approximately constant and the DC current gain of the transistor may be written in the form

$$\beta = \frac{\beta_o}{1 + Ki_c} \tag{B.14}$$

This expression is quite useful, since in many transistors (notably large area types) the self-bias effect becomes important well before other large signal phenomena. Under these conditions it is often not desired to operate the transistor at current levels where β has fallen to a very great extent and hence a specification of K and β_o completely characterizes the large signal operation of the device. In other cases, high-level effects also become important and may not be ignored. These effects will be considered shortly. First, however, we shall derive the limiting form of equation B.10 where severe crowding exists, yet the assumptions made are still valid.

Severely crowded approximation. As the density $\rho(0)$ becomes large compared with $\rho(d)$, θ approaches $\pi/2$. If we let

$$\theta = \pi/2 - \epsilon$$

and expand the trigonometric functions in powers of ϵ, retaining only the first three terms, we obtain

$$\rho^* \approx i_c^{*2} + \frac{\pi^2}{4} \qquad \rho^* \gg 1 \qquad \text{(B.15)}$$

In the region where this expression is valid, the constant term may be ignored without appreciably affecting the accuracy. Under these conditions,

$$G_d^* \approx \frac{2}{3\,Ki_c} \qquad \text{(B.16)}$$

which has been recognized as the established limiting form for the variation of current gain of a junction transistor for many years.[11,14] It is interesting that this expression agrees with the high current limit of equation B.14 within the factor of 3/2.

APPENDIX C

VERY HIGH LEVEL ALLOY TRANSISTOR

For the case $\rho \gg n_n$ the governing equations are

$$i(x) = A \int_d^x \frac{\rho(x)^2}{n_n} \, dx \qquad (C.1)$$

$$-v(x) = \frac{Rn_n}{1+b} \int_0^x \frac{i(x)}{\rho(x)} \, dx \qquad (C.2)$$

$$\rho(x) = \rho(0)\, e^{-qv(x)/kT} \qquad (C.3)$$

These equations may be converted to a single differential equation as follows:

From equation C.3

$$-\,\mathrm{Ln}\, \frac{\rho(x)}{\rho(0)} = \frac{q\,v(x)}{kT}$$

or

$$\frac{q}{kT}\,\frac{d\,v(x)}{dx} = -\,\frac{1}{\rho(x)}\,\frac{d\,\rho(x)}{dx}$$

which, when substituted into equation C.2 becomes

$$\frac{1}{\rho(x)} \frac{d\,\rho(x)}{dx} = \frac{qR\,n_n}{(1+b)kT} \frac{i(x)}{\rho(x)}$$

or

$$i(x) = \frac{(1+b)kT}{qR\,n_n} \frac{d\,\rho(x)}{dx}$$

which, from equation C.1 becomes

$$\frac{A}{n_n} \rho(x)^2 = \frac{(1+b)kT}{qR\,n_n} \frac{d^2\,\rho(x)}{dx^2}$$

which may be written in the more convenient form

$$\frac{d^2\,\rho(x)}{dx^2} = \frac{qRA}{(1+b)kT} \rho(x)^2 \qquad \text{(C.4)}$$

We shall consider two approximate solutions to equation C.4. The first applies to cases where the density crowding is not too severe. In this case a solution of the form

$$\rho(x) = \rho(d)\left[1 + \frac{\rho(d)\,qRA}{2(1+b)kT}(d-x)^2\right] \qquad \text{(C.5)}$$

It is good to first order.[24] The collector current may be written

$$i_c = F\int_0^d \rho(x)\,dx = F\,d\,\rho(d)\left[1 + \frac{qRA\,d^2}{6(1+b)kT}\rho(d)\right] \qquad \text{(C.6)}$$

Again assuming the correction term small, we may write i_c in terms of $\rho(0)$ from equations C.5 and C.6

$$i_c \approx \rho_1 \frac{G_{do}}{1 + K\rho_1} \tag{C.7}$$

where

$$K = \frac{qRAd^2}{3(1 + b)kT}$$

and ρ_1 has been written for $\rho(0)$. These expressions are seen to be identical in form with those given in the low level, slightly crowded case.

Under some conditions the crowding will be very severe and we may write a solution to equation C.4 in the form

$$\rho(x) = \frac{1}{(C + Dx)^2} \tag{C.8}$$

where

$$C = \frac{1}{\sqrt{\rho_1}}$$

$$D = \sqrt{\frac{qBA}{6(1 + b)kT}}$$

It is clear that equation C.8 does not rigorously match the boundary conditions of the problem unless d is very large compared with the distance in which ρ becomes quite small. Hence, although it is an exact solution to the differential equation, it is only an approximate solution to the problem, applicable under severely crowded conditions.

In this case we may assume d large and extend the limit of the integration in equation C.6 to infinity

$$i_c \approx F \int_0^\infty \rho(x)\, dx = 2F \sqrt{\frac{6(1+b)kT}{qBA}}\ \rho_1 \qquad (C.9)$$

which again predicts a collector current which increases as the square root of the injected density ρ_1, and hence a diffusion conductance inversely proportional to the collector current.

The lateral base current i(0) may be determined from the integration of equation C.1

$$i(0) = \frac{A}{5n_n} \sqrt{\frac{qBA}{6(1+b)kT}}\ \rho_1^{\frac{5}{2}} \qquad (C.10)$$

The total base current contains in addition a surface recombination term proportional to ρ_1

$$i_b = \rho_1 \left(G_s + \frac{A\ \rho_1^{\frac{3}{2}}}{5n_n} \sqrt{\frac{qBA}{6(1+b)kT}} \right) \qquad (C.11)$$

where G_s is the effective surface recombination conductance.

APPENDIX D

SATURATION - ALLOY TRANSISTOR

We shall consider the low level case where $\rho \ll n_n$. The governing equations are then similar to equations 31, 32 and 33 except the contribution of the forward biased collector must be taken into account. If A is assumed the same at emitter and collector;

$$i(x) = A \int_{d}^{x} \rho_e(x') + \rho_c(x')\, dx' \qquad (D.1)$$

$$-v(x) = R \int_{0}^{x} i(x')\, dx' \qquad (D.2)$$

The junction voltage equations become

$$\rho_e(x) = p_n \left(e^{\frac{q}{kT}\left[v_e(o) - v(x)\right]} - 1 \right) \qquad (D.3a)$$

$$\rho_c(x) = p_n \left(e^{\frac{q}{kT}\left[v_c(o) - v(x)\right]} - 1 \right) \qquad (D.3b)$$

If we assume that ρ_1 and ρ_2 are both large compared with p_n, we may neglect the unity in the brackets. At first glance this procedure may not seem legitimate since the form of the expression for ρ will not be accurate as ρ becomes of the order of p_n. However, it should be borne in mind that all the results of the analysis involve integrals of the density curves and since the density in the region of interest is much greater than p_n, the region where the density is of the order of p_n will not make an appreciable contribution to the integral. Making this assumption, from equation D. 3

$$\rho_e(x) = \rho_1 e^{-\frac{qv(x)}{kT}} \qquad (D.4a)$$

$$\rho_c(x) = \rho_2 e^{-\frac{qv(x)}{kT}} \qquad (D.4b)$$

where, as in the linear theory

$$\rho_1 = \rho_e(o)$$

$$\rho_2 = \rho_c(o)$$

If we substitute $\rho_e + \rho_c$ for ρ we see that these equations are identical in form with equations of the low level forward bias case. We may therefore write the solution directly from equation B. 1

$$\rho_c = \rho_c(d) \sec^2 K_2(d - x) \qquad (D.5a)$$

$$\rho_e = \rho_e(d) \sec^2 K_2(d - x) \qquad (D.5b)$$

where now

$$K_2^{\,2} = \frac{\left[\rho_e(d) + \rho_c(d)\right] ABq}{2kT}$$

The collector diffusion current is proportional to the area under the curve of the difference in densities.

$$i_d = F \int_0^d \left[\rho_e(x) - \rho_c(x)\right] dx \qquad (D.6)$$

upon the substitution

$$\rho(0) = \rho(d) \sec^2 K_2 d$$

equation D.6 becomes

$$i_d = \frac{G_{do}}{2K_2 d} (\rho_1 - \rho_2) \sin 2 K_2 d \qquad (D.7)$$

The total collector current also includes the recombination term due to G_2

$$i_c = i_d - G_2 \rho_2$$

which may be appreciable in some applications.

If we assume surface recombination dominant, the base current may be written

$$i_b = G_1 \rho_1 + G_2 \rho_2 \tag{D.8}$$

Near-uniform approximation. If the density of carriers is not too much larger at the edge of the emitter than at the center, we may expand the trigonometric functions in terms of their argument, retaining only the first two terms. This condition is equivalent to the approximation

$$K_2 d \ll 1$$

From equation D.7

$$i_d \approx G_{do}(\rho_1 - \rho_2) \frac{1}{1 + \frac{2}{3} K_2^2 d^2} \tag{D.9}$$

where $G_{do} = Fd$ as before

We must now evaluate the factor K_2 in terms of the approximation

$$\sec^2 K_2 d \approx 1$$

Since this approximation only appears in the correction term, the expression for i_d will still be accurate to first order.

$$K_2^2 \approx \frac{(\rho_1 + \rho_2) ARq}{2kT}$$

Therefore the collector diffusion current may be written

$$i_d \approx G_{do}(\rho_1 - \rho_2) \frac{1}{1 + K(\rho_1 + \rho_2) G_{do}} \qquad \text{(D.10)}$$

where $K = \frac{qRAd}{3kTF}$ as in equation 38. Equation D.10 is seen to be identical with equation 39 if ρ is replaced by $\rho_1 + \rho_2$.

Again we may approximate G_d by G_{do} in the correction term, in order to evaluate the correction directly in terms of the collector and base currents. If $\beta \gg 1$

$$i_d \approx G_{do}(\rho_1 - \rho_2) \frac{1}{1 + K \frac{i_c(\beta - \beta_i) + \beta i_b (2\beta_i - 1)}{\beta + \beta_i}} = G_{do}(\rho_1 - \rho_2) \frac{1}{1 + c} \qquad \text{(D.11)}$$

The quantities β and β_i are taken as the low level values.

Upon the substitutions

$$i_d = i_c + \rho_2 G_2$$

and

$$\rho_2 G_{do} = \frac{\beta_i}{\beta + \beta_i + 1} (\beta i_b - i_c)$$

equation D.11 becomes

$$i_c \approx \frac{(\beta i_b - i_c)}{\beta + \beta_i + 1} \left[1 + \beta_i \left(\frac{\rho_1}{\rho_2} - 1 \right) \frac{1}{(1 + c)} \right] \qquad \text{(D.12)}$$

where c is the correction factor given above.

In order to find the saturation voltage, we must solve for the ratio $\frac{\rho_1}{\rho_2}$ in equation D.12.

$$\frac{\rho_1}{\rho_2} \approx 1 + \frac{i_c(\beta + \beta_i) + \beta i_b}{\beta_i(\beta i_b - i_c)} \quad (1 + c) \tag{D.13}$$

where β has been assumed large compared with unity. It is seen that this expression agrees with equations 11 and 12 when K equals zero. As before, the saturation voltage may be calculated from the relation

$$v_c = \frac{kT}{q} \text{Ln} \frac{\rho_1}{\rho_2} \tag{D.14}$$

These equations provide a solution for the saturation voltage under any combination of collector and base current provided the transistor is well saturated and the crowding is not severe.

Severe crowding approximation. Under the condition of extreme crowding K_2d approaches $\frac{\pi}{2}$ and we may again simplify the analysis. If we let

$$K_2d = \frac{\pi}{2} - \epsilon$$

we again may approximate the trigonometric functions

$$i_d \approx \frac{G_{do}(\rho_1 - \rho_2)}{\pi} \epsilon$$

We may obtain ϵ in terms of the densities

$$\epsilon^2 = \frac{kT\pi^2}{(\rho_1 + \rho_2)\, 2ABqd^2} = \frac{\pi^2}{6K\, G_{do}(\rho_1 + \rho_2)}$$

In cases of interest the base drive current is small compared with the collector current and we may approximate the collector current by the diffusion current, neglecting the $\rho_2 G_2$ term.

$$i_c \approx \frac{G_{do}(\rho_1 - \rho_2)}{\sqrt{6K\, G_{do}(\rho_1 + \rho_2)}}$$

$$i_b = G_{do}\left(\frac{\rho_1}{\beta} + \frac{\rho_2}{\beta_i}\right)$$

These equations serve to determine ρ_1 and ρ_2 and hence the device saturation voltage for any combination of collector and base currents.

REFERENCES

1. Ebers, J. J. and J. L. Moll, "Large Signal Behavior of Junction Transistors," *Proc. IRE*, Vol. 42, (December 1954) p. 1761.

2. Middlebrook, R. D., *An Introduction to Junction Transistor Theory*, John Wiley and Sons, Inc., New York, (1957).

3. Linvill, J. G., "Lumped Models of Transistors and Diodes," *Proc. IRE*, Vol. 46, (June 1958) p. 1141.

4. Moll, J. L., "The Evolution of the Theory for the Voltage-Current Characteristic of P-N Junctions," *Proc. IRE*, Vol. 46, (June 1958) p. 1076.

5. Kingston, R. H., "Switching Time in Junction Diodes and Junction Transistors," *Proc. IRE*, Vol. 42, (May 1954) p. 829.

6. Early, J. M., "Effects of Space Charge Layer Widening in Junction Transistors," *Proc. IRE*, Vol. 40, (November 1952) p. 1401.

7. Gaicoletto, L. J., "Study of p-n-p Alloy Junction Transistors from D-C through Medium Frequencies," *RCA Review*, Vol. 15, (December 1954) p. 506.

8. Moll, J. L., "Large Signal Transient Response of Junction Transistors," *Proc. IRE*, Vol. 42, (December 1954) p. 1773.

9. Hurley, R. B., *Junction Transistor Electronics*, John Wiley and Sons, Inc., New York, (1958) Chapt. 19.

10. DeWitt, D. and A. L. Rossoff, *Transistor Electronics*, McGraw-Hill Book Co., Inc., New York, (1957) Chapt. 11.

11. Webster, W. M., "On the Variation of Junction Transistor Current Amplification Factor with Emitter Current," *Proc. IRE*, Vol. 42, (June 1954) p. 914.

12. Fletcher, N. H , "Some Aspects of the Design of Power Transistors," Proc. IRE, Vol. 43, (May 1955), p. 551.

"Self-Bias Cutoff Effect in Power Transistors," Proc. IRE, Vol. 43, (November 1955), p. 1669.

13. Rudenberg, H. G., "On the Effect of Base Resistance and Collector-to-base Overlap on the Saturation Voltages of Power Transistors," Proc. IRE, Vol. 46, (January 1958) p. 1304.

14. Rittner, E. S., "Extension of the Theory of the Junction Transistor," Phys. Rev., Vol. 94, (June 1954) p. 1161.

15. Sah, C. T., R. N. Noyce and W. Shockley, "Carrier Generation and Recombination in P-N Junctions and P-N Junction Characteristics," Proc. IRE, Vol. 45, (September 1957) p. 1229.

16. Shockley, W., and W. T. Read, Jr., "Statistics of the Re-Combinations of Electrons and Holes," Phys. Rev., Vol. 87, (September 1952) p. 835.

17. Moore, A. R., and J. I. Pankove, "The Effects of Junction Shape and Surface Recombination on Transistor Current Gain. Part I," Proc. IRE, Vol. 42, (June 1954) p. 907.

Stripp, K. F., and A. R. Moore, "The Effects of Junction Shape and Surface Recombination on Transistor Current Gain. Part II," Proc. IRE, (July 1955) p. 856.

18. Hall, R. N., "Power Rectifiers and Transistors," Proc. IRE, Vol. 40, (November 1952) p. 1512.

19. Clark, M., "Power Transistors," Proc. IRE, Vol. 46, (June 1958) p. 1185.

20. Looney, D. H., "Base Resistance Considerations in Transistor Design," Ch. 8, p. 76, Bell Telephone Labs., 14th Interim Technical Report on Transistor Services, (January 1955).

21. Shockley, W., "The Theory of p-n Junction in Semiconductors, and p-n Junction Transistors," Bell System Technical Journal, Vol. 28, (July 1949) p. 435.

22. Fletcher, N. H., "The High Current Limit for Semiconductor Junction Devices," Proc. IRE, Vol. 45, (June 1957) p. 862.

23. Hall, R. N., "Electron-Hole Recombination in Germanium," Phys. Rev., Vol. 87, (July 1952) p. 387.

24. Mortenson, K. E., "High Level Transistor Operation and Transport Capacitance," IRE Transaction on Electron Devices, Vol. ED-6, April 1959, p. 174.

25. Griffith, P. G., "Lumped Models of Drift Transistors for Large Signals," Stanford Electronics Laboratories Technical Report No. 1501-1, April 20, 1959.

26. Huang, C., "A Physical Theory of Junction Transistors in the Collector-Voltage-Saturation Region," IRE Transaction on Electron Devices, Vol. ED-6, April 1959, p. 141.