Transistor Switching Analysis

Dr. C. A. MEAD*

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 allow the analysis of complex switching problems with the ease of linear circuit calculations. 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.

Part 1

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.

Requirements of the Analytical Method

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, e.g., 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.

*California Institute of Technology, Pasadena, California.

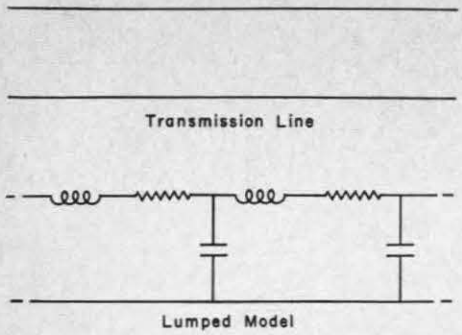


Fig. 1—Transmission line and equivalent lumped element representation.

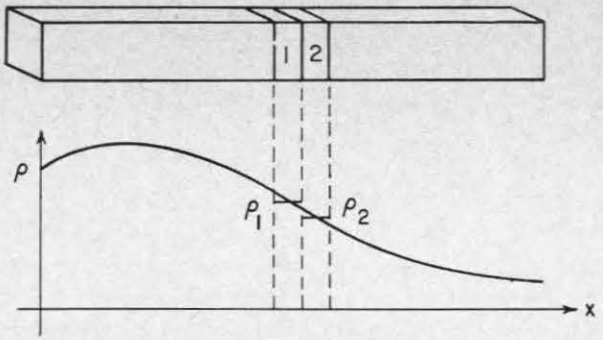


Fig. 2—Semiconductor bar with excess minority carrier distribution.

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

$$\frac{\partial N}{\partial t} = \frac{N_0 - N}{\tau} + D\frac{\partial^2 N}{\partial x^2}$$

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

N_0 is the density of minority carriers at thermal equilibrium

τ is the "lifetime" of minority carriers

D is the diffusion constant of minority carriers

x is the distance through the semiconductor

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

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.

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 time 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.

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*, $N - N_0$.

$$\frac{\partial (N - N_0)}{\partial t} = -\frac{(N - N_0)}{\tau} + D\frac{\partial^2 (N - N_0)}{\partial x^2}$$

For a p-type semiconductor: $N = n$, $N_0 = n_p$
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: $N = p$, $N_0 = n_n$
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.

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

Recombination. 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 $(\rho_1 - \rho_2)$.

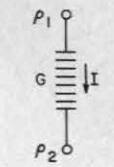
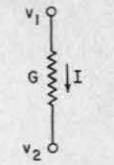
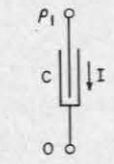
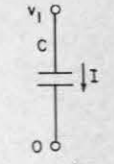
We may state the continuity equation, then, as follows:

> 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.

TABLE I

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

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.

As a check on the lumped model, let us assume a uniform excess minority carrier density in the semi-

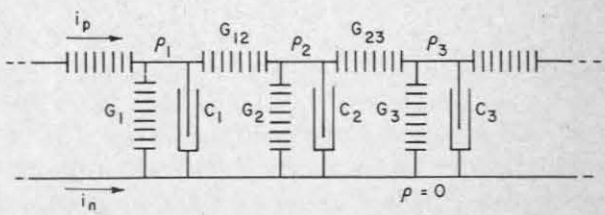
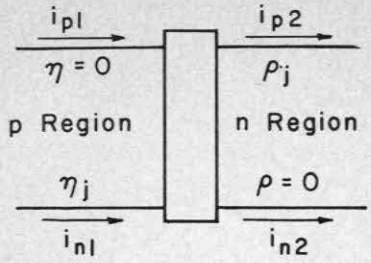


Fig. 3—Lumped model of semiconductor bar.



$$i_{p1} = i_{p2}$$
$$i_{n1} = i_{n2}$$
$$\rho_j = p_n(e^{qv/kT} - 1)$$
$$\eta_j = n_p(e^{qv/kT} - 1)$$

Fig. 4—Model of *p-n* junction.

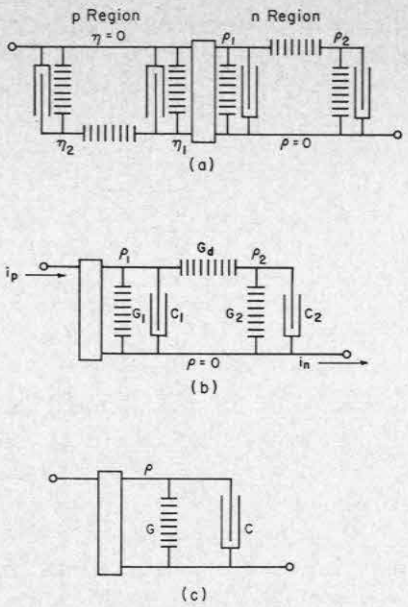


Fig. 5—Development of diode model.

conductor bar. Under these conditions the continuity equation becomes

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

which has the solution

$$\rho = \rho_0 e^{-t/\tau}$$

From the lumped model

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

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

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

which has the solution

$$\rho_2 = \rho_0 e^{-\frac{G_2}{C_2} t}$$

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

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

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 semiconductor 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:

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

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

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.

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.

Diode Small Signal Response. 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_0, we may write the injected excess density

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

where $v = v_0 + v_1 e^{j\omega t}$.

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

$$\rho \approx p_n \left[e^{qv_0/kT} \left(1 + \frac{q}{kT} v_1 e^{j\omega t} \right) - 1 \right]$$

Thus the *a-c* component of excess density is

$$\rho_1 e^{j\omega t} \approx \frac{q}{kT} v_1 p_n e^{qv_0/kT} e^{j\omega t}$$

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

$$\rho_0 \approx p_n e^{qv_0/kT} \tag{2}$$

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

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

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.

(To be continued)