ELECTRONIC NUCLEAR INSTRUMENTATION GROUP

SERVOMECHANISMS LABORATORY
DEPARTMENT OF ELECTRICAL ENGINEERING

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE - 39 - MASSACHUSETTS

Massachusetts Institute of Technology
Department of Electrical Engineering
Servomechanisms Laboratory

ELECTRONIC NUCLEAR INSTRUMENTATION GROUP

D.I.C. Project No. 6986

Technical Report No. 2

MODULATION AND DEMODULATION WITH SEMICONDUCTORS

by

Dale P. Masher

Office of Naval Research
Contract N5ori-07876
NR-025-164

Abstract: In a continuation of the present trend toward the development of more reliable and trouble-free control systems, this report discusses several modulation and demodulation systems which employ transistors. To broaden the foundation upon which the work is built, Chapter II is concerned with classifying such systems in a manner which permits logical study and analysis. Certain of the characteristics which obtain in optimum systems are also considered. The modulators and demodulators realized in Chapter III comprise, as nearly as possible, examples of linear systems in the optimum category of the classification.

JUN 15 1953

Approved: T. S. Gray

The research described in this document was made possible through support extended the Massachusetts Institute of Technology by the Office of Naval Research and the Navy Bureau of Ships.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

1.1 An Assessment of Present Treatments

Perhaps a subject as old as modulation and demodulation would seem to offer little more to the investigator than engineering improvement of present systems or the application of new devices to an old problem. Though this work is pointed to the latter phase, a close study of the problem discloses a more profound need. A survey of the literature which treats modulators and demodulators, including the engineering texts, reveals that most systems are treated as individual units having little or no relation to other systems. Furthermore, the methods of operation are commonly given in descriptive forms--an approach which can hardly be termed analytical. It is especially discouraging to find nothing concerning linearity, or worse, to find a statement to the effect that the presence of nonlinear elements at once voids any sort of linear analysis. This kind of statement is certainly misdirected, for it will be shown that any modulation or demodulation system must be approximately linear in its over-all operation in order to perform effectively. But such is the situation.

1.2 Purpose

The appearance of the transistor and other semiconductor devices has imparted an entirely new dimension to system reliability standards. The primary objective here is to extend the application of these semiconductors to low maintenance modulation-demodulation systems which may prove useful in nuclear instrumentation applications.

However, the general picture, as outlined above, indicates a fundamental need for a generalized approach to the entire problem. Accordingly, the work is conveniently divided into three phases:

(1) Creation of an orderly system of classification under which the heterogeneous mass of existing modulators and demodulators may be logically studied and analyzed.

(2) Establishment of the criteria or characteristics necessary for an optimum modulation or demodulation system.

(3) Realization of several of the systems in the optimum category using semiconductors, principally transistors.

1.3 Definition of Terms

Amplitude modulation is defined as the process in which the amplitude of a carrier signal is varied with time in accordance with an intelligence signal. Since this investigation is concerned solely with amplitude modulation, the qualifying adjective is hereafter omitted. A carrier is a wave suitable for being modulated. Examples of carriers include direct current, a series of pulses, and a sine wave, but the sinusoid is here picked as the most general carrier function and is used in all the analytical work.

Two basic types of modulation systems exist. If the intelligence signal is represented by $f(t)$, and the carrier by $\cos \omega_c t$, the modulated carrier, $f(t) \cos \omega_c t$, is obtained by multiplying the carrier and intelligence. The result is called carrier-suppressed modulation since an intelligence signal of a single angular frequency ω_m results in two components, one of frequency $\omega_c + \omega_m$ and one of frequency $\omega_c - \omega_m$. The carrier frequency is absent. If the carrier is added to the product above, the result is ordinary or carrier-present modulation which may be expressed as $\left[1 + f(t)\right] \cos \omega_c t$. A component at the carrier frequency is always present. Therefore, the coefficient of $\cos \omega_c t$ may or may not contain unity as a normalized term. The unit term indicates the presence of a zero-signal output level in the system and may represent nothing more than a d-c source in series with the intelligence.

It is fairly well-known that time-varying circuit elements, such as synchronous commutators and vibrating armatures, can be used to convert alternating current to direct current. Perhaps not so fully recognized is the fact that time-varying parameters may also be used to effect modulation. In the latter application, the time-varying parameter may be any physical device in which the conductance can be controlled as a function of time. Since the current flowing through such a device is the product of voltage and conductance, the input voltage is effectively multiplied by a time-varying function (the conductance) to produce a modulated current. This time-function multiplier is called the <u>sensitivity-function</u> of the device. Either the carrier or the intelligence may be used to control the sensitivity-function. If the carrier is used, the sensitivity-function is periodic at the carrier frequency. If the intelligence is used, the sensitivity-function is rarely periodic, usually being aperiodic or random. A common sensitivity-function is the square pulse, a periodic function which is easily generated with bistable elements--e.g., diodes, triodes, relays, and switches. Sinusoidal functions are possible, but are normally obtained only in the presence of pulsed inputs (e.g., in Class C modulators).

Filtering is almost invariably associated with the processes of modulation and demodulation. Although the sensitivity-function is often not physically separable from the filter, separation for the purpose of analysis is not precluded. The operation of multiplication is then assigned uniquely to the sensitivity-function, multiplication being considered synonymous with that of <u>frequency translation</u>. An example of sensitivity function analysis is given in the next chapter.

<u>Demodulation</u> or detection is defined as the process by which the intelligence or message is recovered from the modulated carrier. Demodulation thus also involves frequency translation and subsequent discussion shows that sensitivity-function analysis is equally applicable to demodulators where time-varying parameters are employed.

CHAPTER II

THEORY OF OPERATION

2.1 General Discussion

2.11 Linearity and Multiplication

As stated previously, the over-all operation of both modulators and demodulators must be approximately linear. Ideally the process of modulation involves the multiplication of the intelligence, $f(t)$, by the carrier, $\cos \omega_c t$, and subsequent addition of the carrier if carrier-present modulation is desired. Conversely, demodulation involves subtraction of the carrier, if it is present, and division by $\cos \omega_c t$ (i.e., multiplication by $\sec \omega_c t$). The contention made earlier that all these operations are linear is easily verified. It is only necessary to show that the superposition theorem is applicable. In other words, it is necessary to show that the sum of the responses to two separate excitations, each acting independently, is the same as the response obtained when both excitations are acting together. In this connection the zero-signal value of the response must be treated as though it resulted from the action of a separate source.

Let $\theta_o(t)$ be defined as the output of a modulator so that:

$$\theta_o(t) = \left[1 + f(t)\right] \cos \omega_c t \tag{2.1}$$

The zero-signal value of the output is then:

$$\cos \omega_c t;$$

the response to a signal, $f_1(t)$, is:

$$f_1(t) \cos \omega_c t;$$

and the response to a signal, $f_2(t)$ is:

$$f_2(t) \cos \omega_c t.$$

The sum of these responses is:

$$\left[1 + f_1(t) + f_2(t)\right] \cos \omega_c t.$$

Now consider the case where both excitation sources are acting together. Whether these sources produce currents or voltages, their combined input is that of their sum. This means that the intelligence is represented by the sum of $f_1(t)$ and $f_2(t)$. Substituting $f_1(t) + f_2(t)$ for $f(t)$ in equation (2.1) gives:

$$\theta_o(t) = \left[1 + f_1(t) + f_2(t)\right] \cos \omega_c t.$$

Comparing the two results, it is apparent that superposition applies, and since a similar argument holds for the demodulation process, the ideal modulator or demodulator is certainly linear. Hence, the output of a physical system must be approximately linear whether or not nonlinear elements are present. Such a conclusion also seems obvious from the fact that any product is linear if multiplicand and multiplier are independent. In modulation (and most demodulation) systems, not only are the factors independent, but one, the carrier, is an unchanging periodic function. It is true that superposition does not apply to one class of existing systems, but even then the operation must be approximately linear over some range if the apparatus is to perform as expected. This matter is discussed further in section 2.3.

2.12 Operational Diagrams

In order to facilitate the classification scheme which is to follow, let us consider in more detail the actual operation of modulators and demodulators. Figure 2.1(a) illustrates the operation of one type of modulator. There are two inputs to the system. The symbol $m(t)$, which characterizes one of the inputs, is used to represent either $1 + f(t)$ or $f(t)$ alone. The other input is the carrier itself. The block containing $S(t)$ constitutes the frequency translating or multiplying section of the system. The multiplying function, or sensitivity-function, is denoted by $S(t)$. Since the sensitivity-function may be controlled by either the carrier or the modulation, it is functionally related to one of the two. This functional dependence is indicated by the subscripts c and m respectively. Correspondingly, there are two possible outputs from the multiplier: $m(t)S_c(t)$ or $\cos \omega_c t S_m(t)$. The filter in the system is characterized by $h(t)$, its impulse response in the time domain. The output of the filter is then related to the input through the superposition integral. In practice, a band-pass filter centered at the carrier frequency is often used. Under proper conditions, the filter may be so designed that the system output approximates $m(t)\cos \omega_c t$ to any degree desired.

The operation of the demodulator in figure 2.1(b) follows in an analogous way the operation of the modulator. However, the input shown as $\cos \omega_c t$ may not be present at all, particularly if the other input is a carrier-present signal. In such a case the sensitivity-function may involve both carrier and intelligence. This problem is treated in the appendix.

It should be emphasized that the schematics in figure 2.1 characterize a large class of existing systems. In fact, the so-called square law devices constitute the only class which cannot be so embodied. Figure 2.2 illustrates the operation of a modulator and demodulator in this category. In both cases some sort of nonlinear element is employed. That is, the frequency translating section multiplies by virtue of a nonlinear device rather than a time-varying one. Though it is customary to utilize parabolic nonlinearities, other more abruptly changing volt-ampere characteristics are also useful. The significant feature in every case is the existence of the second order or squared term in the power series expansion of the nonlinearity.

The demodulator shown in figure 2.2 is effective only with carrier-present inputs, but the modulator may be so designed that the carrier is balanced out. Hence the output is represented by $m(t)\cos\omega_c t$, signifying either a carrier-present or carrier-suppressed signal. It should be noted that neither output is assumed to be distortionless since the nature of the operation precludes such an assumption. The outputs of the nonlinear devices are not indicated due to the complexity of the result. These matters are discussed more fully in sections 2.3 and 2.4.

2.13 The Principle of Equivalence

Upon close inspection of figures 2.1 and 2.2, one feature stands out significantly. Neglecting the filter, there is no essential difference between the frequency translating section of a modulator and demodulator in a given class. The implication is clear that any modulator will theoretically operate as a demodulator, and vice versa, with proper change of filter. The principle can be stated formally as follows:

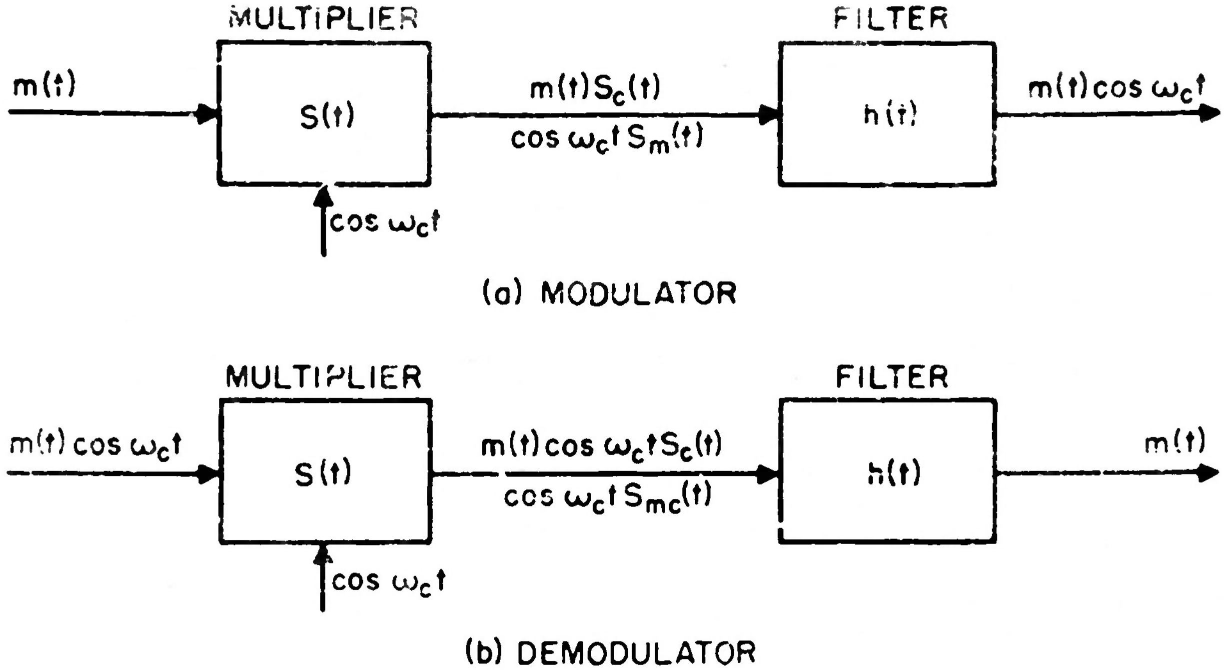


FIGURE 2.1 MODULATOR AND DEMODULATOR EMPLOYING SENSITIVITY-FUNCTIONS

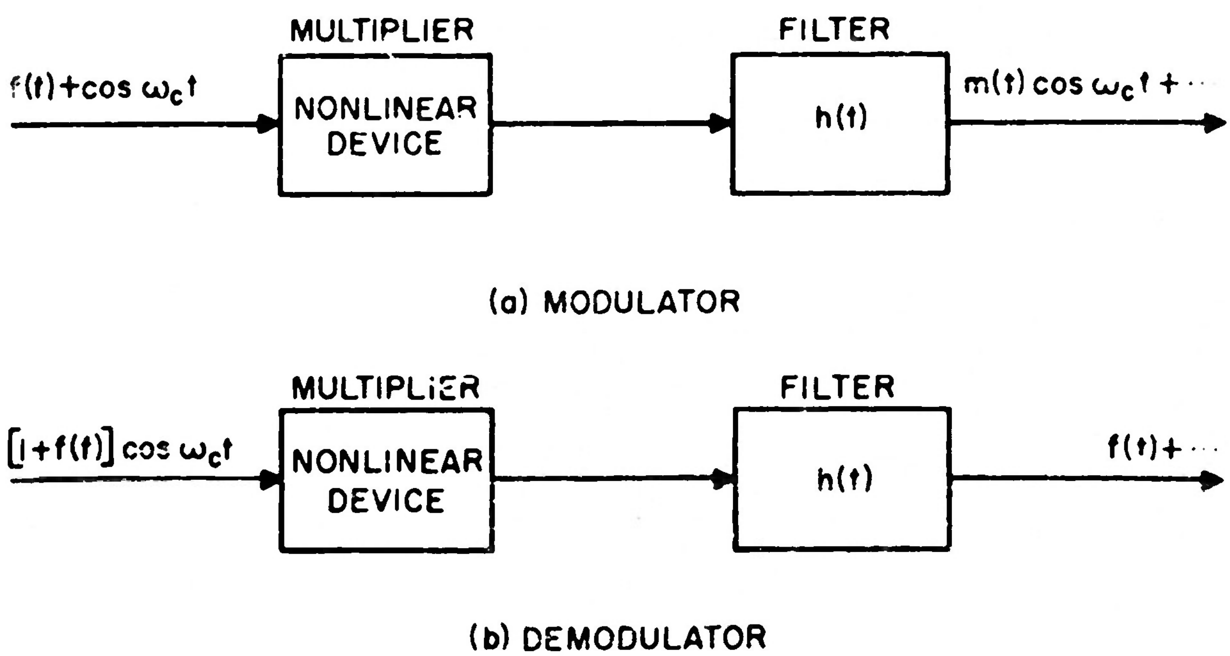


FIGURE 2.2 "SQUARE LAW" MODULATOR AND DEMODULATOR

Any modulator, exclusive of the associated filter, will demodulate a signal of the same form as its output if the input corresponding to the intelligence is activated with that signal. The converse is also true.

It is not to be inferred that a system designed as a modulator will necessarily provide a satisfactory demodulator. Practical considerations may show, for instance, that to demodulate a carrier-present signal with a carrier-present modulator requires that the carrier be available separately. Obviously, this is not a practical situation.

2.2 Classification

As indicated by the principle of equivalence, it is not necessary to make an individual classification for both modulators and demodulators. One treatment is sufficient. Furthermore, the preceding study has made it reasonably evident that the classification of modulating and demodulating systems logically proceeds from the operation of the frequency translating device. It has already been pointed out that two basic methods of frequency translation exist. One calls for a time-varying parameter, and the other, a nonlinear device. Present nomenclature lists a few systems which use a time-varying parameter, particularly demodulators, as linear, but the term is usually used with an apology, since it is meant to apply only to the piecewise linearity of a volt-ampere characteristic. Some authors even express regret that a power series expansion is not applicable. No such reservations in the use of the term linear are necessary, but in an attempt to avoid further confusion this writer chooses to call systems employing time-varying parameters first order systems. Analogously, those systems which rely upon square law nonlinearities for frequency translation are called second order systems. The names stem from the dependence of the system equations on either linear (first order) terms or square (second order) terms.

A further distinction seems necessary. Until now little has been said concerning the actual nature of the sensitivity-function. This function may possess a variety of forms, the square pulse and square wave being quite common. Impulse functions also appear in practice; in fact, almost any wave shape is conceivable. In order to partially differentiate among these waveforms, two subclassifications are made. A sampled data system is said to exist when the sensitivity-function possesses some characteristic distribution of finite intervals where the function is zero. Sensitivity-functions which exhibit only discrete zero-crossings distinguish the continuous data systems. The periodic square pulse and the square wave, respectively, are examples of the two types of sensitivity-functions.

Examples of the Classification

I. First Order Systems

1. Continuous Data Systems

The examples chosen are from modulating systems alone since, as is explained later, most demodulators of the continuous data type consist of two sampled data systems whose outputs are added.

a. microphone in antenna system

b. White system (vacuum tube in series with antenna)

c. magnetic amplifier in antenna system

d. amplitude-modulated oscillator.

2. Sampled Data Systems

Modulators	Demodulators
Class C Amplifier	a. diode detector
a. plate modulated	b. plate detector
b. grid modulated	c. grid-leak detector
c. cathode modulated	d. bridge detector
d. suppressor grid modulated	e. ring detector
e. screen grid modulated	f. phase sensitive detectors

II. Second Order Systems

Modulators	Demodulators
a. van der Bijl modulator	a. diode detector
b. balanced modulator	b. triode detector
c. bridge modulator	c. bridge detector
d. ring modulator	d. ring detector

Note: The listing of a particular device as a modulator or demodulator corresponds to conventional application. The principle of equivalence is in no way violated though practical considerations may restrict the inverse application. The listing of some devices under both sections I and II illustrates that these possess more than one mode of operation.

2.3 Analysis

Before it is possible to specify the prevailing characteristics of the two types of systems, a more thorough analysis is necessary. A second order system is treated first, followed by a first order system.

2.31 Analysis of Second Order System

Consider the diagram in figure 2.3. The resemblance to figure 2.2(a) is apparent. The symbol $y\left[e_1(t)\right]$ is used to denote the operation of expanding a nonlinear volt-ampere characteristic in a power series. Hence,

$$y\left[e_1(t)\right] = a_o + a_1 e_1(t) + a_2 e_1{}^2(t) + \ldots \quad .$$

The power series expansion results in a current, i(t), so that i(t) and $y\left[e_1(t)\right]$ are equivalent. Thus a_o has the dimensions of current, a_1 has the dimensions of conductance, etc.

$$\text{Let } e_1(t) = f(t) + \cos \omega_c t.$$

In order to show that the output is truly nonlinear, assume:

$$f(t) = \frac{1}{2} \cos pt + \frac{1}{2} \cos qt.$$

Now the output, $e_o(t)$, when i(t) is a unit impulse of current, $u_o(t)$, is defined as h(t). That is,

$$h(t) \triangleq e_o(t)\Big|_{\,i(t) = u_o(t)}.$$

In accordance with the superposition principle, the output for an input i(t), is given in terms of h(t) by:

$$e_o(t) = \int_0^t i(\tau)h(t-\tau)d\tau.$$

Since i(t) and $y\left[e_1(t)\right]$ are equivalent,

$$e_o(t) = \int_0^t y\left[e_1(\tau)\right] h(t-\tau)d\tau. \qquad (2.2)$$

Correspondingly, in the frequency domain:

$$E_o(s) = \mathcal{L}\left\{ y\left[e_1(t)\right]\right\} H(s). \qquad (2.3)$$

Equations (2.2) and (2.3) completely describe the system in terms of $h(t)$, $y\left[e_1(t)\right]$, and their Laplace transforms.

As a specific example, let $h(t) = Ru_o(t)$ so that the filter is a pure resistance, R. Equation (2.2) becomes:

$$e_o(t) = R\int_0^t y\left[e_1(\tau)\right] u_o(t-\tau)d\tau .$$

Because of the impulse present, the integral is easily evaluated as,

$$e_o(t) = Ry\left[e_1(t)\right].$$

Substituting for $e_1(t)$ in the series:

$$e_o(t) = R\Big[a_o + a_1(\tfrac{1}{2}\cos pt + \tfrac{1}{2}\cos qt + \cos \omega_c t)$$
$$+ a_2(\tfrac{1}{2}\cos pt + \tfrac{1}{2}\cos qt + \cos \omega_c t)^2 + \ldots\Big]. \qquad (2.4)$$

Expanding equation (2.4) and collecting terms:

$$
\begin{aligned}
e_o(t) = R\Bigg\{ & a_o + \frac{3a_2}{4} + a_1\frac{\cos pt + \cos qt}{2} + a_2\frac{\cos 2\omega_c t}{2} \\
& + \Bigg[a_1 \cos \omega_c t + a_2\frac{\cos(\omega_c + p)t + \cos(\omega_c - p)t}{2} \\
& + a_2\frac{\cos(\omega_c + q)t + \cos(\omega_c - q)t}{2}\Bigg] + a_2\frac{\cos 2pt + \cos 2qt}{8} \\
& + a_2\frac{\cos(p+q)t + \cos(p-q)t}{4} + \dots \Bigg\} \qquad (2.5)
\end{aligned}
$$

It is now apparent that the output, $e_o(t)$, does not satisfy the linearity requirements of the superposition theorem, since a simple addition of the outputs resulting with cos pt and cos qt considered separately would not yield any intermodulation products such as those appearing in the last term of equation (2.5). In fact, the only desirable terms in this equation are those shown in square brackets, because they result from pure multiplication of $1 + f(t)$ and $\cos \omega_c t$. The other terms represent distortion of one form or another.

The entire spectrum of equation (2.5) is shown in figure 2.4. The frequency spread between the dotted lines at $\frac{\omega_c}{2}$ and $\frac{3\omega_c}{2}$ represents the maximum permissible bandwidth of the modulator. This limitation is imposed for reasons now to be given. It is evident that the frequencies p and $\omega_c - p$ are both present in the system. If $p > \frac{\omega_c}{2}$, then $\omega_c - p < p$. In order to pass the complete modulated wave, the filter must be designed to include $\omega_c - p$, and hence would include p in the pass band. But p, itself, is an extraneous frequency; to eliminate it from the system, the lower frequency limit of the modulated carrier is fixed at $\frac{\omega_c}{2}$. Correspondingly, the upper frequency limit is set at $\frac{3\omega_c}{2}$. Had third order or cubed terms in the series expansion been considered, the frequency $2\omega_c - p$ would have appeared. If $p > \frac{\omega_c}{2}$, then $\omega_c + p > 2\omega_c - p$. The latter frequency, being extraneous, is eliminated only by limiting the upper frequency of the modulated-carrier to $\frac{3\omega_c}{2}$.

A filter can presumably be designed to operate within the limits $\frac{\omega_c}{2}$ and $\frac{3\omega_c}{2}$, but notice that the intermodulation term involving $p + q$, and the second harmonic terms involving $2p$ and $2q$ may easily fall within this range. The intermodulation terms are half as large in magnitude as the desired sum and difference frequency terms, but the filter cannot discriminate against them. Apparently, no limitation on the bandwidth or filter response of a second order system can ensure distortionless operation.

2.32 Analysis of First Order System

Figure 2.5 illustrates a first order modulator. The sensitivity function, $S(t)$, may be regulated by either the carrier or the intelligence. If $S(t)$ is controlled by the carrier, then $e_1(t)$ represents the intelligence, and vice versa. The sensitivity function is conveniently expressed in the time domain and is equivalent to the total input admittance presented at the terminals 1-1'. The effect of the filter must be included. Defining $i(t)$ and $h(t)$ as before:

$$i(t) = e_1(t)S(t) \text{ and}$$

$$e_o(t) = \int_0^t i(\tau)h(t-\tau)d\tau. \tag{2.6}$$

$$\therefore e_o(t) = \int_0^t e_1(\tau)S(\tau)h(t-\tau)\,d\tau. \tag{2.7}$$

In the frequency domain:

$$E_o(s) \quad = \quad E_i(s) \circledast S(s) H(s), \qquad (2.8)$$

where the symbol $\circledast$ is used to denote complex convolution. Equations (2.7) and (2.8) both give the complete solution.

As an example, consider the simple keyed modulator in figure 2.6. The key is assumed to be closed during the positive half of the carrier cycle and open during the negative half, yielding a sensitivity-function like that shown in figure 2.7. The impulse response of the resistive filter is:

$$h(t) \quad = \quad Ru_o(t).$$

Let $e_i(t) \quad = 1 + \frac{1}{2}\cos pt + \frac{1}{2}\cos qt.$

From equation (2.7):

$$\begin{aligned} e_o(t) &= \int_0^t e_i(\tau) S(\tau) Ru_o(t - \tau) d\tau \\ &= Re_i(t) S(t). \end{aligned}$$

Expanding S(t) in a Fourier Series,

$$S(t) \quad = \frac{1}{R}\left[\frac{1}{2} + \frac{2}{\pi}(\sin \omega_c t + \frac{1}{3} \sin 3\omega_c t + \ldots)\right]$$

$$\therefore e_o(t) = (1 + \frac{1}{2} \cos pt + \frac{1}{2} \cos qt)\left[\frac{1}{2} + \frac{2}{\pi}(\sin \omega_c t + \frac{1}{3} \sin 3\omega_c t + \ldots)\right]$$

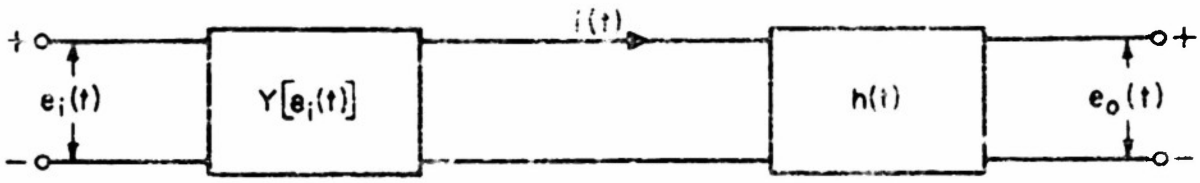


FIGURE 2.3 SECOND ORDER MODULATOR

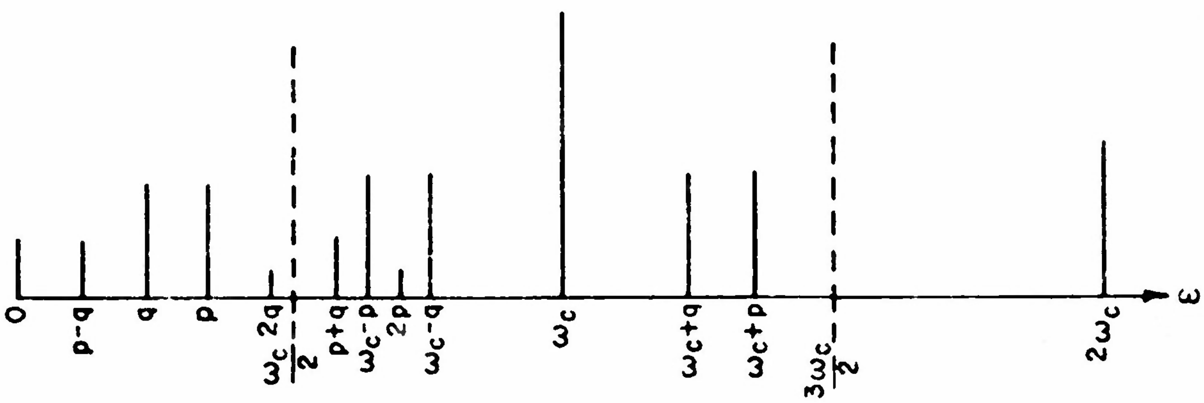


FIGURE 2.4 FREQUENCY SPECTRUM OF SECOND ORDER MODULATOR

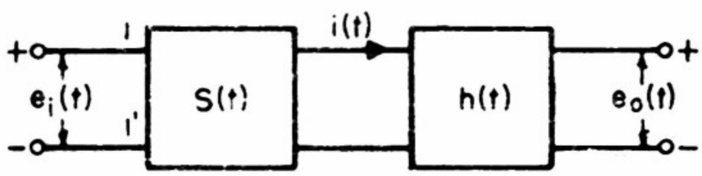


FIGURE 2.5 FIRST ORDER MODULATOR

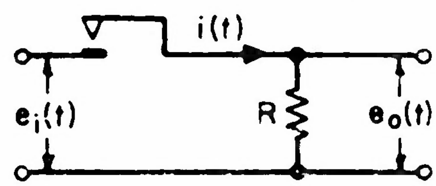


FIGURE 2.6 FIRST ORDER MODULATOR USING KEYED PARAMETER

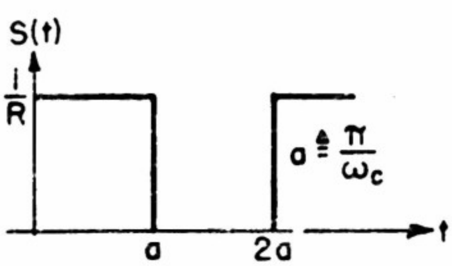


FIGURE 2.7 SENSITIVITY FUNCTION OF FIRST ORDER, KEYED MODULATOR

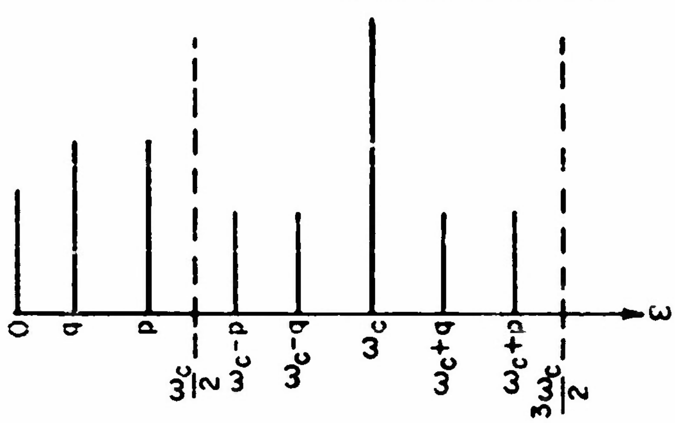


FIGURE 2.8 FREQUENCY SPECTRUM OF FIRST ORDER MODULATOR

Multiplying out:

$$e_o(t) = \frac{1}{2} + \frac{1}{4}\cos pt + \frac{1}{4}\cos qt + \frac{2}{\pi}\sin \omega_c t + \frac{1}{2\pi}\Big[\sin(\omega_c + p)t$$

$$+ \sin(\omega_c - p)t + \sin(\omega_c + q)t + \sin(\omega_c - q)t + \ldots\Big] . \quad (2.9)$$

The spectrum of equation (2.9) is shown in figure 2.8 The bandwidth is still fixed between $\frac{\omega_c}{2}$ and $\frac{3\omega_c}{2}$ for reasons comparable to those given for the second order system, but note that harmonic distortion and interdemodulation distortion are not present. The modulator is truly linear.

2.4 Optimum Systems

2.41 Application

The choice of modulation-demodulation system is governed to a considerable extent by the application. Ordinary or carrier-present systems are employed where distant communication with a simply tuned receiver is desired. The zero-signal level of the carrier-present wave corresponds to a d-c level in the output of the demodulator. Therefore, the null characteristics of the system are quite sensitive to fluctuations of the carrier amplitude. In servomechanisms employing a-c data transmission, the carrier-present system is not a good choice because of the poor null characteristics. Carrier-suppressed systems are necessary. Additional complexity is encountered in servo applications because of the lag introduced by the filter. In general, the application determines the choice of system and the filter design.

2.42 General Considerations

Without reference to any particular application, much can be said about optimization of the systems. First of all, a résumé of the salient features of both types of systems is given.

Characteristics of Second Order Systems

(1) Dependence upon a nonlinear volt-ampere characteristic.

(2) Superposition does not hold.

(3) Intermodulation and harmonic distortion are inherently present and cannot be filtered out.

Characteristics of First Order Systems

(1) The existence of a sensitivity-function implies that the system equations involve time-varying coefficients.

(2) Superposition is applicable.

(3) Undesired frequencies generated can readily be filtered out if the intelligence frequency does not exceed $\frac{1}{2}\omega_c$.

Characteristics (2) and (3) under second order systems are certainly undesirable. In addition, the useful output of such a system is nearly always low because the nonlinearity involved often amounts to no more than a perturbation correction.. For these reasons second order processes are ruled out of consideration for high-performance optimum systems.

Since first order units may also contain nonlinear elements, it is clear that intelligence and carrier must be isolated if intermodulation distortion is to avoided (i.e., no voltage proportional to the sum of $f(t)$ and $\cos \omega_c t$ should be present anywhere in the frequency translating section of the modulator or demodulator). This criterion is usually satisfied by suppressed-carrier modulators and phase-sensitive detectors which are thus indicated as the most desirable of first order systems.

If carrier-present modulation is used, the carrier itself is not available as a separate input to the demodulator. Therefore, the sensitivity function in a first order demodulator may be functionally related to both carrier and intelligence. When such a functional relationship exists, distortion due to nonlinearities present may be reduced in two ways: "swamping" the nonlinear resistance with a linear one, or keeping the percentage modulation so low that $S(t)$ is nearly independent of $f(t)$. Both methods reduce the useful output--the final design must represent a compromise.

2.43 Sensitivity-Functions

The ideal modulator might seem to be one in which $S(t) = \cos \omega_c t$. With such a sensitivity-function, no filter would be required. However, the difficulties in obtaining a physical realization of this function are overwhelming, at least they are greater than those encountered with a d-c amplifier--a unit the designer may be trying to avoid with a modulator. On the demodulation side, the ideal $S(t)$ would seem to be $\sec \omega_c t$ since, again, no filter would be necessary. But consider what happens when this function is used to demodulate a carrier-suppressed signal which contains noise. The system input becomes $f_n(t) + f(t) \cos \omega_c t$, where $f_n(t)$ represents the noise component of the input. Multiplying by $\sec \omega_c t$,

$$\left[f_n(t) + f(t)\cos \omega_c t\right] \sec \omega_c t = f_n(t) \sec \omega_c t + f(t).$$

The secant function becomes periodically infinite, producing an infinite error in the presence of any noise, no matter how small. Hence this function is ruled out from the start.

A great number of sensitivity-functions could be considered, but from the point of view of physical realization with predominantly bistable elements, those functions which produce a sampled data system are the most feasible. The outputs of two such systems may be so added that a continuous data system is achieved. For example, suppose that the sensitivity-functions of the two systems are periodic square pulses with period T and pulse-width $\frac{T}{2}$. If one sensitivity-function is the negative of the other and 180° out of phase with it, the combined system, where the two outputs are added, will possess a square wave sensitivity-function. As pointed out before, the square wave characterizes a continuous data system.

CHAPTER III

TRANSISTOR MODULATORS AND DEMODULATORS

3.1 The Transistor as a Controlled Sensitivity Device

3.11 Grounded Base Operation

The large-signal equivalent circuit of the grounded base transistor, shown in figure 3.1, is now common to the literature. The loop equations describing the behavior of this network are:

$$V_\varepsilon = (r_\varepsilon + r_b)I_\varepsilon + r_b I_c \tag{3.1}$$

$$V_c = (r_b + \alpha_\varepsilon r_c)I_\varepsilon + (r_b + r_c)I_c \tag{3.2}$$

where: $\alpha_\varepsilon = \alpha + \frac{r_b}{r_c}(1 - \alpha)$

and $\alpha \triangleq -\left(\frac{\partial I_c}{\partial I_\varepsilon}\right)_{V_c}$

From these equations and the equivalent circuit in figure 3.1, it is possible to determine four complete sets of characteristic curves. Of chief interest here are the linearized collector characteristics shown in figure 3.2. Applying the method of break-point analysis, one observes that the emitter diode switches when the current through it is zero. When $I_\varepsilon = 0$, equation (3.2) becomes:

$$V_c = (r_b + r_c)I_c. \tag{3.3}$$

Equation (3.3) is the line OC in figure 3.2.

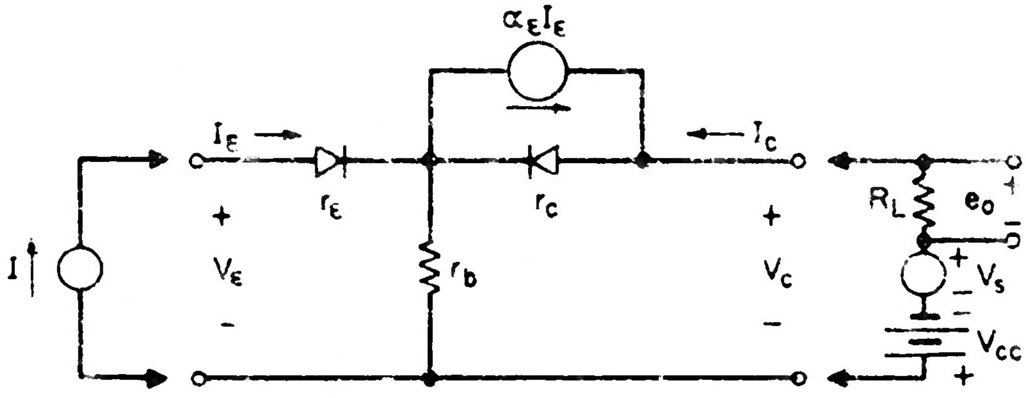


FIGURE 3.1 LARGE-SIGNAL EQUIVALENT CIRCUIT OF GROUNDED BASE TRANSISTOR

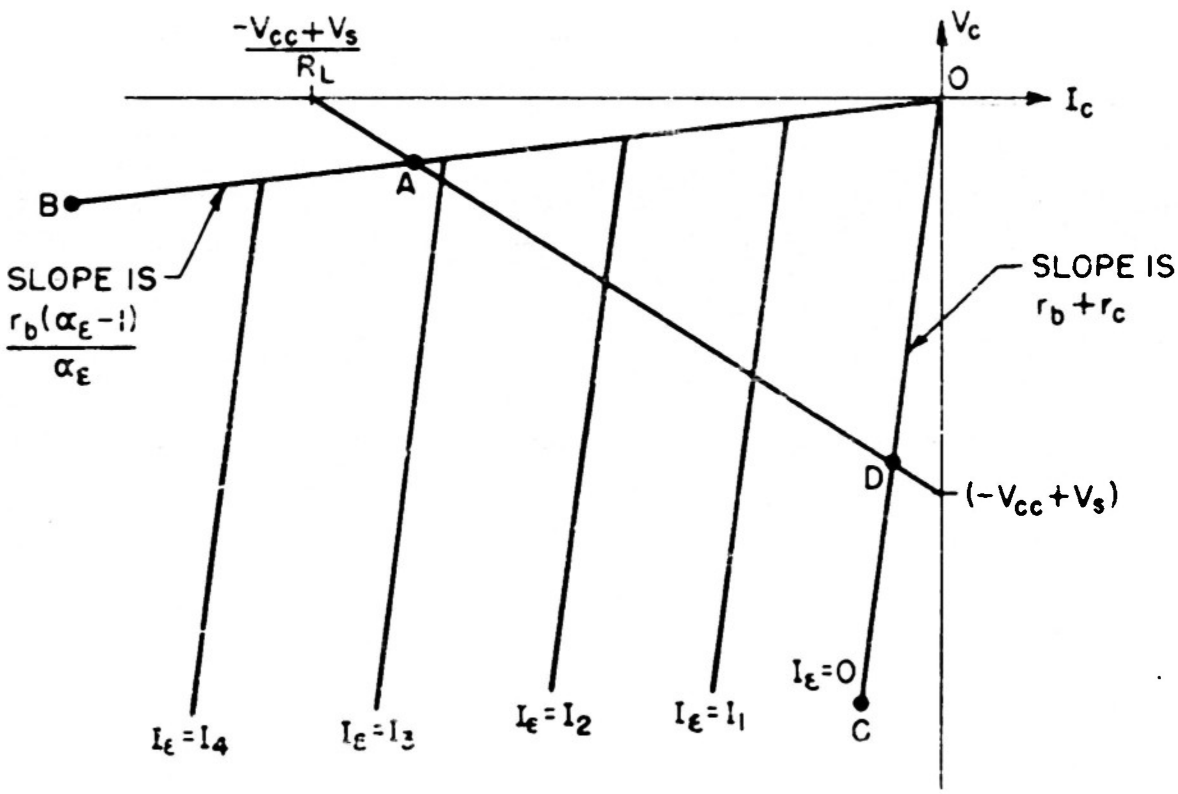


FIGURE 3.2 LINEARIZED COLLECTOR CHARACTERISTICS

The collector diode switches when the current through it is zero. This condition implies that:

$$I_c = \alpha_\varepsilon I_\varepsilon .$$

Accordingly, equation (3.2) yields:

$$V_c = -(r_b + \alpha_\varepsilon r_c)\frac{I_c}{\alpha_\varepsilon} + (r_b + r_c)I_c$$

or:

$$V_c = \frac{I_c r_b(\alpha_\varepsilon - 1)}{\alpha_\varepsilon} . \qquad (3.4)$$

Equation (3.4) is represented by the line OB in the diagram. The area within the lines OB and OC corresponds to the active region of the transistor.

Now consider that the sources shown in figure 3.1 are actually attached to the transistor. The source, I, is meant to represent only a quasi-current source. That is, it specifies the current in the positive direction of I_ε, but cannot force negative current through the emitter diode. For the collector circuit, one can now write the equation:

$$V_c = -I_c R_L - V_{cc} + V_s. \qquad (3.5)$$

Equation (3.5) is shown as the load line AD in figure 3.2. It is seen that a change in source voltage, V_s, shifts the load line parallel to itself.

At any instant, the operating point is specified by the intersection of this line and the corresponding instantaneous emitter current. If the emitter current is large, however, the operating point is determined by the intersection of the load line and the line OB--point A, for example. From equation (3.4) and (3.5), the collector current is then determined as:

$$- I_c R_L - V_{cc} + V_s = I_c r_b \frac{(\alpha_\varepsilon - 1)}{\alpha_\varepsilon}$$

or:

$$I_c = \frac{(- V_{cc} + V_s)\, \alpha_\varepsilon}{\alpha_\varepsilon R_L + r_b(\alpha_\varepsilon - 1)}\,, \tag{3.6}$$

and the output voltage is:

$$e_o = - I_c R_L \,.$$

Equation (3.6) is valid [from equations (3.2), (3.5), and (3.6)] if:

$$I_\varepsilon \geq \frac{V_{cc} - V_s}{\alpha_\varepsilon R_L + r_b(\alpha_\varepsilon - 1)} \cong \frac{V_{cc} - V_s}{\alpha_\varepsilon R_L}\,. \tag{3.7}$$

One concludes that if I_ε is sufficiently large, the maximum excursion of the collector current is linearly related to the source voltage V_s. Such a source may be said to control the sensitivity-function of the transistor. If the current source, I, produces a sinusoidal waveform, the emitter current is rectified to a periodic pulse. The output, e_o, is then a periodic pulse whose peak excursions are controlled by the voltage V_s --- a process of controlled pulse clipping. Though the sensitivity-function is continuous, the presence of the pulses is indicative of a sampled data system.

3.12 Grounded Emitter Operation

The large-signal equivalent circuit of the grounded emitter transistor is shown in figure 3.3. The pertinent loop equations are:

$$V_b = (r_\varepsilon + r_b)I_b + r_\varepsilon I_c \tag{3.8}$$

$$V_{c\varepsilon} = (r_\varepsilon - \alpha_\varepsilon r_c)I_b + \left[r_\varepsilon + (1 - \alpha_\varepsilon)r_c\right] I_c \tag{3.9}$$

The linearized collector-emitter characteristics appropriate to equation (3.9) appear in figure 3.4. Noting that the emitter diode switches when:

$$I_\varepsilon = -(I_b + I_c) = 0,$$

or when: $I_b = -I_c$,

equation (3.9) becomes:

$$V_{c\varepsilon} = r_c I_c. \tag{3.10}$$

The slope of line OC is thus r_c. If $I_b = 0$ in equation (3.9), the slope of all constant base current characteristics is found as $r_\varepsilon + (1 - \alpha_\varepsilon)r_c$.

The collector diode switches when:

$$I_c = -\alpha_\varepsilon I_\varepsilon ,$$

or: $$I_b = \frac{(1 - \alpha_\varepsilon)I_c}{\alpha_\varepsilon}. \tag{3.11}$$

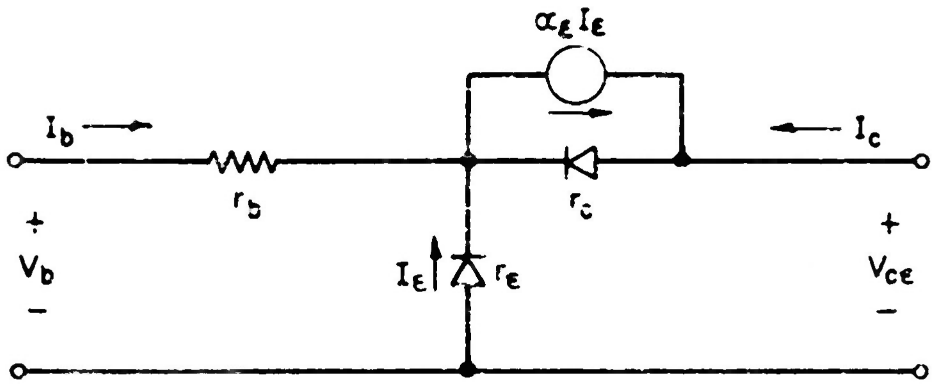


FIGURE 3.3 LARGE-SIGNAL EQUIVALENT CIRCUIT OF GROUNDED EMITTER TRANSISTOR

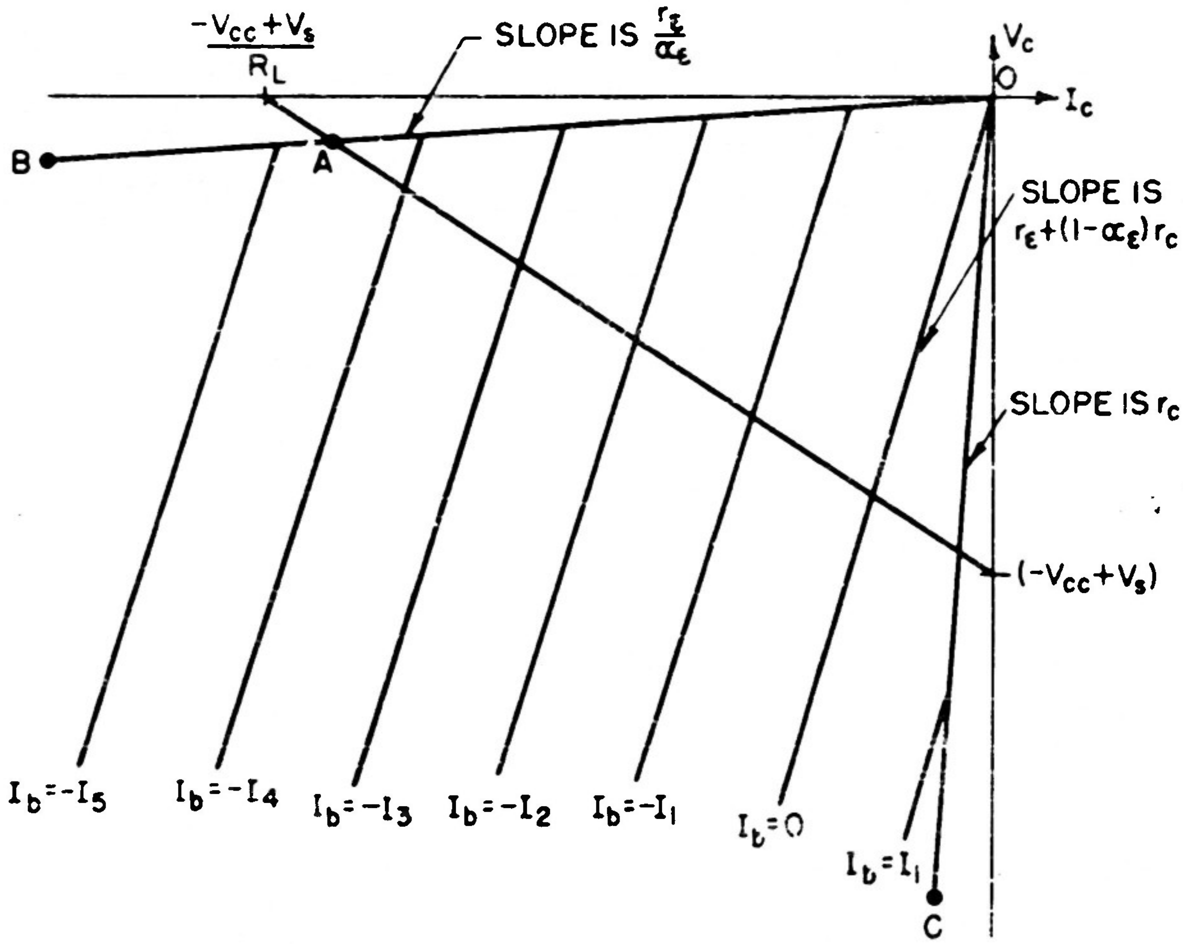


FIGURE 3.4 LINEARIZED COLLECTOR-EMITTER CHARACTERISTICS

Substituting for I_b in equation (3.9):

$$V_{c\varepsilon} = \frac{r_\varepsilon I_c}{\alpha_\varepsilon}, \tag{3.12}$$

indicating that the slope of line OB is $\frac{r_\varepsilon}{\alpha_\varepsilon}$.

Applying the same sources used in figure 3.1 to the circuit of figure 3.3, one may write:

$$V_{c\varepsilon} = -I_c R_L - V_{cc} + V_s. \tag{3.13}$$

Simultaneous solution of (3.11) and (3.12) yields a collector current at point A of:

$$I_c = \frac{(-V_{cc} + V_s)\alpha_\varepsilon}{\alpha_\varepsilon R_L + r_\varepsilon}. \tag{3.14}$$

Equation (3.14) is valid [from equations (3.9), (3.12), and (3.14)] if:

$$I_b \leq \frac{(1 - \alpha_\varepsilon)(-V_{cc} + V_s)}{\alpha_\varepsilon R_L + r_\varepsilon} \cong \frac{(1 - \alpha_\varepsilon)(-V_{cc} + V_s)}{\alpha_\varepsilon R_L} \tag{3.15}$$

The sense of the inequality is determined by the condition that I_b is negative.

The voltage source, V_s, is seen to control the sensitivity-function of the transistor in a fashion exactly analogous to that previously explained--that is, it controls the peak collector current when I_b is sufficiently large.

Equation (3.11) indicates that if $\alpha_E > 1$, the base current must be positive before the collector diode switches--I_c being always negative. But in this region of figure 3.4, the load line does not intersect the line OB. Therefore, point contact transistors (for which α_E is usually > 1) are not really suited to this kind of application. This fact has been verified experimentally.

3.2 Carrier-Present Systems

3.21 Single Transistor Modulators

The design of the circuit in figure 3.5 is based upon the concepts just discussed. In the emitter circuit a current source is simulated with an oscillator and 27KΩ resistor. The input current, since no bias battery is used, is rectified by the emitter diode to provide a periodic pulse as shown in figure 3.7(c). The transient decay evident there is due to the capacitance inherently associated with the emitter diode. At higher frequencies, this capacitance begins to nullify the presence of the diode.

The source of modulation in the collector circuit operates about a constant level of three volts which provides a zero-signal output level and indicates the existence of a carrier-present system. The modulated pulse which results from the action of this source is converted to a sinusoid by a simple parallel-tuned circuit. Sharp tuning and wide bandwidth are not available independently. Both are functions of the damping ratio, ζ, which is defined as $\frac{1}{2R}\sqrt{\frac{L}{C}}$. For practical applications, a damping ratio of 0.2 is found to strike an adequate compromise, giving a bandwidth of about $0.4\omega_c$.

Typical output voltages of the system are shown in the oscillographs of figures 3.6 and 3.7. The waveforms in figure 3.6 were obtained with a Western Electric 1698, type A, point contact transistor; those in figure 3.7 with a Raytheon CK721 junction transistor. There appeared to be very little difference in the frequency response of the two types of transistors, a carrier of 50KC representing very nearly the upper half power frequency of both units. This frequency may seem unusually low for a point contact transistor, but the lossy nature of the tuned circuit parameters used is at least partially responsible for the low frequency cut-off.

Figure 3.8 illustrates the connection of a junction transistor for grounded emitter operation. The oscillographs in figure 3.9 are typical outputs. One interesting feature of this modulator is the waveform resulting from overmodulation. When V_s exceeds V_{cc}, the collector current finds an easier path to ground through the base connection than through the back direction of the emitter diode. Consequently, the base current remains negative, and V_c and I_c become positive, resulting in the nonsymmetrical, double frequency envelope shown in the sketch of figure 3.10.

In modulators of this type, the envelope of the modulated wave may evidence flattening of the peaks. This effect may be attributed to insufficient emitter or base current. Using the emitter current in illustration, equation (3.7) indicates that the effect may be alleviated in three ways; increasing I_ε, increasing R_L, or decreasing V_{cc}. Allowing 100 percent modulation (where $V_s = -V_{cc}$ is the extreme value of V_c) the peak value of emitter current, $\hat{I}_\varepsilon$, must always fulfill the inequality:

$$\hat{I}_\varepsilon \geq \frac{2V_{cc}}{\alpha_\varepsilon R_L} .$$

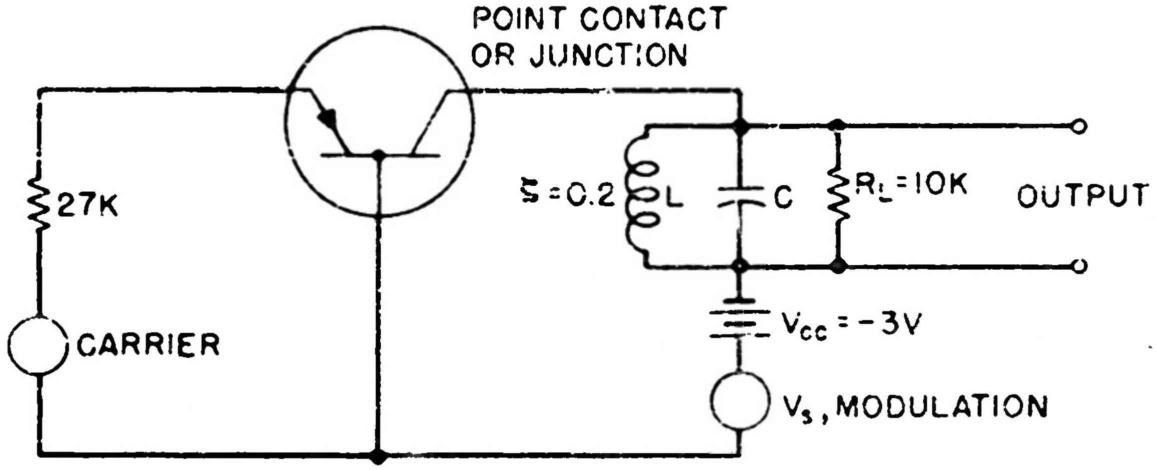


FIGURE 3.5 GROUNDED BASE MODULATOR

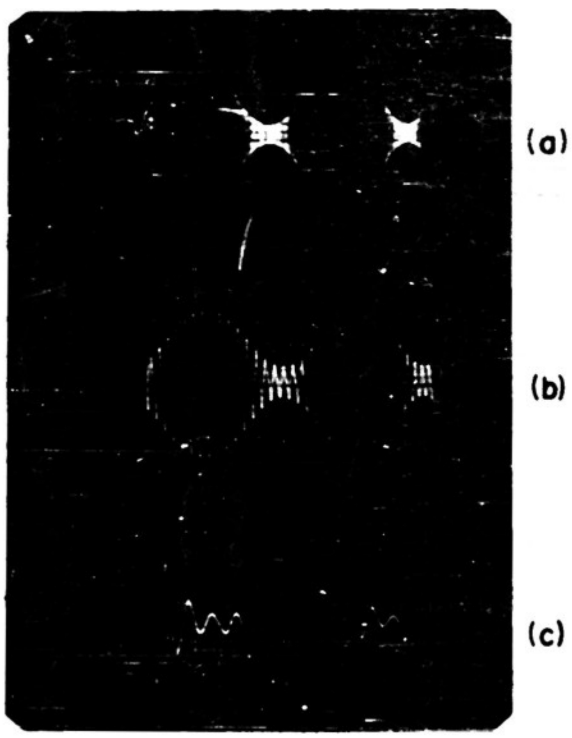


FIGURE 3.6 TYPICAL
POINT-CONTACT WAVEFORMS
16KC CARRIER
(a) 100
(b) 910
(c) 2100
MODULATION FREQUENCIES

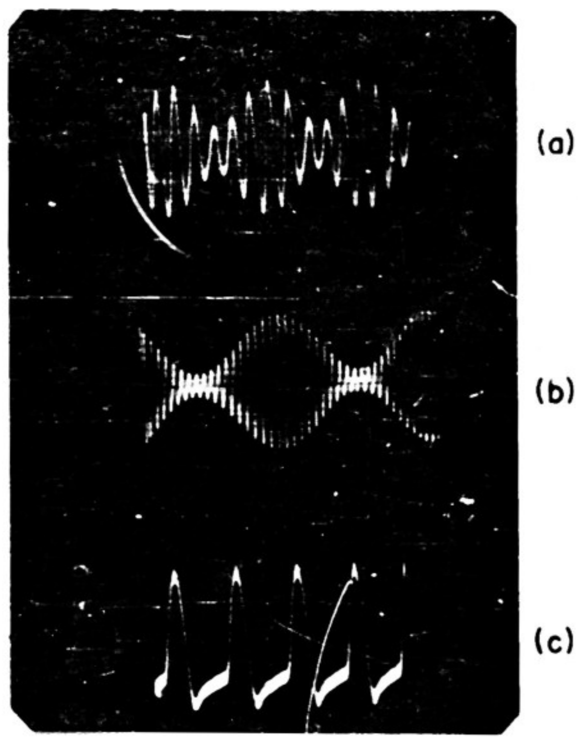


FIGURE 3.7 TYPICAL
JUNCTION WAVEFORMS
50KC CARRIER
(a) 10KC
(b) 2 KC
MODULATION FREQUENCIES
(c) EMITTER CURRENT

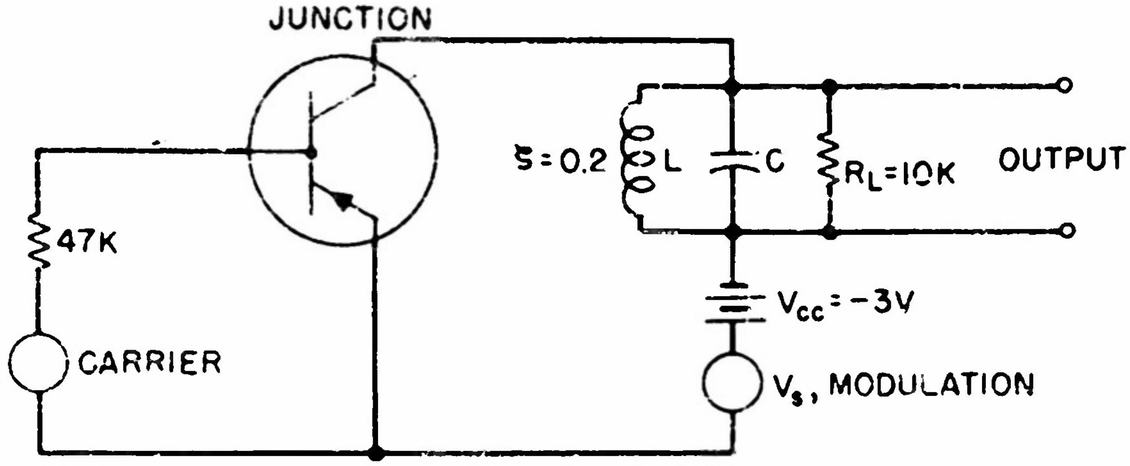


FIGURE 3.8 GROUNDED EMITTER MODULATOR

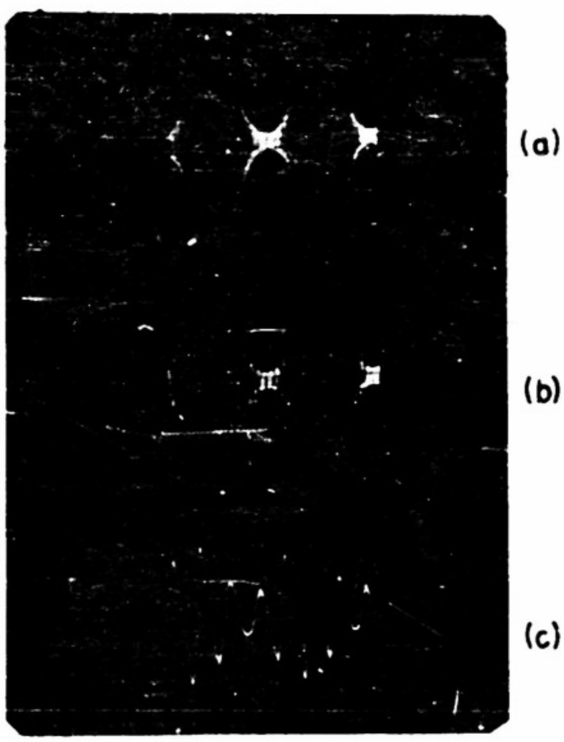


FIGURE 3.9 TYPICAL OUTPUT WAVEFORMS

16KC CARRIER

(a) 100
(b) 920
(c) 4000
MODULATION FREQUENCIES

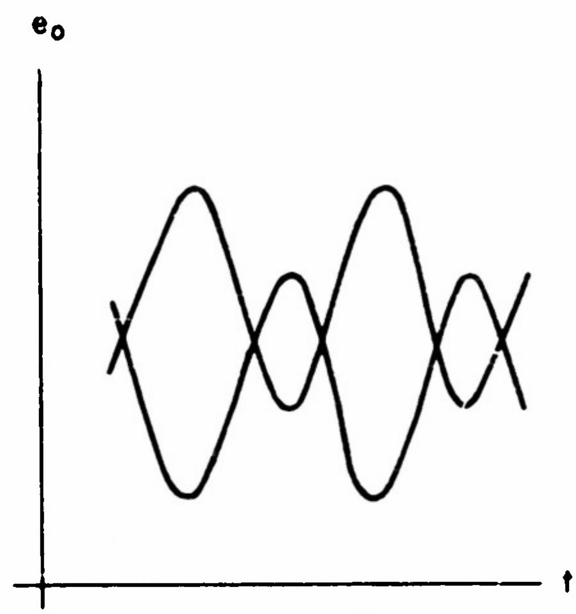


FIGURE 3.10 SKETCH OF ENVELOPE OF OVERMODULATED WAVE

The advantage of the grounded emitter circuit is apparent from the fact that the above value of input current is reduced by a factor of $(1 - \alpha_E)$.

3.22 Distortion

Due to the assumptions made in using linearized characteristics, it might be thought that square law modulation could assume large proportions in the operations described above. However, actual measurements on a General Radio Wave Analyzer have shown that harmonic and intermodulation distortion are less than 2 percent for point contact transistors, and less than 0.8 percent for junction transistors. These measurements were made on 100 percent modulated carriers, indicating much less distortion than would be expected from a second order process. It is also evident that junction transistors excel point contacts in linearity of operation.

3.23 Single Transistor Demodulators

Considering the principle of equivalence, the circuits of figures 3.5 and 3.8 may seem logically capable of demodulation. Unfortunately, they are not, unless the carrier is available separately. In carrier-present systems the carrier is an integral part of the input wave and hence rarely available independently. However, these circuits do possess the property of rectifying the input voltage. By "swamping" the nonlinear resistance variation of the emitter diode with a series resistance, both circuits may be modified for use as demodulators. Appropriate connections are shown in figures 3.11 and 3.12 together with suitable filters. The advantage of these circuits over a simple rectifier circuit is one of gain, the grounded emitter connection being the most useful in this respect. Depending upon the transistor and the connection, tests with a 30KC carrier have indicated that voltage gains between 5 and 20 are possible. At higher carrier frequencies, less filtering is required because of the frequency limitation of the units. With a carrier of 4 megacycles, for example, demodulation of audio frequencies is excellent using no capacitance at all.

3.24 A Modulator Using Complementary Symmetry

The existence of complementary units among transistors is a unique property of junction types alone. So far, only PNP units have been considered. The circuit of figure 3.13, however, makes use of both a PNP transistor and its complement, an NPN transistor. The two varieties are distinguished by calling the first an n-type and the second a p-type. The difference in the characteristic curves for the two is only a matter of sign. If two such units are perfect complements, the characteristics of the p-type unit may be obtained from those of the n-type unit by prefixing all the parameters involved with a minus sign. The collector characteristics for the p-type transistor, therefore, lie in the first quadrant of the $V_c - I_c$ plane. It is this complementary property which permits construction of a push-pull circuit without a phase inverter on the input. Figure 3.13 shows such a circuit. Notice that an output transformer is missing. The transformer present is necessary for the modulation input.

Either grounded base or grounded emitter (illustrated) connections may be used, but the advantage of greater gain in the latter makes it preferable. The operation of the circuit is essentially the same as that of the single transistor modulators previously discussed. As expected, the output is no different in form from those shown in figures 3.6, 3.7, and 3.9. Hence no pictures were taken. The advantages obtained, however, are twofold. First, the output voltage is double that of a

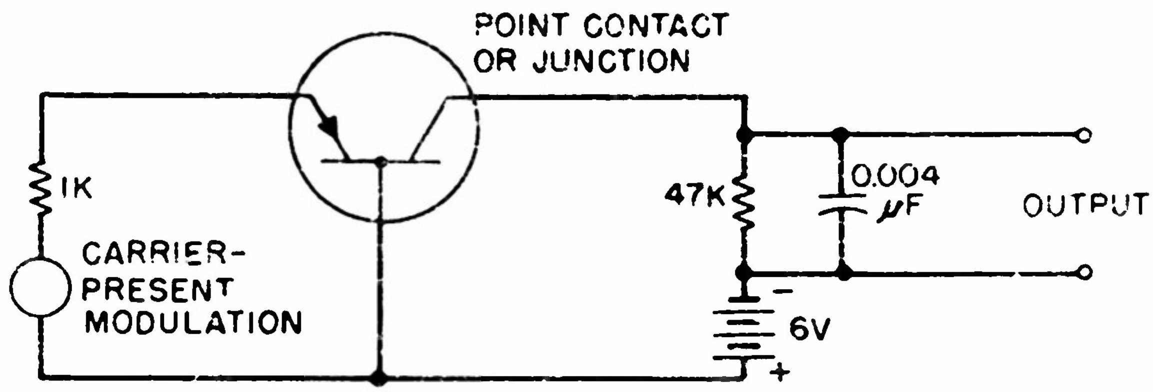


FIGURE 3.11 GROUNDED BASE DEMODULATOR

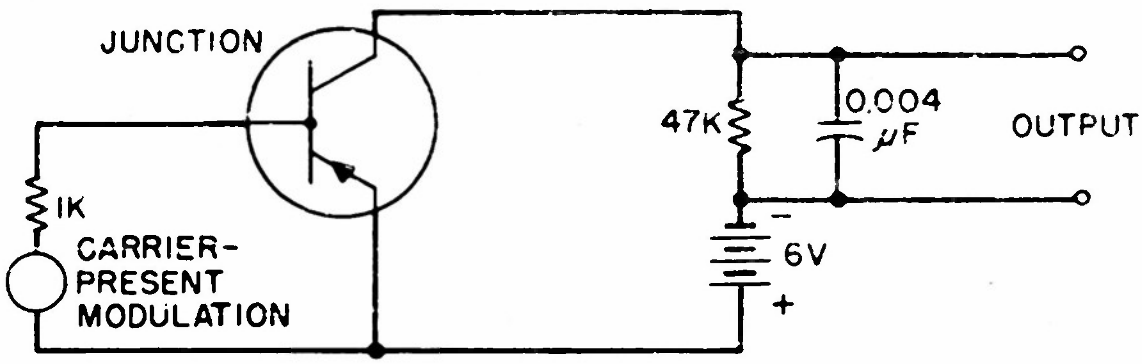


FIGURE 3.12 GROUNDED EMITTER DEMODULATOR

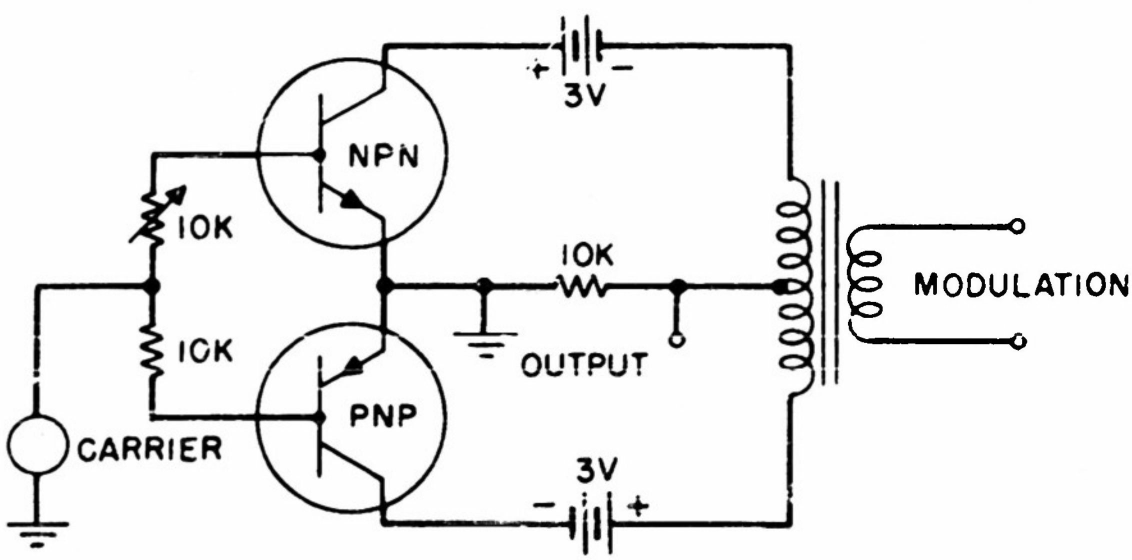


FIGURE 3.13 MODULATOR USING COMPLEMENTARY SYMMETRY

single transistor circuit because control is exerted upon a full wave pulse rather than a half-wave one. In other words, we now have a continuous data system. Second, the filtering problem is simplified. The fact that the input transformer used in the experimental set-up exhibited resonance in the range 30-50KC removed the necessity of other filtering in this band.

There are several disadvantages to consider. Unbalance in the input, which must be compensated by one of the input resistors, precludes free interchange of transistors without rebalancing. The transformer secondary must pass a small amount of quiescent current without core saturation. Lastly, the frequency response is limited by an upper half-power point of about 50KC, at least with the transistors used. (viz. Raytheon CK721 and Federated Semiconductor RD2520). Although the circuit cannot be used as a carrier-present demodulator without a separate carrier input, it does represent, in the writer's opinion, a very fine carrier-present modulator in the optimum category of classification.

3.3 <u>Carrier-Suppressed Systems</u>

3.31 <u>Balanced Circuits</u>

Several attempts were made to design circuits to give a balanced output; that is, to balance the carrier from the output completely. In every case, two transistors are required. Furthermore, accurately balanced circuits call for closely identical units. Unfortunately, production standards for transistors have not reached the point where one unit is a reliable replica of the next. Parameter values may vary considerably between units. All the circuits constructed to provide a balanced output clearly exhibited unbalance, and no amount of compensation, either on the inputs or on the outputs, completely eliminated it. Hence, none of the systems designed gave any promise of practical usefulness. Attempts to balance out the carrier were dropped in favor of other methods.

3.32 The Complementary Symmetry Circuit

The circuit of figure 3.13, which was used to obtain carrier-present modulation, can also be employed as a carrier-suppressed modulator if carrier and modulation are interchanged. Again either grounded base or grounded emitter connections may be used, but the latter is preferred as before. For clarity, the circuit is repeated in figure 3.14 showing the proper connection of sources. Control of the sensitivity-function remains with the modulating signal. To understand how this comes about, consider the set of collector characteristics shown in figure 3.15. It is necessary to consider only one of the transistors at a time, because the modulation specifies which unit may conduct. If the modulation voltage is positive, the p-type transistor permits collector current to flow. If the modulation voltage is negative, the n-type transistor permits collector current to flow. Assume that the modulation voltage is positive so that the n-type transistor maintains control of the output. The characteristic curves of figure 3.15, and the associated load line, are then applicable. The carrier voltage at each half of the transformer secondary must be maintained at a constant peak-to-peak value of $2V_{cc}$ volts. This restriction is imposed to insure that the output from each transistor reaches ground level. The carrier forces the load line to move parallel to itself with a peak to peak excursion on the voltage axis of $2V_{cc}$ volts. Now if the base current has some value, $-I_2$, the collector current will periodically saturate at some value, $-I_{c2}$ (assuming infinite slope for the constant base current curves). It follows that the maximum excursion of the collector current is linearly related to the base current, in so far that the constant base current curves are equally spaced on the collector plane. The base current thus controls the sensitivity-function of the transistor. The operation of the p-type transistor, when the base current is positive, is precisely that of the n-type if the signs of all quantities are reversed.

Typical output waveforms of the half-wave modulator are shown in the oscillographs of figure 3.16(a) and (b). [A full-wave output could be obtained with an appropriate filter (e.g., a parallel-tuned circuit).] It is apparent from these pictures that conduction of collector current takes place during a considerable portion of the carrier cycle. The "off" time, or the time during which the collector current is unsaturated, is the time taken for the carrier to move the load line from point A (in figure 3.15) to point O and back again. The remaining portion of the carrier cycle is called "on" time. The complete distribution of "on" and "off" times is shown in figure 3.17.

The modulator will perform equally well as a phase-sensitive detector, if the input used for the modulation is used as the input for the carrier-suppressed signal, and a low pass filter is placed at the output. The way in which the circuit operates is unchanged, the base current simply varies at a faster rate. A typical output from the demodulator is shown in figure 3.16(c). The long sampling time, corresponding to the long "on" time, still obtains. In the modulator, this property introduces no other difficulty than that involved in filtering the output to a sine wave (if such is desired). In a demodulator, however, this long sampling time may be a serious drawback if quadrature voltage is a problem. Quadrature voltage, which is usually introduced by phase lag in the modulator, is a maximum when the carrier voltage is zero. If a demodulator conducts during this portion of the carrier cycle, as this one does, the output is

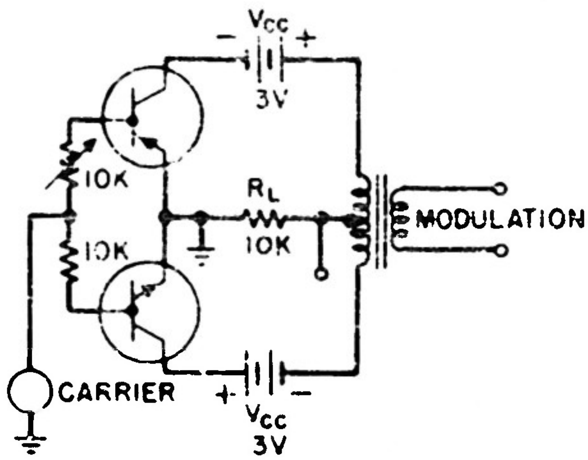


FIGURE 3.14 A CARRIER-SUPPRESSED MODULATOR (AND DEMODULATOR) USING COMPLEMENTARY SYMMETRY

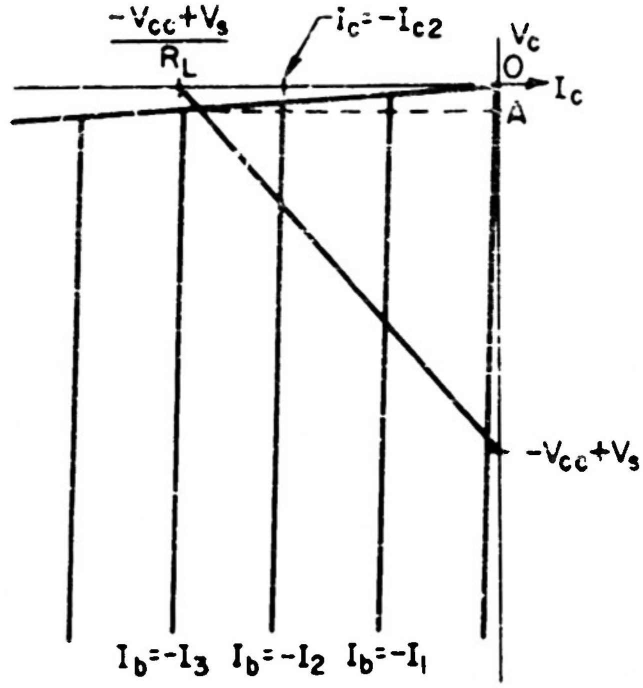


FIGURE 3.15 LINEARIZED SET OF CONSTANT BASE CURRENT CHARACTERISTICS

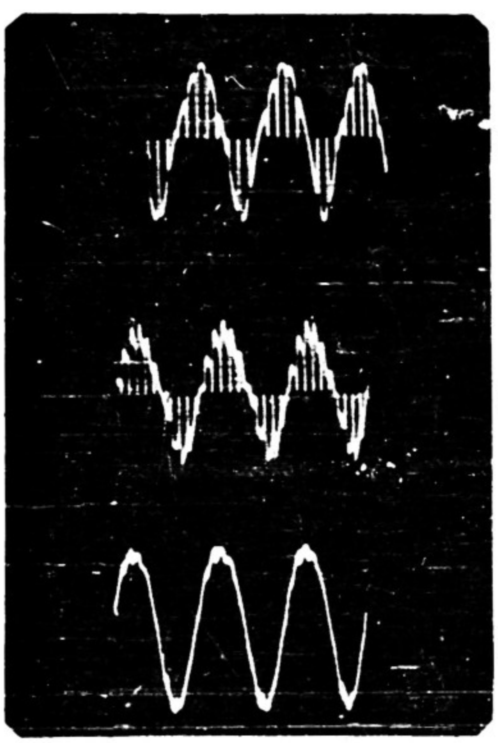


FIGURE 3.16 TYPICAL WAVEFORMS FROM THE SYSTEM USED AS A MODULATOR AND AS A DEMODULATOR

(a) 400 CYCLES MODULATED WITH 50 CYCLES
(b) 5000 CYCLES MODULATED WITH 500 CYCLES
(c) DEMODULATION OF 500 CYCLES ON 5KC

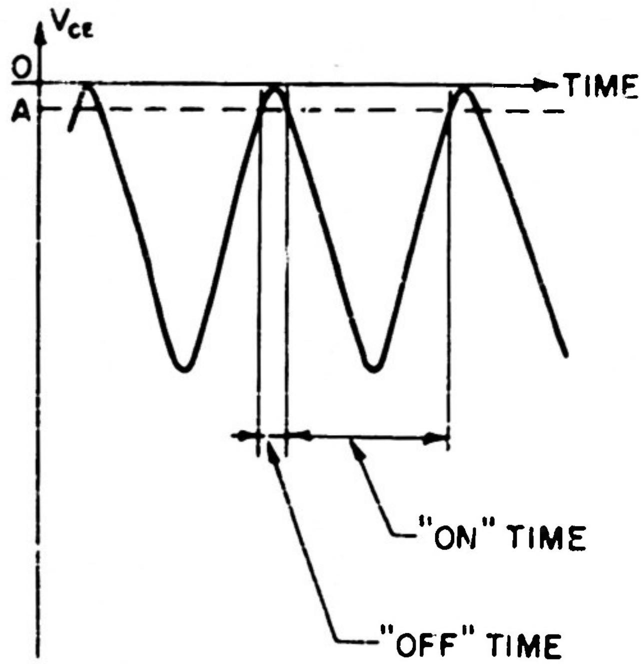


FIGURE 3.17 DISTRIBUTION OF "ON" AND "OFF" TIMES

adversely affected by the quadrature voltage present. If quadrature voltage is not present, long sampling time is an advantage, since the filtering problem is eased. That is, there is more information available per unit time, and subsequent demands on the filter are reduced. The demodulator then, is primarily useful where quadrature voltage is not important. The gain obtained by using the base as the input for the carrier-suppressed signal may also prove advantageous.

3.33 The Schreiner Dual Circuit

The Schreiner circuit, shown in simplified form in figure 3.18, has found fairly wide application as a phase-sensitive detector. The essential portion of the circuit consists of two triodes connected plate to cathode and cathode to plate in parallel. The dual configuration in transistor circuitry is simply two transistors connected collector to base and base to collector in series. The dual concept employed considers the following terminals as analogous: collector and plate, emitter and grid, base and cathode. The complete dual circuit appears in figure 3.19.

The operation of the circuit as a modulator is best understood by considering the switching characteristics of the transistor combination involved. It is necessary to keep in mind the large-signal equivalent circuit of the grounded base transistor shown in figure 3.1. When the carrier voltage is negative, no emitter current can flow in either transistor, Consequently, if the modulation signal is positive, it is faced by the forward resistance of the first collector diode, the two base resistances in series, and the back resistance of the second collector diode. The latter resistance is so large that no appreciable current can flow to the load. If the modulation signal is negative, the roles of the two transistors are reversed. In either case, no appreciable output is obtained. When the carrier voltage becomes positive, emitter current can flow, and either transistor permits negative collector current to flow. Regardless of the sign of the modulation voltage, it is faced now only with the sum of one base resistance, the forward resistance of one collector diode, and the small amount of resistance indicated as $r_b\frac{(\alpha_\varepsilon - 1)}{\alpha_\varepsilon}$ in figure 3.2. If the emitter current does not limit the amount of modulation current flowing, the transistor combination behaves very much like a switch being turned off and on at the carrier frequency. The carrier, then, controls the sensitivity-function of the system.

It is not possible to obtain any sort of gain from the circuit, but the peak output voltage may be expected to very nearly approximate the peak input voltage. Some typical output waveforms for the half-wave modulator are shown in figure 3.20. The last two photographs there indicate what happens as the carrier frequency is increased. That is, the back resistance of the collector diodes begins to decrease with increasing frequency. This apparent decrease in resistance may be attributed to the capacitance inherently associated with the collector diode. Experimentally, the combination of transistors performed well to only 10KC.

As expected from the principle of equivalence, the circuit also operates as a demodulator. It is only necessary to use the modulation input as the input for the carrier-suppressed wave. No change in the reference input, or carrier input, is required. Hence the carrier maintains control of the sensitivity-function. Typical waveforms from the output of the demodulator are shown in figure 3.21. Though not indicated in the diagram, a 0.04 μf capacitor was used as a filter across the load.

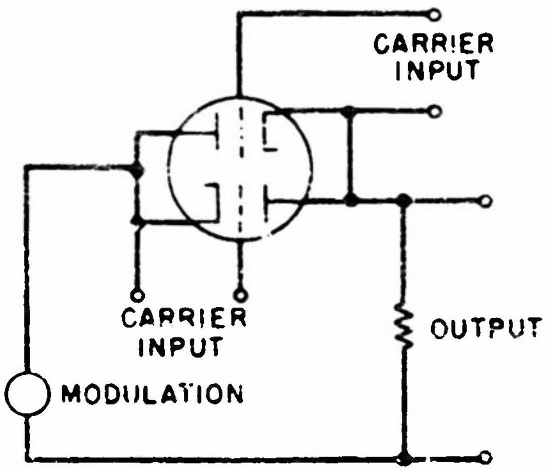


FIGURE 3.18 SIMPLIFIED SCHREINER CIRCUIT

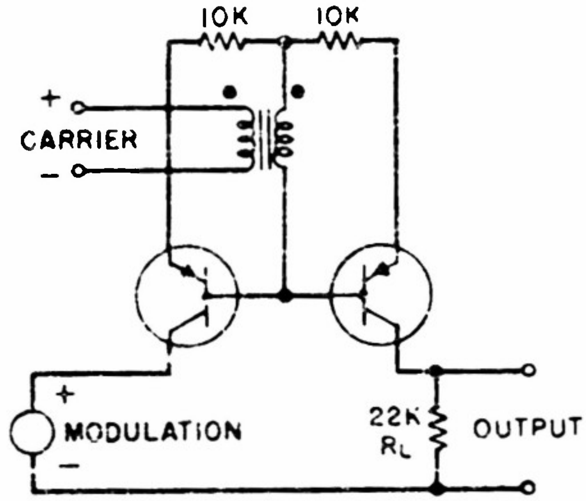


FIGURE 3.19 SCHREINER DUAL CIRCUIT

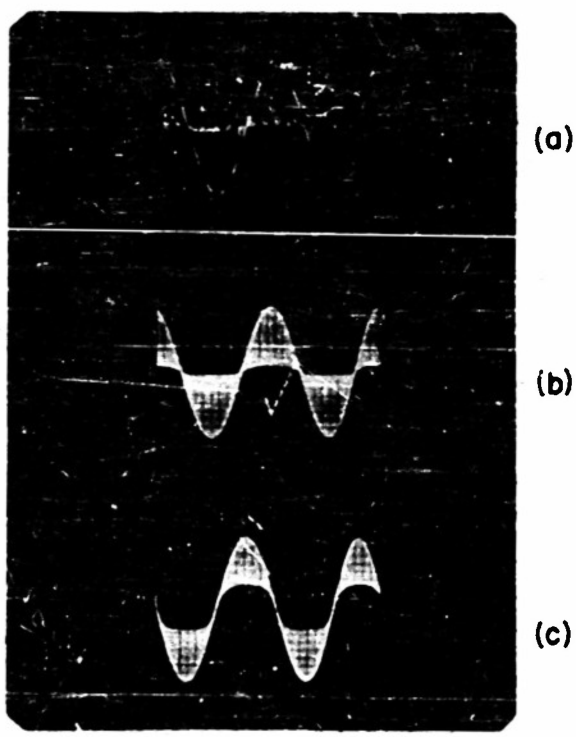


FIGURE 3.20 TYPICAL WAVEFORMS FROM HALF-WAVE MODULATOR

(a) 500 CYCLES ON 5KC
(b) 500 CYCLES ON 10KC
(c) 500 CYCLES ON 20KC

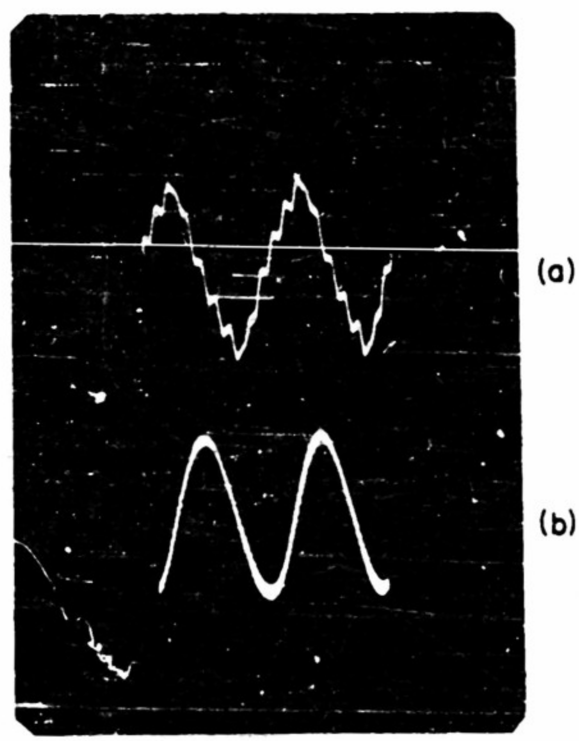


FIGURE 3.21 TYPICAL OUTPUTS FROM HALF-WAVE DEMODULATOR

(a) 500 CYCLES DEMODULATED FROM 5KC CARRIER
(b) 100 CYCLES DEMODULATED FROM 5KC CARRIER

If quadrature voltage is a problem, the demodulator is especially useful, as is the Schreiner demodulator, if the system is switched by an impulse at the peak of the carrier voltage (i.e., where quadrature voltage is zero). An impulse of emitter current may be approximated with a long-time-constant R-C circuit. The capacitance is placed in series with the emitter, and the resistance is bridged from emitter to base. The capacitor maintains its voltage by drawing a spike of current at the carrier peak, thus actuating the emitter.

The full-wave complement of the preceding circuit is shown in figure 3.22. This system was evolved simply by adding the outputs of two half-wave systems with due consideration for the sign of the voltages produced. The result, of course, is a continuous data system. Figure 3.23 illustrates the type of modulation to be expected and, incidentally, the excellent results obtained at the lower carrier frequencies. At higher frequencies (about 10KC), dissymmetry between successive positive and negative peaks of voltage becomes a problem. The cause may be ascribed to differences in parameter values among the transistors and, particularly, to differences in the capacitance associated with the various collector diodes.

No filter is shown in any of the circuits and none was used in any except that of the demodulator. However, simple band-pass filters centered at the carrier frequency should suffice for both modulators. The advantage of the full-wave modulator over the half-wave one rests in the fact that filtering is simpler, since more information is provided per unit of time. The full-wave circuit may be used as a demodulator, but the additional information obtained is probably not worth the extra equipment required.

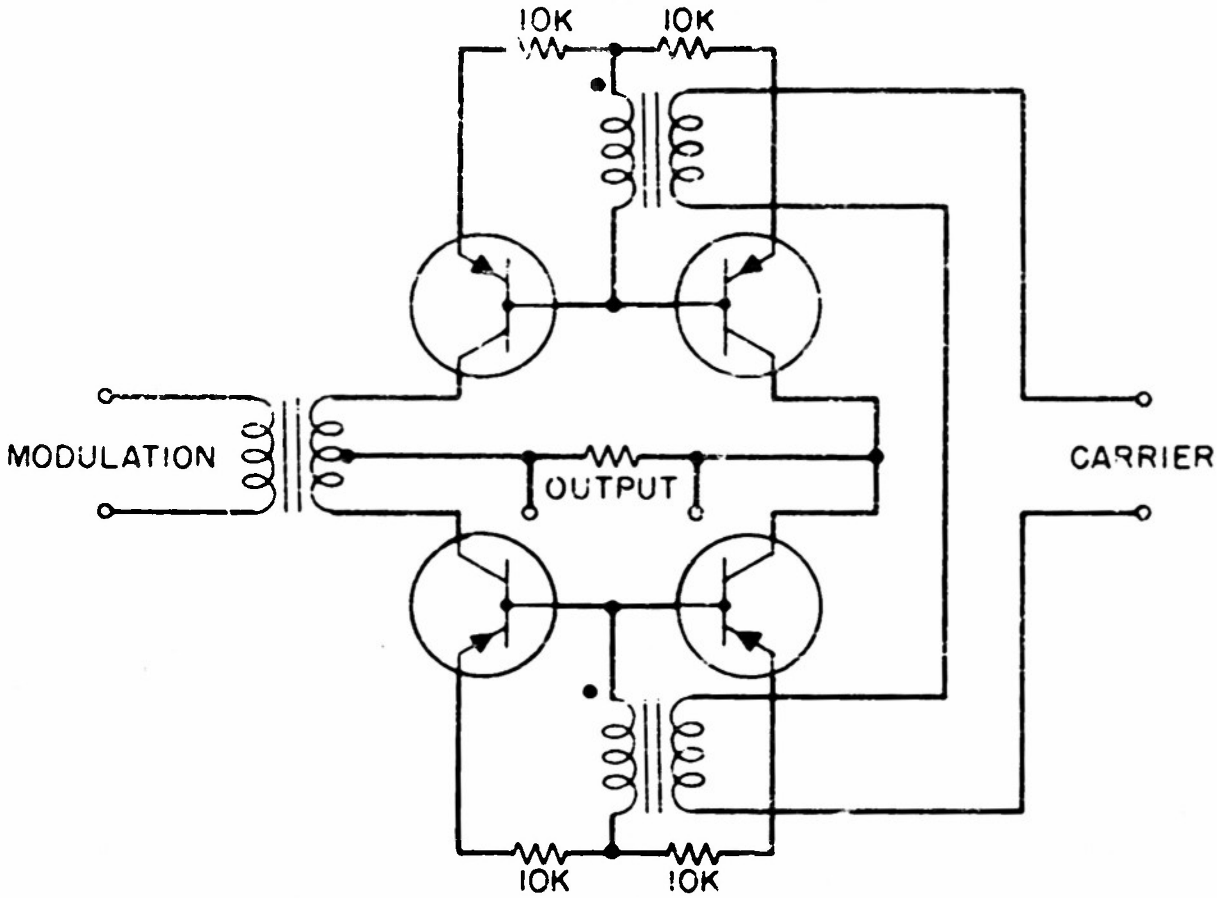
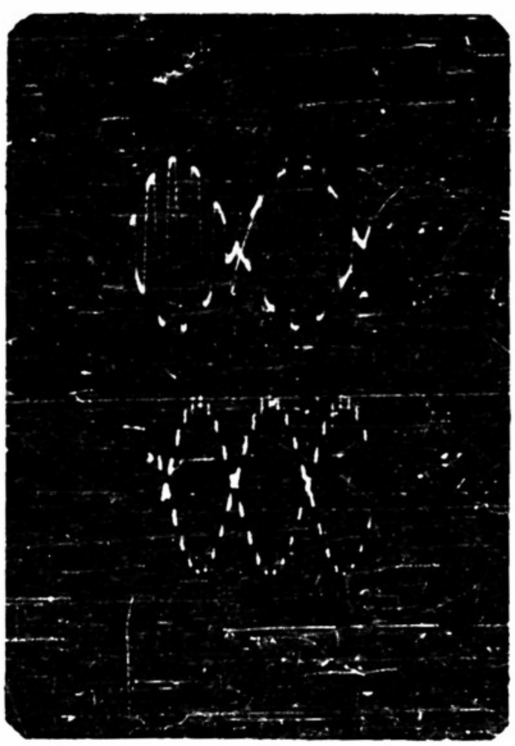


FIGURE 3.22 FULL WAVE SCHREINER DUAL



(a)

(b)

FIGURE 3.23 TYPICAL OUTPUTS FROM FULL WAVE CIRCUIT

(a) 400 CYCLE CARRIER
40 CYCLE SIGNAL

(b) 400 CYCLE CARRIER
20 CYCLE SIGNAL

Interchange of modulation and carrier in the full-wave circuit will also produce a carrier-suppressed output, but no decided advantage is gained in making the change. From the point of view of obtaining good null characteristics, the method indicated is preferred.

CHAPTER IV

CONCLUDING REMARKS

The experimental work described in Chapter III was concerned mainly with realizing a few modulation-demodulation systems which utilize transistors as time-varying parameters. In order to be of practical value, these systems must at least compare favorably with other systems in current use. In this chapter, such a comparison is undertaken.

4.1 Frequency Response

The frequency response of the circuits developed must presently stand as a fundamental limitation in their application. All of the circuits employing single junction transistors exhibited an upper half-power frequency of approximately 50KC. Admittedly, no special precautions were taken with regard to careful wiring, short leads, and shielding in order to ensure good frequency response. For this reason the half-power frequency of the circuits described could probably be increased by a factor of two or more. Beyond this point the response is limited by the intrinsic nature of the transistors themselves. Although this frequency limitation represents a fairly serious drawback in comparison to most other systems available, present indications are that future units will possess a greatly extended response range. Hence the half-power point limitation may be interpreted as a temporary situation, the circuits in question finding use in the higher frequency ranges when suitable transistors become available.

Currently marketed point-contact transistors do not show cut-off until the megacycle range is reached. Unfortunately, these units cause considerably more distortion in the systems described than do junction transistors. For this reason, their use is not recommended.

In circuits employing more than one transistor, a very serious difficulty is encountered. Differences between units become accentuated with increasing frequency. This fact resulted in a frequency limitation for the Schreiner dual circuit of about 10KC, and prohibited the use of transistors in any sort of circuit intended to balance the carrier from the output. But considering the state of technology in the area of transistor manufacture, it is not unreasonable to expect that more uniformity of product is forthcoming.

In summary, no real apology need be made for the fact that the circuits developed are at present useful only in low frequency applications. Considering the comparatively short time that transistors have been available, their development is about as advanced as can be expected.

4.2 Engineering Advantages and Disadvantages

4.21 Comparison with Vacuum Tubes

With respect to linearity and distortion-free operation, transistors offer no decided advantages over vacuum tubes. In unfavorable comparison, transistors are also temperature sensitive and restricted in power handling capacity and frequency response. The extent to which these limitations can be removed remains to be seen. On the favorable side, the advantages of transistors in terms of size, weight, over-all efficiency, and life expectancy are well-known. Their use in low-maintenance systems where vacuum tubes seriously limit trouble-free operation is certainly indicated.

4.22 Comparison with Other Semiconductors

One may well ask what advantages transistors offer over other semiconductors, diodes in particular. Carrier-suppressed modulators and phase-sensitive detectors may easily be constructed from diodes alone. In fact, ring and bridge circuits of this sort have long been used in telephone systems. An answer to this question is now attempted.

In Chapter II it was established that one of the criteria necessary for a linear modulator or demodulator is the effective isolation of carrier and intelligence. With a two-terminal device, such as a diode, no such isolation is possible. Hence any diode circuit so used suffers from a fundamental defect; distortion from nonlinearities present cannot be completely avoided. An added disadvantage of diodes is that they are not capable of any sort of amplification.

As an example for further comparison, let us consider the relative merits of a ring modulator, which employs two transformers and four diodes, and the carrier-suppressed systems discussed in Chapter III. In the ring modulator or demodulator (it will work as either), the carrier must be large in comparison with the input signal. In this way, the carrier may be assumed to control the time at which the diodes switch (i.e. the sensitivity-function). No such assumption is necessary in the systems of Chapter III. The sensitivity-functions there are actually under the control of a separate input. If quadrature voltage is present, little can be done in the ring demodulator to control the instant at which the diodes conduct. Thus the ring demodulator, or any demodulator employing only diodes, is not very practical if quadrature voltage is large. The Schreiner dual circuit, however, has no such limitation.

The effect of quadrature voltage there can be effectively nullified by properly shaping the emitter currents. If quadrature voltage is not present, the half-wave circuit employing complementary symmetry supplies nearly as much information per unit time as does the full-wave diode circuit. In addition, the complementary symmetry circuit provides amplification with a saving of one transformer. The diode circuit is outclassed at every turn.

4.3 Suggestions for Further Work

The entire program of experimental work has been concerned with adapting transistors to practical amplitude modulation and demodulation systems. Certainly not all the possibilities in this respect were investigated. Further work remains to be done in determining other schemes for doing the same job. The largest field open for exploration, however, is that of frequency modulation. A study should be made to determine how transistors may be applied to frequency modulators, limiters, discriminators and ratio detectors.

Acknowledgment

The author wishes to express his sincere appreciation to Dr. Truman S. Gray for supervising the work and suggesting a major part of the formulation.

Credit is also due to Professor A.B. Van Rennes, Professor Thomas F. Jones, Jr., and the entire staff of the Electronic Nuclear Instrumentation Group of the Massachusetts Institute of Technology Servomechanisms Laboratory for their continued suggestions and interest.

BIBLIOGRAPHY

1. Electrical Engineering Staff of M.I.T., Applied Electronics, New York: John Wiley and Sons, Inc., (1943).

2. M.F. Gardner and J.L. Barnes, Transients in Linear Systems, New York: John Wiley and Sons, Inc., (1942).

3. M.H. Goldstein, Jr., "Sensitivity-Function Analysis of Modulation Systems with Statistical Inputs," M.I.T. Department of Electrical Engineering Thesis, (1951).

4. K.L. Schreiner, "High-Performance Demodulators for Servomechanisms," Proceedings of the National Electronics Conference, 2, (1946), 1--11.

5. W. Shockley, Electrons and Holes in Semiconductors, New York: D. Van Nostrand Co., Inc., (1950).

6. Transistor Issue, Proceedings of the I.R.E., 40, (Nov. 1952).

7. R.M. Ryder and R.J. Kircher, "Some Circuit Aspects of the Transistor," Bell System Technical Journal, 28, No. 3, (July, 1949), 367--400.

8. R.L. Wallace, "Duality, a New Approach to Transistor Circuit Design," Proceedings of the I.R.E., 39, (June, 1951).

9. R.L. Wallace and G. Raisbeck, "Duality as a Guide in Transistor Circuit Design," Bell System Technical Journal, 30, (April, 1951), 331--417.

APPENDIX

Mention was made in Chapter II that the sensitivity-function of a carrier-present demodulator may be functionally related to both carrier and intelligence. In fact, it might seem strange that a sensitivity-function could exist at all, since only one input to the demodulator is available. To show that both contentions are valid, consider the simple peak detector shown in figure 1.

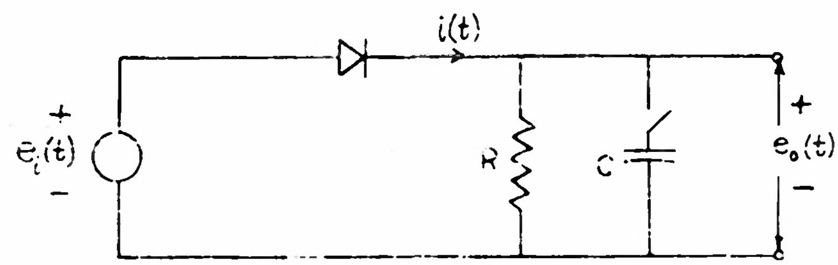


Figure 1: Simple Peak Detector

Assume first that the capacitance, C, is disconnected so that the filter is purely resistive. Since the input, $e_1(t)$, is a carrier-present wave, we may write:

$$e_1(t) = \left[1 + f(t)\right] \cos \omega_c t.$$

If $f(t)$ is always less than unity in absolute value (i.e. modulation always less than 100 percent), $1 + f(t)$ is always greater than zero, and $e_1(t)$ is zero when $\cos \omega_c t$ is zero. Hence, the time during which the diode conducts, corresponding to the positive portion of the input wave, depends only upon the time during which the carrier is positive. The sensitivity-function is thus a periodic square pulse of height $\frac{1}{R}$, period $\frac{2\pi}{\omega_c}$, and pulse width $\frac{\pi}{\omega_c}$.

The presence of a storage element in the filter may radically alter the sensitivity-function. If the capacitance shown in the diagram is switched into the system, the time during which the diode conducts becomes dependent upon the output, $e_o(t)$. Specifically, conduction takes place when $e_i(t) - e_o(t) > 0$. The equation determining the actual times at which conduction begins and ends is involved with difficult transcendental functions. No explicit solution is attempted. To show that a sensitivity-function dependent upon both carrier and intelligence exists, several approximations are made. The time constant of the R-C circuit is assumed to be so long that the natural decay of voltage across the combination is negligible. When the input to the system is an unmodulated carrier, the output is constant at the peak carrier voltage. When intelligence is present in the input wave, the current into the filter is assumed to be a series of impulses occurring at the positive peaks of the carrier. The height of the impulse varies in accordance with the derivative of f(t). For example, if f(t) is a ramp function, all impulses are of a fixed height. Heuristically then, if we let p(t) represent a series of unit impulses which occur at the positive peaks of the carrier,

$$i(t) = \left[\frac{d}{dt} f(t)\right] p(t)$$

From Chapter II, section 2.32:

$$i(t) = e_i(t) S(t)$$

$$\therefore S(t) = \frac{\left[\frac{d}{dt} f(t)\right] p(t)}{\left[1 + f(t)\right] \cos \omega_c t}$$

Both carrier and intelligence appear in the expression.