# FREQUENCY MODULATION ENGINEERING

*by*

CHRISTOPHER E. TIBBS

M.I.E.E., M.Brit.I.R.E.

*Joint Managing Director, Radio Heaters Ltd., Wokingham, England*

*and*

G. G. JOHNSTONE, B.Sc.

*BBC Engineering Training Department*

*foreword by*

K. R. STURLEY

Ph.D., B.Sc., M.I.E.E., Sen.M.I.R.E.

*Head of BBC Engineering Training Department*

*SECOND EDITION REVISED*

*NEW YORK*

JOHN WILEY AND SONS INC.

*440 FOURTH AVENUE*

1956

*First Published 1947*
*Second Edition Revised 1956*

Printed in Great Britain by Jarrold and Sons Ltd., Norwich

# FOREWORD

Of the two carrier characteristics, amplitude and time, that are capable of being modulated, amplitude change has until recently been the only type used for broadcasting purposes.

Frequency modulation, which is one aspect of time modulation (the other is phase modulation) was thought at one time to require less band-width for a given information rate than amplitude modulation, but this myth was exploded by Carson's theoretical analysis of 1922. As stated in the introduction, interest in frequency modulation died down and was only revived by Armstrong's pioneering work (1936) which showed that frequency modulation could provide improved signal-to-noise ratio for a given transmitting power compared with a similar service using amplitude modulation. In fact, he proved that with frequency modulation the effective signal strength is a function of power and band-width and that both random and impulse noise could be reduced by using a wide pass band receiver containing an amplitude limiter.

The war prevented, in England, the same exploitation of frequency modulation for broadcasting as occurred in America. After the war the BBC carried out a prolonged series of comparative a.m./f.m. tests at very high frequencies, which confirmed Armstrong's contentions and added much to our knowledge of high power v.h.f. broadcasting.

Interest in frequency modulation has been given a considerable impetus by the Government's decision, in July 1954, to accept the BBC's recommendations for a v.h.f. broadcasting service using frequency modulation, and the bringing into operation during 1955/6 of the first group of v.h.f. (f.m.) stations included in the BBC's plan. A revised edition of this book is therefore timely, and the increased information made available on interference, aerials, limiters and discriminators, and frequency modulated receivers, will be welcome. The co-author who has undertaken the task is well qualified to do so because he has been responsible for writing instructions concerning the BBC's v.h.f. (f.m.) sound broadcasting station at Wrotham in Kent, which came into regular service on 2nd May 1955, but perhaps above all because he is a keen experimenter with frequency modulation receiver circuits.

K. R. STURLEY

*December 1st, 1955*

# PREFACE TO SECOND EDITION

SINCE the publication of the first edition, there have been extensive additions to the literature dealing with many aspects of frequency modulation engineering, and the engineering practice associated with frequency modulation systems has developed considerably. The book has therefore been enlarged and altered substantially. Whilst the first edition was the work of C. E. Tibbs alone, pressure of business prevented him from devoting the necessary time to the preparation of the second edition, and the present co-author is largely responsible for the revision.

Acknowledgment is due to the BBC for the use of much of its published material, and the revising author is grateful to the Chief Engineer of the BBC for permission to use much valuable information contained in unpublished BBC Reports. Particular thanks are due to Dr. R. D. A. Maurice and G. F. Newell, of the BBC Research Department, for helpful advice and discussions on a number of points. Thanks are also due to S. W. Amos, of the BBC Engineering Training Department, for his help and assistance at all times, and to Dr. K. R. Sturley, Head of the Engineering Training Department, for the interest he has shown in the project, and for his kindness in contributing a foreword.

G. G. JOHNSTONE

TWICKENHAM,
MIDDLESEX
*December 1955*

# PREFACE TO FIRST EDITION

THIS book is intended to provide students, engineers and all those interested, with a concise and readily digestible survey of the whole field of frequency modulation engineering. A number of the classical papers dealing with the theory of this subject are written in such an advanced style that they are almost unintelligible to the average radio engineer. The present work re-presents the basic theory in a form which the author hopes will be more readily understandable. After an introduction the basic properties of a wave modulated in frequency are discussed. As the reader should have at least a nodding acquaintance with the different types of interference and noise structure the third chapter is devoted exclusively to this subject. The means by which interference is suppressed in a frequency modulation system is treated in some detail in the following chapter.

It would not be difficult to write a complete book on the propagation of radio signals in the ultra-short waveband. The chapter on this subject is therefore only a synopsis of the more important points. The same remarks are applicable to the chapter on aerials. The remainder of the book is devoted to a description of the technique and circuits employed for frequency modulation and reception. Wherever possible, circuits of actual commercial equipments have been described and component values indicated. The reader who has a primarily practical outlook will be interested to find that not only has the theory of such components as the discriminator been treated in reasonable detail, but that working designs together with measured response curves are included.

The delay in the issue of this book, due to present difficulties in printing and publishing, coinciding as it did with the Federal Communications Commission's and the British Broadcasting Corporation's choice of the 90 Mc/s region for frequency modulation broadcasts, placed the author in a rather difficult position. Should the book be delayed until sufficient information was available to describe only 90 Mc/s equipment or should it be released for publication substantially as it now stands? In view of the absence of any other satisfactory work covering the same ground the author not only feels justified in offering the present

work, but believes it will make a material contribution towards the progress of frequency modulation in this country. Few if any basic changes have resulted due to the alteration in transmission frequency from 40/50 Mc/s to 90/100 Mc/s. The commercial equipment described is in no way out of date, but rather is suitable for a lower frequency than that now employed. Provided that the reader bears this in mind he will find it detracts little from the usefulness of the book.

A volume of this type is only possible as a result of the efforts of the many authors upon whose work it is based. The present author therefore wishes to make grateful acknowledgment to all those engineers and companies who have published the results of their work in the field of frequency modulation engineering. Many of the names connected with this field will be found in the index at the end of the book. A more personal appreciation is that due to Mr. G. D. Clifford, Secretary of the British Institution of Radio Engineers, to whose lively encouragement the commencement of this book was directly due. The author would also like to offer his warm thanks to Mr. L. H. Bedford for a foreword which is all the more valued for its frankness. Acknowledgment is also made to the *Wireless World* and the *Journal of the British Institution of Radio Engineers* for permission to use both diagrams and material which the author had previously published.

In general, acknowledgment of the source of diagrams and illustrations has been made individually. The author would, however, apologise in advance if in any case credit has been either incorrectly allocated or omitted. If any errors of this or any other type should be found by the reader, he is invited to write the author, care of Messrs. Chapman & Hall, in order that such errors may be corrected in later editions.

C. E. TIBBS

BANSTEAD,
SURREY
*April 1947*

# CONTENTS

## *Chapter One*

# INTRODUCTION

THE use of a frequency modulated carrier wave for the transmission of radio signals is not new. The Poulsen arc, developed well before 1914, transmitted continuous wave signals which were shifted from one frequency to another when the telegraph key was depressed. Since that time the use of frequency modulation has been proposed more than once, as a method of overcoming various difficulties which have occurred during the growth of radio-telephony and broadcasting. Interest in its possibilities was shown when it became apparent that only a rigidly limited number of channels could be accommodated within the medium and long wavebands, which were at that time considered to be the only bands on which a practical broadcast service could be operated. It was suggested that if the carrier wave was maintained at a constant amplitude and modulated with very small frequency swings or "deviations", it would be possible to convey the desired intelligence and at the same time use only a fraction of the band-width necessary to pass the side bands of an amplitude modulated station. It was contended that it would in this way be possible substantially to increase the number of broadcast channels which could be accommodated within any given frequency band.

Serious thought along these lines was, however, brought to a conclusion in 1922, by the publication of one of the first mathematical treatments of frequency modulation. This paper, by J. R. Carson, demonstrated that these ideas were based on a fallacy, and gave for the first time a solution for the spectrum distribution when a wave is modulated in frequency. Carson not only proved that side bands are produced, but also showed that the band-width occupied by these side bands is at least double that of the highest audio modulating frequency. In short, he showed that no reduction in the band-width required for any given station could be obtained by modulating the carrier frequency instead of its amplitude.

For a number of years after the publication of this paper frequency modulation was regarded as of little or no practical

value. However, in 1936, E. H. Armstrong published a paper in which he presented frequency modulation not as a method of cramming more stations into the broadcast band, but as a means of reducing the level of every type of interference. He demonstrated that the earlier mathematical analyses had overlooked the very important point that it is possible to distinguish, at the receiver, between a carrier wave which is frequency modulated and any other undesired signals occupying the same frequency spectrum. It is due to this property—the reduction in level of every type of interference—that frequency modulation or "F.M." for short, has been so rapidly developed during the last few years.

It is perhaps advisable to note at this point that the use of frequency modulation does not in itself result in an improved standard of reproduction, except in so far as it reduces the general noise background. Even before the construction of the first frequency modulation broadcasting station, the sound channel of the BBC television station in London offered similar reproduction fidelity. Reception could, however, be marred by the staccato stutter of ignition interference from passing cars. In changing a very high frequency broadcasting station to frequency modulation, this and all other forms of interference are reduced by some 20 db, which for all practical purposes means that they may be regarded as suppressed.

The medium waveband has for long been used almost exclusively for the transmission of programmes intended primarily for entertainment. This band, however, suffers from many drawbacks, not the least of which is the impossibility of transmitting a satisfactory complement of side bands within the band-width available for each station. The selectivity necessary to separate one station from the next results in the majority of medium-wave broadcast receivers cutting off all side bands—and therefore all audio signals—beyond some 3,000 to 5,000 c/s. The change from a system giving an audio response of this order to one working on the very high frequency band, where it is possible to have an overall characteristic which is flat up to 15,000 c/s, produces a marked improvement in fidelity. When this is combined with the virtual elimination of all types of interference, there is an unanswerable case for the almost universal adoption of frequency modulated transmission for all local high fidelity broadcast stations.

The development of frequency modulation broadcasting was fostered by the conditions which exist in many modern cities. The screening produced by immense steel frame buildings, together with extremely high static levels and the lack of satisfactory aerial arrangements, provided the background against which it was developed. It alone can provide satisfactory reception, in flats which are part of a vast honeycomb packed with every imaginable type of electrical equipment, from hundreds of vacuum cleaners to express lifts.

Tests carried out by the BBC have shown that providing a well-designed receiver is employed, a satisfactory broadcast service in the band 90–100 Mc/s can be obtained with frequency modulation at field strengths as low as 50 $\mu$V/m, whereas, with amplitude modulation and the same amount of noise, the field strength would need to be at least 900 $\mu$V/m. These figures, however, relate to the limitations of receiver noise. When other sources of noise are taken into account, the BBC considers that a minimum field strength of 250 $\mu$V/m is required.

The advantages of frequency modulation are not confined to broadcasting alone. Very greatly improved results are obtained with every type of short-range mobile communication equipment. Tests have been carried out by the International General Electric Company, in which two transmitters were used; the first having a power of 150 watts was situated at Albany, and the second a 50-watt transmitter, was located at Schenectady, some 14½ miles away. Both stations operated on the same wavelength, with both frequency and amplitude modulation. In driving a car equipped with a receiver along a direct route between the two stations the following results were obtained:

| Type of modulation | Interference-free range of 150-watt station | Transitional distance with interference | Interference-free range of 50-watt station |
|---|---|---|---|
| Amplitude . . | 2·3 miles | 11·7 miles | 0·5 mile |
| Frequency . . | 10·5 miles | 1·0 mile | 3·0 miles |

In the following chapters, the complex theory and the engineering technique which lie behind such results as these will be investigated and discussed in some detail.

*Chapter Two*

# THE FREQUENCY MODULATION OF A CARRIER WAVE

BEFORE commencing a detailed examination of the structure of a frequency modulated wave, it will be found helpful to have a general idea of the way in which intelligence may be conveyed by a carrier wave-form. The two principal methods by which a wave may have a second signal impressed upon it are indicated in Fig. 2.1. The first diagram illustrates the application

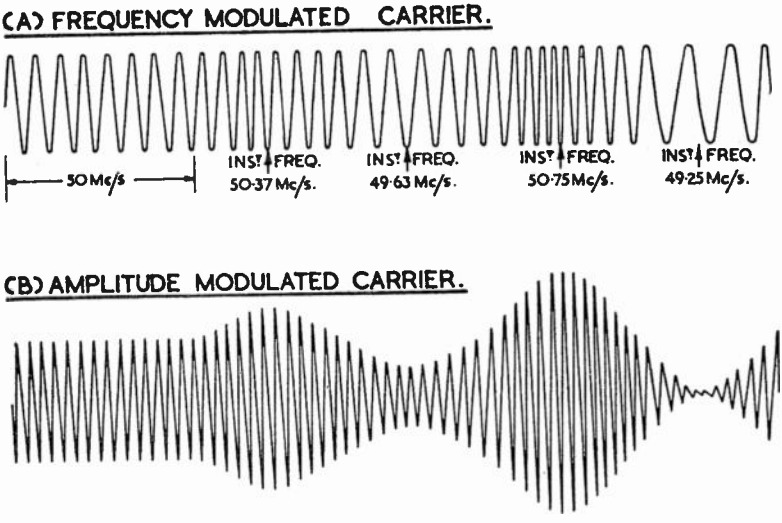



FIG. 2.1.—The general nature of a frequency modulated carrier is compared with that of an amplitude modulated carrier.

*(By courtesy of the British Institute of Radio Engineers.)*

of frequency modulation to the carrier, whilst the second depicts the effect of amplitude modulation. In both cases the same modulating audio signal is applied—two cycles of a sine wave-shape. The amplitude of the first cycle is such that it results in 50 per cent modulation, and that of the second cycle in the maximum permissible modulation; that is to say 100 per cent. In order to recover this audio signal wave at the receiver it is

necessary, in the case of frequency modulation, to provide a demodulation circuit (or discriminator), in which the audio output voltage is directly proportional to the frequency variations of the carrier. In the case of amplitude modulation the detector output voltage must be proportional to the changes in carrier amplitude.

With the aid of Fig. 2.1 it is also possible to make a number of deductions relating to the general nature of a frequency modulated carrier. In the first place, the carrier is steady at its mean or unmodulated frequency until modulation commences. It then swings above and below its mean frequency. The number of excursions which it makes on either side of this mean frequency is directly governed by the frequency of the modulating signal.

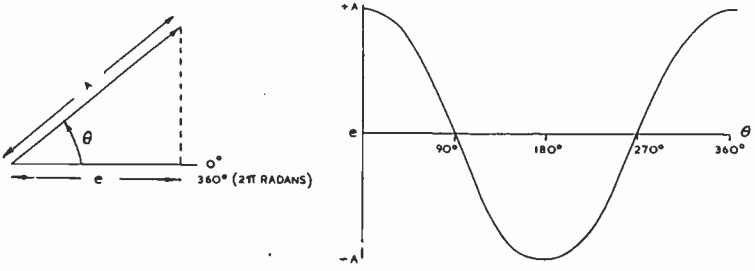



FIG. 2.2.—A simple alternating wave may be represented by the equation $e = A \cos \theta$.

The extent of the frequency swing is directly proportional to the amplitude of the modulating signal. It should be particularly noted that the actual frequency shift has no connection with the frequency of the modulating wave, but is entirely dependent upon its amplitude. One of the most important points which should be brought out at this stage is the fact that the carrier amplitude remains constant regardless of the modulation depth.

In summing up, the general nature of a frequency modulated transmission may be defined as one in which there is no amplitude modulation of the carrier, and in which its frequency faithfully follows the amplitude changes of the modulating wave-shape. In the case of amplitude modulation the carrier amplitude is varied without producing any frequency variation. The amplitude changes are in direct proportion to the modulating signal's amplitude and frequency.

## Modulation

Having now outlined the general form of amplitude and frequency modulated carriers, it is possible to pass on to a more detailed consideration of the whole process of modulation.

The modulation of a wave may be defined as the process by which some characteristic is altered in accordance with the variations of a second signal, such as the voltage fluctuations associated with speech, music, television or telegraph signals. It is proposed, firstly, to establish which of the basic characteristics of a wave can be modulated.

A simple alternating voltage may be represented by the equation:

$$e = A \cos \theta, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.1)$$

where $e$ = the instantaneous voltage amplitude of the wave;
$A$ = the peak voltage amplitude of the wave;
$\theta$ = the instantaneous value of the angle of rotation of the wave vector. This may also be expressed as

$$\theta = \int_0^t \omega dt, \quad . \quad . \quad . \quad . \quad . \quad (2.2)$$

where $\omega = \frac{d\theta}{dt}$ is the instantaneous value of the angular velocity of rotation of the wave vector.

It may therefore be said that

$$e = A \cos \int_0^t \omega dt. \quad . \quad . \quad . \quad . \quad . \quad (2.3)$$

The two basic methods of modulation can be identified from this equation as:

1. Amplitude modulation in which $A$ is varied, and $\omega$ is constant. In this case, expression (2.3) becomes

$$e = A(t) \cos (\omega t + \phi), \quad . \quad . \quad . \quad . \quad (2.4)$$

where $A(t)$ indicates that $A$ varies with time; $\phi = \theta$ at $t = 0$.

2. Angular modulation in which $\omega$ is varied, and $A$ is constant. In this case expression (2.3) becomes

$$e = A \cos \int_0^t \omega(t) dt, \quad . \quad . \quad . \quad . \quad (2.5)$$

where $\omega(t)$ indicates that $\omega$ varies with time.

These two basic modulation groups are in turn divided into a number of different sub-groups each with its particular merits and characteristics. In the first group there is simple amplitude modulation and all the various forms of pulse amplitude modulation. Falling within the second group are phase and frequency modulation—both being special forms of angular modulation.

## Amplitude Modulation

Let it be supposed that a regular periodic change is made about the mean carrier amplitude, at a rate which is slow compared with the carrier frequency. The signal, and it should be observed that the term signal is used in this chapter to denote the modulating wave-form and not the complete modulated carrier, can be expressed as:

$$A_a \cos \omega_a t,$$

where $A_a$=the peak signal voltage;
$\omega_a = 2\pi f_a$, where $f_a$ is the modulating signal frequency;
$\omega_a t$=the signal voltage vector rotation measured in radians.

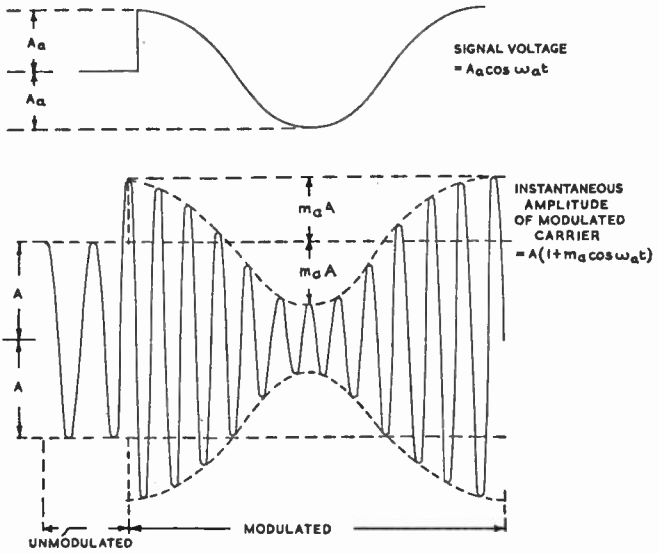



FIG. 2.3.—Carrier amplitude modulated with a cosine wave signal. A modulation factor $m_a$=0·5, results in 50 per cent modulation.

If now the percentage amplitude modulation is made equal to a modulation factor $m_a$, multiplied by 100, then it follows from the definition of an amplitude modulated wave that $m_a A \equiv A_a$.

Under these conditions $A(t)$ in equation (2.4) becomes

$$A(1+m_a \cos \omega_a t). \quad . \quad . \quad . \quad . \quad . \quad (2.6)$$

It will be seen that this indicates a periodic amplitude change about the value of the unmodulated carrier amplitude $A$, the extent of this change being determined by the modulation factor $m_a$. If $A$ had been merely modified by $m_a \cos \omega_a t$, this would have indicated a change about a zero datum line.

By combining expressions (2.4) and (2.6), and taking $\phi = 0°$, an expression for an amplitude modulated carrier is obtained.

$$e = A \cos \omega t(1+m_a \cos \omega_a t). \quad . \quad . \quad . \quad (2.7)$$

This formula indicates that the wave consists of a high-frequency carrier, $A \cos \omega t$, which is constant in frequency, but which is varied in amplitude in accordance with the signal wave, about the mean carrier amplitude $A$.

Expression (2.7) can be expanded to give the full spectrum distribution as follows:

$$\begin{aligned} e &= A \cos \omega t + A m_a \cos \omega_a t \cos \omega t \\ &= A \cos \omega_a t + \frac{m_a A}{2} \cos (\omega - \omega_a)t + \frac{m_a A}{2} \cos (\omega + \omega_a)t. \end{aligned} \quad (2.8)$$

From this it will be seen that the same modulated carrier may also be considered as being built up of a spectrum of constant amplitude, constant frequency waves. This spectrum consists of the original carrier, $A \cos \omega t$, and two sets of high-frequency waves, $\frac{m_a A}{2} \cos (\omega - \omega_a)t$ and $\frac{m_a A}{2} \cos (\omega + \omega_a)t$, known as the side bands, and spaced $f_a$ cycles on either side of the carrier. The amplitude of these side bands will be dependent on the modulation factor $m_a$, and will at 100 per cent modulation (i.e. when $m_a = 1$) reach a maximum of one-half the carrier amplitude.

The magnitude of the modulated wave at any instant is given by the sum of the projections on the reference axis $\theta = 0$ of the three vectors corresponding to the components of the wave, as shown in Fig. 2.4(*a*). The instantaneous wave magnitude can also be found by considering the projections of the side band vectors

on the carrier vector. This leads to the vector diagram of Fig. 2.4(*b*); the upper side band vector rotates in the positive (anti-clockwise) direction relative to the carrier vector, whilst the lower side band rotates in the negative direction. The instantaneous magnitude of the carrier vector is thus $A(1+m_a \cos \omega_a t)$ as given in expression (2.8).

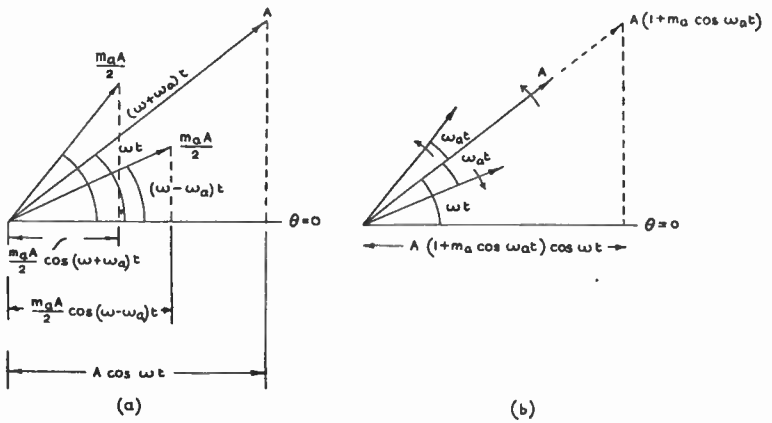



FIG. 2.4.—Diagram (*a*) shows the wave magnitude as the sum of the projections of the side band and carrier vectors on the axis $\theta=0$. Diagram (*b*) shows the variation of the carrier vector magnitude as the sum of its unmodulated magnitude and the projections of the side band vectors upon it.

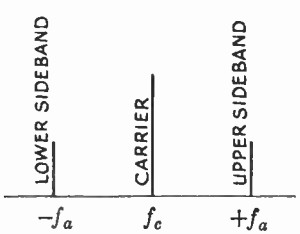



FIG. 2.5.—The side band spectrum of simple amplitude modulated wave.

The total radiated power contained in the side bands at 100 per cent modulation will be half the carrier power, which remains unchanged under all conditions. It will be shown in the following section that matters are entirely different for all forms of angular modulation including, of course, frequency modulation, where the

total radiated power remains constant, and a large proportion of this power is contained in the side bands. It is even possible for the carrier amplitude to fall to zero. It is this important difference which makes a frequency modulation transmitter so much more efficient than its amplitude modulation counterpart.

### Angular Modulation

The general expression for all forms of angular modulation is given by (2.5),

$$e=A \cos \int_0^t \omega(t)dt,$$

where $\omega(t)$ is the instantaneous value of the angular velocity of the wave vector. This can be expressed as the sum of two components, one constant and equal to the angular velocity ($\omega_c$) of the unmodulated carrier vector, and the other varying with time, related to the modulating signal amplitude. Then

$$\omega(t)=\omega_c+\omega_1(t). \quad . \quad . \quad . \quad . \quad . \quad (2.9)$$

The actual value of $\omega_1(t)$ will be considered in detail in the discussion of the various types of angular modulation.

Combining expressions (2.5) and (2.9), the instantaneous value of the wave amplitude is given by

$$e=A \cos \int_0^t \{\omega_c+\omega_1(t)\}\, dt$$

$$= +A \cos \{\omega_c t+\int_0^t \omega_1(t)dt\} \quad . \quad . \quad (2.10)$$

$$=A \cos \{\omega_c t+\phi(t)\}, \quad . \quad . \quad . \quad . \quad . \quad (2.11)$$

where

$$\phi(t)=\int_0^t \omega_1(t)dt, \quad . \quad . \quad . \quad . \quad . \quad (2.12)$$

is the instantaneous value of the wave phase angle $\phi$, the angle between the modulated carrier vector and the mean or unmodulated carrier vector.

From expression (2.11), it can be seen that angular modulation can also be defined in terms of variation of the wave phase angle $\phi$. If the wave frequency is made to vary directly with the amplitude of the modulating signal, frequency modulation results; if the wave phase angle is made to vary directly with the amplitude of the modulating signal, phase modulation results. Before

discussing the forms of angular modulation in particular, it is necessary to elaborate on the meaning of wave frequency and phase angle, and the relationship between the two.

### Wave Frequency and Phase Angle

The frequency of a wave is normally defined as the number of rotations of the wave vector (cycles) in a given period of time, generally expressed in cycles per second, or multiples of this unit. Where, however, the wave angular velocity is not constant, as in the case of angular modulation, the frequency as estimated by the number of vector rotations in a period of time yields only an average value. In order to define the instantaneous value of the wave frequency, the angle swept out per rotation ($2\pi$ radians) must be divided by the instantaneous value of the wave vector velocity. This then, is the time of rotation the vector wave would have if the instantaneous value of the angular velocity $\omega(t)$ were maintained over a period; consequently the corresponding instantaneous value of the wave frequency is the inverse of this. Designating the instantaneous wave frequency $f(t)$,

$$f(t)=\frac{\omega(t)}{2\pi}. \qquad . \quad . \quad . \quad . \quad . \quad (2.13)$$

When the wave vector angular velocity has a fixed and a variable component, as defined in expression (2.9),

$$\begin{aligned} f(t) &= \frac{\omega_c}{2\pi}+\frac{\omega_1(t)}{2\pi} \\ &= f_c+f_1(t), \qquad . \quad . \quad . \quad . \quad . \quad (2.14) \end{aligned}$$

where $f_c$ is the carrier frequency,

$f_1(t)$ is the instantaneous frequency corresponding to $\omega_1(t)$, i.e. $2\pi f_1(t)=\omega_1(t)$.

Expression (2.14) states that the instantaneous value of the wave frequency shift, i.e. the departure of the wave frequency from its unmodulated value, is equal to $f_1(t)$. If, then, $f_1(t)$ is directly proportional to the modulating signal magnitude, the wave frequency shift is proportional to the modulating signal magnitude, and hence this type of angular modulation is termed frequency modulation.

The instantaneous value of the wave phase shift is defined as the angle between the instantaneous position of the wave vector and

the position it would occupy if unmodulated. If this phase shift $\phi(t)$ as defined in expression (2.11) is made directly proportional to the magnitude of the modulating signal, the form of angular modulation termed phase modulation results.

The relationship between the instantaneous value of the wave frequency shift $f_1(t)=(f(t)-f_c)$ and the instantaneous value of the phase shift $\phi(t)$ can be found by combining expressions (2.12) and (2.14),

$$\phi(t) = \int_0^t 2\pi f_1(t)dt, \quad . \quad . \quad . \quad . \quad . \quad (2.15)$$

or, alternatively, by differentiating (2.15),

$$\frac{d}{dt}\left\{\phi(t)\right\} = 2\pi f_1(t). \quad . \quad . \quad . \quad . \quad (2.16)$$

These expressions are of fundamental importance, since they show that frequency shift and phase shift are inseparable, and the relationship between them. Expressed in words, it may be stated that the instantaneous value of the wave frequency shift is equal to $1/2\pi$ times the instantaneous rate of change of phase angle.

### Frequency Modulation

As stated above, if the wave frequency shift is made proportional to the modulating signal magnitude, frequency modulation ensues. With a cosinusoidal modulating signal, the resultant wave will have alternate "compressions" and "rarefactions", to borrow from the sound-wave analogy. The degree of "compression" and "rarefaction" will be proportional to the amplitude of the modulating signal whilst the occurrence of the "compressions" and "rarefactions" will correspond to the signal frequency.

It is convenient at this point to define the terms used in connection with frequency modulation; in particular the meaning assigned to frequency shift, frequency swing and frequency deviation. The term frequency shift is used to describe the departure of the signal frequency from its unmodulated value. The term frequency swing is reserved for the maximum value of frequency shift with a sinusoidal input signal, i.e. the frequency swing corresponds to the amplitude of the modulating signal. The term frequency deviation is a parameter of a given transmitting

system, and is the maximum value of frequency shift permitted; this point is discussed further later.

If the signal applied to the input of the modulating system is $A_a \cos \omega_a t$, and $b$ is a constant, equal to the frequency shift occurring per volt of applied signal,

$$f_1(t) = bA_a \cos \omega_a t. \quad . \quad . \quad . \quad . \quad (2.17)$$

From expression (2.14) $f_1(t) = f(t) - f_c = \omega_1(t)/2\pi$ and combining this with expression (2.10), the expression for a frequency modulating wave becomes

$$\begin{aligned} e &= A \cos \left\{ \omega_c t + \int_0^t 2\pi b A_a \cos \omega_a t dt \right\} \\ &= A \cos \left\{ \omega_c t + \frac{2\pi}{\omega_a} b A_a \sin \omega_a t \right\} \\ &= A \cos \left\{ \omega_c t + \frac{bA_a}{f_a} \sin \omega_a t \right\}, \quad . \quad . \quad . \quad (2.18) \end{aligned}$$

since $2\pi f_a = \omega_a$.

This expression may be rearranged into a more general form by eliminating $b$ and $A_a$. These terms are associated with the modulating system, and it is more convenient generally if the wave frequency swing is introduced. If $f_s$ is the frequency swing corresponding to the amplitude of the modulating signal, $f_s = bA_a$, expression (2.18) can be rewritten as

$$e = A \cos \left\{ \omega_c t + \frac{f_s}{f_a} \sin \omega_a t \right\}. \quad . \quad . \quad . \quad (2.19)$$

By analogy with the case of amplitude modulation, it might be expected that 100 per cent modulation would occur when the maximum value of the frequency swing equalled the unmodulated carried frequency; in this case, the carrier frequency would be swept between the limits 0 and $2f_c$. Such a system is, however, completely impracticable.

In practice, an arbitrary upper limit $f_d$ is set for the frequency swing and this is called the frequency deviation. This upper limit may be considered the equivalent of 100 per cent modulation. The choice of this limit is governed by two primary factors, signal to noise ratio and the band-width required for transmission. As will be shown later, the limit is required to be as high as possible to secure a good signal to noise ratio. The limit is required to be

as low as possible to reduce the band-width required for transmission. The compromise value generally adopted for broadcasting systems is 75 kc/s; for communications systems this is often reduced to 15 kc/s.

Since $f_d$ corresponds to $A_{a\ max}$, the maximum amplitude of the modulating signal, it is possible to introduce a modulation factor defined by

$$m = \frac{A_a}{A_{a\ max}} = \frac{f_s}{f_d} \quad . \quad . \quad . \quad . \quad (2.20)$$

and combining this expression with expression (2.19),

$$e = A \cos \left\{ \omega_c t + \frac{m f_d}{f_a} \sin \omega_a t \right\} . \quad . \quad . \quad . \quad (2.21)$$

**Phase Modulation**

If, as stated above, the wave phase angle is made directly proportional to the modulating signal amplitude, phase modulation ensues. If a cosinusoidal modulating signal is considered, the wave vector will swing about its mean or unmodulated position in such a manner that the instantaneous value of the angle between the vector and its unmodulated position is proportional to the modulating signal magnitude. The frequency of the fluctuations about the mean position will be equal to the frequency of the modulating signal. With a constant frequency input, the angular deviations increase linearly with the modulating signal amplitude. If the signal applied to the input of the modulating system is $A_a \cos \omega_a t$, and $b_1$ is a constant, equal to the phase shift in radians per volt of applied signal,

$$\phi(t) = b_1 A_a \cos \omega_a t. \quad . \quad . \quad . \quad . \quad (2.22)$$

Combining this expression with expression (2.11), the expression for a phase modulated wave becomes

$$e = A \cos \{\omega_c t + b_1 A_a \cos \omega_a t\}. \quad . \quad . \quad . \quad (2.23)$$

By analogy with the frequency modulation case, $b_1 A_a$ may be replaced by $m\phi_d$, where $\phi_d$ is the phase shift produced by the maximum amplitude of the modulating signal, and $m$ is the modulation factor defined by $m = A_a / A_{a\ max}$. Whence

$$e = A \cos \{\omega_c t + m\phi_d \cos \omega_a t\}. \quad . \quad . \quad . \quad (2.24)$$

**Relationship between Frequency and Phase Modulation**

It was shown in expressions (2.15) and (2.16) that any frequency shift of a wave is accompanied by phase shift, and conversely. Thus a frequency modulated wave may be considered in terms of the phase shift of the carrier vector; similarly, a phase modulated wave may be considered in terms of the wave frequency shift.

Consider firstly a frequency modulated wave. The instantaneous value of the phase shift can be seen directly from expression (2.21) to be

$$\phi(t)=\frac{mf_d}{f_a}\sin\omega_a t. \quad . \quad . \quad . \quad . \quad . \quad (2.25)$$

This expression shows that, with a constant amplitude modulating signal, i.e. $m$ constant, the wave phase shift is swept between the limits inversely proportional to $f_a$ in contrast to the analagous case in phase modulation, where the limits are constant. It also shows that the instantaneous value of the phase shift for a frequency modulated wave is in quadrature with the modulating signal magnitude. Both of these effects are due to the fact that the phase shift is proportional to the integral of the frequency deviation. If the signal applied to the modulating system had been made proportional to the differential coefficient of the modulating signal, the processes of differentiation and integration would nullify each other, and a phase modulated signal would result. Since the process of differentiating a signal wave-form can be achieved in practice, a frequency modulation system can be made to produce a phase modulated wave.

Considering now a phase modulated wave in terms of the accompanying frequency shift, expression (2.16) shows that

$$\begin{aligned} f_1(t) &= \frac{1}{2\pi}\frac{d}{dt}(m\phi_d\cos\omega_a t) \\ &= -\frac{1}{2\pi}m\phi_d\omega_a\sin\omega_a t \\ &= -m\phi_d f_a\sin\omega_a t, \quad . \quad . \quad . \quad . \quad (2.26) \end{aligned}$$

since $2\pi f_a=\omega_a$.

This expression shows that, with a constant amplitude modulating signal, i.e. $m$ constant, the frequency shift is swept between limits directly proportional to $f_a$, in contrast to the analagous case

in frequency modulation, where the limits are constant. The expression also shows that the instantaneous value of the frequency swing is in quadrature with that of the modulating signal magnitude. These effects arise from the fact that the frequency deviation is proportional to the differential coefficient of the phase shift.

If the signal applied to the modulating system had been made proportional to the integral of the modulating signal, the processes of differentiation and integration would nullify each other, and a frequency modulated signal would have resulted. Since the

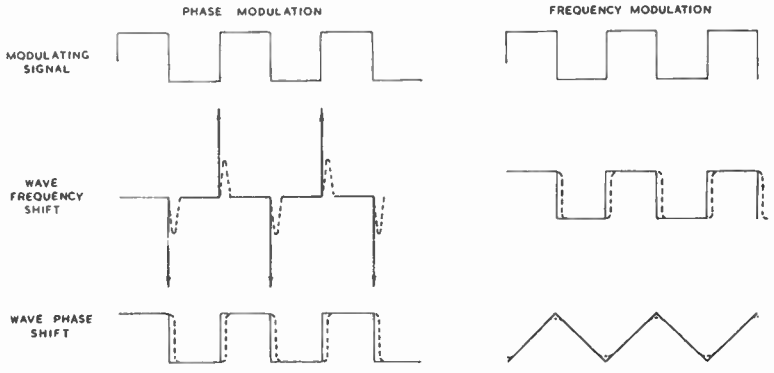



Fig. 2.6—The effect on the carrier wave of a square wave modulating signal. The dotted lines indicate the practical effects obtained.

process of integration of a signal can be achieved in practice, a phase modulation system can be made to produce a frequency modulated wave. This fact is often utilised in practical systems.

It can thus be seen that frequency and phase modulation are very closely related; in fact, without some information as to the nature of the modulation, it is impossible to distinguish a frequency modulated wave from a phase modulated wave by inspection of the wave-form.

The differences between frequency and phase modulation can be shown most clearly by considering a non-sinusoidal modulating signal. When modulation of sinusoidal type is considered, the differences are not clearly marked since the integral and differential coefficients have the same wave-shape. The differences are made most apparent perhaps by considering a rectangular wave modulation wave-form, as suggested by Professor G. W. O. Howe. The resultant frequency shift and phase shift characteristics for

frequency and phase modulation are shown in Fig. 2.6. Here the integral of the modulating signal has a triangular wave-shape, and consequently, from expression (2.15), the phase shift characteristic for a frequency modulated wave has this shape. The differential coefficient of the modulating signal is a series of alternate positive-going and negative-going spikes, of infinite amplitude, since the modulating signal amplitude is assumed to change by a finite amount in an infinitely short time. From expression (2.16), the frequency shift characteristic of a phase modulated wave also has this wave-shape.

In practice, these wave-shapes with discontinuities would be impossible to realise since they would require infinitely large band-widths for their transmission; the practical results of applying such a rectangular wave modulating signal to practical systems are indicated by the dotted lines of Fig. 2.6.

### Other Forms of Angular Modulation

Phase and frequency modulation are not the only possible types of angular modulation; they are only two members of an infinitely large group. Another member of the group is angular acceleration modulation. Whereas in phase modulation, the phase shift is directly proportional to the modulating signal magnitude, and in frequency modulation the first differential coefficient of the phase shift is proportional to the modulating signal magnitude, in angular acceleration modulation, the second differential coefficient of the phase shift is proportional to the modulating signal magnitude. With an input signal $A_a \cos \omega_a t$ applied to the modulating system, the instantaneous wave magnitude would be given by

$$e = A \cos\left(\omega_c t + \frac{b_2 A_a}{\omega_a{}^2} \cos \omega_a t\right), \quad . \quad . \quad . \quad (2.27)$$

where $b_2$ is a constant associated with the modulating system. In this type of modulation, the phase shift is inversely proportional to the square of the modulating signal frequency. It will be noted in passing that by analogy with the name of angular acceleration modulation, frequency modulation could be termed angular velocity modulation.

It will be seen that further forms of angular modulation can be derived by making higher differential coefficients of the phase shift proportional to the modulating signal magnitude. Similarly,

yet further forms could be derived by making successive integrals of the phase shift proportional to the modulating signal magnitude. However, there is no real need to consider such systems, since in practice frequency modulation is generally considered the most satisfactory type of angular modulation. This can be shown by comparison with phase and angular acceleration modulation; the successive forms suggested above merely have the relative defects of these latter types in more accentuated form.

### The Relative Merits of Frequency and Phase Modulation

In view of the number of different types of angular modulation, those factors which have led to the general use of frequency modulation rather than one of the other relationships, are at least worthy of note.

There are two factors which, taken together, for all practical purposes decide the issue. Firstly, whatever method or form of modulation is employed, the limits of the channel allocated to any given transmitter must be defined in terms of frequency. The method of modulation which makes the best use of the frequency band available will therefore have much in its favour. The second deciding factor again arises from limitations which are met in practice. Up to the present all the circuits available for the demodulation of angular modulated carriers have produced an audio output voltage which is directly proportional to the variations in carrier frequency.

As the consideration of the advantages and disadvantages of frequency and phase modulation will very largely centre around these two controlling factors, it is suggested that the reader should, for convenience, also think in terms of frequency; and when considering phase or any other angular modulation visualise it as a special form of frequency modulation.

In order to assist in the building of such a mental picture, it is suggested that reference is made to the three diagrams given in Fig. 2.7. In these diagrams the frequency deviation resulting from 100 per cent modulation has been indicated for the three principal forms of angular modulation. It does not require a very close examination of these diagrams to show that the relationship which results in the greatest overall efficiency in the use of the frequency band employed is undoubtedly frequency modulation. By efficient use of a band, it is meant that the frequency space

necessary is all employed to an equal extent in conveying the signal.

It has already been stated that the practical demodulation circuits available have a direct frequency to output voltage relationship. As most normal programme material produces maximum modulation depths over the band from 100 to 1,000 c/s, it is obvious that the demodulator circuit should be supplied with a signal which will allow it to produce its full voltage output over this region. Normally, the signal voltages over the remainder of the audio band will be of smaller amplitude. As the

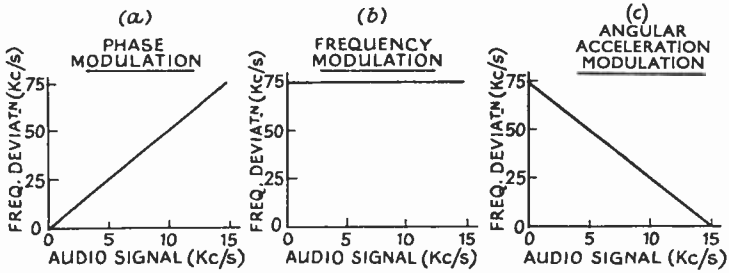



FIG. 2.7.—The above diagrams show the carrier frequency deviations resulting from 100 per cent modulation (arbitrarily fixed at 75 kc/s) at frequencies within the audio band, for the three principal types of angular modulation.

discriminator (the frequency modulation detector circuit) output voltage is the direct resultant of the carrier frequency deviations, it is apparent that if the full output is to be usefully employed, the modulation system adopted must be one in which this band of audio frequencies produces the maximum frequency deviation which can be permitted. Reference to Fig. 2.7 shows that frequency modulation alone fulfils these conditions.

If the use of phase modulation is considered the comparison will be found somewhat unfavourable. In order to reproduce a phase modulated transmission without audio amplitude distortion, its demodulated signals must be corrected to produce a constant relationship between the output voltage and the carrier frequency variations. The only way in which this can be achieved is to attenuate the higher audio frequencies, as shown in Fig. 2.8. In the example given, the correction necessary will result in the output actually available from the discriminator being approximately one-three-hundredth part of the maximum voltage it

develops. This figure assumes an audio characteristic which is flat from 15,000 c/s down to 50 c/s. In order to be comparable with a frequency modulation system this means that either the field strength or the receiver gain will have to be increased by some 300 times. As the phase modulation relationship offers no apparent advantage over frequency modulation, it may be said that on the ground of practical economy it is ruled out for any normal applications.

Angular acceleration modulation may be discounted for the same reasons, as its demodulated signals would also have to be attenuated in order to produce a level audio response.

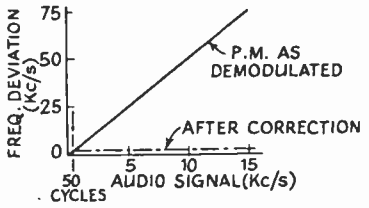



FIG. 2.8.—As the discriminator will demodulate a phase-modulated signal with a rising audio frequency characteristic, it is necessary in order to avoid amplitude distortion, to correct this characteristic in the manner shown above.

Although, as discussed in a later chapter, slightly improved results can be obtained with a relationship which is between frequency and phase modulation, this system—known as transmitter pre-emphasis—is definitely based on the fundamental frequency modulation relationship.

## Frequency Modulation Side Bands

Whether amplitude, frequency or phase modulation is employed, the process will be found to produce a number of side band frequencies. If the intelligence impressed on the carrier is to be faithfully reproduced at the receiver, it is essential that these side bands are not suppressed or altered in their relative amplitudes at any point in the system. If for any reason the side bands of a modulated carrier are suppressed, then the intelligence they carry will be eliminated.

Before it is possible to commence the design of any equipment for use with a practical frequency modulation system, it is essential that the band-width necessary to pass the side bands is first

established. The only satisfactory method of arriving at the actual frequency spectrum of any modulated wave-form is mathematically. This is especially the case with frequency modulated transmissions, which result in a large number of side bands of an extremely complex nature. In order to establish the spectrum of a frequency modulated carrier it is necessary to develop its voltage distribution equation. In so doing it will be necessary to employ an integral equation which was first obtained by Bessel in 1824—while he was investigating planetary motion. Bessel's equation may be defined as follows:

$$J_n(x) = \frac{1}{2\pi}\int_0^{2\pi} \cos\,(n\theta - x\sin\theta)d\theta.$$

The value of $J_n(x)$ is known as a Bessel function of the first kind, and of order $n$.

It was shown earlier that the expression for a wave, modulated in frequency by a single cosine wave-form, was as follows:

$$e = A\cos\left(\omega_c t + \frac{mf_d}{f_a}\sin\omega_a t\right).$$

The term $mf_d/f_a$ is termed the modulation index, and will be designated $m_p$; from the expression it is obvious that $m_p$ is equal to the peak value of the phase shift. Since $f_d$ is fixed, $m_p$ varies directly with the modulating signal amplitude, and inversely with its frequency. Introducing $m_p$ in the expression above,

$$e = A\cos\,(\omega_c t + m_p\sin\omega_a t).$$

This expression can now be developed into the form of a spectrum of constant amplitude, constant frequency waves as follows.

Expanding the expression above,

$$e/A = \cos\omega_c t\cos\overline{(m_p\sin\omega_a t)} - \sin\omega_c t\sin\overline{(m_p\sin\omega_a t)}.$$

By employing the two expansions

$$\cos\overline{m_p\sin\omega_a t} = J_0(m_p) + 2J_2(m_p)\cos 2\omega_a t + 2J_4(m_p)\cos 4\omega_a t\ldots$$

and

$$\sin\overline{m_p\sin\omega_a t} = 2J_1(m_p)\sin\omega_a t + 2J_3(m_p)\sin 3\omega_a t\ldots,$$

it may be shown that

$$e/A = J_0(m_p)\cos\omega_c t + 2J_2(m_p)\cos\omega_c t\cos 2\omega_a t\ldots$$
$$-2J_1(m_p)\sin\omega_c t\sin\omega_a t - 2J_3(m_p)\sin\omega_c t\sin 3\omega_a t\ldots \quad (2.28)$$

These terms may be expanded further to give

$$\begin{aligned} e/A = & J_0(m_p)\cos\omega_c t \\ & + J_1(m_p)[\cos(\omega_c+\omega_a)t - \cos(\omega_c-\omega_a)t] \\ & + J_2(m_p)[\cos(\omega_c+2\omega_a)t + \cos(\omega_c+2\omega_a)t] \\ & \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \\ & + J_{2n-1}(m_p)\{\cos[\omega_c+(2n-1)\omega_a]t - \cos[\omega_c-(2n-1)\omega_a]t\} \\ & + J_{2n}(m_p)[\cos(\omega_c+2n\omega_a)t + \cos(\omega_c-2n\omega_a)t], \quad . \quad (2.29) \end{aligned}$$

where $A$ = unmodulated carrier amplitude;

$J_n(m_p)$ = Bessel function of the first kind, of order $n$ for the argument $m_p$;

$m_p = \dfrac{mf_d}{f_a}$ = the modulation index; $f_d$ is the frequency deviation, $m$ is the modulation factor, i.e. the ratio of the modulating signal amplitude to the peak modulating signal amplitude, and $f_a$ is the modulating signal frequency. Also equal to peak phase shift.

By using the property of Bessel function that $J_n = (-1)^n J_{-n}$, where $n$ is integral, the above expression reduces to the very simple form

$$e = A\sum_{n=-\infty}^{\infty} J_n(m_p)\cos(\omega_c+n\omega_a)t. \quad . \quad . \quad . \quad (2.30)$$

It will be seen from expression (2.29) that for any given value of $m_p$, there is a carrier component, of amplitude $J_0(m_p)$, and an infinite number of side bands at frequencies which are integral multiples of the modulating signal frequency removed from the carrier. The amplitudes of these side bands individually are determined by the corresponding Bessel coefficient. It will be noted that for the side bands which are at frequencies corresponding to odd multiples of the modulating signal frequency removed from the carrier, the upper and lower side bands have opposite signs. The significance of this fact is discussed in the next section.

The relative amplitudes of the side bands themselves ($J_1(m_p)$, $J_2(m_p)$, etc.), can be ascertained from a suitable table of Bessel function values. All the values of $J_n(m_p)$ which are likely to be required in practice have been given in Tables 1 and 2.

TABLE 1

*Bessel function values for modulation indices less than unity*

| $n$ | $J_n(0{\cdot}1)$ | $J_n(0{\cdot}2)$ | $J_n(0{\cdot}3)$ | $J_n(0{\cdot}4)$ | $J_n(0{\cdot}5)$ | $J_n(0{\cdot}6)$ | $J_n(0{\cdot}7)$ | $J_n(0{\cdot}8)$ | $J_n(0{\cdot}9)$ | $J_n(1{\cdot}0)$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0·9975 | 0·9900 | 0·9776 | 0·9604 | 0·9385 | 0·9120 | 0·8812 | 0·8463 | 0·8075 | 0·7652 |
| 1 | 0·0499 | 0·0995 | 0·1483 | 0·1960 | 0·2423 | 0·2867 | 0·3290 | 0·3688 | 0·4059 | 0·4401 |
| 2 | — | — | 0·0112 | 0·0197 | 0·0306 | 0·0437 | 0·0588 | 0·0758 | 0·0946 | 0·1149 |
| 3 | — | — | — | — | — | — | — | 0·0102 | 0·0144 | 0·0196 |

NOTE.—Only those values greater than 0·0100 are given.

## TABLE 2

*Bessel function values for modulation indices up to 15*

| $n$ | $J_n(1)$ | $J_n(2)$ | $J_n(3)$ | $J_n(4)$ | $J_n(5)$ | $J_n(6)$ | $J_n(7)$ | $J_n(8)$ | $J_n(9)$ | $J_n(10)$ | $J_n(11)$ | $J_n(12)$ | $J_n(13)$ | $J_n(14)$ | $J_n(15)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0·7652 | 0·2239 | −0·2601 | −0·3971 | −0·1776 | 0·1506 | 0·3001 | 0·1717 | −0·0903 | −0·2459 | −0·1712 | 0·0477 | 0·2069 | 0·1711 | −0·0142 |
| 1 | 0·4401 | 0·5767 | 0·3391 | −0·0660 | −0·3276 | −0·2767 | −0·0047 | 0·2346 | 0·2453 | 0·0435 | −0·1768 | −0·2234 | −0·0703 | 0·1334 | 0·2051 |
| 2 | 0·1149 | 0·3528 | 0·4861 | 0·3641 | 0·0466 | −0·2429 | −0·3014 | −0·1130 | 0·1448 | 0·2546 | 0·1390 | −0·0849 | −0·2177 | −0·1520 | 0·0416 |
| 3 | 0·0196 | 0·1289 | 0·3091 | 0·4302 | 0·3648 | 0·1148 | −0·1676 | −0·2911 | −0·1809 | 0·0584 | 0·2273 | 0·1951 | 0·0033 | −0·1768 | −0·1940 |
| 4 | — | 0·0340 | 0·1320 | 0·2811 | 0·3912 | 0·3576 | 0·1578 | −0·1054 | −0·2655 | −0·2196 | −0·0150 | 0·1825 | 0·2193 | 0·0762 | −0·1192 |
| 5 | — | — | 0·0430 | 0·1321 | 0·2611 | 0·3621 | 0·3479 | 0·1858 | −0·0550 | −0·2341 | −0·2383 | −0·0735 | 0·1316 | 0·2204 | 0·1305 |
| 6 | — | — | 0·0114 | 0·0491 | 0·1310 | 0·2458 | 0·3392 | 0·3376 | 0·2043 | −0·0145 | −0·2016 | −0·2437 | −0·1180 | 0·0812 | 0·2061 |
| 7 | — | — | — | 0·0152 | 0·0534 | 0·1296 | 0·2336 | 0·3206 | 0·3275 | 0·2167 | 0·0184 | −0·1703 | −0·2406 | −0·1508 | 0·0345 |
| 8 | — | — | — | — | 0·0184 | 0·0565 | 0·1280 | 0·2235 | 0·3051 | 0·3179 | 0·2250 | 0·0451 | −0·1410 | −0·2320 | −0·1740 |
| 9 | — | — | — | — | — | 0·0212 | 0·0589 | 0·1263 | 0·2149 | 0·2919 | 0·3089 | 0·2304 | 0·0670 | −0·1143 | −0·2200 |
| 10 | — | — | — | — | — | — | 0·0235 | 0·0608 | 0·1247 | 0·2075 | 0·2804 | 0·3005 | 0·2338 | 0·0850 | −0·0901 |
| 11 | — | — | — | — | — | — | — | 0·0256 | 0·0622 | 0·1231 | 0·2010 | 0·2704 | 0·2927 | 0·2357 | 0·0999 |
| 12 | — | — | — | — | — | — | — | — | 0·0274 | 0·0634 | 0·1216 | 0·1953 | 0·2615 | 0·2855 | 0·2367 |
| 13 | — | — | — | — | — | — | — | — | 0·0108 | 0·0290 | 0·0643 | 0·1201 | 0·1901 | 0·2536 | 0·2787 |
| 14 | — | — | — | — | — | — | — | — | — | 0·0119 | 0·0304 | 0·0650 | 0·1188 | 0·1855 | 0·2464 |
| 15 | — | — | — | — | — | — | — | — | — | — | 0·0130 | 0·0316 | 0·0656 | 0·1174 | 0·1813 |
| 16 | — | — | — | — | — | — | — | — | — | — | — | 0·0140 | 0·0327 | 0·0661 | 0·1162 |
| 17 | — | — | — | — | — | — | — | — | — | — | — | — | 0·0149 | 0·0337 | 0·0665 |
| 18 | — | — | — | — | — | — | — | — | — | — | — | — | — | 0·0158 | 0·0346 |
| 19 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | 0·0166 |

NOTE.—Only these values greater than 0·0100 are given.

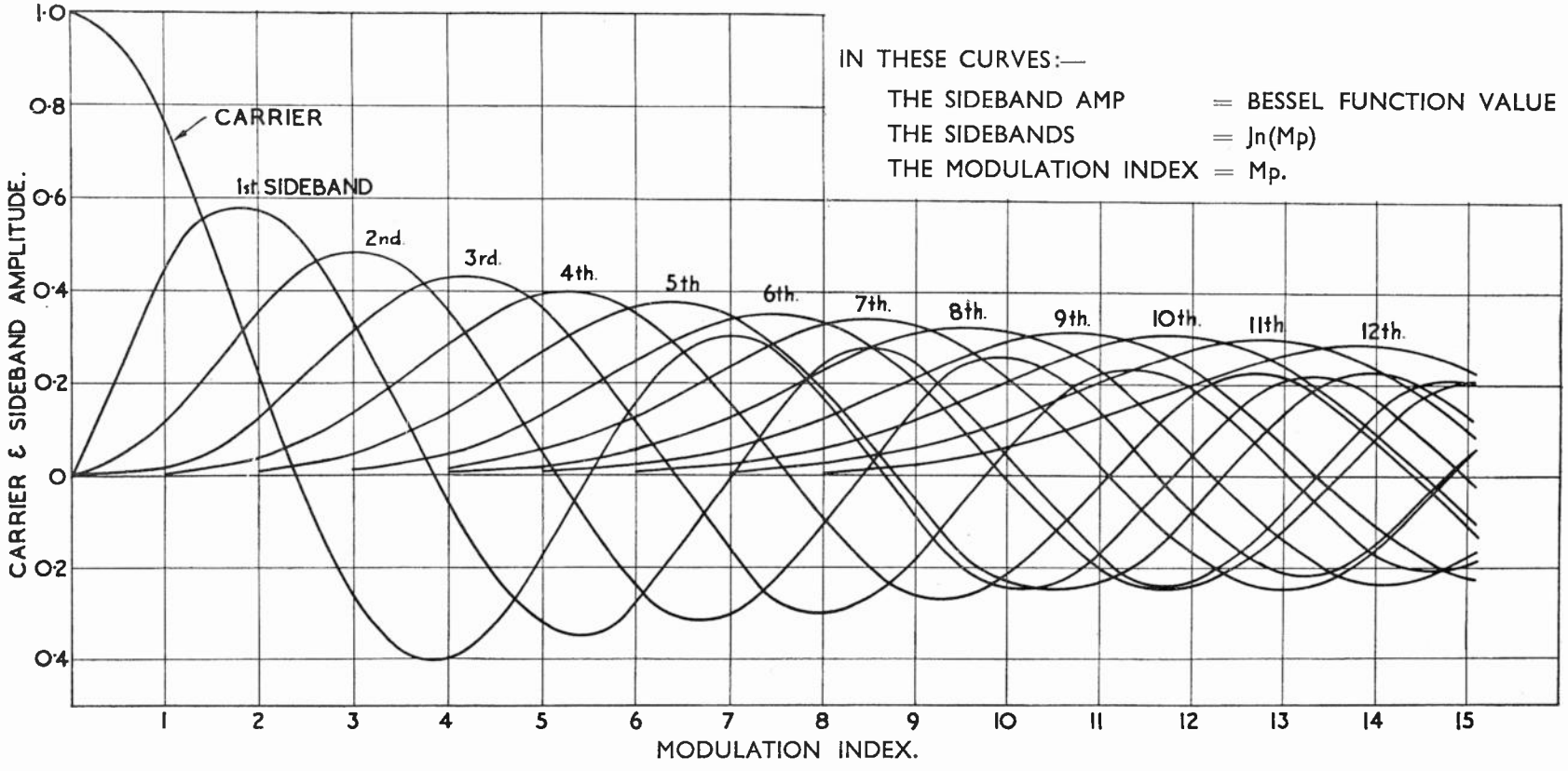



FIG. 2.9.—Curves showing the variation of carrier and side band amplitude with modulation index.

*(By courtesy of the British Institute of Radio Engineers.)*

[*Facing p.* 24

To use them it is only necessary to note that the columns indicated as $J_n(1)$, $J_n(2)$, etc., represent definite modulation indices, and that the values of $n$ given in the first column represent the carrier (0), and the various side bands (1·0, 2·0, etc.). By running the finger along the line of figures against the value $n=0$ it is therefore possible to read off the relative carrier amplitudes for increasing modulation indices. The amplitude of the various side bands may be read off in the same way. As only those side bands with an amplitude greater than 1 per cent of the unmodulated carrier amplitude need to be considered in practice, only these side band values have been included.

If it is ever necessary to determine the side band amplitudes for modulation indices which are not whole numbers, it will be found very convenient to present the Bessel function values in a more directly useful form. The curves shown in Fig. 2.9 are drawn to show the Bessel function values for all modulation indices up to 15. With their aid it is possible to read off directly the amplitude of the carrier and significant side bands. It will be noted that both in these curves and in the tables some of the side bands appear as negative quantities. If two side bands have Bessel function values of opposite sign this indicates that their vectors have an opposite polarity. This difference in polarity need only be taken into account when vectors are being added; for all practical purposes it may be disregarded.

An example of a typical side band spectrum, as determined from the Bessel function tables, is shown in Fig. 2.10. This figure also shows an equivalent Bessel function curve which has been drawn for the one fixed modulation index. Although it is of interest to note that the function values can be presented in this form, a curve of this type is of limited practical value owing to its restricted field of application.

It will be seen from Tables 1 and 2 that the number of significant side bands (i.e. those with amplitudes greater than one per cent) increase with the modulation index. The modulation index is directly proportional to the modulating signal amplitude and inversely proportional to its frequency; thus with a constant amplitude modulating signal, the number of significant side bands decreases as the modulating signal frequency increases. The bandwidth occupied by the significant side bands is equal to twice the frequency of the highest side band, which is given by the number

of significant side bands multiplied by modulating signal frequency. Thus, as the number of significant side bands falls with increasing frequency, the band-width tends to remain constant. This is a very important property of frequency modulation; by comparison,

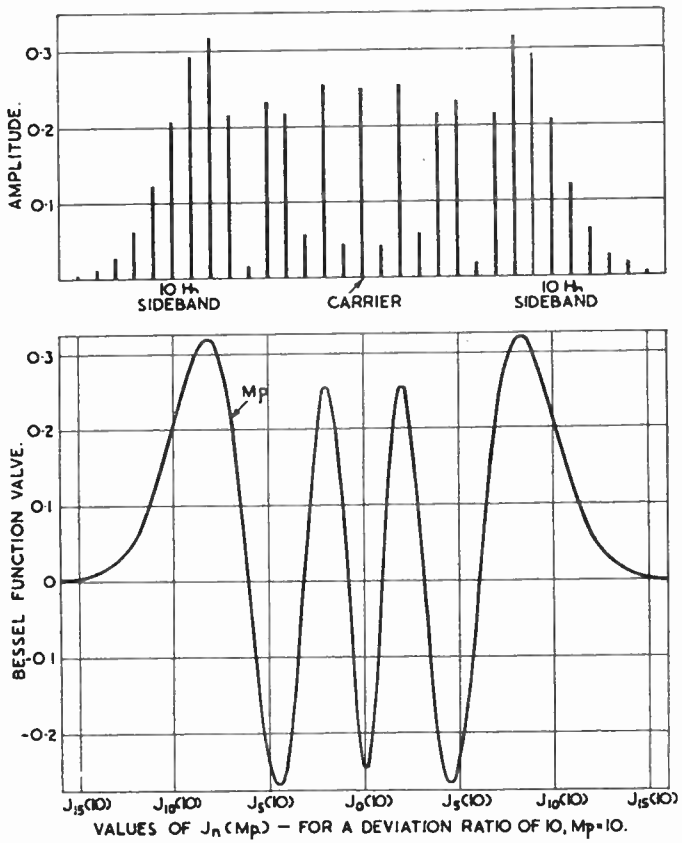



FIG. 2.10.—The side band spectrum distribution for a deviation ratio of 10. Also the equivalent Bessel function curve for a fixed value of $m_p = 10$.
*(By courtesy of the British Institute of Radio Engineers.)*

the number of significant side bands in phase modulation is independent of the modulating signal frequency and the band-width occupied by the side bands increases linearly with increasing frequency for a constant amplitude input signal.

The above discussion has assumed the presence of a single modulation frequency only. For a full understanding of the

subject, it is advisable to consider the side band distribution when the modulating signal comprises two cosinusoidal components. The expression for the modulated wave then becomes

$$e=A\cos\left(\omega_c t+\frac{m_1 f_d}{f_{a1}}\sin\omega_{a1}t+\frac{m_2 f_d}{f_{a2}}\sin\omega_{a2}t\right),$$

where $m_1$ and $m_2$ are the ratio of the amplitude of the two components of the modulating signal to the amplitude of a single signal necessary to produce the frequency deviation, and $\omega_{a1}$ and $\omega_{a2}$ are the angular velocities of the two components of the modulating signal. Of necessity, $m_1+m_2<1$, or the transmitter will be overmodulated at the instants when the two signals are in phase.

By a process similar to that employed with a single frequency modulating signal, the side band spectrum can be found. The manipulation is rather lengthy, and it will suffice here merely to quote the result. Side bands exist at frequencies removed from the carrier frequency by multiples of the individual component frequencies, as would be expected, and additionally at all frequencies of the form $n\omega_{a1}\pm m\omega_{a2}$, where $n$ and $m$ are integral. Expressed more simply, the side bands produced are the same as those which would result if each of the side bands and the resultant carrier produced by one modulating signal were modulated as a carrier by the other modulating signal. Thus, if one signal produced $p$ significant side bands when impressed alone (counting the carrier as one side band) and the other $q$ (also counting the carrier as one side band), the total resultant number of side bands would be $pq$. The amplitude of any side band, $n\omega_{a1}\pm m\omega_{a2}$ is given by $J_n(m_{p1}).J_m(m_{p2})$, where $m_{p1}$ and $m_{p2}$ are the modulation indices for the two signal components. The carrier amplitude is given by $n=m=0$, i.e. $J_0(m_{p1}).J_0(m_{p2})$. Not all of the total number of side bands $pq$ will be of significant amplitude. Where for example $J_n(m_{p1})$ is only just large enough to be considered significant almost all of the side bands of which this term forms one component of the amplitude will be below significant value. It will be seen from the symmetry of the expressions quoted, that it is immaterial which component of the modulating signal is considered initially applied, to give the side bands which form the "sub-carriers" for the other signal. It will be appreciated that in the presence of a complex modulating signal, the side band distribution becomes very complex indeed.

## Frequency Modulation Side Band Vectors

In the solution of a number of practical problems, it will be found necessary to visualise the way in which the side bands combine with the carrier to produce the frequency modulated wave. For the purpose of the present discussion, the carrier will be considered in an arrested condition. This most convenient state of affairs may be reached if the reader visualises that the carrier is actually rotating on the page at its normal angular velocity $\omega_c$ and that the whole book is rotating in the opposite direction with the same velocity. Under these conditions, the

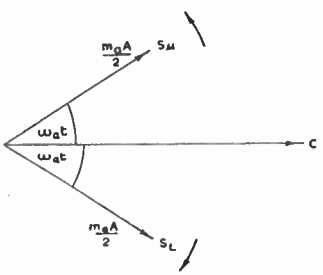



Fig. 2.11—Positions of side band vectors in amplitude modulated wave; carrier vector shown in arrested condition.

carrier vector will appear stationary, whilst the upper and lower side bands will be rotating in anti-clockwise and clockwise directions respectively. The procedure we shall adopt will be to determine the resultant vector of the summation of the side bands and carrier, $R$; the instantaneous wave magnitude is then given by

$$e = R \cos \omega_c t.$$

In order to clarify the procedure, consider first an amplitude modulated wave, modulated by a single frequency component (angular velocity $\omega_a$) to a modulation depth $m$. For convenience, we shall assume that the carrier vector magnitude is unity. The positions of the three vectors, upper side band, $S_u$, carrier $C$ and lower side band $S_l$ at any instant are shown in Fig. 2.11. The respective magnitudes are derived from expression (2.8). The resultant of the addition of the upper and lower side bands is obviously in line with the carrier vector, and equal in magnitude

to $m \cos \omega_a t$. The resultant of the addition of all three vectors is $(1+m \cos \omega_a t)$, and hence the instantaneous wave magnitude is

$$e=(1+m \cos \omega_a t) \cos \omega_c t.$$

Consider now a frequency modulated wave; modulated by a single frequency signal (angular velocity $\omega_a$), of modulation index $m_p$. The side bands are as given in expression (2.29). Consider firstly the two side bands having angular velocities $(\omega_c+\omega_a)$ and $(\omega_c-\omega_a)$. These would appear to correspond to the upper and lower side bands in the case of an amplitude modulated wave.

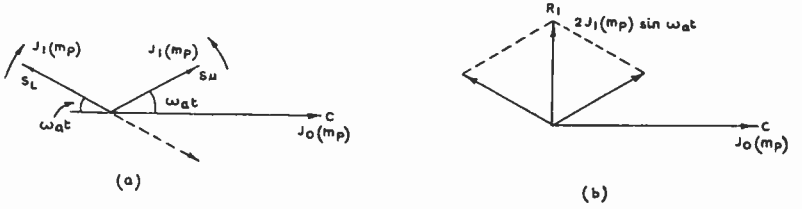



FIG. 2.12.—Positions of side band vectors of angular velocities $\omega_c \pm \omega_a$ in frequency modulated wave; carrier vector shown in arrested condition.

There is, however, one important difference. The sign of the lower side band is negative, and this indicates that the direction of its vector must be reversed. This is shown in Fig. 2.12($a$). In this figure the amplitude of the unmodulated carrier vector is taken as unity, so that the carrier vector magnitude is $J_0(m_p)$. Here the resultant $R_1$ of the upper and lower side bands is at right angles to the carrier vector and equal to $2J_1(m_p) \sin \omega_a t$ as shown in Fig. 2.12($b$). Hence the resultant of the addition of the carrier and $R_1$ is a vector of varying magnitude and phase angle relative to the unmodulated carrier vector. The resultant wave vector magnitude is given by $R=\sqrt{J_0(m_p)^2+4J_1(m_p)^2 \sin^2 \omega_a t}$, and the phase angle $\phi$ by $\tan \phi = \dfrac{2J_1(m_p) \sin \omega_a t}{J_0(m_p)}$. Since the vector resulting from the addition of the two side band vectors is at right angles to the carrier vector, it must be considered to act upon not $\cos \omega_c t$ but $\cos (\omega_c t+\pi/2)$ i.e. $-\sin \omega_c t$. This leads to the expression for the resultant vector

$$e=J_0(m_p) \cos \omega_c t-2J_1(m_p) \sin \omega_a t \sin \omega_c t,$$

which agrees with expression (2.28) derived earlier.

If now the next pair of side bands, having angular velocities $(\omega_c+2\omega_a)$ and $(\omega_c-2\omega_a)$, are considered, these are of precisely the same form as the side bands of an amplitude modulated carrier, and, therefore, their resultant vector is in line with the carrier vector, and its magnitude is given $2J_2(m_p)\cos 2\omega_a t$. The resultant of the addition of this latter resultant and the carrier vector is

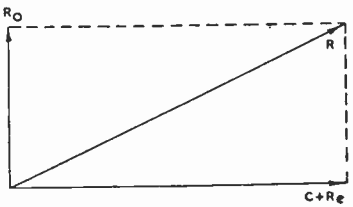



FIG. 2.13—Showing how the resultant $R_0$ off all side bands at *odd* multiples of $\omega_a$ from the carrier, and $R_e$ the resultant of all side bands at even multiples of $\omega_a$ from the carrier, are related to the carrier vector position.

$J_0(m_p)+2J_2(m_p)\cos 2\omega_a t$. The resultant of the addition of the three components is thus of magnitude

$$\sqrt{[J_0(m_p)+2J_2(m_p)\cos\omega_a t]^2+4J_1(m_p)^2\sin^2\omega_a t}$$

and of phase angle $\phi$ given by

$$\tan\phi=\frac{2J_1(m_p)\sin\omega_a t}{J_0(m_p)+2J_2(m_p)\cos\omega_a t}.$$

Generalising, the resultant $R_0$ of all side bands spaced at *odd* multiples of $\omega_a$ from the carrier is at right angles to the carrier vector; the resultant $R_e$ of all side bands spaced at even multiples of $\omega_a$ from the carrier is in line with the carrier vector $C$. This is shown in Fig. 2.13.

If all the side bands are considered, the resultant vector $R$ is of constant magnitude, since there is no amplitude modulation, and, therefore,

$$(C+R_e)^2+R_0^2=1,$$

also

$$\tan\phi=\frac{R_0}{C+R_e}. \qquad \ldots\ldots (2.31)$$

Thus, the actual phase deviation of the carrier is the result of the presence of the *odd* numbered side bands.

If the modulation index, due to a modulating signal, is small (less than 0·2 approximately) $J_0(m_p) \simeq 1$, and $J_1(m_p) \simeq \frac{m_p}{2}$, $J_2(m_p) \simeq J_3(m_p) \simeq$, etc.$=0$. In this case the vector diagram reduces to a unit amplitude carrier vector and a resultant vector at right angles to it due to the side bands $(\omega_c+\omega_a)$ and $(\omega_c-\omega_a)$ of magnitude $m_p \sin \omega_a t$. From expression (2.31),

$$\tan \phi = m_p \sin \omega_a t,$$

and since $m_p$ is small, $\tan \phi \simeq \phi$, whence

$$\phi = m_p \sin \omega_a t,$$

differentiating this to obtain $\frac{d\phi}{dt} = 2\pi f_1(t)$, where $f_1(t)$ is the frequency shift,

$$2\pi f_1(t) = m_p \omega_a \sin \omega_a t,$$

and since $m_p = mf_a/f_a$

$$f_1(t) = mf_a \cos \omega_a t,$$

as would be expected.

**Band-width Occupied by the Significant Side Bands**

Before proceeding to a discussion of the band-width required for transmission, it is necessary to introduce the deviation ratio. This is the particular value of the modulation index for $m=1$, and $f_a$ at the highest value of modulating signal to be transmitted, i.e. it is equal to $f_d/f_{a\ max}$; it is equal to the peak phase shift (in radians) occurring for the signal conditions specified. The deviation ratio is selected in the course of the design of a frequency modulation system; its chief importance lies in the fact that it determines the band-width required for the transmission of the significant side bands, i.e. those of amplitude greater than 1 per cent of the unmodulated carrier amplitude.

The curve given in Fig. 2.14 indicates the band-width occupied by the significant side bands, related to the frequency swing, modulation index and modulating signal frequency. It will be seen that, for a given value of frequency swing the band-width is proportional to the modulating signal frequency. Similarly, for a

given value of modulating signal frequency, the band-width increases with the frequency swing. Hence, for a given system, the widest band-width is required when the modulation index is equal to the deviation ratio, at the highest modulation frequency. With the aid of Fig. 2.14 it is possible to assess the band which will be required to pass them, with any given deviation ratio. As an example, the side bands of a system with a 3 kc/s maximum audio frequency and a deviation ratio of 5, would occupy a band-width of some $3{\cdot}4 \times 15$ kc/s$=51$ kc/s, as against the 30 kc/s ($\pm 15$ kc/s) over which the carrier frequency actually deviates. Similarly, for a maximum audio frequency of 15 kc/s and a deviation ratio of 5, the band occupied is 240 kc/s, as compared with the 150 kc/s band over which the carrier frequency deviates.

As the deviation ratio is increased it will be noted that the band-width occupied by the significant side bands drops towards that over which the carrier frequency actually deviates. It may therefore be deduced that the small deviation ratios are relatively more extravagant in band-width occupied than are the larger ratios.

Having established that the normal commercial deviation ratio of 5 results in significant side bands extending some 70 per cent beyond the actual frequency deviation, the question directly arising is whether or not it is necessary to pass the whole of this band through the various circuits in the system.

In practice there are a number of factors which influence the position. It is shown in Fig. 2.14 that the proportion of side band coverage to carrier frequency swing falls as the modulation index is increased. In practice this will mean that if a $\pm 75$-kc/s carrier swing is produced by an audio signal of 75 c/s (i.e. a modulation index of 1,000), the band occupied by the side bands is for all practical purposes the same as that covered by the actual carrier frequency peak-to-peak swing.

Passing to the other end of the audio frequency scale, let it be assumed that the $\pm 75$ kc/s swing is produced by an audio signal frequency of 15 kc/s. The modulation index is now only 5. If some of the side bands on the outer margin are suppressed by perhaps an over-selective amplifier, this will have the effect of producing harmonic distortion. Owing to the larger frequency coverage resulting from the higher modulating signals, these

signals will be distorted before those of lower frequency are effected.

As 15 kc/s is considered to be the highest frequency which is normally audible, its harmonics must therefore be inaudible. From this it follows that the distortion resulting from side band "clipping" may be ignored provided that the harmonics of the audio signal fall beyond the audible band. Further, it may be stated that if the harmonics of the lowest audio frequency to be

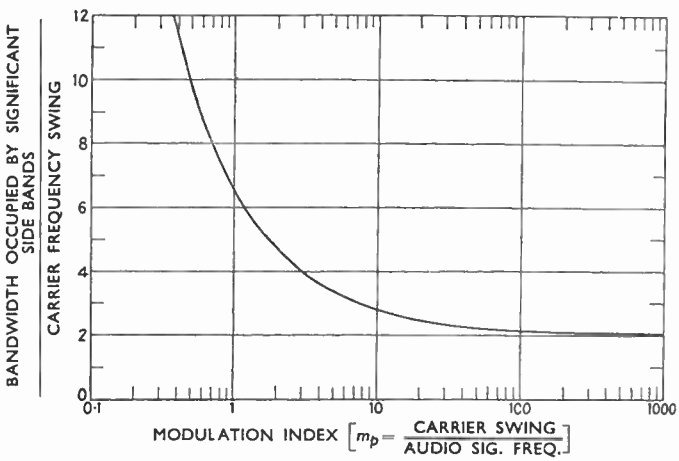



FIG. 2.14.—This curve shows the band occupied by the significant side bands (i.e. those greater than one per cent of the unmodulated carrier amplitude), plotted against varying modulation index. (See also Fig. 4.4.)

so distorted also fall outside the audible band, then the listener will remain unaware of any harmonic distortion. From this it is apparent that no harmonic distortion will be audible, providing that all the side bands associated with the signal whose second harmonic frequency is equal to that of the highest audible frequency, are amplified within the receiver passband available. There will, of course, be a small amount of amplitude distortion, but as this only occurs at maximum outputs and on the highest audio frequencies, it may be ignored. If the case of a system with a $\pm$ 75 kc/s deviation and a 15 kc/s maximum audio signal is again considered, the minimum frequency band which must be passed in order to avoid harmonic distortion may safely be reduced from 240 kc/s to some $75 \times 2 \cdot 8 = 210$ kc/s.

In practice it is very doubtful whether even this passband would be required. The above considerations have assumed a single frequency modulating wave, whereas in practice there will be a complex multi-tone signal. In such a complex signal no one frequency component can be allowed to produce the maximum permissible carrier frequency swing; otherwise the addition of the other frequency components would result in over modulation. It may, therefore, be stated that the greater the number of frequency components present in a modulating signal the smaller must be the average amplitude of each individual component signal.

Crosby has examined this situation mathematically and shown that the greater the number of modulating frequencies present, the more closely will the band occupied by the side bands approach that over which the carrier frequency swings. This is especially true for programme material where the lower modulating frequencies have the largest amplitudes.

It is very difficult to lay down a precise value for the minimum receiver passband which should be allowed. It will be shown later that for the best signal/noise ratio, the receiver passband should be as small as possible. With a frequency deviation of 75 kc/s, the lower limit is obviously 150 kc/s; a passband of between 160 and 180 kc/s would appear desirable. The matter is, however, further complicated by considerations of local oscillator stability, and the desirability of allowing some latitude for receiver mistuning. While this rule may be used as a general guide to the passband required at the receiver, the transmitter circuits should be capable of passing all significant side bands due to the maximum swing at the highest modulating frequency.

## SELECTED REFERENCES

CARSON, J. R., Notes on the Theory of Modulation, *Proc. I.R.E.*, February 1922.

VAN DER POL, Frequency Modulation, *Proc. I.R.E.*, July 1930.

HANS RODER, Amplitude, Phase and Frequency Modulation, *Proc. I.R.E.*, December 1931.

CROSBY, M. G., Carrier and Side-Frequency Relations with Multi-Tone Frequency or Phase Modulation, *R.C.A. Review*, July 1938.

HOWE, G. W. O., Frequency or Phase Modulation? *Wireless Engineer*, November 1939.

EVERITT, W. L., A Clarification and Comparison of the Characteristics of Amplitude and Frequency Modulation, *Proc. A.I.E.E.*, November 1940.

KEALL, O. E., Interference in Relation to Amplitude, Phase and Frequency Modulation Systems, *Wireless Engineer*, January 1941.

ROBINSON, JAMES, Aspects of Modulation Systems, *J. Brit. I.R.E.*, September 1942.

BELL, D. A., Frequency Modulation Communication Systems, *Wireless Engineer*, May 1943.

MCLACHLAN, N. W., *Bessel Functions for Engineers*, Oxford University Press.

WARREN, A. G., Mathematics Applied to Electrical Engineering (*Bessel Functions*, pp. 231–64). Chapman and Hall, London.

BLOCK, A., Modulation Theory, *Journal I.E.E.*, 1944, Vol. 91, Part III, p. 31.

VAN DER POL, The Fundamental Principles of Frequency Modulation, *Journal I.E.E.*, 1946, Vol. 93, Part III, p. 153.

*Chapter Three*

# INTERFERENCE AND NOISE STRUCTURE

THE principal advantage which frequency modulation shows over amplitude modulation lies in the greatly reduced interference level which results from its use. This chapter is devoted to an investigation of the way in which interference effects a carrier, and to the study of the characteristics of the principal types of noise. Although at first sight interference and noise may seem to be one and the same thing, this is not necessarily the case. Interference may be defined as any signal, other than that to which it is intended that the receiver should respond. It normally arises from three main sources:

1. The signals from an unwanted station.
2. Static disturbances (lightning, etc.), radiation from electrical equipment, and motor-car ignition systems.
3. Thermal agitation and valve noise produced in the early stages of the receiver.

Although interference may be produced by any or all of the above types of signal, only the latter two are normally classed as noise; being known respectively as impulsive and fluctuational. As these forms of interference each have their own characteristics and produce different effects on the wanted carrier, they will be examined in turn.

## Continuous Wave Interfering Signals

The effect which an interfering carrier has on the desired carrier will be considered in this section, which will also serve to establish the fundamental relationship existing between the amplitude and angular modulation components resulting from this type of interference. In order to simplify the discussion as far as possible it will be assumed that both interfering and wanted carriers are unmodulated. As the interfering signal always results in some audible or visual output, it may be regarded as causing an interfering modulation to the wanted carrier. As a first step it will be necessary to develop expressions for the amplitude and phase

modulation components of this interference modulation. Once having established the phase modulation component, the equivalent frequency modulation component, being its first differential, may be directly derived.

If the wanted carrier wave-form is represented as $A \sin \omega t$ and the interfering signal as $B \sin \omega_1 t$, then the combined resultant wave-form may be expressed as the sum of the two signals:

$$e = A \sin \omega t + B \sin \omega_1 t. \quad . \quad . \quad . \quad (3.1)$$

This is shown vectorially in Fig. 3.1, where $R$ is the resultant vector.

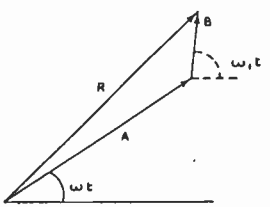



FIG. 3.1.—Vector addition of wanted and interfering signals.

As the interfering signal may be regarded as a single side band associated with the wanted carrier, the above equation may be rewritten in terms of the wanted signal variations only:

$$e = A\{\sin \omega t + x \sin (\omega + 2\pi b)t\}, \quad . \quad . \quad (3.2)$$

where $b$ = the frequency difference between the wanted and interfering signals (i.e. the beat frequency);

and $x$ = the ratio of the interfering signal amplitude $B$ to the wanted signal amplitude $A$.

For clarity $A$ is omitted from the next few steps. Equation (3.2) may be expanded to give terms for the amplitude and angular interference modulation components.

$$e = \sin \omega t + x \sin \omega t \cos 2\pi bt + x \cos \omega t \sin 2\pi bt \quad . \quad (3.3)$$

$$= \sin \omega t(1 + x \cos 2\pi bt) + \cos \omega t(x \sin 2\pi bt), \quad . \quad (3.4)$$

or for simplicity:

$$e = \sin \omega t(P) + \cos \omega t(Q)$$

$$= (P^2 + Q^2)^{\frac{1}{2}} \sin (\omega t + \phi), \quad . \quad . \quad . \quad (3.5)$$

where

$$\phi = \tan^{-1} \frac{Q}{P} = \tan^{-1} \frac{x \sin 2\pi bt}{1 + x \cos 2\pi bt}.$$

Expanding equation (3.5),

$$e=(1+x^2+2x\cos 2\pi bt)^{\frac{1}{2}}\sin(\omega t+\phi). \quad . \quad . \quad (3.6)$$

Reintroducing the amplitude of the wanted signal $A$, the modulation resulting from the interfering signal is as follows:

$$e=A(1+x^2+2x\cos 2\pi bt)^{\frac{1}{2}}\sin(\omega t+\phi). \quad . \quad (3.7)$$

This result can, of course, be derived directly from the vector diagram.

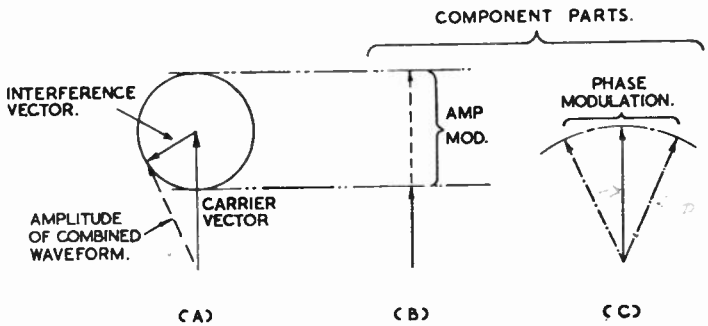



FIG. 3.2.—Shows in (*A*) the arrested carrier vector with superimposed interfering carrier vector. Diagrams (*B*) and (*C*) show respectively the interfering amplitude and phase modulation components which result.
(*By courtesy of the British Institute of Radio Engineers.*)

If reference is made to expressions (2.4) and (2.5) (the equations for the basic modulation forms), it will be apparent that the wanted carrier $A$ is subjected to an interfering amplitude modulation to the extent indicated by the term $(1+x^2+2x\cos 2\pi bt)^{\frac{1}{2}}$ and is angular or phase modulated by an amount equal to:

$$\phi=\tan^{-1}\frac{x\sin 2\pi bt}{1+x\cos 2\pi bt}.$$

The way in which these two modulation components are produced is illustrated in the vector diagrams shown in Fig. 3.2. Each of the modulation components will now be considered separately.

### Equivalent Amplitude Modulation

The detector in an amplitude modulation receiver is only able to respond to changes in carrier amplitude; carrier phase changes having no effect on its output wave-form. In the case of an

amplitude modulation receiver it is therefore only necessary to consider the amplitude modulation component which results from an interfering signal. If this component $(1+x^2+2x \cos bt)^{\frac{1}{2}}$ is expanded it may be shown to consist of a spectrum comprising the original difference frequency beat note and its harmonics.

The expansion of $(1+x^2+2x \cos 2\pi bt)^{\frac{1}{2}}$ is

$$\sqrt{1+x^2}\left[1-\frac{X^2}{4}-\frac{15}{64}X^4-\ldots\right.$$
$$+\left(X+\frac{3X^3}{8}+\cdots\right)\cos 2\pi bt$$
$$-\left(\frac{X^2}{4}+\frac{15X^4}{16}+\cdots\right)\cos 4\pi bt$$
$$+\left(\frac{X^3}{8}+\cdots\right)\cos 6\pi bt$$
$$-\left(\frac{5}{64}X^4+\cdots\right)\cos 8\pi bt$$
$$\left.\cdots\right], \quad \ldots\ldots\ldots \quad (3.8)$$

where $X=\dfrac{x}{1+x^2}$ ($x$=the ratio of interfering to wanted signals);

$b$=the difference frequency between the two signals.

When the amplitude ratio between the interfering and wanted signals is less than 0·5 this equation may be written with sufficient accuracy as

$$\sqrt{1+x^2}\left(1+X\cos 2\pi bt-\frac{X^2}{4}\cos 4\pi bt+\frac{X^3}{8}\cos 6\pi bt-\ldots\right). \quad (3.9)$$

The terms $\sqrt{1+x^2}\,X$, $\sqrt{1+x^2}\dfrac{X^2}{4}$, $\sqrt{1+x^2}\dfrac{X^3}{8}$, etc., represent the modulation depth due to the interference beat frequency and its harmonics referred to the amplitude of the unmodulated wanted carrier. It should be noted that the equivalent modulation depths, given by $X, \dfrac{X^2}{4}, \dfrac{X^3}{8}$ are referred to a carrier of amplitude $\sqrt{1+x^2}$ greater than the original carrier. The presence of the interfering signal increases the wanted carrier amplitude to this apparent value, since the modulating wave-form has a peaky character, so

that its average value is not equal to the average of the positive and negative going peak values. In order to illustrate this point, reference is made to Fig. 3.4, which shows the extreme case when the wanted and interfering signals are of equal amplitude. Under these conditions the resultant modulation envelope is not sinusoidal, but has the peaked form shown. An examination of the conditions which exist in the regions where the vectors of the two waves add together, and that in which they cancel out (Fig. 3.3), reveals the reason for this wave-shape. In the zone in which the vectors

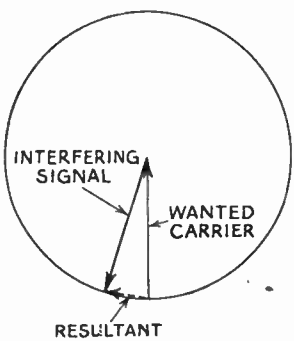



FIG. 3.3.—The shape of the interfering amplitude modulation which is produced when the amplitude of the interfering signal approaches that of the wanted carrier.

are adding there is a period during which the amplitude of the resultant wave increases slowly, remains constant, and then slowly begins to fall. Conditions in the region in which the two vectors tend to cancel each other out are, however, very different. Here the resultant gets rapidly smaller right up to the instant at which the two carriers cancel out. Instantly at this point the amplitude of the resultant wave begins to again increase at the same rate as it was previously falling.

This peaked type of wave-form is very pronounced when the wanted and interfering carriers are approaching the same amplitude. As soon as one signal becomes smaller than the other the modulation gradually approaches a pure sine wave-form.

The average value of the modulation envelope is shown in Fig. 3.4 as a dotted line. Since $x=1$, the apparent carrier amplitude is increased by a factor of $\sqrt{2}\left(1-\frac{1}{16}-\frac{15}{64}\frac{1}{16}\right)\simeq 1{\cdot}32$ over

its unmodulated value. The modulation depth of the harmonic components of the modulation envelope, referred to the amplitude of the unmodulated carrier, is shown in Fig. 3.5. These curves, which have been drawn up to the third harmonic, were calculated from expression (3.9).

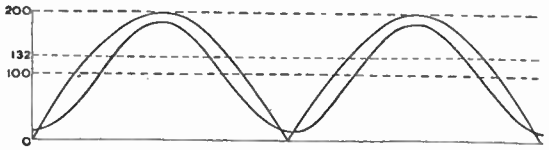



FIG. 3.4.—The shape of the carrier envelope obtained when the amplitude of the interfering signal equals that of the wanted carrier. The middle dotted line shows the mean value of the carrier envelope.

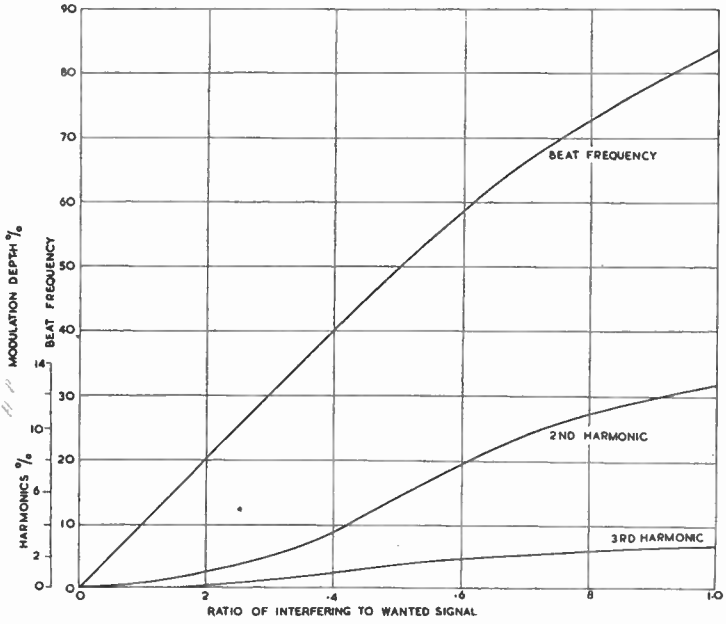



FIG. 3.5.—These curves analyse the interference amplitude modulation produced by an unwanted signal to a wanted carrier. The beat-note frequency is the difference between the interfering and wanted signals, the harmonics are those of the difference frequency.

**Equivalent Phase Modulation**

A receiver designed for the reception of any angular modulated transmission (either phase or frequency modulation) should not be responsive to changes in the carrier amplitude. The only interfering modulation component which need be considered is, therefore, that of the resultant phase modulation. It has earlier been shown that this component, when arising from the interference produced by one unmodulated wave acting upon another, may be expressed in the following terms:

$$\phi = \tan^{-1} \frac{x \sin 2\pi bt}{1 + x \cos 2\pi bt}.$$

Keall has shown that this term may be expanded into the form of the following series:

$$\phi = x \sin 2\pi bt - \frac{x^2}{2} \sin 4\pi bt + \frac{x^3}{3} \sin 6\pi bt - \ldots . \qquad (3.10)$$

This analysis shows that the interfering phase modulation component is built up of a series of waves having frequencies equal to $b$, the beat frequency between the interfering and wanted signals, and its harmonies ($2d$, $3b$, etc.). The magnitude of these component waves is determined by the terms $x, \frac{x^2}{2}, \frac{x^3}{3}$, etc.

A number of single cycles of the phase modulation component of the resultant interference have been drawn in Fig. 3.6. These waves are actually a graphical presentation of equation (3.10) for different ratios of interfering to wanted signal amplitude (i.e. for different values of $x$). The curve for a 1 : 1 ratio is only of theoretical interest as the limiter will have ceased to function before this point is reached. Once the interfering signal has exceeded the wanted signal it is necessary to change one's viewpoint and consider the interference produced by the wanted signal to the interfering signal. Reference to the inserted vector diagram makes clear the reason for the non-sinusoidal phase variation which is produced when the interference signal approaches the wanted carrier amplitude. When the wanted and unwanted carrier vectors are adding together, the resultant phase modulation will be changing relatively slowly, whereas when they are tending to cancel each other out the resultant vector will swing very rapidly through up to 180°.

In order to compare the interfering phase and amplitude modulation components it is necessary to express the series of waves, from which the curves in Fig. 3.6 are derived, in terms of the equivalent modulation depth. This may be done by adding a term $\phi_0$ to indicate the maximum phase displacement in radians employed in the transmission system concerned.

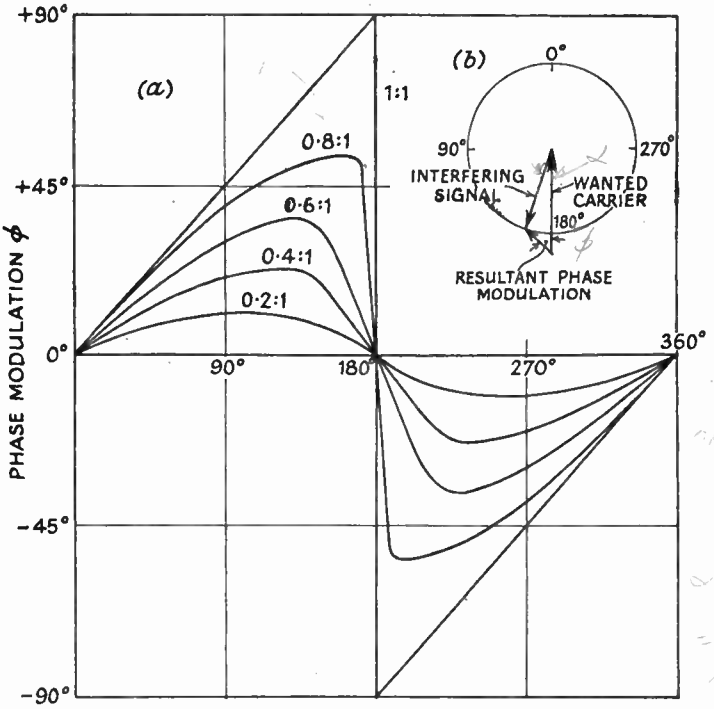



FIG. 3.6.—The above group of curves illustrate a series of single cycles of interfering phase modulation produced by different ratios of interfering to wanted signal amplitude. Diagram (*b*) shows the vector diagram for a 0·8 : 1 ratio.

The equivalent modulation depth of the interfering phase modulation components may therefore be expressed as

$$\frac{x}{\phi_0}\sin 2\pi bt - \frac{x^2}{2\phi_0}\sin 4\pi bt + \frac{x^3}{3\phi_0}\sin 6\pi bt \ldots .$$

In order to express the magnitude of the modulation terms $\left(\frac{x}{\phi_0}, \frac{x^2}{2\phi_0}, \frac{x^2}{3\phi_0}, \text{etc.}\right)$ as a percentage, $\phi_0$ can be assigned a value

of 1 and each term multiplied by 100. They then express directly the equivalent phase modulation depths of the beat frequency ($b$) and its harmonics referred to a maximum phase shift of one radian. In order to complete the comparison of the amplitude and phase modulation interference components a set of equivalent phase modulation depth curves is given in Fig. 3.7, with the maximum phase deviation taken as 1 radian.

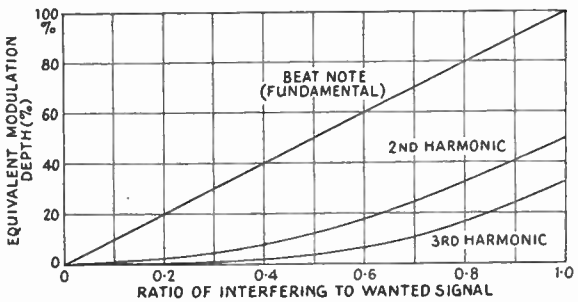



FIG. 3.7.—These curves show the equivalent phase modulation components produced at varying ratios of interfering to wanted signal amplitude. One hundred per cent phase modulation is assumed to be a deviation of 1 radian.

The band-width occupied by the significant side bands of a signal with the maximum value of phase shift of 1 radian can be found from Table 1 of Chapter Two, by using the fact that $m_p$, occurring in the table, is equal to the phase swing of the signal. At the maximum value of phase swing, $m_p=1$, and it will be seen that the significant side bands extend up to the third order. By comparison of the terms involved in both amplitude and phase modulation cases, it will be seen that the magnitudes of the fundamental beat notes are almost equal in both systems, for small values of $x$, whilst the amplitude of the harmonics can be neglected, as they are vanishingly small. Thus there is no improvement obtained by employing phase modulation with a maximum phase swing of 1 radian, although an increased band-width is required. The magnitude of the second and third order side bands, are, however, small for all values of $m<1$, and hence may be suppressed, with only slight spurious amplitude modulation of the signal, and some distortion. If these higher order side bands are suppressed, the amplitude modulated and phase modulated signals require

identical band-width for transmission, and hence no significant improvement can be expected.

If a maximum phase swing, $\phi_0$, other than 1 radian is employed, then the relative magnitudes of the beat notes in the phase modulation and amplitude modulation systems is directly proportional to $\phi_0$. It can be seen from Table 2 of Chapter Two that the number of significant side bands increases relatively slowly as $\phi_0$ is increased; from 3 at $\phi_0=1$, to 4 at $\phi_0=2$, 6 at $\phi_0=3$ and 7 at $\phi_0=4$. Thus, for values of $\phi_0$ greater than unity, the improvement of a phase modulation system over an amplitude modulation one, becomes more nearly commensurate with the greater band-width required.

The above comparison is confined to small values of $x$, the ratio of the interfering to wanted signal amplitude. Inspection of the fundamental and harmonic terms will show that as $x$ increases, the terms associated with the phase modulation components increase more rapidly than those of the amplitude modulation series. This relative deterioration with increasing values of $x$ is common to all forms of angular modulation, and is discussed in detail for the frequency modulation case.

### Equivalent Frequency Modulation

In the last chapter it was shown that frequency modulation is proportional to first differential of phase modulation. This being so, we can derive the carrier frequency shift due to the presence of the interfering signal by differentiating expression (3.10). This gives

$$2\pi f_s = d\phi/dt = 2\pi bx \cos 2\pi bt - 2\pi bx^2 \cos 4\pi bt + 2\pi bx^3 \cos 6\pi bt \ldots$$

or

$$f_s = bx \cos 2\pi bt - bx^2 \cos 4\pi bt + bx^3 \cos 6\pi bt \ldots . \quad (3.11)$$

This shows that the interfering frequency modulation component is built up of a series of waves having frequencies equal to multiples of $b$, the beat frequency. The magnitude of these component waves is proportional to $bx$, $bx^2$, $bx^3$, etc.

A number of single cycles of the frequency modulation component are shown in Fig. 3.8. These correspond to the values of $x$ in Fig. 3.6. It should, however, be noted that the ordinate is now proportional to $b$. Thus, for $x=0{\cdot}8$ and $b=1$ kc/s, the maximum value of the frequency deviation is $4\times 1$ kc/s$=4$ kc/s. The curves for high values of $x$ are of theoretical interest only, as the limiter will have ceased to function.

If the vector diagram of Fig. 3.6 is again referred to, the reason for the rapid change in $f_s$ in the region of 180° phase difference between the carrier and interference vector will be apparent. As the angle approaches 180°, the resultant vector velocity will increase very rapidly and for $x=1$ will be infinite at 180°, since the resultant vector phase changes instantaneously by 360° at this point. This point is, however, of academic interest only, since the amplitude of the resultant vector goes to zero at 180°.

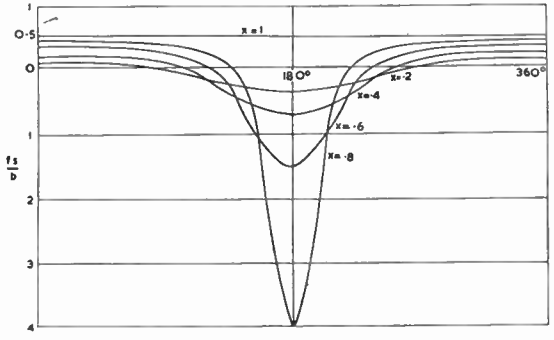



FIG. 3.8.—Showing variation of frequency swing with variation of interfering to wanted carrier amplitude.

The equivalent modulation depth of the interfering frequency modulation component may be expressed as

$$\frac{b}{f_d}\left[ x \cos 2\pi bt - x^2 \cos 4\pi bt + x^3 \cos 6\pi bt \ldots \right]. \qquad (3.12)$$

In order to express the amplitude of the modulation terms as a percentage, $f_d$ can be assigned an arbitrary value, and each term multiplied by 100. They then express directly the equivalent frequency modulation depth of the beat frequency $b$, and its harmonics. Fig. 3.9 shows this result graphically. It will be noted that the equivalent modulation depth varies directly with $b$, in distinction to the case of amplitude modulation and phase modulation when the equivalent modulation depth is independent of $b$.

The foregoing considerations have shown that both the interfering amplitude and frequency modulation components consist of a series of harmonics of somewhat similar form. In order to

complete this investigation, the total equivalent modulation depth of the two components will be plotted on the same graph.

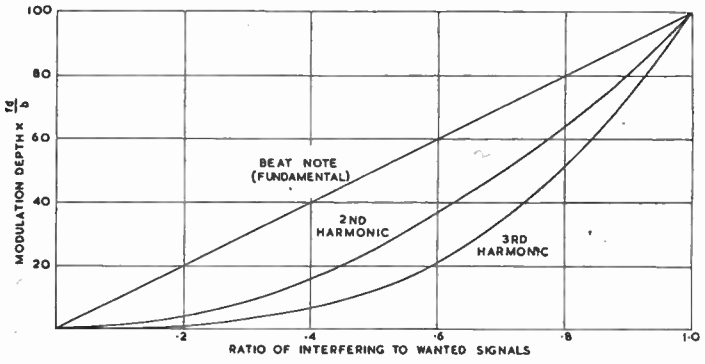



FIG. 3.9.—The curves show the interference frequency modulation produced by an unwanted signal to a wanted carrier. The beat-note frequency is the difference between the interfering and wanted signals; the harmonics are those of the frequency difference.

The total equivalent frequency modulation depth is

$$\frac{b}{f_d}\left\{x^2+x^4+x^6\ldots\right\}^{\frac{1}{2}}$$

and that for the total equivalent amplitude modulation depth is

$$\left\{(1+x^2)\left([X]^2+\left[\frac{X^2}{4}\right]^2+\left[\frac{X^3}{8}\right]^2\ldots\right)\right\}^{\frac{1}{2}} \quad \text{for } x<0{\cdot}5, \quad (3{\cdot}13)$$

where $x =$ the ratio of interfering to wanted signal amplitudes and $X=\dfrac{x^2}{1+x^2}$. Expressed differently, these expressions give the ratio of r.m.s. noise to peak signal amplitudes of the demodulated outputs.

The curves in Fig. 3.10 show these results graphically; the curves for the frequency modulated case have been plotted for a range of $b/f_d$ values. As $x$ approaches 1, the total equivalent modulation depth for frequency modulation increases rapidly, but over the usable range of interfering to wanted signal amplitudes, the frequency modulation curves are lower than the amplitude modulation curves except for high values of $b/f_d$ and $x$. A more realistic comparison perhaps is to compare the equivalent modulation depths for the fundamental beat frequency

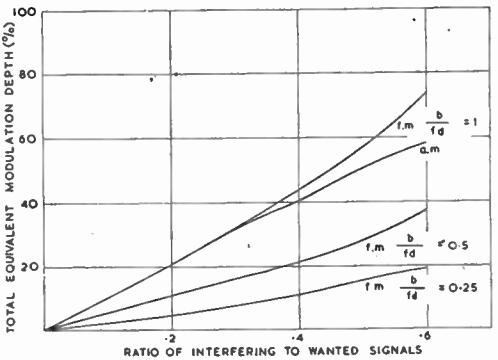



FIG. 3.10.—Showing the total equivalent interfering frequency and amplitude modulation components produced by varying ratios of interfering to wanted signal amplitudes.

terms only, because, since the value of $b$ is unrestricted, many of the harmonic components may be inaudible. For frequency modulation this reduces to

$$bx/f_a,$$

and for amplitude modulation to

$$\sqrt{1+x^2}\,X = \frac{x}{\sqrt{1+x^2}} \simeq x$$

(for small values of $x$, i.e. $<0{\cdot}5$ approximately).

The equivalent modulation depths for the range $x < 0{\cdot}5$ which allows of intelligible reception, are plotted in Fig. 3.11. It will

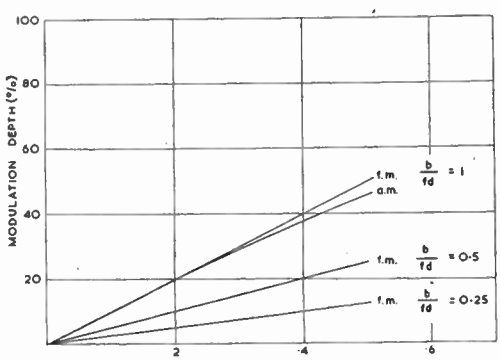



FIG. 3.11.—Showing the fundamental frequency interfering frequency and amplitude modulation components produced by varying ratios of interfering to wanted signal amplitudes.

be seen that the signal to noise ratio for frequency modulation is always better than the signal to noise ratio for amplitude modulation, except in the extreme case of $b$ approaching $f_d$. If the case of a wide band frequency modulation transmission is considered, where $f_d$=75 kc/s, and the maximum audible value of $b$ is 15 kc/s, the frequency modulation system is always at least better by a factor of 5 or 14 db. At worst, the improvement factor is given by $f_d/b_{max}$, and this is normally equal to the deviation ratio. Hence, it is apparent that a large deviation ratio is desirable. This is, however, only true so long as $x$ is small.

### Impulsive Noises

In the study of the improvements resulting from the use of frequency modulation it is necessary to distinguish between impulsive and fluctuational noise. The former is defined as noise made up of definite impulses each of short duration and occurring at relatively widely separated intervals. In practice this type of interference is almost always produced by some form of spark discharge. The most common sources are motor-car ignition systems, commutator-type electric motors, and natural atmospheric static.

Impulses of this type, if subjected to Fourier analysis, may be shown to consist of an infinite sum of sine wave components. The part of this spectrum which is amplified in any one radio receiver is a small band of high order components. Since in practice the difference between the highest and the lowest frequency in such a band is small compared with the mid-frequency, all the components received will be of substantially equal amplitude. As these individual voltages are equally spaced throughout the band being amplified, the number of voltages included in any given band is proportional to its width. This being so, the peak amplitude of the amplified impulse is proportional not only to the amplitude of the original signal, but also to the receiver's band-width.

### The Shape of an Impulsive Wave Train

The solution to the problem of the mathematical examination of discontinuous functions is due to Heaviside. He evolved an operational calculus which is today known by his name. His operational method employs a simple form of discontinuity, which

may be typified by that occurring when a switch is closed. The voltage across a load, which was previously zero, at time $t=0$ suddenly becomes $E$ and remains at this value. A discontinuity of this type is termed a Heaviside unit function, or a unit step. By employing this function as an operator in derived equations, the results produced in transmitting any type of discontinuity signal through a modifying network may be explored mathematically.

Usually the circuits in the receiver are considerably sharper than any present in the noise source. Hence, the shape of the envelope of the wave-train which emerges from the output of the receiver amplifier is usually almost independent of the wave-form of the original impulse signal—provided that the pulses are separated sufficiently in time to avoid overlapping decay trains. The shape of the envelope of the wave-train emerging at the radio receiver's output terminals will therefore be a function of the circuits of the receiver, and as such is a suitable subject for calculation on the basis of Heaviside's operational calculus or the methods of the Laplace transform. These methods can be applied to evaluate the response of a receiver to either of the two basic types of discontinuous function, the unit step or the unit impulse. The former was defined above. The latter may be defined as the limiting case of a rectangular pulse having a constant product of amplitude $\times$ time equal to unity, as $t$ approaches zero. The impulse is thus of infinite amplitude for an infinitely short duration.

These functions may be shown to be equivalent to the sum of sinusoidal components of all frequencies. The components are spaced infinitely closely together, and it is not possible to allocate an amplitude to a component at any given value of frequency $f$, but only for the contribution of all the components in the region between $f$ and $f+df$.

For periodically repeated wave-forms, the distribution of amplitude with frequency is a discontinuous function, i.e. components exist only at frequencies harmonically related to the fundamental period. In general, however, repeated noise wave-forms are of short duration, have low repetition frequencies (generally less than 1,000 per second) and are generally subject to "jitter". This being so, it is permissible to consider each noise pulse separately, and to consider it to have a continuous spectrum. It

is then possible to evaluate the distribution of amplitude with frequency for a single pulse, and to compare its value in the region of interest, i.e. the portion of the spectrum to which the receiver is tuned, with that of unit step or a unit impulse in the same region. If, then, the response of the receiver to either unit impulse or unit step is known its response to the noise pulse can be found by scaling the result appropriately. The resultant amplitude of the

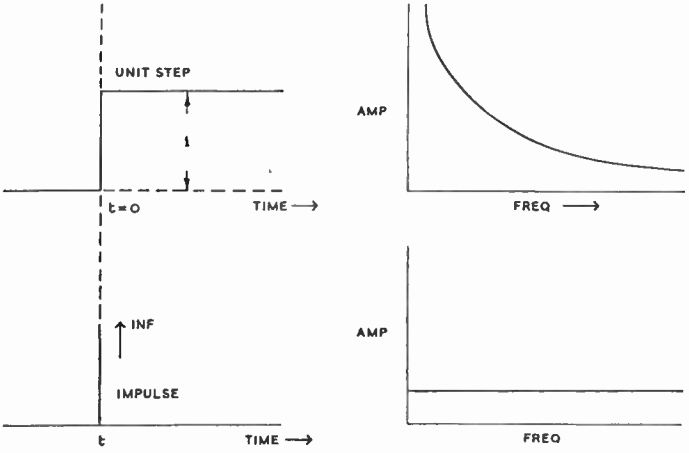



FIG. 3.12.—Wave-forms of unit step and impulse, together with amplitude/frequency distributions.

components in a region $df$ about a centre frequency $f$ for the unit step and the unit impulse is given by $\frac{df}{2\pi f}$ and $df$ respectively. The components of the unit step are all sine waves, whilst those of the unit impulse are all cosine waves. The distribution of amplitude with frequency is shown in Fig. 3.12. If a step of amplitude $E$ is considered, the distribution of amplitude with frequency is given by $\frac{E}{2\pi f}$; correspondingly, if an impulse whose product of amplitude and time $E$ is considered, the distribution of amplitude with frequency is given by $E$.

By virtue of the close relationship of the unit step and the unit impulse, it is permissible to work with either. If the unit impulse is chosen, the corresponding result for the unit step can be found by dividing the result obtained by $2\pi f_0=\omega_0$, where $f_0$ is the

frequency at the centre of the region of interest. This assumes that the band-width of the region of interest is very small compared with $f_0$.

As stated earlier the response of a receiver is dependent upon its band-width. This being so, the shape of the noise output will be determined almost entirely by the stages of the receiver having the narrowest band-width. Neglecting the audio amplifier, the effect of which we shall consider later, these stages will in general comprise the i.f. stages of the receiver.

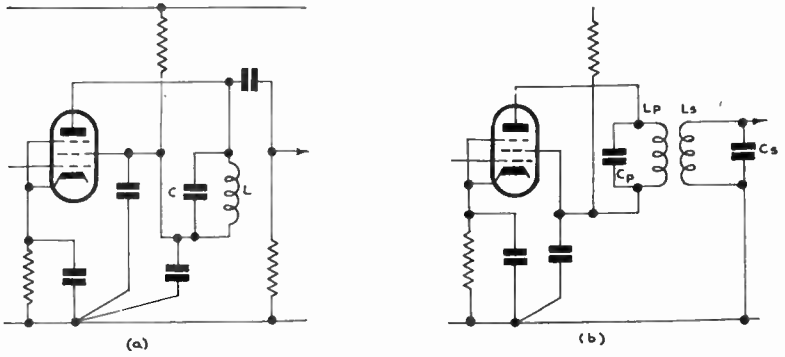



FIG. 3.13.—Receiver i.f. stages: (*a*) single tuned-circuit type, (*b*) coupled tuned circuits.

Such a stage generally comprises a pentode valve feeding a single tuned circuit or two coupled circuits, as shown in Fig. 3.13. In order to secure the band-width required for f.m. systems, it is usual for the centre frequency of the i.f. stages to be in the region 5–15 mc/s. The valve anode impedance is normally sufficiently large for its damping effect to be negligible. Similarly, at the frequencies normally encountered, the input damping of the following stage is generally sufficiently high to be ignored also.

The response of a number of stages of i.f. amplification employing single tuned circuits to an impulse at the grid of the first stage has been examined in detail by Smith and Bradley. In order to provide a fair basis of comparison, Smith and Bradley assumed that the overall band-width (i.e. the frequency separation between the points at which the steady state response is 3 db below that at the centre frequency) is maintained constant irrespective of the number of stages considered. This means, of course, that as the

number of stages considered increases, the damping of each individual circuit must be increased. Then if $R$ is r.f. resistance of the inductor $L$, and assuming the tuning capacitor to be lossless, the damping factor $\alpha$ is defined by $\alpha=\frac{R}{2L}$. The band-width for a single stage is given by

$$\text{Band-width}=\frac{f_0}{Q},$$

where $Q=\frac{L\omega_0}{R}$, and $f_0=\frac{\omega_0}{2\pi}$ is the resonant frequency of the circuit.

Designating the half band-width $\Delta f$,

$$\Delta f=\frac{f_0}{2Q}.$$

The corresponding angular velocity $\Delta\omega=\frac{\omega_0}{2Q}=2\pi\Delta f$.

Where $n$ identical stages are employed, the value of $\alpha$ must be reduced to maintain the overall band-width constant, and the relationship between $\Delta\omega$ and $\alpha$ for $n$ stages under these conditions is given by

$$\alpha=\frac{\Delta\omega}{(2^{1/n}-1)^{n/2}}. \quad . \quad . \quad . \quad . \quad (3.14)$$

Smith and Bradley show that for $n$ stages, the output response to a noise signal comprising a unit impulse applied to the grid of the first stage is given approximately by

$$N_n(t)=2\alpha(g_m R_0)^n\frac{(\alpha t)^{n-1}}{(n-1)!}e^{-\alpha t}\cos\omega_0 t, \quad . \quad . \quad (3.15)$$

where $g_m$ is the mutual conductance of each valve and $R_0$ is the dynamic resistance of each tuned circuit given by $R_0=\frac{L}{CR}=\frac{1}{2\alpha C}$. Since the damping has to be increased as the number of stages increases to maintain the band-width constant, $R_0$ will be reduced progressively as $n$ increases.

The output with a steady signal $E\cos\omega_0 t$ applied to the grid of the first stage is given by

$$e=E(g_m R_0)^n\cos\omega_0 t. \quad . \quad . \quad . \quad . \quad (3.16)$$

So that from (3.15) and (3.16), the instantaneous value of the ratio of noise output envelope to the signal output envelope $\left(\frac{N}{S}\right)$ is given by

$$\left(\frac{N}{S}\right)=\frac{2a}{E}\frac{(at)^{n-1}}{(n-1)!}e^{-at}. \quad . \quad . \quad . \quad . \quad (3.17)$$

The peak value of this ratio occurs when $(at)^{n-1}\,e^{-at}$ reaches its maximum value; this occurs when $t=\frac{n-1}{a}$. Substituting this value in (3.17), the peak value of the noise envelope to signal envelope ratio is given by

$$\left(\frac{N}{S}\right)_{max}=\frac{2a}{E}\frac{(n-1)^{n-1}e^{(1-n)}}{(n-1)!}$$

and substituting for $a$ from (3.14)

$$\left(\frac{N}{S}\right)_{max}=\frac{2}{E}\frac{\Delta\omega}{(2^{1/n}-1)^{n/2}}\frac{(n-1)^{n-1}e^{(1-n)}}{(n-1)!}. \quad . \quad (3.18)$$

This expression is practically independent of $n$ for $n>3$, showing that under these conditions, the peak value of the ratio is practically constant.

Defining $$A=\frac{(n-1)^{n-1}e^{(1-n)}}{(2^{1/n}-1)^{n/2}[(n-1)!]},$$

Smith and Bradley give the following values for $A$ for various values of $n$:

| $n$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $A$ | 1·0 | 0·57 | 0·53 | 0·51 | 0·50 | 0·50 |
| $n$ | 7 | 8 | 9 | 10 | 11 | 12 |
| $A$ | 0·50 | 0·050 | 0·049 | 0·049 | 0·049 | 0·049 |

The actual instantaneous value of the noise to signal envelope ratio of the output wave-form is given by

$$\frac{2\Delta\omega}{E}\frac{(\Delta\omega t)^{n-1}\,e^{-\Delta\omega t/(2^{1/n}-1)^{1/2}}}{(2^{1/n}-1)^{n/2}[(n-1)!]} \quad . \quad . \quad . \quad (3.19)$$

Expression (3.19) is plotted in Fig. 3.14, for $n=2$, and $n=4$, in terms of $\frac{2\Delta\omega}{E}$, i.e. $\frac{2\Delta\omega}{E}$ has been taken equal to unity. The

magnitude of any point of the envelope can be obtained, therefore, by multiplying by the value of $\frac{2\Delta\omega}{E}$.

The time axis is plotted in terms of $1/\Delta\omega$, the value of 1 corresponding to a time of $\frac{1}{2\pi} \times$ the duration of a cycle of a wave of frequency $\Delta f = \frac{\Delta\omega}{2\pi}$, i.e. the time for a vector at the half band-width frequency to rotate through 1 radian.

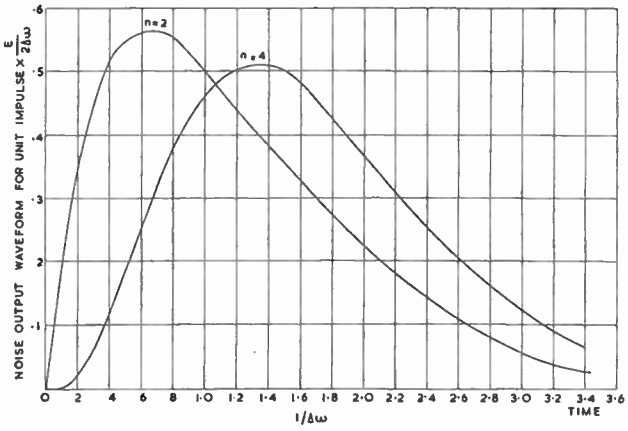



FIG. 3.14.—Noise output wave-form for i.f. amplifier employing single tuned circuits. The magnitude of the output is related to the peak value $E$ of a sinusoidal input at the centre frequency, and the i.f. half band-width $\frac{\Delta\omega}{2\pi}$.

From the expressions for the time for the peak output to occur, $t = \frac{n-1}{a}$, and $A$, it will be seen that the effect of adding successive stages to increase the time for the peak of the output noise pulse to be reached, without appreciably altering the peak noise to signal envelope ratio for $n > 3$.

We may, therefore, take the value of $A$, for $n > 3$ as 0·5 with negligible error and also for $n = 2$ with only a small error. This result is of considerable importance, since it shows that the peak noise to signal envelope ratio remains practically constant, with only a small error for two or more stages, so that it is possible

to generalise about the response of i.f. amplifiers of this type to impulsive interference. It must, however, be remembered that this result is only true provided that the overall band-width is maintained constant.

Since the peak value of the noise envelope is reached when $t=\frac{n-1}{\alpha}$, and $\alpha=\frac{\Delta\omega}{(2^{1/n}-1)^{n/2}}$, the time for the peak value to be reached is inversely proportional to $\Delta\omega$ for a given value of $n$. Also, from expression (3.18), the actual peak magnitude of the noise envelope alone is directly proportional to $\Delta\omega$.

The response of a single i.f. stage employing two coupled circuits to a unit voltage impulse at the grid of the first stage is given below. It is assumed that the two circuits are identical, and coupled by mutual inductance. The circuit parameters employed are:

$\alpha=\frac{R}{2L}$, where $R$ is the r.f. resistance of the inductor $L$;

$M=KL$;

$f_0=\frac{\omega_0}{2\pi}$ = resonant frequency of either tuned circuit alone $=\frac{1}{\sqrt{LC}}$;

$Q=\frac{L\omega_0}{R}$.

The response to the unit voltage impulse is given approximately by

$$-g_m\omega_0^2 Le^{-\alpha t}\sin\frac{K\omega_0 t}{2}\sin\omega_0 t. \quad . \quad . \quad . \quad (3.20)$$

The output with a steady signal $E\cos\omega_0 t$ applied to the grid of the stage is given by

$$E=g_m EQ\omega_0 L\,\frac{KQ}{1+K^2Q^2}\cos\omega_0 t. \quad . \quad . \quad . \quad (3.21)$$

So that from (3.20) and (3.21), the instantaneous value of the ratio of the noise envelope to the signal envelope $\left(\frac{N}{S}\right)$ is given by

$$\left(\frac{N}{S}\right)=\frac{\omega_0}{QE}\,\frac{1+K^2Q^2}{KQ}\,e^{-\alpha t}\sin\frac{K\omega_0}{2}t.$$

$$=\frac{2\alpha}{E}\,\frac{1+K^2Q^2}{KQ}\,e^{-\alpha t}\sin\frac{K\omega_0}{2}t. \quad . \quad . \quad (3.22)$$

It will be seen that this expression is generally similar to the corresponding expression for an i.f. amplifier employing single tuned circuits (expression 3.17, $n=2$) except that an oscillatory term $\sin \frac{K\omega_0}{2}t$ replaces the term in $at$ and an additional factor $\frac{1+K^2Q^2}{KQ}$ is introduced. It will be shown later that this is of relatively minor importance, since, for the coupled circuits, the response after the first half cycle of the oscillatory term is small, and during this first period the envelopes are of similar shape.

Returning to expression (3.22), the peak value of the noise envelope to signal envelope ratio occurs when $e^{-at}\sin\frac{K\omega_0}{2}t$ reaches its maximum value. This occurs when $\tan\frac{K\omega_0}{2}t=\frac{K\omega_0}{2a}=KQ$ (for practical circuits of this type $1<KQ<2$). Hence $t=\frac{2}{K\omega_0}\tan^{-1}KQ$, $at=\frac{1}{KQ}\tan^{-1}KQ$, and $\sin\frac{K\omega_0}{2}t=\frac{KQ}{(1+K^2Q^2)^{\frac{1}{2}}}$. Substituting these values in (3.22),

$$\left(\frac{N}{S}\right)_{max}=\frac{2a}{E}\left(1+K^2Q^2\right)^{\frac{1}{2}}e^{-1/KQ\tan^{-1}KQ}.$$

The value of $\Delta\omega=2\pi\,\Delta f$, where $\Delta f$ is the half band-width, for two coupled circuits is given approximately by $\Delta\omega=\frac{K\omega_0}{\sqrt{2}}$. Introducing this term,

$$\left(\frac{N}{S}\right)_{max}=\Delta\omega\frac{\sqrt{2}}{K\omega_0}\cdot\frac{2a}{E}\left(1+K^2Q^2\right)^{\frac{1}{2}}e^{-1/KQ\tan^{-1}KQ}$$

and rearranging, and using $\frac{2a}{\omega_0}=\frac{1}{Q}$,

$$\left(\frac{N}{S}\right)_{max}=\Delta\omega\cdot\frac{\sqrt{2}}{E}\frac{(1+K^2Q^2)^{\frac{1}{2}}}{KQ}e^{-1/KQ\tan^{-1}KQ}. \qquad (3.23)$$

From this expression it will be seen that, for a given value of $KQ$, the peak ratio of noise envelope to signal envelope is directly proportional to the band-width. Further, since the time at which

this peak is reached is inversely proportional to $K\omega_0$ and therefore inversely proportional to $\Delta\omega$ also.

If the particular value of $KQ=1$ is considered, expression (3.22) reduces to

$$\left(\frac{N}{S}\right) = \frac{4a}{E} e^{-at} \sin \frac{K\omega_0}{2} t.$$

For this value of $KQ$, $\Delta\omega = \sqrt{2}a$.

Substituting these values,

$$\left(\frac{N}{S}\right) = \frac{2\sqrt{2}\Delta\omega}{E} e^{-\Delta\omega t/\sqrt{2}} \sin \frac{\Delta\omega}{\sqrt{2}} t. \quad . \quad . \quad (3.24)$$

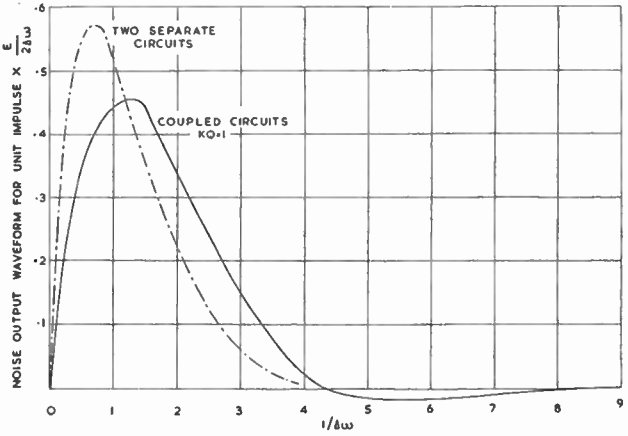



FIG. 3.15.—Noise output wave-form for i.f. amplifier employing coupled circuits. The magnitude of the wave-form is related to the peak value $E$ of a sinusoidal input at the centre frequency, and the i.f. half band-width $\frac{\Delta\omega}{2\pi}$.

This expression is plotted in Fig. 3.15, and to the same scale as Fig. 3.14, i.e. the time axis is plotted in terms of $1/\Delta\omega$, the value of 1 corresponding to the time for a vector having angular velocity $\Delta\omega$ to sweep out 1 radian. Similarly, the ordinate has been plotted to a value of $\frac{2\Delta\omega}{E}$ equal to unity. As a basis of comparison, the corresponding response of an i.f. amplifier employing two single tuned circuits having the same overall band-width is plotted in the same figure.

It will be seen that the bandpass circuit gives a lower peak value

of noise envelope to signal envelope ratio, whilst the time for this peak value to be reached is greater for the bandpass circuit.

It is interesting to note in passing that if $KQ$ is made very small, i.e. the coupling between the circuits of the bandpass filter is very small, expression (3.22) approximates to

$$\frac{2a}{E}\,\frac{1}{KQ}\,e^{-at}\,.\,\frac{K\omega_0}{2}\,t = \frac{2a}{E}\,at\,e^{-at},$$

which is identical with the response of the i.f. amplifier employing two single tuned circuits. This is, of course, an impractical condition, since with very weak coupling, the steady state gain falls to a low value.

It can be shown that the response of an i.f. amplifier employing two stages of coupled circuits is similar to that employing one such circuit, if the overall band-width is maintained constant. It differs from the response obtained by adding additional stages in i.f. amplifier with single tuned circuits, in that, whilst the time for the peak value of the noise envelope to be reached increases as before, the actual magnitude of the peak value of the ratio of noise envelope to signal envelope tends to increase slightly with the addition of further stages. This is the converse of what occurs with single tuned circuits.

The oscillatory term in the envelope of the response to an impulse, $\sin \frac{K\omega_0}{2}t$, goes to zero when $\frac{K\omega_0}{2}t = \pi,\ 2\pi,\ 3\pi$, etc. If all other circuit parameters are held constant, and only $K$ is varied, it is to be expected that as $K$ is increased, the duration of the "lobes" of the envelope will decrease. Also, as $K$ is increased the peak amplitude of the second and later lobes will increase, if the value of $a$ is unchanged.

This is well illustrated in the oscillograms shown in Fig. 3.16. The first shows the impulse wave-train emerging from a receiver having a double-peaked response characteristic (i.e. $K$ large and $KQ > 1$); it is possible to identify three successive maxima. A flat-topped response curve ($KQ = 1$) results in a main lobe with a trace of a second, while a round nosed response ($KQ < 1$) results in practically no identifiable secondary lobe. By mistuning so as to produce a pronounced second peak on the skirt of the response curve, the main and secondary lobes are run together, as shown in the fourth oscillogram.


FIG. 3.16.—Oscillograms showing noise output wave-forms for varying degrees of coupling between circuits. (*By courtesy of "Electronics".*)

As the frequency of the oscillatory component of the envelope is given by $2\pi f = \frac{K\omega_0}{2}$, it follows that the duration of the main lobe will be given by $\frac{2\pi}{K\omega_0}$.

It is of interest to note that the frequency of the oscillatory term $= \frac{1}{2\pi}\frac{K\omega_0}{2} = \frac{K}{2}f_0$, where $f_0$ is the centre frequency of the passband. For values of $KQ > 1$, $\frac{K}{2}f_0$ is approximately the separation between the centre frequency and the "ears" of the steady state response characteristic (the accuracy of this result increases with increasing values of $KQ$). In relation to the half band-width, i.e. the frequency separation between the centre frequency and the point at which the response is 3 db down, $\Delta f$ is given approximately by

$$\Delta f = \frac{Kf_0}{\sqrt{2}}.$$

Hence, in terms of the half band-width, the frequency of the oscillatory term is $\frac{\Delta f}{\sqrt{2}}$.

Thus for $\Delta f = 10$ kc/s, a typical figure for a high fidelity a.m. broadcast receiver, the frequency of the oscillatory component is 7 kc/s, and the duration of the main lobe 70 micro-seconds. For $\Delta f = 100$ kc/s, a typical figure for a v.h.f. receiver, the frequency of the oscillatory component is 70 kc/s, and the duration of the main lobe 7 micro-seconds.

### Addition of Carrier and Impulsive Interference Signal

An impulsive interfering signal adds to an existing carrier to produce amplitude and phase modulation of the carrier in a manner similar to the way in which a continuous wave signal does. There are, of course, two important differences: (1) the amplitude of the interfering signal is varying continuously; (2) the duration of the interference period is short. In order to simplify the analysis, we shall assume initially that the carrier and the response to the impulsive interference are at the same frequency. Then considering the output from a single i.f. stage employing coupled tuned circuits, the response to an impulse may be written as

$$Ae^{-at} \sin \omega_1 t \sin \omega_0 t,$$

where $\omega_1 = \frac{K\omega_0}{2}$, $\frac{\omega_0}{2\pi}$ = the mid-frequency of the passband, $A = g_m \omega_0^2 L$. The output due to an applied carrier may be written as

$$e = B \sin (\omega_0 t + \phi),$$

where $B = g_m EQ\omega_0 LKQ/1 + K^2Q^2$.

The phase angle $\phi$ is introduced here to allow for the fact that the impulse may be applied at any time in the course of the carrier-cycle; if the time at which the impulse is applied is taken as $t = 0$, $B \sin \phi$ gives the instantaneous magnitude of the carrier at this instant. As will be shown, the initial phase displacement of the carrier is of considerable importance. Since the time of application of the impulse is purely random, $\phi$ may have any value from 0 to $2\pi$.

Fig. 3.17 shows the situation of the carrier and interference vectors at time $t$ after the application of the impulse. The resultant vector of the addition of the two components $R$, is given by

$$R = C \sin (\omega_0 t + \theta),$$

where $$C = [(B \cos \phi + Ae^{-at} \sin \omega_1 t)^2 + B^2 \sin^2 \phi]^{\frac{1}{2}}$$

and $$\tan \theta = \frac{B \sin \phi}{Ae^{-at} \sin \omega_1 t + B \cos \phi}.$$

These results are obtained by geometrical considerations.

The resultant amplitude modulation is given by $|R|$, and the phase modulation by $\phi - \theta$.

Considering firstly the amplitude modulation component,

$$|R|=[B^2+A^2e^{-2at}\sin^2\omega_1t+2BAe^{-at}\sin\omega_1t\cos\phi]^{\frac{1}{2}},$$

and putting $\frac{A}{B}e^{-at}\sin\omega_1t=x$,

$$|R|=B[1+x^2+2x\cos\phi]^{\frac{1}{2}}.$$

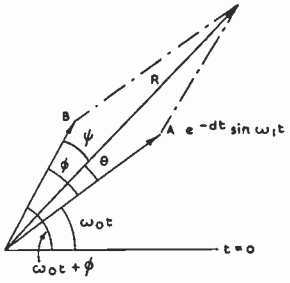



FIG. 3.17.—Showing vector addition of wanted signal and interference.

This is similar to expression (3.7) derived earlier, except that $x$ is now a function of time, and the angle $\phi$ replaces $2\pi bt$. The demodulated output is given by

$$B[(1+x^2+2x\cos\phi)^{\frac{1}{2}}-1] \quad . \quad . \quad . \quad . \quad (3.25)$$

and from the form of the expression will comprise a pulse, the exact shape of which will depend upon the value of $\phi$. Since the envelope of the response to the impulsive noise can be considered zero after time $t=\frac{\pi}{\omega_1}$, the duration of the output pulse is constant. The evaluation of the above expression for a train of impulses, as, for example, will occur if with motor-car ignition interference, involves consideration of the mean value of the output for all values of $\phi$. This involves considerable difficulties, and for our present purpose we shall limit our consideration to the case where the peak value of the impulsive interference signal envelope is very much smaller than that of the carrier envelope, i.e. $x\ll1$.

Then expression (3.25) can be simplified by neglecting $x^2$, and expanding $(1+2x\cos\phi)^{\frac{1}{2}}$, ignoring terms in $x^2$ and above. Then

$$|R|=B(1+x\cos\phi)$$

and the demodulated output

$$Bx\cos\phi=Ae^{-at}\sin\omega_1t\,.\cos\phi.$$

The shape of the output envelope is thus identical in shape with that of the i.f. output due to the application of the impulse.

All values of $\phi$ will arise with time if a series of impulses are applied, if it is assumed that the impulses are quasi-regular, i.e. the intervals at which they occur are approximately uniform, but not so exact that the value of $\phi$ is the same for each impulse. This condition is fulfilled for the type of interference met in practice. If the impulses recur at precisely regular intervals, the value of $\phi$ is consequently the same for all impulses, and the demodulated output will comprise a series of uniform pulses. The duration of each pulse being $\frac{\pi}{\omega_1}$, Fourier analysis of the output would yield components of discrete frequencies, $\frac{\omega_1}{2\pi}, \frac{2\omega_1}{2\pi}, \frac{3\omega_1}{3\pi}$, etc. This case is, however, trivial.

Returning to the case where the demodulated output comprises a series of pulses of random amplitude, we shall assume that the repetition rate of the pulses is such that the output due to each impulse is effectively zero before the next occurs. The pulses will all be of the same shape, but of different magnitude. To assess the audible output, we must determine the frequency spectrum of a single pulse.

We shall assume that the audio amplifier corresponds to an ideal low pass filter of cut-off frequency $f_c$. In order to provide a comparison between amplitude and frequency modulated receivers, we shall consider both types of receiver to be tuned to a very high frequency. With both amplitude and frequency modulation receivers working at these frequencies it is customary for the half i.f. band-width $\left(\frac{\Delta\omega}{2\pi} = \frac{\sqrt{2}\omega_1}{2\pi}\right)$ to exceed considerably the audio band-width. With this condition, the demodulated interference signal in the amplitude modulation case can be treated as being equivalent to an impulse, since its effective duration $\frac{\pi}{\omega_1}$ is very small compared with the duration of a cycle of the highest audio frequency. The magnitude of the impulse is given by

$$\int_0^{\pi/\omega_1} Ae^{-at} \sin \omega_1 t \,.\, \cos \phi \, dt,$$

i.e. the area under the pulse envelope. This, with sufficient

accuracy, is given by $A\frac{\omega_1}{a^2+\omega^2}\cos\phi$. As pointed out earlier, the spectrum of an impulse is uniform with frequency; and therefore the resultant of the components of the demodulated output in the audio frequency range between frequencies $f$ and $f+df$ is given by $F(f)=\frac{A\omega_1}{a^2+\omega_1^2}\cos\phi\,df$.

To determine the r.m.s. value of the audio output, we must evaluate the contribution of $F(f)$ at all frequencies in the working range. The sum of all components up to $f_c$ is

$$\int_{-f_c}^{f_c} F(f)^2\,df.$$

Therefore the mean square value calculated over the interval $1/f_r$, where $f_r$ is the repetition rate of the pulses, is

$$f_r\,2\int_0^{f_c} A^2\,\frac{\omega_1^2}{(a^2+\omega_1^2)^2}\cos^2\phi\,df$$

$$=2f_rA^2\,\frac{\omega_1^2}{(a^2+\omega_1^2)^2}\cos^2\phi\,f_c$$

$$=2f_rf_c\,\frac{\omega_1^2}{(a^2+\omega_1^2)^2}\,A^2\cos^2\phi.$$

The r.m.s. value is therefore

$$\sqrt{2(f_rf_c)^{\frac{1}{2}}A}\,\frac{\omega_1}{(a^2+\omega_1^2)}\cos\phi. \quad . \quad . \quad . \quad (3.26)$$

In order to complete the investigations, we shall evaluate the r.m.s. signal to noise ratio. Assuming the wanted carrier to be modulated 100 per cent by a sinusoidal signal, the r.m.s. value of the audio output is

$$\frac{1}{\sqrt{2}}B$$

and therefore the r.m.s. signal to noise ratio is

$$\frac{S}{N}\text{ r.m.s.}=\frac{1}{2}\,\frac{B}{A}\,\frac{a^2+\omega^2{}_1}{\omega_1}\,\frac{(f_rf_c)^{-\frac{1}{2}}}{\cos\phi}.$$

Now $B=EQ\omega_0 L \cdot \dfrac{KQ}{1+K^2Q^2}$ and for a unit impulse $A=\omega_0^2 L$,

$$\therefore \frac{S}{N}\text{ r.m.s.} = \frac{E}{2}\frac{Q}{\omega_0}\frac{KQ}{1+K^2Q^2}\cdot\frac{a^2+\omega_1^2}{\omega_1}\frac{(f_r f_c)^{-\frac{1}{2}}}{\cos\phi}$$

$$= \frac{E}{2}\frac{1}{2a}\frac{a^2+\omega_1^2}{\omega_1}\frac{KQ}{1+K^2Q^2}\frac{(f_r f_c)^{-\frac{1}{2}}}{\cos\phi}.$$

Now
$$\frac{a^2+\omega_1^2}{a\omega_1} = \frac{1+K^2Q^2}{KQ}$$

and thus
$$\frac{S}{N}\text{ r.m.s.} = \frac{E}{4}\frac{(f_r f_c)^{-\frac{1}{2}}}{\cos\phi}.$$

$\cos\phi$ can take any value from 0 to 1, and for a succession of impulses, it can be replaced by its mean value $\dfrac{2}{\pi}$.

Substituting this value $\dfrac{S}{N}\text{ r.m.s.} = E\dfrac{\pi}{8}(f_r f_c)^{-\frac{1}{2}}$. . . (3.27)

It is interesting to note that this value is independent of the overall band-width of the i.f. amplifier, provided this exceeds twice the audio band-width, and depends solely upon the value of $f_r$ and $f_c$. From this expression it will be seen that the signal to noise ratio deteriorates at the rate of 3 db per octave of pulse repetition frequency. Although this result is evaluated for low values of i.f. output signal to noise ratios, further analysis shows that the result holds good generally.

Consider now the frequency modulation component, or rather the phase modulation component, given by expression

$$\tan\phi = \frac{B\sin\phi}{Ae^{-at}\sin\omega_1 t + B\cos\phi}.$$

For the present purpose it is more convenient to rearrange this expression employing the angle $\psi$, the phase shift between the resultant vector and the carrier (*see* Fig. 3.17). From this figure $\psi=\phi-\theta$,

$$\tan\psi = \frac{Ae^{-at}\sin\omega_1 t\sin\phi}{B+Ae^{-at}\sin\omega_1 t\cos\phi}.$$

If the same assumption is made as in the amplitude modulation case, namely, that $B \gg A$, $\psi$ will be very small, so that $\tan \psi \simeq \psi$, and the above expression reduces to

$$\psi = \frac{A}{B} e^{-at} \sin \omega_1 t \sin \phi. \quad . \quad . \quad . \quad (3.28)$$

This shows that the phase deviation output is identical in shape with the a.m. output as given by expression (3.25). It is of interest to note that expression (3.25) includes $\cos \phi$, whilst expression (3.28) includes $\sin \phi$. Thus when $\phi = 0$, there is no phase modulation, and maximum amplitude modulation. Conversely, when $\phi = \pi/2$, there is maximum phase modulation and no amplitude modulation. This is obvious from inspection of Fig. 3.17.

It is of interest to observe at this point, that in a phase modulation system, the actual magnitude of the demodulation noise output is as given by expression (3.28) above, and it varies inversely with $B$, the amplitude of the carrier output. In an amplitude modulation system, the noise output is constant irrespective of carrier amplitude. However, in the amplitude modulation case, the demodulated wanted signal output is directly proportional to $B$, so that the ultimate signal to noise ratio is inversely proportional to B. In the phase modulation case, the demodulated wanted signal output is independent of carrier amplitude, so that, subject to the condition that $B \gg A$, the ultimate signal to noise ratio is also inversely proportional to $B$.

If an amplitude modulation system is compared with a phase modulation system having a maximum phase deviation of 1 radian, it will be apparent from the foregoing that the ultimate signal to noise ratio is the same for both. Further, if the phase modulation system has a maximum phase deviation of $n$ radians, the signal to noise ratio for the phase modulation system will be better than that of the amplitude modulation system by a factor of $n$ times in voltage, and $n^2$ in power.

For a frequency modulation system, the demodulated output is proportional to the derivative of the phase deviation. Consequently, the shape of the output is markedly different from that met in the cases of amplitude and phase modulation. This is shown in Fig. 3.18; it is this difference which determines the superiority of a frequency modulation system as compared with an amplitude or a phase modulation system.

The demodulated output from a frequency modulated discriminator is proportional to $\frac{1}{2\pi}\frac{d\psi}{dt}$; if $H$ is the sensitivity of the discriminator, i.e. volts out/cycle of frequency shift, the actual output is

$$e = \frac{H}{2\pi}\frac{d\psi}{dt}.$$

Now as we have already shown, the envelope phase deviation $\psi$ is identical with the a.m. output except for the factors $B$ and $\cos\phi$,

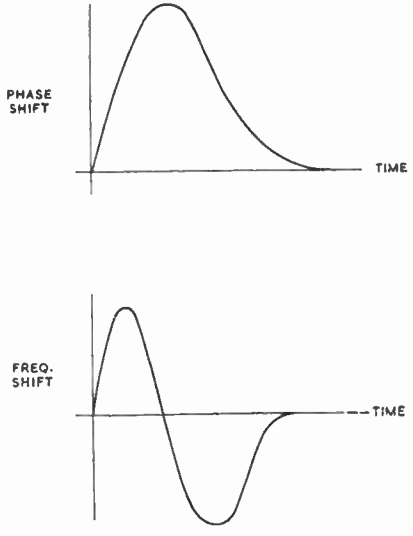



FIG. 3.18.—Illustrating demodulated noise output in an f.m. system.

and by a well-known property of the Fourier integral the distribution of amplitude with frequency for the frequency modulation demodulated output $F_1(f)$ is given by

$$F_1(f) = \frac{H}{2\pi} \times 2\pi f \times \left[ F(f) \text{ a.m.} \times \frac{\sin\phi}{\cos\phi} \times \frac{1}{B} \right],$$

i.e. the resultant of the components at frequencies between $f$ and $f+df$, $F_1(f)$ is directly proportional to $f$. Hence

$$F_1(f) = \frac{AH}{B}\frac{\omega_1}{a^2+\omega_1^2}\sin\phi \,.\, f.\, df.$$

Assuming as before that the half i.f. band-width considerably exceeds the audio band-width, the r.m.s. value of the noise output in the a.f. output can be evaluated as before and is given by

$$\text{Noise r.m.s.} = \left(\frac{2}{3}\right)^{\frac{1}{2}} \frac{AH}{B} \frac{\omega_1}{a^2+\omega_1^2} \sqrt{f_c^3 f_r} \sin\phi. \qquad (3.29)$$

Now the signal output is dependent upon $f_d$, the frequency deviation. In order to compare the frequency modulation with the amplitude modulation case we shall consider the carrier to be 100 per cent modulated by a sinusoidal signal. In this case the r.m.s. value of the demodulated output is given by

$$\frac{1}{\sqrt{2}} Hf_d.$$

And the r.m.s. value of the signal to noise ratio is, therefore,

$$\frac{S}{N} \text{ r.m.s.} = \frac{\sqrt{3}}{2} \frac{BHf_d}{AH} \frac{a^2+\omega_1^2}{\omega_1} \frac{(f_c^3 f_r)^{-\frac{1}{2}}}{\sin\phi}$$

$$= \frac{\sqrt{3}}{4} E \frac{f_d}{\sin\phi} (f_c^3 f_r)^{-\frac{1}{2}},$$

employing

$$\frac{B}{A} \frac{a^2+\omega_1^2}{\omega_1} = \frac{E}{2}.$$

Since $\phi$ can take all values from 0 to $2\pi$, $\sin\phi$ can be replaced for a succession of impulses by its mean value $\frac{2}{\pi}$, hence

$$\frac{S}{N} \text{ r.m.s.} = \sqrt{3} \frac{\pi}{8} Ef_d (f_c^3 f_r)^{-\frac{1}{2}}. \qquad (3.30)$$

Thus the ratio of the r.m.s. values of the signal to noise radio for comparable amplitude modulation and frequency modulation systems is

$$\frac{S/N \text{ r.m.s. f.m.}}{S/N \text{ r.m.s. a.m.}} = \sqrt{3} \frac{f_d}{f_c}. \qquad (3.31)$$

It is the fact that for a frequency modulation system $f_d$ is usually much greater than $f_c$ that leads to the superior performance of frequency modulation in this respect. The subjective assessment of annoyance due to impulsive interference would appear to indicate that the improvement is greater than the comparison of r.m.s. values would suggest. This may be ascribed to the fact that

$$\frac{S}{N} \text{ r.m.s.} = \frac{\pi\sqrt{3}}{8} \cdot \frac{E f_d}{(f_c^3 f_r)^{\frac{1}{2}}}$$

linearly rising distribution of amplitude with frequency in a frequency modulation system is much less disturbing to the listener than the uniform distribution associated with an amplitude modulation system.

A frequency modulation system does not, however, maintain its superiority over an amplitude modulation system at the same value for all magnitudes of the impulsive interfering signal. In particular,

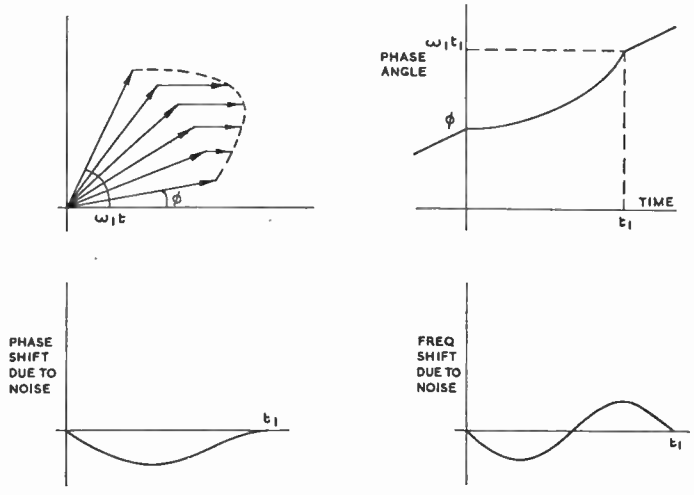



FIG. 3.19.—Vector addition of wanted signal and interference, with relative rotation, generating "click".

the signal to noise ratio deteriorates rapidly when the peak value of the interference at the i.f. output exceeds the carrier amplitude. In order to explain this phenomenon, we must examine the vector diagrams illustrating the relationship between the carrier and interference vectors in this region.

We assumed in the earlier discussion that the interference vector maintained a constant direction with respect to the carrier vector throughout its period of effective duration. If the carrier is modulated, or if the carrier is not at the i.f. centre frequency, the vectors will have a relative rotation during the course of the period of the impulsive interference, and the locus of the resultant vector will therefore lie on a curve, as shown in Fig. 3.19. This shows the instantaneous resultant vector at successive instants; $\phi$ is the initial phase angle between the carrier vector and the

interference vector and $t_1$ is the period of effective duration of the interference. It is assumed that $\omega_1$ is constant during the interval $t_1$, which is true if the carrier is mistuned only, and approximately true if the relative rotation is due to modulation. If the initial value of $\phi$ is in the region of $\pi$ radians, the locus of the resultant vector can enclose the origin 0, provided that the amplitude of

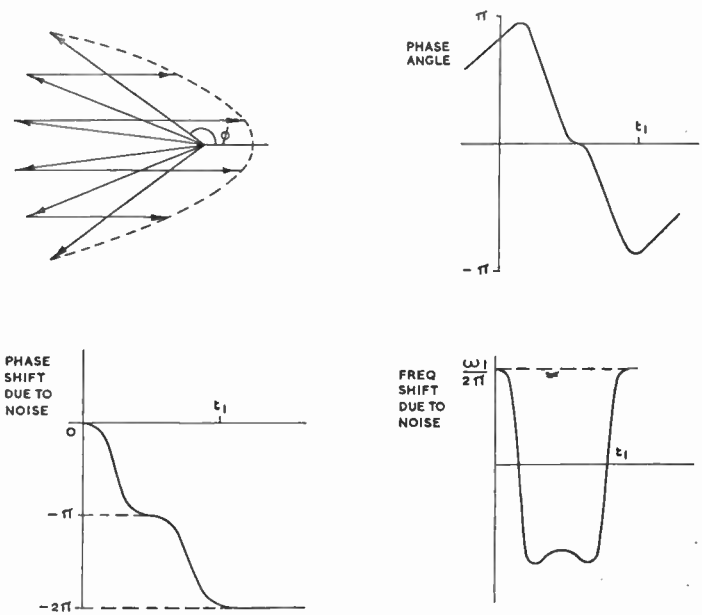



FIG. 3.20.—Generation of "pop" in an f.m. system, with relative rotation of wanted and interfering signals.

the interference exceeds that of the carrier at the instant when the angle between the vector is precisely $\pi$ radians. This is depicted in Fig. 3.20 from which it will be seen that the frequency deviation output differs appreciably from that of Fig. 3.19. In fact, the output approximates to a double peak, with both peaks of the same polarity, combined with a square pulse, corresponding to the removal of the modulation for the interference period. The distribution of amplitude with frequency of such an output tends to a uniform spectrum, and, consequently, the audible noise output is greatly increased, since the demodulated noise output approaches that occurring in the amplitude modulation case. The character of

the noise also changes, being of the nature of a "pop" rather than the "click" heard when this effect does not occur.

If the resultant wave-form of the i.f. signal is observed, the corresponding effect is of the signal having slipped or gained a single cycle in the period $t_1$. There is an upper limit to the deterioration of the signal to noise ratio since when the interference peak amplitude is sufficiently great, all the demodulated interference outputs will be of this form. It is thus of interest to note that the signal to noise ratio then tends to a constant value.

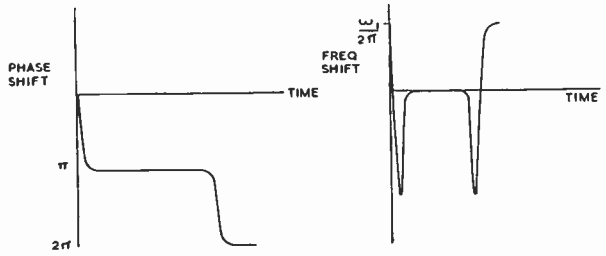



FIG. 3.21.—As Fig. 3.20 but with greater relative velocity.

The effect of further increase of the interference peak amplitude is to modify the shape of the unpolar noise output, to that shown in Fig. 3.21, where the peaks are more sharply defined. The contribution to the noise output due to this cause is, however, of minor importance compared with the contribution due to the mechanism discussed below.

In the foregoing analysis, certain simplifications were made about the nature of the response to the interfering signal. A detailed analysis reveals, however, that in certain cases the output from an i.f. stage due to an interfering impulsive signal can have a component which produces a phase rotation of the output vector. For example, if the input signal is of the form of a unit step, and the i.f. stage has a pair of coupled tuned circuits, the interference output is of the form

$$-\omega_0 L e^{-\alpha t}\left(\sin\frac{K\omega_0}{2}\,t\cos\omega_0 t-\frac{K}{2}\cos\frac{K\omega_0}{2}\,t\sin\omega_0 t\right).$$

The second term of this expression is of small magnitude, and can generally be neglected. However, in the consideration of threshold effects, it is of considerable importance, since it produces a

rotation of the interference vector during the course of its period, and hence an effect similar to that previously discussed can occur. If $\tan\theta = \frac{2}{K}\tan\frac{K\omega_0}{2}t$, the above expression reduces to

$$\omega_0 L e^{-\alpha t}\sin(\omega_0 t - \theta).$$

The value of $\theta$ for $K = 0{\cdot}01$ is plotted in Fig. 3.22.

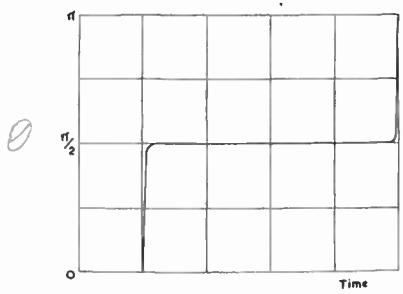



FIG. 3.22.—Variation of phase angle of interference vector due to application of unit step input to a pair of coupled circuits.

Since $\theta$ varies with time, the expression shows that the interference vector is not stationary with respect to a vector rotating at the centre frequency of the passband. The relative rotation of the vectors can produce a double peaked unipolar output in the same way as described in the previous section, but due to the much greater rate of relative velocity, the output from this cause usually greatly exceeds that due to modulation or mistuning effects.

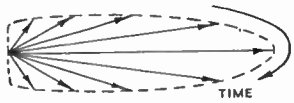



FIG. 3.23.—Relative rotation of interference vector, due to variation of phase angle with time shown in FIG. 3.22.

The position of the interference vector at successive instants is shown in Fig. 3.23 and it will be seen that the angle between the initial and final positions of the vector is $\pi$ radians. The formation of a "pop" and a "click" due to the interference vector rotation is shown in Fig. 3.24. Due to the very rapid changes of phase angle, the peak frequency shift tends to a very large value. Since the

relative positions of the interference vector is 180° at the beginning and end of the cycle, the ultimate ratio of "pops" to "clicks" from this cause tends to unity. Even when allowance is made for

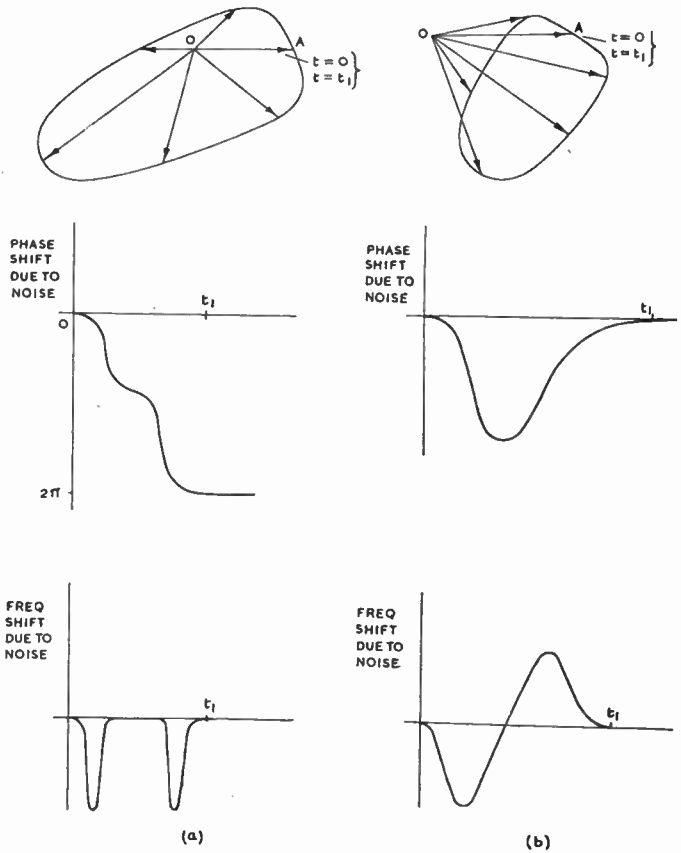



FIG. 3.24.—Production of (*a*) "pop" and (*b*) "click" due to rotation of interference vector; *OA* shows initial position of carrier vector, assumed unmodulated.

this phenomenon, the signal to noise ratio of a frequency modulation system with a frequency deviation of 75 kc/s is at worst better by a factor of 10 db than a corresponding amplitude modulation system. This is shown in Fig. 3.25 which shows comparative figures for signal to noise ratios for an amplitude modulation and frequency modulation system to impulsive interference, found by the BBC.

Any further increase in interference amplitude beyond the carrier amplitude tends only to alter the shape of the pulses obtained with a "pop" output, the pulses becoming taller and narrower. The actual magnitude × time area under the pulse envelopes tends to a constant value and, therefore, as the noise

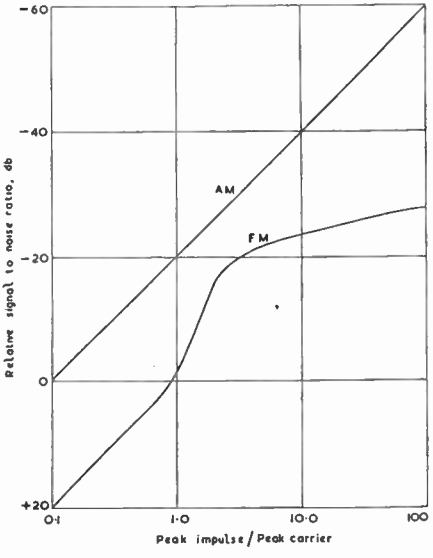



FIG. 3.25.—Output signal to noise ratio for f.m. and a.m. systems for varying levels of interference to wanted signal ratio.

output is computed by assuming the two pulses to be approximations to unit impulses, the noise output tends to a constant value as the pulses tend more nearly to the ideal shape.

By contrast, the output from an amplitude modulation detector can increase indefinitely. It is at the point where the rapid fall in signal to noise ratio occurs that the difference between amplitude modulation and frequency modulation is least marked. In general, however, even in this region, the performance of a frequency modulation system is superior to that of an amplitude modulation system. It is possible, if the interference signal is increased sufficiently, to observe a second region where the signal to noise ratio again falls rapidly. This happens when the vector phase angle changes by $4\pi$ radians, i.e. the resultant vector encloses the origin twice.

However, this region is of minor practical interest.

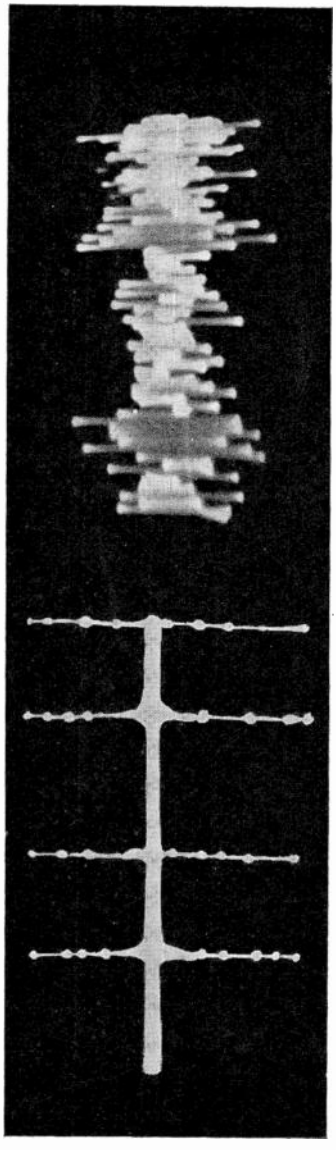


FIG. 3.26.—The first oscillogram shows a typical impulsive noise voltage and the second a fluctuational noise voltage. (*By courtesy of the British Institute of Radio Engineers.*)

### Fluctuation Noise

It has earlier been indicated that fluctuation noise originates in the early stages of a receiver, the two principal causes being thermal agitation and Shot Effect. Noise due to the first source, which normally produces the larger interference voltages, arises as a result of the thermal agitation of the electrons in the conductors forming the receiver's input circuit. The conductivity of metals is dependent on the presence of free electrons, which are continuously moving about within the conductor at a velocity which is dependent upon its temperature. At any one instant there will ordinarily be more electrons moving in one direction than the other, with the result that a voltage is developed across the ends of the conductor. This voltage will vary from instant to instant in an irregular manner, in accordance with the predominant electron motion. The resultant energy is uniformly distributed over the entire frequency spectrum from zero up to frequencies far above those which have as yet been employed for radio communication.

The voltage produced in any specified frequency band as the result of thermal agitation tends to a constant mean square value

$E^2$ when averaged over a long period, and this value is given by the formula:

$$E^2 = 4kTR(f_1 - f_2), \quad . \quad . \quad . \quad . \quad (3.32)$$

where

$f_1$ and $f_2$ are the limits of the frequency band;

$k$ = Boltzmann's constant = $1{\cdot}374 \times 10^{-23}$ joules per degree centigrade;

$T$ = absolute temperature (273 + °C.);

$R$ = resistance component of the impedance producing the thermal agitation voltages. Account must be taken of the increase in resistance due to skin effect. The resistance component is assumed to be constant over the range $f_1$ to $f_2$.

When this expression is applied to determine the noise transmitted through actual bandpass circuits, $f_1$ and $f_2$ are the frequency limits of the equivalent perfect bandpass network, i.e. one having uniform transmission in the passband, and no response outside it. The equivalent bandpass circuit is assumed to give the same output as the actual circuit at its centre frequency, and have the same area under its power-gain/frequency characteristic.

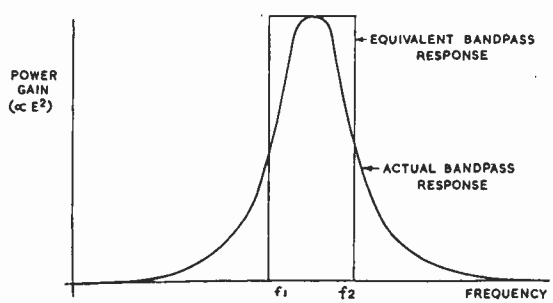



FIG. 3.27.—Equivalent ideal bandpass response to an actual bandpass response, for determining magnitude of fluctuation noise.

As a general indication it may be stated that the thermal agitation voltage developed in the grid circuit of the first stage of a very high frequency broadcast receiver will be in the order of 1 microvolt.

The Shot Effect is due to the fact that the stream of electrons flowing from cathode to anode of a valve is made up of a very

large number of individual electrons, each of which results in a minute voltage impulse. As the electrons arrive at the anode in a random manner, slight irregularities in the plate current result. The voltage fluctuations produced by Shot Effect are uniformly distributed over the whole of the radio frequency spectrum. The voltage produced in the anode circuit of the first amplifier stage due to this cause is determined by the following equation:

$$E^2 = 3{\cdot}18 \times 10^{-19} IZ(f_1 - f_2), \quad . \quad . \quad . \quad (3.33)$$

where

$E^2$ = the square of the effective voltage produced by components lying in the frequency band between $f_1$ and $f_2$;

$I$ = the electron current;

$Z$ = the resonant impedance of the anode circuit, which is assumed to be constant over the frequency range $f_1 - f_2$.

(The formula ignores any improvement due to cathode space charge.)

By way of comparison it may be stated that the noise voltage resulting from Shot Effect, if referred to the grid of the first valve in the receiver, will be the equivalent of an input voltage of approximately 1 microvolt or less. This figure will, of course, be dependent upon the valve gain, the cathode space charge, and the impedance of the anode circuit.

The noise output due to fluctuation noise may be evaluated by the methods applied to continuous wave interference, provided that the magnitude of the interference is small compared with the carrier amplitude. If this assumption is made, then the mean square value of the noise e.m.f. $(\Delta e)^2$ associated with a very narrow band of frequencies $df$ centred about a frequency $f$ can be treated as the mean square value of a continuous wave component of frequency $f$. The equivalent amplitude modulation depth $x$ has, therefore, a mean square value given by $\left(\frac{\Delta e}{e}\right)^2$, where $e$ is the carrier amplitude. Since $x$ is assumed very small, the demodulated output will comprise a component at frequency $f$ only, the magnitude of the harmonic components being negligible. The total mean square value of the output due to all such components between frequencies $f_0 - f_c$ and $f_0 + f_c$ (where $f_0$ is the

carrier frequency and $f_c$ is the highest audible frequency) is therefore

$$E_1^2 = \int_{-f_c}^{+f_c} (\Delta e)^2$$

but $(\Delta e)^2 = Cdf$, where $C$ is a constant, the value of which depends upon the source of the noise, so that

$$E_1^2 = 2Cf_c.$$

For the frequency modulation case, with the same assumptions, the mean square value of the equivalent modulation depth is, from expression (3.12), given by

$$\left(\frac{\Delta e}{e}\frac{f}{f_d}\right)^2$$

and the mean square value of the audible output, employing (3.11), is given by

$$E_2^2 = \int_{-f_c}^{+f_c} \left(\frac{\Delta e}{e} fH\right)^2,$$

where $H$ is the sensitivity of the discriminator, i.e. volts out/cycle of frequency shift.

Here $$(\Delta e)^2 = Cdf,$$

whence $$E_2^2 = \frac{2C}{e^2} H^2 \frac{f_c^3}{3}.$$

The r.m.s. value of the noise for each case is, therefore, as follows:

a.m. $\quad E_1 = \sqrt{2Cf_c}$

f.m. $\quad E_2 = \sqrt{2CH^2 f_c^3/3e^2}.$

Assuming equal amplitude a.m. and f.m. carriers, the signal to noise ratio referred to 100 per cent modulation is given in the amplitude modulation case by

$$(S/N) \text{ a.m.} = e/\sqrt{2Cf_c},$$

and for frequency modulation by

$$(S/N) \text{ f.m.} = \frac{Hf_d}{\sqrt{2CH^2 f_c^3/3e^2}}$$

$$= \frac{ef_d\sqrt{3}}{\sqrt{2Cf_c^3}}.$$

The ratio of the signal to noise ratios is, therefore,

$$\frac{(S/N) \text{ f.m.}}{(S/N) \text{ a.m.}} = \sqrt{3}\,\frac{f_d}{f_c}.$$

That is, the improvement for a frequency modulation system over a corresponding amplitude modulation system is the same as that in the case of impulsive interference. As also with impulsive interference, the subjective annoyance due to fluctuation noise is better than the above comparison of signal to noise ratio would suggest; the linearly rising distribution of amplitude with frequency for frequency modulation being apparently less objectionable than the uniform distribution associated with amplitude modulation.

### Fluctuation Noise Crest Factor

The crest factor may be defined as the ratio of the amplitude of the highest peaks to the r.m.s. voltage. A number of experimental attempts have been made to ascertain whether a definite value can be given for the crest factor of fluctuation noise. The various values obtained by different workers have ranged from 3·4 to 4·5. For convenience in making calculations and plotting curves, a compromise value of 4 will be assumed throughout this book. From the theoretical standpoint it is, however, impossible to place any finite value on the crest factor of fluctuation noise.

The reasons for this require some further explanation. Considering Shot Effect as a typical example of fluctuation noise, the basic noise unit is due to the arrival of one electron at the anode. This may be considered to be an infinitesimal unit impulse. A unit impulse is known to have a uniform distribution of energy over the frequency spectrum. The component waves of a single unit impulse are all in phase at the instant the impulse occurs. However, if two unit impulses occur at different times, the frequency components will add together with various phases. When the impulses occur at random times and the number of impulses is increased to an indefinitely large value, the relative phases of the various components will have no discernible relationship and may be said to be at random.

If all these component waves could at any time come into phase, the voltage would rise to an indefinitely large value;

however, the probability of this happening is infinitesimal. Nevertheless, if any finite crest factor is chosen, the probability that it will be exceeded at a given instant is a finite function. This probability is another name for the fraction of the time that the voltage will exceed the chosen value in a given period of time. It has been shown by Landon that it is possible to express this probability in the form of the curve given in Fig. 3·28. To take an example of the way in which this curve may be used, the point at which the crest factor $=0$ and the probability factor $=0{\cdot}5$

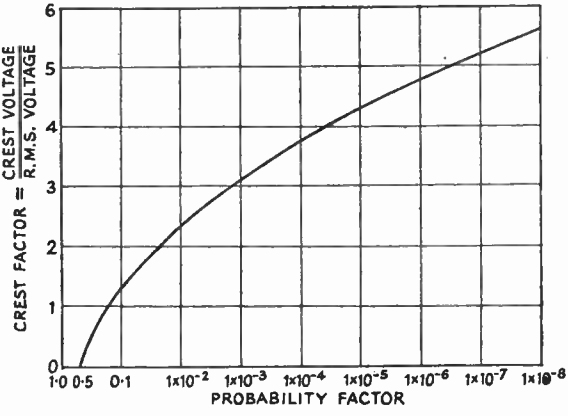



FIG. 3.28.—The fraction of time during which the fluctuation noise voltage will probably exceed any given crest factor may be determined from the above curve.

indicates that the voltage is above the zero value for half the time. In other words, it is positive for half the time and negative for the other half. Again, the point at which the crest factor $=1$ and the probability $=0{\cdot}16$ indicates that the voltage exceeds the r.m.s. value for 16 per cent of the time. To take one further example, the point at which the crest factor $=4$ and the probability $=0{\cdot}000032$ indicates that the voltage is in excess of 4 times the r.m.s. value for 3·2 seconds out of 100,000.

The curve is plotted for the amplitude modulation case in which it is only necessary to consider voltage fluctuations in one direction. When it is necessary to consider whether peaks of either polarity or direction will exceed a certain value, the probabilities given in the curve should be doubled.

It may be shown mathematically that with the receiver band-widths used in practical communication systems, the probability of any noise peak substantially exceeding some four times the r.m.s. level is very remote. The reason for this is the very low probability that any noise peak or impulse will be of sufficiently long duration to be amplified at its full relative amplitude. As the receiver band-width is increased the amplitude at which short duration noise peaks will be reproduced will be increased slightly. The variations in the experimental measurements mentioned at the beginning of this section were possibly due to observers setting a limit to the crest factor without taking into account the receiver band-width used.

### Threshold of Improvement

As in the case of impulsive interfence, the improvement of a frequency modulation system over an amplitude modulation system with respect to fluctuation noise, also falls abruptly in the region where the carrier and interference are of approximately equal magnitudes. The question of the crest factor is of importance in the matter, in that in order to determine the value at which the carrier and interference are of equal magnitude, we must take not the r.m.s. value of the noise wave-form, but rather a value which takes into account the peak interference amplitude. For this reason, it is convenient to take the factor 4 as the fluctuation noise crest factor.

The threshold of improvement with respect to fluctuation noise coincides with the boundary below which reception is not possible. This follows from the fact that once the signal amplitude is less than that of the fluctuation noise, the modulation will be almost entirely masked by the noise. In this, there is marked difference between fluctuation and impulse noise; due to the discontinuous nature of the latter, reception is possible despite the fact that the peak noise magnitude exceeds that of the carrier.

As in the cases of impulsive and continuous wave interference, the signal to fluctuation noise ratio deteriorates abruptly for a frequency modulation system in the region of the threshold of improvement. This is due to two facts: firstly, the greatly increasing-magnitude of the phase deviation as the interference amplitude approaches that of the carrier. Secondly, the production of "pops" by the same mechanism as discussed in the case of impulsive interference.

## SELECTED REFERENCES

LANDON, V. D., A Study of the Characteristics of Noise, *Proc. I.R.E.*, November 1936.

CROSBY, M. G., Frequency Modulation Noise Characteristics, *Proc. I.R.E.*, April 1937.

KEALL, O. E., Interference in Relation to Amplitude, Phase and Frequency Modulation Systems, *Wireless Engineer*, January 1941.

LANDON, V. D., Distribution of Amplitude with Time in Fluctuation Noise, *Proc. I.R.E.*, February 1941.

LANDON, V. D., Impulse Noise in F.M. Reception, *Electronics*, February 1941.

BELL, D. A., F.M. Communication Systems, *Wireless Engineer*, May 1943.

TURNEY, T. H., *Heaviside's Operational Calculus Made Easy*. Chapman and Hall, London. Second edition, 1946.

MAURICE, R.D.A., *Les Parasites Artificiels Dans Les Systèmes de Modulation Par Variation De L'Amplitude* (mA) *Par Variation De La Fréquence. Onde Electrique*, March 1954.

*Chapter Four*

# INTERFERENCE SUPPRESSION

It is as well to start this chapter by stating that frequency modulation does not, in the absolute sense of the word, suppress interference. Under favourable conditions its use will very greatly reduce the audio disturbances resulting from an interfering signal. The only satisfactory basis upon which to assess the improvements arising from the use of frequency modulation is therefore that of a comparison between the interfering audio disturbance reproduced by a frequency modulation receiver and that reproduced by an amplitude modulation receiver, operating on the same waveband and subjected to the same interfering signal. When such a comparison is made it is found that the improvement is not constant but varies with different forms of interference. Further, it is found that there is always an increase in the improvement figure as the deviation ratio of a frequency modulation system is raised. However, regardless of the type of interference or the deviation ratio employed, the improvement has a minimum value at a point known as the "threshold of improvement". Normally, this point occurs when the interference peak amplitude equals that of the wanted signal. Once the threshold of improvement has been reached, the difference between frequency modulation and amplitude modulation depends on the nature of the interference. Due to causes which will be discussed later, the carrier level at which the threshold of improvement occurs is raised as the deviation ratio increases.

The foregoing remarks give a very general introduction to the more important of the factors involved in a study of the way in which interference is suppressed in a frequency modulation system. A careful examination of the conclusions which are reached in this chapter will amply repay anyone who is concerned with frequency modulation equipment. It will ensure that the impossible is not demanded from a frequency modulation receiver and at the same time permit a proper appreciation of the greatly improved results which are possible.

In the last chapter it was shown that the interference modulation

produced to a desired carrier by an unwanted signal is composed of two component parts, namely, an amplitude and an angular modulation component. Now a frequency modulation receiver incorporates a special stage—the limiter—the only function of which is that of suppressing all the amplitude variations

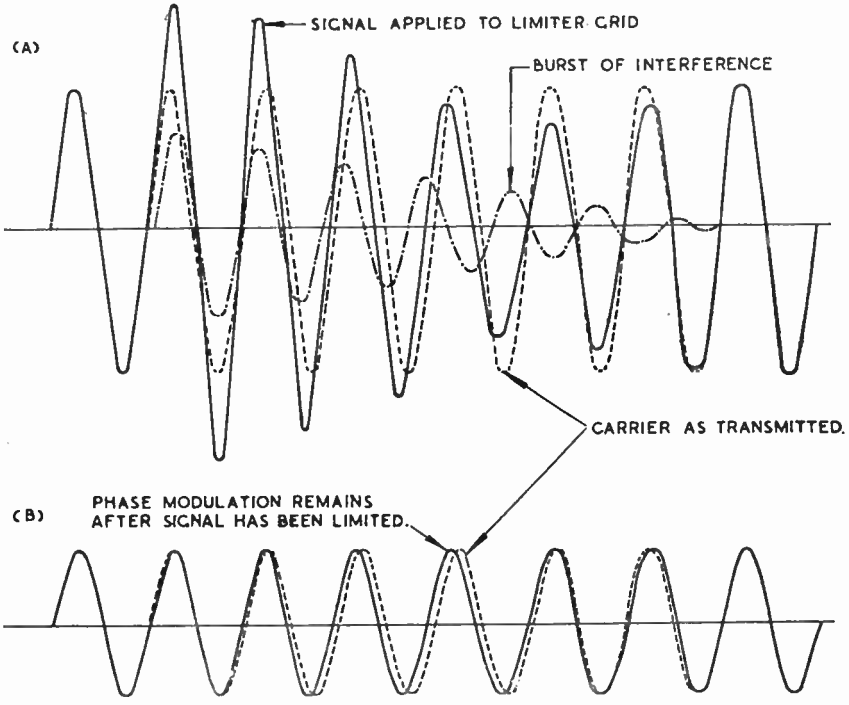



FIG. 4.1.—Diagram (A) shows the combined interference and carrier waves as applied to the limiter valve's grid. Diagram (B) shows that the amplitude changes are absent in the output but the interference phase modulation component remains.

*(By courtesy of the British Institute of Radio Engineers.)*

of the incoming carrier. Should the received signal consist of a carrier with perhaps superimposed impulsive interference or a heterodyne from a second and unwanted station, then the limiter will eliminate the amplitude component, leaving a wave-form of constant level. This action has been illustrated pictorially in Fig. 4.1, from which it is apparent that although the limiter suppresses the amplitude modulation component it does not alter the phase modulation component.

## The Noise Triangle

In the presence of continuous wave interference, where the wanted carrier amplitude exceeds considerably the interfering signal amplitude, the resultant effect in both amplitude modulation and frequency modulation is to produce an output at the difference frequency of the two signals. For amplitude modulation the amplitude of this output is independent of the frequency difference; for frequency modulation the output is directly proportional to it. When the interfering signal is appreciably smaller in amplitude than the wanted signal, the interference amplitude in both is directly proportional to the amplitude of the interfering signal. These statements may be expressed by the following equations:

for a.m. $$E_N = \frac{e_{int}}{e_c};$$

for f.m. $$E_N = \frac{f_c - f_{int}}{f_d} \frac{e_{int}}{e_c};$$

where $E_N$ = amplitude of the interfering signal demodulated output referred to 100 per cent modulation output;
$f_c$ = carrier frequency;
$f_{int}$ = interference frequency;
$f_d$ = frequency deviation;
$e_c$ = peak carrier voltage;
$e_{int}$ = peak interfering signal voltage.

These relationships are shown graphically in Fig. 4.2. This diagram illustrates what has frequently been termed the frequency modulation noise triangle. It assumes that the carrier and the interfering signal are held at the same relative amplitude, only their frequencies being varied. To illustrate the result of this form of noise distribution, an example may be taken of an interfering signal occurring 15 kc/s from the carrier; the resultant carrier frequency modulation will be ten times as great as that which would have occurred had the interfering signal been only 1,500 cycles from the carrier.

It is apparent from Fig. 4.2 that a frequency modulation system, having a peak deviation of 15 kc/s and passing a maximum audio frequency of 15 kc/s (system *b*), will on the average result

in half the noise which would have been reproduced by a comparable amplitude modulation system (system *a*), which would reproduce the noise at its full amplitude over the whole audio band. This direct comparison between amplitude and frequency modulation systems is true only so long as the harmonic frequency

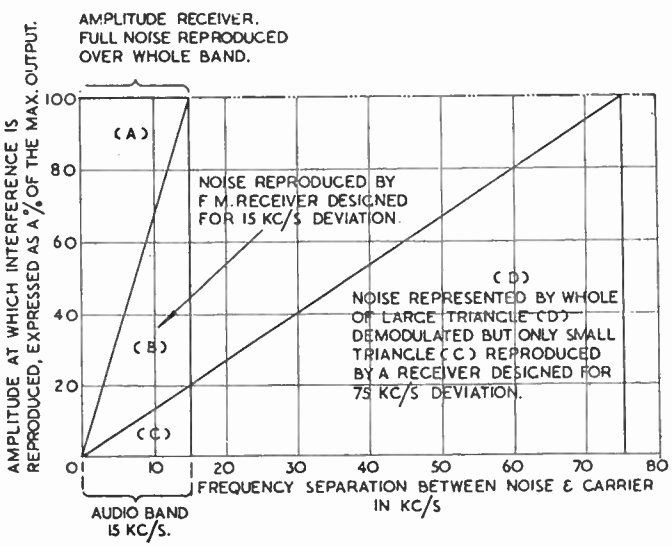



FIG. 4.2.—This diagram illustrates the frequency modulation "noise triangle". The noise spectra for two different frequency modulation deviations are compared with that of an equivalent amplitude modulation system.

*(By courtesy of the British Institute of Radio Engineers.)*

components produced are ignored, which implies that the ratio of the interfering signal to the wanted signal is less than 1 to 3.

In examining the position with deviation ratios other than unity, it becomes necessary to consider the effect of interference occurring at a frequency separation greater than the highest audio frequency. In general, such components will not be audible when either signal is unmodulated; the position when they are is discussed further below. If, however, a frequency modulation system employing frequency deviation of 75 k/cs is employed, the noise output increases with frequency separation as shown by the hypotenuse of triangle *D*. For this it will be apparent that at worst, the signal to noise ratio in the frequency modulation system is

better than that in an amplitude modulation system by a factor of 5, or 14 db.

The above result cannot be applied at random to the interference from a second station. The interfering station's frequency may, for instance, be towards the edge of the frequency band over which the carrier deviates. If it is assumed that the frequency modulation system has a deviation ratio which is greater than unity, then in the case of an amplitude modulation system, an interfering signal occurring in a comparable position would result

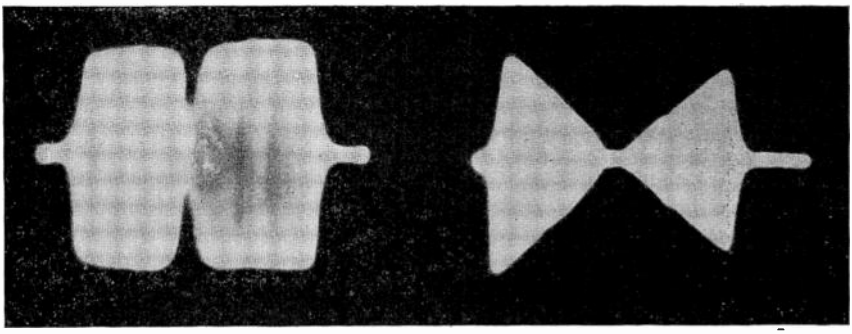


FIG. 4.3.—These oscillograms were obtained by tuning the receiver to a carrier and then manually tuning an interfering carrier signal across the i.f. channel. The first photograph—obtained on an amplitude modulation receiver—shows that the heterodyne beat has a constant amplitude over the audio frequency range; while in the case of a frequency modulation receiver the amplitude increases as the heterodyne beat frequency increases.

(*By courtesy of Murray G. Crosby.*)

in no interference whatsoever. In the case of a frequency modulation system no interference will be audible so long as the carrier is unmodulated, but upon modulation commencing the carrier frequency will swing out until the heterodyne between itself and the unwanted signal falls within the audio frequency range. Under these conditions interference is produced, which in the case of an amplitude modulation system is non-existent.

There can therefore be no question of any reduction (due to the use of frequency modulation) in the level of interference resulting from continuous wave signals separated by a difference of more than the maximum audio frequency from the carrier's unmodulated frequency. It is very difficult to place any definite value on the increased interference resulting from this cause, owing to the

difficulty in assessing the proportion of time that the frequency modulated carrier will be within audio frequency range of the interfering signal. Naturally, the basic improvement figure given above applies directly to any continuous wave interference arising from a second and unwanted carrier having a frequency which is within audio frequency range of the desired carrier.

### Effect of Impulsive Interference

The improvement in the r.m.s. signal to noise ratio of a frequency modulated system over an amplitude modulated system to impulsive interference was calculated in the last chapter, and is given by

$$\sqrt{3}\, f_d/f_c,$$

where $f_d$=frequency deviation;

$f_c$=band-width of receiver audio frequency stages.

Since we are concerned only with these components of the noise output which fall within the audible region, $f_c$ should more properly be taken as the highest audible frequency or the band-width of the receiver audio frequency stages, whichever is the less. Thus in a receiver designed for high quality broadcast reception, there is no need to restrict the band-width of the audio frequency stages of the receiver in order to realise the improvement in signal to noise ratio. In such a system the ratio $f_d/f_c$ would normally be equal to the deviation ratio, since $f_c$ is the assumed equal to the highest audible frequency, and such a system would be designed with this value of deviation ratio. Thus, for a system employing a peak deviation frequency of 75 kc/s, and an audio frequency band-width of 15 kc/s, the improvement is 19 db.

This improvement, however, takes no account of the aural improvement encountered due to the different character of the noise in the two cases. In amplitude modulation the distribution of amplitude with frequency of the components of the audio frequency output is uniform, whilst with frequency modulation the distribution of amplitude is proportional to frequency. The audible result is that ignition interference in an amplitude modulation receiver has the characteristics of a "pop", whilst in a frequency modulation receiver it has the characteristic of a "click". The latter is much less disturbing to the listener, and consequently the subjectively assessed signal to noise ratio is higher than the r.m.s. improvement would indicate.

The improvement is not constant, but deteriorates in the region where the peak interference amplitude exceeds that of the carrier at the output of the i.f. amplifier. When this occurs, the frequency modulation output includes a proportion of "pops", and the signal to noise ratio, both subjective and r.m.s., falls rapidly. However, the fall is limited by the fact that when the output comprises 50 per cent "pops", the signal to noise ratio does not deteriorate further, but tends to a constant value, as explained in the last chapter. Thus at the threshold of improvement, the f.m./a.m. signal to noise improvement ratio is at minimum; above and below the threshold the ratio increases. It is of interest to note that, below the improvement threshold, the f.m./a.m. improvement ratio will ultimately equal and then exceed the value it attains above the improvement threshold. Thus paradoxically, the f.m./a.m. improvement ratio is greatest below the improvement threshold. The importance of this fact should not, however, be over-rated, since reception under such conditions would be very poor in either system. Also, it is to be suspected that the signal level will be so low as to be vanishingly small, and may well be swamped by the receiver fluctuation noise.

In communication systems, where the highest value of signal to noise ratio is desired, it is usual to restrict the audio frequency range; an upper frequency limit of 3 kc/s is frequently adopted. The reduction of the band-width of the audio frequency stages of the receiver thus serves to increase the improvement factor for a given frequency deviation. An alternative advantage is that, for a given improvement factor, the frequency deviation employed may be reduced. This has the effect of reducing the band-width required for transmission, in itself a desirable effect. Further, the i.f. band-width may then be reduced, with a consequent lowering of the input signal level at which the threshold of improvement occurs. This point is discussed in more detail later.

### Effect of Fluctuational Noise

The major difference between impulsive and fluctuational noise is that whereas impulsive interference is discontinuous in time, fluctuation noise is continuous. Thus it is possible for a frequency modulated transmission to be intelligible when the amplitude of the impulsive interfering signal exceeds that of the carrier, whilst it is not in the case of fluctuation noise.

As shown in the last chapter, the voltage reduction in the noise level output due to the use of frequency modulation is given by

$$\sqrt{3}\, f_d/f_c,$$

where $f_d$=frequency deviation;
$f_c$=band-width of receiver audio frequency stages.

As with impulsive interference, the distribution of amplitude with frequency means that the noise is less disturbing to the listener in a frequency modulation system than an amplitude modification system, and the signal to noise ratio assessed subjectively is higher than the expression above indicates. In order to obtain the maximum improvement in a communication system, where the highest modulation frequency is less than the highest audible frequency, the band-width of the audio frequency stages should be limited to the minimum.

### Varying Carrier and Interference Amplitude

As discussed in the previous chapter, the improvement in signal to noise ratio for all three types of interference, continuous wave, impulsive and fluctuation noise, deteriorates as the interference amplitude approaches that of the carrier. In all three cases, the deterioration occurs abruptly. For continuous wave interference, with two unmodulated carriers, the signal to noise ratio in a frequency modulation system falls to zero when the carriers are of equal amplitude; this is not, however, the usual condition encountered in practice, since normally one or both carriers are modulated. This condition is discussed in more detail in the next section.

Where impulsive interference is being considered, the signal to noise ratio falls rapidly in the region where the peak magnitude of the interference i.f. output just exceeds that of the wanted carrier. Additionally, the improvement factor falls to a minimum value in this region. It should be noted that the peak magnitude of the interference is directly proportional to the receiver i.f. band-width, and thus in order to defer the fall in signal to noise ratio to the lowest possible input signal level, the receiver i.f. band-width should be restricted to the minimum possible. It is for this reason that small frequency deviations are usually employed with communication systems.

Where the interference comprises fluctuation noise, the signal to noise ratio for both a frequency modulation system and an

amplitude modulation system fall below unity when the interference amplitude at the i.f. amplifier output exceeds that of the carrier, and under these conditions it may be assumed for the present purpose that the signals in both systems become unintelligible. The input signal level at which the threshold occurs is rather difficult to determine, since it is not possible to assign any definite peak value to the noise signal, by virtue of its nature. It would, however, appear reasonable to take a crest value for fluctuation noise in the region of 4, and assuming this, the threshold would appear to occur when the peak value of the carrier exceeds the r.m.s. value of the noise by the factor of 4. The position of the threshold of improvement in this case is related to the r.m.s. value of the noise signal, and hence the level at which it occurs is proportional to the square root of the receiver i.f. band-width and not directly to the receiver i.f. band-width as in the case of impulsive interference.

The comparison between a.m. and f.m. systems is complicated by the fact that the receiver band-widths required differ appreciably, and hence the r.m.s. value of the fluctuation noise will be different for the two systems. Thus, in determining the carrier level at which the signal becomes unintelligible in the presence of fluctuation noise, it is necessary to know the receiver band-width for both amplitude modulation and frequency modulation systems. In general, by virtue of the smaller band-width necessary, an amplitude modulation system can, therefore, work in a region of lower field strength than a comparable frequency modulation system.

### Suppression of the Weaker Signal

Where two signals are working on a common frequency, or are separated by a relatively small frequency difference, the weaker signal is smothered by the stronger. This effect, which is sometimes referred to as the capture effect, is merely a special example of the signal to noise ratio improvement resulting from the use of frequency modulation.

The effect arises from the phenomenon referred to in the case of continuous wave interference; the rapid fall in the signal to noise ratio as the strength of the interfering signal becomes comparable with that of the wanted signal. Here, however, we are concerned with the converse condition, the rapid rise in signal to noise ratio

which occurs when the wanted signal strength rises above that of the interfering signal. The effect is that the interference is rapidly smothered as the difference increases, the degree of supression being far greater than might at first be suspected from the relative strengths of the two signals.

Where the frequency separation between the carriers exceeds the highest audible frequency, there is no interference in the absence of modulation of the signals. When, however, the unwanted signal is modulated, certain of its side bands will fall within audible frequency separation of the wanted signal carrier, and interference will occur; the magnitude of this interference can be assessed by the methods employed in dealing with continuous wave interference described previously. If the interfering signal is considerably stronger than the wanted signal, it may happen that one of the interfering signal side bands may be comparable in magnitude to the carrier of the wanted signal; in this case the degree of interference will be large. Where the interfering signal side band amplitudes are relatively small compared with the wanted signal carrier, the degree of interference will be small. Consequently, where interference is experienced from an adjacent modulated carrier of relatively large amplitude, the interference has a sporadic nature, being sometimes very disturbing and at others only just perceptible. Interference can also, of course, occur if the interfering signal is unmodulated and is situated within the passband of the receiver; in this case, the side bands of the wanted signal adjacent to the interfering carrier beat with the interfering carrier to produce an interference output. Interference here is, however, generally less perceptible because the amplitudes of the side bands concerned will in general be relatively small and hence also the noise output.

Interference from a modulated carrier generally takes the form of a rasping sound; this is because the side bands of the interfering signal do not in general straddle the frequency of the wanted carrier symmetrically. Consequently the audible output due to each interfering side band individually is not harmonically related to the others.

Within the zone in which two frequency modulated stations are operating on the same channel, and are being received at approximately the same strength, neither station will have any programme value. A ratio of 3 : 1 between the stronger and

weaker signal is sufficient to avoid intelligible break-through, and interference takes the form of an unintelligible rasping sound. A ratio of some 30 : 1 is required for complete elimination of this interference.

### The Threshold of Improvement

The threshold of improvement is reached when the carrier and interference voltages attain the same amplitude. As the noise amplitude is dependent on the receiver band-width, it follows that if equal carrier voltages are fed into two channels of different width, the noise voltage will—if the carrier voltages are gradually reduced—equal the carrier amplitude in the wider channel first. As a result, in the case of a large deviation ratio system, a larger carrier voltage is necessary to reach the threshold of improvement than is necessary in the case of a small deviation ratio system. At low carrier levels the signal to noise ratio may be above the improvement threshold on a low deviation system and below it on a high deviation system. Under these conditions the low deviation ratio system will produce a better output signal to noise ratio than a high deviation ratio system.

In assessing the actual receiver band-width which is necessary in order to pass the full intelligence at any particular deviation ratio, it has already been shown that the receiver band-width cannot simply be increased in proportion to the deviation ratio. Fig. 2.14 indicated that as the deviation ratio is increased the "relative" receiver passband necessary to accommodate the significant frequency modulation side bands is reduced.

From this it follows that the receiver band-width, and with it the noise voltage, is greater than that of a comparable amplitude modulation receiver by an amount equal to the deviation ratio multiplied by one half of a factor which is derived from Fig. 2.14 and its associated text. In order to spare the reader the labour of turning back to Chapter Two, the resultant increase in noise voltage is plotted in Fig. 4.4.

As an example of the effect of this increased noise level the case of a system having a deviation ratio of 4 will be examined. It will be noted that in comparison with a system having a deviation ratio of unity, the impulsive noise peak amplitude has been increased by $\frac{6}{2{\cdot}4} = 2{\cdot}5$ times, while the carrier amplitude

remains unchanged. The threshold of improvement will therefore be reached when the carrier is some 2·5 times the amplitude at which equality would have occurred in the case of a deviation ratio of unity.

In the case of impulsive noise the increase in carrier amplitude at which the threshold of improvement will occur is proportional to the receiver band-width, and may be derived directly from

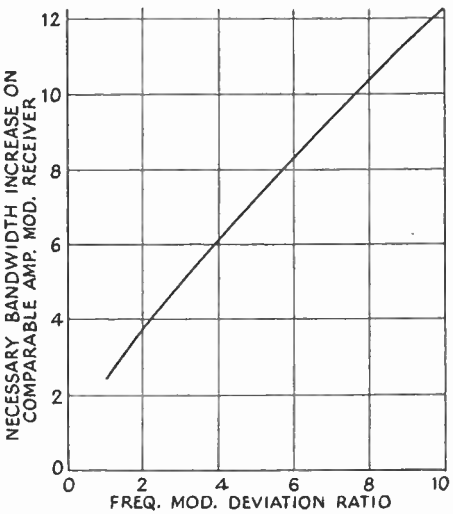



FIG. 4.4.—The increase in band-width above that of a comparable amplitude modulation receiver, which is necessary to accommodate the frequency modulation side bands at various deviation ratios.

Fig. 4.4. It should be noted that any comparison in amplitudes must be that between the peak voltages reached respectively by the carrier and the impulsive interference.

The amplitude of fluctuational noise is proportional to the square root of the receiver band-width. The ratio of the levels at which the threshold of improvement will occur may therefore be obtained by taking the square root of the band-width increase given in Fig. 4.4. Taking the example of a deviation ratio of 4, the threshold of improvement will, in the case of fluctuational noise, occur at some $\sqrt{2{\cdot}5}=1{\cdot}6$ times the carrier amplitude at which it would have occurred in the case of a deviation ratio of unity. In determining the actual magnitude of the carrier at the threshold of improvement, it is necessary to allow for the "Crest Factor" of the fluctuation noise voltage.

It was shown in the last chapter that the peak to mean ratio for fluctuational noise may be taken as 4:1. In the case of a sinusoidal signal such as that of the carrier wave it is, however, only $\sqrt{2}$:1; so that for equal peak voltages the root-mean-square value of the carrier must be greater than that of the fluctuational noise by a ratio of $4:\sqrt{2}$ or some 2·8:1 (i.e. 9 db). It therefore follows that the threshold of improvement will be reached when the mean carrier voltage is 2·8 times greater than the mean fluctuational noise-level.

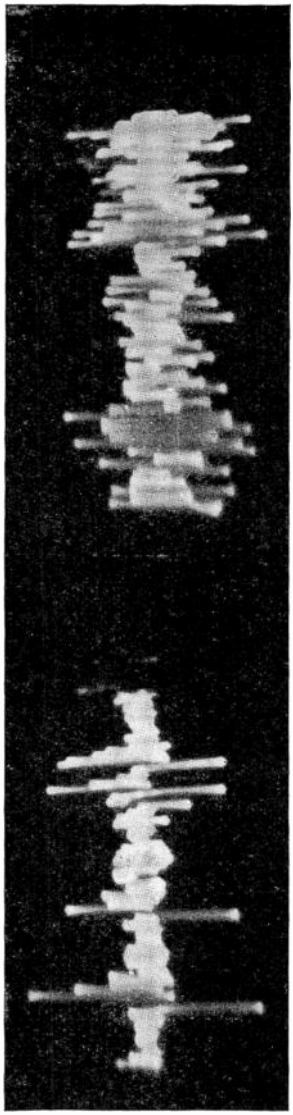


FIG. 4.5.—The first oscillogram shows the fluctuational noise wave-form in the region of the threshold of improvement, while the second illustrates the fluctuational noise output from an amplitude modulation receiver.
*(By courtesy of Murray G. Crosby.)*

The point at which the threshold of improvement occurs has a rather distinctive sound to the ear. With impulsive interference the presence of "pops" in the output is very noticeable; with fluctuation noise, the quality of the "hiss" takes on a more intermittent character, somewhat like that of ignition interference. This latter phenomenon is probably due to the generation of "pops" in the output by the same mechanism as discussed in the case of ignition interference. These aura phenomena provide a good practical indication of the level at which the threshold of improvement occurs.

Fig 4.5 shows oscillograms of the fluctuation noise output of both frequency and amplitude modulation receivers, with the signal to noise ratio adjusted to the sputter point. For

high-fidelity entertainment the desired carrier peak amplitude should exceed that of the peak fluctuation noise-level by at least 13 db, and for speech of high intelligibility, by some 6 db.

The fact that the interference amplitude reaches that of the carrier at a higher field strength-level, as the deviation ratio is increased, results in a reduction to the effective service area of a frequency modulation station of any given power output. As the deviation ratio is increased the boundaries to the service area tend to become very much more sharply defined; reception in this region does not gradually become worse, it suddenly becomes impossible. With the exception of the increased band-width required for transmission, the increased carrier-level at which the threshold of improvement occurs is the only factor which tends to place a limit on the maximum deviation ratio employed.

### Pre-emphasis

It has earlier been shown that the triangular noise distribution of a frequency modulation system results in a progressive increase in the amplitude at which noise is reproduced, from the lower to the higher audio frequencies. This noise distribution is not entirely satisfactory, resulting as it does in the smothering of the upper audio frequencies while the lower still possess a reasonable signal to noise ratio. The position is aggravated as most programme material results in the greatest modulation depths in the band below 1,000 cycles, while above this band the average modulation depth steadily decreases. It is, however, the presence of the relatively small percentage of energy contained in the upper audio frequency band which results in a high standard of reproduction fidelity. The position is therefore somewhat unsatisfactory, the noise being reproduced at the greatest amplitude in the region in which the desired programme material has a minimum amplitude.

From the standpoint of high-fidelity reproduction, the importance of the means employed to produce an effectively constant noise-level distribution over the audio band will at once be apparent. This result is achieved by the accentuating or emphasising, before transmission, of the higher audio frequencies. At the receiver the complementary de-emphasis or restoration to normal is effected by a special filter. This filter normally takes the form of a simple resistance and condenser network connected across the discriminator output and directly preceding the audio amplifier.

The de-emphasis filter attenuates the interference as well as the higher audio frequency components, with the result that while the programme material is merely restored to its original form, a very considerable reduction is made in the level at which the interference is reproduced.

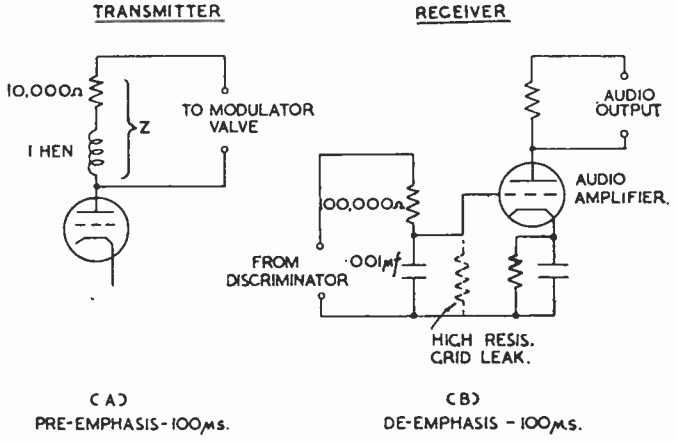



Fig. 4.6.—Pre-emphasis and de-emphasis filters.
(*By courtesy of the British Institute of Radio Engineers.*)

In America the Federal Communication Commission have drawn up standards for the pre-emphasis of a frequency modulated transmission. They originally laid down that it should be standard to pre-emphasise a sound transmission in accordance with the impedance frequency characteristic of a series inductance resistance network having a time constant of 100 micro-seconds. Later this was revised to 75 micro-seconds. The BBC favour a still smaller amount of pre-emphasis; 50 micro-seconds. The arrangements by which a 100 micro-second circuit can be employed both for the pre-emphasis of the audio signal before it is applied to the modulator, and for the de-emphasis of the received signal, are shown in Fig. 4.6. The frequency characteristic of a 100-micro-second pre-emphasis filter is shown in Fig. 4.7, together with those of a 75- and 50-micro-second filter. The figure also shows the corresponding de-emphasis filter curves.

With the uniformly rising noise spectrum associated with a frequency modulation system, the amplitude of the noise at any given frequency $f$ is given by $a\omega$ where $\omega = 2\pi f$, and $a$ is a constant. The effect of a de-emphasis filter is to reduce this in magnitude

by a factor $1/(1+\omega^2C^2R^2)^{\frac{1}{2}}$, where $R$ is the series resistance of the filter, and $C$ the shunt capacitance. As $CR$ is the time constant of the filter, the factor may be written as $1/(1+\omega^2T^2)^{\frac{1}{2}}$. The output noise amplitude at any frequency is thus given by $a\omega/(1+\omega^2T^2)^{\frac{1}{2}}$. This result is shown graphically in Fig. 4.8, for three values of

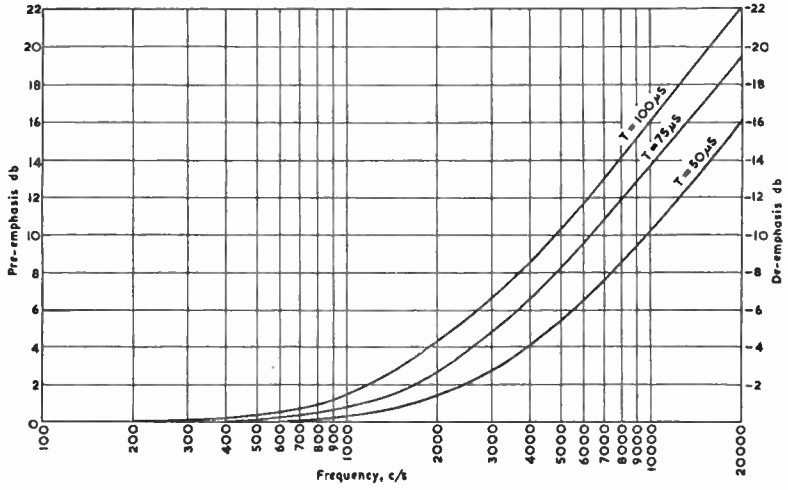



FIG. 4.7—Pre-emphasis and de-emphasis filter responses, for time constants of 50, 75 and 100 micro-seconds.

$T$, 50, 75, and 100 micro-seconds. It will be seen that at very high frequencies the amplitude approaches the constant value $a/T$. That is, the noise output spectrum is similar to that of an amplitude modulation receiver.

The overall effect of de-emphasis may be found by comparing the mean square values of comparative curves of Fig. 4.8. With no de-emphasis the mean square value is proportional to

$$\int_0^{\omega_c} a^2\omega^2 d\omega = a^2\omega_c{}^3/3,$$

where $\omega_c$ is the upper limit of audibility, or the high-frequency cut-off of the receiver a.f. amplifier.

With de-emphasis the mean square value is proportional to

$$\int_0^{\omega_c} a^2\omega^2/(1+\omega^2T^2)d\omega = a^2(\omega_cT-\tan^{-1}\omega_cT)/T^3.$$

The improvement due to de-emphasis is thus

$$\omega_c{}^3T^3/3(\omega_cT-\tan^{-1}\omega_cT).$$

This result is shown graphically in Fig. 4.9, with $f_cT=\omega_cT/2\pi$ as parameter. If an upper limit of audibility of 15 kc/s is assumed, the improvement is thus some 15·5 db for $T=100$ micro-seconds, 13 db for $T=75$ micro-seconds and 10 db for $T=50$ micro-seconds.

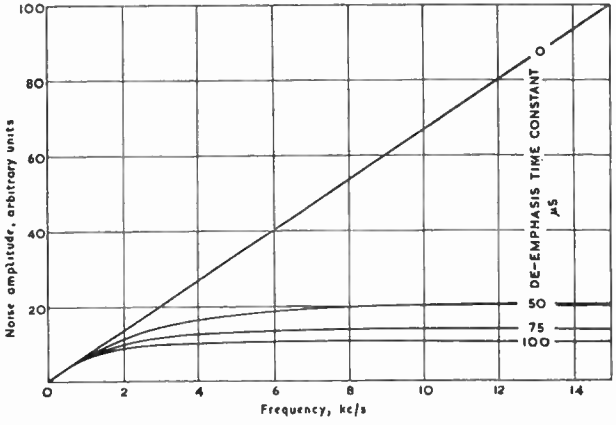



FIG. 4.8—The distribution of noise output with frequency, with and without de-emphasis.

It should, however, be noted that the audible improvement would probably be not so great as the figures would suggest. This because the major part of the improvement is due to the very substantial reduction of noise output at the very high frequencies. The ear is somewhat insensitive to noise in this region, the aural perception of noise being largely governed by the magnitude of the noise output below 5 kc/s.

Since the noise output decreases as $T$ increases, it would appear to be advantageous to make $T$ as large as possible. However, an upper limit is set by the fact that, if too much pre-emphasis is applied at the transmitter, the upper audio-frequency components may become so large as to cause over-modulation. The general modulation level would then have to be reduced to offset this effect. The time constant chosen thus has to be a compromise value.

The figures for the reduction in modulation level determined by various workers do not agree too well. Tests by Crosby using 100 micro-seconds pre-emphasis suggested that an average reduction of 2·5 db was required, although with certain types of instruments, including guitar, harmonica and piano, it was found

that a figure of 4·5 db was more appropriate. This would suggest the overall improvement for 100 micro-seconds pre-emphasis should be about 13 db.

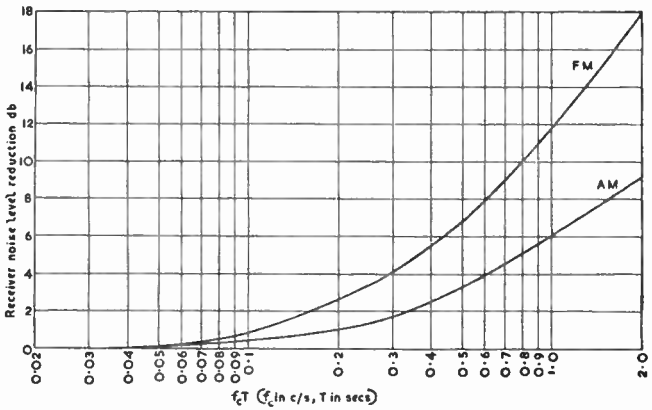



FIG. 4.9—The reduction in receiver noise output for f.m. and a.m. systems with audio frequency band-width ($f_c$) × time constant ($T$).

The results of listening tests carried out by the BBC using various amounts of pre-emphasis and with different types of programme do not agree too well with Crosby's findings. During the BBC tests it was found that a 12 db reduction in audible receiver noise and a 6·5 db reduction in ignition interference was obtained when using 100 micro-second pre-emphasis. To avoid transmitter distortion it was necessary to reduce modulation in some cases by as much as 12 db and generally by 6·5 db. The resulting gain in signal to noise ratio was therefore stated to be only 5·5 db in the case of fluctuational noise and zero in the case of ignition noise. The BBC, however, state that when using 50 micro-second pre-emphasis, the fluctuational noise was reduced by 7·5 db, and the ignition noise by 4·5 db, while it was only necessary to reduce the transmitter modulation by 3 db. The resulting overall gain due to the use of 50 micro-second pre-emphasis is therefore stated to be 4·5 db for fluctuational noise and 1·5 db for impulsive noise.

From Fig. 4.9 it will be seen that the improvement due to de-emphasis is very small for small values of $f_cT$. It is for this reason that pre-emphasis is not used in narrow band commercial frequency modulation radio-telephone links. In such links the

upper frequency limit is usually 3 kc/s. If a pre-emphasis time constant of 100 micro-seconds were used, the receiver noise reduction would be some 4·5 db. However, the highest audio-frequencies transmitted would be "boosted" by some 7 db. As a substantial part of speech energy exists in this upper frequency region, the modulation level would have to be reduced considerably. The overall improvement would then be very small, and might even be negative.

Pre-emphasis and de-emphasis may also be used in an amplitude modulation system. The improvement effected may be calculated as before by a comparison of mean square values. For no de-emphasis, the receiver mean square noise output is proportional to

$$\int_0^{\omega_c} a^2 d\omega = a^2 \omega_c.$$

With de-emphasis this becomes

$$\int_0^{\omega_c} a^2/(1+\omega_c{}^2T^2)d\omega = a^2 \tan^{-1}\omega_c T/T.$$

The improvement is thus $\omega_c T/\tan^{-1}\omega_c T$. This is plotted in Fig. 4.9 with $f_c T = \omega_c T/2\pi$ as parameter. It will be seen from the curves of this figure that the improvement in an amplitude modulation system is considerably less than that in a frequency modulation one. Since the same considerations of reduction of the transmitter modulation level with pre-emphasis apply to both cases, it will be apparent that the employment of pre-emphasis with an amplitude modulation system will give only a small improvement at best.

### Noise Reduction at the Transmitter

One of the most interesting features about a frequency modulated transmitter is its high efficiency in comparison with its amplitude modulated counterpart. With an amplitude modulated transmitter the peak power output at 100 per cent modulation rises to four times the unmodulated carrier power. Should the peak power be the output limiting factor, then the changeover to frequency modulation will allow the carrier power to be increased by four times or some 6 db. If, however, the limitation is on the maximum mean power output, then the permissible increase is only twice, or some 3 db (assuming the most rigorous conditions—a square wave modulation).

From these figures it follows that for any given size of output

valve or power consumption a greater power output can always be obtained by substituting frequency modulation in place of amplitude modulation. This improved output is of very real importance when it is borne in mind that at very high frequencies the output efficiency of large transmitting valves starts to fall fairly rapidly due to the incidence of transit time.

The reason for the reduced energy content of a frequency modulated carrier will be readily appreciated if it is noted that with an amplitude modulated transmission the carrier amplitude is twice that of the side bands at 100 per cent modulation; while in the case of a frequency modulated system the carrier amplitude is almost always less than that of the side bands and can fall to zero with certain deviation ratios.

### Aural Noise Rejection

Quite apart from the various improvements which it is possible to calculate with some accuracy, there is another which is somewhat less tangible. Listening tests carried out by Crosby and confirmed by the BBC have shown that more noise can be tolerated with the triangular frequency modulation noise spectrum than with the rectangular noise spectrum of an amplitude modulation system. This gain would seem to be a measure of the ability of the human ear to unconsciously separate the desired intelligence from interference which is distributed in two different ways. It may in part be due to the fact that the human ear does not give the same sensation of loudness for noises of equal volume but having different frequencies.

In the first instance—that of amplitude modulation—the noise is distributed evenly over the whole audio frequency band, including that part which is conveying the desired intelligence frequencies. In the second form of noise distribution—that of frequency modulation—the noise is concentrated towards the upper end of the audio frequency range. Under these conditions the desired intelligence is concentrated towards the lower end of the audio frequency range, while the interfering signals are concentrated at the upper end. In this way the ear is provided with a natural means of discriminating against the noise and in favour of the desired signal.

The additional noise which can be tolerated in this way is of considerable importance, when an unpre-emphasised system—such as a radio telephone link—is under review.

### Examples of the Improvements due to Frequency Modulation

Taking as an example the case of a system with an audio frequency response up to 15 kc/s, a deviation ratio of 5, the improvement in signal to noise ratio over a comparable amplitude modulation system is as follows:

Basic impulsive and fluctuation noise reduction $= \sqrt{3} \times 5 = 8{\cdot}5$ times in voltage or 19 db.

Gain due to increased transmitter efficiency = 3 db.

Total improvement = 22 db for impulsive and fluctuation noise.

Additionally, this figure is improved by the gain due to the employment of pre-emphasis.

The increase in band-width as ascertained from Fig. 4.4 shows that the impulsive noise voltage reaches that of the carrier, when the carrier is some 7·2 times or 17·2 db above the level at which equality would have been reached in the case of a comparable amplitude modulation system. However, in the case of fluctuational noise the threshold of improvement is reached when the carrier amplitude is $\sqrt{7{\cdot}2}$ or only some 2·7 times or 8·6 db greater.

## SELECTED REFERENCES


Snow, W. B., Audible Frequency Ranges of Music, Speech and Noise, *Bell System Monograph B-591*. Also *J. Acoustical Soc. of America*, July 1931.

Fletcher, Harvey, Physical Characteristics of Speech and Music, *Bell System Technical Journal*, July 1931.

Armstrong, Edwin H., A Method of Reducing Disturbances in Radio Signalling by a System of Frequency Modulation, *Proc. I.R.E.*, May 1936.

Crosby, Murray G., Frequency Modulation Noise Characteristics, *Proc. I.R.E.*, April 1937.

Crosby, Murray G., The Service Range of Frequency Modulation, *R.C.A. Review*, January 1940.

Fletcher, Harvey, Hearing the Determining Factor for High-Fidelity Transmission, *Proc. I.R.E.*, June 1942.

Bell, D. A., Frequency Modulation Communication Systems, *Wireless Engineer*, May 1943.

Tibbs, C. E., A Review of Wideband Frequency Modulation Technique, *J. Brit. I.R.E.*, September 1944.

Kirke, H. L., Frequency Modulation: B.B.C. Field Trials, *The B.B.C. Quarterly*, July 1946.

Standards of Good Engineering Practice Concerning Frequency Modulation Broadcast Stations, *Federal Communications Commission*.

*Chapter Five*

# FREQUENCY MODULATION PROPAGATION

THE question whether a particular frequency band is suitable or not for the propagation of a given signal is determined by a number of factors, of which the most important are:

The distance over which the signal has to be transmitted.
The amount of distortion or interference which can be tolerated.
The frequency band-width necessary to accommodate the signal's side band spectrum.

If on a given band a frequency modulated transmission suffers more severe distortion than an equivalent amplitude modulated transmission, then the fact that the frequency modulated system may show substantial reductions in the received noise-level will probably be more than cancelled by its increased susceptibility to distortion. In the last chapter the improvement in signal to noise ratio was discussed at some length. In this chapter it is proposed to start by examining the increased susceptibility of a frequency modulated signal to certain forms of distortion, and in so doing to determine those bands which are best suited to the transmission of this type of signal.

Practically all distortion which occurs during the propagation of a radio signal has as its root cause the fact that there are several possible paths which that signal may follow in its course from the transmitter to the receiver. The received signal is therefore almost always made up of a number of components which have arrived by different routes. With the exception of the wave which travels directly from the transmitter to the receiver, all other wave components will have been reflected by one medium or another. The frequency band on which the signal is transmitted determines to a very large extent the medium which is most liable to be responsible for these reflections. The principal sources may be very broadly classified under three main headings:

Reflections due to the ionised layers.

Reflections from the boundary layer between two different air masses.

Reflections from solid objects, such as mountains, buildings, gas-holders, or aircraft in flight.

The very fact that one wave may have travelled directly from the transmitter to the receiver, while another has followed an indirect path, makes it clear that the two paths must have different lengths. It is therefore apparent that the various component waves reaching the receiver will, although all having the same frequency, vary in their phase relationships to each other. It can so happen that at one moment all the various component waves will be adding together and that at the next, as a result of changed conditions in the reflecting medium, they will all be tending to cancel each other out. The practical result of this occurrence is normally termed fading.

### Frequency Bands Employed for Frequency Modulation Transmissions

As stated earlier, conditions are most favourable for the employment of wide band frequency modulation at those frequencies where ionospheric reflections are of minor importance. There is thus a low useful frequency limit, and this may generally be taken to exist at about 30 Mc/s, i.e. the lower limit of the very high frequency band. Above 30 Mc/s, the range of frequencies up to 30,000 Mc/s is divided into three groups for the purposes of classification. These groups are as follows:

1. V.H.F. (very high frequency) 30–300 Mc/s.
2. U.H.F. (ultra high frequency) 300–3,000 Mc/s.
3. S.H.F. (super high frequency) 3,000–30,000 Mc/s.

The upper limit of 30,000 Mc/s is arbitrarily taken; at the present time frequencies beyond this limit are not actively employed for communication.

Within each of these bands there are sub-divisions allocated by International Agreement to various classes of service. A list of the broadcast frequency bands are given below; these allocations were determined at the Altantic City Conference in 1947. For the purposes of these allocations, the world is divided into three regions,

broadly as follows. Region 1, Europe, North Africa and Near East; Region 2, North and South America; Region 3, Asia and Australasia.

TABLE 3

| Region 1 | Region 2 | Region 3 |
|---|---|---|
| Mc/s | Mc/s | Mc/s |
| 41–68 | 44–50 | 44–50 |
| | 54–72 | 54–72 |
| 87·5–100 | 76–108 | 87–108 |
| 174–216 | 174–216 | 170–200 |
| 470–585 | 470–940 | 470–585 |
| 940–960 | | 940–960 |

Of the three major frequency bands, the v.h.f. band has been developed rapidly, the period when it is fully utilised appears to be not far removed. In the u.h.f. band, development is proceeding, particularly in the utilisation of the broadcast frequency allocations. The s.h.f. band is being used particularly for short distance radio links and radar applications. Frequency modulation is frequently used for broadcast-chain links in this band, since the type of valve frequently employed for the generation of oscillations in this band, the reflex klystron, lends itself to this type of modulation.

In a short survey of the type here attempted, it is impossible to deal fully with the propagation characteristics in all three bands, and we shall confine ourselves to an account of the more important phenomena associated with the v.h.f. band and the lower portions of the u.h.f. band.

### Selective Fading

Not only is it necessary to consider the effect of variations in the reflected path length, but account must also be taken of the fact that a propagated radio signal is made up of a spectrum of component frequencies occupying a band of slightly differing wavelengths. Assume for a moment that the lengths of the paths followed by the various component waves are held constant, as in fact they would be if there was a direct wave combining with

a reflection from a static object such as a large building or a gas-holder. It is apparent that with two different path lengths there will be at least one signal wavelength at which the number of waves along the direct and reflected routes will be such that they will arrive in phase with each other, so adding directly together. At the same time it is equally apparent that there will be an adjacent signal wavelength at which the two paths will, when expressed in terms of wavelengths, be one-half wavelength either longer or shorter than they were before. Under these conditions the direct and indirect signals will be received in such a phase relationship that they will be tending to cancel one another out.

When these two conditions both occur within the band of frequencies which are being used for a single transmission it is said that selective fading is taking place. Expressed in different words, it may be stated that some of the signal's side bands are being received at a far larger relative amplitude than others. As a result the amplitude relationship of the various side bands is distorted.

Selective fading is by far the most serious form of distortion which is caused during the propagation of a wave. Variations in signal strength due to straightforward fading may be readily corrected at the receiver by means of automatic volume control, or, in the case of frequency modulation, by means of the limiter stage. There is, however, no completely satisfactory answer to selective fading. This form of fading becomes most pronounced when the difference, expressed in wavelengths between the direct and reflected paths, is considerable. On the short waveband where such large differences exist, it is almost entirely responsible for the very low fidelity reproduction which is normally associated with transmission on this band.

### Ionospheric Reflections

On the short waveband between some 10 and 150 metres, signals are reflected back to earth from the various layers of ionised air which exist at considerable height above the earth's surface. Physics has provided a picture of the way in which these layers are formed. The ionising agent is the ultra-violet wavelength light from the sun. This light is absorbed during the production of the ionised air so that the nearer we approach the

earth the feebler the effective ultra-violet radiations will become. On the other hand, the nearer we approach the earth the more molecules there are for a given volume of air, which in turn leads to an increase in the number of ions. These two effects will work against one another, so that it is reasonable to expect a maximum effect at one particular height. There will be little production of ions in the most rarefied air very remote from the earth, where the radiation is strong, but where there are few molecules for it to encounter. There will also be a small production of ions near the surface of the earth, where there are plenty of molecules, but where there is very little effective ultra-violet radiation left. To explain the effect fully it is necessary to take into account the recombination of the ions and electrons, and the rotation of the earth. When this is done, as it has been by Chapman, it is found that there is a very satisfactory agreement between theoretical calculations and the observations of the ionised layers.

The two principal layers in the ionosphere have been named "E" and "F"; the latter, however, often separates into two parts, which are termed the $F_1$ and $F_2$ layers respectively. These names were introduced by Appleton, who discovered the $F$ layer. The lower or $E$ layer remains constant in height, at about 100 km., although its density varies with the sun's altitude. Thus, it is most dense at midday in the summer, and has minimum density during a winter's night. Similarly, the height of the $F_1$ layer remains constant at about 200 km., although it is not always observable as a separate layer because it merges with the $F_2$ layer at various times. As far as can be judged, its density, like that of the $E$ layer, is directly correlated to the sun's altitude. As both the $E$ and $F_1$ layers are absent during winter periods in the Polar regions, it is assumed that both these layers are entirely due to the sun's ultra-violet radiation.

The highest, or $F_2$ layer does not appear to be caused entirely by ultra-violet radiation, as it is observable in the Polar regions during the winter months. Although the complex behaviour of this layer makes a theoretical explanation rather difficult, it is possible to base a fairly satisfactory explanation on the general physics of ionising radiations.

The path followed by a wave on leaving the transmitter is determined by the extent to which the refractive index of the upper atmosphere departs from unity, as a result of the ionised

layers. The refractive index of ionised air relative to un-ionised air may be determined from the formula:

$$\mu = \sqrt{1 - 8{\cdot}1 \times 10^7 \frac{N}{f^2}}, \qquad . \quad . \quad . \quad (5.1)$$

where $N$ = the electron density or number of electrons per cubic centimetre of the ionised layer;
$f$ = the frequency of the signal being considered.

The actual bending effect produced by any given difference in refractive index may be readily determined by the use of normal

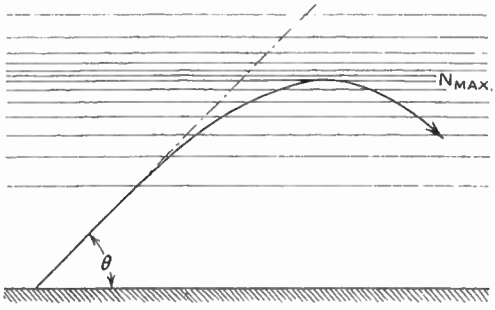



FIG. 5.1.—As the number of electrons increases with height the wave will be bent so that it returns to earth.

optical formula. It is apparent that a wave entering the zone of ionised air (i.e. air having a higher refractive index) will be bent away from the normal and towards the earth's surface, as shown in Fig. 5.1. As the number of electrons increases with height then the wave will travel in a curve, and if the bending effect produced is sufficient it will be directed back to the earth. It is impossible accurately to trace out the path of a wave as it passes through the ionosphere due to the wide variations in electron density with the time of day and other factors.

As the wave enters the ionosphere at a steeper angle it is apparent that it must be bent through a larger angle before it can be returned to earth. No wave will be returned above a certain angle, which is given by:

$$\cos\theta = \sqrt{1 - 8{\cdot}1 \times 10^7 \frac{N_{max}}{f^2}}. \qquad . \quad . \quad . \quad (5.2)$$

It should be noted that if the frequency is low enough $\theta$ may be 90°. This fact is used to determine the value of $N_{max}$ experimentally. If a receiver is situated beside a transmitter, so that $\theta = 90°$, and the frequency is raised steadily, the highest frequency $f_{max}$ received by reflection gives the value of $N_{max}$ from $f^2_{max} = 8{\cdot}1 \times 10^7 N_{max}$. As the frequency is raised a value will be eventually reached from which no wave will be returned by the ionosphere, however oblique the incident. A consideration of the geometry involved will show that, because of the earth's curvature, even a wave which leaves the transmitter tangentially to the earth's surface cannot enter the ionosphere at less than a certain value of $\theta$, which is dependent upon the height of the lower edge of the ionosphere. It will be seen that if a ray travels in a straight line it will make a continually increasing angle with the tangent of the earth below. The smallest angle at which a ray can enter the *E* layer is about 8°, and with the *F* layer about 14°.

To arrive at some idea of the actual frequencies involved in this type of reflection, let it be assumed that $N_{max}$ (the maximum electron density) is $4 \times 10^5$ free electrons per c.c.; then it follows from equation (5·2) that the highest frequency which would be returned for $\theta = 10°$ will be 33 Mc/s, whilst for a vertical incident, frequencies above 5·7 Mc/s would penetrate the ionised layer and escape into free space. It should be noted that these values are only approximate due to the simplified theory. The whole subject has been treated in considerable detail by A. W. Ladner and C. R. Stoner in *Short Wave Wireless Communication.*

### Effect of Ionospheric Reflections

Having now obtained some idea of the nature of the reflections produced by the ionised layers, it is possible to pass on to the consideration of the effect which these reflections will have on frequency modulated signals. In order to do this it is proposed to take an actual example illustrating the distortion which results from selective fading. At one particular frequency, say 15 Mc/s (20 metres), the signals arriving by two different paths may add directly together. If it is assumed that the one path is 30,000 metres longer than the other, then there will be some $\frac{30{,}000}{20} = 1{,}500$ wavelengths extra along the longer path. It will readily be seen that the two signals will be exactly out of phase if there are only

1,499·5 wavelengths extra along the longer path. This will occur if the signal wavelength is altered to $\frac{30,000}{1,499·5}=20·01$ Mc/s; or only 10 kc/s away from the frequency at which the two signals arrive exactly in phase with each other. Although in actual practice the position is considerably more complicated than this it is not unusual for there to be maximum and minimum fading amplitudes even closer than 10 kc/s.

Following on from the above example, it will readily be seen that, if the carrier is frequency modulated with a deviation of ±50 kc/s, then during each cycle of modulation the phase relations of the signals arriving by the two alternative paths will pass through no fewer than five complete cycles of phase reversal. In this particular example the selective fading will result in the fifth harmonic of the audio signal being added to the demodulated intelligence. Other deviations would, of course, result in the production of different harmonics. In short, it will be seen that selective fading will produce a type of distortion very similar to that produced when an amplitude modulated programme is transmitted on the short waveband. As, however, the band occupied by a frequency modulated channel is very considerably greater than that necessary for a comparable amplitude modulated transmission, the susceptibility of the frequency modulated signal to selective fading will be proportionally greater.

It may therefore be stated that a high-fidelity frequency-modulated system is impractical on any band which employs the ionosphere as part of its transmission medium. It should, however, be noted that if a frequency modulation system employs such a small frequency band for its transmission that it is not unduly distorted by violent selective fading, then it may be operated on the short waveband and will in fact show all the advantages normally associated with frequency modulated transmission. In Chapter Eleven it is shown that there are a number of services which do in fact fall into this category, notably sub-carrier picture telegraphy and high-speed telegraphy.

Under normal conditions, the reflections due to ionised layers cease between 30 and 40 Mc/s, and since most frequency modulation systems work at frequencies above this limit, it may be assumed that ionospheric reflection is not important in the propagation of such transmissions.

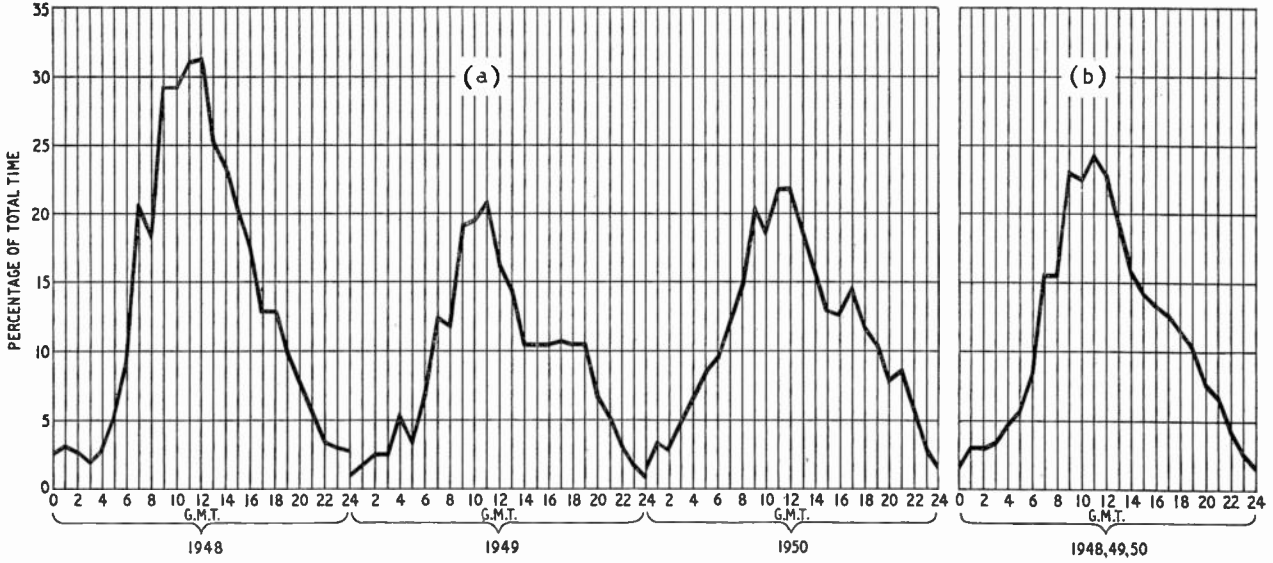



FIG. 5.2.—(*a*) Mean annual 24-hour distribution of Sporadic *E* at Slough, showing percentage of time when frequency of reflection at vertical incidence exceeds 5 Mc/s. (*b*) Mean 24-hour distribution during period 1948, 48, 50.

(*By courtesy of "Wireless World".*)

As a corollary, it may be stated that ignoring the effect of ionospheric reflection, the service area of such a transmitter is determined, therefore, by the area in which the direct ray can be received; outside this area, reception is not normally possible. Thus two stations whose service areas do not overlap may share a common frequency without mutual interference normally occurring. Due, however, to the appearance of occasional regions of abnormally high electron density in the *E* layer, reflections at frequencies higher than those normally affected may result.

These regions of high electron density in the *E* layer are generally of small area, and are of a random and intermittent character; for this reason the phenomenon is termed "Sporadic *E*". T. W. Bennington, in a survey of the subject, states that reflection at oblique incidence occurs frequently on frequencies exceeding 30 Mc/s, and on remote occasions reflections may occur at frequencies approaching 100 Mc/s. The maximum range of interference from this type of reflection is normally limited to 1,400 miles, i.e. one hop, since more than one reflection requires the simultaneous existence of particular Sporadic *E* patches at widely distributed geographical points, a remote possibility.

Bennington cites three distinct types of Sporadic *E*. The first, occurring in high latitudes, is clearly associated with ionospheric and magnetic disturbances, and with auroral activity; its occurrence and intensity has a maximum around midnight and minimum around noon. The second, and most important type, occurs in middle latitudes, and is not related to ionospheric or magnetic disturbances; its maximum occurs around noon, and its minimum at night. Additionally, it exhibits marked seasonal variations, being maximum at mid-summer and minimum during the winter. The first two types are not rigidly confined to the areas specified, and that normally associated with one region is frequently encountered within the other. The third type is observed only in lower latitudes; little information is available about its occurrence.

The daily and seasonal variations of the type of Sporadic *E* encountered in middle latitudes is shown in Figs. 5.2 and 5.3. These figures are based on measurements made by the Slough Station of the Department of Scientific and Industrial Research, and show the percentage of time reflections at vertical incidence for frequencies greater than 5 Mc/s occurred in the period 1948–50. With vertical reflections at 5 Mc/s, it is to be expected that the

maximum frequency at which reflection would occur would be in the region of 26 Mc/s. This would occur with a horizontally propagated wave, and consequently interference could be expected at a range of 1,400 miles. At shorter distances, the

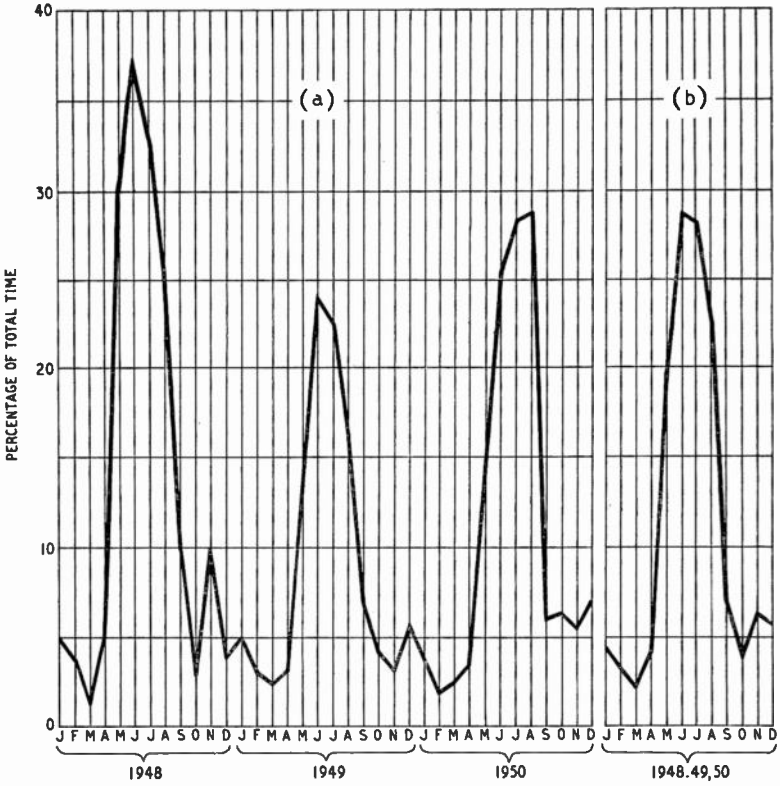



FIG. 5.3.—(*a*) Monthly distribution of Sporadic *E* at Slough, showing percentage of time when frequency of reflection at vertical incidence exceeds 5 Mc/s. (*b*) Mean monthly distribution for years 1948, 49, 50.

(*By courtesy of "Wireless World".*)

maximum frequency on which interference would be expected is correspondingly reduced. Bennington has analysed the results to show the mean percentage of time when Sporadic *E* would sustain propagation over a range of 1,400 miles during daytime in the summer months May–August. This is shown graphically in Fig. 5.4; it will be seen that for frequencies above 100 Mc/s the

percentage of time is vanishingly small. This graph shows the limiting case, since for smaller distances the percentages are correspondingly smaller.

Interference may also be experienced at frequencies up to about 50 Mc/s due to reflections by the $F_2$ layer. Such occasions are infrequent, and are generally associated with periods of maximum sun-spot activity. Reflection occurs in the main with signals

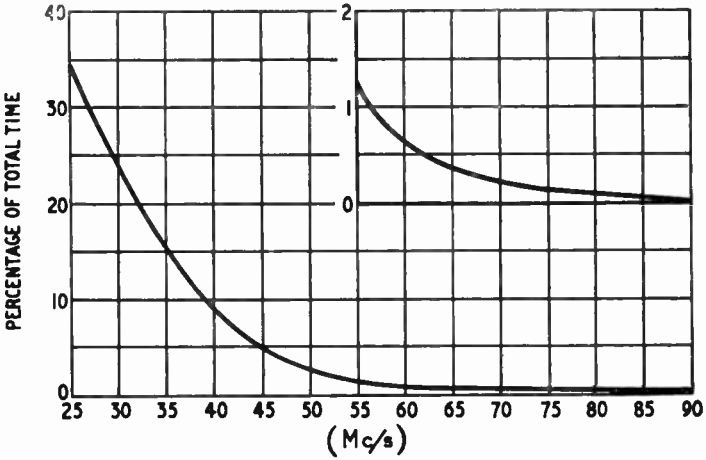



FIG. 5.4—Mean percentage of time when Sporadic *E* would sustain propagation over 1,400 miles during daytime in summer months (May–August inclusive); 0600–2200 G.M.T.
(*By courtesy of "Wireless World".*)

transmitted at low angles to the earth's surface, and under these conditions transmission over a very long distance may be obtained. As an example, during 1947 and 1948 television signals on 45 Mc/s radiated in the United Kingdom were received in South Africa. A peculiar feature of this type of propagation is that reflection may only occur in a limited portion of the v.h.f. band. In the example cited above, for instance, the 45 Mc/s vision signal was sometimes received when the 41·5 Mc/s sound signal was not. The phenomenon is, however, sufficiently rare in occurrence for propagation by this means to be ignored for practical purposes.

### Boundary Layer Reflections

Having already noted that radio signals are reflected on passing into ionised air having a higher refractive index, it will naturally

be expected that if, due to any further factors, other levels of the atmosphere also have different refractive indices, then they also will reflect radio waves. In practice these further refracting layers actually occur, but it is not until the v.h.f. band is reached that their effect begins to assume noticeable proportions. As in the case of such optical illusions as the mirage, these different refractive indices occur between air masses having different densities, temperatures, and water contents. Once the refractive index of the various air masses has been established at the frequency of the signal under consideration, the critical angle and the actual angle through which a wave will be bent can be readily calculated by the normal optical formula.

A very considerable amount of work has been done in order to determine the properties of the lower atmosphere. In one of several papers by C. R. Englund, A. B. Crawford, and W. W. Mumford, the results are published of a two years' study of reflections occurring at the boundaries between different air masses. Their measurements were made over a 70-mile sea-water path at wavelengths within the range of 1·6 to 5·0 metres. On this particular band the bending effect of several typical North American air masses was expressed as a factor which modified the actual radius of the earth—as far as radio transmissions are concerned. In expressing the bending effect in this way much complex calculation can be omitted and the desired information obtained directly. The table below is reprinted from the paper referred to above.

TABLE 4

| Air mass type | Effective earth radius | |
|---|---|---|
| | Summer | Winter |
| Tropical Gulf . . . | $1{\cdot}53 \times R$ | $1{\cdot}43 \times R$ |
| Polar Continental . . | $1{\cdot}31 \times R$ | $1{\cdot}25 \times R$ |
| Superior . . . . | $1{\cdot}25 \times R$ | $1{\cdot}25 \times R$ |

$R$ = the actual earth's radius.

The boundaries between these different air masses provide differences in the refractive indices which are adequate to produce a substantial bending effect on the radiation trajectory. In a

typical case reflections might be caused at a height of 4 to 5 km. at the boundary between a Superior air mass above a wedge of Transitional Polar to Tropical Atlantic air; while at a height of, say, 3·5 km., further reflections may occur at the boundary between the last-mentioned air mass and a Transitional Polar Continental air mass.

TABLE 5

| Altitude | Summer | | | Winter | | |
|---|---|---|---|---|---|---|
| | $S/T_g$ | $S/P_c$ | $T_g/P_c$ | $S/T_g$ | $S/P_c$ | $T_g/P_c$ |
| 1·0 km | 100 | 20 | 80 | 55 | 25 | 30 |
| 2·0 km | 50 | 10 | 40 | 50 | 15 | 35 |
| 3·0 km | 30 | 10 | 20 | 35 | 10 | 25 |

NOTE.—$S$=Superior. $T_g$=Tropical Gulf. $P_c$=Polar Continental.

The above table shows the difference in dielectric constant (times $10^6$) produced at the boundary between the various air-mass types indicated above. It should be noted that the refractive index is the square root of the dielectric constant. This table is also reproduced from the paper referred to earlier.

As might be expected, the shorter wavelengths in general exhibit the worst fading, considered either as a higher rate of fading, a greater amplitude of signal variation, or both. The whole nature and severity of the fading changes enormously from day to day. No sunrise and sunset variations are noticeable, but there is a seasonal falling-off in the average signal strength in the winter. No connection has been established between the visible weather phenomena and the fading experienced. Cloud bottoms which are merely the adiabatic dew-point level do not apparently cause signal reflection.

During their measurements Englund, Crawford, and Mumford noted changes in the received signal strength of up to 40 db; these changes were always slower than those due to the ionised layers on the short waveband. Even under turbulent atmospheric conditions (high wind and convective instability) fading did not exceed a rate of some five cycles per second. Most of the reflections occur at boundaries lying between 5·5 km and 1 to 1·5 km. The majority of the boundaries occur at the lower heights, as would

be expected from Table 5, which indicates that the largest differences in dielectric constant occur in this region.*

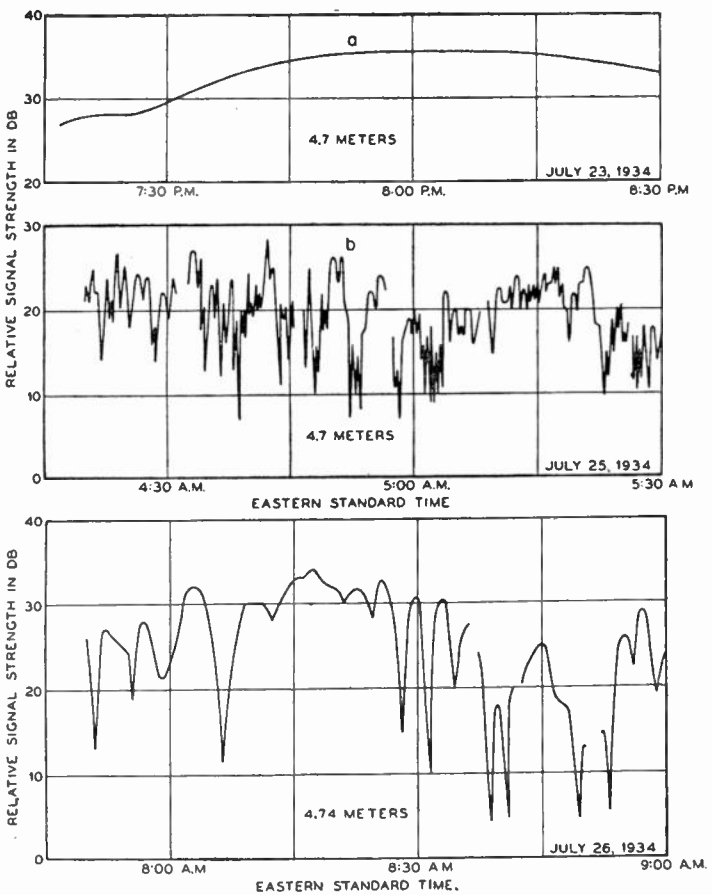



FIG. 5.5.—Recordings of the extreme boundary layer fading conditions observed on 4·7 metres over a 70-mile sea-water path, by Bell System Engineers. The first diagram shows the slowest rate of fading, the second a rapid rate of fading, and the third extreme fading amplitudes.

In the same paper it is also deduced that for this type of fading the difference between the direct and reflected paths was normally between 8 and 550 metres over the 70-mile sea-water path. These

* When the magnetic permeability of a medium is unity the dielectric constant at the frequency being considered is the square of the refractive index. See also the section on horizontal and vertical polarisation.

figures make it possible to assess the severity of any selective fading which may occur due to this cause. Taking the maximum difference in path length (550 metres) and assuming a carrier frequency of 200 Mc/s (1·5 Metres) and a peak-to-peak deviation of 150 kc/s, then at the lower deviation limit (199·925 Mc/s) there will be a difference in path length of $\frac{550}{1{\cdot}501} = 366{\cdot}4$ wavelengths. At the maximum peak deviation frequency (200·075 Mc/s) there will be a difference in path length of $\frac{550}{1{\cdot}499} = 366{\cdot}9$ wavelengths.

It will be seen that the difference between the reflected and direct path lengths at the upper and lower deviation limits will therefore be approximately one-half wavelength. Under these conditions it is possible for the direct and reflected signals to be in phase at one extreme frequency limit and 180° out of phase at the other. As, however, this is an extreme case, both on the length of path and differences in direct and reflected path lengths, it is safe to draw the deduction that the difference in path length will in practice only produce sporadic distortion to a high-fidelity frequency-modulation broadcast service on a carrier frequency of some 200 Mc/s. Below this frequency fading resulting from this cause may be ignored.

**Reflections from Solid Objects**

The third type of selective fading is likely to become troublesome at rather lower frequencies. It results from reflections due to such objects as buildings, mountains, gas-holders, and aircraft in flight. It has already been shown that the factors which determine the amount of distortion resulting from selective fading are, firstly, the difference expressed in wavelengths between the direct and the reflected paths, and, secondly, of course, the strength of the reflected signal. It is very difficult to lay down any definite figures for the distortion which will result. However, particularly in the case of gas-holders, relatively powerful signals can be returned from distances of some miles. By employing reasoning similar to that adopted earlier it can be shown that even before the carrier frequency is raised as high as 200 Mc/s, serious distortion can be caused. In practice it is usually possible to completely eliminate this type of distortion by moving the receiving aerial a few feet in either direction.

In addition to the distortion which may be caused by selective fading, it is possible for the signals reflected from aircraft to produce detrimental results due to the shortening or lengthening of the path taken by the reflected wave. Because of the Doppler Effect, the reflected signal frequency will be increased by an amount determined by the rate at which the transmission path is being shortened; conversely the reflected signal frequency will be lowered when the reflection path is being increased. The result at the receiver is a heterodyne beat note due to the frequency

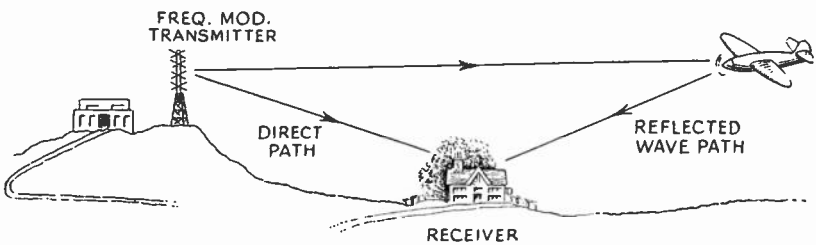



FIG. 5.6.—Reflections from moving objects are received at an altered frequency. The resultant heterodyne beat note between the direct wave and the reflected wave will in many cases be below the limit of audibility.

difference existing between the reflected and the direct waves. Taking the example illustrated in Fig. 5.6, the reflected path is being shortened at a rate which is twice the speed of the approaching aircraft. Assuming that it is travelling at a speed of 300 miles per hour (or some 134 metres per second), the length of the reflected signal's path will be shortened by some 268 metres per second. If the carrier wavelength is taken as 1·5 metres (200 Mc/s) the reflected signal frequency will be raised by some 180 c/s. The difference frequency between the reflected and direct signals will therefore result in a 180-c/s heterodyne beat note. It should, however, be noted that the example taken is extreme, and that in the majority of cases the path length will not alter so rapidly and therefore the heterodyne frequency will be lower —in most cases a "fluffing" noise accompanied by distortion will be heard. It is this same effect which causes a television picture to "flutter" when an aircraft passes overhead.

In summing up the position, it may be stated that low frequency heterodynes, accompanied, of course, by severe selective fading, may be expected as a result of aircraft reflections under conditions

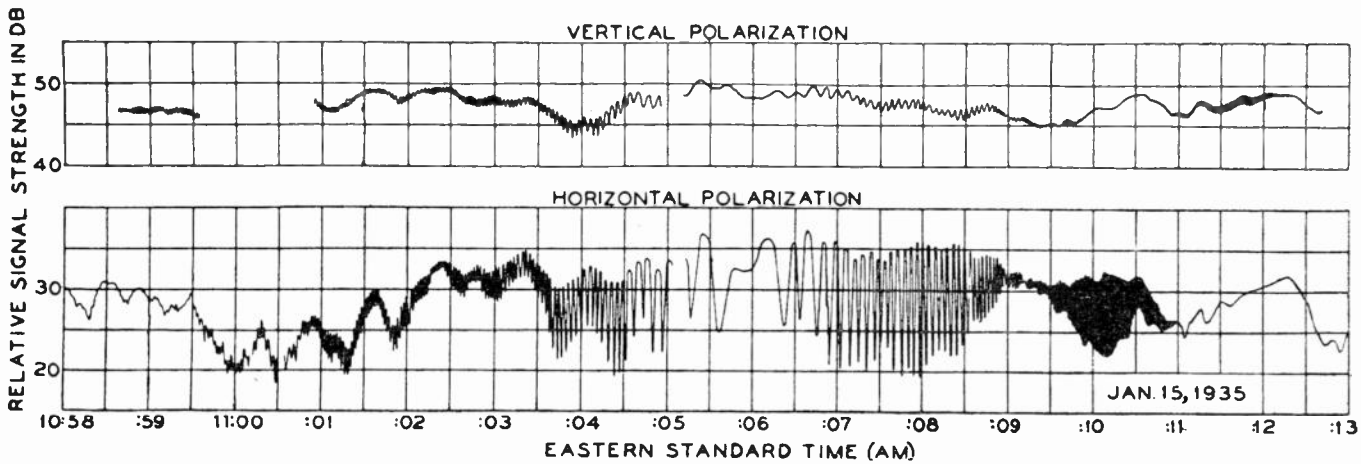



FIG. 5.7.—Records, taken by Bell System Engineers on 4·7 metres and in two polarisations, of high-speed fading attributable to radio reflections from a moving aircraft.

of low ground field strength with high field strengths above the ground. These conditions are liable to occur in valleys or on low-lying ground near aerodromes or other points over which aircraft pass at relatively low altitudes.

### Transmitter Service Range

The service area of a transmitter operating in the frequency bands where reflections from the ionosphere do not normally occur, can be divided into two regions, within the horizon and beyond the horizon. By horizon, we shall mean the boundary of that portion of the earth's surface (assumed perfectly spherical) within which the direct line between transmitting and receiving aerials does not intersect the earth's surface. Within the horizon, the field at the receiver can be considered as the resultant of two components, the direct ray and the ray reflected from the earth's surface. Beyond the horizon, the field consists of a single component, the direct ray, and is due to diffraction effects.

In determining the position of the horizon, allowance must be made for the effect of refraction. Under normal conditions the refractive index of the atmosphere falls with increasing height. A wave transmitted from the earth's surface is, therefore, travelling into a medium the refractive index of which is decreasing; as a consequence, the wave is refracted towards the earth's surface. This effect is closely associated with that of boundary layer reflections, discussed earlier; boundary layer reflections require, however, the presence of air masses having distinct refractive indices one above the other. As explained in the section on boundary layer reflections, it is usual to allow for the effect of refraction by assuming the radius of the earth to be larger than its actual value; and a value of $1{\cdot}33 \times$ the actual radius is usually taken as a mean value, as representative of the atmospheric conditions most likely to be encountered. The table on page 116 indicates the order of deviation from this value which may be expected in practice.

A correction factor must, therefore, be applied in computing the distance to the horizon as defined above. The normal distance to the horizon, as measured by considerations of optical range, is given by:

$$D = 3{\cdot}55\ (\sqrt{h} + \sqrt{a}) \text{ kilometres}$$
$$= 1{\cdot}22\ (\sqrt{h} + \sqrt{a}) \text{ miles},$$

where $h$ is the height of the transmitting aerial and $a$ is the height of the receiving aerial, and $h$ and $a$ are both measured in metres or feet. Allowing for the standard corrections for refraction, the expression becomes:

$$D = 4{\cdot}13\ [\sqrt{h} + \sqrt{a}] \text{ kilometres}$$
$$= 1{\cdot}42\ [\sqrt{h} + \sqrt{a}] \text{ miles.}$$

The phenomenon of diffraction occurs whenever electro-magnetic waves are propagated in the neighbourhood of an obstacle; in the case of determining field strength beyond the horizon, the obstacle is the earth itself. The effect is that a certain portion of the radiated energy enters that region which is "shadowed" by the obstacle. In general, the strength of the field received by diffraction tends to fall ultimately exponentially with distance from the transmitter. There is, however, no abrupt transition in field strength at the horizon for the range of frequencies employed in radio communication.

For the region within the horizon, a good approximation to the median field strength at any point is given by the following expression due to H. H. Beverage:

$$E = \frac{88\sqrt{W}\ ah}{\lambda D^2}, \qquad \ldots\ldots \quad (5.3)$$

where $E$ = the field strength in r.m.s. volts per metre;
$W$ = effective power radiated in watts = power into the aerial times the aerial gain over a half-wave dipole (see page 191);
$a$ = the receiving aerial height in metres;
$h$ = the transmitting aerial height in metres;
$D$ = the distance from transmitter to receiver in metres;
$\lambda$ = signal wavelength in metres.

This expression may be extended to give the approximate field strength beyond the horizon, by multiplying the field strength obtained in expression (5.3) by a factor $(D_h/D)^n$, where $D_h$ is the distance to the horizon. The value of $n$ varies with frequency as shown in Fig. 5.8.

The median field strength is that exceeded at 50 per cent of the receiving sites at a given distance from the transmitter; as the

transmission frequency is raised, so does the median field strength fall below the value given by the expression above. Additionally, the range of variation of field strength in the neighbourhood of a given receiving site is also found to increase as frequency is raised. J. A. Saxton, in discussing the departure of the median field strength from the predicted value, states that, to a first degree of

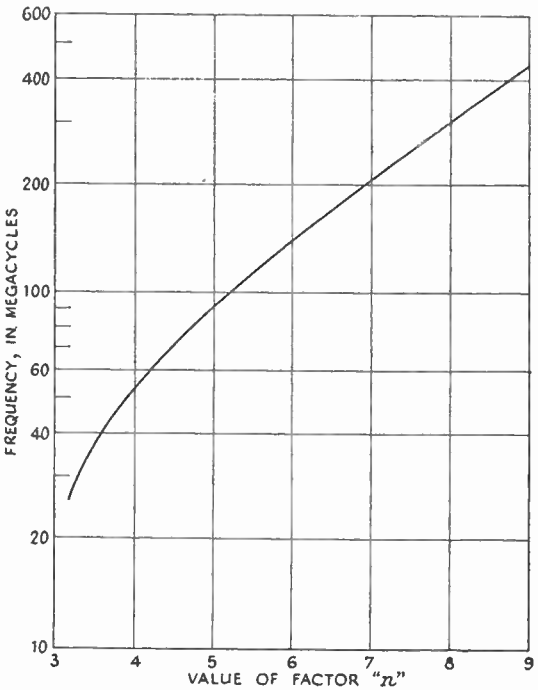



FIG. 5.8.—Variation of the factor "$n$" when considering v.h.f. propagation beyond the horizon.

approximation, the departure in decibels would appear to be constant for all distances up to 50 miles. At frequencies in the region of 50 Mc/s there is close agreement between predicted and actual values; at 200 Mc/s, the difference is in the region of 10 db increasing to 15 db at 500 Mc/s and 20 db at 1,000 Mc/s.

In order to assess the limitations of the expression (5.3) above, when used to compute the field strength within the horizon, it is necessary to set out the assumptions made; these are as follows:

1. The transmissions are above the ionospheric reflection limit, i.e. above 30–40 Mc/s.
2. The received signal consists of two components, a direct component and one received by reflection from the earth's surface; it is assumed that the aerial radiates equally well in the direction of both rays.
3. The surface of the earth is taken as flat.
4. The path length of the direct ray is shorter than that of the reflected ray by less than one-sixth of a wavelength.
5. The coefficient of reflection at the earth's surface is unity, and the phase change produced on reflection is 180°.
6. In the case of vertical polarisation, the aerials are assumed to be at least two wavelengths above the earth's surface.

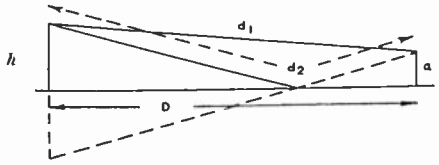



FIG. 5.9.—Paths of reflected and direct rays from transmitter to receiver.

There is no need to comment on the first assumption. The second assumption is the basis on which the calculation rests. Fig. 5.9 shows the paths of the direct and reflected rays, and from the geometry of the diagram, the path length of the direct ray $d_1$, is

$$d_1 = \sqrt{D^2+(h-a)^2},$$

whilst the path length of the reflected ray is

$$d_2 = \sqrt{D^2+(h+a)^2};$$

the path length difference is thus

$$d_1-d_2 = \sqrt{D^2+(h-a)^2} - \sqrt{D^2+(h+a)^2}.$$

The expressions under the square root signs can be expanded by the Binomial theorem if $D \gg h \pm a$, which is normally true. This leads to

$$d_1-d_2 = 2ah/D.$$

The corresponding phase difference $\theta$ between the two components (adding $\pi$ radians for the phase change at reflection) is thus

$$\theta=\pi+\frac{2ah}{D}\frac{2\pi}{\lambda} \text{ radians.}$$

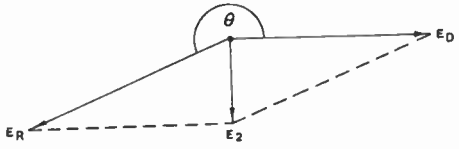



FIG. 5.10.—Vector diagram of received direct and reflected components.

The vector diagram from the two components at the receiving aerial is therefore as shown in Fig. 5.10; the resultant of the two components is therefore

$$E_2=\sqrt{(E_D+E_R\cos\theta)^2+E_R^2\sin^2\theta}$$

$$=\sqrt{E_D^2+E_R^2+2E_DE_R\cos\theta}. \quad . \quad . \quad (5.4)$$

We have assumed that the coefficient of reflection is unity; hence $E_D=E_R$, whence

$$E_2=E_D\sqrt{2(1+\cos\theta)}.$$

Since $\theta=\pi+\frac{2ah}{D}\frac{2\pi}{\lambda}$, $\cos\theta=-\cos\left(\frac{2ah}{D}\frac{2\pi}{\lambda}\right)$.

Thus

$$E_2=E_D\sqrt{2\left\{1-\cos\left(\frac{2ah}{D}\frac{2\pi}{\lambda}\right)\right\}}$$

$$=2E_D\sin\left(\frac{ah}{D}\frac{2\pi}{\lambda}\right). \quad . \quad . \quad . \quad . \quad . \quad (5.5)$$

At small values of $\frac{ah}{\lambda D}$, this is approximately equal to

$$E_2=E_D\frac{2ah}{D}\frac{2\pi}{\lambda}.$$

As shown later, $E_D$ is given by

$$E_D=\frac{7{\cdot}0\sqrt{W}}{D} \quad \text{(see pages 156 and 159),}$$

whence

$$E_2 = \frac{88\sqrt{W}}{\lambda D^2}\, ah.$$

In practice, the approximation is satisfactory for phase differences up to $\pi/3$ radians; this corresponds to a path length difference of $\lambda/6$.

If, however, the condition is not fulfilled, expression (5.5) must be employed instead. It is instructive to consider what happens if the receiving aerial height only is altered. As the aerial is raised, the field strength as predicted by expression (5.3) is found;

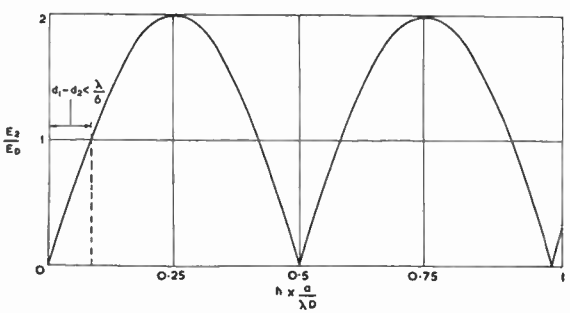



FIG. 5.11.—Variation of field strength with aerial height due to reflected ray. The approximate expression (5.3) is valid over the range $d_1 - d_2 < \lambda/6$, i.e. $ah/\lambda D < 0.085$.

when, however, the path difference exceeds one-sixth wavelength, the signal strength increases more slowly than predicted until the path difference is equal to $\lambda/2$; the signal strength is then equal to twice the free space value. With further increase of height, the signal strength again decreases, and falls to zero when the path length difference is equal to $\lambda$. If the aerial is still further raised, the cycle is repeated. This is illustrated by Fig. 5.11. A similar variation of field strength with distance is also experienced; Fig. 5.12 shows this.

Beverage's expression may be modified for use when the signal frequency is given in Mc/s, $h$ in feet, $D$ in miles and $E$ in microvolts per metre. The expression then becomes

$$E = 0{\cdot}01052 f\sqrt{W}\,\frac{ah}{D^2}.$$

With these units, the requirement that the path length difference should be less than $\lambda/6$ can be stated as

$$\frac{haf}{D} < 0{\cdot}433 \times 10^6.$$

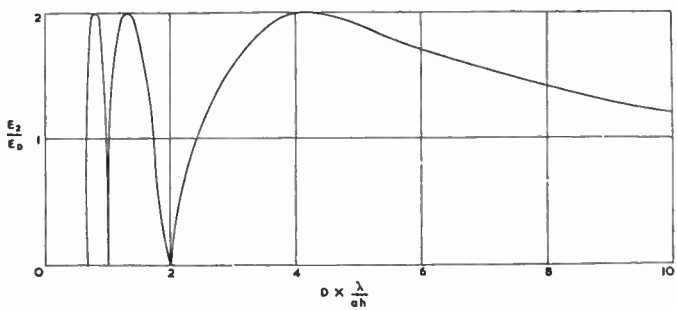



FIG. 5.12.—Variation of field strength with distance due to reflected ray. The approximate expression (5.3) is valid for $d_1 - d_2 < \lambda/6$, i.e. $\lambda D/ah > 12$.

The condition that the earth is considered flat can be modified to take account of the earth's curvature by taking the heights of the transmitting and receiving aerials as less than their actual value. As shown by Fig. 5.13, this assumes reflection at a plane

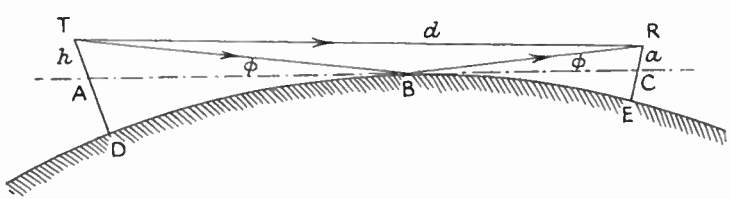



FIG. 5.13.—This diagram illustrates the geometry involved in spherical earth calculations of field strength.

surface $ABC$; the length of the aerial computing field strength must then be taken as $h'(AT)$ and $a'(CR)$.

From a knowledge of the distances $BD$ and $DE$, the necessary reductions of aerial heights (by $AD$ and $CE$) can be calculated. These reductions are given with sufficient accuracy by

$$AD = \frac{BD^2}{2},$$

and

$$CE = \frac{BE^2}{2},$$

where $AD$ and $CE$ are in feet, and $BD$, $BE$ are in miles. This relationship automatically allows for the refraction in the earth's atmosphere referred to earlier.

Passing on to consider the fifth assumption which postulated that the coefficient of reflection was 100 per cent and the phase change on reflection was 180°. For horizontal polarisation these conditions will generally be satisfied when reflection is due to either the surface of earth or to water and the carrier frequency is above some 50 Mc/s. For vertical polarisation these requirements will be satisfied for reflections from either the earth or fresh water at angles of reflection $\phi$ below approximately 0·5°. However, more comprehensive formulae must be employed for cases where a vertically polarised wave is reflected from salt water, or where the angle of reflection $\phi$ exceeds about 0·5°.

The properties of the earth which produce these varying phase changes and coefficients of reflection are expressed as a dielectric constant $\varepsilon$ and an ohmic resistance. For the earth this ohmic resistance is conveniently expressed in terms of resistivity per centimetre cube. The reciprocal of the restivity is the specific conductivity $\sigma$, which is expressed in mhos per centimetre cube. It should be noted that the specific conductivity is expressed in electro-magnetic units in this chapter. These em-cgs system units should not be confused with the es-cgs system units which give values $9\times10^{20}$ times larger.

TABLE 6

*Typical figures for the dielectric constant and resistivity of ground*

| Type of ground | Specific conductivity $\sigma$ em-cgs units | Dielectric constant— $\varepsilon$ |
|---|---|---|
| Sea water | $4\times10^{-11}$ | 80 |
| River water | $45\times10^{-15}$ | 80 |
| Dry soil | $10^{-16}$ | 3 to 5 |
| Farm soil—fertile | 5 to $15\times10^{-14}$ | 10 to 30 |
| Sandy soil close to the sea | $10^{-15}$ | 8 to 10 |
| Moist ground | $30\times10^{-14}$ | 30 |
| Inland soil | $10^{-13}$ | 15 |

Measurements made by Barfield and Smith-Rose have shown that there are large variations in the conductivity of the ground with moisture content. Thus, for one sample of loam $\sigma$ was $9\times 10^{-15}$ em-cgs units when the moisture content was about 1 per cent, and $1{\cdot}3\times 10^{-13}$ em-cgs units when it was 25 per cent. For the same sample $\varepsilon$ varied from 3 for 1 per cent moisture content to 37 for a 25 per cent moisture content. In addition to these variations the reflection constants changed slightly with frequency, in general the dielectric constant decreased and the specific conductivity increased.

McPetrie and Saxton have shown that, on the v.h.f. band, reflection takes place very near the surface. On one site they found that at a frequency of 60 Mc/s grass-covered ground gave $\sigma=3\times 10^{-11}$ em-cgs units and $\varepsilon=18$; whereas when the 9-inch layer of grass-covered surface soil was removed and the gravel subsoil dug out to a depth of 7 feet, so as to leave a new surface of pure gravel without the slightest sign of humus or decayed vegetation; $\sigma=3\times 10^{-14}$ em-cgs units and $\varepsilon=5$.

### Horizontal and Vertical Polarisation

It is impossible to consider the question of earth reflections without bringing in the differences between horizontal and vertical polarisation. It is apparent that so long as we consider a direct wave alone, there will in fact be no difference between the two polarisations. In practice, however, the received signal will always be made up, at least in part, of a reflected wave component. In so far as it is made up of such a component it may be expected that any differences existing between the reflection of horizontally and vertically polarised waves will in turn result in proportional differences in the received field strengths.

It has already been noted that the magnitude of the reflection coefficient is dependent upon the dielectric constant, the conductivity of the reflecting surface, the incident made by the wave with this surface, and also, what is most important, on whether the wave is vertically or horizontally polarised. Provided that the receiver lies within optical range of the transmitter, i.e. providing that both transmitter and receiver lie above the tangent plane through the point of reflection (see Fig. 5.13), then the resultant field strength will be determined by the vectorial combination of these two components.

The first component, that arriving over the direct path, will be

$$E_0 = \frac{138\sqrt{P} \times 10^3}{D}, \quad . \quad . \quad . \quad . \quad (5.6)$$

where $E_0$ = the field strength due to the direct path wave, in r.m.s. microvolts per metre;
$P$ = transmitting aerial power in kilowatts (half-wave dipole aerial);
$D$ = direct distance in miles between the transmitter and the receiver.

Equation (5.6) will also give the strength of the reflected component, if in place of $D$ the total length of the reflected path is substituted, and the resultant field strength is multiplied by the coefficient of reflection.

To combine vectorially the direct and reflected components it is also necessary to know the phase angle existing between them. This phase angle is made up of two parts, one due to the difference in the path lengths and the other due to the phase change produced upon reflection. While the phase angle introduced by the difference in path length may be calculated by straightforward geometric procedure, that produced upon reflection is rather more difficult to ascertain. This second phase angle is included in the reflection coefficient, a complex relationship which expresses the reduction suffered by the original wave amplitude, as well as the phase shift imparted to it during reflection.

$$\text{The reflection coefficient} = Ke^{ja},$$

where $K$ = the reduction in amplitude upon reflection;
$a$ = the phase shift produced upon reflection.

The reflection coefficient defines the extent to which the amplitude of this wave is altered, by means of the term $K$, while the phase relation in which the reflected wave component (a $j$ component of variable angle) should be added to direct wave component, is indicated by the term $a$ in the index $ja$.

It has been shown by H. O. Peterson that the value of the reflection coefficient can be evaluated from the following two formulae:

$$K_v e^{ja_v} = \frac{\varepsilon_0 \sin\phi - \sqrt{\varepsilon_0 - 1 + \sin^2\phi}}{\varepsilon_0 \sin\phi + \sqrt{\varepsilon_0 - 1 + \sin^2\phi}} \quad . \quad . \quad (5.7)$$

for vertical polarisation, and

$$K_h e^{j\alpha_h} = \frac{\sin\phi - \sqrt{\varepsilon_0 - 1 + \sin^2\phi}}{\sin\phi + \sqrt{\varepsilon_0 - 1 + \sin^2\phi}} \quad . \quad . \quad . \quad (5.8)$$

for horizontal polarisation,

where $\phi$ = the angle of incident with the reflecting medium;
$\varepsilon_0$ = the effective dielectric constant of the reflecting medium.

The value of $\varepsilon_0$ is given by

$$\varepsilon_0 = \varepsilon - j\frac{18\times 10^{14}\sigma}{f},$$

where $f$ = the signal frequency in megacycles;
$\sigma$ = the earth's conductivity in em-cgs units;
$\varepsilon$ = the dielectric constant of the reflecting medium.

To illustrate the extent to which these various factors influence the strength of the reflected wave, a numerical example published by M. Katzin will be given. For simplicity he takes the case of a pure dielectric reflecting medium (which will, in fact, be approached by dry soil) with a horizontal polarisation; such a wave is always reversed in phase by 180° on reflection by a pure dielectric, while the magnitude of the reflected wave will fall from 100 per cent ($K=1$) of its original value with grazing reflection incidence, to a value at perpendicular incidence which is dependent on the dielectric constant. For small angles between the wave and the reflecting surface, the reflected wave will suffer practically no attenuation on reflection, so that $K$, in the coefficient of reflection, differs inappreciably from unity.

For vertical polarisation, the position is somewhat different. Here, for grazing angles of incidence with the reflecting medium, the reflected wave is reversed in phase without any appreciable reduction in amplitude ($\alpha=180°$, $K=1$). However, with increasing angles of incidence the reflected wave's amplitude decreases rapidly and finally becomes zero ($K=0$), at the angle of incidence whose co-tangent is equal to the square root of the dielectric constant of the reflecting medium. This angle is known as the Brewster angle. Above this angle there is no change in the phase of the wave on reflection ($\alpha=0°$), and the amplitude of the reflected wave increases steadily to a value, at a perpendicular

angle of incidence, which is the same as that for a horizontally polarised wave. Fig. 5.14 shows the reflection coefficient of both vertical and horizontal polarisations for ground with zero conductivity and a dielectric constant of 9. In this case the angle at which no reflection takes place for vertical polarisation is some 18·5°.

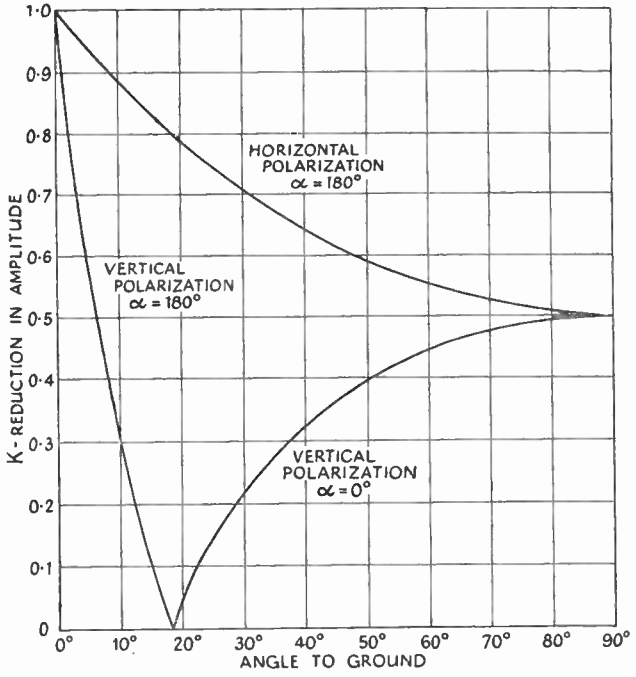



FIG. 5.14.—Reflection coefficient of ground having $\varepsilon=9$, $\sigma=0$.

When the conductivity of the reflecting medium is not negligible the relations are more involved. The phase shift on reflection is in general other than zero or 180°. As in the case of a perfect dielectric, the reflected wave is reversed in phase without reduction in amplitude so long as the angle of incidence is zero, but the amplitude decreases rapidly as the angle of incidence is increased. However, instead of passing through zero it reaches a finite minimum value after which it increases in amplitude again. At the same time the phase shift on reflection, considered as a lag, decreases from 180° at zero incident angle to zero at vertical

incidence, passing through 90° at the angle for which the amplitude of the reflected wave is a minimum. For a given ground dielectric constant, increasing conductivity lowers the angle of incidence at which the amplitude of the reflected wave is a minimum. Fig. 5.15 shows the reflection coefficients for sea-water at a frequency of 50 Mc/s.

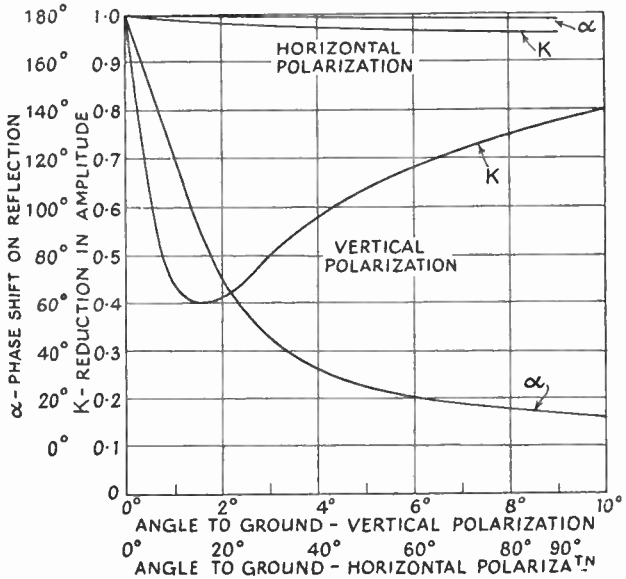



FIG. 5.15.—Reflection coefficient of sea-water at 50 Mc/s. $\varepsilon=80$, $\sigma=4\times10^{-11}$ em-cgs units.

It has already been noted that the difference in the reflection coefficient with horizontal and vertical polarisation is very largely responsible for the difference in behaviour of these two polarisations when propagated over mediums of good conductivity, such as sea-water. In vew of this it is of interest to refer back to the two curves given in Fig. 5.7. It will be noted that these curves were obtained over a 70-mile sea-water path and that the field strength of the vertically polarised signal is very considerably greater than that due to the horizontally polarised signal.

Investigations by Trevor and Carter have shown that such variations are to be expected from theoretical considerations. They have shown that in the case of sea-water at frequencies

above some 200 to 300 Mc/s, the dielectric "earth" current predominates over the conductivity current, and that the sea-water "ground" behaves as a pure dielectric. With vertical polarisation as the frequency is lowered the phase shift produced upon reflection departs from 180° and, the difference in path length being

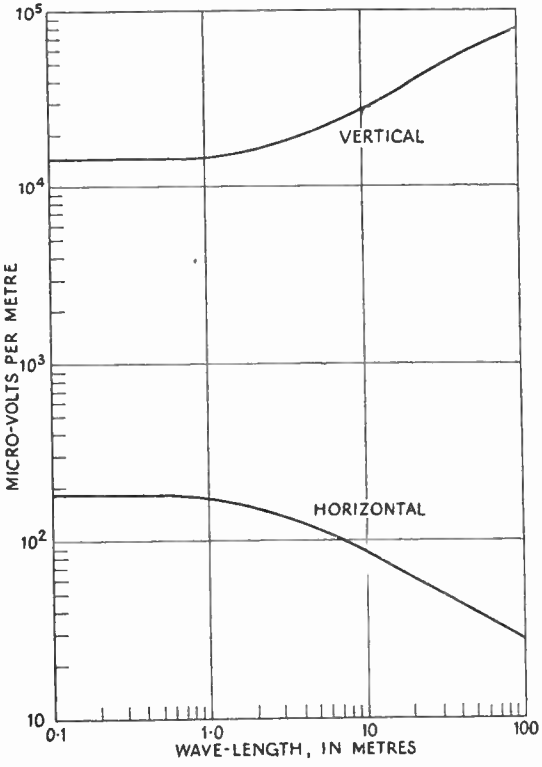



FIG. 5.16.—Theoretical field strength vs. wavelength over sea-water at a distance of 1 km. for a dipole 8 metres high and radiating 1 watt; vertical and horizontal polarisation. Receiving aerial height is zero feet.

only a fraction of a wavelength, the reflected wave does not arrive with an absolutely opposing phase relationship to the direct wave. It will be seen that this will result in an increased field strength for the vertically polarised signal. With horizontal polarisation, on the other hand, there is no appreciable change in the phase of the reflected wave with frequency; at the same time the

magnitude of the reflection coefficient approaches unity more closely as the frequency is lowered, so resulting in reduced field strengths. The limiting ratio to the difference in field strength due to a change from vertical to horizontal polarisation is equal to the dielectric constant of water, which in this case is 80.

### Measurements confirming Differences between Horizontal and Vertical Polarisation

Having now outlined the theory underlying the differences between horizontal and vertical polarisation, it is proposed to see how far this theory is confirmed by practical results. Measurements with this object in view have been made by Katzin, George, and others. The results published by both these investigations agree fairly closely. Measurements were made by George over the bands from 81 to 86 Mc/s and 140 to 145 Mc/s at twenty-one locations in the New York area. He used both horizontal and vertical polarisations, and made measurements to establish the strength of the direct and indirect waves in each case.

The table following summarises the results obtained in the mass plot referred to in the preceding paragraph. It compares the maximum and minimum ratios measured over the two 5-megacycle bands; this variable frequency method of determining the strength of the direct and indirect signals being the inverse of

TABLE 7

*Geometric means of the ratios of maximum field strength to minimum field strength obtained during measurements made at 21 locations*

| | 81 to 86 Mc/s | 140 to 145 Mc/s |
|---|---|---|
| Horizontal polarisation . . | 1·86 | 2·12 |
| Vertical polarisation . . . | 2·97 | 3·38 |

that employed in the study of propagation distortion at the beginning of this chapter. It will be seen that the indirect interfering signals were from 10 to 20 per cent stronger at the higher frequencies and that they were strongest with vertical polarisation on both frequencies. Another interesting point emerging from George's study is that, on the average, horizontal polarisation

produces a field strength about 2 db higher than that resulting from vertical polarisation. This was found on both the frequency bands on which he made measurements.

At first sight it would appear that the larger variations in the field strength which occurred in the case of a vertically polarised signal (i.e. due to a stronger indirect ray), do not agree too well with the theoretical conclusions outlined earlier. However, there is a very reasonable explanation. It has been shown that the reflection coefficient for horizontal polarisation (i.e. the electric field parallel to the reflecting surface) is always greater than that for vertical polarisation (i.e. the electric field perpendicular to the reflecting surface), except for the limiting cases of grazing and perpendicular incidence. In urban areas we are concerned with reflecting surfaces which are predominantly vertical, instead of the horizontal ground surfaces, to which Fig. 5.14 applies. It follows, therefore, that the effective polarisations are interchanged, so that transmission from a vertical aerial corresponds to a horizontal polarisation with respect to vertical buildings and vice versa. It may, therefore, be expected that vertical aerials will result in reflected ray components of greater amplitude, on the average, than would horizontal ones. The very great number and complex nature of these reflections, however, tend on the average to neutralise one another, so that horizontal polarisation with the smaller number of reflections will actually result in a slightly greater field strength.

It should be noted that, in addition to consideration of the field strength, the difference between the motor-car ignition noise pick-up on a horizontal and vertical receiving aerial is of considerable importance. This is discussed in the next section.

### Interference Pick-up on Vertical and Horizontal Dipoles Aerials

At the end of the last chapter it was shown that, on the average, horizontal polarisation may be expected to give a slightly greater field strength in urban areas, and that the amount of selective fading—again in urban areas—will in general be less than that experienced with vertical polarisation. Important though these points are, it is doubtful whether they would in themselves justify the strong preference which exists for horizontal polarisation.

By far the most serious form of interference experienced on the v.h.f. band is that generated by the ignition systems of passing

motor-cars. The field strength of the interfering signals radiated by motor-car ignition systems over the frequency band between 40 and 450 Mc/s has been studied in detail by R. W. George of the Radio Corporation of America. He made his measurements on a receiving aerial situated 35 feet above the ground and 100 feet away from a dual-carriageway arterial road. He measured the

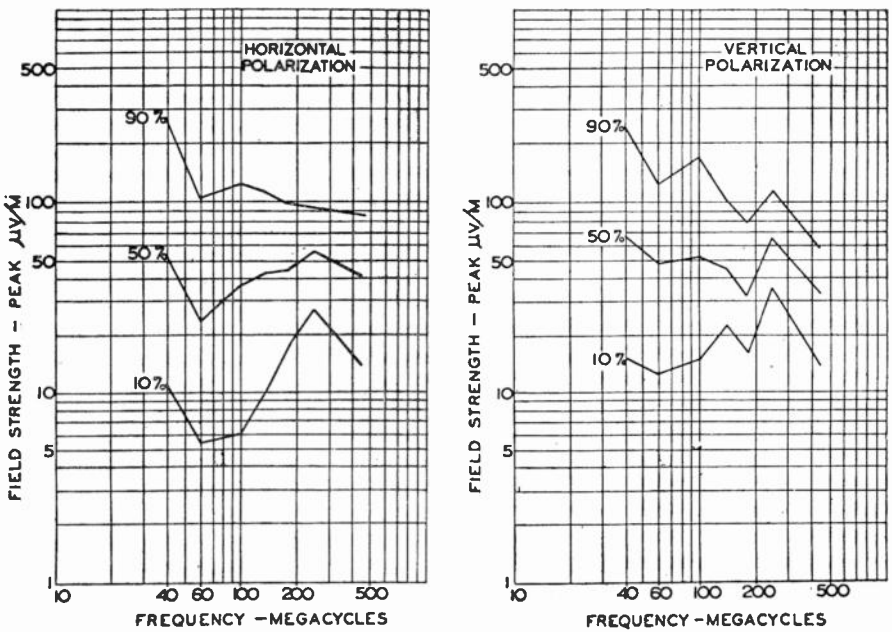



FIG. 5.17.—Motor-car ignition radiation picked up on a receiving dipole, firstly in the vertical and then in the horizontal position.
*(By courtesy of the I.R.E.)*

peak ignition field strength of each car as it passed the nearest point to the receiving aerial, on a specially designed peak voltage indicator. In order to ensure known aerial constants a half-wave dipole receiving aerial was used, the complete receiving equipment being recalibrated for each frequency on which measurements were made.

Fig. 5.17 summarises the results of these measurements. The curves show the motor-car ignition radiation which was picked up on the receiving dipole, firstly in the vertical and then in the horizontal position. The field strength is expressed in peak

microvolts developed within a receiver band-width of 10 kc/s. The percentage given against each curve is the proportion of the total number of vehicles which produced less than the field strength indicated. These measurements show clearly that there is a small difference in the initial level of the horizontally and vertically polarised components of motor-car ignition interference.

*Summary of BBC Measurements on the Range of Ignition Interference*

| Field Strength 45 Mc/s Half-wavelength dipole 30 feet above ground | Extinction Distance | | |
|---|---|---|---|
| | F.M. | | A.M. |
| | Horizontal polarisation | Vertical polarisation | Horizontal polarisation |
| 50 $\mu$V/m | 200 yds. | >200 yds. | At 100 yards the ignition was very disturbing, but merged into the set noise which was very high. |
| 100 ,, | 150 ,, | 200 ,, | |
| 300 ,, | 80 ,, | 120 ,, | As above, but less disturbing. |
| 500 ,, | 60 ,, | 120 ,, | Perceptible at 100 yards but merging into the set noise. |
| 1,000 ,, | 40 ,, | 80 ,, | 190 yards |
| 5,000 ,, | 25 ,, | 50 ,, | 120 ,, |

NOTE—Field strength on 90 Mc/s for same degree of interference is approximately one-third that at 45 Mc/s.

In passing, it should be noted that if these curves are used as a basis for any calculations of the ignition interference field strength under different conditions, it is most important that allowance be made for any difference in the receiver band-width. It will be remembered that it was shown in Chapter Three that the peak amplitude of impulsive noise was directly proportional to the receiver band-width.

Measurements which have been recorded by H. L. Kirke of the BBC Engineering Division are of a more practical nature than those carried out by R. W. George. The BBC measurements were made in order to determine any difference in the range of ignition interference. The relative distances at which the ignition noise is

extinguished were measured. The results are of considerable importance and are summarised in the above table.

The figures given are for substantially complete inaudibility of interference in a condition of low ambient acoustic noise. They may be taken as typical for a fairly well-designed receiver at 45 Mc/s. Further interference tests made by the BBC on 90 Mc/s indicate that the field strength required for the same degree of interference as on 45 Mc/s is only about one-third. It was also found that horizontal polarisation was preferable on 90 Mc/s, the improvement being about the same as on 45 Mc/s (e.g. some 10 db).

## Circular Polarisation

In addition to the measurements referred to above, George made various measurements on waves propagated with a circular polarisation. His measurements, which were taken at three locations, are summarised in Table 8. These measurements show that circular polarisation is slightly less desirable than horizontal and possibly somewhat more desirable than vertical.

TABLE 8

*Circular polarisation—comparisons of measurements made on the band 81 to 86 Mc/s*

| Comparison | Location | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| $\frac{\text{Vert. max./min. ratio}}{\text{Horiz. max./min. ratio}}$ | 1·16 | 1·24 | 1·37 |
| $\frac{\text{Cir. max./min. ratio}}{\text{Horiz. max./min. ratio}}$ | 1·02 | 0·94 | 1·11 |
| $\frac{\text{Avg. vert. mV/m.}}{\text{Avg. horiz. mV/m.}}$ | 0·8 | 0·97 | 1·48 |
| $\frac{\text{Avg. circ. mV/m.}}{\text{Avg. horiz. mV/m.}}$ | 1·0 | 0·91 | 1·16 |

## Received Power

Assuming that a half-wave dipole is used at the transmitter and that the receiving half-wave dipole has a radiation resistance

of 75 ohms and is matched into the load circuit, then the watts developed in that load circuit will be

$$W_r = \frac{3{\cdot}37Ph^2a^2}{D^4} \times 10^{-18} \text{ watts}, \quad . \quad . \quad . \quad (5.9)$$

where $W_r$ = watts absorbed in receiver load circuit;
$P$ = transmitting aerial power in kilowatts (half-wave dipole aerial);
$h$ = height of transmitting aerial above surrounding country;
$a$ = height of receiving aerial above surrounding country;
$D$ = distance in miles between transmitter and receiver.

When aerials other than half-wave dipoles are used at either or both the receiver and transmitter, then the above received power should be multiplied by the power gain of the two aerials.

The attenuation of a radio transmission circuit may be expressed in terms of the ratio of power transmitted to power absorbed at the receiver. By developing the above formula the following relationship is obtained:

$$\frac{P}{W_r} = \frac{2{\cdot}297D^4 \times 10^{11}}{h^2a^2} . \quad . \quad . \quad . \quad . \quad (5.10)$$

The above power ratio may, of course, be converted into decibels if so required. The attenuation should be divided by the power gain of the transmitting and receiving aerials. It should be noted that the above formulae only apply within optical range of the transmitter.

### The F.C.C. Field Strength Charts

With a view to standardising calculations of the field strength which may be expected from ultra-short-wave broadcasting stations, the United States Federal Communications Commission's Engineering Department have published a series of curves from which the service area of such stations may be determined. The first of these curves, which is reproduced as Fig. 5.18, deals with the propagation of a 46-Mc/s signal over land having a compromise but representative conductivity and dielectric constant. This curve, which is calculated from theoretical considerations, assumes a receiving aerial height of 30 feet. The chart reproduced as Fig. 5.18 may be used in order to determine the anticipated

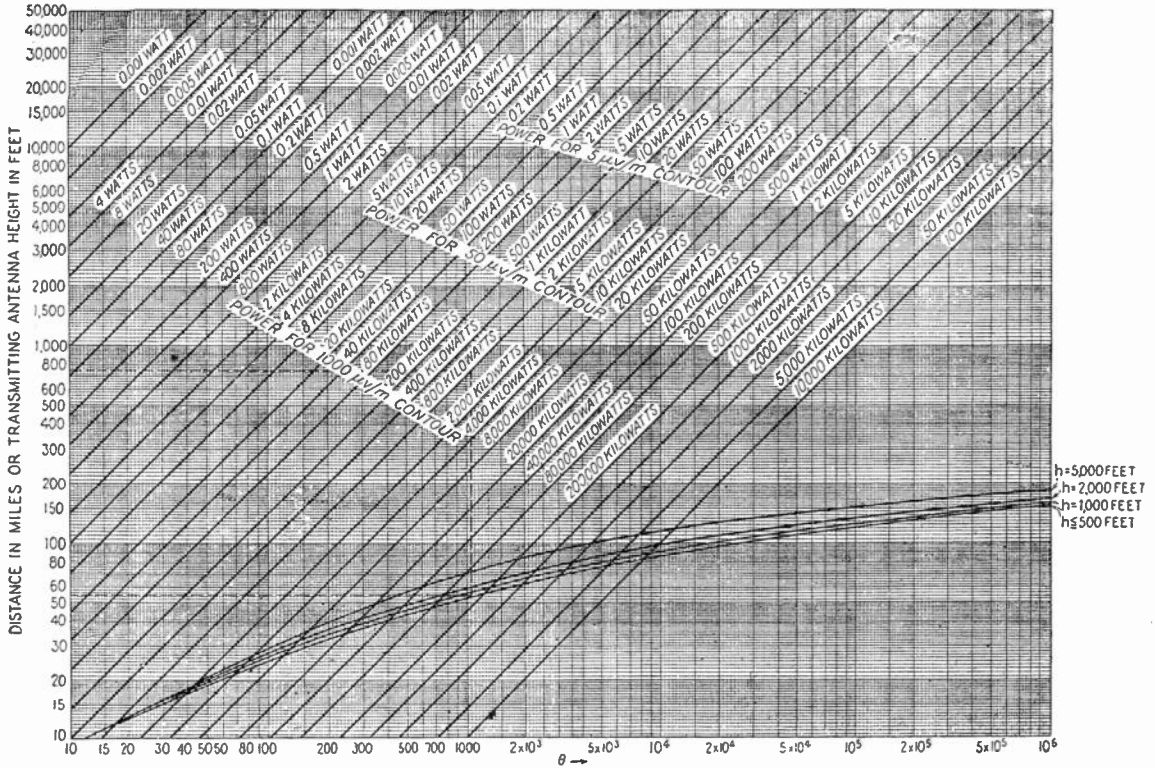



FIG. 5.18.—The F.C.C. field strength chart for determining the signal range of high frequency broadcasting stations. The chart applies to 46 Mc/s propagation over land with a conductivity of $\sigma = 5 \times 10^{-14}$ em-cgs units and a dielectric constant of $\varepsilon = 15$. The receiving aerial height is 30 feet.

*(By courtesy of the Federal Communications Commission, Washington.)*

distances to the 1,000, 50, and 5 microvolts per metre contours at the carrier frequency of 46 Mc/s. These distances are determined by the height of the transmitting aerial above the surrounding country, as well as the transmitting aerial's power and field gain. To determine the anticipated distance to, say, the 5-microvolt per metre contour, it is only necessary to follow the horizontal line corresponding to the transmitting aerial height over to the 45° line for 5 microvolts per metre and corresponding to the effective aerial power thence one proceeds vertically downward to the curved line corresponding to the aerial's height, and then again horizontally to the left to read off the distance in miles to the 5 microvolts per metre contour.

The term $\theta$ at the base of the chart is defined as the Effective Signal Radiated, and is expressed mathematically as

$$\theta = \frac{50hG\sqrt{P}}{F}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.11)$$

where $h$=the height of the transmitting aerial above the surrounding country;
$G$=the transmitting aerial field gain;
$P$=the aerial power in kilowatts;
$F$=the required field strength in microvolts per metre.

By working out the value of the Effective Signal Radiated, the F.C.C. chart can be used to determine the distance to any desired contour. Once having evaluated the Effective Signal Radiated, it is only necessary to run vertically upwards to the transmitting aerial height curve, and then over to read off the distance in miles on the left-hand scale to the selected field strength contour.

As an example of the way in which the chart is used, let it be assumed that the receiving aerial is a half-wave dipole 30 feet above the ground, that the transmitting aerial height is 750 feet, and that it is desired to determine the distance in miles to the 50-microvolt contour for a station in the 46-Mc/s band. Suppose that the aerial power is 500 watts, and the aerial array is such that there is a field gain of 2. As the field gain is the square root of the power gain, it follows that the true aerial power of 500 watts must now be multiplied by $2^2$ in order to obtain the effective aerial power. This results in $4 \times 500 = 2{,}000$, which means that the effective power is 2 kilowatts.

In using the chart it is, firstly, necessary to find the intersection between the horizontal line passing through the 750-foot ordinate and the 2-kilowatt 45° line associated with the 50-microvolt group. The distance to the 50-microvolt contour is found by proceeding

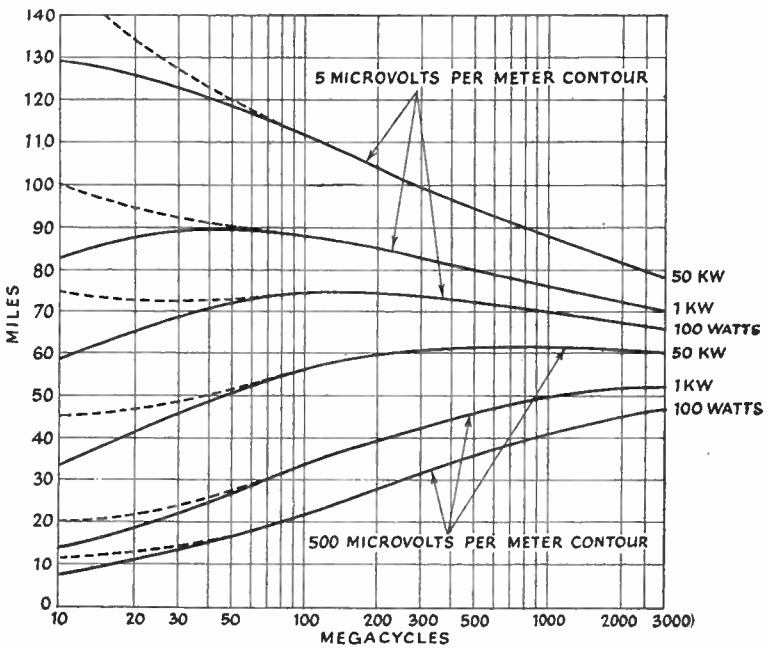



FIG. 5.19.—The F.C.C. chart showing the variation with frequency of the range of a v.h.f./u.h.f. broadcasting station. The chart assumes a half-wave dipole transmitting aerial at a height of 1,000 feet and a similar receiving aerial at 30 feet. Ground characteristics as for Fig. 5.18. The dotted contour is for vertical and the solid line for horizontal polarisation.

*(By courtesy of the Federal Communications Commission, Washington.)*

vertically downwards to the intersection with the 750-foot curve. This curve, although not drawn, lies half-way between the 500-foot and the 1,000-foot curves. The height of this intersection point, when read off on the vertical scale, gives the distance to the 50-microvolt contour as 54.5 miles. If the above procedure is reversed, Fig. 5.18 may be used to find the power required for a given aerial height, in order to cover a certain distance within a 50-microvolt contour.

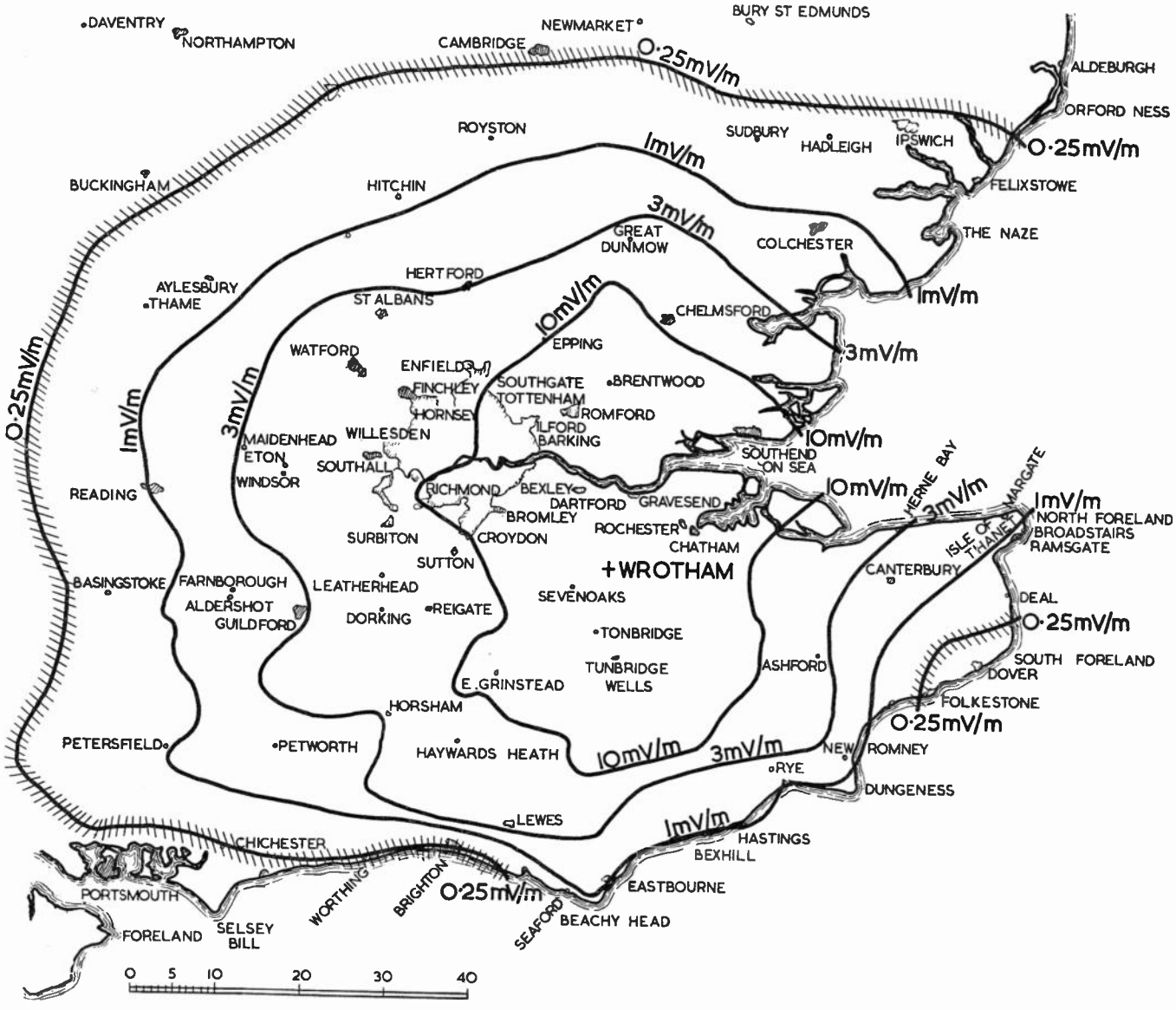



FIG. 5.20.—The field strength contours for the BBC's Wrotham F.M. transmitter.
Site height 725 feet above sea-level. Mast height 406 feet E.R.P. 120 kW. Horizontal polarisation.
*(By courtesy of the British Broadcasting Corporation.)*

[*Facing p.* 144

The curves given in Fig. 5.18 are only strictly correct at one particular frequency—namely, 46 Mc/s. As it is, however, necessary to assess the field strength at widely differing frequencies, the F.C.C. publish another curve which is reproduced as Fig. 5.19. This curve shows the way in which the field strength may be expected to vary as the frequency of the transmission is varied. It will be noted that this chart deals only with one specific case—a dipole transmitting aerial at 1,000 feet above the surrounding country and a receiving dipole at 30 feet. In conclusion it may be stated that field strength tests carried out by the BBC show substantial agreement with the F.C.C. curve; although it is recommended that a value of one-half the idealised field strength should be used.

## Conclusions

As this chapter has covered a very wide field, it is proposed to summarise the more important conclusions which may be drawn from it. Firstly, although the use of the v.h.f. band is essential for all high-fidelity frequency-modulation broadcasts, the short waveband or any other band on which the ionised layers form part of the transmission path, may be used provided that the frequency spectrum occupied by the signal in question is small—preferably less than 2 or 3 kc/s. In cases where this is possible the full gain due to the use of frequency modulation can be realised.

In the case of high-fidelity frequency-modulation broadcasting, reflections from static objects such as buildings and gas-holders start to become troublesome at round about some 150 to 200 Mc/s, although in the case of a frequency-modulated radio telephone system the frequency would have to be raised considerably higher before any noticeable distortion occurred. It therefore follows that the method of polarisation which gives a minimum of reflected signals should be employed. In urban areas it has been shown that horizontal polarisation not only gives a slightly increased field strength, but also results in a smaller amount of selective fading. For radio telephone and other services requiring a comparatively limited frequency band for their transmission, no very great benefit would result from the use of horizontal in place of vertical polarisation. There would be a marked improvement (i.e. a reduction in distortion) on a high-fidelity frequency-modulation transmission.

## SELECTED REFERENCES


TREVOR, B., and CARTER, P. S., Notes on the Propagation of Waves Below Ten Metres in Length, *Proc. I.R.E.*, March 1933.

BARFIELD, Some Measurements of the Electrical Constants of the Ground, *Journal I.E.E.*, August 1934.

SMITH-ROSE, Electrical Measurements on Soil with Alternating Currents, *Journal I.E.E.*, August 1934.

CROSBY, MURRAY G., Frequency Modulation Propagation Characteristics, *Proc. I.R.E.*, June 1936.

BEVERAGE, H. H., Some Notes on Ultra High Frequency Propagation, *R.C.A. Review*, January 1937.

ENGLUND, C. R., CRAWFORD, A. B., and MUMFORD, W. W., Ultra Short Wave Transmission and Atmospheric Irregularities, *The Bell System Tech. Journal*, October 1938.

GEORGE, R. W., A Study of Ultra High Frequency Wide Band Propagation Characteristics, *Proc. I.R.E.*, January 1939.

MACLEAN, K. G., and WICKIZER, G. S., Notes on the Random Fading of 50 Megacycle Signals Over Non-optical Paths, *Proc. I.R.E.*, August 1939.

KATZIN, M., Ultra High Frequency Propagation, *Proc. Radio Club of America*, September 1939.

PETERSON, H. O., Ultra High Frequency Propagation Formulas, *R.C.A. Review*, October 1939.

ENGLUND, C. R., CRAWFORD, A. B., and MUMFORD, W. W., Ultra Short Wave Transmission Over a 39-mile "Optical" Path, *Proc. I.R.E.*, August 1940.

CROSBY, MURRAY G., Observations of Frequency Modulated Propagation on 26 Megacycles, *Proc. I.R.E.*, July 1941.

CHAPMAN, The Sun and the Ionosphere, *Journal I.E.E.*, November 1941.

NORTON, K. A., The Calculation of Ground-Wave Field Intensity over a Finitely Conducting Spherical Earth, *Proc. I.R.E.*, December 1941.

SMITH-ROSE, R. L., and STICKLAND, A. C., A Study of Propagation over the Ultra-Short-Wave Radio Link between Guernsey and England on Wavelengths of 5 and 8 Metres (60 and 37·5 Mc/s), *Journal I.E.E.*, Part III, March 1943.

MCPETRIE, J. S., and SAXTON, J. A., The Determination of the Electrical Properties of Soil at a Wavelength of 5 Metres (Frequency 60 mc/s), *Journal I.E.E.*, Part III, March 1943.

ARMSTRONG, E. H., Discussion of Proposed F.M. Frequencies, *F.M. and Television*, March 1945.

KIRKE, H. L., Frequency Modulation: B.B.C. Field Trials. *BBC Quarterly*, July 1946.

HUND, A., Frequency Modulation (*Wave Propagation in the F.M. Band*, pp. 131–49). McGraw-Hill Book Co.

LADNER, A. W., and STONER, C. R., *Shortwave Wireless Communication*. Chapman and Hall, London.

ALLEN, E. W., Jr., VHF and UHF signal ranges as limited by noise and co-channel interference. *Proc. I.R.E.*, Feb. 1947.

BENNINGTON, T. W., Radio Propagation in the Frequency Range 40–100 Mc/s, *BBC Quarterly*, Vol. II, No. 4, 1948.

BENNINGTON, T. W., Propagation of VHF via Sporadic *E*, *Wireless World*, Jan. 1952.

Tropospheric Propagation: A selected guide to the literature, *Proc. I.R.E.*, May 1953.

MORGAN, H. G., A review of VHF Ionospheric Propagation, *Proc. I.R.E.*, May 1953.

EPSTEIN and PETERSEN, An experimental study of Wave Propagation at 850 Mc/s, *Proc. I.R.E.*, Vol. 41, May 1953.

SAXTON, J. A., Basic ground-wave Propagation Characteristics in the Frequency Band 50–800 Mc/s, *Journal I.E.E.*, Part III, July 1954.

SAXTON, J. A. and HARDEN, B. N., Ground-wave Field-strength Surveys at 100 and 600 Mc/s, *Journal I.E.E.*, Part III, July 1954.

*Chapter Six*

# AERIALS

A VARYING electric field produces a magnetic field in just the same way as an electric current flowing in a conductor sets up a magnetic field. If the current in the conductor is termed conduction current, then the current flowing "through" the dielectric of a perfect condenser (excluding any current due to

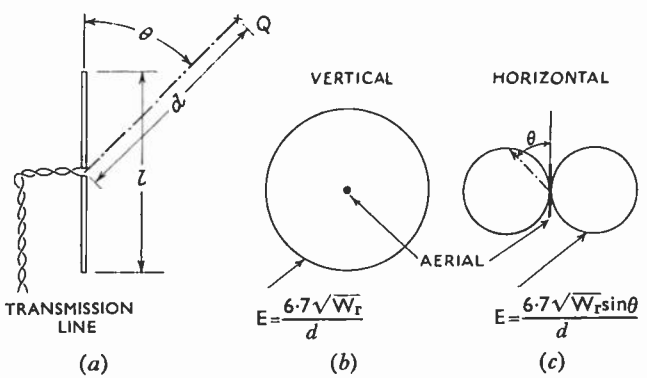



FIG. 6.1.—Diagram (*a*) shows the general form taken by a centre-fed dipole aerial. Diagrams (*b*) and (*c*) show contours of equal field strength when a short aerial of this type is firstly vertical and secondly horizontal.

leakage) may be termed the capacity or displacement current. As far as the resultant magnetic effect is concerned there is no distinguishable difference between that produced by a displacement current and that produced by a conduction current. In the case of displacement current, it should be borne in mind that as the current path lies through a perfect dielectric, then there can be no $I^2R$ loss equivalent to that which occurs when current flows through a conductor of finite conductivity.

If it is now assumed that an alternating current is fed into the centre of a conductor, as shown in Fig. 6.1 (*a*), it will be apparent that conduction current flows into the two rods, and that this current can only find its return path by flowing through the capacity existing between them. In other words, the conduction

current flowing in the rods results in a displacement current in the space outside them. If the frequency of the applied current is the same as that at which the inductive reactance equals the reactance of the total incremental capacity existing between the two rods, then the circulating current, as in any other resonant circuit, will be very large.

Once a varying electric field and its complementary magnetic field have been produced in free space, they will be mutually self-supporting and, there being no finite boundary to limit the space in which they "flow", a self-supporting electro-magnetic wave-motion is set up. This wave-motion, which travels with the velocity of light, has magnetic and electrostatic fields at right angles to each other, and also at right angles to the direction of travel. The total energy content of the electro-magnetic waves set up in this way is termed the radiation loss of the aerial. It will be noted that basically it is possible to consider a radio aerial as a device for producing the largest possible displacement current for a given power input.

Before proceeding to a more detailed study of aerials, it is useful to state some of the basic expressions employed in dealing with electro-magnetic radiation. It is usual to express field strengths in terms of the r.m.s. value of the electric intensity component $E$. This is usually given in volts/metre or microvolts/metre. The r.m.s. value of the magnetic intensity component is however sometimes quoted; this is usually given in ampere-turns/metre or microampere-turns/metre. In a plane polarised plane wave the electric and magnetic vectors are mutually perpendicular and are related by the following expression:

$$E = 120\pi H.$$

Since the dimensions of $E/H$ are ohms, the factor $120\pi$ ($=377$ approximately) is frequently termed the "impedance of free space".

The power associated with such a wave, over a unit area parallel to the plane of the wave vectors, is given by

$$P = E \times H \text{ watts/sq. metre}$$
$$= \frac{E^2}{120\pi} \text{ watts/sq. metre.}$$

By means of this expression, it is possible to calculate the power radiated from an aerial. On the surface of a sphere of very great

radius $R$, centred on the aerial, the wave may be assumed plane; the direction of the electric and magnetic vectors may be taken to lie in the tangent plane at every point. Then the total power radiated by the aerial is given by

$$W_r = \int_0^{\pi} \frac{E_R^2}{60} R^2 \sin \psi \, d\psi, \quad . \quad . \quad . \quad . \quad (6.1)$$

where $\psi$ is the angle made by the radius vector at any point $P$ with an axis arbitrarily taken through the centre of the sphere, and $E_R$ is the electric intensity at the point $P$.

### Field Strength Diagrams of Short Aerials

In the v.h.f. band it is frequently possible to use the simple aerial arrangement illustrated in Fig. 6.1 (*a*). Normally such an aerial has a total length equal to a half wavelength of the radiated signal. This type of aerial, which is termed a half-wave dipole, is normally used as the reference or standard against which the performance of other more complex types is measured. This being so, it provides a suitable point at which to start a general investigation of the behaviour of short-wave aerials.

Ignoring the effect of the earth, the theoretical field strength produced at a point $Q$ situated at a substantial distance from an aerial which is short in comparison with the wavelength being radiated is given by

$$E = \frac{60\pi Il}{\lambda d} \sin \theta \text{ r.m.s. volts/metre}, \quad . \quad . \quad (6.2)$$

where $\theta$=the angle which the line connecting the point $Q$ with the aerial makes with the axis of the aerial;
$d$=the distance in metres to the point $Q$ (it should be noted that $d$ must be large compared with the wavelength);
$I$=r.m.s. current (amps) in the aerial element (assumed constant throughout the element);
$l$=length of aerial element in metres;
$\lambda$=signal wavelength in metres.

This expression may be written using the signal frequency $f$ in megacycles as

$$E = 0{\cdot}628 Ilf \sin \theta / d.$$

The power radiated by such an aerial can be evaluated by means of expression (6.1) above, giving

$$W_r = 0{\cdot}0088 I^2 l^2 f^2.$$

If this field strength distribution is expressed as a polar field strength diagram it will be of the form shown in Fig. 6.1 (*b*). For clarity it will be assumed from now on that the aerial is mounted vertically when this field distribution is under consideration; and waves radiated from an aerial in this position will be referred to as vertically polarised.

The field strength distribution in the plane along which the axis of the aerial lies will vary in accordance with equation (6.2), from which it will be seen that there is a sinusoidal variation from a maximum field strength in the plane perpendicular to the conductor (i.e. the field strength value obtained when the aerial is vertical), to zero along a line corresponding to the axis of the aerial. As this change is sinusoidal, it follows that the polar diagram for a short aerial lying in a horizontal position will be as shown in Fig. 6.1 (*c*).

### Field Strengths produced by Longer Aerials

So far the formulae given are only applicable to short aerials. As soon as the length of the aerial is increased it becomes necessary to take into account its actual length in determining the total field strength it produces at any point. In doing this it is necessary to consider the individual contributions which are made by each element of length *ds* of the aerial. When this is done it is found that at certain angles the signals radiated by the different elementary sections cancel each other out. By totalling up the effect of the signals radiated from all the incremental segments of the aerial, it is possible to calculate the field strength produced at any given point.

### Aerial Current Distribution

In order to calculate the polar diagram of an aerial, it is necessary to make certain assumptions about the distribution of current along the length of the aerial. It is generally assumed, by analogy with transmission line theory, that the current distribution takes the form of a standing wave, the current being zero at the free ends of the aerial. This assumption cannot be entirely accurate,

since if the current distribution conformed to a true standing wave pattern, there could be no energy radiated by the aerial. However, the assumption yields results which are generally very close to those found by experiment, and hence is widely used.

Two types of standing wave pattern are possible; the first is a symmetrical distribution of current about the mid point of the

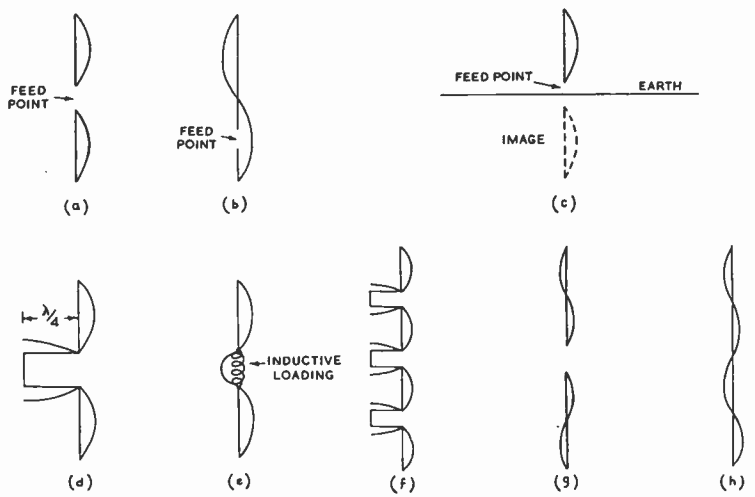



FIG. 6.2.—Current distribution in various types of aerial

(*a*) Symmetrical current distribution
(*b*) Asymmetrical current distribution
(*c*) Unipole aerial and image in ground plane
(*d*) and (*e*) Marconi-Franklin aerial, overall length one wavelength
(*f*) Marconi-Franklin aerial, overall length two wavelengths
(*g*) Symmetrical current distribution in aerial of overall length two wavelengths
(*h*) Asymmetrical current distribution in aerial of overall length two wavelengths

aerial; the second is asymmetrical with respect of the centre. Examples of both types are shown in Fig. 6.2 for an aerial one wavelength long. If the aerial is centre fed, as shown in Fig. 6.2 (*a*), the current distribution is symmetrical; if fed at a point $\lambda/4$ from one end, an asymmetrical distribution results. Symmetrical distribution is encountered wherever an aerial is centre fed, and also if an aerial is end fed over a ground plane (Fig. 6.2 (*c*)). In this case, the current in the aerial and in its image formed in the ground plane have a symmetrical distribution with respect to earth.

Asymmetrical current distribution is generally encountered when the aerial is fed off centre; however, this is not necessarily so, as shown by the two variants of the Marconi-Franklin aerial shown in Fig. 6.2 (*d*) and (*e*); in these examples the overall length of the aerial is one wavelength, and the current distribution is symmetrical, although the aerial may be fed off centre. In example (*c*), the $\lambda/4$ "folded" section is non radiating, and in example (*d*) the inductive loading element can be made sufficiently small to be effectively non radiating.

A Marconi-Franklin aerial which is greater than one wavelength long has a current distribution which falls in neither category; the distribution of such an aerial of overall length $2\lambda$ is shown in Fig. 6.2 (*f*); for comparison the symmetrical distribution is shown in Fig. 6.2 (*g*) and the asymmetrical in Fig. 6.2 (*h*). It is not proposed to deal with the Marconi-Franklin aerial in detail, but attention has been drawn to it to illustrate the fact that the classification is not rigid. The method of determination of the field strength diagram for a Marconi-Franklin aerial will be apparent from a study of the succeeding sections.

It will be appreciated that when the aerial is precisely an odd number of half wavelengths long, the current distribution is the same in both the symmetrical and asymmetrical cases. When the aerial is an even number of half wavelengths long, the difference is most marked; particular care is needed when dealing with a symmetrical distribution aerial if a parasitic element is employed, since the current in the latter will tend to an asymmetrical distribution.

### Dipole with Symmetrical Current Distribution

If it is assumed that the instantaneous value of the current $i$ in each element of length $ds$ of the aerial varies sinusoidally with time ($i=\sqrt{2}\,I\cos\omega t$, where $I$ is the r.m.s. value of the current in $ds$), and that $I$ varies along the length of the aerial sinusoidally also,

$$I=I_{max}\sin\frac{2\pi}{\lambda}\left(\frac{l}{2}-s\right), \quad . \quad . \quad . \quad . \quad (6.3)$$

where $I_{max}$ = the maximum r.m.s. current in the aerial;
$l$ = the overall length of the aerial;
$s$ = the distance from the centre of the aerial to the element $ds$.

This assumption is sufficiently accurate, and provides a working basis on which the field strength at any point $P$ can be computed. The aerial is assumed centre fed; the calculation applies equally well to the field due to an end fed aerial of length $l/2$ above a perfectly conducting surface. In this case, of course, the calculation yields only the field in the hemisphere above the plane; of necessity, the field strength at any point below the plane is zero.

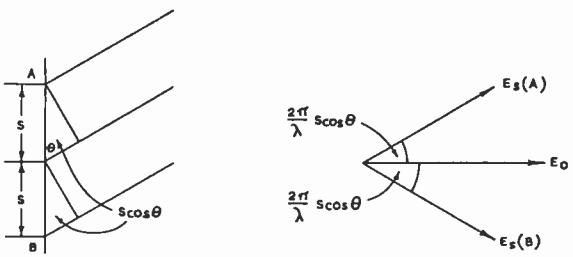



FIG. 6.3—Path length differences for radiation field components due to different points along aerial, and corresponding phase relationships.

With this latter type of aerial, an image of the real aerial is assumed to exist, by virtue of the presence of the conducting plane.

At the point $P$, the components of the resultant field $E_R$ due to each element $ds$ can be considered equal in magnitude, provided that the distance $d$ from the aerial to the point $P$ is large; the r.m.s. magnitude of each such component is given by

$$E = \frac{60\pi I ds}{\lambda d} \sin \theta.$$

The components are, however, not in phase, due to the variation of path length with the position of element $ds$. If the component $E_0$, due to the element at the centre of the aerial ($s=0$) is taken as the reference vector, the component $E_s$ due to any other element differs in phase by an angle $\frac{2\pi}{\lambda} s \cos \theta$. This is shown in Fig. 6.3, for two elements $A$ and $B$ distance $s$ from the centre on opposite sides of the centre.

The resultant of these two latter components is

$$2E_s \cos\left(\frac{2\pi}{\lambda}s\cos\theta\right) = \frac{60\pi I ds \sin\theta}{\lambda d} \cdot 2\cos\left(\frac{2\pi}{\lambda}s\cos\theta\right),$$

and is in phase with the reference vector $E_0$.

The resultant field due to all elements is given by

$$E_R = \int_0^{l/2} \frac{120\pi}{\lambda d} I \sin\theta \cos\left(\frac{2\pi}{\lambda}s\cos\theta\right) ds,$$

but $$I = I_{max} \sin\frac{2\pi}{\lambda}\left(\frac{l}{2} - s\right),$$

$$E_R = \frac{120\pi}{\lambda d} I_{max} \sin\theta \int_0^{l/2} \sin\frac{2\pi}{\lambda}\left(\frac{l}{2} - s\right)\cos\left(\frac{2\pi}{\lambda}s\cos\theta\right) ds$$

$$= \frac{60 I_{max}}{d} \frac{1}{\sin\theta}\left[\cos\left(\frac{2\pi}{\lambda}\frac{l}{2}\cos\theta\right) - \cos\frac{2\pi}{\lambda}\frac{l}{2}\right]$$

$$= \frac{60 I_{max}}{d} F(\theta), \qquad (6.4)$$

where

$$F(\theta) = \frac{\cos\left(\frac{2\pi}{\lambda}\frac{l}{2}\cos\theta\right) - \cos\frac{2\pi}{\lambda}\frac{l}{2}}{\sin\theta} \qquad (6.5)$$

This function shows the variation in magnitude of the radiation field with $\theta$. From expression (6.1), the total power radiated is given by

$$W_r = 60 I_{max}^2 \int_0^{\pi} \frac{\left[\cos\left(\frac{2\pi}{\lambda}\frac{l}{2}\cos\theta\right) - \cos\frac{2\pi}{\lambda}\frac{l}{2}\right]^2}{\sin\theta} d\theta.$$

This expression applies only for a true centre fed aerial; where the

aerial is end fed over a perfectly conducting surface, the power radiated is halved. The expression may be written as

$$W_r = I_{max}^2 R,$$

where
$$R = 60 \int_0^{\pi} \frac{\left[\cos\left(\frac{2\pi}{\lambda}\frac{l}{2}\cos\theta\right) - \cos\frac{2\pi}{\lambda}\frac{l}{2}\right]^2}{\sin\theta} d\theta. \qquad (6.6)$$

The factor $R$ is termed the loop radiation resistance, since it represents the equivalent resistance, which, existing at the point where the aerial current is maximum, would absorb the same power as is dissipated in the radiation field.

From expressions (6.4) and (6.6)

$$E_R = \frac{60\sqrt{W_r}}{d\sqrt{R}} F(\theta). \qquad . \quad . \quad . \quad . \quad (6.7)$$

The expression for the loop radiation resistance may be found from the expression below:

$$R = 30 - \cos\frac{2\pi}{\lambda} l\, S_1\left(\frac{2\pi}{\lambda} 2l\right) + \sin\frac{2\pi}{\lambda} l\, Si\left(\frac{2\pi}{\lambda} 2l\right) + 4\cos^2\frac{2\pi}{\lambda}\frac{l}{2} S_1\left(\frac{2\pi}{\lambda} l\right) - 2\sin\frac{2\pi}{\lambda} l\, Si\left(\frac{2\pi}{\lambda} l\right).$$

The functions $S_1(x)$ and $Si(x)$ are as follows:

$$S_1(x) = \int_0^x \frac{l - \cos u}{u} du = 0{\cdot}5772 + 2{\cdot}303 \log_{10} x - Ci(x),$$

$$Si(x) = \int_0^x \frac{\sin u}{u} du,$$

$$Ci(x) = \int_x^{\infty} \frac{\cos u}{u} du.$$

The table on page 157 indicates the values of $Si(x)$, and $S_1(x)$. If a more complete table is required, reference should be made to *Radio Engineers' Handbook*, by F. E. Terman.

TABLE 9

| $x$ | $Si(x)$ | $S_1(x)$ | $Ci(x)$ | $x$ | $Si(x)$ | $S_1(x)$ | $Ci(x)$ |
|---|---|---|---|---|---|---|---|
| 0·0 | 0·0000 | 0·0000 | — ∞ | 10 | 1·6583 | 2·9253 | —0·0455 |
| 0·2 | 0·1996 | 0·0100 | —1·0422 | 11 | 1·5783 | 3·0647 | —0·0896 |
| 0·4 | 0·3965 | 0·0397 | —0·3788 | 12 | 1·5050 | 3·1119 | —0·0498 |
| 0·6 | 0·5881 | 0·0887 | —0·0223 | 13 | 1·4994 | 3·1154 | 0·0268 |
| 0·8 | 0·7721 | 0·1558 | 0·1983 | 14 | 1·5562 | 3·1469 | 0·0694 |
| 1·0 | 0·9461 | 0·2398 | 0·3374 | 15 | 1·6182 | 3·2390 | 0·0463 |
| 1·5 | 1·3247 | 0·5123 | 0·4704 | 16 | 1·6313 | 3·3640 | — |
| 2·0 | 1·6054 | 0·8474 | 0·4230 | 17 | 1·5901 | 3·4657 | — |
| 2·5 | 1·7785 | 1·2076 | 0·2859 | 18 | 1·5366 | 3·5111 | — |
| 3·0 | 1·8486 | 1·5562 | 0·1196 | 19 | 1·5186 | 3·5166 | — |
| 3·5 | 1·8331 | 1·8621 | 0·0321 | 20 | 1·5482 | 3·5285 | 0·0444 |
| 4·0 | 1·7582 | 2·1045 | —0·1410 | 21 | 1·5949 | 3·5808 | — |
| 4·5 | 1·6541 | 2·2748 | —0·1935 | 22 | 1·6161 | 3·6666 | — |
| 5·0 | 1·5499 | 2·3767 | —0·1900 | 23 | 1·5955 | 3·7484 | — |
| 5·5 | 1·4687 | 2·4240 | — | 24 | 1·5547 | 3·7936 | — |
| 6·0 | 1·4247 | 2·4370 | —0·0681 | 25 | 1·5315 | 3·8029 | —0·0068 |
| 6·5 | 1·4218 | 2·4379 | — | 50 | 1·5516 | 4·4949 | 0·1863 |
| 7·0 | 1·4546 | 2·2464 | 0·0767 | | | | |
| 7·5 | 1·5107 | 2·4765 | — | | | | |
| 8·0 | 1·5742 | 2·5342 | 0·1224 | | | | |
| 8·5 | 1·6396 | 2·6179 | — | | | | |
| 9·0 | 1·6650 | 2·7191 | 0·0553 | | | | |
| 9·5 | 1·6745 | 2·8258 | — | | | | |

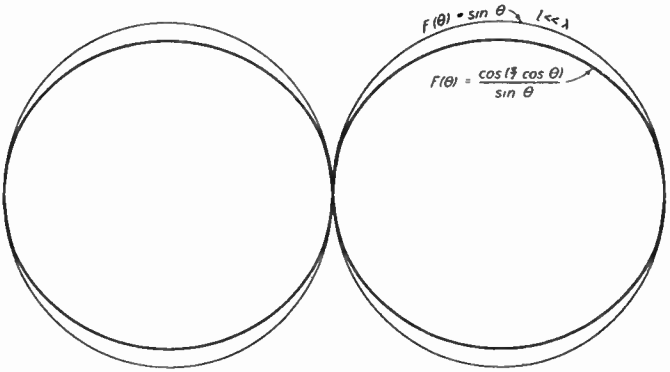



FIG. 6.4 (*A*).

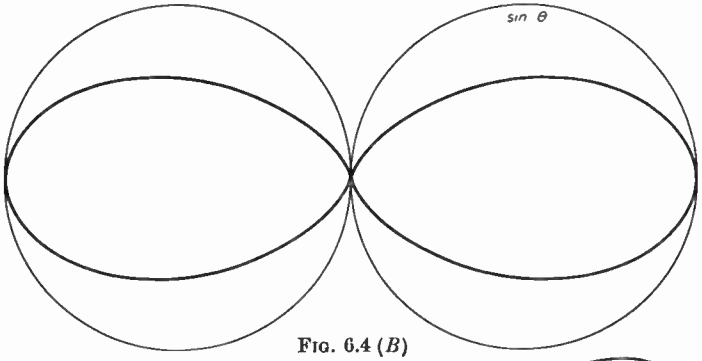



FIG. 6.4 (*B*)

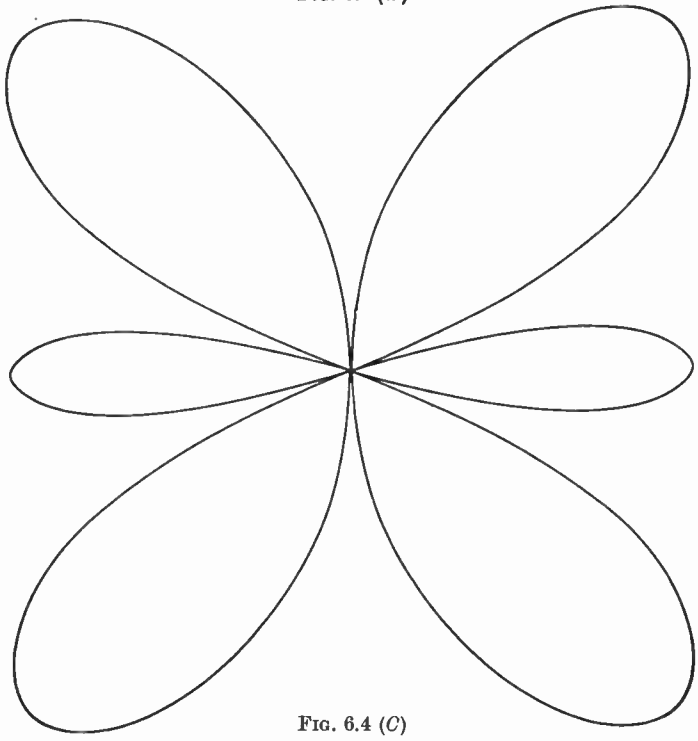


FIG. 6.4 (*C*)

FIG. 6.4.—Calculated polar diagrams for dipoles having symmetrical current distribution in free space. The outer line in (*A*) shows the field strength distribution for a very short aerial ($F(\theta)=\sin\theta$), and the inner that for a half-wave aerial $\left[F(\theta)=\dfrac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}\right]$. Diagrams (*B*) and (*C*) show the distribution for a full wavelength aerial $\left[F(\theta)=\dfrac{\cos(\pi\cos\theta)+1}{\sin\theta}\right]$ and a 1½ wavelength aerial $\left[F(\theta)=\dfrac{\cos\frac{2\pi}{2}\cos\theta}{\sin\theta}\right]$. The axis of the aerial is assumed vertical.

(*From "Ultra High Frequency Techniques". Edited by J. G. Brainerd.*)

The following special cases are of particular interest, and the values of $F(\theta)$ and $R$ are tabulated:

| | $F(\theta)$ | $R$ |
|---|---|---|
| 1. A very short aerial length | $\sin\theta$ | $1{\cdot}25\left(\frac{2\pi}{\lambda}l\right)^4$ |
| 2. A half-wave aerial | $\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}$ | 73 |
| 3. A full wave aerial | $\frac{\cos(\pi\cos\theta)-1}{\sin\theta}$ | 199 |
| 4. One and a half wave aerial | $\frac{\cos\left(\frac{3\pi}{2}\cos\theta\right)}{\sin\theta}$ | 105 |

The polar diagrams for aerials of overall length $\lambda/2$, $\lambda$ and $3\lambda/2$ are shown in Fig. 6.4, together with that of a very short aerial.

### Dipole with Asymmetrical Current Distribution

The polar diagram and radiation resistance of a dipole aerial with this type of current distribution differ appreciably from those of a dipole with symmetrical current distribution, when the overall length of the aerial is in the region of an integral number of wavelengths. Where the aerial is an odd number of half wavelengths long, the polar diagram and radiation resistance are the same, as would be expected, since the current distributions are identical.

It is only possible to deal here with dipoles of overall lengths which are an integral number of half wavelengths; at other lengths the current distribution departs radically from the sinusoidal pattern postulated, and the discussion of such aerials is beyond the scope of this book. With the limitation, then, that the overall length of the aerial is $m\frac{\lambda}{2}$, where $m$ is integral, the polar diagram function $F(\theta)$ can be evaluated by a method similar to that employed with symmetrical current distribution; $F(\theta)$ is given by

$$F(\theta)=\frac{1}{\sin\theta}\left[1+\cos^2(\pi m\cos\theta)-2\cos(\pi m\cos\theta)\cos\pi m\right]^{\frac{1}{2}}. \quad (6.8)$$

The radiation resistance is given by

$$R=30[S_1(2\pi m)]. \quad . \quad . \quad . \quad . \quad . \quad (6.9)$$

For the particular case of $m=2$, i.e. the overall aerial length is one wavelength, the values of $F(\theta)$ and $R$ are given by

$$F(\theta)=\frac{1}{\sin\theta}[1-\cos(2\pi\cos\theta)]$$

$$=\frac{2\sin^2(\pi\cos\theta)}{\sin\theta},$$

$$R=93.$$

The polar diagram function $F(\theta)$ is shown in Fig. 6.5, from which it will be seen that it differs appreciably from the corresponding diagram (Fig. 6.4 (*B*)) for the symmetrical current distribution aerial.

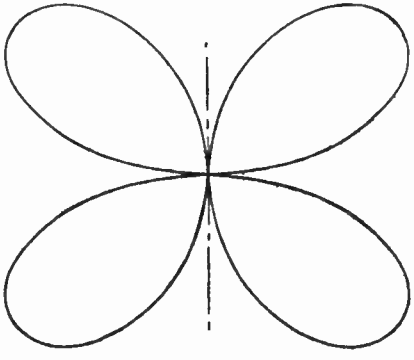


FIG. 6.5.—Calculated polar diagram for an aerial of one wavelength, having asymmetrical current distribution $F(\theta)=2\frac{\sin^2(\pi\cos\theta)}{\sin\theta}$.

(*From "Ultra High Frequency Techniques". Edited by J. G. Brainerd.*)

## Unipole Aerials

The unipole aerial, an example of which is shown in Fig. 6.39, is basically similar to a dipole having symmetrical current distribution, except that one element of the dipole is replaced by an image existing in an earth plane at the foot of the unipole. This type of aerial has the advantage that it is easily fed by a concentric feeder, the outer of the feeder being attached to the ground plane. The ground plane may comprise a circular conducting disc, the

diameter of which is not important provided that it exceeds $\lambda/2$ in radius. More usually, the ground plane consists of two $\lambda/2$ elements at right angles.

The radiation resistance and input impedance of a unipole is half that of the corresponding symmetrical current dipole.

### Accuracy of Assumption of Current Distribution in Aerials

The foregoing expressions for the polar diagrams and radiation resistance of an aerial postulated a current distribution of the form of a sinusoidal standing wave pattern. As pointed out earlier, this assumption cannot be completely accurate, since with such a distribution, no energy could be radiated. Stratton has examined the case of an aerial carrying a travelling current wave, and has found that the radiation resistance tends to a higher value than predicted on the basis of a standing wave pattern. It is therefore to be expected in practice that the radiation resistance will be slightly higher than the values predicted previously.

It should also be noted that the assumption of a standing wave current pattern becomes more in error as the aerial length is increased, and for very long aerials the results are seriously in error. For aerials of the order of less than two wavelengths long, however, the assumption of a standing wave of current leads to results which are close to those found in practice.

### The Receiving Aerial

Whilst this chapter is almost entirely devoted to the discussion of the properties of aerials for transmission, the properties of a given aerial in respect of input impedance and polar diagram are identical whether the aerial is used for transmission or reception. To complete the information, however, it is necessary to know the output voltage obtained from an aerial when in a region of known field strength. The aerial open-circuit r.m.s. voltage is given by

$$e = \frac{E\lambda}{\pi}\sqrt{G},$$

where $E$ is the r.m.s. field strength, $G$ is the power gain of the aerial referred to a half-wave dipole, and $\lambda$ is the signal wavelength.

Whilst it is true that this output falls linearly with $\lambda$, it must be remembered that field strength tends to rise with decreasing $\lambda$ (see expression 5.3). The net result is that, under ideal constant

power transmission conditions, the open-circuit voltage of a resonant half-wave dipole is independent of signal frequency.

## The Input Impedance of Dipole Aerials

In order to efficiently feed power into a dipole for transmission purposes or to draw power from it (as is desired in the case of reception), it is necessary to have a fairly precise knowledge of the aerial's input impedance and the various factors upon which it is

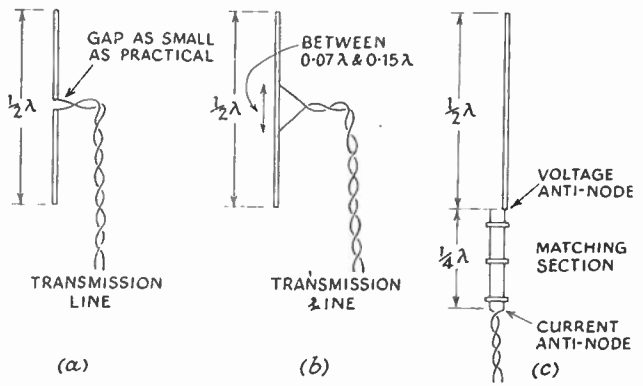



FIG. 6.6.—The three conventional methods of matching a half-wave dipole into the transmission line. Diagram (*a*) shows the split centre-fed arrangement, (*b*) the feeder tapped into the dipole at a point having the same impedance as the feeder, and (*c*) shows the use of a quarter-wavelength matching section in order to feed the aerial at a voltage anti-node.

dependent. Fig. 6.6 shows the three conventional methods of matching the transmission line into a half-wave dipole aerial. For reasons of convenience the first of these methods is used almost exclusively.

A good approximation to the input impedance of a centre fed aerial can be found by treating the aerial, considered a section of transmission line, as matching the radiation resistance to the feeder. The radiation resistance is considered to be situated at the point where $I_{max}$ exists. The characteristic impedance of the transmission line to which the aerial is assumed equivalent is given by

$$Z_0 = 120\left[\log_e \frac{l}{a} - 1\right], \quad . \quad . \quad . \quad . \quad (6.10)$$

where $l$ is the overall length of the aerial, and $a$ is its radius

measured in the same units. The value of $Z_0$ for a range of values of $l/a$ is plotted in Fig. 6.8.

Since the point where $I_{max}$ exists is $\lambda/4$ from the end of the aerial, the equivalent diagram for determining the input impedance is as shown in Fig. 6.7. This equivalent current can also be employed

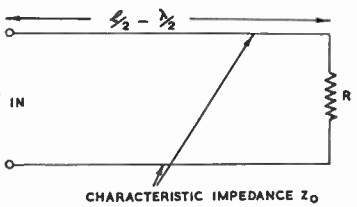



FIG. 6.7.—Equivalent diagram for a centre fed aerial, for determining the value of the input impedance.

when the overall aerial length is less than $\lambda/2$; in this case, the length of the section of transmission line is given by $\lambda/2 - l/2$.

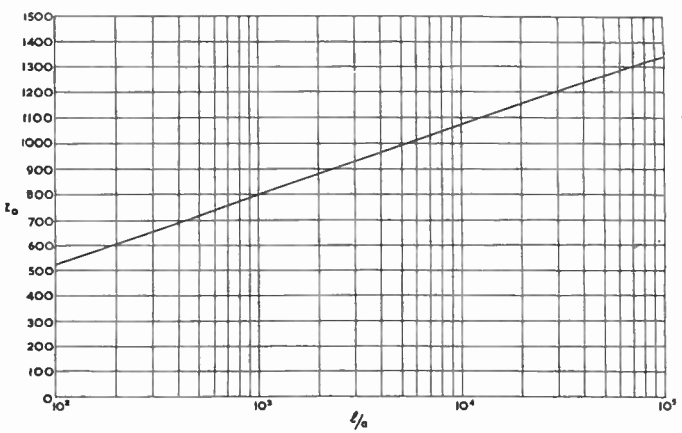



FIG. 6.8.—"Characteristic Impedance" of aerial $Z_0 = 120\left(\log_e \frac{l}{a} - 1\right)$.

It is shown in standard works covering transmission line theory that the input impedance to a transmission line of characteristic impedance $Z_0$, length $L$, terminated by a load impedance $Z_L$ is given by

$$Z_{IN} = Z_0 \cdot \frac{Z_L + jZ_0 \tan \frac{2\pi}{\lambda} L}{Z_0 + jZ_L \tan \frac{2\pi}{\lambda} L} \quad . \quad . \quad . \quad (6.11)$$

A case of particular interest is that of the half-wave dipole, with $l=\lambda/2$. With this particular value of $l$, $L=0$, and hence the input impedance would appear to be equal to the radiation resistance. The input impedance of such an aerial contains, however, a reactive component. This is due to the fact that, more accurately, the radiation resistance should be replaced by an impedance having both resistive and reactive components. For a very thin $\lambda/2$ dipole, the value of this impedance tends to $73{\cdot}2+j42{\cdot}5$ ohms. Both of these components and also $Z_0$ vary slowly with $l$ in the region of $\lambda/2$, and in order to determine at what overall length the dipole input impedance is purely resistive, the impedance $Z_L=73{\cdot}2+j42{\cdot}5$, may be inserted in expression (6.11) and the value of $L$ found for which the reactive component goes to zero. Since $Z_0$ is in general much greater than $Z_L$, and $L$ is very small, expression (6.11) can be simplified for this purpose to

$$Z_{IN}=Z_L+jZ_0 \tan \frac{2\pi}{\lambda} L$$

$$\simeq Z_L+jZ_0 \frac{2\pi}{\lambda} L.$$

The input impedance is thus purely resistive when

$$jZ_0 \frac{2\pi}{\lambda} L+j42{\cdot}5=0,$$

i.e. when $L = -\dfrac{42{\cdot}5\lambda}{2\pi Z_0}$.

The overall length of a resonant dipole in the region of $\lambda/2$ overall length is therefore shorter by an amount $2L$ than $\lambda/2$. That is, the decrease in overall length $dl$, necessary for resonance is

$$dl = \frac{13{\cdot}5\lambda}{Z_0}, \qquad \ldots\ldots (6.12)$$

and the decrease in overall length per cent is thus $2{,}700/Z_0$. The input resistance at resonance will be somewhat lower than the value of 73·2 ohms due to the decrease of aerial length. It will be seen from the values of $Z_0$ in Fig. 6.8, that for practical aerials, the overall length for resonance is less than $\lambda/2$ by a factor of from 2 to 6 per cent.

The foregoing can also be applied to compute the change in input reactance for small changes in working frequency and overall length for a resonant dipole. A change of length of $\frac{2,700}{Z_0}$ per cent produces a change of reactance of 42·5 ohms; therefore, in this region, the reactance increases (decreases) at the rate of $0{\cdot}016Z_0$ ohms for a one per cent increase (decrease) in frequency or overall length.

Obviously, therefore, for wide-band operation, where it is desired to minimise the reactive component throughout the working range of frequencies, it is advantageous to employ elements of large diameter to ensure low values of $Z_0$. The desired result can be obtained with a large continuous conductor surface, or by means of a cage of parallel spaced wires. This latter open type of construction reduces both wind resistance and weight.

For an aerial 10 feet long, diameter $\frac{1}{2}$ inch, and resonant in the region of 49·5 Mc/s, it is of interest to note the rate of change of the resistive and reactive components with frequency. From Fig. 6.8, $Z_0=620$ ohms, and, therefore, the change of input reactance for a one per cent change in frequency is about 10 ohms. By comparison, the figure for the corresponding charge in resistance is about 2·5 ohms. For this particular aerial, the physical length of which is shorter by about $4\frac{1}{2}$ per cent than $\lambda/2$ for resonance, the resistive component at resonance will be in the order of 65–70 ohms. The variation of the resistive and reactive components of the input impedance of this aerial with frequency are shown in Fig. 6.9, together with the variation of the magnitude of the input impedance.

If expression (6.11) is applied to determine the input impedance of a dipole of overall length one wavelength, the input resistance at resonance is given by

$$Z_{IN}=\frac{Z_0^2}{199}. \qquad \ldots\ldots \quad (6.13)$$

This value is commonly in the region of 1–10 kilohms; if the current distribution were truly sinusoidal, it would, of course, be infinite. In this expression for the input impedance, the aerial would appear to be resonant when the overall length is precisely one wavelength; resonance, however, actually occurs at a length slightly shorter than this. R. A. Smith states that, to a first

degree of approximation, the aerial should be shorter than one wavelength by an amount $\frac{40\lambda}{Z_0}$.

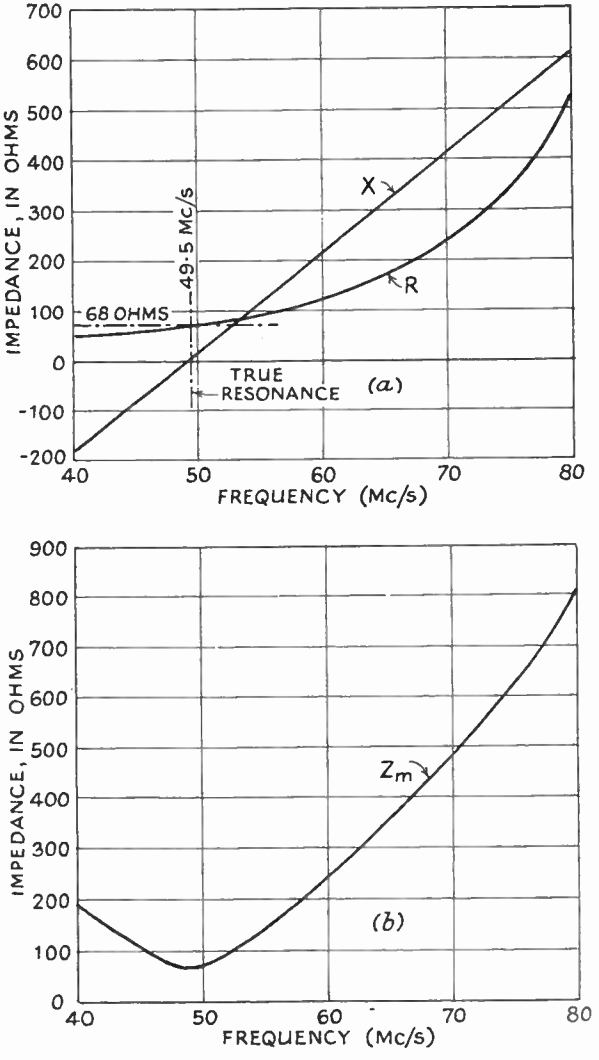



FIG. 6.9.—The upper diagram shows the effective input resistance $R$ and the reactance component $X$ of a centre-fed dipole 10 feet long and $\frac{1}{2}$ inch diameter. The lower diagram shows the variation of the input impedance of the same aerial over the frequency range.

## Folded Dipole

Where it is desired that an aerial shall have a low reactive component of the input impedance over a wide band of frequencies, the folded dipole is frequently employed. This type of aerial is shown in Fig. 6.10 and differs only from the normal dipole in that a second element is present, close to the first, and joined to it at the ends.

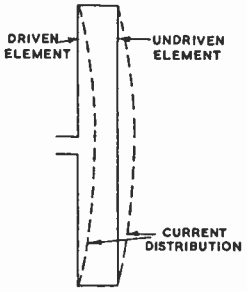



FIG. 6.10.—Folded dipole construction and current distribution (overall length $\lambda/2$).

With equal diameter elements the input resistance at resonance is equal to four times that of a normal dipole. This is due to the fact that the current in the second element will be equal in amplitude and in phase with that of the current in the first; thus, as far as the radiation field is concerned, the aerial behaves as a single dipole fed with a current equal to twice the actual input current. For the same radiated power, therefore, the input resistance of a folded dipole must be four times that of a single dipole. This fact is frequently useful where an aerial employing a number of parasitic elements is used; these latter have the effect of lowering the input resistance of the aerial, so that the resulting value with a normal dipole may be inconveniently small.

In the region near resonance, the folded dipole exhibits very little change of reactance with frequency or length. The reason for this form is that the two folded sections of the aerial behave as short circuited $\lambda/4$ sections of transmission line; the reactance slope of these elements is opposite to that of the reactance slope of a normal dipole.

By judicious choice of element spacing, and/or the diameters of the elements, cancellation of the reactive component of the input impedance can be achieved over a band of frequencies of some $\pm 10$ per cent about the resonant frequency. The input impedance of this type of aerial is considered in more detail later; the wide band characteristics of a number of types of aerial are shown in Fig. 6.11.

Advantage can be taken of the configuration of the folded dipole to support the aerial at the centre of the undriven element; since this point is at zero potential, it can be earthed. A very robust mechanical construction can thereby be achieved.

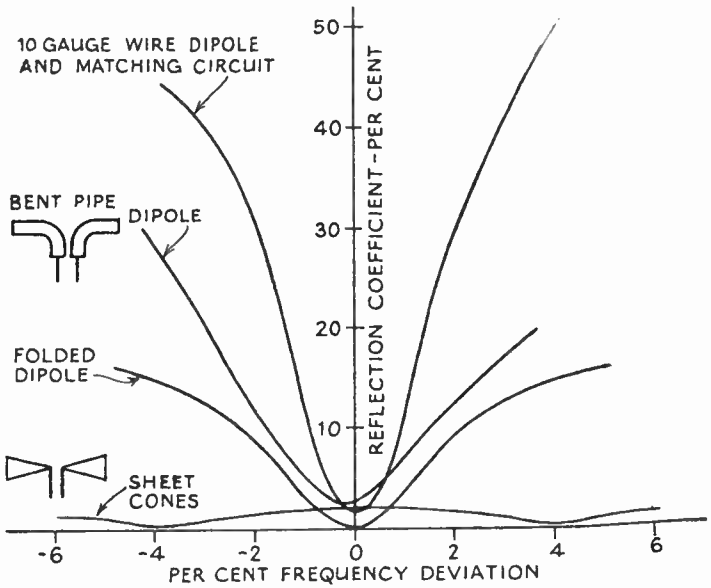



FIG. 6.11.—The selectivity of various types of dipole as measured by P. S. Carter. The values given are those for the reflection coefficient obtained between the aerial and a two-wire transmission line.

## Aerial Radiators and Parasitic Elements

When an undriven aerial element is introduced in the neighbourhood of a driven element, a number of changes occur.

The input impedance of the driven element, and the shape of the polar diagram can be calculated with the aid of the table of mutual impedances given on page 170.

If $i_1$ is the current at the centre of the driven element, and $i_2$ the current at the centre of the parasitic element, then

$$e_1 = Z_{11}i_1 + Z_{12}i_2, \quad . \quad . \quad . \quad . \quad . \quad (6.14)$$

and

$$0 = Z_{21}i_1 + Z_{22}i_2, \quad . \quad . \quad . \quad . \quad . \quad (6.15)$$

where $Z_{11}$ is the centre point impedance of the driven element alone;

$Z_{12} = Z_{21}$ is the mutual impedance between the dipoles (see table for values);

$Z_{22}$ is the centre point impedance of the parasitic element;

and $e_1$ is the driving-point voltage of the driven element.

From expression (6.15),

$$i_2 = -\frac{Z_{21}}{Z_{22}}\, i_1,$$

and hence

$$e_1 = \left(Z_{11} - \frac{Z_{12}^2}{Z_{22}}\right) i_1,$$

whence

$$Z_{IN} = \frac{e_1}{i_1} = Z_{11} - \frac{Z_{12}^2}{Z_{22}},$$

where $Z_{IN}$ is the centre point impedance of the driven element in the presence of the parasitic element.

As an example of the use of these expressions, we may take the case of element spacing $=0{\cdot}12\lambda$, where $Z_{12}$ is purely resistive and equal to 65 ohms approximately. If the parasitic element has a centre point impedance of $73+j68$ ohms (the optimum value for a reflector at this spacing, as shown later), then assuming the driven element to be resonant, and its input impedance equal to 73 ohms,

$$Z_{IN} = 73 - \frac{65^2}{73+j68}$$
$$= 24{\cdot}6 + j20{\cdot}3.$$

In order to determine the polar diagram of the aerial, the phase angle $\phi$ of the current in the parasitic element must be found. This can be done by applying expression (6.15). In the case considered above,

$$i_2 = -\frac{65}{73+j68}\, i_1$$
$$= (-0{\cdot}47 + j0{\cdot}44)\, i_1,$$

approximately, therefore the magnitude of $i_2$ is equal to $0{\cdot}65 i_1$, and the phase angle between the currents is 137°. Thus at a point remote from the aerial system the field strength comprises two components, of relative magnitudes $E$ and $0{\cdot}65E$, having a phase angle between them of 137° plus the phase angle difference due to path length difference. At a point in line with the elements, the phase angle difference due to path length difference is

$$\frac{360}{\lambda} \times 0{\cdot}12\lambda \simeq 43°.$$

*Table of Mutual Impedances*

Parallel $\lambda/2$ dipoles; distance between dipoles $b$

| $\frac{b}{\lambda}$ | $R_{12}$ | $X_{12}$ |
|---|---|---|
| 0·00 | 73·3 | 42·2 |
| 0·05 | 71·7 | 24·3 |
| 0·10 | 67·4 | 7·6 |
| 0·15 | 60·4 | −7·1 |
| 0·20 | 51·4 | −19·1 |
| 0·25 | 40·8 | −28·3 |
| 0·30 | 29·2 | −34·6 |
| 0·35 | 17·5 | −37·4 |
| 0·40 | 6·0 | −37·6 |
| 0·45 | −3·4 | −34·8 |
| 0·50 | −12·5 | −29·9 |
| 0·55 | −19·0 | −23·4 |
| 0·60 | −23·2 | −16·5 |
| 0·65 | −25·2 | −8·0 |
| 0·70 | −24·6 | −0·1 |
| 0·75 | −22·5 | 6·6 |
| 0·80 | −18·5 | 12·2 |
| 0·85 | −13·0 | 16·3 |
| 0·90 | −7·5 | 18·5 |
| 0·95 | −1·5 | 19·0 |
| 1·00 | 4·0 | 17·3 |
| 1·1 | 12·3 | 11·2 |
| 1·2 | 15·2 | 2·0 |
| 1·3 | 12·6 | −6·7 |
| 1·4 | 6·0 | −11·8 |
| 1·5 | −1·9 | −12·6 |
| 1·6 | −8·1 | −8·4 |
| 1·7 | −10·9 | −2·0 |
| 1·8 | −9·1 | 4·5 |
| 1·9 | −5·4 | −5·3 |
| 2·0 | 1·1 | 9·2 |
| 2·1 | 6·0 | 6·7 |
| 2·2 | 8·2 | 1·8 |
| 2·3 | 7·5 | −3·3 |
| 2·4 | 4·0 | −6·8 |
| 2·5 | 0·9 | −7·3 |
| 2·6 | −4·8 | −5·2 |
| 2·7 | −6·7 | −1·7 |
| 2·8 | −6·3 | 2·6 |
| 2·9 | −3·4 | 5·6 |
| 3·0 | −0·3 | 5·8 |

With the driven element nearer, the total phase difference is 94°, and the magnitude of the resultant field strength is thus increased by a factor of approximately $\sqrt{1+(0{\cdot}65)^2}=1{\cdot}18$ approximately. At a point in line with the elements, but with the parasitic element nearer, the phase difference is 180°, i.e. the components are in opposition, and the field strength is multiplied by a factor $1-0{\cdot}65=0{\cdot}35$ approximately. The front-to-back ratio of such an aerial is therefore about 3·4 or 10·6 db. In order to obtain the requisite centre impedance of the parasitic element of $73+j68$ ohms, the element must be longer than the resonant length; the actual length can be calculated by the methods described earlier. In order to obtain a substantially resistive centre impedance at the driven element, this may be made slightly longer than its resonant length so that the reactance component will cancel the reactance component due to the presence of the parasitic element ($+j20{\cdot}3$ ohms).

If the calculation is repeated for a parasitic element, cut to a length shorter than its resonant length, the parasitic element will be found to act as a director.

Distribution curves covering a large number of practical cases have been published by G. H. Brown (*Proc. I.R.E.* for January 1937). Fig. 6.12 shows a series of curves for a quarter-wave dipole and parasitic element, both the phase angle of the current in the parasitic element and the spacing being varied.

The presence of the parasitic element causes the input impedance of the aerial to vary more rapidly with frequency; additionally, the impedance changes much more rapidly on one side of the resonant frequency than the other. The region of rapid impedance change is above or below the resonant frequency according to whether the parasitic element is a director or reflector. This is shown in Fig. 6.13, where the voltage delivered to a load, matched at the resonant frequency, is plotted against frequency.

### Slot Aerials

The basic type of slot aerial comprises a slot cut in an infinite sheet of conducting material, the feed points being situated on opposite sides of the slot, as shown in Fig. 6.14 (*a*). The characteristics of the aerial are closely related to those of the corresponding strip dipole aerial which would close the slot. The electric and

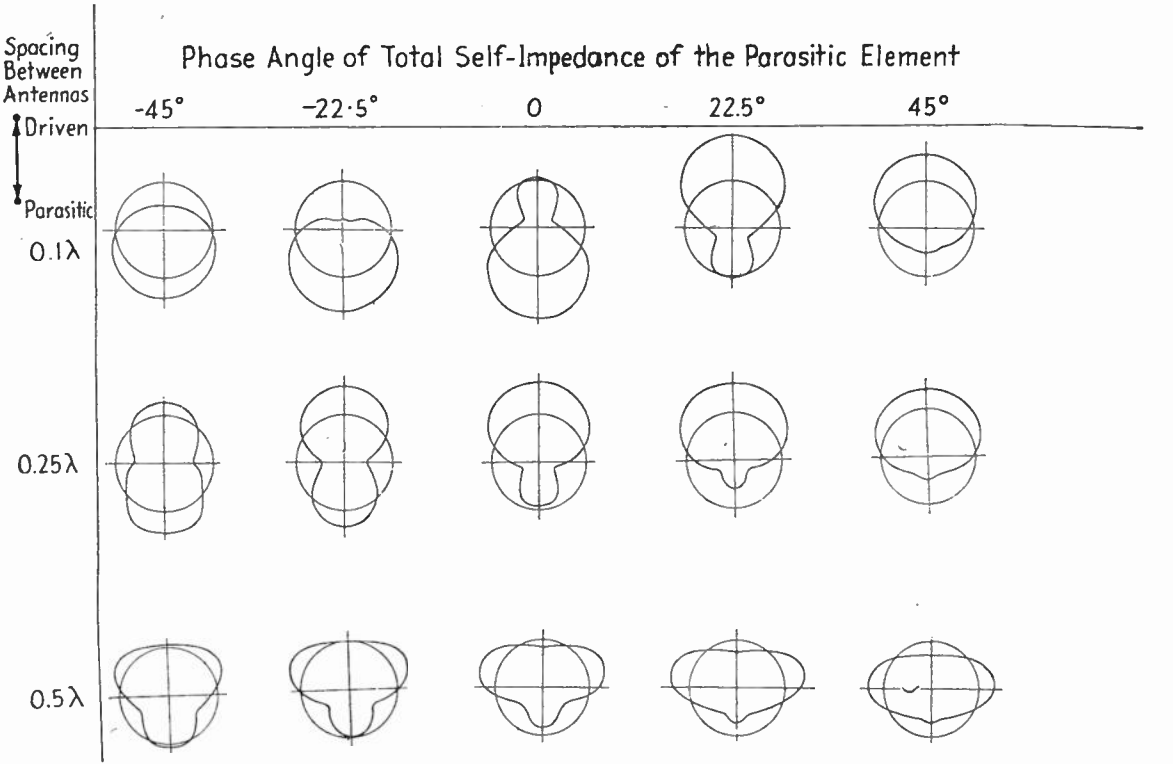



FIG. 6.12.—The horizontal patterns of two vertical quarter-wave aerials, one driven and the other parasitic. The distance between aerials is indicated on the left, the phase angle of the self-impedance of the parasitic aerial at the top. Note that for the tuned case (phase angle at zero), maximum radiation is in the direction of the parasitic aerial, which is then called a director. For a phase angle of +22·5° the converse is true and the parasitic aerial is called a reflector.

*(From G. H. Brown, Proc. I.R.E., Jan. 1937.)*

magnetic components associated with the two aerials are interchanged; whereas, with the dipole the electric field component is parallel to the axis of the dipole, with the slot aerial it is perpendicular to the axis of the slot. The vertical slot aerial therefore has the very valuable property of producing a horizontally polarised

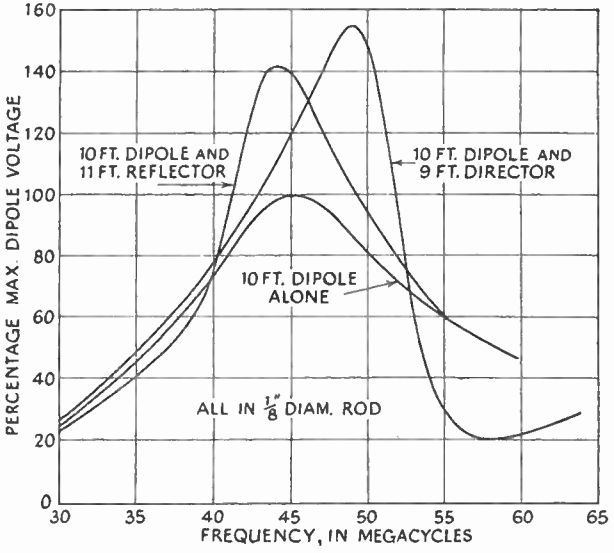



FIG. 6.13.—The selectivity characteristics of a simple dipole when loaded with a matched transmission line—as measured by Holmes and Turner. A reflector increases the slope of the selectivity characteristic on the low frequency side of resonance and the director similarly steepens the high frequency side.

radiation field. The shapes of the polar diagrams of the slot aerial and its corresponding dipole aerial are precisely similar.

That the radiation field from such an aerial is horizontally polarised may be seen from the fact that the currents in the conducting sheet in the neighbourhood of one side of the slot are everywhere flowing in the opposite direction to the currents at the other side of the slot, and hence the vertically polarised components due to these currents tend to cancel. At the ends of the slots, however, where the currents are at maximum, the currents flow in the same direction. Thus the major contribution to the radiation field is by the currents flowing in a direction perpendicular to the length of the slot, and hence the electric intensity is everywhere at right angles to the slot axis.

The distribution of current and voltage along a slot aerial are again similar to those existing on the corresponding dipole, except that the standing current and voltage waves are interchanged. Thus, whereas the half-wave dipole has a relatively low centre

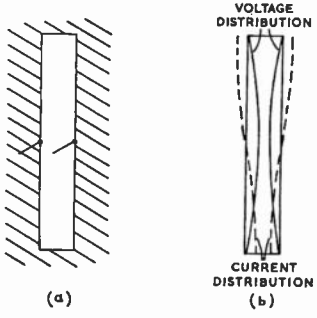



FIG. 6.14.—(*a*) Slot aerial cut in infinite conducting sheet. (*b*) Current and voltage distribution at edges of slot aerial. Instantaneous direction of current flow shown by arrows.

point impedance, the half-wave slot aerial has a high impedance. Similarly, the slot aerial of overall length one wavelength has a low centre point impedance, whereas that of the corresponding dipole impedance is high.

It can be shown that the input impedance of a slot aerial $Z_S$ and its complementary dipole aerial $Z_D$ are related by the expression

$$Z_S Z_D = \tfrac{1}{4}(377)^2. \quad . \quad . \quad . \quad . \quad . \quad (6.16)$$

In this expression, the term 377 is the "impedance of free space" referred to earlier.

It should be noted that the dipole aerial related by this expression to the slot aerial is a "strip" dipole. The "characteristic impedance" of this type of dipole is given by the expression

$$Z_0 = 120\left(\log_e \frac{41}{w} - 1\right), \quad . \quad . \quad . \quad (6.17)$$

where $w$ is the width of the strip and $l$ is its overall length, as before. The factor 4 is introduced because the strip dipole behaves as a cylindrical dipole having a radius equal to one quarter of the width of the strip.

As suggested by expression (6.16), the slope of the reactance-frequency curve of a slot aerial is opposite in sign to that of a dipole aerial. The reactive component of the input impedance of a slot aerial precisely one half wavelength long is capacitive; as the length of the slot is decreased, the reactive component decreases to zero, and thus, for resonance, the slot is somewhat less than $\lambda/2$ long. At resonance the input resistance can be determined from the input impedance at resonance of the corresponding dipole. If a value of 73 ohms is assumed for the latter, the theoretical input impedance of the slot aerial is in the region of 485 ohms. In practice, the value found is somewhat lower than this, being in the region of 350–400 ohms. As with a dipole, the reactive component of the input impedance over a range of frequencies can be minimised by increasing the slot width.

With a slot cut in a finite sheet, the polar diagram departs from the circular pattern in the plane perpendicular to the axis of the slot, due to diffraction effects. Provided that the sides of the sheet are greater than $5\lambda$, the polar diagram does not depart appreciably from that obtained with an infinite sheet, except near the plane of the sheet. With sheets of dimensions smaller than this, the polar diagram tends to a figure-of-eight pattern in the plane at right angles to the slot axis, accompanied by a narrowing of the pattern in the plane of the axis.

### Boxed Slot Aerial

For practical reasons, it is frequently inconvenient to employ the basic type of slot aerial described above, and alternative types of slot aerials have been evolved. It is often desired to suppress radiation from one side of a slot aerial, and to achieve this one side of a slot aerial may be enclosed or "boxed". The enclosure is frequently of skeleton form comprising bars perpendicular to the axis of the slot, spaced at intervals of $\lambda/10$ or less, as shown in Fig. 6.15. The effect of enclosing a slot aerial in this fashion is to double the resistive component of the input impedance, and also to provide an additional reactive component to the input impedance. The enclosure serves effectively to load the slot reactively

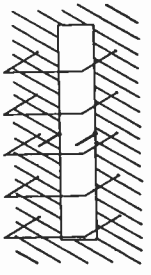


FIG. 6.15.—Boxed slot aerial, skeleton enclosure.

throughout its length, and generally necessitates an alteration of slot length to maintain resonance.

### Slotted Cylindrical Aerial

In this type of aerial, the slot is cut in a cylinder, the axis of the slot being parallel to that of the cylinder itself. In this type of aerial, the slot is "boxed" by the cylinder itself. The polar diagram in the plane at right angles to the slot axis is heart shaped, the direction of maximum radiation being along the radial containing the slot itself. As the diameter of the cylinder is reduced the polar diagram becomes more nearly circular, and for cylinders of diameters between $\lambda/10$ and $\lambda/8$, the radiation pattern is very nearly uniform. The cylinder provides reactive loading of the

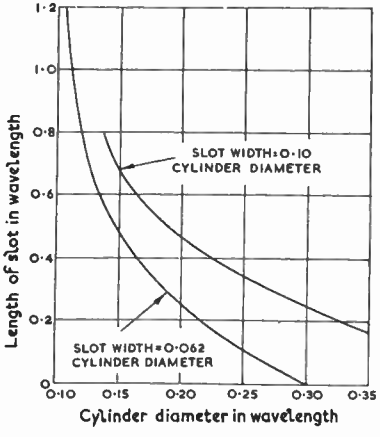



FIG. 6.16.—Length of slot in cylinder for resonance for varying values of cylinder diameter and two values of slot width.
(*By courtesy of "Electronics".*)

slot, and with small cylinders, of diameter of the order of $0{\cdot}1\lambda$ to $0{\cdot}15\lambda$, the length of the slot must be increased appreciably in order to maintain resonance.

Jordan and Miller have given data relating slot length to wavelength for nominal "half wavelength" slots cut in cylinders for a range of cylinder diameters. These data are given graphically, and reproduced in Fig. 6.16, for two values of slot width. Jordan and Miller state that with this type of aerial the extremes of the working band of frequencies (determined by a standing wave ratio of

two to one on the feeder) is between four and eight per cent of the working frequency.

### Folded Slot Aerial

This type of aerial is complementary to the folded dipole discussed earlier. It comprises a slot aerial with a conductor placed centrally in the slot; the aerial is fed between one side of the slot and the central conductor. The input resistance is reduced by a quarter, to a theoretical value of 120 ohms. By appropriate choice of the dimensions of the central conductor, the reactive component of the input impedance can be made substantially zero over a wide band.

### An Equivalent Circuit for the Folded Dipole and Folded Slot Aerials

A very elegant approach to the determination of the input impedance of these types of aerials has been developed by G. D. Monteath. The folded dipole and folded slot are regarded as three

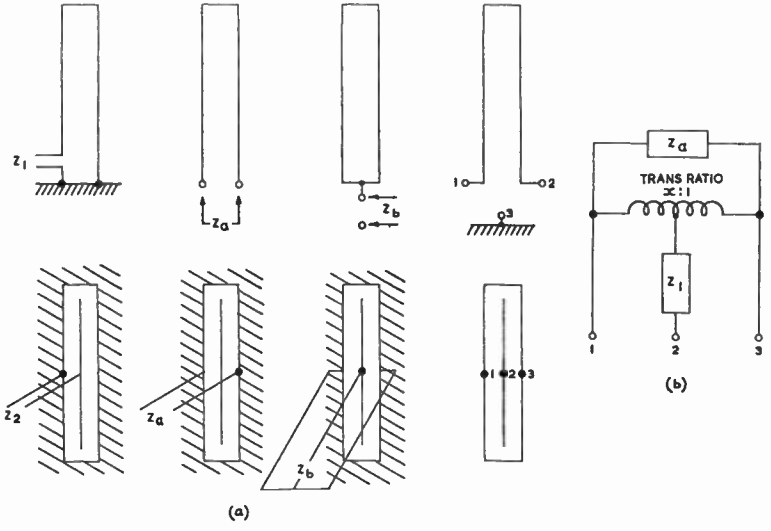



FIG. 6.17.—Equivalent circuits for folded unipole and folded slot aerials.

terminal networks, as shown in Fig. 6.17, and the three impedances $Z$, $Z_a$, and $Z_b$ are defined as shown in the figure. Employing these impedances, the equivalent circuit of Fig. 6.17 (*b*) can be constructed, where the auto-transformer is assumed to have infinitely

high inductance and perfect coupling. The ratio $k$ may be considered either as the ratio in which the potential difference between terminals 1 and 3 is divided by terminal 2 when this is free, or the ratio in which the current divides between terminals 1 and 3 when these are joined together. This equivalent diagram is only valid provided that the aerial satisfies one of the three conditions following:

1. The aerial is symmetrical about terminal 2 ($k=1$).
2. $Z_b$ is a pure reactance.
3. $Z_a$ is a pure reactance.

Thus for the unipole $Z_b$ is purely reactive, and where terminals 2 and 3 are connected together, as is usual,

$$\frac{1}{Z}=\frac{1}{Z_a}+\frac{1}{(k+1)^2 Z_b}.$$

$Z_b$ is the input impedance of the aerial treated as a simple unipole; $Z_a$ is the input impedance of the two aerial elements treated as a short circuit terminated transmission line, and is given by

$$Z_a=jZ_0 \tan \frac{2\pi}{\lambda} l,$$

where $Z_0$ is the characteristic impedance of the transmission line so formed, and $l$ is its length. $Z_0$ can be found from expression (6.10).

$k$ can be determined from electrostatic considerations provided that the elements are not too closely spaced. This has been considered by W. Van Roberts, who considers the capacitance of each element of the aerial; the result is given in the form

$$k=\frac{Z_1}{Z_2},$$

where $Z_1$ = characteristic impedance of the feeder formed by two elements equal in diameter to the diameter of the driven element, and spaced at a distance equal to that between the actual aerial elements;

and $Z_2$ = characteristic impedance of the feeder formed by two elements equal in diameter to the diameter of the undriven element, also spaced at a distance equal to that between the actual aerial elements.

Where the elements are of the same diameter, obviously $Z_1=Z_2$, and $k=1$. In this case the input resistance of the aerial at resonance is quadrupled. By a suitable choice of diameters, the input resistance at resonance can be varied over a wide range.

In the case of the folded slot, Monteath has shown that the input resistance at resonance can be varied appreciably by the position of the conductor in the slot. If the conductor is assumed to be very thin, and displaced by a distance $x$ from the centre of the slot in the direction of the undriven terminal (3), $k$ is given by

$$k=\frac{1+\sin^{-1}\frac{2x}{D}}{1-\sin^{-1}\frac{2x}{D}},$$

where $D$ is the slot width.

The input impedance, $Z_2$, when the aerial is fed between one side (terminal 1) and the central conductor (terminal 2) is given by

$$Z_2=Z_b+\frac{k^2}{(k+1)^2}Z_a,$$

where $Z_a$ is the normal input impedance.

$Z_b$ is the input impedance of the two $\lambda/4$ open circuit terminated sections of feeder formed by the central conductor and the slot, and is given by

$$Z_b=\tfrac{1}{2}j\,Z_0\tan\frac{2\pi}{\lambda}\frac{l}{2},$$

where $l$ is the overall length of the central conductor. $Z_0$ is given when the central conductor is narrow by

$$Z_0=138\log_{10}\frac{8W}{D},$$

where $W$ is the width of the central conductor, and $D$ is the slot width.

Monteath states that, in general, the cancellation of the reactive component of the input impedance of a slot aerial over a band requires impractically high values for $Z_0$. In order to reduce the reactive component, one half of the central conductor may be omitted, thus doubling the reactance slope of $Z_b$. The reactance slope may be still further increased by terminating the central

conductor in a length of co-axial transmission line, $\lambda/4$ long with a short circuit termination, as shown in Fig. 6.18. This co-axial line provides a reactive for the transmission line formed by the central conductor and the slot, and it can be shown that if the characteristic impedance of the co-axial line is $Z_1$, $Z_b$ is given by

$$Z_b = jZ_0(1+Z_0/Z_1) \tan \frac{2\pi}{\lambda} \frac{l}{2}.$$

As an example of the use of this method of reactive compensation, Monteath quotes a case in which the reactive component of $Z_a$ changed by 80 ohms for a 5 per cent change in frequency. With a $\lambda/2$ central conductor in the slot, $Z_0$ would have to be 2,000 ohms for exact compensation, calling for a central conductor diameter of $10^{-13}$ inches in a 10-inch slot. By

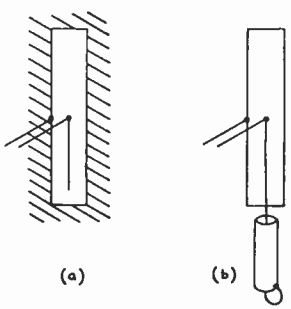



FIG. 6.18.—Two methods of improving reactance compensation in slot aerials. (*a*) half of central conductor omitted. (*b*) $\lambda/4$ section of co-axial line (short circuit termination) added.

employing the arrangement of Fig. 6.18 (*b*), reactance compensation was achieved by employing a central conductor of 0·5 inch diameter terminated by a co-axial line of characteristic impedance 62 ohms ($Z_{11}$). The characteristic impedance of the central conductor and slot section ($Z_{10}$) was 220 ohms.

## Transmission Lines

In order to convey power from a transmitter to the aerial or, conversely, from the aerial to the receiver, it is necessary to employ a transmission line. It has been already shown that an aerial has a definite characteristic impedance. If the full power

in the transmission line is to be transferred to the aerial, then its input impedance must be matched to that of the transmission line; in other words, the line must have the same characteristic impedance as the aerial's input impedance.

The characteristic impedance of a transmission line is defined as $Z_0 = \sqrt{L/C}$, where $L$ is the inductance and $C$ the capacity per unit length (expressed respectively in henries and farads). The impedance $Z_0$ is equal to that value of pure resistance which if placed across the end of the transmission line would absorb all the power being brought up by the waves travelling along it, without setting up any reflections. In other words, the whole of the power would be converted into heat. As the characteristic impedance is a pure resistance value, it follows that the line is aperiodic. The relatively small attenuation produced only varies with frequency as a result of the increasing losses occurring as the frequency is raised.

The characteristic impedance of a co-axial transmission line having an air dielectric with inner and outer radii of $d$ and $D$ respectively is

$$Z_0 = 138 \log_{10} \frac{D}{d} \text{ ohms.} \quad . \quad . \quad . \quad (6.18)$$

For parallel wire transmission lines consisting of wires having a radius $d$ and separated by a distance $D$:

$$Z_0 = 276 \log_{10} \frac{D}{d} \text{ ohms.} \quad . \quad . \quad . \quad (6.19)$$

It should be noted that owing to the slow way in which $\log_{10} \frac{D}{d}$ changes, even when the ratio $\frac{D}{d}$ is greatly changed, the characteristic impedance of lines cannot be varied over very wide limits. Practical values lie between some 30 and 600 ohms. The effect of introducing a dielectric other than air, and having a constant $K$, is to reduce the characteristic impedance by a factor $\frac{1}{\sqrt{K}}$ and to decrease the wavelength and velocity of propagation by a factor of $\frac{1}{\sqrt{K}}$.

For any given diameter of outer conductor, lines having a

characteristic impedance of 30 ohms will give a maximum power-carrying ability.

The high frequency resistance $R$ of a concentric line is

$$R=2\sqrt{\varrho\mu F}\left(\frac{1}{d}+\frac{1}{D}\right)10^{-9} \text{ ohms per cm., . . (6.20)}$$

where $D$=inner diameter of outer conductor in cm.;
$d$=outer diameter of inner conductor in cm.;
$\varrho$=specific resistance, in em-cgs units (some 1,700 for copper);
$\mu$=magnetic permeability;
$F$=frequency in cycles per second.

With the aid of this formula and equation (6.18) (and making the approximation $\alpha=\frac{R}{2Z_0}$), the formula for the attenuation in decibels per centimetre length for an air-spaced co-axial cable can be developed as follows:

$$\alpha=6{\cdot}3\times10^{-11}\frac{\sqrt{\varrho\mu F}\left(\frac{1}{d}+\frac{1}{D}\right)}{\log_{10}\frac{D}{d}} \text{ db per cm., . (6.21)}$$

where $\alpha$ is the attenuation constant.

The attenuation due to resistance losses is a minimum when $\frac{D}{d}=3{\cdot}6$ at which value $Z_0=76{\cdot}8$ ohms. It is of interest to note that in the 40 to 50 mc/s band the attenuation of ordinary twisted lamp-flex is some 3 to 10 db per 100 feet, and that it has a characteristic impedance of between 80 and 140 ohms. The way in which the attenuation of various types of transmission line varies with frequency is illustrated in Fig. 6.19.

### Transmission Line Termination Losses

If a line is not terminated in a pure resistance of value $R=Z_0$, then the relationship between the current and voltage will be different—both in magnitude and in phase—from the relationships in the line, with the result that all the power will not be

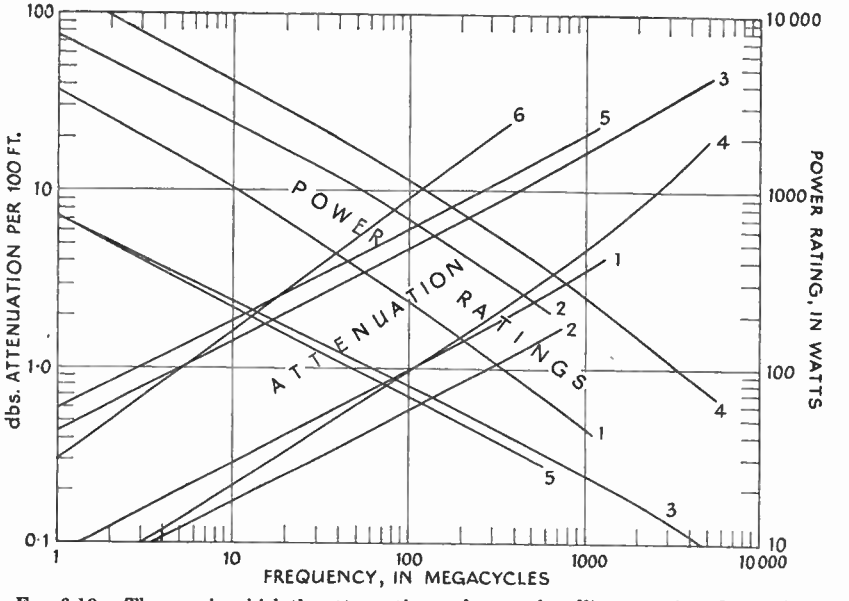



FIG. 6.19.—The way in which the attenuation and power-handling capacity of a number of different cables varies with frequency is shown above.

| Cable no. | $Z_0$ (ohms) | Type | Dielectric | Sheath | Conductor diameter | Sheath inner diameter |
|---|---|---|---|---|---|---|
| 1 | 75 | Coaxial | Air (Disc Spacers) | Lead | 0·103″ | 0·385″ |
| 2 | 100 | Coaxial | ,, ,, | Lead | 0·128″ | 0·75″ |
| 3 | 70 | Coaxial | Polythene | T/C Braid | 0·022″ | 0·133″ |
| 4 | 70 | Coaxial | Polythene | Lead | 0·144″ | 0·820″ |
| 5 | 100 | Screened Twin | Polythene | T/C Braid | 0·029″ | 0·170″ |
| 6 | 110 | Twisted Pair | Rubber | None | 0·042″ | 0·106″ |

absorbed by the terminating impedance. The net power transferred will be the difference between the power brought up by the incident wave and the power taken away by the reflected wave.

It is frequently desirable to have a knowledge of the loss which results from the incorrect termination or mismatching of a transmission line. It has already been pointed out that all the electrical oscillations in a line are built up from the propagated waves which

are travelling along it. Thus, any sinusoidal wave travelling along a line may be represented by

$$V=m \sin 2\pi\left(ft-\frac{x}{\lambda}\right)+n \sin\left\{2\pi\left(ft+\frac{x}{\lambda}\right)+\phi\right\}, \quad . \quad (6.22)$$

where $m$ and $n$=the amplitudes of the forward and reflected waves respectively;

$\phi$=the arbitrary phase difference between the two waves;

$x$=the distance from the source of origin, measured along the line;

$f$=frequency;

$t$=time.

The corresponding current $i$ is given by

$$Z_0 i=m \sin 2\pi\left(ft-\frac{x}{\lambda}\right)-n \sin\left\{2\pi\left(ft+\frac{x}{\lambda}\right)+\phi\right\}. \quad (6.23)$$

From these two equations it is possible to calculate the effect of terminating the line with an impedance $Z$, which may be either reactive or resistive. The point at which the line is terminated may conveniently be taken as the origin, at which point $x=0$. Under these conditions equations (6.22) and (6.23) may be rewritten as follows:

$$V = m \sin 2\pi ft + n \sin (2\pi ft+\phi), \quad . \quad . \quad (6.24)$$

$$Z_0 i = m \sin 2\pi ft - n \sin (2\pi ft+\phi). \quad . \quad . \quad (6.25)$$

At this point the current $i$ is flowing through the terminating impedance $Z$ and the voltage $V$ is developed across it. Thus, the current $V$ flowing in the impedance must also equal the current in the line at the point under consideration (i.e. $x=0$) both in magnitude and in phase. This condition can only be satisfied by a given ratio of $\frac{n}{m}$, and a given value for $\phi$, which will now be the phase change on reflection.

These equations may conveniently be solved graphically by the vector construction set out in Fig. 6.20. Referring to this diagram, if two vectors $OE$ and $OD$ are constructed from the point $O$, the vector $OE$ represents a voltage acting on the impedance $Z$ to

produce a current $i$ lagging by the angle $\theta$, and the vector $OD$ represents $Z_0 i$. Thus, if the terminating resistance $Z$ and the line impedance $Z_0$ are known it is always possible to construct the vectors $OE$ and $OD$. Having done so, join $DE$ and bisect it at $A$. Next construct a circle with $DE$ as its diameter and with its

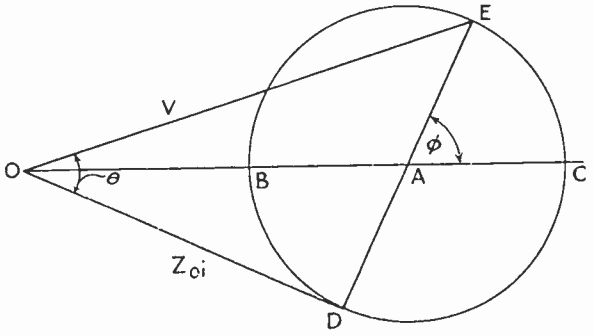



FIG. 6.20.—This diagram illustrates the vector construction for determining the termination conditions at the end of a transmission line.

centre at $A$. Then draw a line joining $OA$, and continue it on to cut the circle at $C$.

With the aid of the diagram thus constructed all the main properties of a terminated line can be deduced. The vector $OA=m$, and vector $AC=n$, while the angle $EAC=\phi$.

If, as an example, it is assumed that the terminating impedance $Z$ is a pure resistance having a value $R$, then $\theta=0$ and $\phi$ (the phase change produced upon reflection) is either 0 or $\pi$ radians. Under these conditions $Z_0 i=OB$ and $V=OC$, also

$$\frac{m-n}{m+n}=\frac{Z_0 i}{V}=\frac{Z_0}{R}.$$

If now $R=Z_0$, then $\dfrac{m-n}{m+n}=1$, from which it follows that $n=0$ and that there is no reflected wave, all the energy being absorbed by the resistance $R$. It should be noted that if $R>Z_0$ the phase change on reflection is zero, and if $R<Z_o$ the phase change is $\pi$ radians.

Let it now be assumed that instead of the terminating resistance $R$ equalling the line impedance $Z_0$, that it is half the value. Suppose that the characteristic impedance of the line is 80 ohms

and that of the terminating resistance is only 40 ohms. What proportion of the original line energy is absorbed by the resistance? Let the original energy $=\frac{m^2}{2Z_0}$ watts, and the reflected energy $=\frac{n^2}{2Z_0}$ watts. The fraction absorbed will therefore be $1-\frac{n^2}{m^2}$; however, $\frac{m-n}{m+n}=\frac{Z_0}{R}=\frac{80}{40}=2$, from which it follows that $\frac{n}{m}=-\frac{1}{3}$. (The negative sign indicates a phase change of $\pi$ radians on reflection.) The fraction of energy absorbed in the terminating impedance is therefore $1-\frac{1}{9}$ or 88 per cent.

The reflected energy may be dissipated as heat in the line if it is a long one; it may be reabsorbed in the generator, or it may be again reflected so as to form part of the original or incident wave. It should be noted that as the terminating resistance is less than the line impedance there will be a phase change of $\pi$ radians on reflection.

The case of a terminating impedance $Z$, which is a pure inductance, will next be considered. Under these conditions $\phi=90°$ thus 0 lies on the circle drawn on $ED$ as a diameter, and $OB=BC$ or $m=n$. The reflected wave has under these conditions the same amplitude as the original wave. As a pure inductance has no means of dissipating energy, it is, of course, to be expected that all the energy brought up to the terminating impedance $Z$ will under these conditions be reflected back again. The phase change $\phi$ will be dependent upon the ratio of $OE$ to $OD$ and will lie between 0 and $\pi$. If $Z$ should be a pure capacity, $\phi$ will lie between 0 and $-\pi$.

### Input Impedance of loaded Transmission Line

The input impedance of a loss free transmission line of characteristic impedance $Z_0$ when terminated at its far end by an impedance $Z_L$ is given by

$$Z_{IN}=Z_0\,\frac{Z_L+jZ_0\tan\frac{2\pi}{\lambda}l}{Z_0+jZ_L\tan\frac{2\pi}{\lambda}l},$$

where $l$ is the length of the line.

A number of special cases are of practical interest. Firstly, if the line is short circuited at its far end,

$$Z_{IN\ SC}=jZ_0 \tan \frac{2\pi}{\lambda}l,$$

i.e. the line behaves as a pure reactance, the magnitude of the reactance depending upon the length of the line. Correspondingly, if the line is open circuit at its far end,

$$Z_{IN\ OC}=-jZ_0 \cot \frac{2\pi}{\lambda}l,$$

i.e. the input impedance is again purely reactive, but of opposite sign to the input reactance to the short circuited line. Sections of transmission line terminated in one of these two ways are frequently employed to provide reactance compensation in matching circuits. Additionally, at frequencies above 100 Mc/s, sections of transmission line are frequently used as reactive components in preference to the equivalent "lumped" component; high $Q$ values can be achieved by this means.

The third special case occurs when the line is $\lambda/4$ long; there the input impedance is given by

$$Z_{IN\ \lambda/4}=Z_0^2/Z_L.$$

Owing to this impedance transforming property, $\lambda/4$ sections are widely employed in aerial matching networks.

Where this type of transformation is required and wide band matching is needed, it is common to employ two $\lambda/4$ sections in series; if the characteristic impedance of the section nearest the load is $Z_{01}$, and that of the section nearest the input $Z_{02}$, the input impedance is given by

$$Z_{IN}=\frac{Z_{02}^2}{Z_{01}^2}\cdot Z_L.$$

This expression determines the ratio of $\frac{Z_{02}}{Z_{01}}$ given $Z_{IN}$ and $Z_L$.

If now $\frac{Z_{IN}}{Z_{02}}$ is made equal to $\frac{Z_{01}}{Z_L}$, combination of the two $\lambda/4$ sections acts as a wide band matching network.

If a single section is used, a reactive component due to the line

appears at the input at those frequencies at which the line is not $\lambda/4$ long; when a double $\lambda/4$ transformer is employed, the characteristic impedances chosen being in accordance with the expressions above, the reactive components due to the two sections tend to cancel.

### Balance to Unbalance Networks (Baluns)

It is frequently required to feed a dipole aerial from a concentric feeder, the outer of which is earthed. Some form of balance to unbalance matching network is therefore required at the aerial.

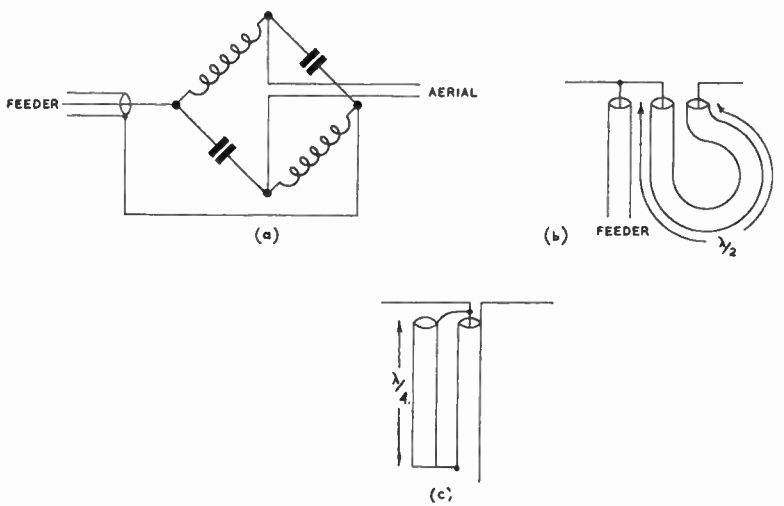



FIG. 6.21.—Three types of Balun: (*a*) Bridge circuit, (*b*) $\lambda/2$ section (*c*) Pawsey Stub.

This may take one of many forms, some of which are shown in Fig. 6.21. In example (*a*), if the aerial input impedance is resistive and equal to $r$, the input impedance presented to the line is

$$Z = \frac{L}{2Cr},$$

and hence by choice of suitable values of $L$ and $C$, the network can also be used for impedance matching.

In example (*b*), the $\lambda/2$ section of transmission line produces a change of phase of 180°; the signals from the two halves of the dipole aerial are thus fed in phase to the feeder. In addition, the arrangement transforms the aerial impedance down in the ratio

4:1; this arrangement is thus eminently suitable for matching a folded dipole to a 70-ohm cable. Example (*c*) is the well-known "Pawsey Stub". Here the two outer sections of the concentric feeders form a short-circuit terminated $\lambda/4$ stub. The stub thus does not appreciably load the circuit, and enables the element of the dipole attached to the outer conductor of the main feeder to be driven without appreciable radiation occurring from the feeder itself; additionally, the stub provides reactance compensation, the reactance slope of the stub being opposite to that of a simple dipole. This arrangement is thus especially suitable for wide band aerials.

Where a folded slot aerial is to be fed, difficulties may be encountered since the load is not truly balanced, and neither is one side at earth potential. For feeding the slotted cylinder type of aerial, the arrangement shown in Fig. 6.22 may be employed. Here a co-axial feeder is used, and the outer is connected to the neutral point opposite the slot. The feeder then runs in close proximity to the outer wall inside the cylinder, and the inner conductor is connected to the conductor lying in the slot. Over part of this section of feeder, the outer conductor is omitted; in this section the inner conductor forms one conductor of a transmission line with its image in the surface of the cylinder. By this means the generator is "floated".

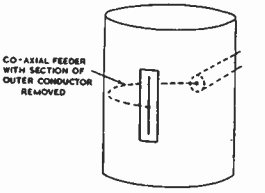



FIG. 6.22—Method of coupling co-axial feeder to folded slot aerial.

## Multi-element Transmitting Aerials

At the transmitter it is possible to obtain increased effective aerial power, and with it increased coverage by the use of multi-element aerials. Such aerials consist essentially of from two to ten or more separate aerial elements arranged in some fixed configuration and all fed with power effectively in parallel.

It should be noticed that the total power radiated is not increased by the use of a multi-element aerial, as it is obvious that this power can only be as great as the transmitter's power—less transmission line losses. Rather the gain achieved is a gain in the effective or useful power. It is obtained by reducing the power radiated in the upward direction and increasing the power radiated in the horizontal direction—i.e. along the earth's surface.

The manner in which this extra coverage is obtained is illustrated in Fig. 6.23. Diagram (A) shows the radiation pattern in the vertical plane, from a single horizontal half-wave aerial. As the arrows indicate, power is radiated equally at all angles to the horizontal. Of this power only that radiated horizontally or at very small angles to the horizontal serves any useful purpose; all the rest travels out into space and is lost. Diagram (B) illustrates

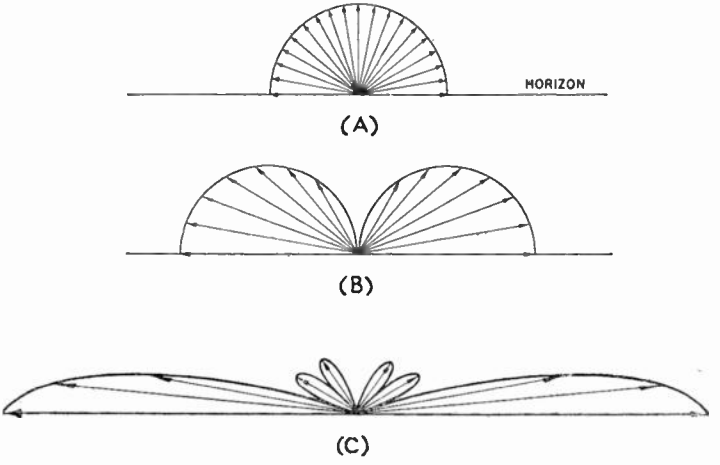



FIG. 6.23.—The way in which extra coverage is obtained with multi-element transmitting aerials is illustrated above.

the vertical radiation pattern obtained from an aerial consisting of two elements stacked vertically. In this case there is no radiation directly upward, and that at high angles has been greatly reduced, while the radiation at low angles and along the horizon has been greatly increased. Thus, the effective or useful power is much greater even though the total power radiated is the same.

Diagram (C) shows the vertical radiation pattern for a six-element aerial. The pattern has been still further squashed down and the radiation along the horizontal still further increased. As more elements are added beyond six, the horizontal radiation continues to increase. However, the amount of increase per added layer decreases, so that the diminishing return hardly justifies going beyond ten layers, and in many instances six layers is considered the best practical choice.

## Field and Power Gain

In comparing the advantages of multi-element aerials, the terms "field gain" and "power gain" are used. The field gain is defined as the ratio of the field intensities which results at a point one mile from the aerial when a vertical half-wave dipole is replaced by a multi-element aerial. Thus:

$$\text{Field gain} = \frac{\text{Field intensity with multi-element aerial}}{\text{Field intensity with vertical half-wave aerial}}.$$

It is of interest to note that a half-wave vertical aerial fed with a power of 1 kW. will produce a signal of approximately 137 millivolts per metre at one mile. Hence, the field gain of any particular aerial can be expressed as the ratio of the signal it produces per kilowatt at a distance of one mile, to the 137 millivolts per meter level.

The power gain is defined as the ratio of the powers that would be required to give the same field intensity at a point one mile distant. Thus:

$$\text{Power gain} = \frac{\text{Power required with a vertical half-wave aerial}}{\text{Power required with a multi-element aerial}}.$$

From this it follows that:

$$\text{Power gain} = (\text{Field Gain})^2.$$

## Practical Frequency Modulation Transmission Aerials

Both vertically and horizontally polarised propagation has been used for the transmission of frequency modulation broadcasts. However, for broadcasting purposes, horizontal polarisation appears to be standard. This means that the aerial elements must lie in a horizontal position. The practical transmission aerials of this type so far devised fall into five general categories:

(*a*) the original Brown turnstile and the improved "co-axial" versions;
(*b*) modifications such as the De Mars, the three-quarter wave-spaced and the folded turnstiles;
(*c*) variations of the circular or ring aerial;
(*d*) variations of the Alford square loop aerial;
(*e*) pylon aerials.

The first aerial designed specifically to provide directivity in the vertical plane (as contrasted with the communications type of aerial which is designed for directivity in the horizontal plane) was the original "turnstile" aerial. This aerial was developed by

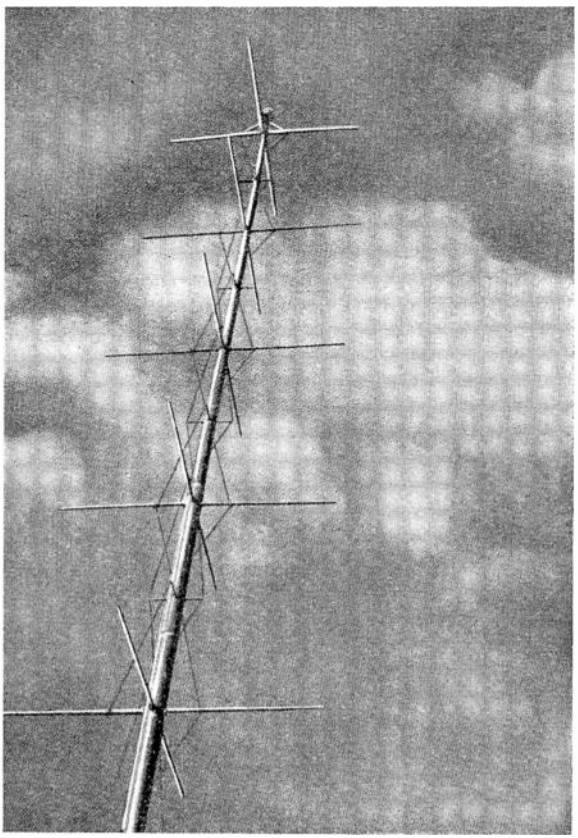


FIG. 6.24.—The original turnstile aerial as developed by G. H. Brown.

G. H. Brown and first described in 1936. As will be apparent from Fig. 6.24, it has derived its name from its striking similarity to the moving element of a turnstile gate. In its original form, the quarter-wave radiator rods were attached directly to the supporting pole. Four such rods arranged at 90° spacing around the mast made up each "layer" or "bay", the complete aerial being

composed of from two to ten such layers, depending upon the gain required and the supporting structure.

Each pair of oppositely placed rods forms a half-wave dipole aerial. Since the centre of the dipole is at zero voltage with respect to ground, the rods can be fastened directly to the grounded pole at this point. Power is fed to the dipoles by the open transmission line visible in Fig. 6.24, the various connections being spaced the proper distance from the pole to provide correct impedance matching. The horizontal field of a single bay is shown in Fig. 6.25, the first diagram of which shows the radiation patterns when the currents in the two dipoles are equal and their phase is varied. The second diagram shows the polar diagrams with a fixed 90°-phase difference and varying currents in the two arms. It will be noted that the combined field, as shown by the solid line, is very nearly a circle. In practice the field strength diagram will not be truly circular, as this would only occur if the field distribution function $F(\theta)$ for the individual dipoles was sin $\theta$. This would, of course, apply in the case of doublets of no finite length, under which conditions the combined field distribution function

$$F(\theta)=\sqrt{\sin^2\theta+\cos^2\theta}=1,$$

and the field strength would be uniform in all directions. The amount by which the actual field strength distribution will depart from this ideal can be judged from a comparison of the two profiles in Fig. 6.25.

In order to achieve the uniform distribution of field strength noted in Fig. 6.25, it follows that all the dipoles in one plane must be fed with equal amounts of power 90° out of phase. If the layers are spaced a half-wavelength apart, this can be done conveniently by means of the transmission line already mentioned. This line is crossed over between each layer, thereby counterbalancing the phase shift that occurs along the line between layers. Two such lines, one for each set of dipoles, run up the mast, twisting around it as they go and, in the case of the original Brown turnstile, being set off from it on stand-off insulators. At the base of the tower the two lines are fed with oppositely phased currents.

The early field experience with the original turnstile aerials brought to light one or two minor drawbacks. One of these was that the feeder line matching was extremely critical and required

adjustment in the field. The second, that the open wire transmission lines invited the formation of ice which tended to increase the wind resistance of the aerial and generally to detune the

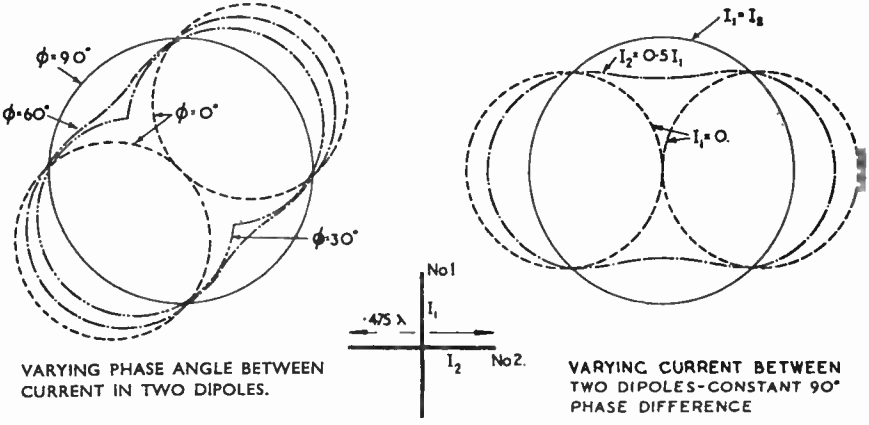



FIG. 6.25.—The horizontal radiation patterns for a turnstile aerial. Firstly, the patterns obtained when the currents in the two arms are equal and their phase is varied, and secondly, the polar diagrams with a fixed 90° phase difference and varying currents in the two arms.
*(By courtesy of the British Institute of Radio Engineers.)*

radiating system. To overcome these difficulties a modification of the original turnstile was developed, in which co-axial transmission lines replaced the open wire lines.

A close-up photograph of a later design is shown in Fig. 6.26. The arrangement of the radiators and lines in this version of the turnstile has several advantages. Firstly, the aerial can be completely "pre-tuned" during fabrication. Secondly, as the phasing is accomplished at the radiators, there are no phasing adjustments to be made at the bottom of the tower. All line impedances are exactly matched, and there are no standing waves on the lines. Thirdly, as the aerial's frequency range is much wider than that required for wide-band frequency modulation, the whole system is not critical in any respect.

There are a number of other modifications of the turnstile, the best known being the De Mars aerial. The essential difference between this and the Brown turnstile lies in the use of a separate co-axial transmission line to each radiator. Thus, for a six-bay aerial there are 24 feed lines, which run all the way down the tower to a "phasing room" at the base. The advantages claimed

for this system are that it enables the phasing to be done at a sheltered and convenient point, and that a more accurate matching is obtained. However, on the disadvantage side there is the cost and work of installing the greater number of lines and the extra wind resistance and ice hazard which they form.

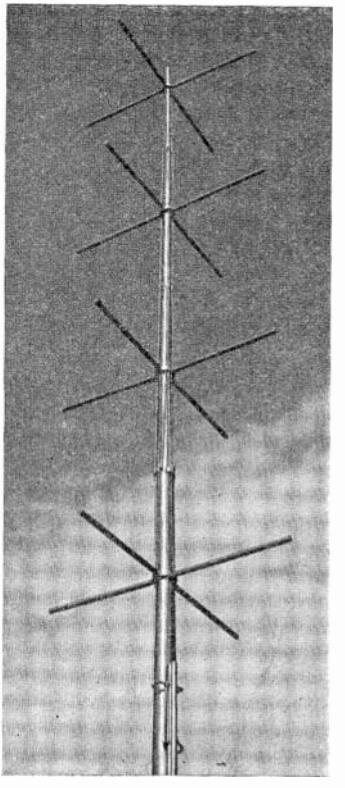


FIG. 6.26.—Close-up of an R.C.A. turnstile aerial.

Another variation of the turnstile which had a short vogue employed a between-layer spacing of three-quarters of a wavelength, instead of the normal half-wavelength spacing. A three-quarter-wave spacing gives a slightly greater gain than the half-wave spacing, and therefore an aerial of this type has a higher gain per layer.

It should, however, be noted that the gain per layer is not the

TABLE 10A

*Gain of turnstile aerial*

| Number of layers | Power gain db | Field-strength gain db | Distance between top and bottom layers in wavelengths |
|---|---|---|---|
| 1 | 0·50 | 0·707 | 0 |
| 2 | 1·25 | 1·12 | 0·5 |
| 3 | 2·00 | 1·41 | 1·0 |
| 4 | 2·75 | 1·66 | 1·5 |
| 5 | 3·50 | 1·87 | 2·0 |
| 6 | 4·24 | 2·06 | 2·5 |
| 7 | 5·05 | 2·26 | 3·0 |
| 8 | 5·75 | 2·40 | 3·5 |
| 9 | 6·25 | 2·5 | 4·0 |
| 10 | 6·76 | 2·6 | 4·5 |

only indication of worth. Actually, extra layers add little to the cost or weight of the whole aerial assembly; the overall height of the supporting pole is a more important factor, as it is the weight of this pole and the means of mounting it that determine what can and what cannot be used on any given structure. In this respect the three-quarter-wave spacing offers no advantage, as the gain per unit length of pole is the same.

### Circular or Ring Aerials

The circular or ring aerial is essentially a folded dipole aerial which has been bent around into a circle. In Fig. 6.31 (A) the folded dipole is shown in its simplest form. It consists of two half-wave radiators, one of which is broken at the centre, where it is fed from a balanced transmission line. The instantaneous currents in both units are in the same direction, and therefore the current distribution does not differ greatly from that of an ordinary half-wave dipole (which is approximately sinusoidal). As the voltage to ground at the centre is zero, the unbroken radiator can be attached directly to the supporting pole at this point.

In order to approach a more uniform field strength distribution than that of an ordinary dipole aerial, the folded dipole may be bent around into a circle, as shown in Fig. 6.31 (B). This, however, will not in itself give a circular pattern as the current distribution

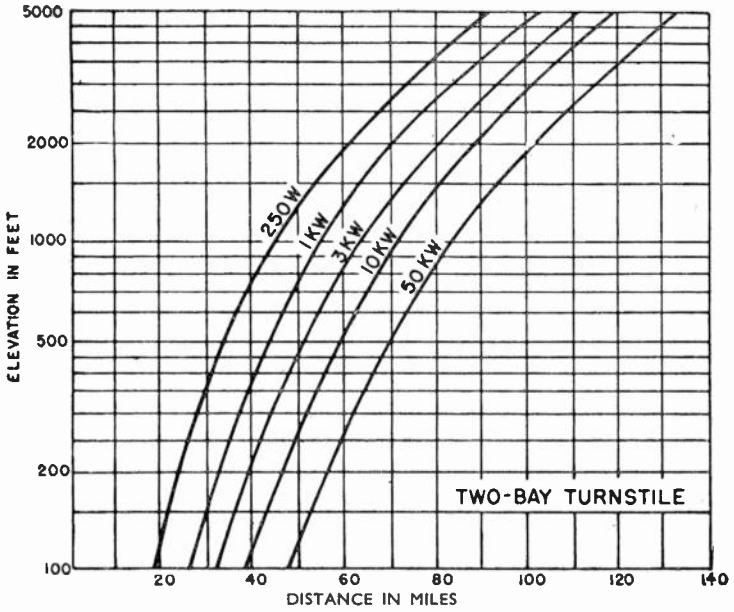



FIG. 6.27.—Curves showing the relationship of height, power and distance to the 50 microvolt per metre contour line for a two-bay R.C.A. turnstile.

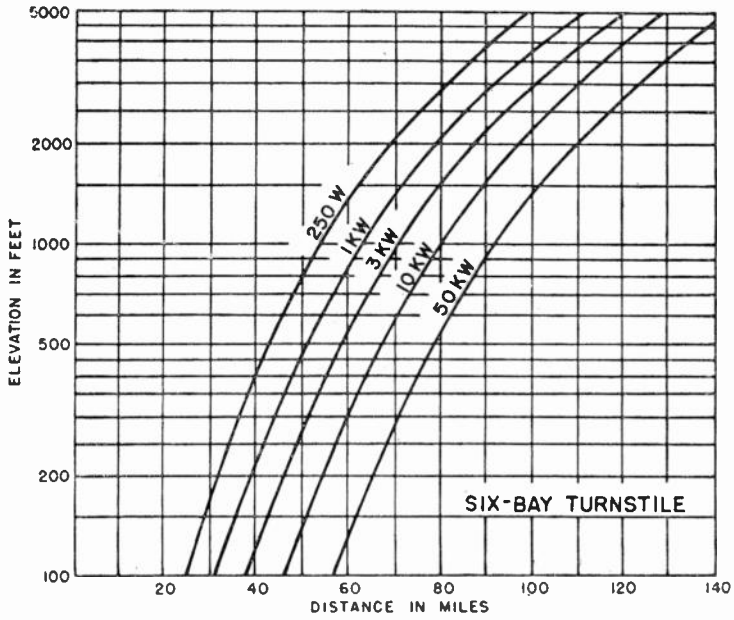



FIG. 6.28.—Curves showing the relationship of height, power and distance to the 50 microvolt per metre contour line for a six-bay R.C.A. turnstile.

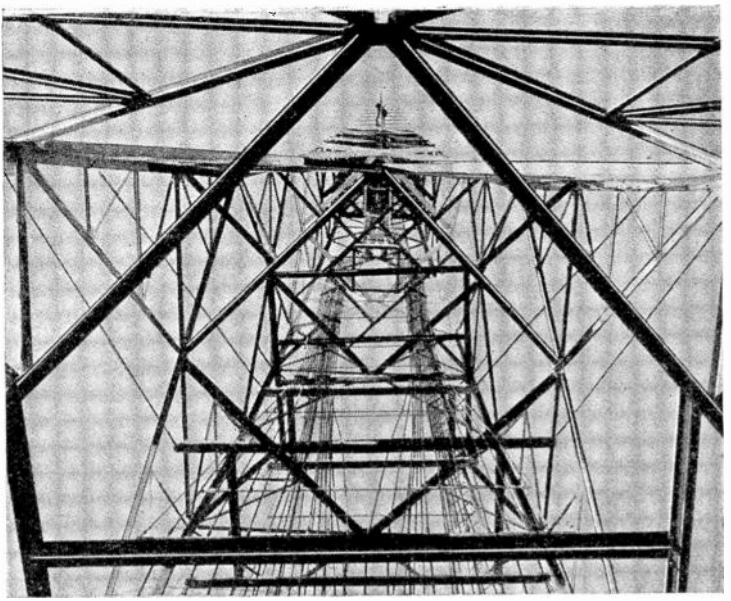


FIG. 6.29.—The co-axial cables coming down the aerial tower from the 10-bay De Mars turnstile at the 50 kW. Yankee Network F.M. station at Paxton.
*(By courtesy of "F.M. and Television".)*

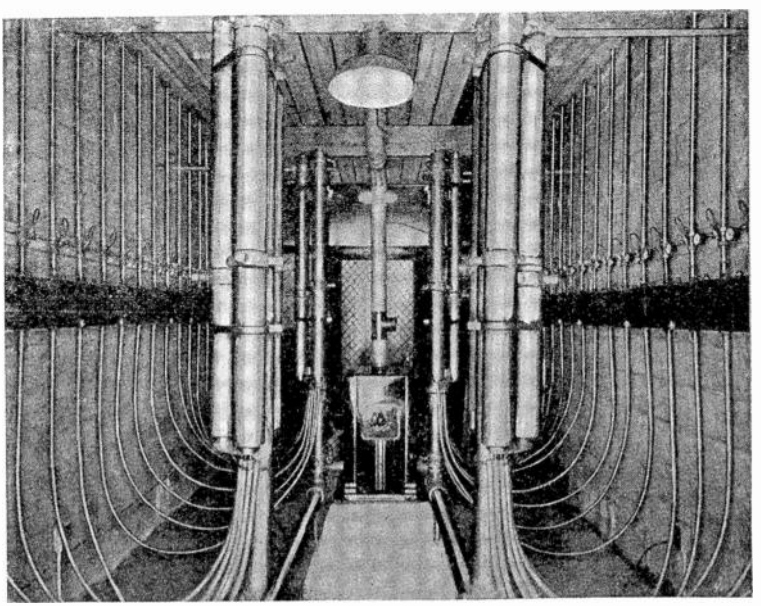


FIG. 6.30.—The matching and phasing room at the Paxton station.
*(By courtesy of "F.M. and Television".)*

is not uniform around the radiator. To improve this situation a pair of metal plates are fastened at the folded points, as shown in Fig. 6.31 (C). These plates have the effect of adding end capacity to the radiators and change the current distribution to something like that shown in Fig. 6.31 (D). The current being now approximately uniform around the loop, the signal radiated approaches a circular pattern to the same degree.

A circular aerial presents a neat appearance and has a higher gain per layer than a turnstile. However, in order to keep down the mutual impedance, the layers must be placed a full wavelength apart. Thus, the gain for a given height of mast is less than that obtained with a turnstile—provided that more than one layer is used. For example, a three-bay circular aerial is two wavelengths high and has a power gain of 2·6, whereas a five-bay turnstile having the same height has a power gain of 3·5.

TABLE 10B

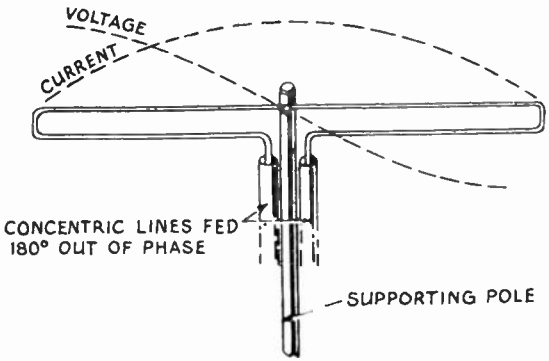
*Gain of Circular Aerial*

| Number of layers | Power gain db | Field strength gain db | Distance between top and bottom layers in wavelengths |
|---|---|---|---|
| 1 | 0·79 | 0·89 | 0 |
| 2 | 1·7 | 1·3 | 1·0 |
| 4 | 3·63 | 1·9 | 3·0 |
| 6 | 5·5 | 2·35 | 5·0 |
| 8 | 7·24 | 2·7 | 7·0 |

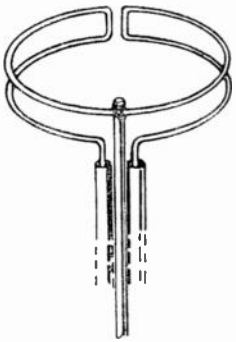
As noted earlier, it is the height which is the important parameter since it is the weight and upsetting moment of the supporting pole which determine the practicality of any given design. The off-centre mounting of the rings is also a disadvantage in that it makes for mechanical dissymmetry. Thus, while the loops are of the same approximate weight as the turnstile elements, the fact that they are off-centre requires a stronger supporting pole.

## Square-Loop Aerials

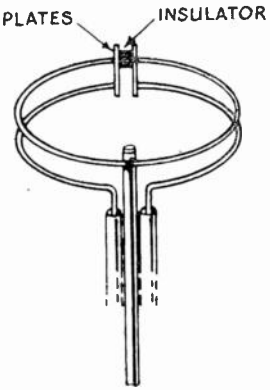
The broadcasting aerials discussed up to this point have all been mounted on supporting masts of the flag-pole variety. Where such a pole can be mounted on an existing structure, or



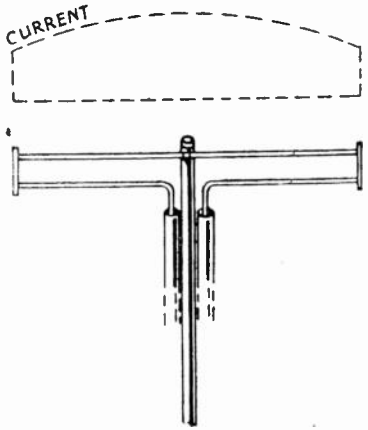

(A) FOLDED DIPOLE



(B) FOLDED DIPOLE BENT AROUND INTO A CIRCLE




(C) PLATES ADDED AT FOLDED ENDS




(D) EFFECT OF ADDING PLATES

FIG. 6.31.—The evolution of the circular aerial.

where the ground height is in itself sufficient, one of these standard types would normally be used. In some cases it is, however, not possible to mount a flag-pole on the building chosen—either because the building structure will not support it, or because of

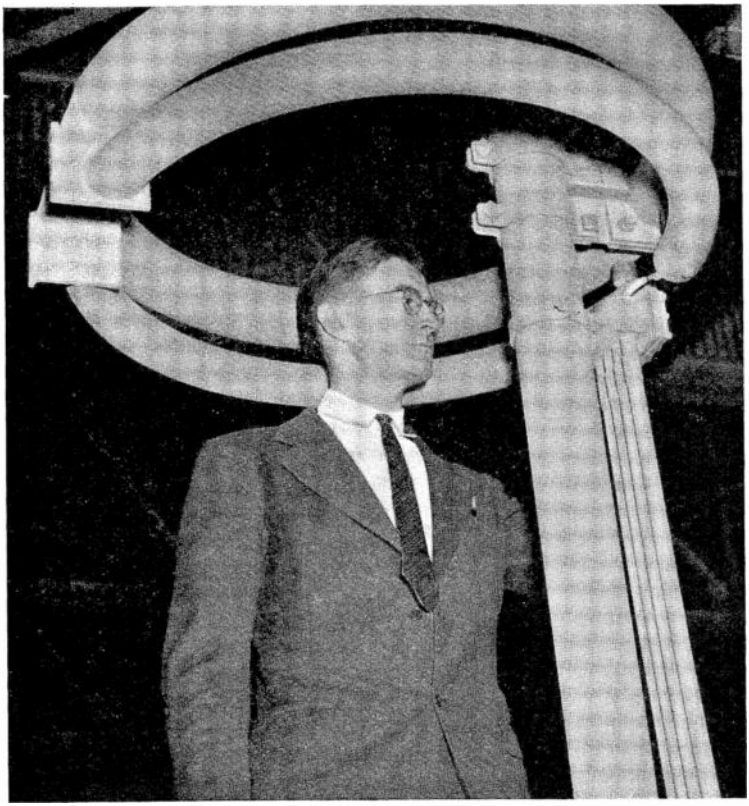


FIG. 6.32.—A close-up view of one bay of the General Electric Co., U.S.A., circular aerial, with its designer, M. W. Scheldorf.

the configuration of the building itself. Similar difficulties sometimes arise when it is desired to mount an f.m. aerial on an existing a.m. tower. The majority of such towers were not built for, and will not support, the heavy pole used with multi-element turnstiles or ring aerials. In such cases, several variations of what for want of a better name may be called a square-loop aerial have been used with success.

The square-loop aerial consists of four dipole radiators arranged in the form of a square which may or may not be closed at the corners. In the case of a large building tower, the dipoles may project from the four sides. They may be in the form of folded dipoles or of simple dipoles fed at the centre, according to how impedance matching is to be obtained.

A type of square-loop aerial which can be conveniently mounted around a standard amplitude modulation broadcast tower has

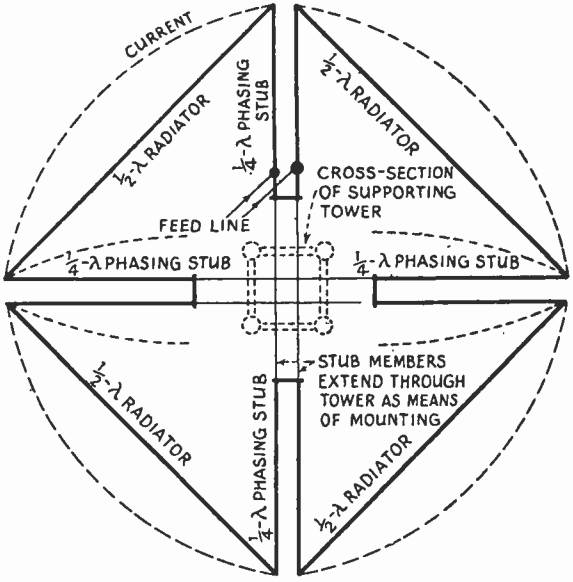



FIG. 6.33.—Each layer of the square-loop aerial consists of four half-wave radiators arranged in a square. Correct phasing is provided by the use of quarter-wave stubs which also form the radiator supports.

(*By courtesy of "Electronics".*)

been described both by A. Alford and G. H. Brown. While various configurations are possible—including a three-sided type—the most usual arrangement is that developed by R.C.A. and shown in Fig. 6.33. The radiators are half-wavelength sections supported at their ends by lengths of tubing which run diametrically across the square and are attached near the centre to the framework of the broadcasting tower. These supports have shorting bars placed at points a quarter wavelength in from the corners. The current distribution is shown by the dotted line. Since the points at which

the shorting bars are located represent voltage nodes, the supports can be at ground potential.

The gain per layer of the square-loop aerial is greater than that of either the turnstile or the ring aerial. The reason will be evident when it is noted that each layer has effectively twice as many radiators as the turnstile. Moreover, as the vertical radiation is very low, the layers can be mounted at half-wave intervals. Comparative gains of the various types of aerial discussed are shown in Fig. 6.35.

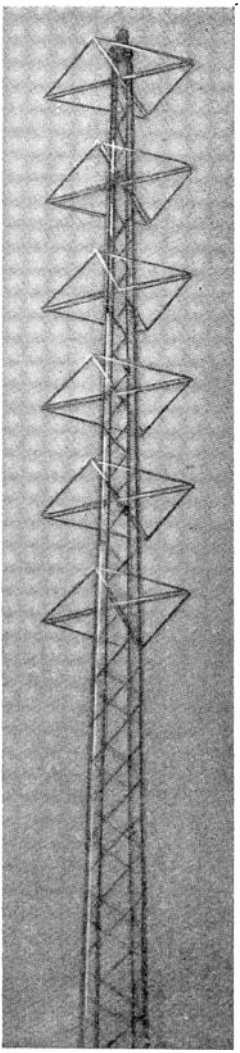


FIG. 6.34.—The six-layer square-loop aerial at the f.m. station at Baton Rouge, Louisiana.

### Slotted Cylinder Aerials (Pylon Aerials)

This type of aerial employs one or more slots cut in a cylinder of diameter of the order of $\lambda/10$. As explained earlier, a vertical slot aerial has a horizontally polarised radiation field, and is thus particularly useful for f.m. applications, since it has a relatively low wind resistance, and the feeders may be accommodated inside the cylinder itself. The slots may be filled with a low loss material, and hence this type of aerial is relatively little troubled by weather conditions, since the feeders and feed points are totally enclosed.

Due to the reactive loading effects on the cylinder itself, the slots are generally considerably larger than $\lambda/2$ for resonance, the actual increase in length being determined by the slot width and cylinder radius, as shown in Fig. 6.16. The feed point impedance of a single slot at half wave resonance is in the region 300–1,000 ohms; with a multi-element aerial of this type the slots may be fed in parallel, to reduce the magnitude of the load presented to the main feeder to reasonable proportions.

Jordan and Miller have described the characteristics of a

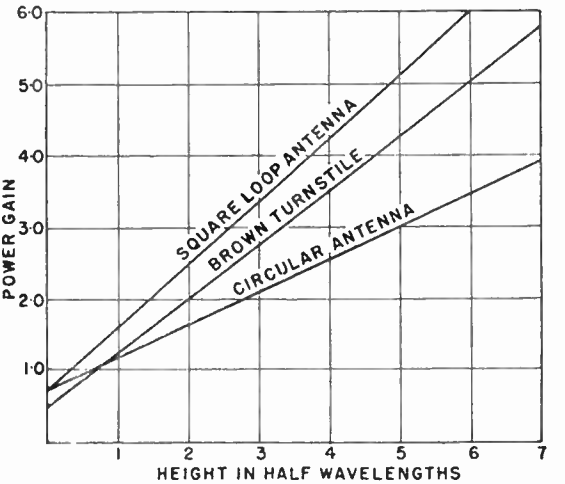



FIG. 6.35.—Comparative gains of the various types of horizontally polarised multi-layer transmission aerial.

four-slot aerial of this type, which may be considered typical. The slots are spaced at one wavelength intervals, in a cylinder of diameter $\lambda/8$. Each slot, for resonance, is $3\lambda/4$ long, and the slot width is $\lambda/60$. The vertical and horizontal radiation patterns of this aerial are shown in Fig. 6.37, from which it will be seen that the radiation is substantially omni-directional in the horizontal plane. The aerial gain over a half-wave dipole is 5 db.

Whilst the radiation pattern in the horizontal plane becomes more nearly omni-directional as the cylinder diameter is decreased, a lower practical limit of cylinder diameter is set by the need to increase slot length to maintain resonance; with cylinder diameters in the region of $\lambda/10$, the increase in length becomes inconveniently large, and a figure of $\lambda/8$ would appear to represent a good compromise value.

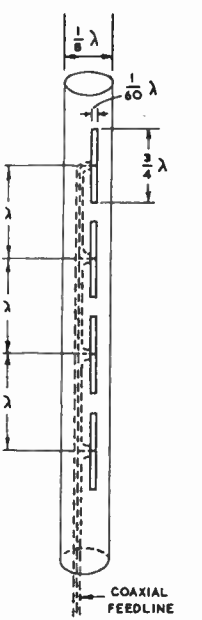



FIG. 6.36.—Construction of slotted cylinder aerial.
(*By courtesy of "Electronics".*)

## Vertically Polarised Transmission Aerials

All the aerials so far discussed have radiated

horizontally polarised signals. However, for some special purposes it may be desirable to radiate vertically polarised signals. Fig. 6.38 shows a group of three aerials which can be employed for this purpose.

The first has already been discussed earlier in this chapter, and it is not therefore proposed to say anything further at this point.

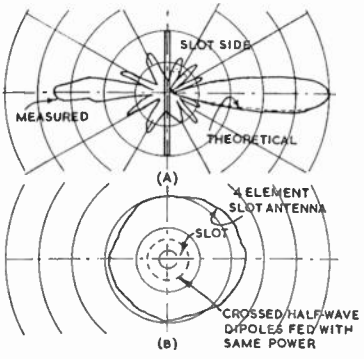



FIG. 6.37.—(*A*) Vertical and (*B*) horizontal radiation patterns of the aerial of Fig. 6.36.
(*By courtesy of "Electronics".*)

The second aerial, like the first, is suitable for flag-pole mounting. The outer conductor of the concentric transmission line is extended and folded back on itself in the form of a cylinder of relatively large diameter and a quarter of a wavelength long. The outer cylinder, together with the extension of the central conductor, then function as a half-wavelength radiator which, when at resonance, has the same input impedance as a centre-fed, half-wave dipole.

The third aerial is due to G. H. Brown and J. Epstein. This aerial is excited from a concentric transmission line which is connected directly to the end of a vertical quarter-wave radiator element. The lower end of this quarter-wave radiator element is continued into a screened section which permits a rigid mounting to be obtained without any detrimental effects on the electrical characteristics of the radiator itself—the impedance at the connection point is practically infinite. As the radiation resistance of a quater-wave aerial is some 37 ohms, a co-axial cable of this impedance must be used to feed the aerial. Normally, this low impedance will make it necessary to use a polythene or other solid

dielectric cable. As an alternative, however, a quarter-wave section of line may be used as a matching section between the transmission line and the aerial's 37-ohm input impedance. The more normal practice is, however, to use a cable of some 70 ohms impedance and eliminate the matching section by shortening the aerial and the matching section, so making the aerial impedance

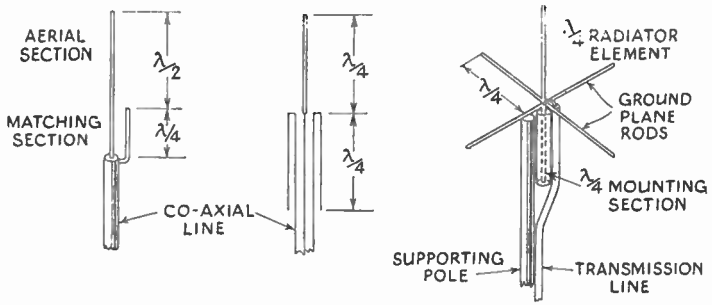



FIG. 6.38.—Three aerials which can be used for the transmission of vertically polarised signals.

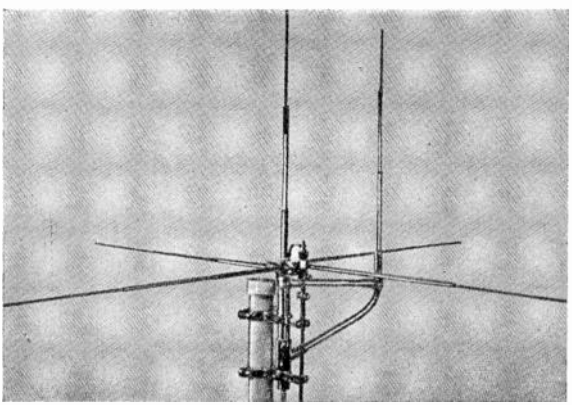


FIG. 6.39.—A vertically polarised four rod "ground plane" aerial with reflector.

*(By courtesy of R.C.A.)*

equal to that of the transmission line. The four horizontal rods are used to provide the "ground plane" necessary to ensure the provision of a strong image of the aerial—so producing the overall effect of a vertical half-wave dipole.

Fig. 6.39 shows an R.C.A. aerial of this type. This aerial consists of a vertical half-wave dipole and a four-rod "ground

plane", which gives a field gain of some 2·8 db. Such a light aerial is not suitable for handling an output of more than some 3 kW. Where a certain amount of directivity is required a reflector can be added, as is shown in the illustration. This particular reflector gives a field gain of 1·3 db in the forward direction and a reduction of some 1·7 db on the back side.

### Tilted Wire Aerials

The tilted wire family of aerials, in general, have high gain and directivity combined with very little selectivity. However, to be of much value they must be several wavelengths long. Consequently they have not been very widely used up to the present date. The correct angle of tilt or slope is such that the length of wire is one-half wavelength greater than its projection along the direction of propagation. Thus, the longer the wire the smaller the angle it should make with the direction of wave propagation in space.

Two sloping wires may be combined to form a balanced V-aerial lying in the plane of polarisation, and two V-aerials may be joined to form a diamond or rhombic pattern. With the far ends of these tilted-wire aerials open, they receive signals from the back nearly as well as from the front. This is because the energy from one direction travels out to the open ends and is reflected back to the receiver. Resistive termination may be applied to the open ends if reception from the back direction is undesired.

In Britain E.M.I. have designed aerials which increase the phase velocity along the tilted wire by the insertion of capacities at regular intervals along it. This permits a greater angle of tilt and consequently exposes the wire to a longer wave-front.

### Slotted Cylinder Aerial with Multiple Slots

At very high frequencies, the requirement of the Pylon aerial that the cylinder diameter shall be of the order of $\lambda/10$, may lead to an inconveniently small value from the point of view of mechanical stability. To overcome this, the cylinder diameter may be increased, and more than one slot used. An example of this type of aerial is that employed by the BBC for its v.h.f. transmissions.

The aerial comprises eight stacks, each stack having four slots cut in a cylinder slightly greater than $\lambda/2$ in diameter. The overall

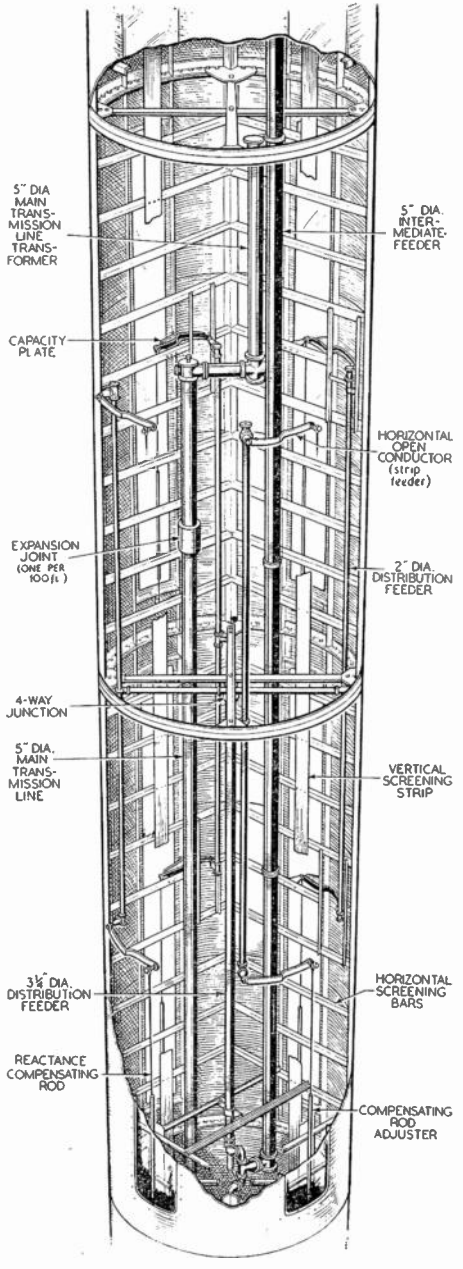



FIG. 6.40.—Cut-away view of BBC Multi-slot cylindrical aerial.
*(By courtesy of "Wireless World".)*

height of the aerial is eight wavelengths, and its gain over a half-wave dipole is 9 db. The slots are operated at half-wave resonance; as each slot is boxed, the actual length for resonance is approximately $3\lambda/4$. The aerial is designed for wide-band working, and

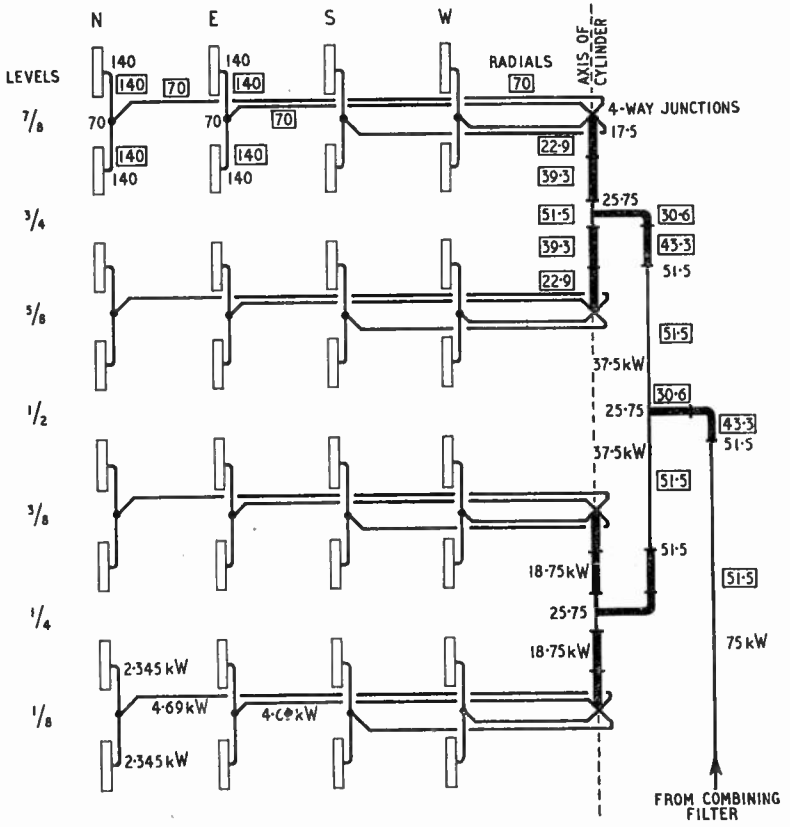



FIG. 6.41.—Distribution feeder system for the aerial of Fig. 6.40, shown opened out into flat plane for clarity. Radial feeders are all of equal length with point impedances shown as plain figures and characteristic impedances in "boxes".

*(By courtesy of "Wireless World".)*

each slot is "folded" and has reactance compensation of the type illustrated in Fig. 6.22.

The construction of each stack is shown in Fig. 6.40; as will be seen, the "boxing" of each slot is in skeleton form, comprising conductors spaced at approximately $\lambda/10$ intervals. Additionally,

a vertical conducting strip is situated in contact with the box bars behind each slot, to minimise fields in the enclosure at the centre of the cylinder. The outer of the coaxial feeder to each slot is terminated at the box bar opposite the centre of the slot, and the inner conductor is continued parallel to the cylinder surface to the central conductor of the slot, in the type of Balun circuit described previously. At the edge of each slot, capacitance coupling to the feeder is provided, to give a pre-set degree of reactance compensation.

The 32 slots are fed by successive bifurcation. At each point of division double $\lambda/4$ matching networks are employed, to provide the necessary impedance matching. A schematic diagram of the feeder arrangement is shown in Fig. 6.41.

## SELECTED REFERENCES


CARTER, P. S., Circuit Relations in Radiating Systems and Applications to Antenna Problems, *Proc. I.R.E.*, June 1932.

WHEELER NAGY, A., An Experimental Study of Parasitic Wire Reflectors on 2–5 Metres, *Proc. I.R.E.*, February 1936.

BROWN, G. H., A Turnstile Antenna for use on Ultra High Frequencies, *Electronics*, April 1936.

BROWN, G. H., Directional Antennas, *Proc. I.R.E.*, January 1937.

LINDENBLAD, N. E., Television Transmitting Antenna for the Empire State Building, *R.C.A. Review*, April 1939.

CARTER, P. S., Simple Television Aerials, *R.C.A. Review*, October 1939.

BROWN, G. H., and EPSTEIN, J., An Ultra-High Frequency Antenna of Simple Construction, *Communications*, July 1940.

ALFORD, A., and KANDOIAN, A. C., Ultra-High-Frequency Loop Antennas, *Trans. A.I.E.E.*, 1940 (page 843).

GEORGE, R. W., Field Strength of Motor-car Ignition Between 40 and 450 Megacycles, *Proc. I.R.E.*, September 1940.

BROWN, G. H., and KING, RONALD, High Frequency Models in Antenna Investigations, *Proc. I.R.E.*, April 1943.

KING, RONALD, and HARRISON, CHARLES W., The Distribution of Current Along a Symmetrical Centre-Driven Antenna, *Proc. I.R.E.*, October 1943.

KING, RONALD, Coupled Antennas and Transmission Lines, *Proc. I.R.E.*, November 1943.

NEIMAN, M. S., The Principle of Reciprocity in Antenna Theory, *Proc. I.R.E.*, December 1943.

CARTER, P. S., Antenna Arrays Around Cylinders, *Proc. I.R.E.*, December 1943.

KING, RONALD, and HARRISON, CHARLES W., The Receiving Antenna, *Proc. I.R.E.*, January 1944.

WELLS, N., The Quadrant Aerial—An Omni-Directional Wide-Band Horizontal Aerial for Short Waves, *Journal I.E.E.*, Part III, December 1944.

TAYLOR, JOHN P., A Square-Loop F.M. Antenna, *Electronics*, March 1945.

SCHELDORF, M. W., Circular Antennas for F.M. Broadcasting, *F.M. and Television*, May 1945.

KIRKE, H. L., Frequency Modulation—BBC Field Trials, *BBC Quarterly*, July 1946.

HOLZ, R. F., Characteristics of the Pylon F.M. Antenna, *F.M. and Television*, September 1946.

Papers on Metre-Wave Aerials, *Journal I.E.E.*, Part IIIA, No. 3, 1946. (A group of important papers well worth reference.)

JORDAN, E. C., and MILLER, W. E., Slotted Cylinder Antenna, *Electronics*, February 1947.

HOLMES, R. S., and TURNER, A. H., Simple Antennas and Receiver Input Circuits for Ultra-High Frequencies. *Radio at Ultra High Frequencies* (book published by R.C.A. Institutes Technical Press).

BRAINERD, KOEHLER, REICH and WOODRUFF, *Ultra High Frequency Techniques*. Chapman and Hall.

PUTNAM, J. L., Input Impedances of Centre fed Slot Aerials near Half-wave resonance, *Journal I.E.E.*, Part III, July, 1948.

PUTNAM, J. L., RUSSELL, B., and WALKINSHAW, W., Field distributions near a centre fed half wave radiating slot, *Journal I.E.E.*, Part III, July 1948.

BOOKER, H. G., Slot aerials and their relation to complementary wire Aerials (Babinet's principle), *Journal I.E.E.*, Part IIIA, March–May, 1946.

BAILEY, C. E. G., Feeders and Slot Aerials, *Journal I.E.E.*, Part IIIA, March–May, 1946.

JORDAN, E. C., and MILLER, W. E., Slotted-Cylinder Antenna, *Electronics*, February 1947.

SMITH, R. A., *Aerials for Metre and Decimetre Wavelengths*, Cambridge University Press, 1949.

MONTEATH, G. D., Wide band fold slot aerials, *Journal I.E.E.*, Part III, November 1950.

GILLAM, C., Wrotham Aerial System, *Wireless World*, June and July, 1953.

*Chapter Seven*

# FREQUENCY MODULATION TRANSMITTERS

ONE of the most interesting differences between a frequency modulation transmitter and its amplitude modulated counterpart is its higher efficiency. When employing amplitude modulation the peak power output at 100 per cent modulation rises to four times the unmodulated carrier power. It therefore follows that if the peak power output is the limiting factor, by changing to frequency modulation the carrier power may be increased by four times or some 6 db. However, if the limitation is the maximum r.m.s. power output, then the permissible increase is only twice or some 3 db (assuming the most rigorous conditions—a square-wave modulating signal). From these figures it will be apparent that for any given size of power output valve or specified power consumption, a greater signal output can always be obtained by substituting frequency modulation for amplitude modulation. This higher efficiency is due to the reduction in the carrier component's energy content which occurs when it is modulated in frequency—as opposed to the constant power content of the carrier component of an amplitude modulated signal.

In this chapter the general principles of frequency modulated transmitters will be considered, while the detail changes which have to be made for the various different uses to which frequency modulation is put will be considered in Chapter Eleven. It will, however, be necessary to outline the general requirements of typical frequency modulation systems in order to obtain a general quantitative background against which to discuss the various circuits involved. As the largest use to which frequency modulation has yet been put is that of high-fidelity broadcasting, and, further, as definite requirements in this connection have been laid down by at least one country, it is proposed to start by reviewing these requirements. In America the Federal Communication Commission believe that frequency modulation is capable of providing higher fidelity transmissions than those of the normal medium-wave broadcasting stations; accordingly, they have

specified standards which are far more rigorous. The more important of these requirements have been summarised in the table following.

TABLE 11

*Frequency modulation broadcast transmitter performance standards*

| Characteristic | F.C.C. overall requirements | Transmitter measurements | Studio equipment audio measurements | Relay circuit requirements |
|---|---|---|---|---|
| Audio Frequency | ±2 db of 1 kc/s level 50 c/s to 15 kc/s | Better than ±1 db of 1 kc/s level 30 c/s to 16 kc/s | Better than ±1 db of 1 kc/s level 30 c/s to 15 kc/s | ±1 db of 1 kc/s level 30 c/s to 15 kc/s System must be compensated overall |
| Freq. Mod. noise-level | 60 db below 100 per cent mod. 50 c/s to 15 kc/s | Better than 70 db below 100 per cent mod. 30 c/s to 16 kc/s | Better than 65 db below level of 1 mv. input to pre-amplifier | Should be better than 65 db |
| Distortion | Less than 2 per cent (r.m.s.) 50 c/s to 15 kc/s | Less than 1·5 per cent 30 c/s to 15 kc/s | Less than 0·5 per cent 30 c/s to 15 kc/s | Less than 1·5 per cent 30 c/s to 15 kc/s |

It will be evident from column three that in order to comply with these requirements a very high standard of performance must be maintained at the transmitter. At first sight it might appear that such high standards are unnecessary; however, it is on record that observers have commented on distortion as low as 2 per cent, and have frequently objected to noise-levels of 60 db below 100 per cent modulation. These standards may, therefore, as their name implies, be accepted as Standards of Good Engineering Practice for high-fidelity broadcast stations.

One of the earliest decisions to be made when starting the design of a frequency modulated transmitter is the frequency deviation or swing which is to correspond to 100 per cent modulation. In the case of a high-fidelity broadcast station, a swing of ±75 kc/s is normally employed with a maximum audio frequency of 15 kc/s—i.e. a deviation ratio of 5:1. In the case of radio-telephone links it is usual practice to employ a deviation of ±15 kc/s and a maximum audio frequency of 3 kc/s—again a

deviation ratio of 5:1. Although deviation ratios as low as 1:1 have been employed for telephony transmissions having a maximum audio frequency of 3 kc/s, the considerations at the end of Chapter Two show that the use of such low deviation ratios results in little if any improvement in noise-level over amplitude modulation. In such cases the reduction in noise-level does not therefore provide a valid technical reason for the use of frequency modulation. In the case of frequency modulated telegraph transmitters a frequency shift of 850 cycles is frequently adopted, the band-width occupied by the signal's side bands being some 1,100 cycles. Even with high-speed Morse this gives a deviation ratio of between 3:1 and 4:1, while for manual signals the ratio is often as high as 10:1 or 20:1.

Regardless of the purpose for which the frequency modulated signal is being employed, it is of prime importance that the carrier frequency should be as stable as is possible. In the case of frequency modulated broadcast stations the Federal Communication Commission specify that the carrier's mean must be held on its allotted frequency to with $\pm 2$ kc/s (i.e. to within one part in 40,000, at the carrier frequencies normally employed). The majority of frequency modulated transmitters are arranged so that the power output stage is supplied with its drive from a chain of frequency multiplying stages, which are in turn driven from the frequency modulator—the controlling and by far the most important stage in the transmitter. Although a wide variety of circuits have been employed for the frequency modulation of the carrier, the common aim of all these circuits is always the achievement of the highest possible mean frequency stability, coupled with minimum distortion. The frequency modulator circuits most widely employed up to the present fall largely into three groups—the variable reactance valve, the modulation circuit due to E. H. Armstrong, and the "Phasitron" (see selected references at end of chapter). Although there have been many alternative circuits developed from time to time they have not been widely used in practice.

### The Reactance Valve Modulator

The operation of the variable reactance valve type of frequency modulator can be followed with the aid of Fig. 7.1. The master oscillator circuit has connected across it a resistance $R$ and a

capacitor $C$. The value of these components is so arranged that the resistance is high in comparison with the capacitor's impedance. Under these conditions the voltage across the capacitor lags almost 90° behind that across the tuned circuit. This lagging voltage is applied to the grid of $V_1$. As the anode current drawn by a valve is in phase with its grid voltage, it follows that the

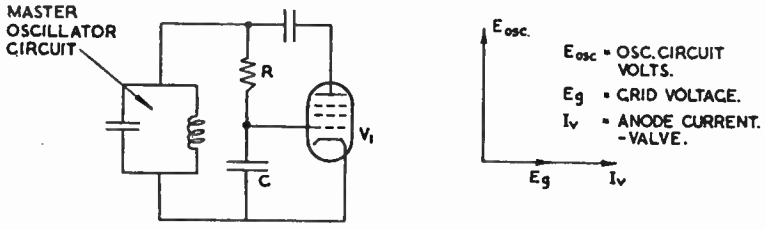



FIG. 7.1.—The basic circuit of the variable reactance valve modulator.
*(By courtesy of the British Institute of Radio Engineers.)*

current flowing through $V_1$ lags almost 90° behind that across the tuned circuit. As the valve fulfils all the necessary conditions, it may therefore be regarded as an inductance shunted across the tuned circuit. The value of this "inductance" will, of course, be varied by altering the valve anode current. It therefore follows that the application of an audio signal to the valve's grid will in

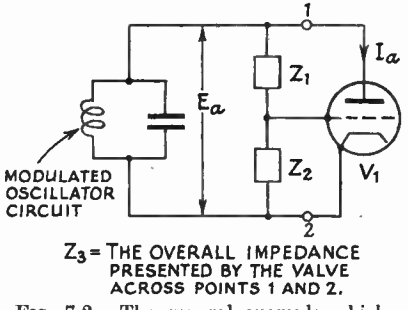



FIG. 7.2.—The general example which illustrates the controlling impedances in a reactance modulator circuit.

effect cause alterations to the "inductance" shunted across the tuned circuit. The oscillator will therefore be modulated in frequency as a result of signals applied to the grid of the reactance valve.

If now, instead of the particular example illustrated in Fig. 7.1, the general example given in Fig. 7.2 is considered, it will be

found that the reactance $Z_3$ presented by the valve across points 1 and 2 is as follows:

$$Z_3 = \frac{E_a}{I_a} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7.1)$$

$$= \frac{E_a}{g_m E_g} . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7.2)$$

Expressing this in terms of the phase shifting impedances:

$$\therefore \; Z_3 = \frac{Z_1 + Z_2}{g_m Z_2} , \quad . \quad . \quad . \quad . \quad . \quad . \quad (7.3)$$

where $E_a$=the r.m.s. voltage developed on the valve anode;
$I_a$=the r.m.s. anode current in amps;
$E_g$=the r.m.s. voltage applied to the grid;
$g_m$=the valve's mutual conductance in amps per volt at the operating point.

The valve reactance $Z_3$ can be further resolved into a resistance $R_3$ in parallel with a reactance $X_3$ which may be either inductive ($L_3$) or capacitive ($C_3$). The table below gives values of these components in terms of those used to produce the phase-shifted voltage applied to the valve's grid.

TABLE 12

*The equivalent impedance presented by a reactance valve expressed in terms of the phase shifting network $Z_1$ and $Z_2$.*

| $Z_1$ | $Z_2$ | $Z_3$; equivalent parallel resistive and reactive components | |
|---|---|---|---|
| | | $R_3$ | $L_3$, $C_3$ |
| $R$ | $C$ | $\frac{1+(\omega CR)^2}{g_m}$ | $L_3 = \frac{1+(\omega CR)^2}{g_m \omega^2 CR}$ |
| $C$ | $R$ | $\frac{1+(\omega CR)^2}{g_m(\omega CR)^2}$ | $C_3 = \frac{g_m CR}{1+(\omega CR)^2}$ |
| $R$ | $L$ | $\frac{R^2+\omega^2 L^2}{g_m \omega^2 L^2}$ | $C_3 = \frac{g_m LR}{R^2+\omega^2 L^2}$ |
| $L$ | $R$ | $\frac{R^2+\omega^2 L^2}{g_m R^2}$ | $L_3 = \frac{R^2+\omega^2 L^2}{g_m \omega^2 LR}$ |

As the voltage applied to the valve grid is limited by the length of its grid base this determines the minimum ratio of the reactances $Z_1$ and $Z_2$, the phase-shifting components. With the aid of the above table and a knowledge of the valve's mutual conductance at the bias value being used, it is possible to calculate the impedance it presents across the tuned circuit under all conditions.

In order to produce a variable reactance, it is apparent from the table that the mutual inductance of the valve must be variable. This may be achieved in one of two ways; if the modulation signal is applied to the same grid as the r.f. input, the valve must be biased to a point of the $i_a - v_g$ characteristic where appreciable curvature exists, so that the mutual conductance varies with modulation signal amplitude. Alternatively, the modulation may be applied to an outer grid of a multi-grid valve; in this case the modulation signal controls the proportion of the electron stream through the valve reaching the anode, and hence the effective mutual conductance from the inner grid to which the r.f. signal is applied.

In all four cases set out in the table, it is of interest to note that $X_3$, the reactive element in parallel with the main tuned circuit, is inversely proportional to $g_m$. This type of modulator should therefore more correctly be termed a susceptance modulator, since it is $jB_3 = 1/jX_3$ which varies linearly with $g_m$.

### Reactance Modulator Sensitivity

For simplicity it will be assumed in this section that the variable reactance valve has the modulation and r.f. signals applied to its control grid, and that the mutual conductance characteristic varies linearly from cut-off to zero bias. It will also be assumed that the phase shift network always supplies the grid with a voltage which has been shifted by a full 90°, and therefore that the in-phase component of the power drawn by the reactance valve may be neglected. Under these conditions it is possible to calculate with sufficient accuracy for all practical purposes the sensitivity and band-width coverage of a reactance valve modulator.

Firstly, as has been shown by Winlund, the maximum modulator sensitivity, expressed in terms of radio-frequency deviation to audio input voltage, occurs when the peak audio and radio frequency voltage swings applied to the modulator grid are equal.

That is to say when

$$e_a = e_r = \tfrac{1}{2} V_g, \qquad (7.4)$$

where $e_a$ = peak audio voltage applied to modulator grid;
$e_r$ = peak radio frequency voltage applied to modulator grid;
$V_g$ = modulator bias voltage at the operating point.

The peak radio frequency current drawn by the reactance valve when there is no modulation is

$$i_m = e_r g_m. \qquad (7.5)$$

The peak oscillator tank capacity current is

$$i_f = \frac{e_0}{X_0} = e_0 \omega C_f, \qquad (7.6)$$

where $e_0$ = the peak oscillator tank voltage;
$C_f$ = ocillator tank fixed capacity;
$\omega = 2\pi f$;
$X_0$ = reactance of each of the two oscillator tank circuit components.

The fraction of the tank current flowing through the modulator without modulation (i.e. zero percentage modulation) is, then,

$$\left(\frac{i_m}{i_f}\right)_0 = \frac{e_r g_m}{e_0 \omega C_f}. \qquad (7.7)$$

At full deviation (i.e. at 100 per cent modulation or peak $e_a$) $e_r$ becomes 50 per cent greater, or,

$$\left(\frac{i_m}{i_f}\right)_{100} = \frac{3}{2} \frac{e_r g_m}{e_0 \omega C_f}. \qquad (7.8)$$

The change—expressed as a fraction of the total current flowing round the oscillator tank circuit, from zero to full modulation—will be the difference between equations (7.8) and (7.7). The resultant fractional frequency deviation will be one-half of this (due to the fact that the frequency changes as the square root of the capacitance change), providing that these deviations are only a relatively small percentage of $f$ the oscillator frequency. Thus,

$$\frac{\Delta f}{f} = \frac{1}{2}\left[\left(\frac{i_m}{i_f}\right)_{100} - \left(\frac{i_m}{i_f}\right)_0\right] \qquad (7.9)$$

$$= \frac{e_r g_m}{4 e_0 \omega C_f} = \frac{e_r g_m}{8 \pi f e_0 C_f}. \qquad (7.10)$$

The maximum band-width $\omega_m$, over which the modulator is capable of swinging the oscillator frequency is, then,

$$\omega_m = 2\Delta f = \frac{e_r g_m}{4\pi e_0 C_f} = \frac{V_g g_m}{8\pi e_0 C_f} = \frac{k_m V_g g_m}{e_0 C_f}. \quad . \quad (7.11)$$

In the above formula $k_m$—the "modulation constant"—is $\frac{1}{8}\pi$ or less, depending upon the percentage of the modulator valve's grid voltage to anode current ($g_m$) characteristic which is linear and therefore usable for the *sum* of the peak audio modulation voltage, the peak radio frequency voltage and the direct-current bias change necessary to make allowance for the extreme automatic frequency control correction. In a practical valve the truly linear part of the mutual conductance characteristic may well be but a fraction of the total. Under these conditions the modulator sensitivity in cycles per volt is, then,

$$S_m = \frac{\omega_m}{V_g}. \quad . \quad . \quad . \quad . \quad . \quad (7.12)$$

This derivation is only applicable to single valve class A modulators and for the assumptions specified. For push-pull Class A modulators, the value given in equation (7.11) for $\omega_m$ must be multiplied by two, while that for $S_m$ in equation (7.12) must also be doubled.

### Distortion in Reactance Modulators

If it is assumed that the capacitance of a tuned circuit remains constant and the inductance element only is varied, then it follows from the formula for the resonant frequency of a tuned circuit $\left(f = \frac{1}{2\pi\sqrt{LC}}\right)$ that the frequency will vary as the square root of the change in inductance. From this it follows that even if the inductance change is strictly proportional to the modulating signal amplitude, the resultant frequency variations will not be so. It is therefore apparent that a variable reactance valve must produce some distortion in frequency modulating an oscillator.

In a circuit embodying a reactance valve the inductive reactance so formed does not behave as in a normal tuned circuit; instead, a reactance is produced in which the current is substantially independent of the frequency variations. This abnormal condition

has been examined by Winlund, in order to determine the distortion which results from the use of a reactance valve as a frequency modulator. The table below shows the conclusions reached.

TABLE 13

*Distortion due to reactance valve modulators*

| Fractional band-width (per cent) | Distortion (per cent) | Distortion / Fractional band-width |
|---|---|---|
| 100 | 13·6 | 0·136 |
| 40 | 5·3 | 0·132 |
| 20 | 2·8 | 0·14 |
| 10 | 1·4 | 0·14 |
| 4 | 0·56 | 0·14 |
| 2 | 0·28 | 0·14 |
| 1 | 0·14 | 0·14 |
| 0·4 | 0·056 | 0·14 |
| 0·2 | 0·028 | 0·14 |
| 0·1 | 0·014 | 0·14 |

NOTE—Fractional band-width = Peak to peak frequency swing expressed as a percentage of the centre-operating frequency.

A study of the above table reveals the somewhat surprising result that the distortion is an approximately fixed percentage of the fractional band-width, i.e. some 14 per cent. Assume as an example a swing of plus and minus 75 kc/s on a mean frequency of 50 Mc/s. The fractional band-width is therefore some 0·3 per cent, from which it follows that the distortion due to reactance curvature is 0·04 per cent. If this swing is obtained by multiplying up from a lower centre frequency, the fractional band-width, and therefore the distortion, remain the same. As the distortion is well below that required for even a high-fidelity transmission, it follows that it is permissible to work with a lower reactance modulator frequency and, if so desired, use a heterodyne frequency changer to obtain the final carrier frequency. If this circuit arrangement is employed care must, however, be taken not to exceed—in the modulated oscillator alone—the total distortion which can be permitted.

Even if a fairly difficult case is taken the result is still not too unsatisfactory. Take, for example, the transmission of frequency modulated signals over either power or telephone lines. If a

carrier of 30 kc/s and a deviation of $\pm 6$ kc/s is assumed, the distortion arising from the use of a reactance modulator will be some 6 per cent. It should, of course, be borne in mind that this figure assumes a ruler-straight reactance valve characteristic. In practice the valve characteristic curvature will add to the distortion figure. The distortion produced by variable reactance valves is usually far too great to permit their use for the direct production of extremely large fractional band-widths, such as those encountered in sub-carrier frequency modulation transmission.

### Push-Pull Reactance Modulators

If a simple oscillator is frequency modulated by means of a reactance valve, the question of frequency stability immediately crops up. As the reactance valve is a device which makes the frequency of an oscillator dependent upon the voltages applied to its electrodes, it follows that voltage variations other than the desired modulation signal will also cause the frequency to vary. Unless suitable steps are taken this will result in very poor frequency stability. The first and obvious method of improving the stability of a simple reactance valve modulator is that of using a voltage-regulated power supply. In practice such a step is an absolute necessity in any circuit in which the simple modulator is employed.

Another and, in general, more satisfactory method of reducing the modulator's susceptibility to power supply variations is by means of a push-pull reactance valve circuit, such as that illustrated in Fig. 7.3. In this circuit the two valves are arranged so that they produce opposite reactance variations. It follows that they must be supplied with push-pull modulating signals in order that their reactive effects may add together. Any unwanted voltage drift or variation (including power supply ripple) will then be cancelled out as in an audio push-pull amplifier.

Valves $V_1$ and $V_2$ are the reactance modulators and $V_3$ is the oscillator. The use of heptode valves is suggested because of the convenience of having an extra control grid for applying the modulation. The phase shift network feeding valve $V_1$ is arranged in the same way as that used for automatic frequency control in broadcast receivers. The resistance $R_1$ and the grid to cathode capacity ($C_{gc}$) of the valve forms the phase shift network supplying the grid feed voltage. $R_1$ is made large in comparison with the

reactance of $C_{gc}$, so that the phase of the voltage developed at the valve grid lags by 90° that across the oscillator circuit. It therefore follows that this valve's anode current will also lag the voltage across the oscillator circuit. As the valve is drawing a lagging current it is thus effectively a shunt inductance across the oscillator circuit.

The valve $V_2$ uses a phase-shifting network which supplies it

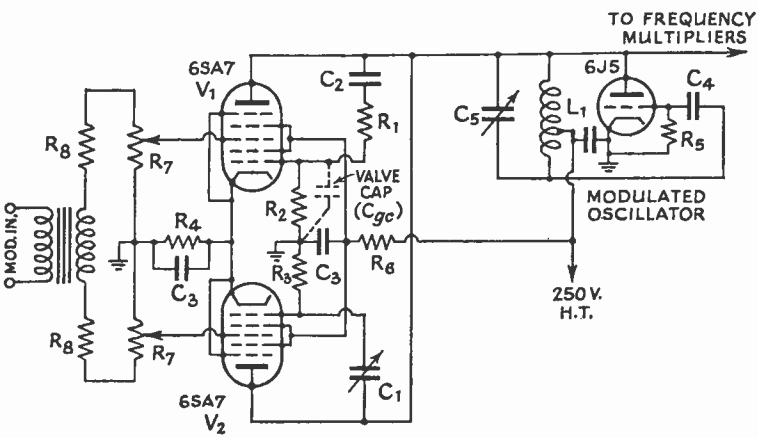



FIG. 7.3.—The circuit of a push-pull reactance valve modulator. Valve $V_1$ is fed with a phase-shifted voltage via $C_2$ and $R_1$ so that its anode circuit acts as a shunt inductance the value of which is dependent upon the gain of the valve. Similarly the valve $V_2$ is fed via $C_1$ with a signal which causes the valve to present the effect of a shunt capacity across the oscillator circuit. The application of a push-pull audio modulation therefore produces additive frequency modulation of the oscillator's frequency.

with a leading voltage and thus results in its presenting a capacitive reactance effect instead of an inductive effect in the anode circuit. In the phase-shift network the capacitor $C_1$ is small enough to have a reactance which is high in comparison with the resistance $R_3$—the capacitor being the controlling factor it therefore follows that the current flowing in this circuit leads the voltage. As it is this current which results in the voltage developed across the resistance $R_3$, it follows that the voltage applied to the valve grid must also lead the voltage across the oscillator circuit. This being so, this valve's anode current will also be leading, so causing the valve to present an effective shunt capacitance across the oscillator circuit. Any change in voltage which is applied to

both modulator valves will therefore produce reactance variations which will be cancelled out at their anodes.

It will be noted that a Hartley oscillator circuit is shown in Fig. 7.3, and that the reactance valves are only connected across a part of the oscillator circuit. This reduces to some extent the effectiveness of the reactance valves. However, in oscillator circuits in which one end of the tuned circuit is grounded, the cathode is usually at a radio frequency potential from ground. Such an arrangement invariably leads to hum in the form of frequency modulation. This may be avoided by choosing a circuit in which the cathode is grounded. As an alternative the heater may be raised to the same radio-frequency potential as the cathode by means of radio frequency chokes.

With the aid of a balanced reactance valve modulator it is possible to neutralise all frequency instability due to power supply variations. If, as frequently occurs, the oscillator which is being modulated itself varies in frequency as a result of power supply changes, then this may also be neutralised by means of the balanced reactance valves. When used for this purpose, the valves are adjusted so that they are slightly off-balance; in this way they produce a residual reaction which is equal and opposite to the oscillator's reaction to power supply variations.

When it is desired to use the reactance valves for this overall balancing, the circuit may be adjusted by deliberately switching a series resistance in and out of the power supply lead and at the same time varying $C_1$ until the frequency variations resulting are negligible. This frequency variation may be best observed by noting the beat-note it produces when heterodyned against a stable oscillator.

When it is merely desired to balance the two reactance valves so that they are themselves unaffected by power supply variations, the grids of the two modulator valves may be tied temporarily together so that they are modulated in parallel. Modulation is then applied and $C_1$ adjusted for a minimum frequency modulation output. After the balance has been obtained the modulator grids can be connected back in push-pull.

The push-pull reactance modulator has an additional advantage in that the resistive components in parallel with the main oscillator circuit due to the two valves vary in opposite senses when modulation is applied. With a single-ended modulator, the variation of

the resistive term tends to produce unwanted amplitude modulation. T. P. Flanagan has shown that the condition for constant resistance is the same as that for reactance balance at the centre frequency. In terms of the components in Fig. 7.3, this condition is $\omega_0{}^2=1C_1R_3C_{gc}R_1$. Under these conditions of operation, the reactance modulator has no effect in determining the centre frequency, this being determined solely by the oscillator elements

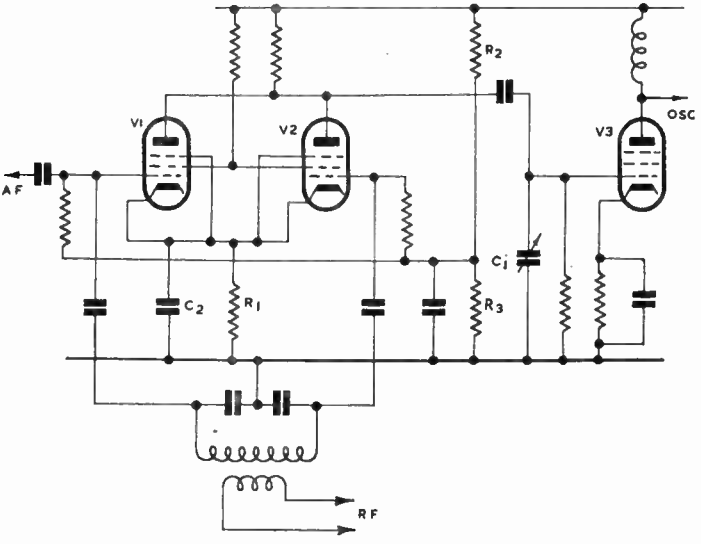



FIG. 7.4.—Balanced reactance modulator of type employed in Marconi FMQ circuit.

($L_1$ and $C_5$ of Fig. 7.3). If the values of $C_1$, $R_3$, $R_1$ are chosen so that $1/\omega_0C_{gc}R_1=1/\omega_0C_1R_3=k$, the shunt resistance presented to the oscillator tuned circuit is given by $(1+k^2)/2g_mk^2$.

An alternative form of push-pull reactance modulator has been developed by Marconi for use in the FMQ oscillator; a simplified circuit diagram of the arrangement is shown in Fig. 7.4. Valves $V_1$ and $V_2$ form a "long-tailed pair", sharing a high valve common cathode load $R_1$. As the cathodes of the two valves are consequently at a relatively high potential, the grid resistors are returned to a potential divider $R_2$, $R_3$ connected across the h.t. supply. The value of $R_1$ is such that the bias applied to the two valves $V_1$ and $V_2$ brings the anode current near cut-off value, and the mutual conductance varies linearly with small charges of grid

bias. Capacitor $C_2$ serves to decouple the cathodes of $V_1$ and $V_2$ to ground at r.f., but not at a.f. A portion of the r.f. signal from the oscillator circuit is fed to the grids of both valves $V_1$ and $V_2$ in anti-phase; consequently with no modulation applied, the sum of the r.f. components of the two anode currents is zero, and there is no signal fed to the grid of $V_3$.

When, however, modulation is applied to the grid of $V_1$, the cathode potential tends to follow the input signal, and hence $V_2$ anode current alters, tending to maintain the cathode potential constant. $R_1$ is sufficiently large for near-equality variations of anode currents in the valves to be achieved, the difference between the two currents being such as to maintain the drive to $V_2$ at approximately half the input signal amplitude. Thus when modulation is applied the grid-cathode bias applied to the two valves alters differentially, and hence the effective mutual conductance of the two valves varies differentially also. An r.f. voltage is therefore developed across $C_1$ which forms the major portion of the common anode load at r.f. This output is proportional to the modulating signal amplitude and, additionally, is in quadrature with the input voltage. The voltage across $C_1$ is fed to the grid of $V_3$; the anode of $V_3$ is connected in parallel with the oscillator tuned circuit. Thus when modulation is applied, $V_3$ behaves as a reactive circuit element, its alternating component of anode current being proportional to the a.f. input signal. This circuit arrangement has a number of advantages over that conventionally employed. Since $V_3$ is purely an amplifying stage, its performance is relatively non-critical with regard to operating potentials and it can therefore be made to deliver a relatively large r.f. current without overloading. Conversely, the anode current swings in $V_1$ and $V_2$ can be kept to relatively low magnitudes in order to ensure maximum linearity.

Capacitor $C_1$ is made variable to provide a control of deviation sensitivity; for a constant r.f. input its value can be altered to vary the drive to $V_3$, and hence the frequency swing.

### Stabilised Reactance Modulators

In practice there are other factors besides the variations in supply voltage which affect the oscillator frequency. There are changes occasioned by differences in temperature and humidity. While these variations are largely a matter of mechanical design

they still remain a problem to be surmounted. While they may be very largely overcome by placing the components involved in a simple thermostatically controlled oven, this still does not overcome effects due to such causes as the ageing of valves. In order to be certain of holding all frequency drift within a tolerance of $\pm 1{,}000$ or $\pm 2{,}000$ cycles in some 100 Mc/s, further and more positive methods of stabilisation have to be resorted to.

The representative circuit shown in Fig. 7.5 was first described

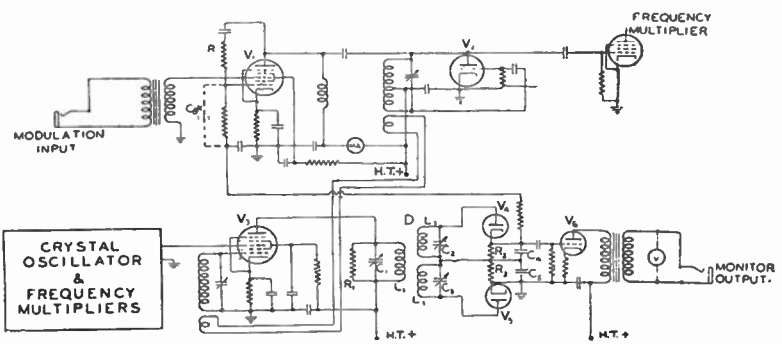



FIG. 7.5.—A reactance valve frequency modulator with automatic frequency control. The A.F.C. Circuits serve the dual function of maintaining frequency stability and providing a monitoring circuit.

$I.F.$ = 460 kc/s.
$L_1 = L_2 = L_3 = 2{\cdot}5$ Mh.
$R_2 = R_3 = 200{,}000$ ohms.
$C_1 = C_2 = C_3 = 0-75$ pf midget condenser.
$R_1 = 50{,}000$ ohms.
$C_4 = C_5 = 200$ pf.
Coupling between $L_1$ and $L_2$ = 14 per cent.
Coupling between $L_2$ and $L_3$ = 2·5 per cent.

(*By courtesy of "R.C.A. Review".*)

by Crosby. It will be noted that this circuit uses a simple oscillator valve $V_2$ and a single reactance valve $V_1$. Normally such an arrangement would have a bad frequency stability if it were not for the automatic frequency control system consisting of the heterodyne frequency converter $V_3$ and the discriminator $D$, with the two detectors $V_4$ and $V_5$. This automatic frequency control system operates off the relatively low frequency obtained from the heterodyned output of the frequency changer stage. By using a heterodyne frequency changer to convert the modulated oscillator's frequency to a much lower value, any drift—expressed in terms of cycles "off" the mean frequency—remains unchanged. At the lower frequency the discriminator circuit will be very much more stable—in terms of cycles actually drifted—than would have

been the case if it had been operating at the same frequency as the reactance modulated oscillator. The crystal oscillator supplying the signal to the frequency changer valve must have a stability higher than that desired for the final carrier frequency. For maximum stability the intermediate frequency at which the discriminator $D$ operates should be as low as possible; 450 kc/s (the broadcast i.f.) is frequently used for this purpose.

Stabilisation in the above arrangement occurs in the following manner. Assume that the oscillator frequency of valve $V_2$ is low; this will result in the frequency applied to the discriminator also being low. The discriminator output instead of being zero will now be a slight positive voltage, which when applied to the grid of the variable reactance valve $V_1$ causes it to draw a larger lagging current and so present an appearance of a smaller value of shunt inductance across the oscillator circuit. This results in the oscillator frequency being raised.

There are many possible variations of the typical reactance valve circuits already described; it is, for example, possible to apply direct-current amplification to the voltage output of the discriminator before applying it to the variable reactance valve. Normally, when a stabilised oscillator is employed, a balanced modulator is used and both the h.t. and l.t. supplies are stabilised. It is also normal to enclose the critical components in a thermostatically controlled oven. As the discriminator's action forms the subject-matter for a later chapter it will not be discussed at this point.

### FMQ Modulator (Frequency Modulated Quartz)

The FMQ oscillator circuit has been developed by Marconi for the purpose of effecting frequency modulation of a crystal controlled oscillator. It employs a reactance modulator of the type described in the section on push-pull reactance modulators. The method of modulating the crystal itself is of sufficient importance to warrant special attention. It is not normally possible to obtain satisfactory results by connecting a reactance modulator to a circuit employing crystal control; the reasons for this will be apparent from an inspection of the equivalent network of a crystal given in Fig. 7.6. The capacitor $C_h$ is the crystal holder capacitance and that between the leads etc., the components $L_c$, $R_c$, and $C_c$ being those which would give precisely the same electrical

performance as the crystal itself at the operating frequency. Typical values of $L_c$ run into henries whilst those of $C_c$ run in fractions of pico-farads; it is therefore obvious that any charge in the reactance in parallel with $C_h$ has very little effect on the resonant frequency of the circuit, since this is dominated by $L_c$ and $C_c$. Expressed alternatively, it may be said that $C_h$ and $C_c$ form tapping points to the crystal circuit, and their relative magnitudes are such that any external element is tapped across an extremely small portion of the circuit.

In order to effect modulation, an impedance transformation is necessary, and for this purpose a quarter-wave network is

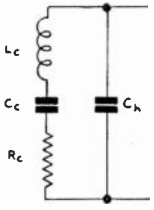



FIG. 7.6.—Equivalent circuit of crystal.

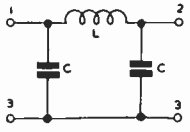



FIG. 7.7.—Quarter wave $\pi$-section network.

employed. This comprises a $\pi$-section network, of the type shown in Fig. 7.7; if the values of $L$ and $C$ are chosen so that $f_0=\frac{1}{2\pi\sqrt{LC}}$, where $f_0$ is the working frequency, then the impedance measured between terminals 1 and 3 ($Z_{13}$), when a load impedance $Z_L$ is connected to the terminals 2 and 3, is given by

$$Z_{13}=Z_0{}^2/Z_L, \quad . \quad . \quad . \quad . \quad . \quad (7.13)$$

where $Z_0{}^2=L/C$.

If then the load comprises a crystal, and the value of $C$ connected between terminals 2 and 3 is adjusted to incorporate $C_h$, the load is equivalent to a series tuned circuit comprising $L_c$, $R_c$, and $C_c$. Then input impedance is given by (7.13) above; it, is, however, more convenient to calculate the input admittance given by $Y_{13}=1/Z_{13}$.

$$Y_{13}=\frac{j\omega L_c+\frac{1}{j\omega C_c}+R_c}{Z_0{}^2}.$$

That is, the network input admittance is the same as that of a

parallel tuned circuit comprising three branches, $L$, $C$, and $R$, the relationship being given by:

$$L = Z_0^2 C_c;$$
$$C = L_c / Z_0^2;$$
$$R = Z_0^2 / R_c. \quad . \quad . \quad . \quad . \quad . \quad (7.14)$$

To the input terminals 1 and 3 can be therefore connected a maintaining amplifier and a reactance modulator. The circuit has the advantage of retaining the high centre-frequency stability of a crystal oscillator.

Considerable care is necessary in mounting the crystal, to prevent operation in spurious modes under conditions of varying frequency; additionally, precautions must be taken to ensure that oscillation is not controlled by the quarter-wave section itself, the resonant frequency of which is $f_0/\sqrt{2}$.

## Armstrong's Frequency Modulator

E. H. Armstrong is responsible for developing the method of frequency modulation which will now be discussed. The chief advantage of his method is that, being based on a crystal controlled

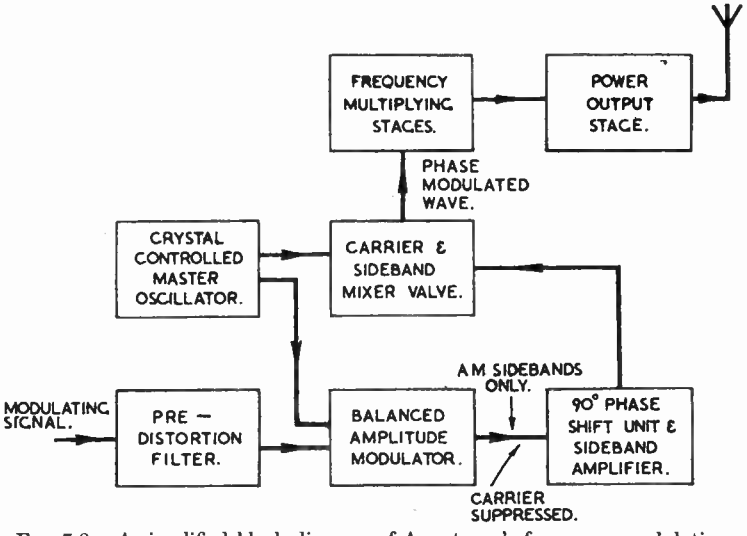



FIG. 7.8.—A simplified block diagram of Armstrong's frequency modulation transmitter. The mixing of a carrier and phase shifted amplitude modulation side bands produces phase modulation. As a result of the pre-distortion of the incoming audio signal the overall effect is that of frequency modulation.

*(By courtesy of the British Institute of Radio Engineers.)*

oscillator, it has an inherent frequency stability which is probably higher than any other type of modulator.

A block circuit diagram of a transmitter incorporating Armstrong's modulator is shown in Fig. 7.8. In such a transmitter

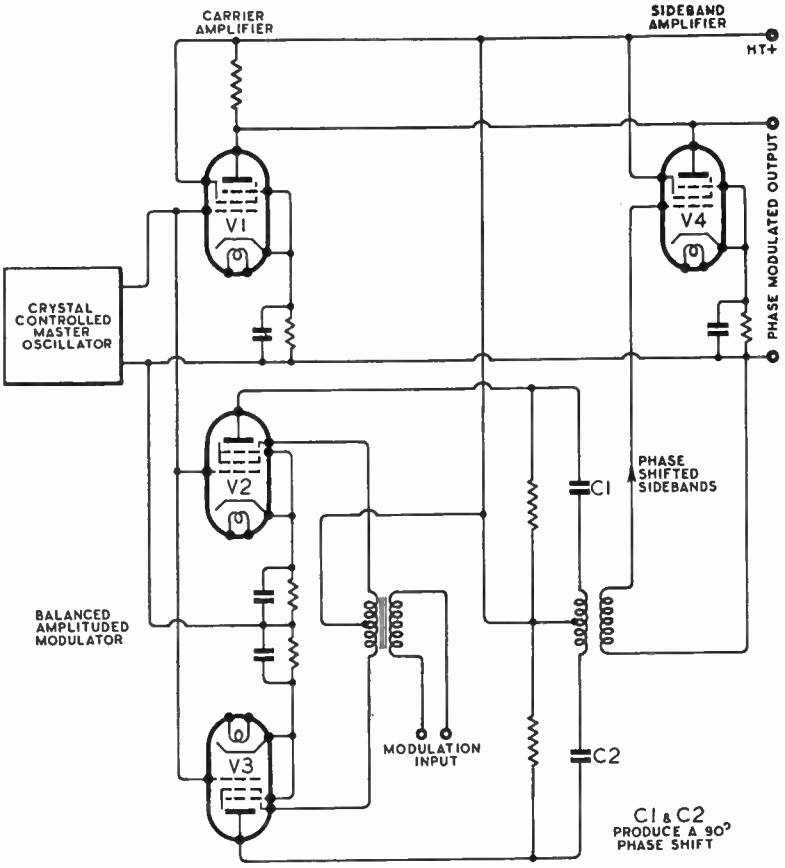



FIG. 7.9.—An outline circuit of Armstrong's phase modulator.

the output from a very stable crystal controlled oscillator is fed through two channels. The first channel starts with a pair of valves ($V_2$ and $V_3$, Fig. 7.9) operating as a balanced amplitude modulator. The carrier is cancelled out across the output transformer in the anode circuit of these two valves, so leaving the amplitude modulation side bands only. By feeding these side bands

through the small condensers $C_1$ and $C_2$, a phase shift of 90° is produced. The second outlet channel from the crystal oscillator, after being amplified by $V_1$, is combined with the phase-shifted amplitude modulation side bands on the anode of $V_4$.

The effect of combining a carrier with amplitude modulation side bands which have been shifted in phase by 90° is shown in Fig. 7.10. These vector diagrams show, firstly, how the side bands

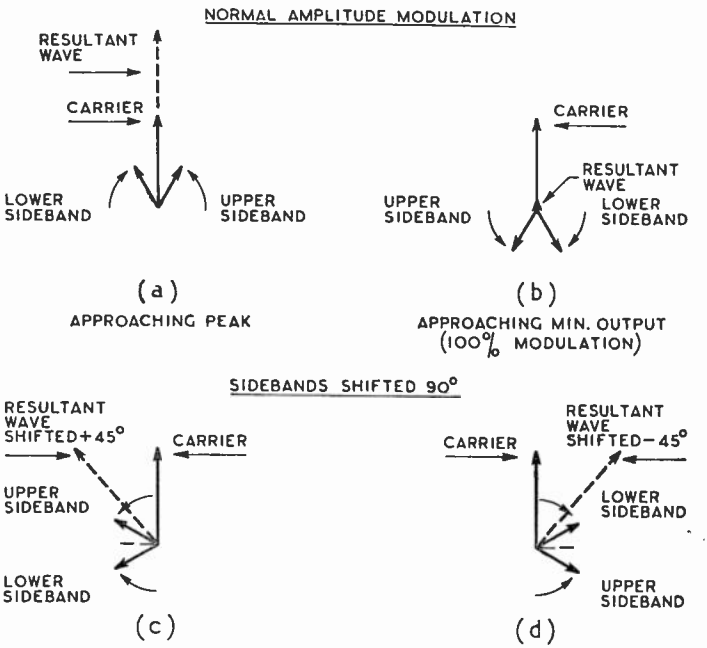



FIG. 7.10.—Vector diagrams (*a*) and (*b*) show the normal phase relations between the side bands and carrier of an amplitude modulated wave. Diagrams (*c*) and (*d*) show that by shifting the side bands by 90° relative to the carrier, phase modulation is produced.

normally combine with the carrier to produce amplitude modulation, and, secondly, how when shifted in phase they produce phase modulation. It will be recalled from the discussion in Chapter Two that the relationship between phase and frequency modulation is simply that in the former case the frequency deviation of the carrier is directly proportional to the differential of the audio signal rather than, as in the case of frequency modulation, being proportional to the audio signal itself. This being

so, it follows that, if it is desired to obtain a frequency modulated signal from a modulator producing a phase-modulated signal, it is only necessary to integrate the audio signal before applying it to the modulator. When used with such a pre-distortion circuit Armstrong's phase modulator gives a frequency modulation output.

### Distortion Produced by Armstrong's Modulator

If, in a modulator producing phase-modulation, the angle $\theta$ is made to vary linearly with the side band vector amplitude $E_s$, then it follows that there will be no distortion produced at this point in the system. It is therefore necessary that $\theta$ should be related to $E_s$ by the equation $\theta = KE_s$, where $K$ is an arbitrary constant. However, it is apparent from Fig. 7.11 that they are in fact related by the equation $\theta = \tan^{-1}\left(\frac{E_s}{E_c}\right)$; it is therefore obvious that the modulator introduces distortion equivalent to the difference between the desired relationship and the actual relationship. Jaffe has investigated this position and shown that this distortion takes the form of the production of odd harmonics and that the amplitude of these harmonics can be expressed by the following equation:

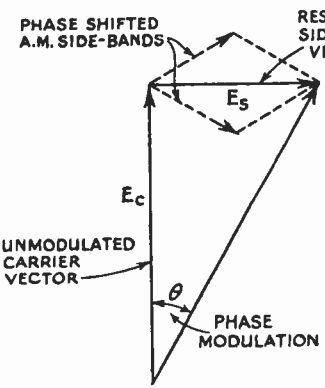



FIG. 7.11.—The carrier vector $E_c$ and the resultant side band vector $E_s$.

$$A_n = \frac{2}{p^{n+1}}(\sqrt{1+p^2}-1)^n, \qquad . \quad . \quad . \quad (7.15)$$

where $A_n$ = the amplitude of the $n$th harmonic;
$p$ = the tangent of the maximum phase shift expressed in degrees.

Fig. 7.12, which is prepared from the above formula, gives the percentage harmonic distortion in terms of the fundamental, against phase shift values ($\theta$) of up to 45°. It should be noted that the bulk of the distortion is made up of third harmonic.

As the audio signal is integrated before being applied to

Armstrong's modulator, it follows that the resulting phase modulation deviation will be progressively decreased as the frequency of the audio signal is increased. This being so, it is obvious that the larger phase excursions will only occur in the lower frequency region, with the result that the maximum distortion will also occur in this region.

This is well illustrated in Fig. 7.13, which shows the distortion

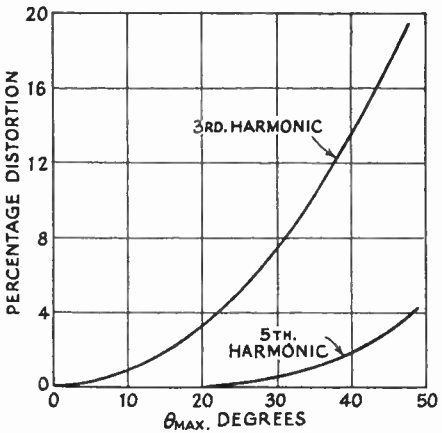



FIG. 7.12.—Percentage distortion versus the maximum phase deviation in degrees, which results from Armstrong's Modulator.

produced over the audio band when the phase modulation has been made equal to some 25·5° at 20 cycles. It will be noted that the distortion is 5 per cent at this frequency, but that it falls to negligible proportions above 100 cycles.

It is interesting to note that neither the percentage frequency change nor the percentage harmonic distortion is affected by frequency multiplication. If heterodyning is employed to alter the frequency, the percentage frequency change will be modified, but the percentage distortion will be unaffected. It is common practice to use a maximum phase shift of 30° for an audio signal of 30 cycles. The distortion will then be some 7·2 per cent at 30 cycles and less than 0·05 per cent at 400 cycles. Under these conditions the actual modulation produced would be some $\frac{30°}{57·3°} = 0·524$ radians, which results in a frequency swing of only $30 \times 0·524 = 16$ cycles for a modulating frequency of 30 cycles.

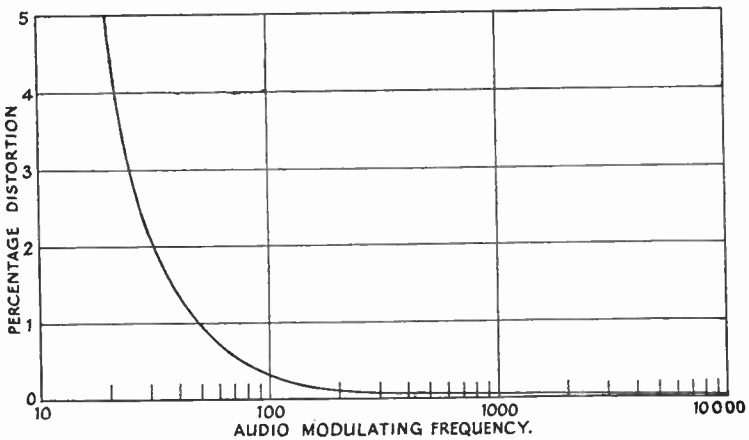



FIG. 7.13.—The percentage harmonic distortion over the audio frequency range for the case when a phase change of some 25·5° is produced by a modulating frequency of 20 cycles.

## Minimising Distortion in Armstrong's Modulator

In the last section it was shown that whereas $\theta$, the phase deviation in Armstrong's modulator, should be strictly proportional to variations in the amplitude of the phase-shifted side bands —i.e. $\theta = KE_s$ (see Fig. 7.11)—it is in fact proportional to

$$\theta = \tan^{-1}(E_s/E_c).$$

It will be apparent that the desired linear relationship can be obtained by varying the amplitude of the carrier vector $E_c$. In this way it is possible to make

$$\theta = KE_s = \tan^{-1}(E_s/E_c), \quad . \quad . \quad . \quad . \quad (7.16)$$

or

$$\frac{E_s}{E_c} = \tan KE_s. \quad . \quad . \quad . \quad . \quad (7.17)$$

From this last equation the desired relationship which the amplitude of $E_c$ should bear to that of the side bands can be derived as

$$E_c = E_s \cot KE_s. \quad . \quad . \quad . \quad . \quad (7.18)$$

Bertram has shown that for small values the cotangent can be expressed by the series

$$\cot x = \frac{1}{x} - \frac{x}{3} - \frac{x^3}{45} - \frac{2x^5}{945}, \quad . \quad . \quad . \quad (7.19)$$

so that

$$E_c = E_s\left[\frac{1}{KE_s} - \frac{KE_s}{3} - \frac{(KE_s)^3}{45} - \frac{2(KE_s)^5}{945}\right] \quad . \quad (7.20)$$

$$= \frac{1}{K} - \frac{KE_s^2}{3} - \frac{K^3E_s^4}{45} . \quad . \quad . \quad . \quad . \quad . \quad (7.21)$$

It will be shown later that by amplitude modulating the carrier with a wave having the above form, it is possible to operate up to some 60° phase shift without appreciable distortion. Accepting this figure for the moment and making $E_s=1$ for a 60° (or 1·048 radian) phase shift, then $K=\frac{\theta}{E_s}=1{\cdot}048$. Hence it follows from equation (7.21) that for complete correction of all distortion up to this phase shift the carrier $E_c$ must be amplitude modulated with a wave of the following form:

$$E_c = 0{\cdot}955 - 0{\cdot}349E_s^2 - 0{\cdot}025E_s^4. \quad . \quad . \quad (7.22)$$

This shows that if the carrier amplitude (the factor $K$) is taken as 0·955, then the carrier should be amplitude modulated to a relative depth of 0·349 with the side bands' second harmonic and also to a relative depth of 0·025 with their fourth harmonic. Expressed in different terms, this means that when phase modulating to 60° (by Armstrong's method) the carrier should also be amplitude modulated to $\frac{0{\cdot}349}{0{\cdot}955}=36$ per cent with the audio signal's second harmonic and to 2·6 per cent with its fourth harmonic.

In practice it is, however, difficult to produce accurately more than the second harmonic of the audio signal. Pieracci has therefore given empirically derived values for the second harmonic correction which should be applied to a carrier which is phase modulated to a maximum of 60°. Under these conditions:

$$E_c = 0{\cdot}765 + 0{\cdot}188 \cos 2\omega_a t, \quad . \quad . \quad . \quad (7.23)$$

where 0·765=the relative mean carrier amplitude;
0·188=the relative amplitude of the second harmonic of the audio signal $\omega_a$.

Comparing this figure with equation (7.22)

$$E_c = 0{\cdot}765 + 0{\cdot}188(1 - 2\sin^2 \omega_a t), \quad . \quad . \quad (7.24)$$

$$= 0{\cdot}953 - 0{\cdot}376 \sin^2 \omega_a t, \quad . \quad . \quad . \quad . \quad (7.25)$$

or

$$E_c = 0{\cdot}953 - 0{\cdot}376 E_s{}^2. \quad . \quad . \quad . \quad . \quad . \quad (7.26)$$

It will be noted that this equation compares very closely with that derived theoretically by Bertram.

It therefore follows that if a simple sine wave having twice the frequency of the audio signal is applied as amplitude modulation to the carrier, then there will be a very close approximation to a linear relationship between $\theta$ and $E_s$. The higher coefficient of $E_s$ used in the practical case compensates to some extent for the neglected higher order terms. As it is possible to correct for a sine wave of any frequency, it follows that the distortion can be corrected however complex the modulating signal may be.

TABLE 14

*Distortion remaining after second harmonic correction of 60-degree phase modulation*

| θ assumed (degrees) | $KE_s$ | 0·765+0·188 cos $2\omega_a t$ | θ actual (degrees) | Error remaining (degrees) |
|---|---|---|---|---|
| 6 | 0·952 | 0·949 | 6·03 | 0·03 |
| 12 | 0·940 | 0·938 | 12·0 | 0·0 |
| 18 | 0·922 | 0·919 | 18·0 | 0·0 |
| 24 | 0·899 | 0·892 | 24·1 | 0·1 |
| 30 | 0·866 | 0·859 | 30·2 | 0·2 |
| 36 | 0·825 | 0·822 | 36·0 | 0·0 |
| 42 | 0·778 | 0·770 | 42·2 | 0·2 |
| 48 | 0·720 | 0·713 | 48·2 | 0·2 |
| 54 | 0·652 | 0·649 | 54·2 | 0·2 |
| 60 | 0·577 | 0·577 | 60·0 | 0·0 |

Table 14 indicates the theoretical error resulting from this practical correction. It should be noted that it is not practical to correct the distortion on phase deviations very much greater than some 60°. This is due to the fact that the limiting value of 90° phase modulation is being approached.

## Distortion Correction Circuits Applied to Armstrong's Modulator

In addition to a theoretical examination of the correction of distortion in Armstrong's modulator, Pieracci has also described a practical method of carrying it into effect. Fig. 7.14 shows a typical circuit diagram of a modulator embodying this principle.

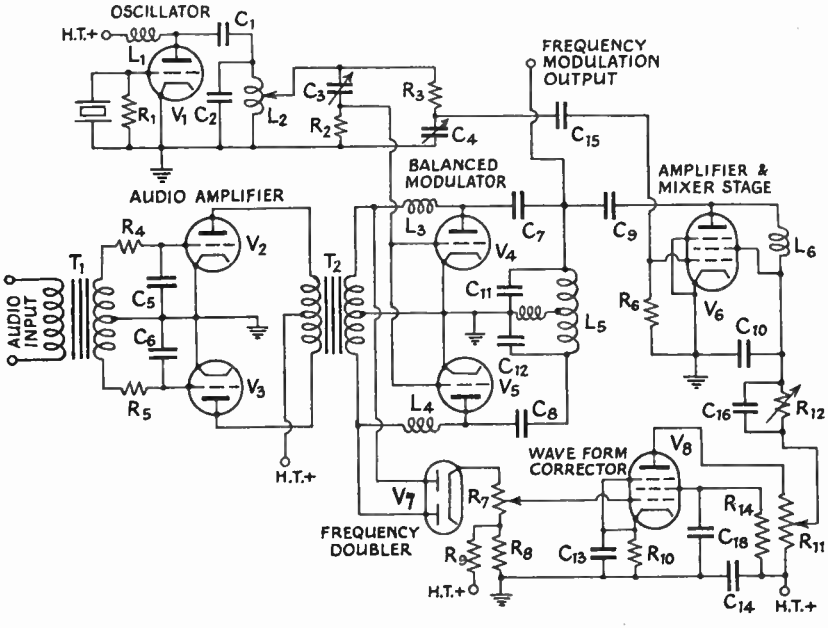



FIG. 7.14.—A circuit diagram of Armstrong's modulator with the addition of distortion correction circuit. This arrangement minimises the distortion produced at large-phase deviations.

The audio amplifier stage, including $V_2$ and $V_3$, the balanced modulator $V_4$ and $V_5$, the oscillator $V_1$ along with the phase-shifting network $R_2$, $C_3$, $R_3$, and $C_4$, and the carrier amplifier and mixer stage, $V_6$, provide frequency modulated signals in the general manner already described. The new feature introduced in this circuit is the amplitude modulation of the phase-modulated carrier with a wave-form having twice the frequency of the original modulating signal. The valve $V_7$ full wave rectifies the audio signal, so providing a basic double frequency voltage. The wave-form of the voltage delivered by $V_7$, however, has a high harmonic

content. It is therefore passed through $V_8$ which shapes the wave-form so that it emerges within 5 per cent of a true sine wave.

The operation of this valve is illustrated in Fig. 7.15. Diagram (*a*) shows the wave-form which is normally produced by a full-wave rectifier. In the case of $V_7$ the cathode is raised to a sufficient positive voltage—with resistances $R_8$ and $R_9$—to cause the actual output wave-form to be of the shape shown in Fig. 7.15 (*b*).

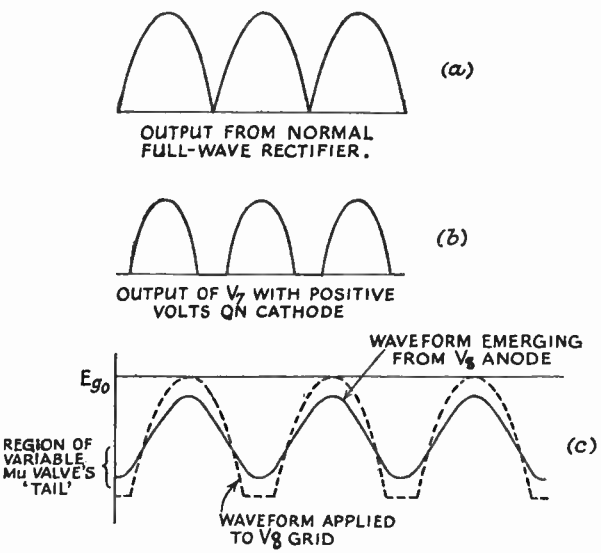



FIG. 7.15.—This diagram illustrates the means by which the audio signal is rectified and corrected to within 5 per cent of a true sine wave-form having double frequency.

This wave-form, which is one step nearer to a sine wave, is applied to the grid of $V_8$, a valve with a variable-*mu* characteristic. Fig. 7.15 (*c*) shows that while the positive half of the wave-form applied to the grid emerges at this valve's anode in a substantially unchanged form, the negative half is distorted by the tailing variable-*mu* characteristic so that it becomes a close approximation to a sine wave. By adjustment of the resistors $R_8$ and $R_9$ the value of the positive bias applied to the cathode of $V_7$ can be set to give the smoothest wave-shape obtainable.

As the output wave-form of $V_8$ has been shifted by 180° in its passage through the valve (by only 90° in terms of the original audio modulating signal), it is in the correct phase to apply as

anode modulation to the carrier amplifier and mixer valve $V_6$. This is done by the simple expedient of taking the anode feed for $V_6$ from a variable resistance in the anode circuit of $V_8$. In this way the phase-modulated radio frequency output voltage from $V_6$ is amplitude modulated with double the original audio frequency, this modulation being applied in the correct phase to substantially cancel out the inherent amplitude modulation which arises as a result of Armstrong's method of producing phase modulation.

As would be expected, the practical results obtained with this system of correction fall a little short of those which are theoretically possible. In a test in which a maximum phase shift of 54° was produced by a 50-cycle modulating signal, the measured results showed a reduction from 21 to 3 per cent distortion. In a second test a phase shift of 32° was also produced by a 50-cycle modulating wave; under these conditions distortion was reduced from $7\frac{1}{2}$ per cent to less than 2 per cent.

The chief advantage in being able to produce increased phase deviation without distortion lies in the fact that the subsequent frequency multiplication required is reduced to half. As has already been explained, after the audio input has been integrated to provide a frequency modulation deviation characteristic which is level over the audio band, Armstrong's modulator produces a maximum frequency deviation of 20 to 25 cycles. If the transmitter is to operate at a maximum deviation of, say, 75 kc/s, these frequency changes will have to be multiplied some 3,000 times. This multiplication can be reduced to 1,500 times if the system of correction indicated above is used. It should, however, be borne in mind that the same result can be obtained by the use of an extra doubler stage in the frequency multiplying chain. It would appear that the only conditions under which the above system of correction would be warranted is in cases where its use would avoid heterodyning the signal to a lower value before multiplying it to the final carrier frequency.

In the case instanced above this would still be necessary; the oscillator frequency might have to be multiplied from 100 kc/s to perhaps 10 Mc/s, at which it could be heterodyned against a crystal oscillator to produce some lower frequency, in the region of 1 or 2 Mc/s, and then multiplied up to, say, 40 to 50 Mc/s. The question of frequency multiplication is discussed in a later section.

**Cathode Ray Frequency Modulator**

In addition to the types of frequency modulator already described, there are a number of other types which for one reason or another have not been widely used, or which have only restricted uses. While it is not proposed to discuss each of these types in any great detail it is felt that an outline of the most interesting will be useful. The first and perhaps the most ingenuous of these is the cathode ray frequency modulator. Like

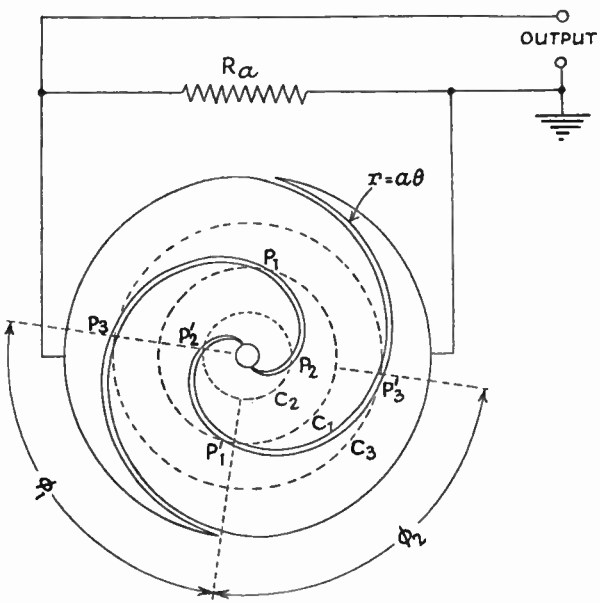



FIG. 7.16.—A typical target anode of a cathode ray frequency modulation generator.

Armstrong's modulator, it also produces phase modulation, but instead of a maximum phase deviation of some 30° to 60°, this method is capable of producing a distortionless phase shift of several times 360°.

This modulator consists of an electrostatically deflected cathode ray tube, with a special target anode. Fig. 7.16 illustrates one form which the target anode may take. It consists of two plates formed by depositing a conducting coating directly upon the inner surface of the "screen end" of a cathode ray tube. The

tube illustrated in Fig. 7.17 is constructed in this way, and it will be noted that in this case the target electrode consists of five complete circles.

In operation the electron stream is deflected in such a way that it traces out a circle on the target anode. Such a trace can be obtained by applying to the two sets of deflector plates two waves which have been derived from the same oscillator, but differ in phase by 90°. In the case under consideration the oscillator is crystal

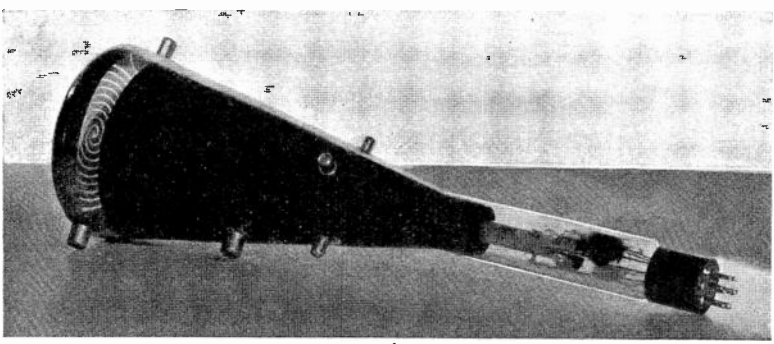


FIG. 7.17.—An R.C.A. cathode ray frequency modulation tube. The target anodes of this tube consist of five complete circles.

controlled in order to ensure a high frequency stability for the final frequency modulated signal. It will be readily seen that if the amplitude of the signal producing the circular trace is varied, then the diameter of the trace itself will also be varied in strict proportion.

If reference is again made to Fig. 7.16 it will be noted that as the beam follows its circular trace it falls first on one anode plate spiral and then on the other. While the beam is drawing current from the anode spiral connected to the right-hand terminal there will be no voltage drop between the target and earth. However, as soon as it commences to draw current from the other target spiral there will be a voltage drop due to the current flowing through the resistance $R_a$. It therefore follows that as the beam traces its circular course a square-wave signal of the same frequency as that of the initial oscillator will be developed in the target anode circuit. This is illustrated in Fig. 7.18, in which the first diagram illustrates the wave-form generated when the beam is scanning the circular trace indicated by the central dotted line

in Fig. 7.16. If now the amplitude of the scanning signal is reduced, then the diameter of the circular trace will also be reduced, as indicated by the inner dotted line in Fig. 7.16. Under these conditions, however, the point at which the beam switches from one anode spiral to the other will be retarded by 90°. Similarly, if the diameter of the circular trace is increased to the

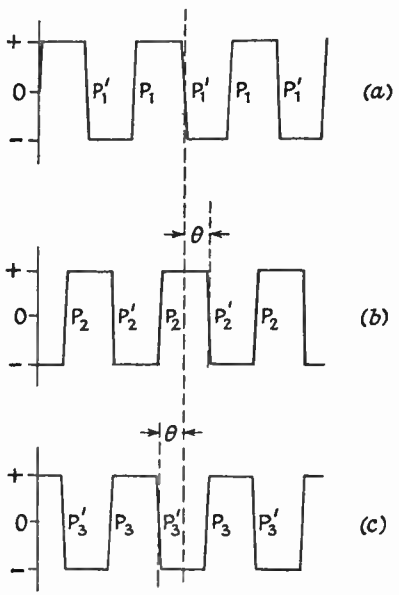



FIG. 7.18.—The wave-form produced at the target anode shown in Fig. 7.16 is illustrated above. (*a*) shows the wave produced when scanning the central dotted line. When the signal amplitude is increased so as to scan the outer dotted line, the phase of the resultant wave is shifted by 90° as shown in (*b*). Similarly if the inner line is being scanned the resultant wave will be shifted in phase by 90° in the other direction as shown in (*c*).

size indicated by the outer dotted line, then the point of change-over will have been advanced by 90°. It therefore follows that by amplitude modulating the signal producing the circular trace —and so causing its diameter to vary—it is possible to phase modulate the square wave-form produced in the cathode ray tube's anode circuit. It follows that with the target illustrated in Fig. 7.16 it would in this way be possible to produce up to 360° phase modulation—without any distortion whatsoever.

Fig. 7.19 illustrates the way in which the cathode ray frequency modulator would be employed in a practical circuit. The two target sections are connected to the two ends of a circuit tuned to the frequency of the crystal controlled oscillator. As the beam changes from one anode to the other it sets the circuit in oscillation, the phase of which is shifted in direct ratio to the change in amplitude of the amplitude modulated scanning signal. The

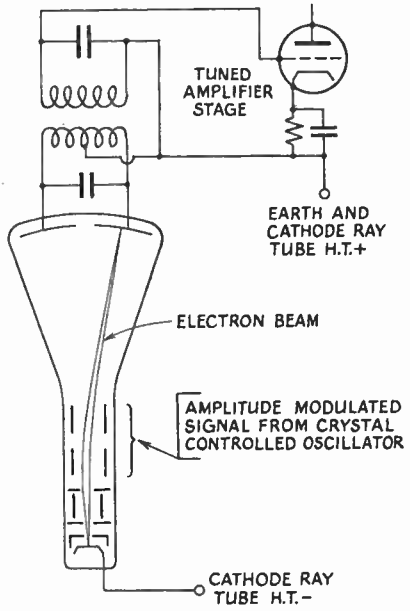



FIG. 7.19.—An outline circuit of the cathode ray frequency modulator.

practical modulator tube illustrated in Fig. 7.17, having five complete turns to its spiral anode, will be capable of producing a maximum phase modulation of $360° \times 5 = 1{,}800°$, some sixty times more than the uncorrected Armstrong modulator.

In an experimental transmitter in which this tube was employed the initial crystal oscillator frequency was some 1·7 Mc/s. The oscillator was followed by a chain of multiplying stages which produced a final frequency of 41 Mc/s. As in the case of Armstrong's modulator, the overall result of frequency modulation was obtained by integrating or pre-distorting the audio signal so that its amplitude falls progressively as its frequency is increased.

### Suppressor Grid Modulator

This type of modulator was described by K. C. Johnson for a "Wobbulator" design; its application as a modulator is, however, obvious. It has the merit of affording a very wide range sweep with good linearity, and freedom from spurious amplitude modulation. A simplified circuit diagram is shown in Fig. 7.20.

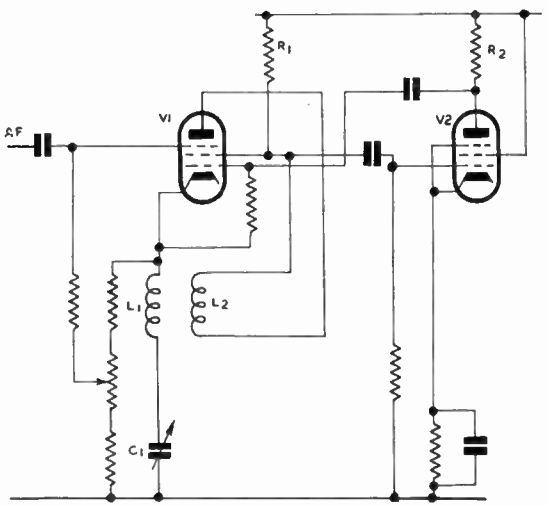



FIG. 7.20.—Modulator employing variation of anode current with suppressor grid bias due to K. C. Johnson.

The circuit relies for its operation upon the fact that, for given anode, screen and control grid potentials, the total cathode current of a pentode is independent of the bias applied to the suppressor grid to a good degree of approximation. The result of applying bias to the suppressor grid is to vary the proportion of this constant cathode current flowing to the anode and to the screen grid respectively.

In the circuit, valves $V_1$ and $V_2$ are cross-connected in a multivibrator-type circuit, the loop gain, ignoring the feedback from $V_1$ cathode circuit, being somewhat greater than unity. $V_1$, for this purpose, operates as a triode, since the load resistor $R_1$ carries both anode and screen currents. In order that $V_1$ shall retain its pentode characteristics $R_1$ must be small so that the screen potential variations are kept to a low value. Since, however, the

loop gain is low, this condition does not require unduly high values of $R_2$. The circuit is constrained to operate at a frequency determined by the series tuned circuit $L_1$, $C_1$ in the cathode of $V_1$; the resistive network in parallel with the tuned circuit serves to provide d.c. continuity to earth, and is of such value as to prevent oscillation in the absence of the tuned circuit.

Coupled to $L_1$ is a second winding $L_2$, which carries the anode current of $V_1$; this current is a fraction $k$ of the cathode current,

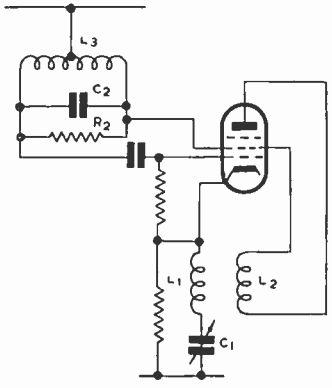



FIG. 7.21.—Single valve version of circuit of Fig. 7.20.

determined by the suppressor grid bias. If the mutual inductance between the two windings is $M$, the e.m.f. induced in $L_1$ by the current in $L_2$ is $j\omega Mki_c$, where $i_c$ is the alternating component of the cathode current.

The e.m.f. across $L_1$ in the absence of $L_2$, is given by $j\omega Li_c$; and hence the e.m.f. developed between the ends of $L_1$, under working conditions, is $j\omega i_c(L_1+kM)$, and thus, in so far as the series tuned circuit is concerned, the value of $L_1$ is altered to an apparent value $L_3$ given by

$$L_3=L_1+kM.$$

As $k$ can be varied with suppressor grid bias (as described above) the resonant frequency of the cathode circuit varies correspondingly. By choice of winding direction the sign of $M$ can be made negative, and $M$ can be made, if necessary, larger than $L_1$. The value of $k$ can be varied between 0 (anode current cut off) and 0·6 approximately (this latter figure, corresponding to zero suppressor grid bias, depends upon valve design). Hence large variations of

$L_3$ can be produced, and very large frequency shifts. The circuit has the property that it is not greatly troubled by spurious amplitude modulation, since the input to $V_2$ is independent of suppressor grid bias.

An alternative single valve circuit is shown in Fig. 7.21; in this the feedback to the grid is supplied by the tuned circuit $L_3C_2$. This circuit is heavily damped by resistor $R_2$, to ensure that the oscillation frequency is governed by $L_1$, $C_1$.

### Condenser Microphone Frequency Modulator

In the case of small portable transmitters of the "walkie-talkie" type, the various forms of modulator already described are rather difficult to accommodate on account of the large number of valves involved. Various alternative circuit types have therefore been proposed, the simplest of which makes possible wide-band frequency modulation with only a single valve transmitter.

The capacity variation of a condenser microphone could, in theory, if connected across an oscillator's tank circuit, produce frequency modulation of the oscillator. This has often been used as an example in making simple explanations of the way in which frequency modulation is produced. Due to the high stray capacities involved in practice, this method is, however, not of any real value. The frequency variations produced would be far too small to be of practical use.

However, if a condenser microphone is used with a simple inductively coupled circuit, the capacity changes can be made to result in very considerable frequency deviations, the extent of which will be determined by the circuit constants and the operating frequency. An outline circuit of such an arrangement is illustrated in Fig. 7.22. The winding $L_p$ is both the oscillator tank circuit inductance and the primary of an r.f. transformer. The winding $L_g$ is merely the oscillator grid feed winding and plays no other part in the circuit operation. The secondary winding $L_s$ is coupled fairly closely into the oscillator tank winding $L_p$, and the capacity $C$ represents that of the condenser microphone. This microphone is directly shunted with an inductance $L_c$, which has such a value that the combination is made parallel resonant in the region of the oscillator's normal operating frequency. The impedance at the terminals of the primary winding of the coupled

circuit, taking into account the effect of the secondary winding, the microphone capacity and shunting inductance, will always be inductive. It will be this inductive reactance in combination with the tuning condenser $C_1$ which determines the frequency of oscillation. However, any variation in capacity of the microphone produces a variation in the impedance coupled into the primary winding, and therefore varies its effective inductance and, with it, the frequency of oscillation.

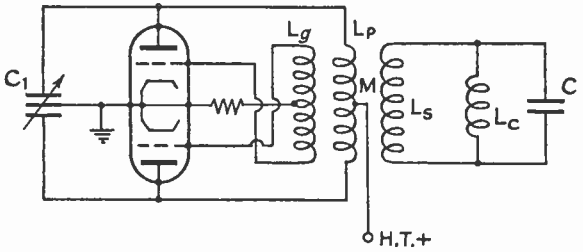



FIG. 7.22.—The outline of a condenser microphone coupled circuit frequency modulator. The condenser microphone is represented by the condenser $C$.

It has been shown by E. J. O'Brien that when the microphone capacity and its shunting inductance are at resonance at the oscillator's operating frequency, that the effective inductance of the primary after an incremental change $\Delta C$ in the microphone capacity is

$$L_e \equiv L_p \pm \omega^2 M^2 \Delta C, \quad . \quad . \quad . \quad . \quad (7.27)$$

where $M$ = the mutual inductance between the primary and secondary windings;
$\omega = 2\pi f$ ($f$ = the frequency of oscillation).

From this equation it is apparent that in order to obtain the greatest change in effective inductance, the mutual inductance and frequency of operation should be as high as possible—without introducing any large direct capacity between the primary and the secondary—and also that the primary inductance should be as small as possible. It should also be noted that the above equation ignores a very small distortion factor $\omega^2 L_c \Delta C$, which should be included to give a precise value for $L_e$. When this factor is included equation (7.27) becomes

$$L_e = L_p \pm \left\{ \frac{M^2 \omega^2 \Delta C}{\pm \omega^2 L_s \Delta C - 1} \right\}. \quad . \quad . \quad . \quad (7.28)$$

When the above formula is expanded to take account of conditions other than resonance it becomes

$$L_e = L_p - \frac{M^2(\omega^2 L_c C - 1 \pm \omega^2 L_c \Delta C)}{L_s(\omega^2 L_c C - 1 \pm \omega^2 L_c \Delta C) - L_c} \quad . \quad (7.29)$$

These equations do not take account of the resistance of the primary and secondary circuits. As, however, any practical circuit must have appreciable resistance, it follows that the peak of the resonance curve will be somewhat flattened. Accordingly, it will be found that the maximum variation in the effective oscillator circuit inductance will be obtained by tuning the parallel combination of microphone capacity and shunting inductance to a frequency somewhat displaced from the normal operating frequency of the oscillator circuit; that is to say, the parallel microphone circuit will be tuned to such a point that the steepest part of its response curve corresponds to the operating frequency of the oscillator.

Arising from the fundamental properties of coupled tuned circuits, it follows that the $\omega^2 M^2 \Delta C$ term will have a greater positive than negative value. It therefore follows that, as in the case of a reactance valve modulator, some distortion must therefore be produced. It will, however, be small, providing that the carrier frequency is high in comparison to the deviation frequency. Again, as in the case of a reactance valve modulator, it will be almost entirely second harmonic distortion. The percentage second harmonic distortion can in this case be simply expressed as

$$\frac{\frac{1}{2}(f_{max} + f_{min}) - f_{carrier}}{(f_{max} - f_{min})} \times 100. \quad . \quad . \quad (7.30)$$

Using the above formula, it can be shown that the basic second harmonic distortion at a carrier frequency of some 40 Mc/s will be in the order of 0·4 per cent with a deviation of $\pm 100$ kc/s.

In the paper by O'Brien in which this type of modulator is described, it is claimed that with a good microphone, high fidelity wide-band frequency modulation has been obtained.

### Variable Resistance Frequency Modulator

The frequency at which a tuned circuit will maintain self-oscillation is termed the natural frequency. The relevant formula

$$\left\{f_n = \frac{1}{2\pi}\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}\right\}$$

indicates that an oscillator's frequency is to some extent dependent upon the circuit resistance, and can therefore be varied by altering it. It should be noted that normally the term $\frac{R^2}{4L^2}$ is negligible compared with $\frac{1}{LC}$, and therefore the error caused in taking the frequency of free oscillation as one and the same thing as a circuit's resonant frequency

$$\left\{f_r = \frac{1}{2\pi}\sqrt{\frac{1}{LC}}\right\}$$

may be neglected. The latter frequency is, of course, that at which the applied voltage will result in the largest circulating current, i.e. that at which forced oscillations will have a maximum amplitude.

It will be apparent that if the circuit resistance component is varied, then the frequency of oscillation will also be changed. This fact provides a further fundamental means of producing frequency modulation. There are, of course, several possible ways in which a variable resistance component could be introduced into an oscillatory circuit. One such way has been described by B. E. Montgomery. He overcomes the various difficulties arising in the application of the resistance change to the oscillator circuit by injecting it into the main tuned circuit by means of inductive coupling.

Referring to the circuit given in Fig. 7.23, it will be noted that

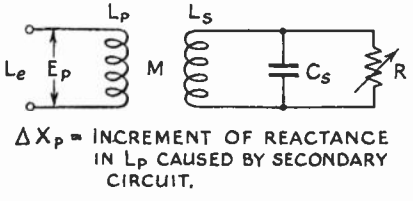



FIG. 7.23.—The outline of the fundamental circuit for inductively coupling a variable resistance into the controlling circuit of a frequency modulator.

there is an inductance, a capacity, and a resistance in parallel and inductively coupled to a second inductance which forms part of the oscillatory circuit of a self-excited oscillator. For this

particular circuit the greatest rate of change in the effective primary inductance $L_e$ is caused by the incremental change of $R$ when its value is

$$R = \pm \frac{X_{L_s} X_{C_s}}{\sqrt{3}\,(X_{L_s} - X_{C_s})} . \quad . \quad . \quad . \quad (7.31)$$

In the above formula the positive sign should be used when $(X_{L_s} - X_{C_s})$ is positive and the negative sign when it is negative —in order that $R$ may at all times be positive.

The actual incremental change $\Delta R_p$, in the primary resistance which is caused by the presence of the secondary circuit, is

$$\Delta R_p = \frac{(\omega M)^2\,(X_{C_s}{}^2 R)}{X_{L_s}{}^2 X_{C_s}{}^2 + R^2 (X_{L_s} - X_{C_s})^2} . \quad . \quad . \quad (7.32)$$

And the incremental change $\Delta X_p$ made to the effective primary reactance by the presence of the secondary is

$$\Delta X_p = \frac{(\omega M)^2 \{ X_{L_s} X_{C_s}{}^2 + R^2 (X_{L_s} - X_{C_s}) \}}{X_{L_s}{}^2 X_{C_s}{}^2 + R^2 (X_{L_s} - X_{C_s})^2} \quad . \quad . \quad (7.33)$$

Assume as a practical example a frequency modulated oscillator of this type operating at 2·5 Mc/s and with the following circuit constants:

$M = 0{\cdot}75$ microhenry or $X_M = 11{\cdot}8$ ohms.
$L_s = 37{\cdot}4$ microhenries or $X_{L_s} = 588$ ohms.
$C_s = 96$ micromicrofarads or $X_{C_s} = 662$ ohms.
$R$ = variable between 0 and 20,000 ohms.

In this case the value of $M$ was determined by assuming that the maximum value of $\Delta R_p$ coupled into $L_p$ by the secondary circuit should be of such a magnitude as to cause the $Q$ of $L_p$ to fall from 200 to an effective value of 100. Fig. 7.24 shows a calculated curve of the variation in $\Delta R_p$ and $\Delta X_p$ when the net reactance of $L_s C_s$ is inductive.

The calculated curves have been confirmed in tests made by Montgomery with the circuit illustrated in Fig. 7.25. His measurements show that a frequency variation as indicated in Fig. 7.26 is obtainable. While it will be noted that a total frequency variation of 24·5 kc/s is actually obtained, it should be pointed out that out of this total variation only some 17 kc/s is suitable

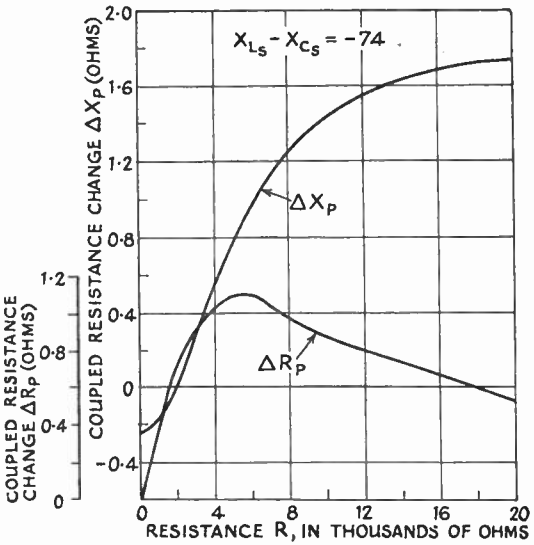



FIG. 7.24.—The above calculated curves, which refer to the coupled resistance type of frequency modulator, show the variations in primary resistance ($\Delta R_p$) and variations in primary reactance ($\Delta X_p$) when the net reactance of the secondary ($L_s$ $C_s$) is inductive.

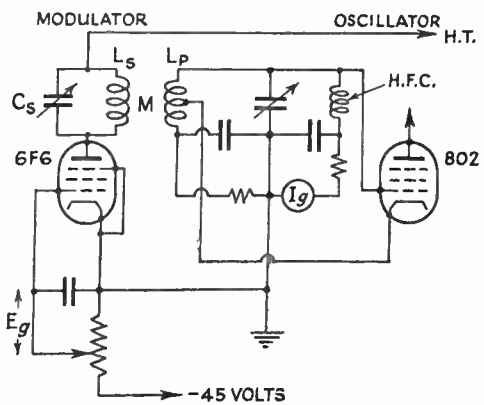



FIG. 7.25.—An experimental coupled variable resistance type of frequency modulator; the self-excited oscillator operates at 2·5 megacycles.

for frequency modulation purposes. If, however, the carrier frequency of 2·5 Mc/s is multiplied up to some 100 Mc/s this yields a deviation of ±350 kc/s—considerably more than is normally employed.

The variation in the oscillator grid current ($I_g Osc.$) as the modulator grid voltage ($E_g$) changes is also shown in Fig. 7.26. This variation in grid current is caused by the resistance coupled into the oscillating circuit by the modulator. As the coupled resistance increases, the amplitude of oscillation decreases, and

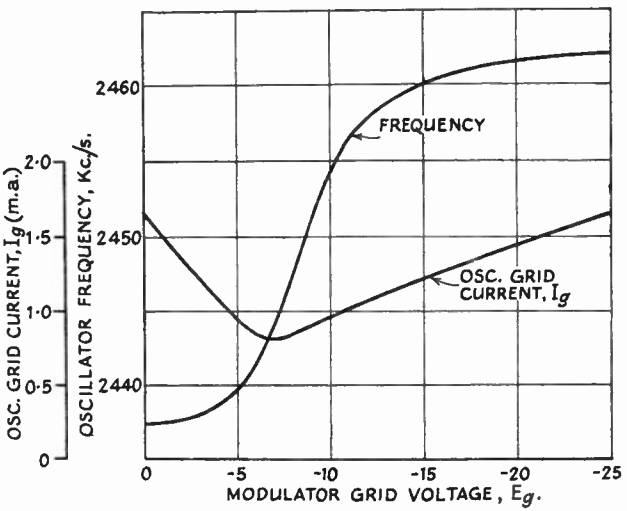



FIG. 7.26.—The variations in the oscillator frequency and grid current obtained with the circuit illustrated in Fig. 7.25.

thus the rectified oscillator grid current falls. This effect produces amplitude modulation of the frequency modulated output; however, the limiting action of the frequency multiplier stages which follow effectively remove this amplitude modulation.

### Balanced Phase Modulators

As an alternative to Armstrong's phase modulator there is a somewhat simpler circuit which may be used in cases where a fair degree of distortion can be tolerated without detrimental results. This circuit, which is used fairly extensively in communication transmitters of the radio-telephone variety, is capable of producing some $\pm 45°$ (i.e. 0·785 radians) phase modulation. This deviation coupled with the fact that such communication systems do not require an audio response going below some 250 cycles, means that a fairly substantial frequency deviation is obtained directly from the modulator. In practice it has been found that speech

intelligibility is not impaired if the audio response starts falling off from some 500 or even more cycles downwards. Taking this figure as that above which there should be a substantially level audio response, it follows that the resultant frequency deviation which can be obtained at 500 cycles is $500 \times 0{\cdot}785 = \pm 393$ cycles.

The maximum frequency swing required for a radio-telephone

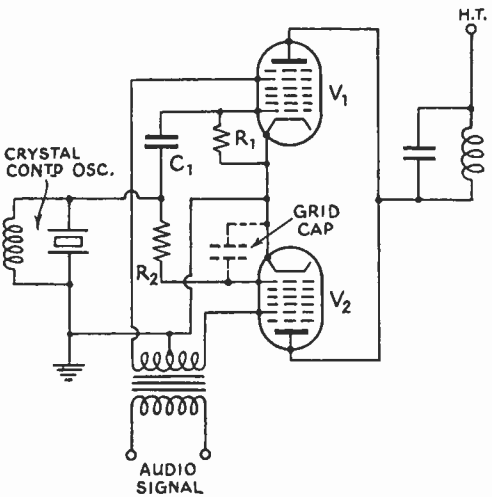



FIG. 7.27.—The basic circuit arrangement of a balanced phase modulator.

system is only some $\pm 15$ kc/s. From this it follows that a subsequent frequency multiplication of some 38 times will in this case be sufficient to produce the final carrier deviation. This order of multiplication may be obtained without any difficulty from three valves. These features coupled with the fact that, like Armstrong's modulator, the balanced phase modulator is also based on a high stability crystal controlled oscillator, explains its popularity for the purpose indicated.

Referring to Fig. 7.27, it will be noted that the signal from a crystal controlled oscillator is applied to the control grids of two heptode valves. The feed circuit to each of these valves consists of a resistance in series with a capacity of comparable reactance. The current flowing through such a circuit will lead the applied voltage by some 45°. In the case of the resistance the voltage will be in phase with the current; it therefore follows that the

grid voltage applied to $V_1$ will lead that of the oscillator by 45°. The voltage across the condenser, however, lags 90° behind the current, with the result that the voltage developed at the grid of the valve $V_2$ will lag 45° behind that across the oscillator circuit.

If the two valves pass equal currents it is apparent from Fig. 7.28 (*a*) that the resultant voltage developed across the common tuned anode circuit will be in phase with that across the oscillator circuit. If, however, the two valves have a push-pull

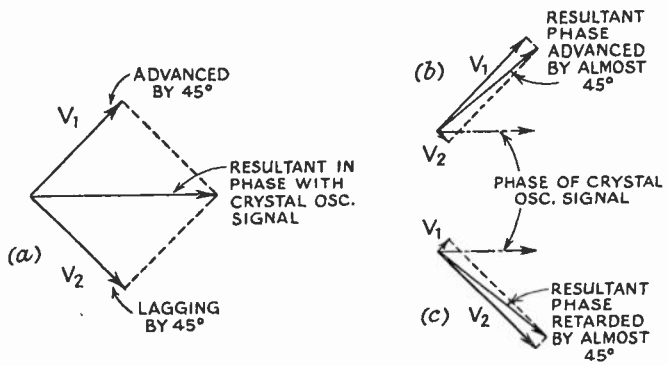



FIG. 7.28.—Vector diagrams of the balanced phase modulator shown in Fig. 7.27.

audio signal applied to their two second control grids, then, as shown in Fig. 7.28 (*b*) and (*c*), the phase of the resultant signal will be alternatively advanced and retarded by 90°. The exact amount of distortion produced by this type of modulator is very largely dependent on the grid voltage/anode current characteristic of the modulator valves used. The desirable features to look for in these valves are a sharp cut-off, followed by a smoothly curved characteristic.

## Frequency Modulation of Resistance Capacity Oscillators

As is discussed in a later chapter, there has been a very considerable increase in the use of frequency modulated sub-carriers superimposed on amplitude modulated transmissions. Such sub-carriers are used for facsimile systems as well as for both ordinary and picture telegraph services. It is normal to use a sub-carrier frequency in the audio band and employ deviations of up to some

$\pm$30 to 40 per cent of the sub-carrier frequency. These deviations are such a large percentage of the mean frequency that, owing to the very high distortion which would result, it is not possible to use a reactance valve or any other normal type of frequency modulator directly. While it is possible to use any of the modulators already discussed by heterodyning their output signals against a stable oscillator, this will in general not be found entirely satisfactory due to the extreme precautions which must be taken against frequency drift, coupled with the difficulty of adjustment.

It is possible to replace the beat-oscillator system by a frequency modulated resistance capacity oscillator and thus obtain directly the large frequency swings required, and at the same time a far higher stability than is possible with the heterodyning system. The frequency of a resistance capacity oscillator is determined by the constants of its phase shift network, and it therefore follows that any change in either a resistance or capacity value will alter the frequency. If one of the resistance elements is replaced with a valve, then this valve can be used to control the oscillator's frequency.

Resistance capacity oscillators fall under two main headings—those having zero phase shift in the coupling network, and those embodying 180° phase-shifting ladder networks. Valves may be used to replace one or more of the resistors in oscillators of either type, although for the purpose under consideration there are definite advantages associated with the use of the latter type. In either case changing any single element will generally alter the network loss which will in turn result in some undesired amplitude modulation as well as the desired frequency modulation. It is much easier to eliminate these amplitude variations in oscillators with ladder phase shift networks than in those incorporating the zero phase shift type. If this latter type is employed, resort will probably have to be made to an automatic volume control circuit which will place a definite limit on the speed of the modulator's response. By proper choice of the circuit constants and operating conditions the amplitude modulation can, however, be reduced to negligible proportions, even with deviations as high as $\pm$40 per cent of the carrier frequency.

Detailed particulars of the theory of this type of modulator, together with its amplitude and harmonic distortion characteristics, have been published in a paper by M. Artzt.

### Frequency Multiplication to Produce the Final Deviation and Carrier Frequency

It has already been noted that the output frequency of most of the modulators described has to be multiplied many times before the final deviation and carrier frequency is obtained. It has been shown that in order to avoid exceeding the maximum distortion which can be permitted in high-fidelity broadcasting, the maximum frequency deviation which can be produced by Armstrong's modulator is only some $\pm 25$ cycles or with distortion correction some $\pm 50$ cycles.

In the case instanced earlier the crystal oscillator frequency was assumed to be 100 kc/s and the final carrier frequency 40 Mc/s. If the oscillator frequency and its variations are simply multiplied up to the desired carrier frequency, the maximum deviation would only be some $400 \times 25 = \pm 10$ kc/s, while the deviation normally required is some $\pm 75$ kc/s. In order to produce this deviation it may therefore be necessary to multiply up to, say, 10 Mc/s, then heterodyne change back to 1 Mc/s before multiplying up to the final carrier frequency. Most other types of frequency modulator produce a greater frequency deviation than Armstrong's, and, therefore, although multiplication is necessary in almost all cases, it will normally be possible to obtain the desired final deviation by simply multiplying the frequency modulated oscillator's signals up to the desired carrier frequency.

It is at this point opportune to note that when a frequency is multiplied, any variations will remain as a fixed proportion of the mean frequency. Similarly, if for any reason it should be divided, the frequency variations again remain as a fixed proportion of the signal frequency. It therefore follows that in the first case the variations—expressed in terms of cycles deviation—are directly multiplied, and in the second divided. However, if the frequency is changed by the heterodyne method, any variations remain constant—only the mean frequency about which these deviations are occurring is changed.

### Frequency Multipliers

The majority of frequency multiplier circuits are based upon a class C amplifier or a series of such amplifiers. As shown in Fig. 7.29, the normal system is to arrange for the grid circuit of a triode valve to be tuned to the oscillator's fundamental frequency

whilst the anode circuit is tuned to the desired (or $n$th) harmonic. The valve is biased well below the cut-off point and driven so that during the peaks the grid is slightly positive, with the result that the anode current is in the form of pulses—as shown in diagram (*b*). The period $\theta$ is the time during which the anode current flows; it is normally expressed in degrees, as a proportion

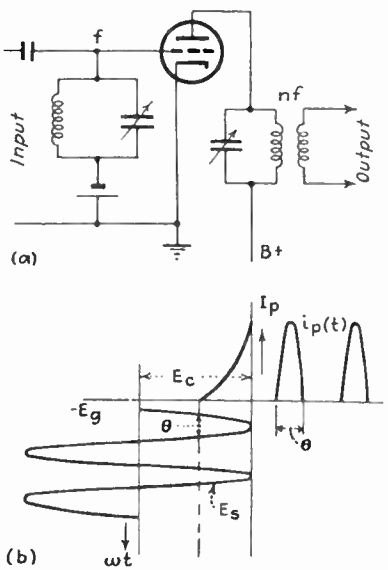



FIG. 7.29.—Diagram (*a*) shows the basic circuit arrangement of a frequency multiplier, while (*b*) shows the voltage and current wave-forms from which it is apparent that there is appreciable distortion in the output.

of the total time or, in other words, as a proportion of 360°. Under these conditions it is termed the "angle of flow" or the "angle of drive". The anode current wave-form $i_a(t)$ has a very high harmonic content, the disposition of which is illustrated in Fig. 7.30, which shows the magnitudes of the d.c. and harmonic components expressed in terms of the peak anode current against the angle of drive. It will be noted that this angle has to be selected fairly carefully if the desired harmonic's output is to have a maximum amplitude. The angle of drive $\theta$ can be reduced by increasing both the negative bias and the signal voltage applied to the grid.

It will be apparent from Fig. 7.30 that the maximum harmonic output occurs when the angle of drive is $\theta = \frac{270^\circ}{n}$, where $n$ is the desired harmonic. It is usually desirable to operate at a somewhat smaller angle of drive as the anode efficiency can be increased by reducing $\theta$ (this is due to the fact that under these conditions the anode current is flowing over a smaller part of the

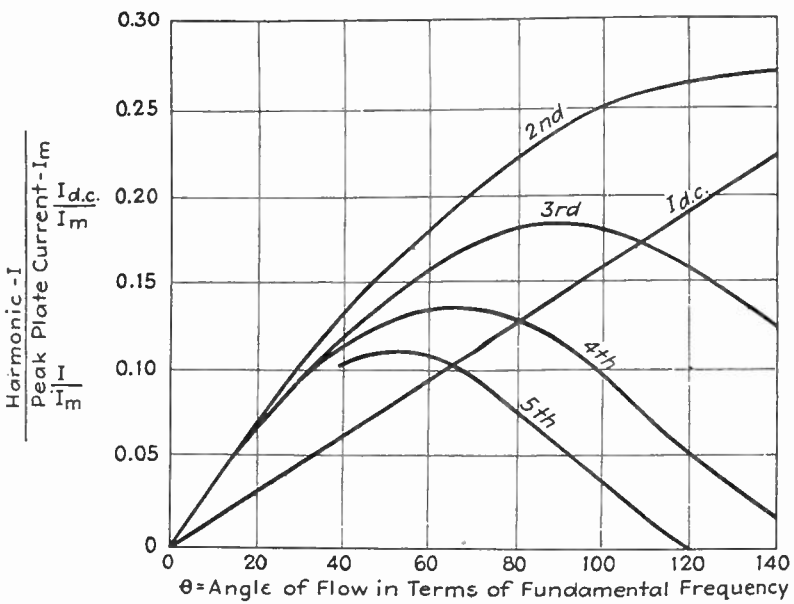



FIG. 7.30.—The magnitudes of the d.c. and harmonic components of a class C amplifier.

(*By courtesy of "Electronics."*)

cycle). In this connection it has been claimed that the most efficient angle of drive is in the region of $\frac{180^\circ}{n}$. In cases where the grid driving power is limited it is, however, permissible to increase the angle of flow up to as much as $\frac{360^\circ}{n}$, although under these conditions the anode efficiency will be considerably reduced. The output from an harmonic generator, as compared with that obtained from the same valve used as a class C amplifier is shown

in Fig. 7.31. The angle of drive for a class C amplifier is taken as 140°.

The two practical circuits which are most commonly used for the production of harmonics are illustrated in Fig. 7.32. The "push-pull" arrangement shown in diagram (*a*) produces odd harmonics, the even ones tending to cancel each other out across the load circuit. In the case of the "push-push" type of frequency multiplier the grids are fed in push-pull while the anodes are connected in parallel. As in the case of a full-wave rectifier, the

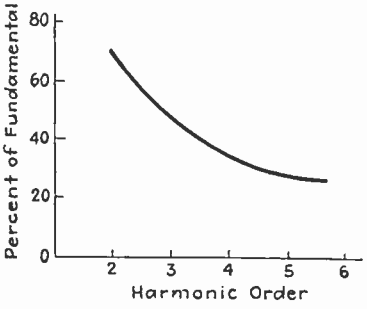



FIG. 7.31.—The output amplitude of an harmonic generator expressed in terms of a percentage of the output at the fundamental frequency.

odd harmonics tend to cancel each other out, so leaving a high proportion of even harmonics in the output.

D. L. Jaffe has given the following information on the design of practical harmonic generators. He suggests that, having selected a valve with a high mutual conductance and a sharp cut-off, the first step is to determine the maximum safe peak anode current $I_m$. This current may be determined from a knowledge of the type of filament and its heating power. The emission current in milliamps per watt of heating power can be taken as approximately 10 for tungsten, 62·5 for thoriated tungsten, and 100 for oxide-coated emitters. However, for the latter two types it is necessary to use factors of safety varying between 3 and 7 for the thoriated tungsten type and at least 10 for the oxide-coated type.

Having determined $I_m$, the corresponding minimum anode voltage and the maximum grid drive voltage can be determined from the valve characteristics. Then the angle of anode current

flow can be determined from the formula $\theta=\frac{180^\circ}{n}$ and, assuming that the anode current follows a 3/2 power law, the ratio $I_{dc}/I_m$ and the harmonic output ratio $I/I_m$ can be determined from Fig. 7.30. Let these values be $K_{dc}$ and $K_h$ respectively, and let $I_{dc}$ be the zero frequency anode current and $I_h$ the harmonic anode current. Then:

$$I_{dc}=K_{dc}I_m, \quad . \quad . \quad . \quad . \quad . \quad (7.34)$$

and

$$I_h=K_hI_m. \quad . \quad . \quad . \quad . \quad . \quad (7.35)$$

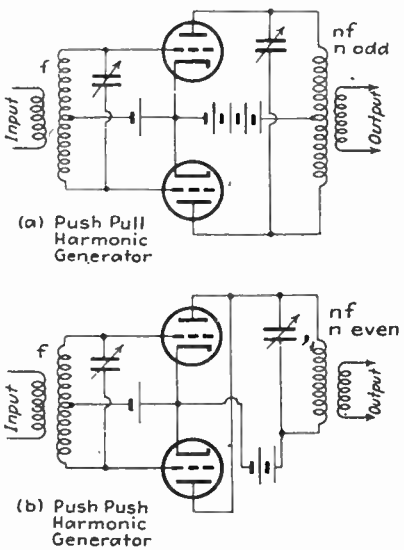



FIG. 7.32.—These two diagrams show the differences in circuit layout between a push-pull and a push-push harmonic generator. The former type produce odd harmonics and the latter even harmonics.

If now the grid current is assumed to be 15 per cent of the total space current, then

$$I_{g\,dc}=0{\cdot}15K_{dc}I_m \quad . \quad . \quad . \quad . \quad (7.36)$$

and

$$I_{a\,dc}=0{\cdot}85K_{dc}I_m, \quad . \quad . \quad . \quad . \quad (7.37)$$

where $I_{g\,dc}$ and $I_{a\,dc}$ are respectively the d.c. grid and anode currents.

The harmonic component of the anode current is

$$I_h=(K_h-0{\cdot}3K_{dc})I_m. \quad . \quad . \quad . \quad . \quad (7.38)$$

The anode input power is

$$P_a = E_B I_{a\,dc} \text{ watts}, \quad . \quad . \quad . \quad . \quad (7.39)$$

where $E_B$ is the anode supply voltage.

The power delivered to the load is

$$P_l = \tfrac{1}{2}(E_B - E_{a\,min}) I_h. \quad . \quad . \quad . \quad . \quad (7.40)$$

The anode efficiency is

$$\eta_a = 100 \times \frac{P_l}{P_a}. \quad . \quad . \quad . \quad . \quad . \quad (7.41)$$

The anode loss is:

$$A_p = P_a - P_l \text{ watts}. \quad . \quad . \quad . \quad . \quad (7.42)$$

The tank circuit impedance is:

$$Z = \left( \frac{E_B - E_{a\,min}}{I_h} \right) \text{ohms}. \quad . \quad . \quad . \quad (7.43)$$

The grid bias can be calculated from Terman's formula

$$E_c = \frac{E_B[1 - \cos \tfrac{1}{2}(n\theta)] + E_{a\,min} \cos \tfrac{1}{2}(n\theta)}{\mu(1 - \cos \tfrac{1}{2}\theta)} + \frac{E_{g\,max} \cos \tfrac{1}{2}\theta}{1 - \cos \tfrac{1}{2}\theta}. \quad . \quad . \quad . \quad (7.44)$$

The grid excitation voltage is

$$E_s = (E_c + E_{g\,max}). \quad . \quad . \quad . \quad . \quad (7.45)$$

And, finally, the grid driving power is

$$P_g \equiv E_s I_{g\,dc}. \quad . \quad . \quad . \quad . \quad . \quad (7.46)$$

## Frequency Modulation Transmitters

Those sections of the frequency modulation transmitter which differ materially from normal amplitude modulation transmitter practice have now been described, and it is now, therefore, intended to outline representative commercial transmitter practice. A review of the various designs shows that reactance modulators are widely employed although in later equipments special circuits, such as the phasitron, have been used. Two of the transmitters which have been selected for description are indicative of broadcast transmitter practice, and the other of communication transmitter practice.

Any frequency modulation transmitter can be divided into three fundamental sections: the first is concerned with the generation of the frequency modulated signal, the second with the multiplication of that signal, and the third with its power amplification. There is very little fundamental difference between the last two sections of the various manufacturers' transmitters—the arrangement of the power output sections only differ materially when larger or smaller power outputs are required. The usual arrangement is to have a range of standard transmitter power amplifier output sections, starting perhaps with a 250-W section which can either be employed as the final output stage in a small transmitter or as the driver stage in a transmitter with a power output of say 3 kW. In turn the 3-kW output stage can also be employed as the driver stage for a 50-kW transmitter. The various different manufacturers will obviously follow power progressions of their own choice rather than that suggested above.

### RCA BTF.3B Transmitter

This transmitter is capable of a power output of 3 kW at any frequency between 88 and 108 Mc/s; it incorporates its own

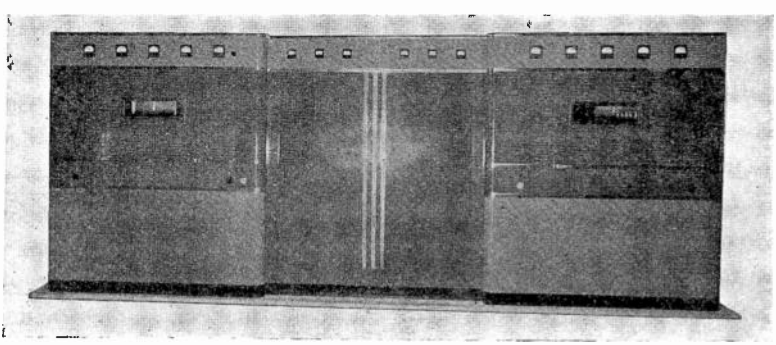


FIG. 7.33.—RCA transmitter type BTF.3B.
*(By courtesy of RCA.)*

modulator giving an f.m. output at half the final carrier frequency. The transmitter is housed in a three-bay cabinet, as shown in Fig. 7.33.

The technical summary of the transmitter is as follows:

| | |
|---|---|
| Frequency range | 88 to 108 Mc/s |
| Power output (into transmission line) | 1,000 to 3,000 W |

| | |
|---|---|
| Output impedance | 51·5 ohms (standing wave ratio 1·75 to 1 or less) |
| Carrier frequency stability, deviation less than | 1,000 c/s |
| Modulation capability | ±100 kc/s |
| Method of modulation | Reactance tubes |
| Audio input impedance | 600/150 ohms |
| 100 per cent modulation level* | +10±2 db |
| Audio frequency response† 30 to 15,000 c/s (1,000 c/s reference) | flat within ±1 db |
| Audio frequency distortion‡ 30 to 15,000 c/s (including all harmonics up to 30 kc/s at 75 kc/s swing) | not more than 1·0 per cent |
| FM noise level (reference ±75 k/cs swing‡) | not more than 65 db |
| AM noise level (reference 100 per cent amplitude modulation‡) | not more than 50 db |
| Power line requirements—transmitter | |
| Line voltage | 230/208 V |
| Phase | 3 |
| Frequency | 50 or 60 c/s |
| Line regulation (maximum) | 5 per cent |
| Power consumption (approximate) | 7·7 kW |
| Power factor (approximate) | 90 per cent |
| Power line requirements—crystal heaters | |
| Line voltage | 100 to 130 V a.c. or d.c. |
| Power consumption | 28 W |

The essential circuit features of the transmitter and its modulator are shown in Fig. 7.34 and Fig. 7.35. In Fig. 7.34 is shown the modulator, together with the centre frequency control circuit. Push-pull reactance valves $V_1$ and $V_2$ are used to modulate the Hartley type oscillator ($V_3$). An r.f. input, phase shifted by 90° from the oscillator tank circuit voltage, is applied in anti-phase to the grids of the reactance valves. The a.f. input is also applied to the valves in push-pull, to vary the mutual conductance differentially, and hence achieve modulation.

The oscillator is arranged to operate at the eighteenth subharmonic of the carrier frequency, and to achieve the desired final

* Level at input of 600 ohms pre-emphasis network. Insertion loss of this network is approximately 24 db.

† Audio frequency response is referred to a standard 75 micro-second curve when measured using pre-emphasis.

‡ Distortion and noise are measured following a standard 75 micro-second de-emphasis network.

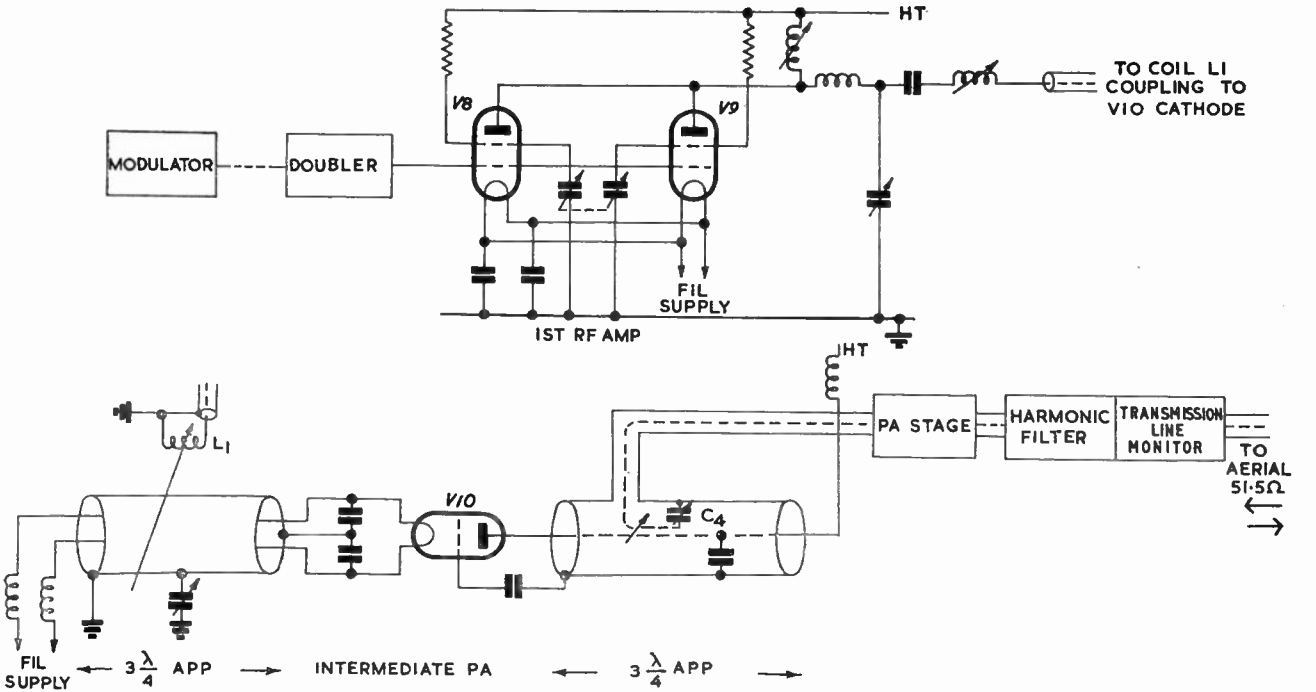



FIG. 7.35.—Simplified circuit diagram of r.f. stage of the RCA transmitter type BTF.3B.

range is tunable from 4·9 to 6·0 Mc/s. The output from the oscillator is multiplied nine times to give an output to the transmitter proper at half the final carrier frequency.

The automatic centre frequency control is achieved by comparing the oscillator frequency with that of a crystal controlled

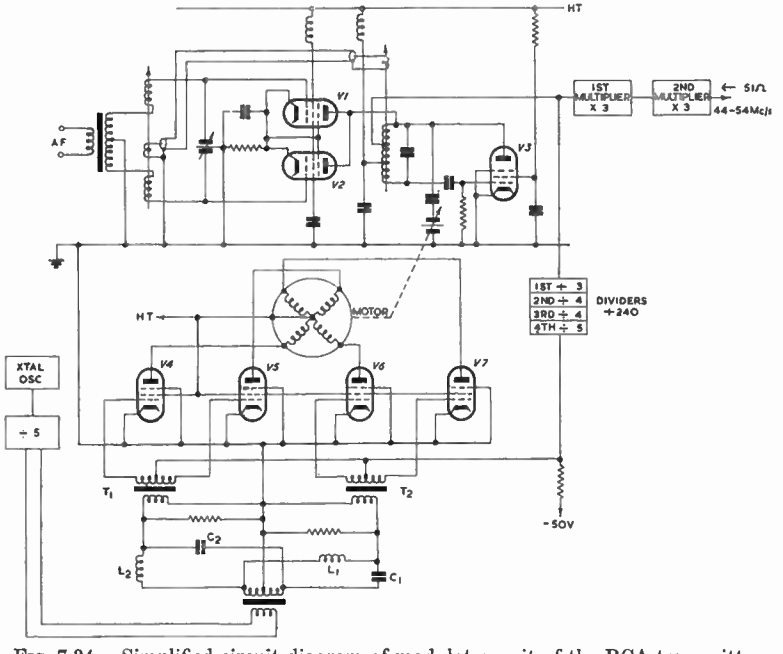



FIG. 7.34.—Simplified circuit diagram of modulator unit of the RCA transmitter type BTF.3B.

oscillator. The modulator output frequency is divided by 240, by a chain of locked-in oscillator dividers, and hence provides an output signal having a frequency in the region of 25 kc/s. The ratio of the modulated oscillator frequency to that of the crystal is 48:1; the output of the crystal oscillator is divided by 5 to reduce the reference frequency to that of the compared signal. The comparison circuit is essentially similar to that of a phase difference discriminator. The compared signal input is applied in-phase to the grids of the pairs of valves $V_4$, $V_5$ and $V_6$, $V_7$. The crystal controlled reference signal is applied in push-pull, and a phasing network $L_1$, $C_1$, $L_2$, $C_2$ is employed, to ensure that the input to transformer $T_1$ is 90° out of phase with that to $T_2$.

The voltage at each grid is thus the sum of the negative bias, which fixes the operating point near cut-off, and the vector sum of the reference and compared signal voltages. Since each valve is heavily biased, a charge of the magnitude of the latter signal voltage alters appreciably the d.c. component of the anode current. Any difference in frequency between reference and compared signals thus results in a rotation of the magnetic field in the two-phase control motor. This causes the motor shaft to rotate in the direction necessary for the capacitor $C_1$ to re-establish the correct frequency.

The transmitter power stages comprise a doubler followed by the first r.f. power amplifier. This uses two tetrodes in parallel ($V8$ and $V9$); neutralising is effected by series-tuning the screen grid lead inductances by a twin-gang capacitor. The anode load of this stage comprises a **T**-section matching the output to the co-axial line feeding the next stage.

The intermediate P.A. stage $V10$ comprises an earthed-grid triode. The heater leads are a.c. coupled to an open wire unbalanced line of approximate length $3\lambda/4$, and run inside this line to the earthy end. This arrangement ensures that the heater supply leads are free of r.f. and at earth potential. The cathode line is tuned by capacitor $C_2$; the output from the first r.f. amplifier is fed to the cathode circuit by the inductor $L_1$, which effects magnetic coupling to the cathode line.

The valve is mounted inside the inner conductor of the anode tuned circuit, which comprises a co-axial line of length $3\lambda/4$ approximately; the inner conductor provides a duct for the valve forced air-cooling supply. The load is tuned to resonance by altering the position of the capacitor $C_3$ which acts as an r.f. short circuit. The end of the inner conductor remote from the anode is thus at earth potential, and the h.t. supply is introduced at this point.

The output is taken by a coupling loop inside the co-axial line; the spacing of this loop from the inner conductor can be varied to alter the degree of coupling. The capacitor $C_4$ is introduced to resonate with the inductance of the pick-up loop, so that the output impedance is resistive.

The final P.A. stage is identical in circuit arrangement with the intermediate power amplifier shown. The coupling loop here is connected to the output feeder, which is of 51·5 ohms characteristic impedance. A harmonic filter and transmission line monitor

are inserted in the feeder; this latter circuit removes the h.t. supply from the final P.A. stage if the standing wave ratio exceeds a pre-set value.

### Marconi BD.306 Transmitter

This transmitter is capable of a power output of 10 kW at any frequency in the range 88–108 Mc/s. The modulator circuit is the

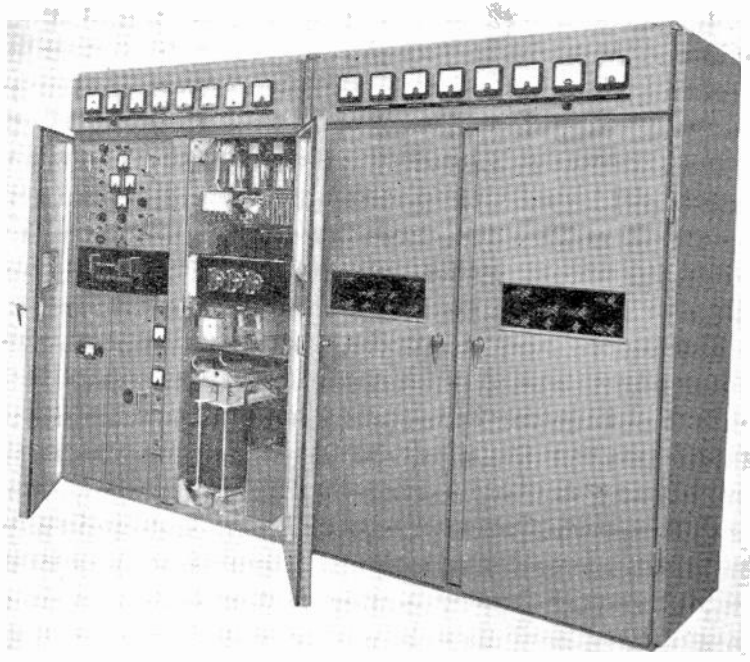


FIG. 7.36.—Marconi transmitter type BD.306
*(By courtesy of Marconi's Wireless Telegraph Co.)*

FMQ type, the essential features of which were discussed earlier. The transmitter is housed in a four-bay cabinet with doors front and rear; a view of the complete transmitter is shown in Fig. 7.36.

The technical summary of the transmitter is as follows:

| | |
|---|---|
| Power rating | 10 kW |
| Frequency range | 88–108 Mc/s |
| Frequency stability | ±0·002 per cent |
| Frequency deviation | ±75 kc/s |

| | |
|---|---|
| FM noise level | 65 db below the level corresponding to ±75 kc/s deviation |
| AM noise level | 50 db below the level corresponding to 100 per cent amplitude modulation |
| AF distortion | Less than 1·5 per cent from 30 to 100 c/s.<br>Less than 1·0 per cent from 100 to 15,000 c/s for ±75 kc/s deviation |
| Audio pre-emphasis | 75 micro-second network |
| AF response | ±1 db from 30 to 15,000 c/s measured at the output of a standard de-emphasis network. Reference level 400 c/s. |
| AF input level | 10 (±2) db in 600 ohms balanced for ±75 kc/s deviation |
| Output impedance | 51·5 ohms unbalanced |
| Power supply | 380–440 V, 40–60 c/s, three-phase, four-wire a.c. mains |
| Power consumption | 24 kW at 0·9 power factor |

| Dimensions | Height | Width | Depth | Weight |
|---|---|---|---|---|
| | 7 ft (213 cm) | 10 ft (305 cm) | 2 ft 6 in (76 cm) | 3,500 lb (1,589 kg) |

The essential features of the transmitter circuit are shown in Fig. 7.37. The output from the FMQ modulator is at carrier frequency, and is applied in push-pull to the double pentodes forming the first r.f. stage $V_1$. The anode circuit comprises a tuned short-circuited line, the position of the shorting bar providing coarse adjustment of the tuning, and the setting of the capacitor fine adjustment. The output from this stage is fed to the grids of the tetrodes $V_2$ and $V_3$, which together form the second r.f. stage. The anode load of this stage also comprises a tuned short-circuited line, which is inductively coupled to a second tuned line which feeds the unbalanced co-axial feeder to the next stage.

The third r.f. stage $V_4$ is of the earthed-grid type. The output from the second r.f. stage is fed to the inner conductor of the cathode co-axial line tuned circuit; this inner conductor comprises the two filament leads to the valves, effectively in parallel at r.f. The circuit is tuned by the ganged capacitors $C_1$, $C_2$; the capacitors $C_3$, $C_4$ adjust the input resistance presented to the feeder by altering the effective length of the cathode line.

The valve $V_4$ is situated inside the inner conductor of the anode co-axial line tuned circuit; this inner conductor provides a duct for the cooling air. The line is tuned by adjustment of the position

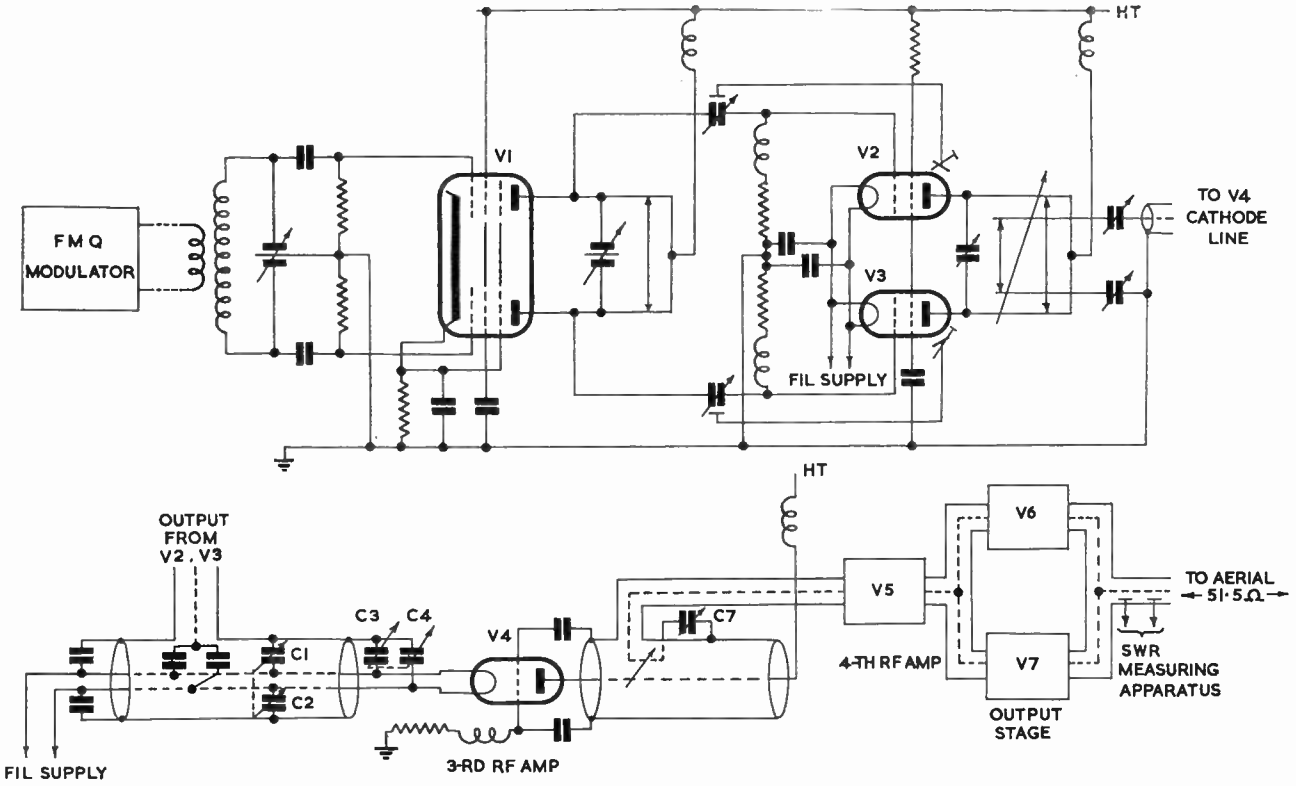



FIG. 7.37.—Simplified circuit diagram of r.f. stage of the Marconi transmitter type B.D. 306.

of the r.f. short circuiting capacitors $C_5$, $C_6$. The output is taken by a pick-up loop; the spacing of the loop from the inner conductor can be varied to secure maximum power output. The reactance of the pick-up loop is resonated by the capacitor $C_7$ to ensure that the output impedance of the stage is resistive.

The fourth r.f. amplifier ($V_5$) is identical in circuit arrangement with the third. Its output is split into two, to feed the two valves ($V_6$, $V_7$) of the output stage in parallel. The circuit of each of the latter valves individually is again identical with that of $V_3$.

### The Link Type 50-U.F.S. Frequency Modulation Transmitter

This equipment has been selected as typical of the smaller types of communication transmitter. It is normally combined to form a complete transmitter-receiver station, as illustrated in Fig. 7.38. The transmitter in this photograph is the upper of the three chassis, those below being respectively the receiver and the power unit.

The technical summary of the performance of the Type 50-U.F.S. Frequency Modulation Transmitter is as follows:

| | |
|---|---|
| Power Output | 50 watts (nominal) |
| Frequency Range | 30 to 40 Mc/s |
| Frequency Deviation | ±15 kc/s |
| Audio Frequency Range | 300 to 3,000 cycles with high frequency pre-emphasis |
| Power Input (whole equipment) | Stand-by (receiver only), 125 W Transmitting 320 W—from 115 V a.c |
| Output Impedance | Any—usually fed into concentric line |

In essence, the circuit of this transmitter consists of a crystal controlled oscillator ($V_1$—see Fig. 7.39), which is followed by a balanced phase modulator and three stages of frequency multiplication, giving a total of thirty-two times increase in frequency. The third of these multiplication stages ($V_6$) feeds directly into the power output stage which consists of two 807 valves in parallel.

Going through the circuit in greater detail, it will be noted that the crystal oscillator, $V_1$, is a pentode receiving valve connected as a triode. As the output frequency range lies between 30 Mc/s. and 40 Mc/s, the crystal frequency will lie between 937·5 kc/s and 1,250 kc/s.

The crystal oscillator utilises a resistance coupled circuit in order to avoid the necessity of oscillator tuning when changing frequency. The crystal is connected between the grid and anode of the oscillator valve and $R_2$ acts as the anode circuit load.

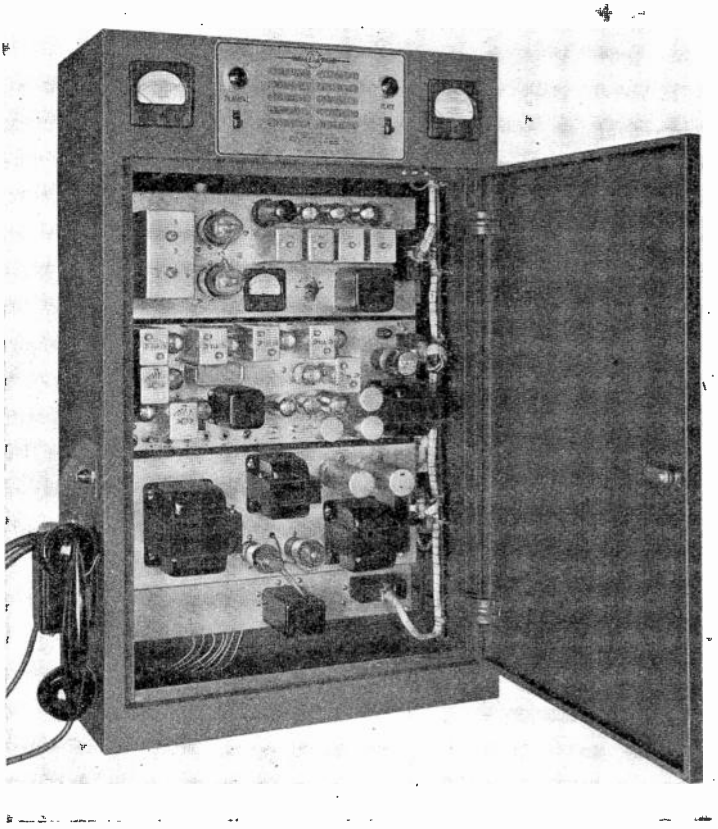


FIG. 7.38.—The Link Type 50–U.F.S. frequency modulation transmitter-receiver. The transmitter chassis is uppermost and delivers a nominal 50 W output.

The injection grids of the two balanced phase modulator valves $V_2$ and $V_3$ are driven from the plate of the oscillator through phase-shifting networks designed to advance the phase of one grid by approximately 45° and to retard the phase of the other by approximately 45°. The anode currents of $V_2$ and $V_3$ will therefore be about 90° out of phase and are so proportioned as

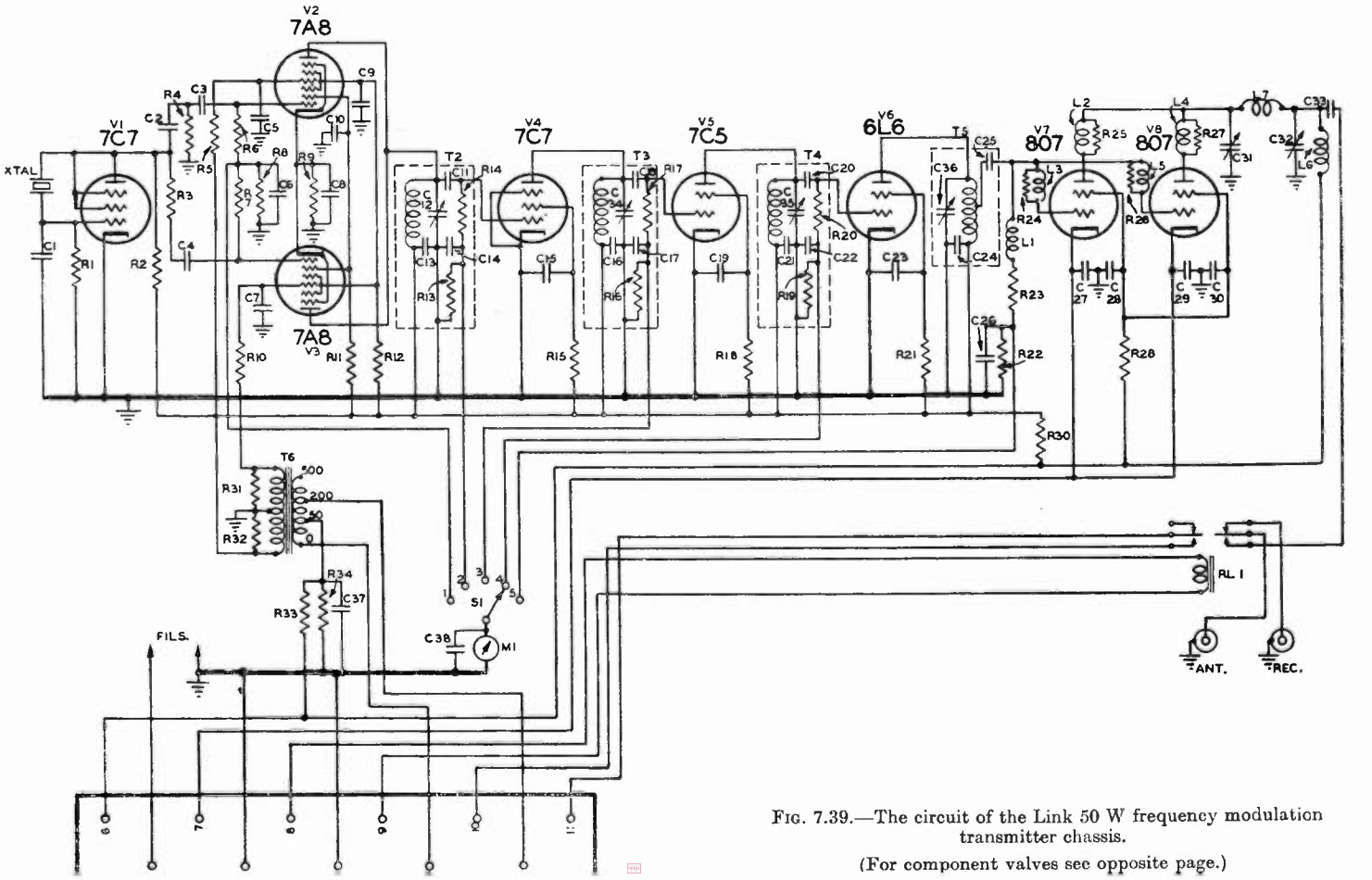



FIG. 7.39.—The circuit of the Link 50 W frequency modulation transmitter chassis.

(For component valves see opposite page.)

# LIST OF COMPONENTS USED IN LINK TYPE 50-U.F.S. TRANSMITTER

C 1—150 pfd. mica
C 2—10 pfd. mica
C 3—100 pfd. mica
C 4—100 pfd. mica
C 5—0·002 mfd. mica
C 6—0·05 mfd. 600 V. block
C 7—0·002 mfd. mica
C 8—0·05 mfd. 600 V. block
C 9—0·05 mfd. 600 V. block
C10—0·05 mfd. 600 V. block
C11—100 mmfd. mica
C12—100 mmfd. variable
C13—0·05 mfd. paper 600 V.
C14—0·002 mfd. mica
C15—0·002 mfd. mica
C16—0·002 mfd. mica
C17—0·002 mfd. mica
C18—100 mmfd. mica
C19—0·002 mfd. mica
C20—100 pfd. mica
C21—0·002 mfd. mica
C22—0·002 mfd. mica
C23—0·002 mfd. mica
C24—0·002 mfd. mica
C25—0·002 mfd. mica
C26—0·002 mfd. mica
C27—0·002 mfd. mica
C28—0·002 mfd. mica
C29—0·002 mfd. mica
C30—0·002 mfd. mica
C31—25 pfd. variable
C32—140 pfd. variable
C33—0·002 mfd. mica
C34—44 pfd. variable
C35—44 pfd. variable
C36—44 pfd. variable
C37—25 mfd. 50 V electrolytic
C38—0·002 mfd. mica

T2—Amplifier plate tank
T3—Multiplier plate tank
T4—Multiplier plate tank
T5—Doubler plate tank
T6—Audio trans.

S1—S.P. 5 Pos. Metering Switch
M1—0-5 ma. meter

R 1—0·5 megohm ½ W
R 2—50 kilohms 1 W
R 3—20 kilohms 1 W
R 4—20 kilohms 1 W
R 5—50 kilohms ½ W
R 6—50 kilohms 1 W
R 7—50 kilohms 1 W
R 8—1,000 ohms 1 W
R 9—250 ohms 1 W
R10—50 kilohms ½ W
R11—50 kilohms 1 W
R12—50 kilohms 1 W
R13—1,000 ohms 1 W
R14—0·25 megohm ½ W
R15—100 kilohms 1 W
R16—1,000 ohms 1 W
R17—0·25 megohm ½ W
R18—50 kilohms 1 W
R19—1,000 ohms 1 W
R20—0·25 megohm ½ W
R21—50 kilohms 1 W
R22—1,000 ohms 1 W
R23—10 kilohms 1 W
R24—100 ohms ½ W
R25—100 ohms ½ W
R26—100 ohms ½ W
R27—100 ohms ½ W
R28—7,500 ohms 10 W
R30—3,000 ohms 25 W
R31—25 kilohms ½ W
R32—25 kilohms ½ W
R33—40 kilohms 10 W
R34—1,000 ohms 1 W

L1—2·5 mH. choke
L2—Paras. suppressor choke
L3—Paras. suppressor choke
L4—Paras. suppressor choke
L5—Paras. suppressor choke
L6—2·5 mH. choke
L7—R.F. Tank inductance

RL1—Aerial Relay

to be equal in magnitude. The two currents add vectorally to produce a resultant phase-modulated voltage across $T2$.

The control grids of the balanced modulators $V_2$ and $V_3$ are connected to the secondary of the push-pull audio transformer $T_6$. This transformer is driven directly from the microphone, which derives its current from the voltage divider and filter network $R_{33}$, $R_{34}$, and $C_{37}$.

The modulator grids are fed through the frequency correction networks $R_{10}$, $C_7$, and $R_5$, $C_5$. These $RC$ combinations attenuate the audio frequency range (above 2,000 cycles), so the excessive frequency deviation is not obtained. Resistors $R_{31}$ and $R_{32}$ are terminating resistors for the secondary of the microphone transformer $T_6$. As the audio voltages are applied in push-pull to the control grids of the modulators $V_2$ and $V_3$, their plate currents vary about mean values, and as one increases the other decreases. The resultant current and voltage in $T2$ also varies in phase with these changes. In addition to the frequency modulation resulting from these phase variations there is also a small amount of amplitude modulation; this is, however, removed by the limiting action of the subsequent frequency multiplication stages.

To obtain sufficient deviation ($\pm 15$ kc/s) the frequency of the modulated wave is multiplied thirty-two times. This is accomplished by two quadruplers ($V_4$ and $V_5$), followed by a doubler ($V_6$). All three valves act as grid leak biased Class C amplifiers. The grid drive in each case is well above saturation so that slight changes in tuning or reductions in valve emission can have little effect on succeeding stages. Up to this point all stages have employed receiving valves working at relatively low anode and heater currents. The final power amplifier stage, however, utilises two 807 beam valves in a parallel Class C amplifier circuit. Grid leak bias is used and, as in the preceding stages, provision is made for the metering of grid current. Finally, the anode tank aerial circuits are of the *Pi* type in order to secure a high harmonic suppression ratio and also ease of adjustment.

## SELECTED REFERENCES


TERMAN, F. E., and FERNS, J. H., The Calculation of Class C Amplifier and Harmonic Generator Performance, *Proc. I.R.E.*, March 1934.

TERMAN, F. E., and ROOKE, W. C., Calculation of Class C Amplifiers, *Proc. I.R.E.*, April 1936.

JAFFE, D. L., Armstrong's Frequency Modulator, *Proc. I.R.E.*, April 1938.

SHELBY, R. E., A Cathode-Ray Frequency Modulation Generator, *Electronics*, February 1940.

SHEAFFER, C. F., Frequency Modulator, *Proc. I.R.E.*, February 1940.

CROSBY, M. G., Reactance-Tube Frequency Modulators, *R.C.A. Review*, July 1940.

MORRISON, J. F., A New Broadcast Transmitter Circuit Design for Frequency Modulation, *Proc. I.R.E.*, October 1940.

WINLUND, E. S., Drift Analysis of the Crosby Frequency Modulation Transmitter Circuit, *Proc. I.R.E.*, July 1941.

THOMAS, H. P., and WILLIAMSON, R. H., A Commercial 50 Kilowatt Frequency Modulation Broadcast Station, *Proc. I.R.E.*, October 1941.

MONTGOMERY, BRUCE E., An Inductively Coupled Frequency Modulator, *Proc. I.R.E.*, October, 1941.

PIERACCI, ROGER J., A Stabilized Frequency-Modulator System, *Proc. I.R.E.*, February 1942.

A Modern 10 kW. Frequency Modulation Transmitter, *Electronics*, March 1942.

JAFFE, D. L., Wide-Band Amplifiers and Frequency Multiplication, *Electronics*, April 1942.

DUENO, B., F.M. Carrier Current Telephony, *Electronics*, May 1942.

Pennsylvania Turnpike U.H.F. Traffic Control System, *Electronics*, May 1942.

SKENE, A. A., and OLMSTEAD, N. C., A New Frequency Modulation Broadcast Transmitter, *Proc. I.R.E.*, July 1942.

HUND, AUGUST, Reactance Tubes in F.M. Applications, *Electronics*, October 1942.

CHANG, C. K., A Frequency Modulation Resistance Capitance Oscillator, *Proc. I.R.E.*, January 1943.

BERTRAM, S., Correction of F.M. Distortion, *Proc. I.R.E.*, April 1943, p. 186.

GOETTER, W. F., Frequency Modulation Transmitter and Receiver for Studio to Transmitter Relay System, *Proc. I.R.E.*, November 1943.

O'BRIEN, ELWIN J., A Coupled Circuit Frequency Modulator, *Proc. I.R.E.*, June 1944.

ARTZT, MAURICE, Frequency Modulation of Resistance-Capacity Oscillators, *Proc. I.R.E.*, July 1944.

F.M. Carrier Telephony for 230 kV. Lines, *Electronics*, December 1944.

STURLEY, K. R., Frequency Modulation, *Journal I.E.E.*, Part III, 1945.

BAILEY, F. M., and THOMAS, H. P., Phasitron F.M. Transmitter, *Electronics*, October 1946.

ADLER, ROBERT, A New System of Frequency Modulation (The Phasitron), *Proc. I.R.E.*, January 1947.

BRADFORD, H. K., Wide-Angle Phase Modulator, *Electronics*, February 1947.

MORTLEY, W. S., F.M.Q., *Wireless World*, October 1951.

FLANAGAN, T. P., Spurious A.M. in F.M. Signal Generators, *Marconi Instrumentation*, December 1953.

*Chapter Eight*

# LIMITERS AND DISCRIMINATORS

It has already been shown that it is necessary to suppress the incoming signal's amplitude variations. It has also been shown that a very large part of the improvement in signal to noise ratio can be ascribed to this.

All the earlier theoretical discussions on the improvement in signal to noise ratio were based on the assumption that ideal limiting is employed. The desirable property of the limiting circuit is that it should eliminate from the output any variations arising from alterations of the amplitude of the input signal. It should (*a*) function for all levels of input signal, and (*b*) its action should be independent of the rate of alteration of signal amplitude. All the practical limiting circuits fall short of the ideal with respect to (*a*) at low signal levels. With respect to (*b*), the circuit usually falls short of the ideal in its ability to handle very rapid or very slow variations of signal amplitude.

## Grid Limiters

A Foster-Seeley discriminator is generally preceded by a grid limiter, and the circuit of a typical grid limiter is shown in Fig. 8.1.

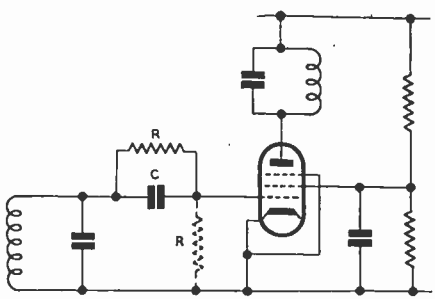



Fig. 8.1.—General form of grid limiter stage.

The valve employed is a pentode operated with a low screen voltage, usually of the order of 50 volts; the screen is generally fed from a potential divider to ensure relative constancy of screen potential under working conditions. The stage is biased by grid

current rectification. Thus as the amplitude of the signal at the grid is increased from zero, the grid bias increases and, since the screen is at a low potential, the standing bias exceeds the cut-off value when the input signal amplitude is of the order of 2 volts. With further increase of signal, the valve is operated under Class C

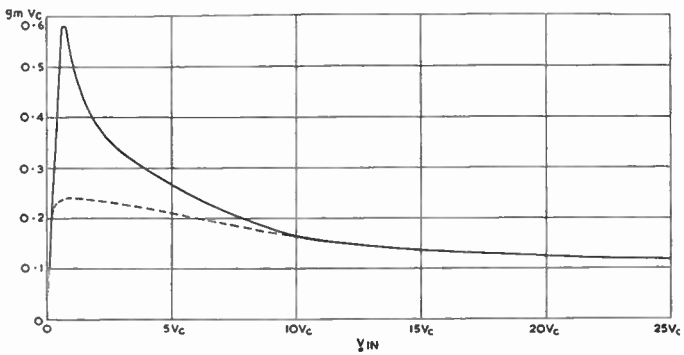



FIG. 8.2.—Fundamental frequency component of anode current of grid leak limiter, for "ideal" $i_a - v_g$ characteristic; practical characteristic shown dotted.

amplifier conditions. The anode current then has a very distorted wave-form, comprising pulses of constant amplitude, with decreasing duration as the signal amplitude increases. The component of the anode current at the fundamental driving frequency

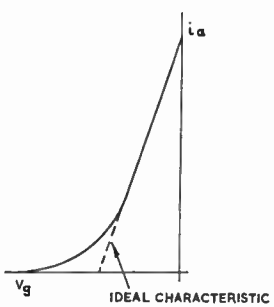



FIG. 8.3.—Practical and "ideal" $i_a - v_g$ characteristics.

varies but slowly with increasing signal amplitude, and it is this phenomenon which provides the limiting action of the stage.

In Fig. 8.2 is shown the fundamental frequency component of the anode current plotted against input signal amplitude for an

"ideal" valve, one in which the $i_a$—$v_g$ characteristic is linear from zero bias to the cut-off bias $V_c$ (see Fig. 8.3). From this figure it will be seen that the limiting action leaves much to be desired. Above the threshold where limiting may be said to commence ($0{\cdot}6V_c$ approximately), the output falls rapidly at first and then more gently. Provided that operation is confined to inputs substantially in excess of $10V_c$, limiting is fairly satisfactory.

In practice, the slope of the $i_a$—$v_g$ is usually relatively linear over a small range near zero bias, with increasing curvature towards cut-off as shown in Fig. 8.3. This appreciably alters the shape of the anode current pulses, with the result that the type of limiting

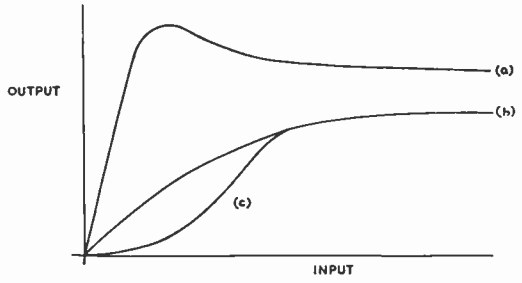



FIG. 8.4.—Limiting characteristic of grid leak limiters: grid resistor returned to (*a*), source of positive bias, (*b*) earth, (*c*) negative bias.

characteristic obtained in practice is more nearly that of the dotted curve of Fig. 8.2, which represents a much more desirable characteristic. Two additional factors contribute to obtaining further improvement of the limiting characteristic. Firstly, although it is assumed that the valve is at zero grid bias in the absence of an input signal, this is generally not so. The random arrival of electrons at the grid means that the grid will take up a standing negative potential in the absence of a signal, and hence the valve will be operating nearer the region of curvature; the rise of the fundamental frequency anode current component with input is thus more gentle and the threshold at which limiting commences is less sharply defined as shown in Fig. 8.4 (*b*). Against this must be set the fact that the maximum anode current output is lower, and if it is desired to retain the sharp threshold, and/or secure increased output, the grid resistor must be returned to a point of positive potential. The slope of the characteristic is then as shown in Fig. 8.4 (*a*).

Secondly, as the signal amplitude increases, the direct current component through the valve decreases, and hence the screen potential tends to rise. This in turn increases the bias required for anode current cut-off. This has the effect of increasing the fundamental frequency anode current output as the signal amplitude increases, and hence opposes the falling tendency exhibited by the curve of Fig. 8.2. By judicious choice of screen components, the curve of fundamental frequency anode current component against input can be made substantially flat above the threshold.

The threshold of limiting can be made to occur at a lower signal input level by employing a lower screen potential. However, the fundamental frequency component of the anode current also decreases under these conditions. The value of screen potential is therefore a compromise between a high value, to secure adequate input to the discriminator, and a low value to secure a low threshold of limiting.

The time constant of the grid circuit is somewhat critical. If it is too long, it will not be possible for the capacitor to discharge sufficiently rapidly to follow the amplitude variations of the input signal. Under these circumstances after a burst of interference, the limiter may be cut off whilst the capacitor discharges. The charging of the capacitor presents no difficulties, since the time constant is determined in this case by the forward resistance of the diode formed by the valve grid and cathode, which has much lower value than the grid leak. If the time constant $RC$ of Fig. 8.1 is reduced to overcome this effect, difficulties arise from two other factors. If the capacitor alone is reduced, it cannot be reduced indefinitely, since it forms a potential divider with the valve input capacitance, and loss of signal will result. If the resistance is reduced, the damping of the i.f. transformer feeding the limiter increases. The actual equivalent damping resistor is given by $R/2$ for the series-fed diode arrangement of Fig. 8.1 and $R/3$ for the shunt-fed connection shown dotted. This type of limiter is therefore unable to suppress completely very rapid changes of input signal amplitude. From cathode ray oscillograph studies of Hobbs and others, it is apparent that the fault can be reduced to reasonable proportions if the grid circuit time constant is kept down to the order of 2·5 micro-seconds or less. Typical values for the grid leak and capacitor are 100 kilohms and 25 pf.

Some degree of interstation noise suppression can be obtained by biasing the limiter towards the cut off value by means of a substantial bleed current through a suitable resistance in the cathode circuit, or by returning the grid leak to a point of negative potential. The effect of this is to reduce the gain at low signal input levels; the approximate shape of the limiting characteristic is shown in Fig. 8.4 (*c*).

In order to avoid overloading the stages which come before the limiter, a small amount of a.v.c. is sometimes employed. The amount of a.v.c. is, however, kept down to a minimum in order that the signal applied to the limiter may be as large as possible. In order to eliminate any unnecessary delay in the operation of the limiter grid circuit it is desirable to provide a series resistance to isolate the limiter grid from the a.v.c. circuit capacity.

### Grid Circuit De-tuning

An unfortunate feature of the grid leak type of limiter is the fact that the damping caused by the flow of grid current results in de-tuning of the input circuit. This is demonstrated by the oscillograms recorded by Landon and reproduced in Fig. 8.5. These oscillograms, which were obtained at the limiter anode, show the wave-train which is produced by an impulsive signal. In oscillogram *A* the input level is too low for limiting to take place. Following the theoretical discussion in Chapter Three, it will be apparent that the narrow neck between the two lobes indicates accurate alignment of all the circuits involved. The oscillograms *B* and *C* were taken at progressively higher impulsive signal levels. As the deep valley between the lobes is missing this indicates that some circuits must have become de-tuned. The corresponding oscillograms for the discriminator output wave-forms are shown as *D*, *E*, and *F*.

The next six oscillograms demonstrate that the trouble is being very largely caused by the de-tuning of the limiter grid circuit. They repeat the conditions of the preceding six, except that the limiter grid is now fed from a circuit tuned with a 600-micro-microfarad condenser in place of the original 100-micromicrofarad tuning condenser. That the de-tuning at high signal levels is correspondingly less is illustrated by the deep valley between the lobes in *H* and *I*, and also the smaller deflection in *K* and *L*.

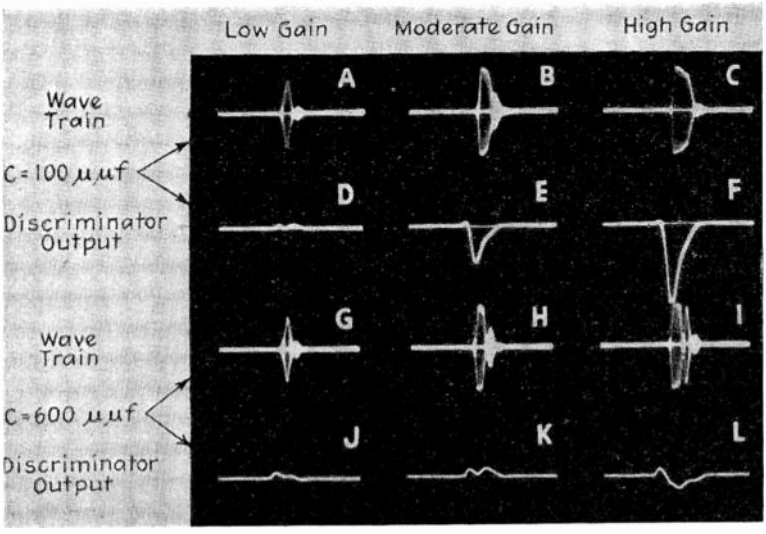



FIG. 8.5.—Oscillograms showing the distortion caused to an impulse wave-train as a result of the de-tuning produced by the flow of limiter grid current.
(*By courtesy of "Electronics".*)

The de-tuning caused by grid current can also be minimised in other ways. For example, the response of the affected transformer can be made broader than that of the preceding stages; under these conditions the de-tuning will not greatly affect the overall receiver characteristic.

## Anode Limiters

Another form of limiter, which is often employed in conjunction with a grid limiter, is the anode limiter. In this type of limiter, the anode is operated at a low potential, with a relatively high value of anode load which we shall assume initially to be resistive. If the limiter is normally biased, then the anode voltage swings in the downward sense are limited by the "knee" of the $i_a - v_a$ characteristic, i.e. the valve "bottoms" at relatively small values of signal input, as shown in Fig. 8.6. The maximum positive going excursions of the anode voltage swings are limited by the low value of h.t. employed. If the valve is biased towards cut-off, it can be made to limit symmetrically, and once the input signal exceeds the amplitude necessary for limiting to commence, the

amplitude of the anode voltage output remains constant, the wave-form becoming more nearly rectangular as the input is further increased as shown in Fig. 8.7.

Where the anode load comprises a tuned circuit, limiting again occurs when the anode voltage is driven below the "knee" of the

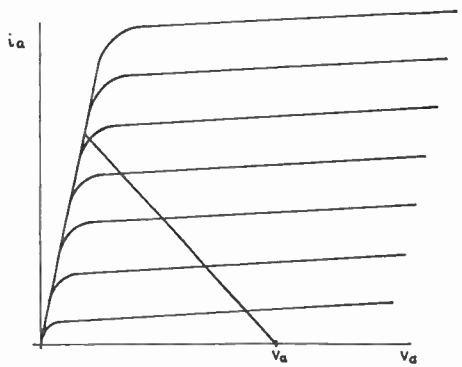



FIG. 8.6.—Anode limiter operated with low h.t. voltage.

$i_a$—$v_a$ characteristic. As the input signal is increased in amplitude, the output tends to remain constant because the damping imposed on the tuned circuit increases; i.e. the load line tends to become progressively more steep, as shown in Fig. 8.8. As the increased damping lowers the $Q$ of the tuned circuit, there is progressive

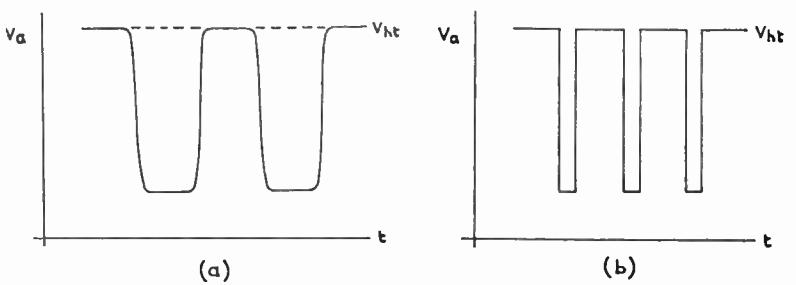



FIG. 8.7.—Anode voltage wave-form; anode limiter with resistive load. (*a*) small input, (*b*) larger input.

degradation of the wave-form from the sinusoidal shape, the negative going peaks becoming more flattened. Because of this damping imposed, the anode limiter is not suitable for use with a discriminator as its anode load. For this reason and because its limiting action commences with a lower input than with a grid

limiter, it is generally employed in the first stage of a two stage limiter, the second limiter being usually of the grid limiter type. In this condition of operation the anode limiter is frequently combined with a grid limiter in the first stage, the anode and screen potentials being both held to a low value. The initial anode limiting action is then supplemented by the grid limiting action occurring at a higher input signal level.

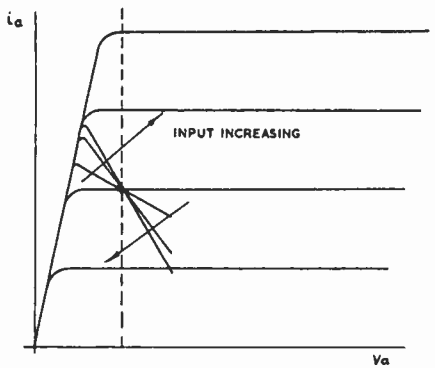



FIG. 8.8.—Anode limiter with tuned circuit load.

When the anode load comprises a tuned circuit, the output necessarily remains approximately sinusoidal. The position of the load line is then very difficult to determine. In fact, the usual way to plot the load line in such circumstances is to invert the normal procedure of drawing the load line to determine the output; the load line is positioned to agree with the output wave-form. This is shown in Fig. 8.9 for a stage employing anode and grid limiting, when driven hard. It will be noted that the peak output cannot exceed the h.t. voltage. For comparison, the load line with anode limiting just commencing ($a$-$a$) is also shown; it will be seen that for the closest approach to ideal limiting action, the portion of the $i_a$—$v_a$ characteristic below the knee should be as steep as possible.

Where such a limiter is employed prior to a grid limiter, the input to the second valve is already substantially limited; the second limiter can therefore be designed to give a higher output from the discriminator.

### Oscillator Limiters

Perhaps one of the most interesting methods of eliminating the amplitude variations from the received carrier is that based

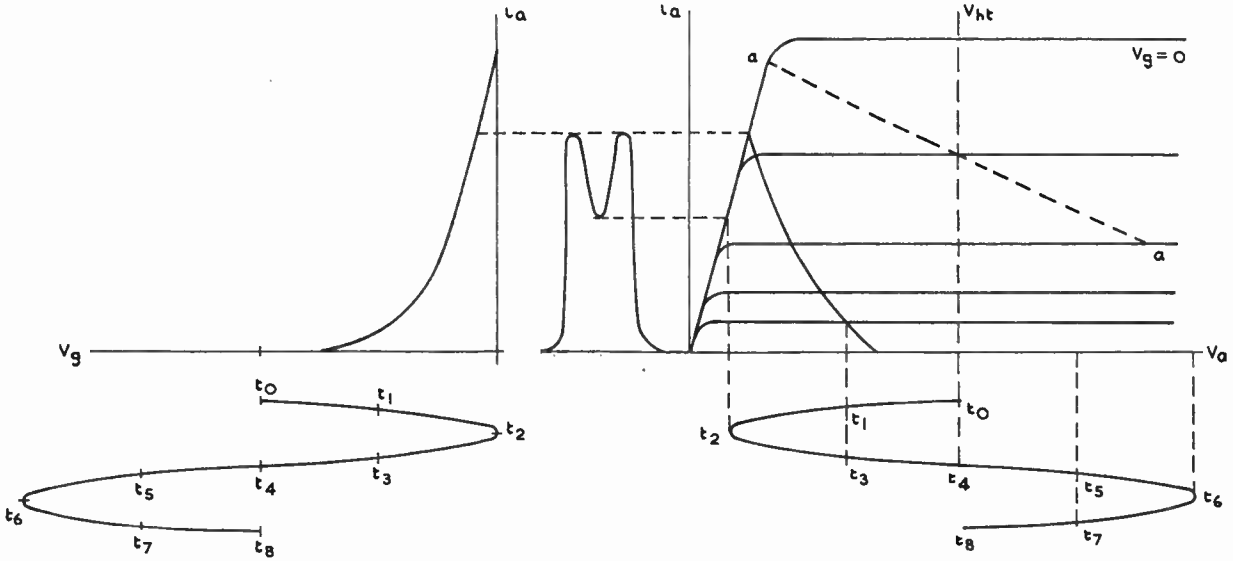



FIG. 8.9.—Derivation of wave-form for combined grid and anode limiter with tuned circuit load.

on a synchronised oscillator. The idea behind this type of limiter is that an oscillator having a constant output voltage should have its frequency controlled by the incoming carrier. It is claimed that a limiter of this type is capable of giving a better amplitude limiting action than that of the conventional grid leak limiter, and at the same time a selectivity to adjacent channel interference equal to that of two extra i.f. stages. It is further claimed that, if properly designed, it is possible to obtain a synchronisation

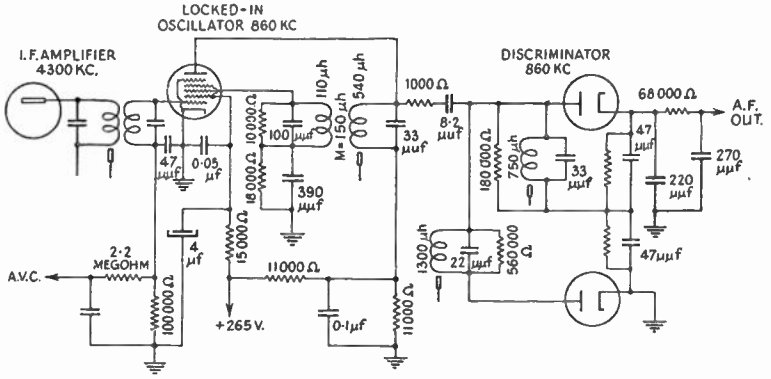



FIG. 8.10.—Complete oscillator-limiter showing the locked-in oscillator and the special double-tuned circuit type of discriminator.

sensitivity high enough to give a voltage gain (when translated into terms of an amplifier) equal to that of a conventional i.f. stage.

A paper describing an experimental receiver using this type of limiter has been published by Beers. This receiver used an i.f. frequency of 4·3 Mc/s with the synchronised or locked-in oscillator working at one-fifth of this frequency. In place of the normal discriminator designed to operate from carrier frequency deviations of $\pm 75$ kc/s the "frequency-dividing locked-in" oscillator was followed by a special double-tuned circuit type of discriminator designed to operate from carrier frequency deviations of $\pm 15$ kc/s.

The general arrangement of the complete oscillator-limiter section of the receiver is shown in Fig. 8.10. With this circuit arrangement an intermediate frequency signal of 1 volt on the first grid of the oscillator valve was required in order to provide the desired "locked-in" range of $\pm 110$ kc/s. The frequency range

in excess of the $\pm 75$ kc/s required for the normal modulation of the received signal was provided in order to take care of mistuning by the user, frequency drift of the heterodyne oscillator and over-modulation at the transmitter. The oscillator voltage developed at the discriminator was between 20 and 30 volts.

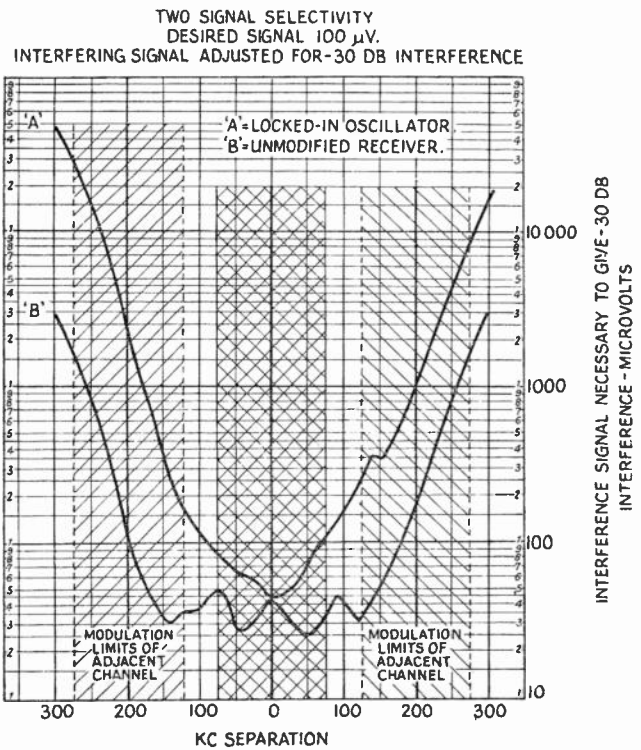



FIG. 8.11.—Illustrating the improved selectivity obtained with an oscillator-limiter.

From the above it is apparent that the receiver must be sufficiently sensitive to always produce 1 volt on the first grid of the oscillator.

The results of selectivity measurements made by the two-signal method are shown in Fig. 8.11. In these tests the receiver was tuned to a desired signal of 100 microvolts, with a 400-cycle modulation and a deviation of $\pm 25$ kc/s. An interfering signal modulated with 1,000 cycles and having a deviation of $\pm 25$ kc/s

was adjusted in signal strength and frequency to give an interference output 30 db below the 400-cycle output. A very considerable improvement in selectivity, especially over the adjacent channel area, is shown by the receiver employing the oscillator-limiter circuit.

It should be noted that, as the interfering signal is increased, a point at which the oscillator tends to break away from the desired signal will ultimately be reached. Field tests indicate that a somewhat higher distortion than the conventional receiver's is encountered when a receiver incorporating an oscillator-limiter is tuned so that the signal is received at the edges of the receiver response characteristic. Various practical difficulties associated with this type of limiter are discussed in an article by C. W. Carnahan and H. P. Kalmus.

If so desired an oscillator-limiter and a phase detector can be combined in a single-valve stage. In an article describing such a circuit W. E. Bradley claims that it is possible to obtain a response to amplitude variations which is 50 db less than that due to frequency changes. The audio output at full deviation from such a stage is claimed as 20 volts peak to peak.

### Series Grid Resistance Type of Limiter

Where the frequency of the signal being limited is relatively low, as in the case of either sub-carrier frequency modulated signals or carriers of a few tens of kilocycles (such as those used for transmission over lines), then it is possible to employ a very simple type of limiter. The circuit consists of a series grid-feed resistance which is high in comparison with the limiter valve's grid/cathode impedance. The valve is then operated under conventional limiter conditions so that it is in cut-off when its grid is only a few volts negative. When a signal is applied to the valve, limiting of the positive half of the signal wave occurs as a result of the voltage drop through the series grid resistance, while that of the negative half of the wave occurs in the normal way, as a result of the grid passing into the cut-off zone.

### Cathode-Coupled Limiter

This type of limiter comprises two valves connected as shown in Fig. 8.12. The common cathode load is of a high value, its actual magnitude being chosen to bias the valves to the mid point of the

grid base approximately. When a signal is applied to the grid of $V1$, the cathode tends to follow the input wave-form. In so doing, it provides an input to $V2$. The consequent change in the anode current of V2 is in the opposite sense to the change occurring in $V1$, and tends to maintain the cathode voltage at the quiescent value. Some change of cathode voltage must, however, occur to

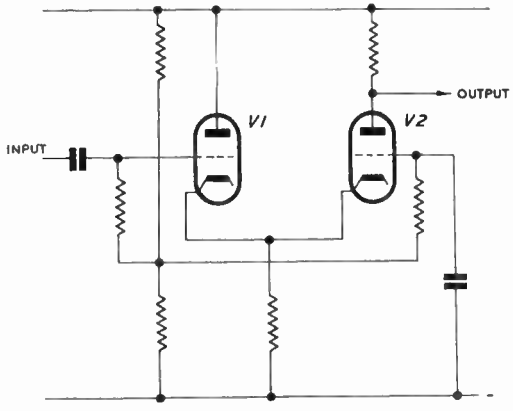



FIG. 8.12.—Cathode-coupled type of limiter.

provide the input to $V2$, and in practice the cathode voltage variations are approximately one half of those at the grid of $V1$. Used in this manner, this type of circuit provides a push-pull output at the anodes, and is frequently used for this purpose.

When, however, the signal amplitude is increased sufficiently, the anode current in $V1$ will be cut off at one signal peak, whilst the anode current in $V2$ will be cut off at the other. When this happens, limiting of the output signal occurs.

This type of circuit suffers from the disadvantage of requiring a rather large input signal for efficient limiting, of the order of 5–10 volts peak. It is, however, used extensively at relatively low frequencies, especially in conjunction with "counter" type discriminators.

### Frequency to Amplitude Conversion

A very large measure of the success attained by wide-band frequency modulation can be attributed to the high efficiency with which it is possible to convert changes in carrier frequency into audio voltages. As late as 1932 a paper was published by

Andrew, in which it was deduced that a receiver designed for frequency modulation would produce less than one-tenth the power output of an amplitude modulation receiver. This author, and others of the same period, based their calculations on the only method then available for the demodulation of a frequency modulated carrier. They used the sloping side of the receiver response curve to convert variations in frequency into amplitude changes. As will be seen from Fig. 8.13, this may be done

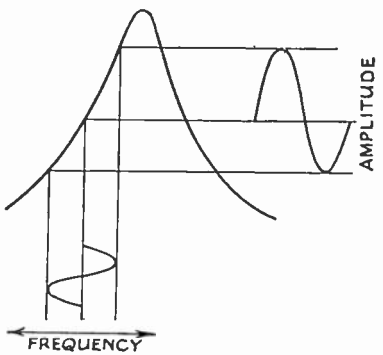



FIG. 8.13.—Illustrating the way in which the sloping side of the receiver's response-curve may be used to produce amplitude variations from carrier frequency changes.

by tuning the frequency modulated carrier about midway up one side of the response curve. In this way the frequency variations of the carrier result in amplitude variations which can be demodulated with a normal detector circuit.

While in an emergency it is possible to use an amplitude modulation receiver for the reception of a frequency modulated transmission, this method is never employed in practice. There are many objections; to start with, less than 50 per cent of the skirt of a tuned circuit's response curve is sufficiently straight to permit of even substantially linear frequency to amplitude conversion. A further loss results from the amplification of the carrier some way down the skirt, rather than at the crest of the response curve. Quite apart from these considerations all the benefits of noise suppression are completely lost.

In 1936 Armstrong's classic paper introducing the basic concepts behind wide-band frequency modulation contained the following observation: "The most difficult operation in the

receiving system is the translation of the changes in the frequency of the received signal into a current which is a reproduction of the original modulating current." Today, although the discriminator is still a most important stage in a frequency modulation receiver, it is hardly fair to describe it as the most difficult.

## The Double-Tuned Circuit Discriminator

The discriminator circuits in current use fall into two main classes. Firstly, those depending on two tuned circuits, one resonant beyond the upper and the other below the lower deviation

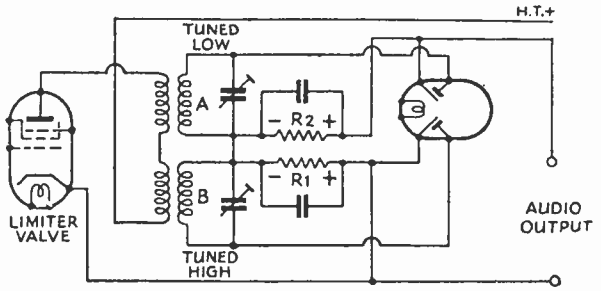



FIG. 8.14.—The double-tuned circuit type of discriminator was first described by Travis. This typical circuit has been used in a Motorola mobile communication receiver.

limit. The second arrangement depends for its functioning upon the phase shifts which occur between the primary and the secondary windings of a tuned transformer. Of these two types the latter is by far the most popular, and can for all practical purposes be regarded as the standard frequency modulation discriminator circuit.

A typical circuit of the first type of discriminator, originally described by Travis, is shown in Fig. 8.14. There is a very wide variety of ways in which this circuit can be arranged, but the basic functioning of all is the same. That illustrated consists of two tuned circuits—one tuned to a frequency above the upper and the other below the lower deviation frequency limit.

As the carrier frequency is modulated over the receiver response band, the voltage characteristics indicated in Fig. 8.15 (*a*) are produced across the two diode loads. These curves are those which the voltage developed across any parallel-tuned circuit will

follow as the applied frequency is varied. This voltage is determined by the equation:

$$E = I \left\{ \frac{(R+j\omega L)}{(1-\omega^2 LC+j\omega CR)} \right\}. \quad . \quad . \quad (8.1)$$

It will be noted that while the voltage produced across $R_2$ is

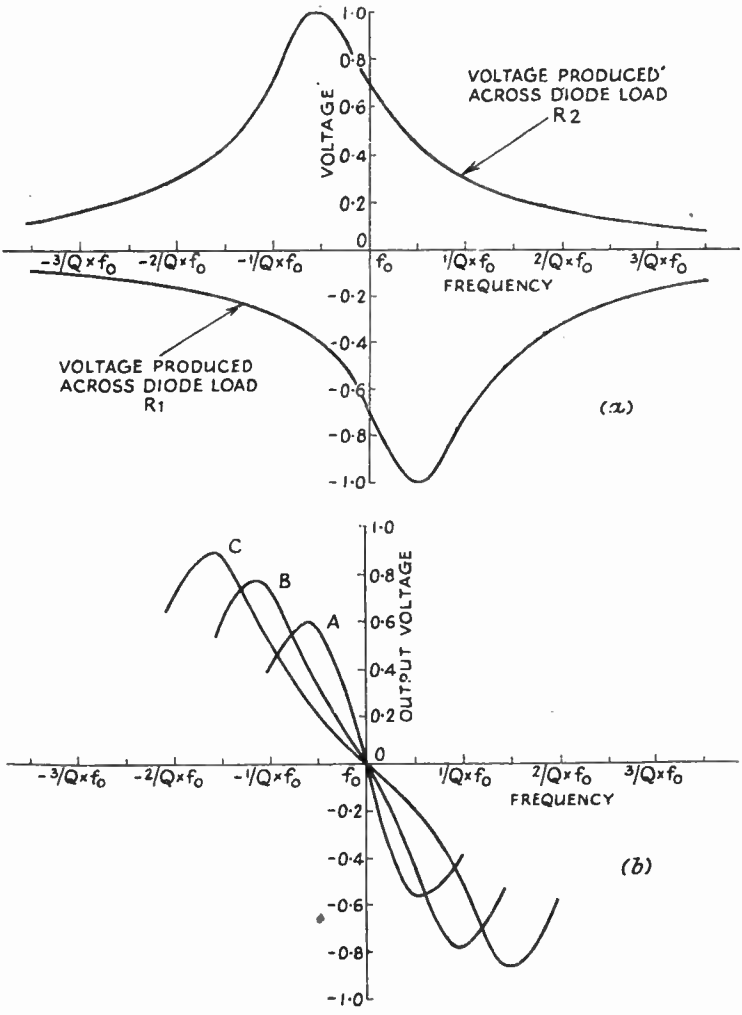



FIG. 8.15.—The response of the two tuned circuits results in voltages being developed across the two diodes loads, which when added together, result in the overall discriminator characteristics shown in (*b*).

positive, that across $R_1$ is negative. The output from the discriminator will therefore be the difference or algebraic sum of these two voltages. Fig. 8.15 (*b*) shows a group of overall characteristics obtained when these curves are combined; these curves illustrate the effect of various different spacings between the frequencies to which the two circuits are tuned. The curve *A*, which is linear over a considerable part of its range, is produced when the two tuned circuits are separated by a frequency equal to $1/Q \times f_0$, where $Q = \omega L/R$, and is the same value for both circuits, and $f_0$ is the frequency midway between those to which the two circuits are tuned.

The effect of increasing the frequency separation is indicated in curves *B* and *C*, being respectively those obtained with a frequency separation between peaks of $2/Q \times f_0$ and $3/Q \times f_0$. It will be noted that although a wider peak separation results in an increased output, this is only obtained at the expense of linearity. It would appear from these curves that the spacing between the two resonant frequencies is not very critical, and that any frequency separation between $1/Q \times f_0$ and $2/Q \times f_0$ would give reasonable results. While this is substantially correct, it is normally desirable to be considerably more exact than this if first-class results are required. The group of curves in Fig. 8.16 indicate the percentage departure from the tangents drawn through the cross-over points of the group of discriminator curves with frequency separations of $1/Q \times f_0$, $\sqrt{1 \cdot 5}/Q \times f_0$, $1 \cdot 5/Q \times f_0$, and $2/Q \times f_0$. From these curves it is apparent that for general purpose working, a peak separation of $1 \cdot 5/Q \times f_0$ gives a response characteristic with a very acute turnover and a maximum departure of just over 1 per cent from the tangent drawn through its cross-over point. This separation will, in general, be the most satisfactory for all normal purposes. Where an even closer approach to a truly linear characteristic is required the separation should lie somewhere between $1 \cdot 5/Q \times f_0$ and $\sqrt{1 \cdot 5}/Q \times f_0$.

The curves given in Figs. 8.15 and 8.16 make it possible to arrive at working values for this type of discriminator. To take one example, assume that a receiver with an 8-Mc/s i.f. is designed for operation on a 75-kc/s deviation system. The discriminator response is to be linear within 1 per cent over a band of $\pm 100$ kc/s. A large-scale plot of the characteristic obtained with a peak separation of $1 \cdot 5/Q \times f_0$ shows that this characteristic is within

these limits over a frequency range of $1/Q \times f_0$. It therefore follows that $1/Q \times f_0 = 200$ kc/s, from which it is apparent that the peak separation is 300 kc/s, and that the $Q$ of the two tuned circuits is

$$Q = \frac{8 \text{ Mc/s}}{200 \text{ kc/s}} = 40.$$

It should be noted that this figure is that obtained under actual working conditions and includes the damping effect of the two diode load circuits.

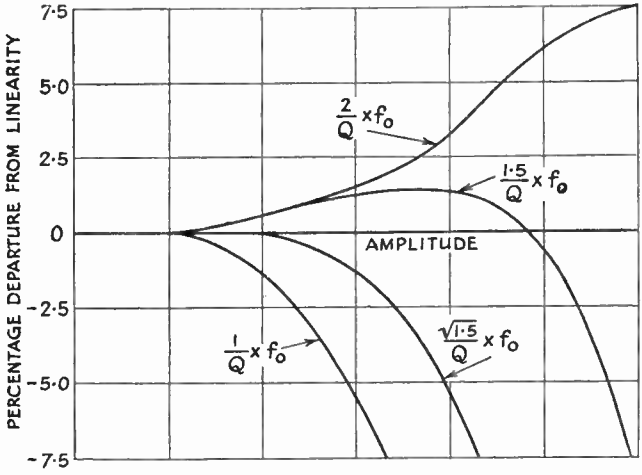



FIG. 8.16.—Percentage departure from the tangents drawn through the cross-over points of discriminator curves which are produced with the peak separations indicated. The scale of the base has been modified to improve readability.

Two alternative circuit arrangements are shown in Fig. 8.17. Both circuits have been used in commercial receivers. When designing discriminators of the type that employ two independently tuned circuits, it should be noted that all the foregoing deductions assume that the coupling between the circuits concerned is kept substantially below the critical value. In addition to those shown in Fig. 8.17, there are many other variations of the double-tuned circuit type of discriminator. Some quite unrecognisable circuits turn out to be variations on the same basic type.

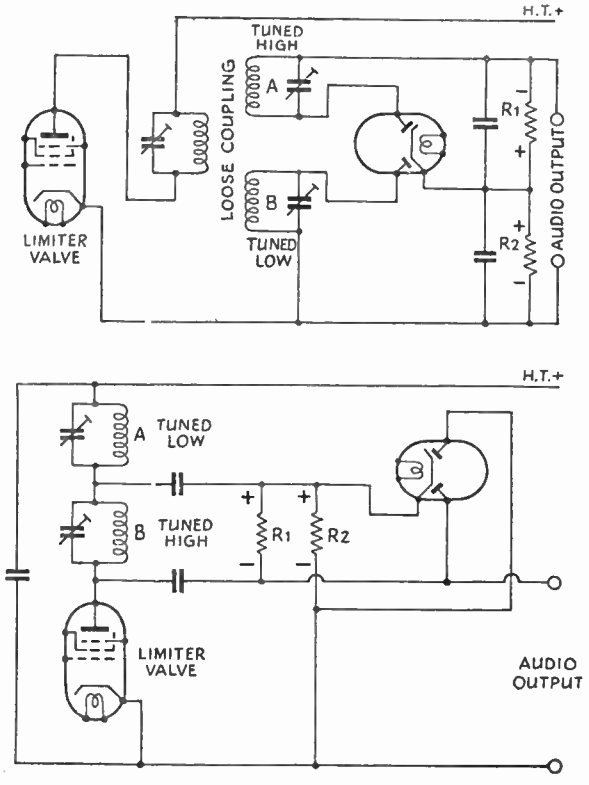



FIG. 8.17.—Alternative arrangements of the double-tuned discriminator circuit.

## Phase Difference Discriminator

The phase difference type of discriminator has become so widely used that it can for most practical purposes be regarded as the standard frequency modulation discriminator. It was first introduced by Foster and Seeley as a means of developing the control voltages required by receivers incorporating automatic frequency control. At a later date a complete mathematical treatment of the circuit theory was published by Hans Roder.

The circuit arrangement of the phase difference type of discriminator is shown in Fig. 8.18. It is based on a tuned primary, tuned secondary i.f. transformer. The voltage developed across the primary is injected into the centre of the secondary winding via $C_2$. At the frequency to which the transformer has been

aligned the voltages applied to the two diodes $D_1$ and $D_2$ are equal. Consequently the rectified output voltages produced across the loads $R_1$ and $R_2$ are also equal, and, being of opposite polarity, cancel each other out, with the result that zero voltage is produced across the output terminals.

If the signal frequency applied to the discriminator transformer

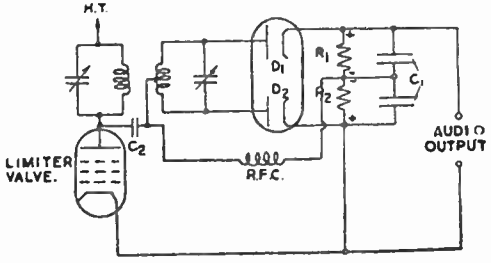



FIG. 8.18.—The circuit arrangement of the phase difference discriminator.
*(By courtesy of the British Institute of Radio Engineers.)*

is varied, then—for reasons which will be discussed presently—the signal applied to one diode, say $D_1$, will be larger than that applied to the second diode $D_2$. This results in a greater voltage being developed across $R_1$ than across $R_2$, with the result that a positive output voltage is produced. If, however, the frequency applied to the transformer is varied in perhaps the opposite

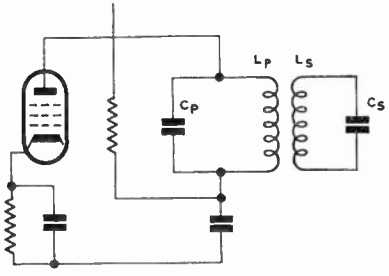



FIG. 8.19.—Two coupled tuned circuits; see Fig. 8.20 for equivalent diagram.

direction, then the voltage applied to $D_2$ will become the larger, with the result that the output voltage will be negative.

In order to describe the various types of phase difference discriminators, it is necessary first to establish the relationship between currents and voltages in two coupled circuits.

Such a coupled pair is shown in Fig. 8.19; for the sake of generality the two circuits are assumed to have dissimilar values of inductance, capacitance and resistance. The equivalent circuit is shown in Fig. 8.20, together with the symbols employed in the following text.

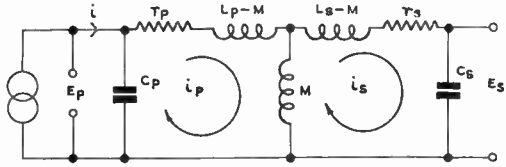



FIG. 8.20.—Equivalent diagram of Fig. 8.19.

The valve feeding the primary circuit is assumed to be a constant current generator, i.e. its a.c. anode resistance is assumed very large compared with the impedance of the load circuit. This is generally true for pentode valves.

The primary and secondary circuits are assumed resonant at the working frequency $f_0$, and the generator output current $i$ is the peak value of the a.c. component of the anode circuit in the region of the resonant frequency. This latter point is stressed since, if the driving stage is a limiter, the anode current wave-form will contain a very high percentage of harmonics. The fundamental circuit equations are:

$$\begin{aligned} E_s &= jX_{cs}i_s, \\ E_p &= jX_{cp}(i-i_p), \\ 0 &= Z_p i_p - j\omega M i_s - jX_{cp}i, \\ 0 &= Z_s i_s - j\omega M i_p, \end{aligned}$$

where $X_{cs} = \dfrac{1}{\omega C_s}$ = reactance of secondary tuning capacitor at applied signal frequency $f = \dfrac{\omega}{2\pi}$;

$$X_{cp} = \frac{1}{\omega C_p};$$

$$Z_p = jL_p\omega + \frac{1}{j\omega C_p} + r_p;$$

$$Z_s = jL_s\omega + \frac{1}{j\omega C_s} + r_s.$$

Since it is intended to confine the examination of circuit relationships to the region close to the resonant frequency of the primary and secondary circuits, it may be assumed that $i_p > i$; this is true provided that the $Q$ values of primary and secondary circuits are large.

With this simplification:

$$E_s = -jX_{cp}X_{cs}\frac{\omega M}{Z_pZ_s+\omega^2M^2}i,$$

$$E_p = X_{cp}{}^2\frac{Z_s}{Z_pZ_s+\omega^2M^2}i.$$

It is instructive to note that the term $X_{cp}/(Z_pZ_s+\omega^2M^2)$ is common to both expressions; the significance of this fact is discussed later. Since it is intended to apply the expressions only in the region of resonance, a further simplification can be employed. At resonance $\omega_0L_p = 1/\omega_0C_p$ and $\omega_0L_s = 1/\omega_0C_s$; at an adjacent frequency, $f = f_0+\delta f$, $\omega_0L_p - 1/\omega C_p$ is approximately equal to $2L_p\delta\omega$. This approximation is in error by only 5 per cent at $\delta f = f_0/10$, and can therefore be employed with negligible error. Similarly, $\omega L_s - 1/\omega C_s = 2L_s\delta\omega$.

Substituting in the expressions above:

$$E_s = jX_{cp}X_{cs}\frac{\omega M}{(r_p+2jL_p\delta\omega)(r_s+2jL_s\delta\omega)+\omega^2M^2}i,$$

$$E_p = X_{cp}{}^2\frac{r_s+2jL_s\delta\omega}{(r_p+2jL_p\delta\omega)(r_s+2jL_s\delta\omega)+\omega^2M^2}i.$$

Additionally we shall assume that $X_{cp}$, $X_{cs}$, and $\omega M$ are constant and equal to their values at $\omega_0$; this again involves a negligibly small error. Dividing each expression by $r_pr_s$, putting

$$\frac{L_p\omega_0}{r_p} = \frac{X_{cp}}{r_p} = Q_p, \qquad \frac{L_s\omega_0}{r_s} = \frac{X_{cs}}{r_s} = Q_s, \qquad 2\frac{\delta\omega}{\omega_0} = x = 2\frac{\delta f}{f_0},$$

and employing $n = K\sqrt{Q_pQ_s}$, where $K = \dfrac{M}{\sqrt{L_pL_s}}$,

$$E_s = \frac{-jQ_pX_{cp}\,KQ_s\sqrt{L_s/L_p}}{(1+jQ_px)(1+jQ_sx)+n^2}i, \qquad . \quad . \quad (8.2)$$

$$E_p = \frac{Q_pX_{cp}(1+jQ_sx)}{(1+jQ_px)(1+jQ_sx)+n^2}i. \qquad . \quad . \quad (8.3)$$

If $L_s=L_p$, $Q_s=Q_p$, the expressions at resonance ($x=0$) simplify to

$$E_s=E_{s0}=-jQ_pX_{cp}\frac{n}{1+n^2},$$

$$E_p=E_{p0}=Q_pX_{cp}\frac{1}{1+n^2}.$$

The magnitudes of $E_s$ and $E_p$ for $L_p=L_s$ and $Q_s=Q_p$ are shown in Fig. 8.21. The coupling between the circuits is given by $n$, and

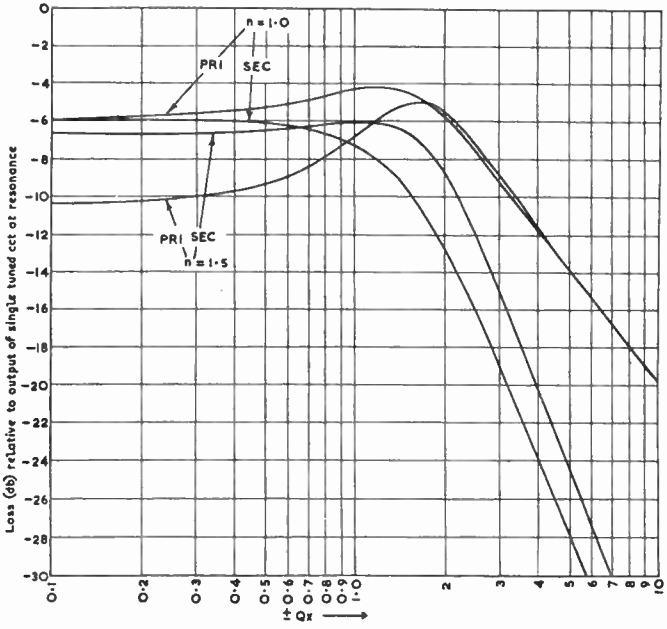



FIG. 8.21.—Primary and secondary voltage curves for a transformer having equal primary and secondary impedances and couplings which first critical ($n=1$) and then $1{\cdot}5\times$ critical.

for critical coupling $n=1$. It will be seen that the primary voltage exhibits a much greater variation in amplitude than the secondary voltage.

It is sometimes convenient to express the secondary voltage in terms of the primary voltage; from (8.2) and (8.3),

$$E_s=-j\frac{KQ_s\sqrt{L_s/L_p}}{1+jQ_sx}E_p. \qquad . \quad . \quad . \quad (8.4)$$

### Foster-Seeley Discriminator

The basic circuit arrangement for this type of discriminator is shown in Fig. 8.18. Although the whole of the primary voltage is employed in the circuit shown in the figure, often only a portion of the primary voltage is used.

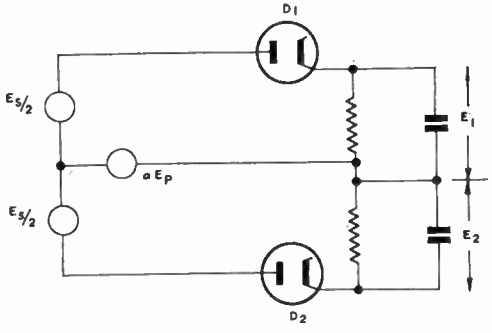



FIG. 8.22.—Equivalent diagram of Foster-Seeley discriminator.

The equivalent circuit for the arrangement is shown in Fig. 8.22; here a proportion $a$ of the primary voltage is employed. The voltage applied to each diode is given by the vector sum of $aE_p$ and $\frac{1}{2}E_s$; if $E_p$ is assumed constant, the vector relationships are as shown

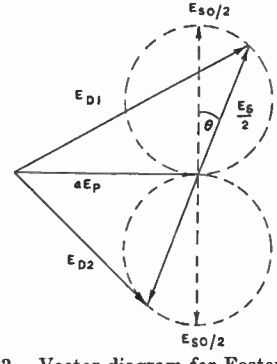



FIG. 8.23.—Vector diagram for Foster-Seeley discriminator, assuming $E_p$ constant.

in Fig. 8.23. At resonance, the vectors $E_s/2$ are perpendicular to the vector representing $aE_p$; this is apparent from expression (8.4) above, the 90 phase rotation being indicated by the $j$ term. The locus traced out by the tips of the two half-secondary voltage

vectors is a circle. This can be shown from geometrical considerations, since the component in phase with the primary voltage is given by

$$\pm KQ_s\sqrt{L_s/L_p}\,\frac{Q_s x}{1+Q_s{}^2x^2}\,E_p$$

and the quadrature component by

$$\pm j\,\frac{KQ_s\sqrt{L_s/L_p}}{1+Q_s{}^2x^2}\,E_p.$$

The voltages applied to the two diodes, $E_{d1}$ and $E_{d2}$, are given by

$$E_{d1}{}^2=\left(aE_p+\frac{1}{2}\,\frac{Q_s x KQ_s\sqrt{L_s/L_p}}{1+Q_s{}^2x^2}\,E_p\right)^2+\frac{1}{4}\,\frac{K^2Q_s{}^2L_s/L_p}{(1+Q_s{}^2x^2)^2}\,E_p{}^2,$$

$$E_{d2}{}^2=\left(aE_p-\frac{1}{2}\,\frac{Q_s x KQ_s\sqrt{L_s/L_p}}{1+Q_s{}^2x^2}\,E_p\right)^2+\frac{1}{4}\,\frac{K^2Q_s{}^2L_s/L_p}{(1+Q_s{}^2x^2)^2}\,E_p{}^2.$$

In order to simplify the calculation involved, we shall now assume that $Q_p=Q_s=Q$; this does not unduly restrict the treatment. Additionally, let

$$E_p{}'=aE_p,$$

$$KQ=n,$$

$$\frac{1}{a}\,KQ\sqrt{L_s/L_p}=b.$$

$E_p{}'$ is obviously the actual magnitude of the primary voltage injected in series with the half-secondary voltages, whilst $b$ is the ratio of the secondary voltage to employed primary voltage at resonance; i.e. at $\delta\omega=0$, $b=E_s/E_p{}'$

Then

$$E_{d1}{}^2=a^2E_p{}^2\left\{\left(1+\frac{1}{2}\,\frac{bQx}{1+Q^2x^2}\right)^2+\frac{1}{4}\,\frac{b^2}{(1+Q^2x^2)^2}\right\},$$

$$E_{d2}{}^2=a^2E_p{}^2\left\{\left(1-\frac{1}{2}\,\frac{bQx}{1+Q^2x^2}\right)^2+\frac{1}{4}\,\frac{b^2}{(1+Q^2x^2)^2}\right\}.$$

Since $aE_p$ is a variable quantity, it is more convenient at this point to postulate that $E_s$ is held constant; we shall consider the effect of the variation of $E_s$ with frequency later. Employing the relationship

$$E_s = -j\frac{b}{1+jQx}aE_p$$

we have

$$E_{d1}^2 = E_s^2\frac{(1+jQx)^2}{b^2}\left\{\left(1+\frac{1}{2}\frac{bQx}{1+Q^2x^2}\right)^2+\frac{1}{4}\frac{b^2}{(1+Q^2x^2)^2}\right\},$$

$$E_{d2}^2 = E_s^2\frac{(1+jQx)^2}{b^2}\left\{\left(1-\frac{1}{2}\frac{bQx}{1+Q^2x^2}\right)^2+\frac{1}{4}\frac{b^2}{(1+Q^2x^2)^2}\right\}.$$

The term $(1+jQx)^2$ indicates that there is a bodily rotation of the vector diagram of Fig. 8.23, with respect to its position at resonance; since this does not affect the magnitude of $E_{d1}$ and $E_{d2}$, to which the diodes are responsive, the term may be replaced by $(1+Q^2x^2)$, its modulus value. Hence,

$$E_{d1}^2 = E_s^2\left\{\frac{1+Q^2x^2}{b^2}+\frac{1}{4}+\frac{Qx}{b}\right\},$$

$$E_{d2}^2 = E_s^2\left\{\frac{1+Q^2x^2}{b^2}+\frac{1}{4}-\frac{Qx}{b}\right\}.$$

Since $E_{d1}$ and $E_{d2}$ are the peak values of the signals applied to the two diodes, the outputs across the two loads will therefore be equal to $E_{d1}$ and $E_{d2}$ respectively assuming 100 per cent rectification efficiency. The difference, $E=E_{d1}-E_{d2}$, represents the net output due to the departure of the carrier frequency from $f_0$.

$$E = E_s\left[\left\{\frac{1+Q^2x^2}{b^2}+\frac{1}{4}+\frac{Qx}{b}\right\}^{\frac{1}{2}}-\left\{\frac{1+Q^2x^2}{b^2}+\frac{1}{4}-\frac{Qx}{b}\right\}^{\frac{1}{2}}\right]. \quad (8.5)$$

The value of $E/E_s$ for various values of $b$ is plotted in Fig. 8.24. It will be seen that as $b$ increases, the linearity is improved. The output for negative values of $Qx$ are, of course, equal in magnitude but opposite in sign to those for positive values. The sensitivity measured in volts per kc/s of frequency shift obviously decreases with increasing $b$. Thus the value of $b$ employed represents a compromise between the requirements of good linearity and high sensitivity. In order to arrive at a quantitative assessment of the departure from linearity, we must take the ratio of the value of $E$ at a given value of $Qx$, and compare it with the value that it would have if the initial slope were maintained. From the expression above, if $Qx$ is considered vanishingly small:

$$E = E_{s0}\,Qx/(1+b^2/4)^{\frac{1}{2}}$$

This then gives the equation of the "ideal" output for a given value of $b$. It may be noted in passing that the maximum slope occurs when $b$ is very small. Thus for the range of values $1 \gg b^2/4$, the value of the initial slope tends to a constant value, and this sets a practical lower limit to the value of $b$

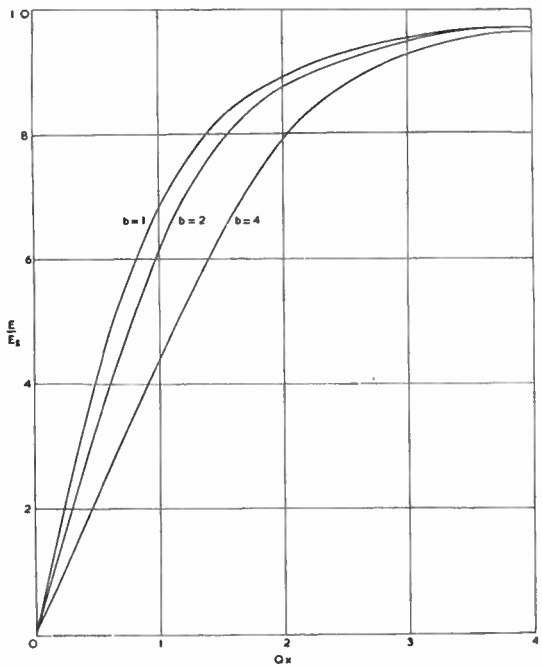



FIG. 8.24.—$E/E_i$ *v* $Qx$ for various values of $b$.

employed, since no further improvement in sensitivity is achieved when $1 \gg b^2/4$, i.e. $b < 0{\cdot}6$.

Reverting to the criterion for linearity, the "ideal" output for any value of $Qx$ is $E_i = E_{s0} Qx/(1+b^2/4)^{\frac{1}{2}}$. If the actual value is $E$, we shall measure the departure from linearity by the quantity

$$20 \log \frac{E_i}{E}.$$

The value of $20 \log E_i/E$ for the same range of values of $b$ used in Fig. 8.24 is shown in Fig. 8.25. Although the values of $20 \log E_i/E$ for a given value of $Qx$ fall with increasing values of $b$, an upper practical limit is set by the fact that the improvement tends to

become progressively less as $b$ increases. Thus the useful practical range of values of $b$ is $0{\cdot}6 < b < 6$.

We shall now consider the effect of variation of $E_s$ with $Qx$. For values of coupling factor $n$ greater than 1, $E_s$ increases over a limited range with $Qx$. Under these conditions, the value of $E$ is greater than the values shown in Fig. 8.24. By judicious choice of $n$, the increase in $E$ due to $E_s$ can be made to offset the fall below the "ideal" value over a range of values of $Qx$; the curve of $E$ against $Qx$ can be made to follow that of the "ideal" within

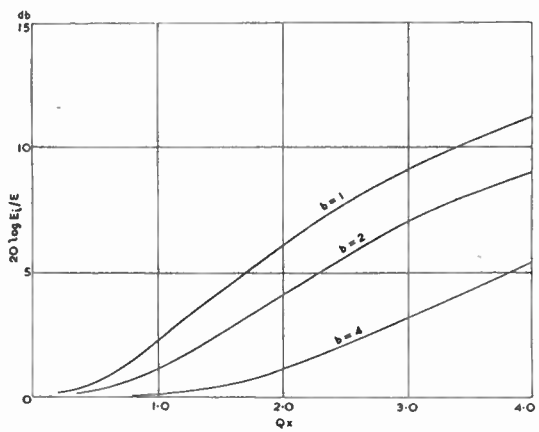



FIG. 8.25.—Showing departure from linearity of $E/E_i$ for values of $b$ used in Fig. 8.24.

close limits over a selected range. In order to select the value of $n$, the departures of $E_s$ from its value at resonance $E_{s0}$ is plotted in Fig. 8.26; the variable here is 20 log $E_s/E_{s0}$. For correction to be achieved at any given value of $Qx$, $E \times \dfrac{E_s}{E_{s0}} = E_i$. It follows therefore that for this condition 20 log $\dfrac{E_s}{E_{s0}} = 20$ log $\dfrac{E_i}{E}$. To satisfy this condition, the value of $n$ must be chosen to give 20 log $E_s/E_{s0}$ equal to 20 log $E_1/E$ at the selected value of $Qx$. In general, however, correction at a particular value of $Qx$ is not required, but correction over a range of values of $Qx$; for this condition, the value of $n$ must be chosen so that the curve of 20 log $E_s/E$ is identical with that of 20 log $E_1/E$ over the range.

It will be seen from Fig. 8.26 that the steepest slope of

20 log $E_s/E_{s0}$ occurs when $n=2$. For values of $n$ greater and less than this, the initial slope is at a lower value. If the curve of 20 log $E_s/E_{s0}$ is compared with those of Fig. 8.25, it will be seen that correction cannot be achieved for values of $b$ less than 2. At

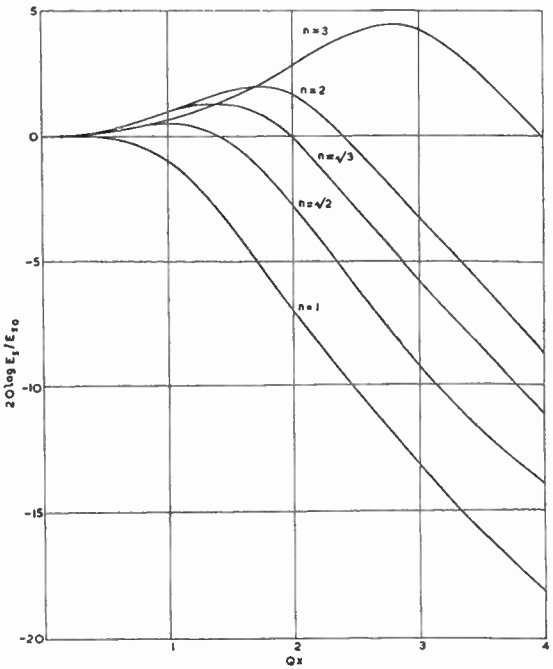



FIG. 8.26.—Variation of magnitude of secondary voltage with departure from centre frequency.

$b=2$, $n=2$, close matching is obtained up to $Qx=1$. The overall discriminator characteristic for this condition is shown in Fig. 8.27. These values represent the optimum for correction over the largest range of $Qx$; smaller values of $b$ require values of $n$ such that the correction is maintained only over a smaller range of $Qx$.

We shall consider a discriminator designed around the values of $b=2$, $n=2$, working at an i.f. of 10·0 Mc/s. For broadcast reception (75 kc/s deviation) it is desirable that the discriminator characteristic is linear over a range of $\pm 100$ kc/s, i.e. over a range of $x$ up to 0·02. Since, with the values chosen, linearity is maintained up to $Qx=1$, this fixes the value of $Q$ at 50. The value of $b$ then determines the remainder of the parameters of the system.

Since $b=\frac{n\sqrt{L_s/L_p}}{a}$ suitable values of $a$ and $\sqrt{L_s/L_p}$ are $a=1$, $\sqrt{L_s/L_p}=1$. The transformer then has identical primary and secondary circuits and the whole of the primary voltage is employed.

The sensitivity of the discriminator, when corrected in the above manner, is, of course, the same as that of the "ideal" over the corrected range. As shown earlier, the sensitivity of the "ideal"

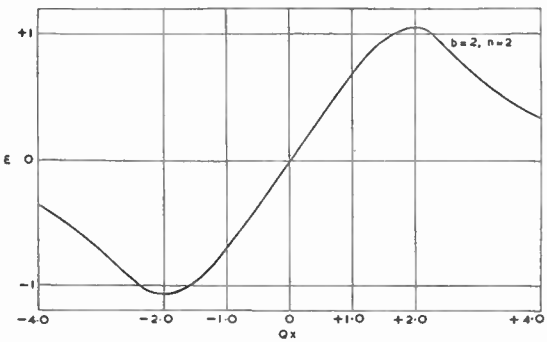



FIG. 8.27.—Discriminator characteristic for $n=2$, $b=2$.

curve is given by $E_{s0}Qx/(1+b^2/4)^{\frac{1}{2}}$. Substituting the value of $E_{s0}$, this is equal to

$$QX_{cp}\frac{n}{1+n^2}\frac{Qx}{(1+b^2/4)^{\frac{1}{2}}}\,i.$$

The maximum value of $E_{s0}$ occurs when $n=1$; thus if a lower range of linearity than given by the values $b=2$, $n=2$, can be tolerated, an increase in sensitivity can be achieved. As explained above, $b$ cannot be decreased below 2 without the loss of linearity at relatively small values of $Qx$. However, if $n$ is decreased, correction can be achieved for $b>2$, but for a smaller range of $Qx$. This leads to a somewhat higher sensitivity since the value of $n/1+n^2$ will increase. However, if such a reduction of the absolute magnitude of the linear range can be tolerated, it would be more advantageous to increase $Qx$ values, which would reduce the value of $x$ for linearity. In this case, the sensitivity increases with $Q^2$. The expression for sensitivity can be put in the more practical form of volts per kc/s of frequency shift given as follows:

$$\text{Sensitivity} = R_D \frac{n}{1+n^2} \frac{Q}{f_0} \frac{1}{(1+b^2/4)^{\frac{1}{2}}} \times 10^{-3} \text{ volts/kc/s/milliamp input current.}$$

Where $R_D$ is the dynamic resistance of either circuit alone in kilohms, $f_0$ is the centre frequency in Mc/s.

In conclusion, it is necessary to estimate the damping effect of the two diode detectors on the tuned circuits. The voltage across two load resistors $R$ is $E_{d1}$ and $E_{d2}$ assuming 100 per cent rectification efficiency; the power absorbed is therefore $\frac{E_{d1}^2}{R} + \frac{E_{d2}^2}{R}$. The equivalent circuit for the discriminator transformer is shown in Fig. 8.28; $R_p$ and $R_s$ are hypothetical resistors in parallel with the

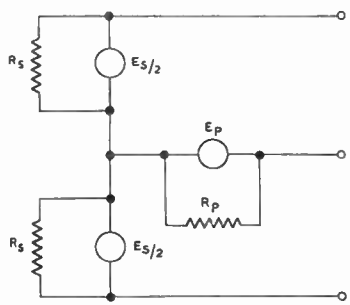



FIG. 8.28.—Equivalent circuit for determining damping of primary and secondary circuits of Foster-Seeley discriminator.

voltage sources which would absorb the same power as the diode loads. It follows therefore that

$$\frac{E_{d1}^2 + E_{d2}^2}{R} = 2\left(\frac{E_s}{2}\right)^2 \frac{1}{2R_s} + \frac{E_p'^2}{2R_p}.$$

The factors of 2 in the denominators of the right hand side are necessary because $E_s$ and $E_p'$ are peak values; in power relationships they must be replaced by $\frac{E_s}{\sqrt{2}}$ and $\frac{E_p'}{\sqrt{2}}$ respectively, the r.m.s. values.

But $$E_{d1}^2 + E_{d2}^2 = 2(E_p'^2 + E_s^2/4).$$

For these two equations to be true for all values of $E_p'$ and $E_s$,

$$R = 2R_s = 4R_p.$$

Since $R_s$ is shunted across only half of the secondary circuit, its value when transformed to be across the whole circuit is $2R$; since there are two such loads, the final load across the whole secondary is $R$, i.e. the same value as it would have in the absence of the primary voltage signal, half the total d.c. load of the diodes ($2R$). For the primary, the equivalent damping resistor $R/4$ must be transformed in the ratio of $1/a^2$, to give the damping on the whole primary circuit, $R/4a^2$. This may differ appreciably from the secondary damping, and hence lead to unequal primary and secondary $Q$ values. If $a=0{\cdot}5$, i.e. only half the primary voltage is employed, the primary damping is $R$, and thus equal to the secondary damping, restoring equality of $Q$ values. The seriousness of the unequal damping effect is, of course, dependent upon the relative values of $R$ and $R_D$, the undamped dynamic resistance of the coupled circuits individually. If comparable, and $a$ does not equal $0{\cdot}5$, it is necessary to introduce additional physical resistance damping across one circuit to equalise the $Q$ values. For example, if $a=1$, and $R=2R_D$ which are realistic circuit values, the primary $Q$ value will be reduced by a factor of $\frac{1}{3}$, whilst that of the secondary will be reduced by $\frac{2}{3}$, and the primary to secondary $Q$ values, assuming initial equality will be now in the ratio 1 : 2. If the value of $a=0{\cdot}5$ is chosen to eliminate this effect, the value of $b$ can be maintained constant at its previous value by choosing $L_p/L_s=4$; however this may introduce further difficulties in obtaining equal $Q$ values for the two values of inductance. Alternatively, given $a=0{\cdot}5$, $L_s/L_p=1$, the values of $b$ and $n$ may be chosen to satisfy the criterion for linearity and $b=2n$. With these limitations, approximate values are $b=2{\cdot}8$, $n=1{\cdot}4$.

The diode loading differs appreciably from the case discussed above when the secondary circuit centre tap is derived by dividing the capacitance branch. Where this is done, the diodes of necessity must be of the shunt-fed type, and a circuit arrangement for this condition is shown in Fig. 8.29. The secondary tuning capacitors $C$ are equal in value, and $a$ is given by $C_1/(C_1+C_2)$. The equivalent circuit is as shown in Fig. 8.28; in addition to the power dissipated by the d.c. outputs of the diodes in the load resistors $R$, these resistors can also be "seen" directly by the generators. The result is that, to the power dissipated by d.c., $\frac{E_{d1}{}^2}{R}+\frac{E_{d2}{}^2}{R}$, must be added

$\frac{E_{d1}^2}{2R}+\frac{E_{d2}^2}{2R}$, the factors of 2 being necessary because it is the r.m.s. values of $E_{d1}$ and $E_{d2}$ which are required to calculate power dissipation. With same notation as previously, therefore,

$$\frac{3}{2}\frac{E_{d1}^2}{R}+\frac{3}{2}\frac{E_{d2}^2}{R}=2\left(\frac{E_s}{4}\right)^2\frac{1}{2R_s}+\frac{E_p'^2}{2R_p}$$

and

$$E_{d1}^2+E_{d2}^2=2\left[\left(\frac{E_s}{4}\right)^2+E_p'^2\right]$$

and, therefore,

$$R_s=\frac{R}{3},\ R_p=\frac{R}{6}.$$

The value of the equivalent damping resistance across the whole secondary is therefore $2R/3$ (i.e. one third of the total d.c. load,

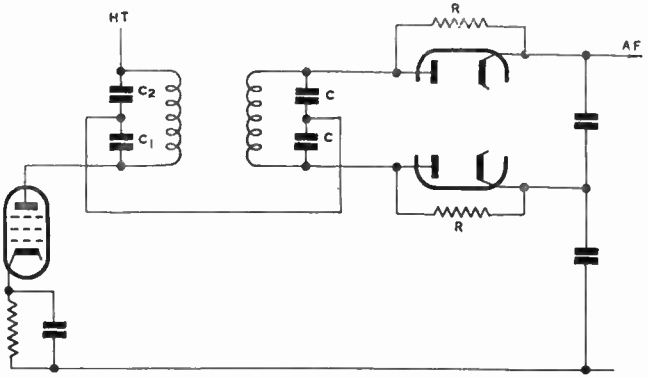



FIG. 8.29.—Foster-Seeley discriminator employing a capacitance tapping of the secondary circuit.

$2R$), and of the equivalent damping resistance across the primary, $\frac{R}{6a^2}$. For equal primary and secondary damping, $a=0{\cdot}5$ as before.

## Practical Design Considerations

In the construction of a practical discriminator transformer, a number of factors must be observed. Firstly, the secondary circuit centre-tap must be situated at the electrical centre of the circuit; in the case of a tapped inductance branch, where the secondary winding comprises a single continuous winding, this may be at an appreciable distance from the physical centre.

This difficulty may be overcome by winding the coil in two equal sections, one situated on top of the other, or interwound with the other. Where the capacitance branch is tapped, this difficulty does not arise.

Secondly, design is generally based upon the assumption that mutual inductance coupling between the two circuits only is employed; it is, however, extremely difficult to eliminate capacitance coupling between the windings. The result of the presence of such coupling is, in general, to distort the discriminator characteristic. As a first precaution, where co-axial windings are employed, the end of the primary winding nearest to the secondary winding should be returned to the h.t. supply, and not to the driving valve anode. Where the best possible performance is desired, it may be necessary to add an electrostatic screen between the windings; this may take the form of a flat spiral of wire between the windings, earthed at one end, or alternatively a mesh of parallel wires joined together at one end only to a further single earthed conductor.

The signal to noise ratio may also be appreciably degraded if the loads of the two diodes are not accurately balanced under dynamic as well as static conditions; this means that the diode capacitors must be equal. It is particularly important to note that the two discriminator reservoir capacitors do not form the whole of the capacitance shunting the diode loads. In practice the diode load system is rarely balanced with respect to ground. Under these conditions it is almost certain that the stray capacities across the two loads will be unequal. This is well illustrated by Fig. 8.30, which shows two oscillograms recorded by Landon. They depict the signal demodulated by the discriminator when supplied with an impulsive wave-train. In the first case it is accurately aligned and the loads and their shunt capacities are carefully balanced. In the second case the capacity across the two diode loads has been deliberately unbalanced by, it is claimed, only 10 micromicrofarads across 100,000 ohms.

In connection with this type of unbalance one important source of trouble is worth noting. If Fig. 8.18 is again referred to it will be noted that as far as the audio frequency side of the discriminator load circuit is concerned the coupling condenser $C_2$ is effectively shunted across the lower diode load, but not across the upper—the impedance of the radio frequency choke is very small

at audio frequencies. To balance up the dynamic impedances of the two loads an additional capacity should be added across the upper diode load. The exact value of the extra capacity can best be determined by oscillographic tests.

Finally, it should be noted that the phenomenon of distortion arising from the a.c. load of the detector differing from that of the d.c. load can also arise. This effect in a.m. detectors is well known; if the diode d.c. load is $R$, and this is shunted by a coupling network $C$, $R_1$, the maximum modulation depth which can be handled

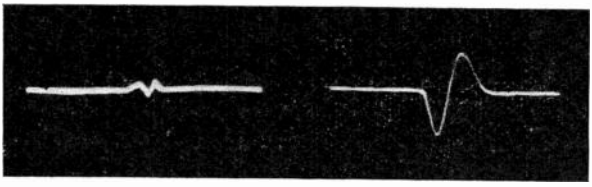


FIG. 8.30.—Signals demodulated when a discriminator is supplied with an impulsive wave-train. In the first case the two diode loads and their shunt capacities are accurately balanced, whilst in the second the capacities have been unbalanced.

(*By courtesy of "Electronics".*)

before distortion occurs is given by $R_1/(R+R_1)$. It will be apparent, however, that the same effect can arise with the type of f.m. detector described above. In this case, however, the phenomenon is not entirely due to the fact that the a.c. and d.c. loads are different, but also that the loads presented to the diodes at frequencies other than the centre frequency may differ from that at the centre frequency. At the centre frequency, the output voltage is zero, and any load may be connected between the a.f. take-off point and earth without affecting conditions at the diodes.

Under working conditions, however, additional damping is applied to the discriminator transformer when modulation is applied, and this may lead to distortion of the discriminator characteristic if the effective $Q$ values of primary and secondary circuits alter appreciably during the modulation cycle.

In this connection it should also be noted that distortion can arise due to de-emphasis components connected directly across the discriminator output; care should therefore always be taken to ensure that the impedance of any network connected to the discriminator output is very large compared with that of the diode loads proper.

### Self-Limiting Phase-Difference Discriminators

The self-limiting type of phase-difference discriminator depends basically for its action upon the properties of outer control electrodes of a multi-grid valve.

If, for example, the suppressor grid of a pentode valve is considered, it will be found that, provided that the control grid and screen grid voltages are maintained constant, the total cathode

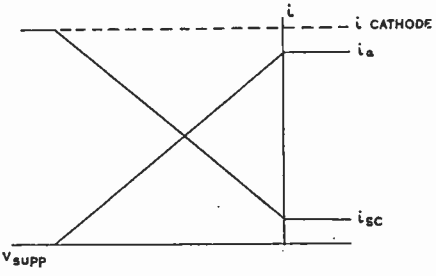



FIG. 8.31.—Anode, screen-grid and cathode currents of an "ideal" pentode with variation of suppressor grid bias.

current through the valve is substantially independent of the suppressor grid bias. The cathode current is determined almost solely by the screen grid and control grid potentials of the valve. The effect of biasing the suppressor grid negatively is to set up a retarding electric field in the valve between screen grid and anode, and hence a proportion of the electron stream which would otherwise have reached the anode returns to the screen grid. Thus, increasing the suppressor grid bias increases the screen grid current, and decreases the anode current, the sum of the two remaining substantially constant; this is shown in Fig. 8.31. At the ultimate limit, anode current is cut off, and the whole of the cathode current flows to the screen.

When the suppressor grid is driven positive, the anode current does not increase appreciably beyond its value at zero suppressor bias; this is due to the fact again that the suppressor grid cannot appreciably influence the total cathode current. The screen grid current under these conditions tends to decrease only slightly, since its current is due largely to collection of electrons by the obstruction it presents. The suppressor grid will, of course, take current in this region, but by careful design, this can be held at a very low value, so that appreciable input damping does not occur.

Thus if the suppressor grid of a pentode is biased mid-way to cut-off, and an input signal is applied, the anode current is a copy of the input signal at low signal levels, provided, of course, that the anode current-suppressor grid bias characteristic is linear. At high signal levels, where the input signal drives beyond cut-off and into the region of positive bias, the anode current wave-form tends to a square wave shape, and the anode current wave-form becomes progressively more nearly independent of the amplitude of the input signal. This is shown diagrammatically in Fig. 8.32 (*a*), (*b*) and (*c*).

In order to achieve demodulation of an f.m. signal, two grids having characteristics similar to those described above are required. The control grid of a pentode cannot normally be used since its limiting action at positive grid bias is generally poor, and grid current damping is generally severe. The circuit arrangement therefore normally employs a nonode valve of the 6BN7 (EQ80) type in which two grids, $g_3$ and $g_5$ function as described above. The description above requires modification in that each grid controls not the anode current direct, but the proportion of the total space current transmitted onwards through the valve. The limiting action is similar, but whereas the anode of a pentode receives all the electrons which pass through the suppressor grid, in a nonode only a proportion reach the anode. Since the two electrodes to be employed are not in the vicinity of a large space charge, input damping on positive grid excursions is not excessive. The inputs applied to the two grids are derived from the primary and secondary windings of an i.f. transformer, as shown in Fig. 8.33. We shall assume that both grids are driven into the regions beyond cut-off and zero bias, and that therefore the anode current wave-form due to the input at each grid separately comprises square waves. It was shown earlier that the primary and secondary voltages of a coupled pair are given by

$$E_s = -X_{cp} X_{cs} \omega M / (Z_p Z_s + \omega^2 M^2)\, i,$$

$$E_p = X_{cp}^2 Z_s / (Z_p Z_s + \omega^2 M^2)\, i.$$

Obviously, anode current can only flow when both grids are positive simultaneously, as shown in Fig. 8.34. At resonance, the voltages are in quadrature, and therefore anode current flows for 90° of the carrier cycle.

Since we are interested only in the relative phase angles of the two voltages, it is permissible to ignore phase shifts common to

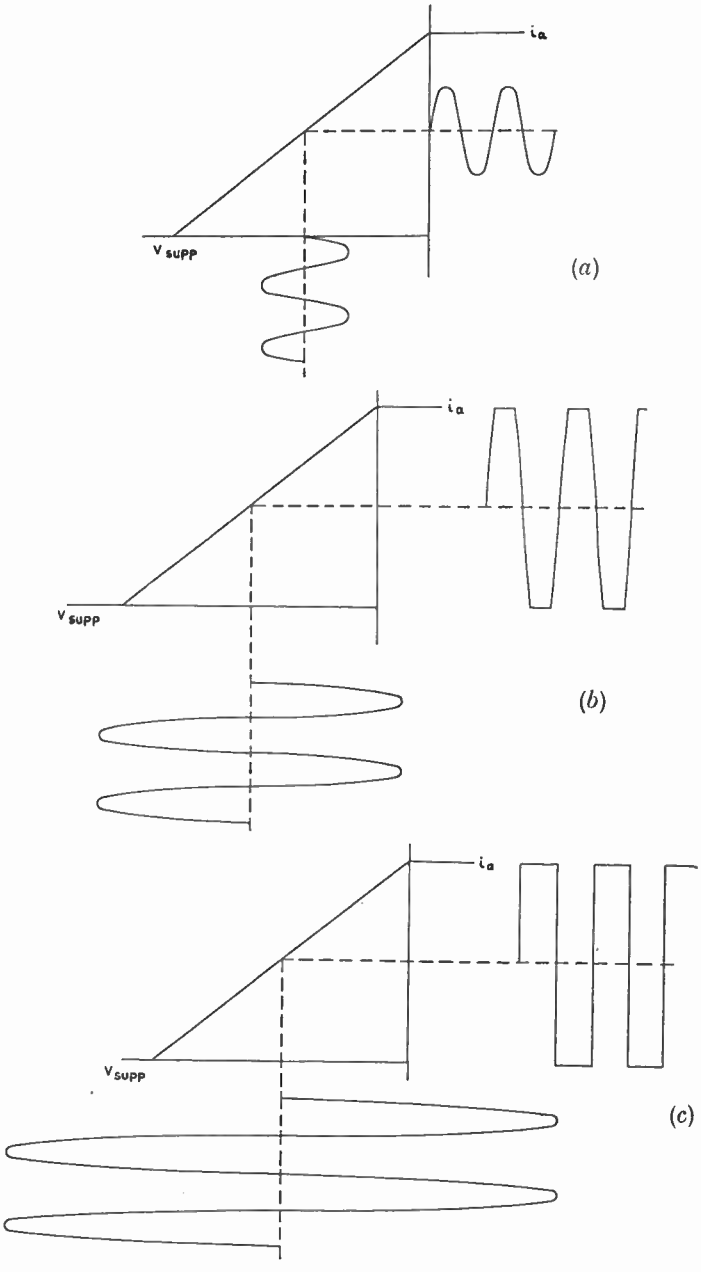



FIG. 8.32.—Showing limiting of anode current wave-form with large input signals to pentode suppressor grid.

both; also, since variations of amplitude of either or both signals are assumed to have no effect on the anode current, we can ignore most of the terms describing the primary and secondary voltages,

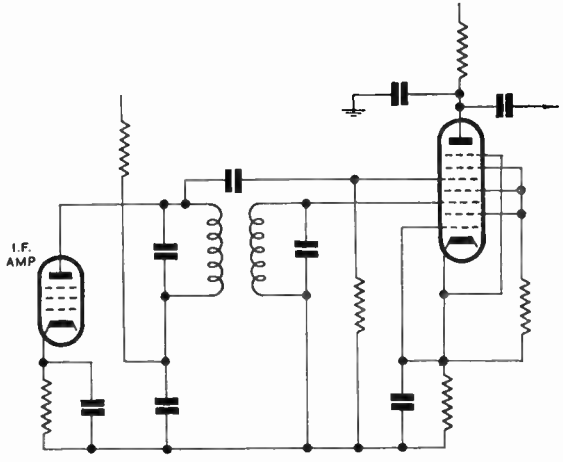



FIG. 8.33.—F.M. limiter-discriminator, employing nonode valve.

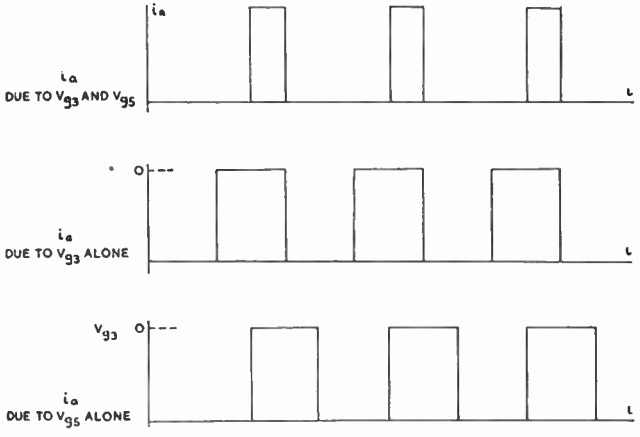



FIG. 8.34.—Anode current in nonode, with inputs to $g_3$ and $g_5$ displaced in phase.

and consider only the phase angle $\phi$ of the primary voltage relative to the secondary voltage. This is given by

$$\phi = 90 + \tan^{-1} Qx,$$

assuming identical primary and secondary circuits, and $x = \dfrac{2\delta f}{f_0}$,

where $\delta f$ is the departure of the carrier frequency from the centre frequency.

It will be noted that the vertical edges of the anode current wave-form due to the signal at either grid alone, are co-incident with the instant at which the signal goes to zero; the period of conduction per cycle due to the signals at both grids is therefore $\frac{\phi}{360}$ of the whole time of one cycle of the carrier wave. The mean value of the anode current, therefore, is directly proportional to $\phi$, and by including a suitable load resistor and integrating capacitor in the anode circuit, a demodulated output is obtained, the magnitude of which is proportional to $\phi$ and independent of the amplitude of the input signals.

Since $\phi = 90 + \tan^{-1} Qx$, the anode current has a d.c. component, equal to $90/360 = \frac{1}{4}$ of its value when both grids are at zero bias. More importantly, there is an a.c. component, the magnitude of which is proportional to $\tan^{-1} Qx$. If, therefore, the frequency deviation of the incoming signal is such that $Qx$ is small, we can make the approximation that $\tan^{-1} Qx = Qx$, i.e. the output amplitude is proportional to frequency deviation.

It will, however, be noted that the output frequency deviation characteristic is not truly linear anywhere, and this limits the usefulness of this type of discriminator. However, the discriminator has the great practical advantage of a high audio output. At an i.f. of 10·7 Mc/s with an anode load of 470 kilohms and an h.t. supply of 250 volts, the peak output for 75 kc/s deviation is about 30 volts. This may be compared with that of a Foster-Seeley discriminator, for which a typical figure may be taken as 5 volts. The linearity for a given frequency deviation and i.f. can be increased by reducing the $Q$ values of the tuned transformer; this, however, leads to lower i.f. gain, and hence raises the receiver input signal necessary for efficient limiting.

As the limiting action does not depend upon circuit time constants, as with the grid limiter, the limiting action is instantaneous, and so "paralysis" of the receiver due to signal surges at the limiter cannot occur. In practice it is general to feed the screen grids $g_2$, $g_4$, $g_6$ from a low resistance potential divider across the h.t. supply; this prevents the voltages at these grids rising when input signals are applied, and hence ensures full limiting efficiency. In order to keep the cut off bias at $g_3$ and $g_5$

to small values, to ensure a low limiting threshold, the voltage at $g_2$, $g_4$, and $g_6$ is kept to a low value, about 20 volts. The control grid $g_1$ is not employed in this circuit arrangement; it is normally biased so that the electrode dissipations are kept within the limits prescribed. In the circuit of Fig. 8.33, $g_1$ is connected direct to cathode; the cathode bias for grids $g_3$ and $g_5$ is largely determined by the bleeder current, and is hence largely independent of any change in electrode potentials under operating conditions.

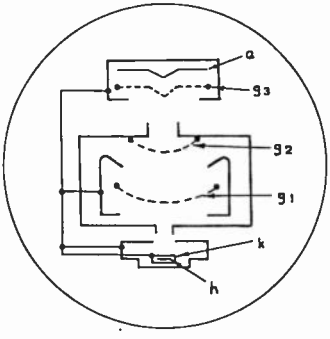



FIG. 8.35.—Plan view of 6BN6 type of gated beam tube.
*(By courtesy of S.T.C.)*

A similar type of circuit is employed with the gated beam tube of the 6BN6 type. This valve is basically a pentode, and the two signal inputs are applied to the control grid and suppressor grid. An electron lens technique is employed in the construction of the valve, to achieve the desired limiting characteristics at the grids; a plan view of its structure is given in Fig. 8.35. The cathode is surrounded by a focusing shield, connected to the cathode. Through an aperture in the shield the electron stream enters a second enclosure containing the control grid, where it is accelerated by the action of the accelerator grid $g_2$, the action of which corresponds to the screen grid of a normal pentode valve. The control grid is relatively isolated from the cathode, and hence from the cathode space charge. Thus on being driven positive, the grid current is low, the minimum input impedance being of the order of 20 kilohms. Because of the effect of the enclosure, the anode current does not increase greatly as the grid is driven positive. From this second enclosure, the electron stream passes through another aperture to the enclosure containing the anode and second control grid. Due to the valve construction, when this

latter grid is biased negatively, that part of the electron stream which would, in a normal pentode, return to the screen grid, is collected by the enclosure walls, which are at cathode potential. Also, since the accelerator grid $g_2$ has comparatively little influence on the field in the space before the second control grid, the electron stream approaching the second control grid does so at relatively low velocity, and hence the second control grid has a relatively short grid base without an unduly heavy mesh.

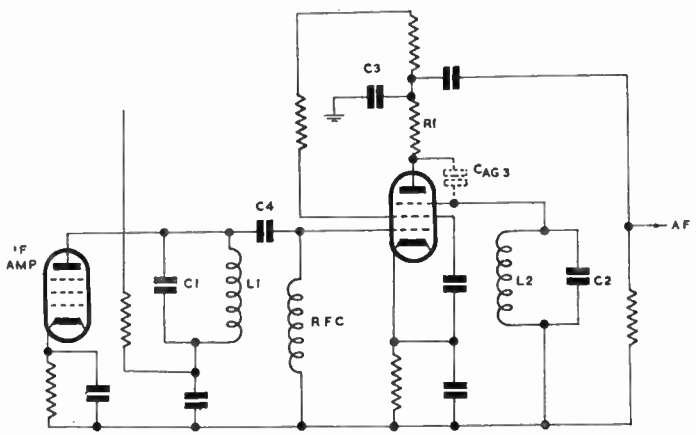



FIG. 8.36.—F.M. limiter-discriminator employing gated beam valve.

A circuit arrangement employing this type of valve is shown in Fig. 8.36. In this circuit, the "primary" circuit $L_1$, $C_2$ of the tuned transformer is not coupled directly to the "secondary" circuit, $L_2$, $C_2$ but by the valve electron stream. The mechanism is as follows. The output of the tuned circuit is applied to $g_1$; coupling capacitor $C_4$ is employed to isolate the anode voltage of the preceding valve, and an r.f. choke is used to connect the grid to ground, this latter being used in preference to a resistor, since a resistor would produce variations of bias under operating conditions. The anode current is therefore in phase with the grid voltage, and a voltage is thus developed across $R_1$ in anti-phase with that across $L_1$, $C_1$. The anode/second control grid capacitance, shown dotted, feeds this signal to the tuned circuit $L_2$, $C_2$.

By varying $R_1$, the r.f. gain from the first control grid to anode can be varied. In this way, an ample voltage swing can be produced to ensure efficient limiting at $g_3$. The performance of the

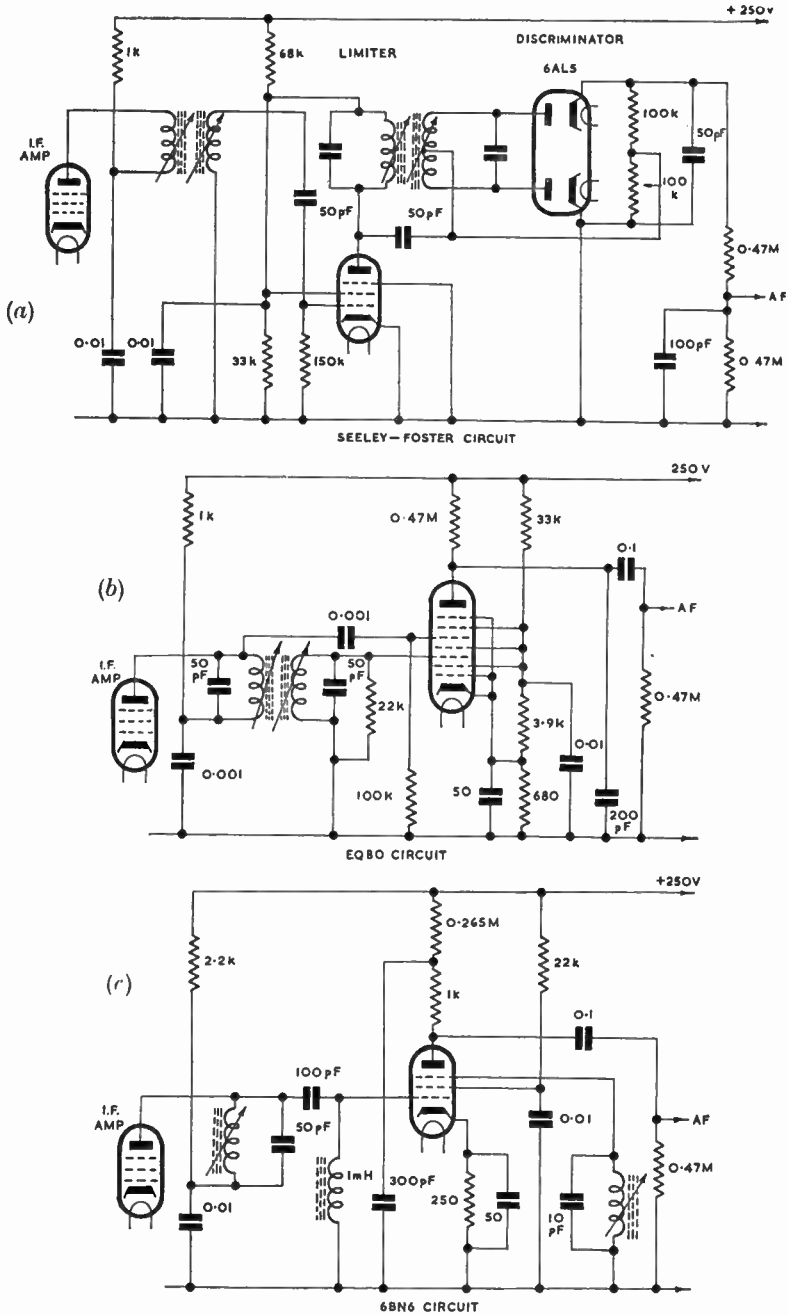



FIG. 8.37.—Practical f.m. discriminator circuits: (*a*) Foster-Seeley, (*b*) EQ80 nonode, (*c*) 6BN6 gated beam valve.

*(By courtesy of S.T.C.)*

circuit is almost identical with that of the nonode discussed earlier in respect of the linearity of its frequency swing-output characteristic, i.e. it is not truly linear anywhere, but has a good approximation to linearity over a limited range. This range can of course be extended by employing lower $Q$ values in the tuned circuits. This, however, leads to a raising of the receiver input signal level at which limiting occurs.

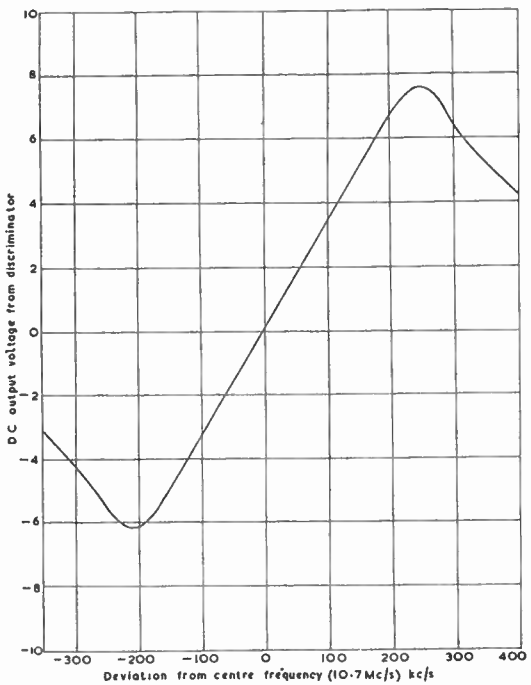
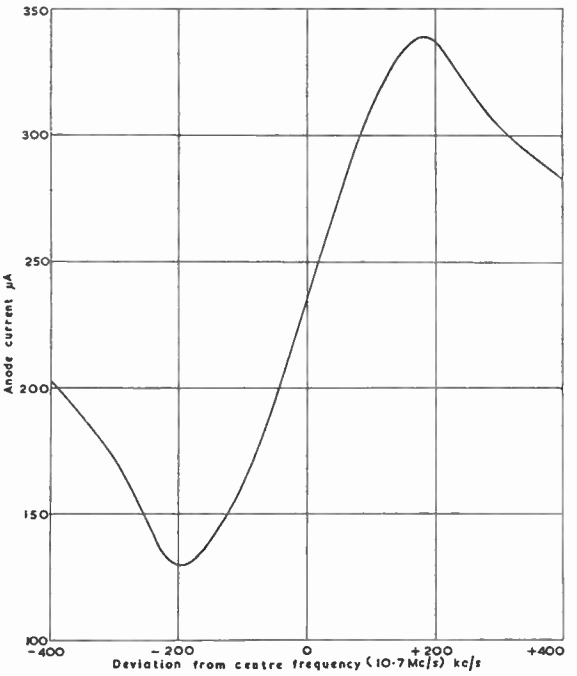


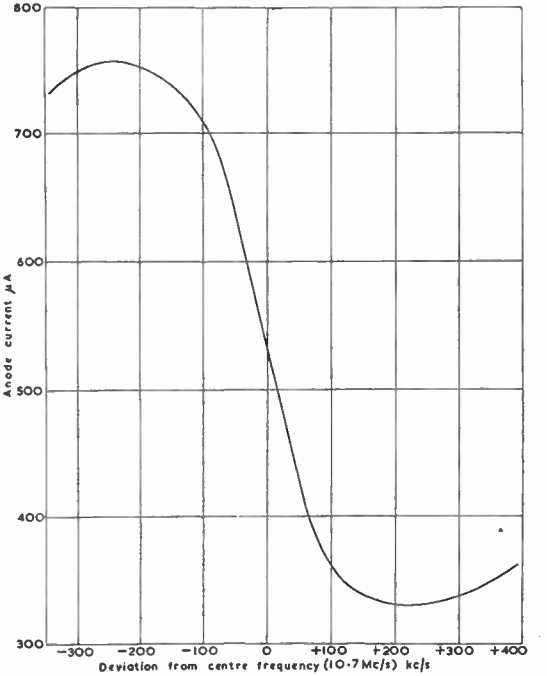

FIG. 8.38.—Response curves of circuits of Fig. 8.37.
(*a*) Foster-Seeley.

In common with the nonode valve, the gated beam tube gives a high level audio output; with an anode load resistor of 250 kilohms, and fed from a supply of 250 volts, the peak output for 75 kc/s deviation is about 40 volts.

In order to provide a basis of comparison of the three types of phase difference discriminator so far described, typical operating circuits and performance curves are shown in Figs. 8.37, 8.38, and 8.39. It will be seen that, in respect of linearity of discriminator characteristic and of a.m. rejection, the Foster-Seeley circuit is the best.




(*b*) EQ80 nonode.




(*c*) 6BN6 gated beam tube.

FIG. 8.38.—Response curves of circuits of Fig. 8.37.
(*By courtesy of S.T.C.*)

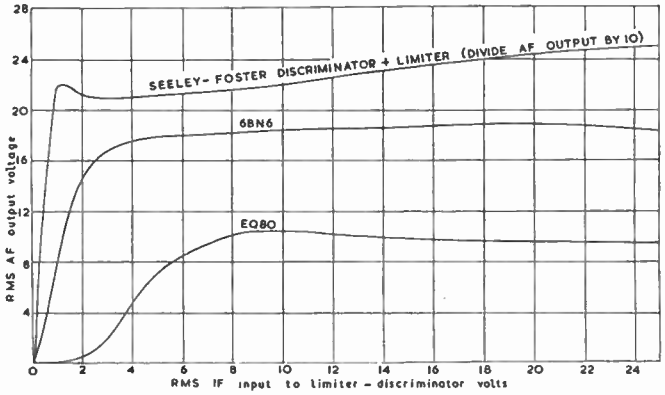



FIG. 8.39.—Limiting characteristics of circuits of Fig. 8.37.
*(By courtesy of S.T.C.)*

However, the higher sensitivity of the others gives the practical advantage that an a.f. amplifying stage may be saved in the receiver.

The circuit of the Foster-Seeley discriminator differs from that of Fig. 8.18 in that the voltage from the primary circuit is injected across the diode load resistors which are effectively in parallel to the r.f. input. It will be noted, therefore, that the primary circuit has additional damping imposed, the equivalent resistor being equal to the sum of the load resistors in parallel.

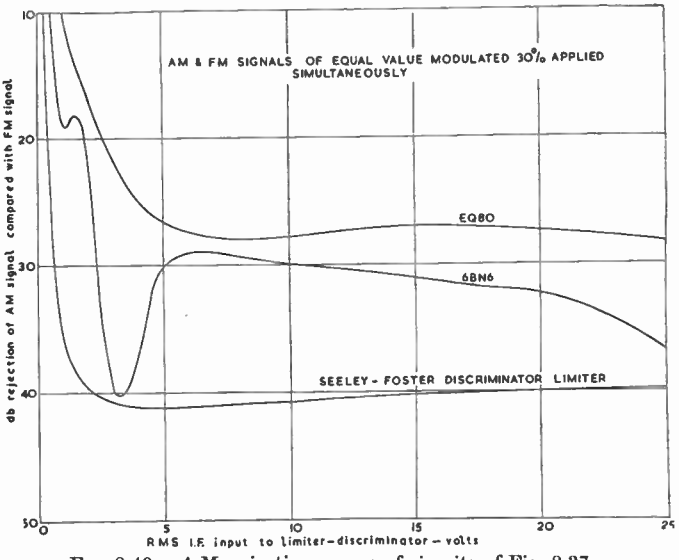



FIG. 8.40.—A.M. rejection curves of circuits of Fig. 8.37.
*(By courtesy of S.T.C.)*

## Frequency Counters

In the case of sub-carrier frequency modulation transmissions, the frequency of the signal is often too low to permit the use of conventional discriminator circuits. Alternative methods of demodulation have therefore to be adopted. While there are many possible circuits available, variations on that outlined in Fig. 8.41 are the most commonly employed.

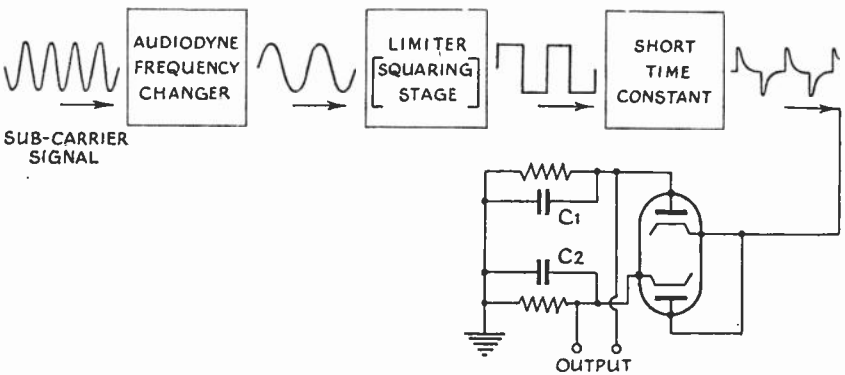



FIG. 8.41.—Illustrating the general arrangement of the frequency-counter circuit used to demodulate sub-carrier frequency modulated signals.

The sub-carrier, which normally lies within the audio range, is first changed to such a frequency that at the peak sub-carrier deviation the frequency of the resultant signal is zero. It now follows that as the sub-carrier is frequency modulated the resultant will vary from zero cycles up to double the peak deviation frequency. This signal is next passed through a limiter stage which also squares up the wave-form, which is then passed through a filter having a short time constant. The signal emerging from this stage takes the form of a series of pulses which are rectified by means of a pair of back-to-back diodes.

The two condensers $C_1$ and $C_2$ function as reservoirs, with the result that the voltage output is directly determined by the number of pulses per second, which are in turn dependent upon the frequency of the original sub-carrier signal. Providing the time constant of the pulse-shaping network is such that the voltage has returned to zero before the start of the following pulse, counter-circuits of this type are capable of giving a perfectly linear relationship between applied frequency and voltage output.

**Dynamic Limiters**

The limiters described earlier all suffered from one major disadvantage, that of a fixed threshold below which limiting action ceased. This implies that there is a fixed input signal level below which the receiver cannot operate satisfactorily at all. Further, the input signal must exceed this threshold value by a substantial amount to achieve satisfactory limiting, as under conditions of multi-path reception and severe interference the instantaneous value of the carrier amplitude may fall well below its mean level. In fact, the greatest amplitude modulation depth that a receiver employing such a limiter will handle is obviously the ratio of the difference between the signal mean amplitude and the minimum amplitude required for efficient limiting, to the signal mean amplitude; this is thus a variable quantity.

The family of dynamic limiters function by providing variable damping of a tuned circuit, the value of the equivalent damping resistor varying in such a way as to maintain constant the output signal amplitude. It is shown in its simplest form in Fig. 8.42, where a diode in series with a battery is connected in parallel with a tuned circuit. The battery may be assumed for the present purpose to have zero internal impedance, whilst its voltage is assumed equal to the peak signal amplitude. Under these conditions, the diode will not take current whilst the signal amplitude remains steady. If, however, the signal amplitude tends to increase, the diode will conduct, and assuming a perfect diode, the tuned circuit will be heavily damped, the equivalent damping resistor being such that the signal amplitude increases by an infinitely small amount. If, however, the signal amplitude tends to decrease, the diode will be cut-off. Therefore the circuit will limit perfectly on outward swings of modulation, and not at all on inward swings. This situation can be remedied if in series with the diode is connected a paralleled combination of a resistor $R$ and a large capacitor $C$ instead of the battery, as shown in Fig. 8.43. The value of $R$ is chosen so that its resistance is small compared with the tuned circuit dynamic resistance, and therefore the tuned circuit is normally heavily damped by the diode circuit, the equivalent damping resistance being $R/2$.

The voltage across the parallel combination of the resistor $R$ and the capacitor $C$ will be equal to the peak signal amplitude,

and because of the long time constant of the combination, behaves in a manner similar to an equivalent battery. If the signal amplitude tends to increase, the diode will take a heavy current on signal peaks. Since the voltage across the capacitor cannot be readily increased, this is equivalent to reducing the value of the equivalent damping resistor, so that under dynamic conditions, the signal amplitude increase is negligibly small. If the increase in signal

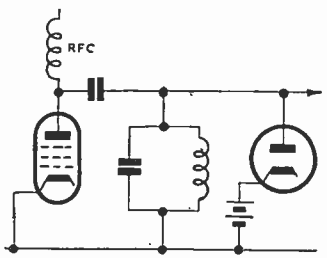



FIG. 8.42.—Simple basic form of dynamic limiter.

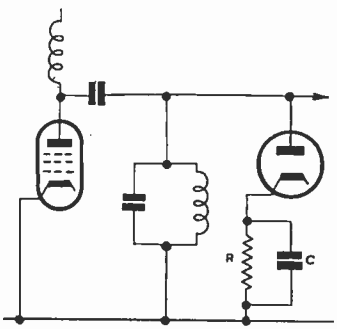



FIG. 8.43.—Practical form of Fig. 8.42.

amplitude is in the nature of a long term change, the capacitor will ultimately change to the new peak value of the signal, and thereby automatically restore the equilibrium conditions.

If, however, the signal amplitude tends to decrease, the diode ceases to conduct, and the damping supplied by the diode circuit is removed. The gain therefore rises, and the reduction in signal amplitude is therefore offset. The degree of "downward" modulation which can be restored in this manner is obviously conditioned by the ratio of the equivalent damping resistance due to the diode circuit under equilibrium conditions and the dynamic resistance of the tuned circuit.

The manner in which the downward modulation depth can be calculated can be seen from Fig. 8.44. The curve of $E=R_d i$ gives the voltage output when the limiter is not in circuit, where $R_d$ is the dynamic resistance of the tuned circuit. The curve of $E=iR_dR'(R_d+R')$ gives the output with the limiting diode in circuit, using a small value capacitor, permitting the rectified voltage to follow rapid changes of the peak input current $i$. $R'$ is the equivalent damping resistor due to the diode circuit,

and if $R$ is the actual magnitude of the diode load resistor $R'=R/2$. With a large value of capacitor, and a steady input current $i_0$, the working peak voltage output is $E_0$; this is necessarily the same as that with a small capacitor under static conditions. When, however, a rapid change of $i_0$ occurs in the downward direction the curve of $E_0$ follows the horizontal dotted line until it rejoins the line $E=R_d i$, at $i=i_1$. For values of $i$ below $i_1$, the output will

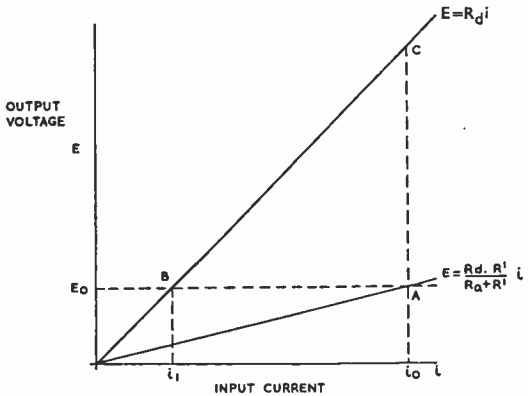



FIG. 8.44.—Method of determining the greatest downward modulation depth which a dynamic limiter will handle.

decrease. Thus the maximum downward modulation depth which the limiter will handle is given by $\frac{i_0-i_1}{i_0}$. From the geometry of the figure this is obviously equal to

$$\frac{R_d i_0 - R_d R'/(R_d+R')i_0}{R_d i_0} = 1 - \frac{R'}{R_d+R'} = \frac{R_d}{R_d+R'}.$$

Thus for efficient downward modulation limiting, $R'$ must be small compared with $R_d$. The expression for the maximum downward modulation depth may also be put in the form $1-Q'/Q$, where $Q$ refers to the circuit in the undamped condition, and $Q'$ to its working condition.

If, for example, the equivalent damping resistance $R'=R/2$ is equal to one ninth of the dynamic resistance $R_d$ of the tuned circuit alone, the load presented to the previous valve is one tenth of the dynamic resistance $R_d$, under equilibrium conditions. With this damping removed, the signal level must fall by a factor of ten before the output signal level falls below the equilibrium

level, and hence, a dynamic limiter designed on such a basis would handle the equivalent of 90 per cent amplitude modulation of the input signal, and transmit a negligibly small quantity to the output. It will be realised, of course, that if the reduction in signal amplitude is in the nature of a long term change, the capacitor will discharge until equilibrium conditions are restored.

The advantages of this type of limiter will thus be apparent; its operation is independent of input signal amplitude, down to the level at which the diode can no longer be considered a very small resistance on the charging stroke. Further it can be designed to reject amplitude modulation of the signal up to a pre-determined depth, which property holds at all levels of the input signal, with the same qualification as before as to the lower limit set by the properties of the diode. Its disadvantages are the absence of a fixed output level, and consequently, long term variations of input signal level cannot be rejected. Also the variable damping of the tuned circuit under operating conditions means that its passband must equally be variable, and the circuit cannot be relied upon to provide selectivity in the receiver. The long term variations of signal level can, however, be reduced to tolerable proportions by the employment of an efficient a.g.c. system, which additionally tends to remove the disadvantage of variations of output signal between two signals of unequal amplitude. The limiting action, since it does not depend upon the rapid charge or discharge of a capacitor, does not therefore suffer from the "blocking" effect encountered with grid leak limiters.

The dynamic limiter is frequently incorporated in a discriminator circuit; one such arrangement is shown in Fig. 8.45. The discriminator is of the double tuned circuit type. A tertiary circuit is coupled tightly to the tuned primary, and the dynamic limiter is fed from this winding. The diode may be actually a germanium crystal type, which has a low forward resistance. The dynamic limiter could, of course, be connected in parallel with the tuned primary circuit directly, but the arrangement shown has the advantage that the limiter supplies a.g.c. voltage.

The limiting action of the circuit is as described above. The two secondary circuits, which form the basis of the discriminator, must be very loosely coupled to the primary circuit, so that the primary circuit is comparatively unaffected by these two circuits. The voltage across the primary winding is thus stabilised against

short term signal amplitude variations, and the circulating current in the primary circuit is also stabilised.

The voltage which is injected in series with each secondary circuit is $j\omega Mi_p$, and hence the voltage applied to each discriminator diode is

$$E=j\omega Mi_p/j\omega C_sZ_s=Mi_p/C_sZ_s,$$

where $Z_s=j\omega L_s+1/j\omega C_s+r_s$, the suffix $s$ referring to the secondary circuits; the value of $C_s$ will be different for the two secondary

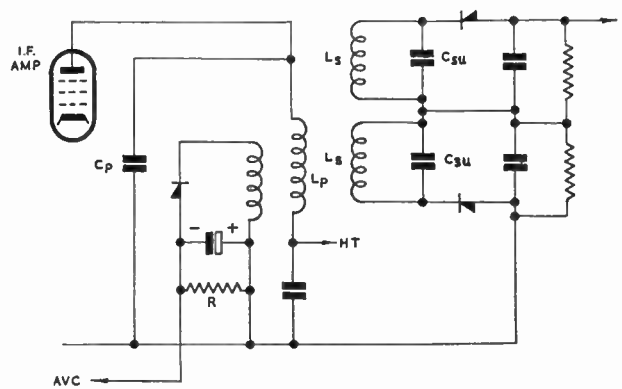



FIG. 8.45.—F.m. discriminator of double tuned circuit type, incorporating dynamic limiter.

circuits individually. The current $i_p$ is related to the anode current of the driving valve by $i_p=Q'i_a$, where $Q'$ is the working $Q$ of the primary circuit.

The voltage applied to each diode when its secondary circuit is resonant is, therefore,

$$E_d=MQ'i_a/C_sr_s$$
$$=\frac{MQ'i_a}{L_s}R_{ds},$$

where $R_{ds}$ is the dynamic resistance of the secondary circuit, and is equal to $L_s/C_sr_s$.

If

$$L_p=L_s,$$
$$E_d=KQ'i_aR_{ds},$$
$$=n\left(\frac{Q'}{Q}\right)^{\frac{1}{2}}i_aR_{ds},$$

where $K=M/L$, and $n=k\sqrt{Q'Q_s}$.

With identical circuits, the undamped primary circuit $Q$ is equal to $Q_s$, and therefore the output voltage is lower by a factor $n(Q'/Q)^{\frac{1}{2}}$ than it would be if the secondary circuit were connected directly in the anode circuit of the driving valve. This apparent loss is, however, offset by the fact that, were the secondary so connected, the driving stage would have to be a conventional limiter, and hence operated under conditions of low gain.

In respect of the discriminator characteristic the circuit performance is identical with that of the conventional circuit described earlier. In order to meet the requirement that the coupling between primary and secondary circuits shall be loose, $n$ should be less than 0·4 of critical coupling.

## The Ratio Detector

The ratio detector belongs to the class of self-limiting discriminators. In its demodulating action, it is closely akin to the Foster-Seeley circuit, which it closely resembles. A circuit diagram of a

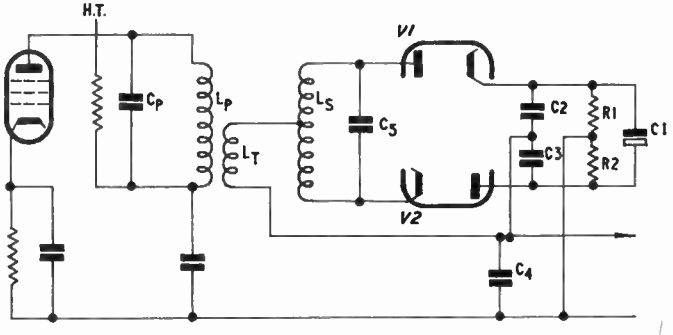



FIG. 8.46.—Circuit of typical ratio detector.

typical ratio detector is shown in Fig. 8.46; if the capacitor $C_1$ is ignored, then the circuit is that of a Foster-Seeley discriminator, with one diode reversed as compared with a conventional arrangement.

The a.f. output is given by the difference of the voltage across the load capacitors $C_2$ and $C_3$. In the circuit shown, the centre point of the loads $R_1$ and $R_2$ is earthed, and the output taken from the junction of the load capacitors $C_2$ and $C_3$. Because of the mode of connection, the a.f. output is actually equal to half of the difference of the voltages across $C_2$ and $C_3$ and in the output is thus

half of that obtainable with the conventional Foster-Seeley arrangement. The tertiary winding $L_t$ is tightly coupled to the primary winding, and provides the voltage normally obtained in a Foster-Seeley circuit directly from the primary winding itself. This form of circuit shown is frequently adopted, since only a small proportion of the primary voltage is usually required, and the tertiary winding provides a convenient means of obtaining this voltage. A small fraction of the primary voltage is employed because of the heavy damping imposed by the diode load resistors; the reasons for this are discussed later.

The shape of the discriminator output-volts/input-frequency is essentially similar to that of a Foster-Seeley circuit. Because the sum of the rectified outputs is substantially constant over a considerable range about the centre frequency, the reservoir capacitor $C_1$ has but little effect on the performance of the circuit when the input signal is free from amplitude modulation. If, however, the signal has an a.m. component, an action similar to that of the dynamic limiter occurs and, provided that the circuit parameters are correctly chosen, the a.m. component produces only a very small output.

It will be seen from inspection of the circuit that the ratio detector has the same inherent property as that possessed by the dynamic limiter, namely, that the degree of downward modulation which can be handled is dependent upon the degree of damping imposed under quiescent conditions. The larger this is, the greater the downward modulation depth which can be handled. Because of this heavy damping, careful adjustment of the primary tertiary ratio and coupling factor to the secondary are necessary to secure adequate sensitivity and high signal level at the diodes.

The voltages applied to the diodes can be found by employing the circuit relationships derived earlier. In dealing with the ratio detector, it does, however, simplify the treatment if the tuning of the primary circuit is ignored. In the ratio detector design, it is not possible to adjust the coupling factor to obtain best linearity as in the Foster-Seeley circuit, these parameters being fixed by considerations of a.m. rejection. Although the selectivity of the primary circuit will produce amplitude modulation of the signal, provided that the a.m. rejection is satisfactory, this will not affect the audio output; in this the ratio detector differs fundamentally from the Foster-Seeley discriminator.

It will be assumed initially that the input signal across the inductor of the primary winding is frequency modulated, and of constant amplitude. The tertiary winding will be assumed to have a coupling co-efficient to the primary of unity, and the voltage across the tertiary will be a fraction $a$ of the primary voltage, given by $a=\sqrt{(L_t/L_p)}$ where $L_t$ is the inductance of the tertiary winding and $L_p$ is that of the primary winding.

The voltage induced in the secondary circuit is given by expression (8.4)

$$E_s=-j\frac{KQ_s\sqrt{(L_s/L_p)}}{1+jQ_sx}E_p$$

where $E_s$ is the voltage across the secondary circuit:

$K$ is the coupling co-efficient between primary and secondary circuits:

$Q_s$ is the $Q$-value of the secondary circuit:

$x=2\Delta f/f$: $\Delta f$ is the departure from the secondary circuit resonant frequency $f_0$:

$E_p$ is the voltage across the primary inductor $L_p$:

$L_s$ is the secondary circuit inductor.

The above expression can be separated into real and imaginary parts as follows:

$$E_s=\frac{E_pKQ_s\sqrt{(L_s/L_p)}}{1+Q_s^2x^2}(Q_sx-j).$$

As explained earlier, heavy damping is usually employed in order that a high degree of downward a.m. can be handled, and this leads to a low working value of $Q_s$; the expression above can thus generally be simplified to

$$E_s=E_p[KQ_s^2x\sqrt{(L_s/L_p)}-jKQ_s\sqrt{(L_s/L_p)}].$$

The equivalent circuit diagram is shown in Fig. 8.47. The vector diagrams for computing the voltages applied to the diode $E_{d1}$ and $E_{d2}$ are shown in Fig. 8.48, and from the figure these voltages are

$$E_{d1}^2=a^2E_p^2\,[(1+ZQ_sx)^2+Z^2]$$

and

$$E_{d2}^2=a^2E_p^2\,[(1-ZQ_sx)^2+Z^2]$$

where $Z=\frac{1}{2}\dfrac{KQ_s\sqrt{(L_s/L_p)}}{a}=\dfrac{\text{half secondary voltage at resonance}}{\text{tertiary voltage}}$.

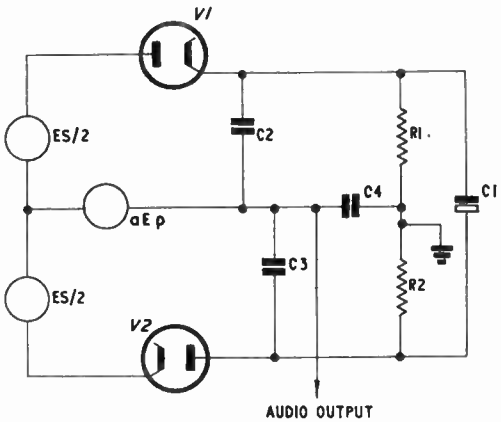



FIG. 8.47.—Equivalent circuit to that of Fig. 8.46.

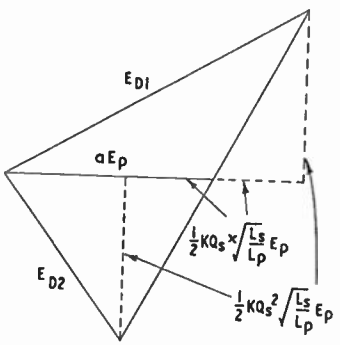



FIG. 8.48.—Vector diagram showing relationships of tertiary and secondary voltages in a ratio detector circuit.

By subtraction, this gives

$$E_{d1}^2 - E_{d2}^2 = a^2 E_p^2 \cdot 2ZQ_s x$$
$$(E_{d1} + E_{d2})(E_{d1} - E_{d2}) = a^2 E_p^2 \cdot 2ZQ_s x.$$

Also, the arithmetic sum of the two voltages $E_{d1}$ and $E_{d2}$ over a limited range of values of $x$ about zero is substantially independent of $x$, and equal to $E_t$,

where $$E_{d1} + E_{d2} = E_t = 2aE_p(1+Z^2)^{\frac{1}{2}} \quad (8.6)$$

whence $$E_{d1} - E_{d2} = a^2 E_p^2 \cdot 2ZQ_s x / E_t$$
$$= aE_p \cdot ZQ_s x/(1+Z^2)^{\frac{1}{2}} \quad (8.7)$$
$$= E_t \cdot ZQ_s x/2(1+Z^2). \quad (8.8)$$

Expression (8.7) shows that the audio output, which is proportional to $E_{d1}-E_{d2}$, is independent of variations of $E_p$ provided that $ZQ_s/(1+Z^2)^{\frac{1}{2}}$ varies inversely with $E_p$. That the latter term does vary with $E_p$ follows from the fact that any change of $E_p$ alters $E_t$, which in turn alters the voltage applied to the load circuit, and hence induces changes of $Q_s$ as the damping varies. Expression (8.8) shows more clearly, however, what happens if the reservoir capacitor is connected across the whole of the d.c. output as shown in Fig. 8.46, and perfect diodes (i.e. having zero forward resistance) are employed. Under these conditions $E_t$ is stablised exactly by the dynamic limiting action of the load circuit, in the manner described earlier. The output is thus only independent of a.m. in the input if any tendency for $E$ to increase produces no change in the value of the r.h.s. of expression (8.8). Since $Z$ contains $Q_s$, this requires that $Q_s^2/(1+Q_s^2Z_0^2/Q_{s0}^2)$ should be constant, where $Q_{s0}$ is the quiescent value of $Q_s$ and $Z_0$ is the value of $Z$ quiescent. This condition is approached if $Z_0^2 \gg 1$. If this condition is satisfied, then the a.f. output to an f.m. input is vanishingly small, and hence this is not a practical condition of operation.

With small values of $Z$, it will be seen that as $E_t$ tends to increase, and $Q_s$ consequently decreases, the a.f. output tends to decrease, since $Q_s^2/(1+Q_s^2Z_0^2/Q_{s0}^2)$ necessarily decreases. Thus with the conditions postulated, over-compensation occurs. If the reservoir capacitor is removed, then the output tends to increase with increasing input. This suggests that this is a condition of operation between the extremes of no dynamic limiting and perfect dynamic limiting at which maximum a.m. rejection exists, and this is in fact so, and is the condition of operation normally encountered. There are three methods commonly employed to give the correct degree of dynamic limiting; the circuit arrangements are shown in Fig. 8.49. Circuits (*b*) and (*c*) are essentially similar, and depend for their action upon the current limiting effect of $R_3$ in (*b*) and $R_3$ and $R_4$ in (*c*). The series resistor $R_3$ in (*a*) achieves a similar effect operating in the r.f. section of the circuit. In examples (*b*) and (*c*), the proportion of the output stabilised by the action of $C_1$ determines the a.m. rejection properties. The fraction stabilised is given for the circuit of (*b*) by $(R_1+R_2)/(R_1+R_2+R_3)$ and for the circuit of (*c*) by $R_a(R_a+R_b)$, where $R_a=R_1+R_3+R_4$ and $R_b=(R_3+R_4)+(R_3+R_4)/(R_1+R_2)$. This fraction is related to a single value of $Z$ (the half-secondary/tertiary voltage ratio) for

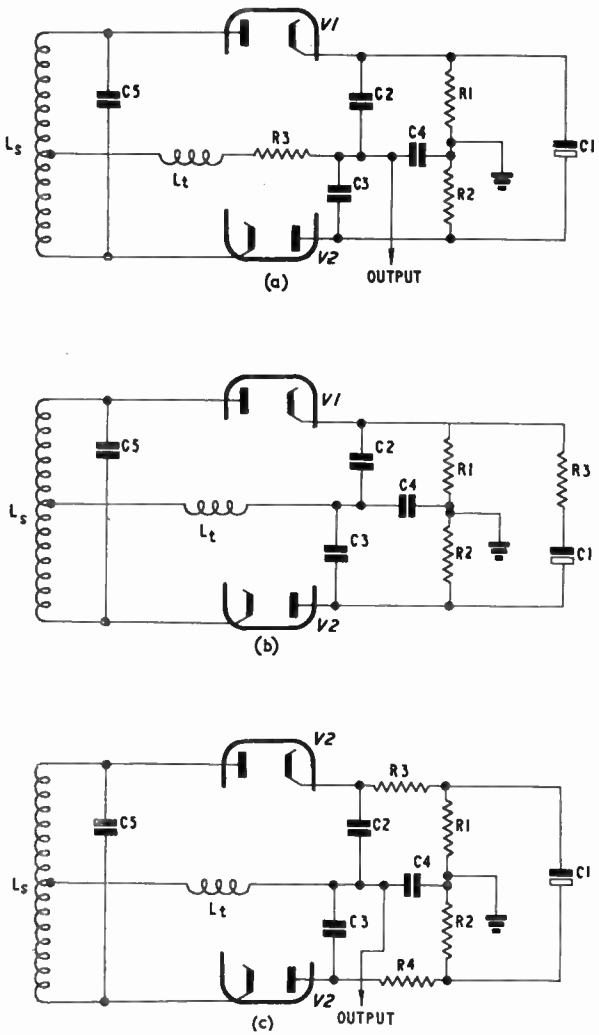



FIG. 8.49.—Three methods of securing maximum a.m. rejection in a ratio detector. Methods (*b*) and (*c*) are essentially similar.

best a.m. rejection. In a practical circuit arrangement where the diode load resistors are much smaller in value than the dynamic resistance of the secondary circuit, the fraction rises almost linearly from 0·5 approximately at $Z=0$, to 1·0 at $Z=1$. Beyond the value of $Z=1$, the total output must be stabilised. In practice,

the adjustment of the fraction stabilised is almost always carried out on an empirical basis. For reasons discussed later, values of $Z$ close to unity are not favoured; and values between 0·5 and 0·9 are generally employed.

It will be seen from expression (8.8), that for a given value of $Z$ and $Q_s$ the maximum a.f. output occurs when $E_t$ is a maximum. This occurs at one value of $a$ only, for which the equivalent resistance in parallel with the primary winding due to secondary circuit losses and diode damping equals that dynamic reactance of the primary winding itself, i.e. power matching occurs. The optimum value of $a$ may be derived as follows. If $E_t$ is the peak value of the sum of $E_{d1}$ and $E_{d2}$, then the voltage across the load resistor is given by $\eta E_t$, where $\eta$ is the rectification efficiency. The power dissipated in the d.c. load ($R_{dc}$) ($R_1+R_2$ of Fig. 8.48 (*b*) and $R_1+R_2+R_3+R_4$ of Fig. 8.49 (*c*)) is thus $\eta^2 E_t^2/R_{dc}$. Since $E_s=E_t Z/2(1+Z^2)^{\frac{1}{2}}$, the power dissipated in the secondary circuit itself is $E_t^2 Z^2/8(1+Z^2)R_{ds}$, where $R_{ds}$ is the dynamic resistance (the further factor of 2 is necessary since $E_t$ is a peak value).

Thus the total power dissipated is given by

$$P=E_t^2[\eta^2/R_{dc}+Z^2/8(1+Z^2)\,R_{ds}].$$

This power loss can be equated to that of a fictitious resistor $R_{eq}$, connected across the primary circuit, the power dissipated being $E_p^2/2R_{eq}$ (assuming $E_p$ to the peak value). Equating the expressions

$$\frac{E_p^2}{2R_{eq}}=E_t^2[\eta^2/R_{dc}+Z^2/8(1+Z^2)R_{ds}].$$

But from expression (8.6)

$$E_t^2=a^2E_p^2(1+Z^2)$$

whence

$$\frac{1}{2R_{eq}}=\frac{\eta^2a^2(1+Z^2)}{R_{dc}}+\frac{a^2Z^2}{8R_{ds}}.$$

For maximum power transfer, $R_{eq}$ must be equal to the dynamic resistance $R_{dp}$ of the primary circuit when tuned and this determines the value of $a$, given the other circuit parameters. If, as is usual, the power loss in the load circuit is much greater than that in

the secondary circuit itself, the expression above can be simplified at the power matching condition ($R_{eq}=R_{dp}$) to

$$\frac{1}{2R_{dp}}=\frac{\eta^2a^2(1+Z^2)}{R_{dc}}$$

$$a=\frac{1}{\eta}\left[\frac{R_{dc}}{2R_{dp}(1+Z^2)}\right]^{\frac{1}{2}}.$$

The values of $Z$ commonly employed lie between 0·5 and 0·9, and the following approximation for $a$ is useful.

$$a=(R_{dc}/3R_{dp})^{\frac{1}{2}}/\eta.$$

At this condition of matching, the $Q$ value of the primary circuit is reduced to one half of its undamped value, and as the principal source of the increased damping is the diode load circuit, the $Q$ value will vary with amplitude modulation of the input signal in such a way as to minimise the modulation. The primary circuit thus usefully supplements the "internal" a.m. rejection action of the circuit.

It follows from expression (8.6) that $E_p$ should be as large as possible for maximum a.f. output, and hence it is usual with this type of circuit to aim for the highest possible primary circuit dynamic impedance, and to achieve this the primary tuning capacitance is reduced to the lowest practicable value, 10 pf of fixed capacitance being a commonly encountered value.

It might also appear at first sight advantageous to make the dynamic resistance of the secondary circuit as high as possible, so that the working $Q$-value of the circuit would be determined almost exclusively by the diode damping. This would imply high circuit reactances, i.e. a small value of tuning capacitance. The value of the tuning capacitances cannot, however, be reduced too much, because of the variation of the equivalent input reactance of the diode circuit with amplitude modulation applied; the capacitance must be sufficiently large to "swamp" these variations.

The input impedance presented by the diodes and load circuit to the r.f. side of the circuit can be determined as follows. Each of the two diodes conducts in pulses, $V_1$ when $E_{d1}$ is at its peak positive value, and $V_2$ when $E_{d2}$ is at its peak negative value. These pulses of current are of relatively short duration, and can be analysed into a d.c. component and a series of a.c. components at multiples of the intermediate frequency. Provided that the durations of the pulses are relatively short, then the magnitude of the d.c. component

is half that of the fundamental frequency a.c. peak value. The a.c. component returns via capacitors $C_2$ and $C_3$ whilst the d.c. components flow via $R_1$ and $R_2$; since this latter is a single continuous path, the d.c. components through $V_1$ and $V_2$ are of necessity equal, and hence the a.c. components are also equal. In this connection the term d.c. component is used somewhat loosely, since with modulation applied the term extends to cover components at modulation frequencies. A very small "difference" d.c. component flows in $C_4$ to produce the a.f. output, but this component is so small as not to invalidate the assumption of equality of d.c. components. On the r.f. side of the circuit, the fundamental frequency a.c. components are in phase with the voltages producing them; i.e. the fundamental frequency a.c. component in $V_1$, $i_{d1}$, is in phase with $E_{d1}$, and similarly $i_{d2}$ with $E_{d2}$. This is shown in Fig. 8.50. The secondary circuit behaves as though each half secondary voltage were supplying one of these currents, since these currents flow out from the secondary circuit. The situation for each half secondary is thus as represented by the vector diagram of Fig. 8.51; diode $V_1$ is drawing a current equivalent to that of a capacitor and resistor in parallel, whilst diode $V_2$ is drawing a current equivalent to that of an inductor and resistor in parallel. Since

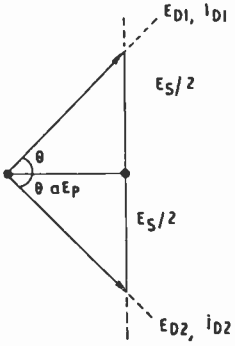



FIG. 8.50.—Vector diagram showing relationship between diode currents ($i_{d1}$, $i_{d2}$) and voltage ($E_{d1}$, $E_{d2}$).

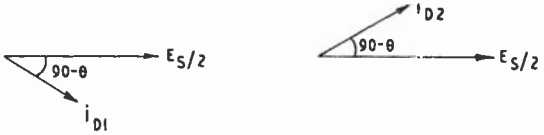



FIG. 8.51.—Showing vector relationships between the two half-secondary voltages and the diode fundamental frequency current components.

$i_{d1}=i_{d2}=2i_{de}$, these components may be found in terms of the direct current in the load circuit by resolving, using the fact that the angle $\theta$ of Figs. 8.50 and 8.51 is equal to $\tan^{-1}Z$. When this is done, the resistive terms are given by $E_{d1}/2i_{de}$ and $E_{d2}/2i_{dc}$. The sum of the resistances giving the equivalent damping resistance across the whole secondary winding is equal to $E_t/2i_{dc}$, or $R_{dc}/2$,

where $R_{dc}$ is the total value of the d.c. load resistance. Under conditions of amplitude modulation, it is the fact that $R_{dc}$ varies with $E_t$ that produce the a.m. rejection action. It is of interest to note that this damping resistance is precisely equal to that obtained if the tertiary winding were not present.

The reactive components can be similarly determined; in the diagrams of Fig. 8.51, the reactive component of $V_1$ input is capacitive, since $E_s/2$ leads $i_{d1}$, whilst that of $V_2$ is inductive since

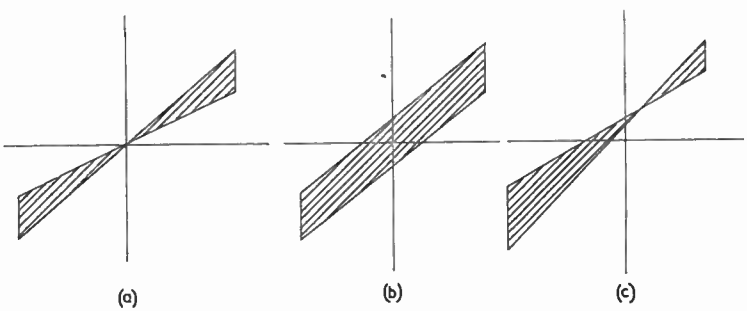



FIG. 8.52.—Showing oscillograms obtained from an f.m. signal with simultaneous a.m., the time base being provided by the f.m. modulating signal (*a*) with a "balanced" a.m., component, (*b*) an "unbalanced" a.m. component, and (*c*) with both "balanced" and "unbalanced" a.m. components.

$i_{d2}$ leads $E_s/2$. The magnitudes of these reactances are given by $ZR_{dc}/4\eta$. If the two halves of the secondary are perfectly linked, i.e. unity coupling exists, the effects of these reactances cancel out. If, however, unity coupling does not exist, the effective centre-tap of the secondary circuit is shifted from the true electrical centre tap of the winding. This results in the production of an "unbalanced" a.m. component in the audio output.

The terms "balanced" and "unbalanced" a.m. components are reserved for the two types of output due to amplitude modulation. In the first, the "balanced" case, the a.f. output is zero at the centre frequency and increases linearity with the frequency shift. In the second, "unbalanced" case, the output is of constant amplitude at all frequencies. Oscillograms in the presence of the two types of a.m. output individually and together are shown in Fig. 8.52; the audio output is applied to the $Y$ plates whilst the frequency modulating signal is applied to the $X$ plates. In the absence of a.m. the oscilloscope beam traces the input/output characteristic of the discriminator. The "balanced" output occurs with a Foster-Seeley

discriminator when the limiter is not fully effective. It also occurs with the ratio detector when the fraction of the d.c. stabilised is not correct. An "unbalanced" output occurs with a Foster-Seeley circuit and a ratio detector circuit if the secondary circuit tap is misplaced from the electrical centre, and with the ratio detector this of course occurs with the diode reactive loading effects discussed above. With a Foster-Seeley circuit, this reactive loading is sufficiently small to be negligible because of the high diode load resistors employed.

The "unbalanced" components can be eliminated quite simply by altering the relative magnitude of the resistors $R_3$ and $R_4$ of Fig. 8.49 (*c*). This adds an a.f. component to the output which can be made to cancel that due to the reactance unbalance. It is for this reason that full stabilisation of the output is generally avoided; if this is done, no such adjustment is possible.

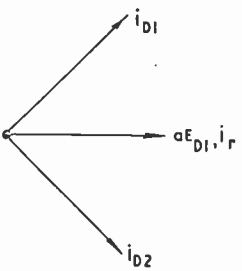



FIG. 8.53.—Showing vector relationships between $i_{d1}$, $i_{d2}$, and their resultant $i_r$, and the tertiary winding voltage.

It is also of interest to evaluate the loading applied to the primary circuit via the tertiary winding. In this case, the voltage $aE_p$ drives the two currents $i_{d1}$ and $i_{d2}$ in parallel. Thus the reactive components cancel, and the resulting current $i_r$ is in phase with $aE_p$, as shown in Fig. 8.53. The magnitude of $i_r$ is given by

$$i_r = 2i_{d1}/(1+Z^2)^{\frac{1}{2}} = 4i_{d1}/(1+Z^2)^{\frac{1}{2}}.$$

The damping resistance is thus equal to $aE_p(1+Z^2)^{\frac{1}{2}}/4i_{dc}$ and since $aE_p(1+Z^2)^{\frac{1}{2}} = E_t/2$ this resistance is given by $R_{dc}/8\eta$. This is equivalent to a resistance of $R_{dc}/8a^2\eta$ in parallel with the primary winding. This is, of course, precisely similar to the result obtained if the voltage source $aE_p$ were considered as driving directly the two diodes circuits in parallel, each diode with a load resistor $R_{dc}/2$.

In conclusion, the circuit of a typical practical ratio detector operating at 10·7 Mc/s is shown in Fig. 8.54. The primary winding is tuned by 10 pf of fixed capacitance, the total capacitance being some 16 pf. The undamped $Q$ value is in the region of 70, reduced to some 40–50 by circuit losses. The tertiary winding has approximately 1/6th of the number of primary turns, and is closely

coupled to the primary, being wound over the "cold" end. The secondary winding is bi-filar wound, to ensure uniform coupling of each half to the primary and good coupling between the halves. The tuning capacitance is 47 pf, and with an undamped $Q$ value of 100, gives a secondary dynamic resistance of 30 kilohms. The total diode load resistance is some 16 kilohms; this is equivalent to a damping resistance in parallel with the whole secondary winding of some 8–10 kilohms. The working $Q$ of the secondary is thus in

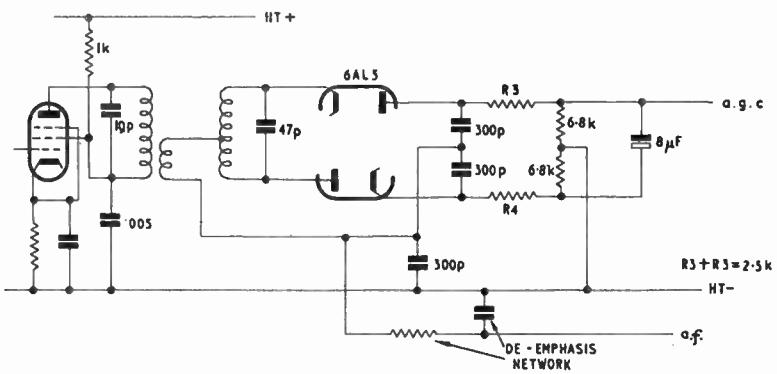



FIG. 8.54.—Typical practical ratio detector circuit with component values.

the region 20–25. The value of the half-secondary/tertiary voltage ratio is in the region of 0·6–0·7 and this determines the coupling between primary and secondary circuits, which is about 0·5 of critical for the values given. At the value of $Z=0{\cdot}7$, the magnitude of the reactances due to each diode across the half-secondary winding is about 3·3 kilohms; this is equivalent to a capacitance of about 5·5 pf on one side and an inductance of some 50 $\mu$H on the other. Because of the tight coupling between the two halves of the secondary circuits, the unbalance is comparatively small. The values of $R_3$ and $R_4$ are used not only to compensate for the reactance unbalance, but for other minor discrepancies and hence are usually adjusted in test; their sum value must be maintained constant to ensure that the correct fraction of the d.c. output is stabilised.

In conclusion, it may be stated that the ratio detector is generally inferior to the Foster-Seeley discriminator in respect of linearity; the distortion can, however, be made reasonably small by

employing a wide range discriminator characteristic. This requires in general a low value of secondary circuit $Q$, which is also necessary for adequate a.m. rejection. The a.m. rejection is usually somewhat inferior to that obtained with a Foster-Seeley circuit and separate limiter.

It has, however, the great advantage that a stage may be saved in the receiver; this is due to the fact that a Foster-Seeley circuit requires a limiter with an input of about 1 volt, whereas the ratio detector gives satisfactory results with a signal of the order of 10–100 millivolts at the grid of its driver stage. The minimum satisfactory signal depends upon the characteristics of the diodes employed; the signal applied to the diodes must be adequate to ensure satisfactory a.m. rejection.

The ratio detector gives, of course, no protection against long term variations of signal strength, nor does it equalise the outputs from two transmissions of unequal signal strengths. Hence the provision of a good a.g.c. system is essential when this type of detector is employed. The voltage across the stabilising capacitor is frequently utilised for this purpose.

## SELECTED REFERENCES

ANDREW, V. J., The Reception of Frequency Modulated Radio Signals, *Proc. I.R.E.*, May 1932.

TRAVIS, CHARLES, Automatic Frequency Control, *Proc. I.R.E.*, October 1935.

ARMSTRONG, E. H., A Method of Reducing Disturbances by a System of Frequency Modulation, *Proc. I.R.E.*, May 1936.

FOSTER, D. E., and SEELEY, S. W., Automatic Tuning, Simplified Circuits and Design Practice, *Proc. I.R.E.*, March 1937.

RODER, HANS, Theory of the Discriminator Circuit for Automatic Frequency Control, *Proc. I.R.E.*, May 1938.

LANDON, V. D., Impulse Noise in F.M. Reception, *Electronics*, February 1941.

SHEAFFER, C. F., The Zero-Beat Method of Frequency Discrimination, *Proc. I.R.E.*, August 1942.

CARNAHAN, C. W., and KALMUS, H. P., Synchronized Oscillators as F.M. Receiver Limiters, *Electronics*, August 1944.

PARKER, WILLIAM H., The Design of an Intermediate-Frequency System for Frequency Modulation-Receivers. *Proc. I.R.E.*, December 1944.

BEERS, G. L., A Frequency-Dividing Locked-in Oscillator Frequency-Modulation Receiver, *Proc. I.R.E.*, December 1944.

ARGUIMBAU, L. B., Discriminator Linearity, *Electronics*, March 1945.

BRADLEY, W. E., Single-Stage F.M. Detector, *Electronics*, October 1946.

## *Chapter Nine*

# FREQUENCY MODULATION RECEIVERS

THERE are now many excellent texts covering the design of radio receivers in general. This chapter will therefore only deal with those features which are peculiar to the design of frequency modulation receivers.

A block circuit diagram of a typical frequency modulation receiver is shown in Fig. 9.1. The circuit follows conventional

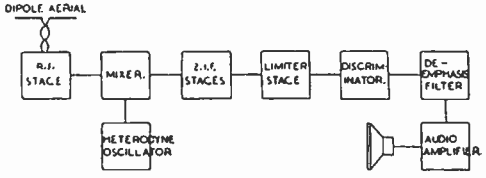



FIG. 9.1.—A block circuit diagram of a frequency modulation broadcast receiver.
(*By courtesy of the British Institute of Radio Engineers.*)

amplitude modulation superheterodyne practice up as far as the limiter stage, although, naturally, it is arranged for reception on the very high frequency band in place of the medium-wave broadcast band. There is little which need be said about the earlier stages. The band-widths of the r.f. and i.f. stages should be wide enough to pass the largest frequency deviation without introducing appreciable amplitude distortion. If, for example, 100 per cent modulation is represented by a deviation of $\pm 75$ kc/s, then a reasonable margin should be allowed on this figure; a passband of over 150 kc/s would be the essential minimum.

Following the intermediate frequency amplifier there is a limiter stage whose function is that of suppressing all amplitude variations of the received carrier signal. The output from this stage passes to a frequency discriminator which takes the place of the normal detector stage. The resultant audio signal then has the upper frequency pre-emphasis removed. This pre-emphasis is—as discussed in Chapter Four—that given to the higher audio frequencies at the transmitter.

After the audio signal has been restored to its original form by the de-emphasis filter, it is amplified in the normal manner. While there is no special technique for either the audio frequency amplifier or the loud-speaker system, they should both be of the most liberal high-fidelity design. Owing to the very low distortion factor which is inherent at almost every stage throughout a frequency modulation system, it is all too easy for the loud-speaker to become the weakest link in the chain.

Among the various refinements which can be fitted to a frequency modulation receiver, perhaps one of the most important is some form of visual tuning indicator. As frequency modulation offers an essentially high fidelity, noise and distortion-free service, it is most important to eliminate every point at which noise or distortion might be introduced. One of the most probable ways in which they may arise is as a result of the user tuning his receiver incorrectly—a tuning indicator is therefore a very practical step towards obtaining the best possible quality. Among the other possible refinements worth mentioning are automatic-frequency control and inter-station noise suppression.

### Essential Receiver Features

It has already been made clear that the only advantage shown by frequency modulation is that for a given interfering signal it is capable of substantially reducing the audio disturbance reproduced. The theoretical improvement which can be obtained has already been established in Chapters Three and Four. **However, the extent to which this improvement is realised in practice rests almost entirely in the hands of the receiver designer. Failure to appreciate the extent to which the improvement can be whittled away by lack of attention to details is frequently responsible for complaints that interference is marring reception.**

Referring back to Fig. 5.17, it will be noted that the peak interference field strength produced by 90 per cent of the vehicles which passed on a main road was less than 250 microvolts over a band 10 kc/s wide. This field strength was that measured on an aerial 100 feet from the road and some 35 feet above the ground. As the normal frequency modulation broadcast receiver bandwidth is some 150 kc/s, it follows that the equivalent peak noise field strength would be some 4,000 to 5,000 microvolts. Even taking the peak interference field strength produced by an average car, the level will still be in the region of 1,000 microvolts.

In America the Federal Communications Commission have laid down the minimum field strength which is to be provided within the service area of a frequency modulation broadcast station. The Commission specify that at all points within the transmitter's service area the minimum signal voltage produced in a receiving aerial, with an effective height of 30 feet, must be at least 1,000 microvolts in urban areas and 50 microvolts in rural areas. It will be apparent from these figures that automobile interference must very frequently have an amplitude which is comparable with that of the desired station. Couple this with the fact that automobile interference is by far the most serious source of interference, and it stands to reason that every possible step must be taken to achieve the full theoretical improvement in signal to noise ratio. If this improvement is realised in the case of impulsive interference it automatically follows that it will be present for the other and less violent forms of interference.

As was shown earlier, the threshold of improvement is determined by the level at which the carrier and interference signals have the same amplitude at the i.f. output. It therefore follows, that for the threshold to occur at the lowest possible signal level, the receiver i.f. band-width should be wide enough to pass the signal but no wider. Some margin must obviously be left in a practical receiver to allow for receiver mistuning and oscillator drift, and it is the latter requirement which generally determines amount by which the i.f. band-width exceeds the minimum permissible. A highly stable oscillator is therefore essential.

Although it was not discussed in detail earlier, it can be shown that receiver misalignment can result in a considerably higher noise output than that obtained with a receiver properly aligned. It is thus essential that the receiver should be as accurately tuned as possible, and some form of tuning indicator is therefore desirable. It also follows, of course, that the drift of the oscillator should be small, to ensure that the signal remains in tune after the initial setting.

Finally, the limiting action of the receiver must be such that the degree of amplitude modulation rejection is high. With the dynamic type of limiter, the degree of amplitude modulation which can be rejected can be pre-determined, and it is desirable that this should be in the region of 75 to 90 per cent. With the grid-leak type of limiter, the receiver must be so designed so that

with the lowest signal input at which the receiver is destined to work, the signal amplitude at the limiter grid exceeds the limiting threshold by a factor of at least three.

### Sensitivity and Selectivity

The sensitivity and selectivity of the receiver will be entirely dependent upon the transmission with which it is to be used. It will depend upon whether broadcasting, long-distance picture or code telegraphy, mobile or fixed communications, sub-carrier line telephony, or some other service is under consideration.

For broadcast receivers, the field strength at a height of 30 feet at the limits of the service area in the U.S.A. is 50 microvolts/metre and in the U.K. 250 microvolts/metre. If, however, an indoor aerial is used, the field strength may fall below this figure by some 30 db, as shown by tests carried out by the BBC. Thus for design purposes, it would appear desirable that a receiver should operate satisfactorily with an input of 1·5 microvolts/metre (U.S.) or 8 microvolts/metre (U.K.). It is doubtful if, in fact, a sensitivity of 1 microvolt/metre can be achieved without the employment of a high gain aerial, and in general, a figure of below 10 microvolts/metre for satisfactory operation may be taken. In general, for a receiver operating in the v.h.f. band, with a half wave dipole aerial, and assuming a signal of about 1 volt amplitude at the demodulator or limiter, the overall gain required is about 200,000.

The requirements of a broadcast receiver in respect of selectivity are that the passband of the r.f. and i.f. stages should be of the order of 150–200 kc/s between the points at which the response is 3 db below that at the centre frequency. In respect of adjacent channel rejection, the required degree of attenuation cannot be stated precisely, since the degree of interference is a function of the field strength of the wanted and unwanted signals. With equal signal strength, attenuation at the adjacent channel carrier frequency should be of the order of 30 db.

### The R.F. Amplifier

Most f.m. receivers include at least one stage of r.f. amplification. Whilst the gain of such a stage is generally low, its presence is generally essential for a number of reasons. Amongst these may be mentioned second channel protection; with a single tuned

circuit preceding the mixer stage, it is almost impossible to secure adequate rejection. Additionally, the mixer stage generally has a comparatively high noise level; the provision of r.f. gain therefore assists materially in maintaining a good signal to noise ratio. Further with the additive type of mixer frequently employed, the r.f. stage serves to isolate the grid of the mixer from the aerial, and thus prevents appreciable radiation at the oscillator frequency.

Three types of r.f. stage are commonly employed for v.h.f. working. These are the single pentode, the earthed-grid triode, and the combination of earthed-cathode and earthed-grid triodes in cascade (cascode circuit). At frequencies below 100 Mc/s, the single pentode or the cascode is usually preferred, although the earthed-grid triode is sometimes employed on the grounds of economy; with a twin triode, the first section can be used as an earthed-grid triode and the second as a self-oscillating mixer. At frequencies above 100 Mc/s, the cascode and earthed-grid triode are more commonly encountered.

The choice of r.f. stage type is generally governed by considerations of signal to noise ratio. Although, at the lower frequencies, the pentode and cascode circuits have the advantage that a reasonable gain can be obtained from the aerial input circuit, at the higher frequencies, this gain becomes very small. Typical specimens of each type of circuit are shown in Fig. 9.2.

In the subsequent text, we shall consider firstly the general properties of each type of circuit, and then the performance of each with respect to signal to noise ratio.

In the circuit of Fig. 9.2 (*a*), the gain from grid to anode is given by $g_m R_d$, where $g_m$ is the mutual conductance of the valve, and $R_d$ is the dynamic resistance of the tuned circuit; this assumes that the anode slope impedance of the valve is very much larger than the load. In order to follow the working of the circuits of Fig. 9.2 (*b*) and (*c*), we shall investigate further the properties of the earthed-grid circuit.

The equivalent circuit for this type of connection is shown in Fig. 9.3 (*a*). The anode, cathode, and grid of the valve are shown as terminals *a*, *c*, and *g* respectively; $\mu$ is the amplification factor of the valve, $r_a$ is its anode slope resistance and $Z_l$ is the external anode load. For comparison, the equivalent circuit for an earthed-cathode connection is shown in Fig. 9.3 (*b*). It will be seen that

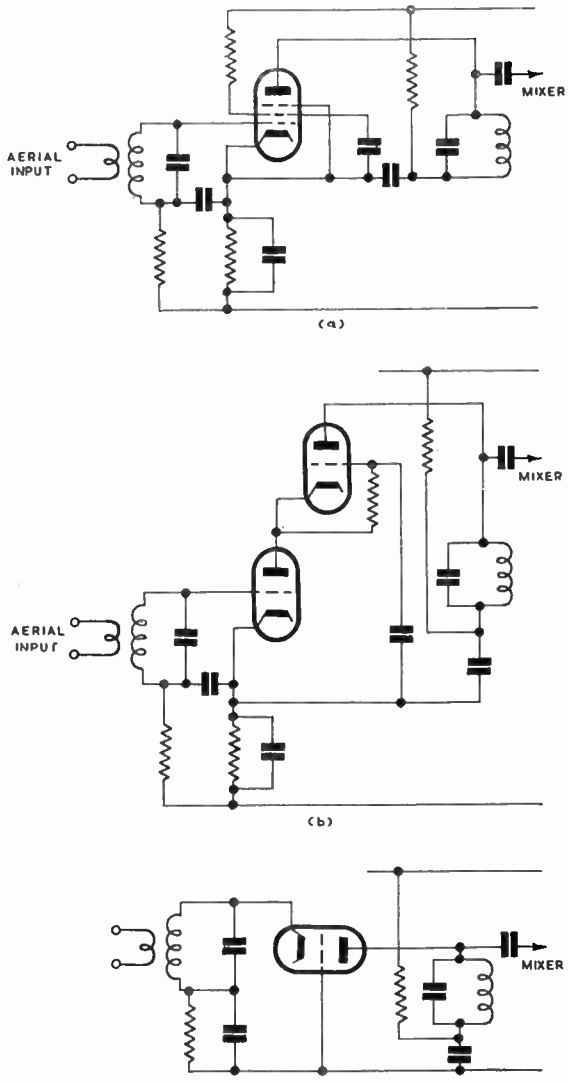



FIG. 9.2.—The three types of r.f. stages commonly employed for v.h.f. receivers: (*a*) earthed-cathode pentode, (*b*) cascode, (*c*) earthed-grid triode.

the essential difference between the circuits is that in the earthed-grid circuit, the anode current flows in the input circuit, whereas it does not for the earthed-cathode circuit. With an input voltage

$e_{gc}$ applied to the earthed-grid circuit, conditions are as shown in Fig. 9.3 (*a*), whence

$$(\mu+1)e_{gc}=(r_a+Z_l)i,$$

and

$$e_{gc}/i=(r_a+Z_l)/(\mu+1),$$

$e_{gc}/i$ is the input impedance $Z_{in}$; for $r_a \gg Z_l$ and $\mu \gg 1$, this reduces to $Z_{in}=r_a/\mu=1/g_m$. The input impedance is thus relatively low; commonly, the value of $1/g_m$ is 200 ohms. The

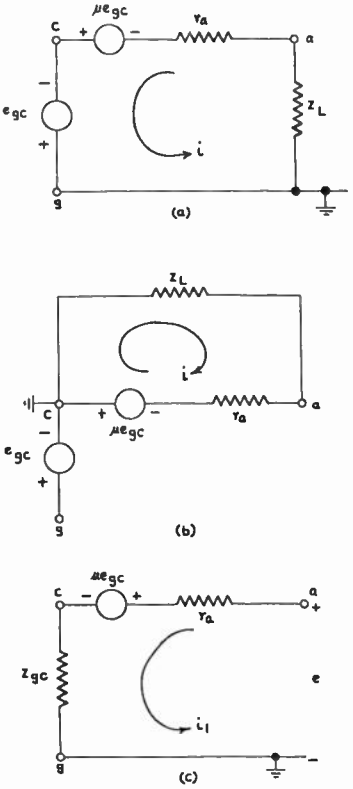



FIG. 9.3.—Equivalent circuits for: (*a*) earthed-grid circuit, (*b*) earthed-cathode circuit, (*c*) earthed-grid circuit (to determine output impedance).

expressions above are, of course, identical with those obtained for the output impedance of a cathode follower circuit. It is this low value of input impedance which accounts for the low overall stage gain generally obtained when this type of circuit is employed,

since it limits severely the input voltage obtainable from the preceding stage.

The gain of the stage is given by

$$A = Z_l i / e_{gc},$$

whence $$A = Z_l(\mu+1)/(r_a+Z_l) = Z_l/Z_{in}.$$

For $r_a \gg Z_l$ and $\mu \gg 1$, this reduces to $A = g_m Z_l$.

The full expression for the stage gain resembles that obtained with an earthed-cathode circuit, except that where $\mu$ is employed for the latter circuit, it is replaced by $\mu+1$. It should be noted further that the output is in phase with the input for the earthed-grid circuit, whereas it is in anti-phase for the earthed-cathode circuit.

The output impedance can be found from the equivalent circuit of Fig. 9.3 (*c*), where $Z_{gc}$ is the impedance of the input circuit. From the figure, assuming a voltage $e$ applied between anode and earth,

$$e = (\mu+1)\, Z_{gc} i + r_a i.$$

The output impedance $Z_{\text{out}}$ is given by $e/i$. Hence

$$Z_{\text{out}} = r_a + (\mu+1) Z_{gc}.$$

It will be seen therefore that the earthed-cathode circuit has a relatively low input impedance, and a relatively high output impedance.

In the circuit of Fig. 9.2 (*c*), it is obvious that the aerial input circuit will be severely damped by the low input resistance of the valve. Further it will be seen from the expressions above that the input impedance is not independent of the anode load impedance, unless $r_a \gg Z_l$; fortunately, at the frequencies at which this type of input is employed, this latter condition is usually fulfilled. Also the output impedance is not independent of the input circuit.

The arrangement of Fig. 9.2 (*b*) combines an earthed-cathode triode feeding into an earthed-grid triode. The anode load of the first valve, $V_1$ is thus $(r_a+Z_l)/(\mu+1)$, the input impedance of $V_2$ as determined above. Its gain $A'$ is therefore $\dfrac{\mu'\,(r_a+Z_l)/(\mu+1)}{r_a'+(r_a+Z_l)/(\mu+1)}$, where $\mu'$ and $r_a'$ relate to $V_1$. This expression reduces to

$$A' = \mu' \frac{r_a+Z_l}{(\mu+1)r_a'+r_a+Z_l}.$$

With two identical valves, $A'=\mu(r_a+Z_l)/(\mu+2)r_a+Z_l$; for $r_a \gg Z_l$, $A'=\mu/\mu+2$, which is generally very close to unity. $V_1$ is thus operated under conditions where Miller effect is not serious.

The overall gain to the anode of $V_2$ is given by $AA'$, where $A$ is the gain of $V_2$, deduced above. Hence

$$AA'=\frac{\mu'(\mu+1)Z_l}{(\mu+1)r_a'+r_a+Z_l},$$

i.e. the valve behaves as a single valve of amplification factor $\mu'(\mu+1)$, and anode slope resistance $(\mu+1)r_a'+r_a$. This output impedance is, of course, that obtained if $r_a'$ is substituted for $Z_{gc}$ in the expression for the output impedance of an earthed-grid amplifier deduced above. The combination of the two valves thus exhibits the properties of a pentode, and with two identical valves corresponds to a single pentode of anode slope resistance of $(\mu+2)r_a$ and mutual conductance $\mu(\mu+1)/(\mu+2)r_a$, which, for $\mu \gg 1$, is approximately equal to $g_m$, the mutual conductance of either valve alone. More significantly, the noise output of the stage is substantially that of $V_1$ alone, since the internal noise in $V_2$ is greatly reduced by the large cathode load, the anode slope impedance of $V_1$. For this reason, the overall combination of $V_1$ and $V_2$ has an inherently lower noise output than the comparable pentode, due mainly to the absence of partition noise in the triode $V_1$. Thus the double triode circuit is preferred to the single pentode where it is imperative that the receiver first stage noise shall be held to a minimum.

Where the cascode circuit is adopted, it is usual to resonate the stray capacitance at the anode of $V_1$ and the cathode of $V_2$ to prevent undesirable loss of signal at this point due to shunt impedance. The circuit frequently takes the form of an inductor connected between the anode of $V_1$ and the cathode of $V_2$ as shown in Fig. 9.4 (*a*). The total capacitance due to $V_1$ and $V_2$ resonating with the inductor is then equal to the sum of the stray capacitance at the anode of $V_1$ in series with the capacitance at the cathode of $V_2$. Provided these are equal, the network acts as a 1 : 1 matching network, the signal at the cathode of $V_2$ being in anti-phase with the signal at the anode of $V_1$. Because of the heavy damping of the circuit introduced at the cathode of $V_2$, the resonant circuit

contributes very little to the selectivity. If, however, the capacitances are unequal, or are deliberately unbalanced by the introduction of additional lumped capacitance, the transformation ratio departs from unity. If the capacitance at the anode of $V_1$ exceeds that at the cathode of $V_2$, a voltage step down is achieved.

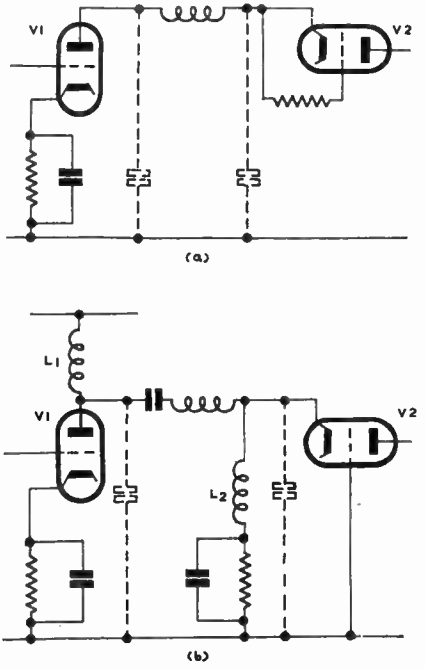



FIG. 9.4.—Cascode circuit interstage coupling by series inductance.

This is sometimes advocated as a means of reducing the amplification of $V_1$ to minimise Miller effect.

Where the two valves are not connected directly in series across the h.t. supply, as shown in Fig. 9.4 (*b*), the chokes $L_1$ and $L_2$ are introduced to complete the h.t. supply paths to the two valves. These chokes could of course be replaced by resistors; this, however, leads to a higher noise output, since the thermal noise in the resistors is then applied directly to the cathode of $V_2$.

It is now necessary to consider the input impedance of all three types of r.f. stage. In general, the input impedance comprises two components, a capacitance and a resistance in parallel. The

capacitive component is generally absorbed in the input circuit tuning; its value sets a lower limit to the tuning capacitance employed. In general, it is desirable to "swamp" this capacitance by lumped tuning capacitance, since the input capacitance of a valve is liable to changes of value as the valve warms up, and with changes of operating conditions.

The input resistive component has a number of sources. With the earthed-grid triode, as shown above, there is a low value of input resistance due to the deliberate feedback between input and output circuits. With the earthed-cathode types of circuit, however, there is generally a resistive input component due to undesired coupling of the input and output circuits (with the earthed-cathode triode, this includes the well-known Miller effect). Coupling ensues through the finite inductance of the cathode lead; the effect of this becomes more serious as frequency is raised. The earthed-grid circuit is not so seriously affected by this form of coupling, and for this reason is often preferred at the higher frequencies.

A second cause of the input resistive component is known as transit time effect. The damping due to this source becomes appreciable when the duration of the input signal cycle becomes comparable with the time taken by the electrons in the anode current stream to traverse the grid region. Whilst a full discussion of transit time effect is beyond the scope of this book, it should be noted that the magnitude of the input resistive term due to this cause decreases with the square of frequency, and hence is of increasing importance with increasing frequency. This resistive damping appears in all three types of r.f. circuit discussed; its effect is naturally least serious with the earthed-grid circuit, where the inherent input impedance is low. A third source of input damping is due to losses in the valve base and socket; the equivalent damping resistance due to this source decreases linearly with frequency. Its effect is therefore less and less marked as frequency increases, as compared with the input damping resistance due to transit time effect and cathode load inductance which decrease with the square of frequency.

The resistive damping of the input circuit due to the fact that the cathode lead is common to both the input and output circuits, is derived as follows.

The input circuit of a valve is shown in Fig. 9.5; the cathode lead

between the points $A$ and $B$ we shall assume to have an inductance $L$; included in this, of course, should be the inductance of the leads of the decoupling capacitor $C$. The valve grid-cathode capacitance $C_{gc}$ is assumed lumped, and connected to $A$.

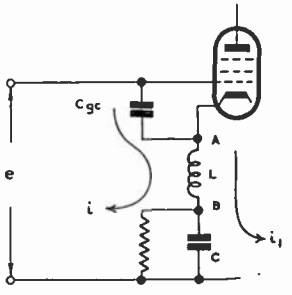



FIG. 9.5.—Input circuit of earthed-cathode stage, showing cathode lead inductance.

Then if a signal of magnitude $e_{gc}$ exists between grid and cathode, a current $i_1$ equal to $g_m e_{gc}$ flows in the cathode circuit. To this must be added the current $i$ in the capacitor $C_{gc}$, to give the voltage across the inductance, $(g_m e_{gc}+i)j\omega L$. The voltage from grid to earth is thus given by

$$e=j\omega L(g_m e_{gc}+i)+e_{gc},$$

but $e_{gc}=i/j\omega C_{gc}$, where $i$ is the input current from the source connected between grid and earth. Hence

$$e=i/j\omega C_{gc}+j\omega Li(1+g_m/j\omega C_{gc}),$$

$$e/i=1/j\omega C_{gc}+j\omega L+g_m(L/C_{gc}).$$

This is the input impedance presented by the valve, and is equal to that of the normal input capacitance, $C_{gc}$, plus two other impedances in series with it. The first modifies the reactance; the second, more importantly, is a resistive term. Making the assumption that $1/j\omega C_{gc} \gg j\omega L$, and also $1/g_m \gg L\omega$, the input impedance can be expressed as a parallel combination of a capacitance $C_g$ and a resistor of value $1/\omega^2 LCg_m$. It will be noted that this input resistance damping decreases with the square of frequency and hence varies in the same way as the damping due to transit time effect.

If practical values are considered, $C_{gc}$ may be of the order of 7 pf, and $g_m$ 8 mA/volt; $L$ obviously depends upon the physical length of the cathode lead. For a lead of diameter 0·0025 in. and of length 1 in., the inductance is 0·02 microhenries. We shall consider this value to estimate the order of magnitude of the quantities involved. From the expressions above, the input resistance of the valve at 100 Mc/s is

$$1/(4\pi^2\times10^{16}\times0{\cdot}02\times10^{-6}\times7\times10^{-12}\times8\times10^{-3})=2{,}200 \text{ ohms}$$

approximately. To determine the inductance of leads of other lengths and diameters, the following expression is given by Terman for the limiting value of inductance as frequency increases:

$$L=0{\cdot}000508l\left(2{\cdot}303\ \log_{10}\frac{4l}{d}-1\right),$$

where $l$ is the wire length and $d$ its diameter.

The effect of cathode lead inductance can be minimised by the employment of valves having two or more cathode leads. These can then be connected in parallel, to reduce the effective inductance. Alternatively, the input circuit can be returned to one lead, and the output circuit to the other. In this way, the input circuit can be divorced from the output circuit, with a consequent reduction of interaction. Another way of reducing the effect of cathode lead inductance is choosing the cathode decoupling capacitor so that series resonance with the cathode lead inductance occurs.

In general, the damping due to cathode lead inductance is more serious than that due to transit time effect; for the 6AK5 pentode, with the two cathode leads connected in parallel, the equivalent resistance due to transit time effect is approximately twice that due to cathode lead inductance.

The following table gives approximate values for the input damping resistance due to the three sources mentioned above at a frequency of 50 Mc/s for valves type 6AK5, EF80, and PCC84, which are representative of typical v.h.f. r.f. stage valves. The 6AK5 and EF80 are both pentodes, having two cathode leads. The PCC84 is a double triode, designed especially for use in cascode circuits. One triode has two cathode leads, and this valve is meant for use in the input stage. It is this triode to which the valves tabulated below relate. In deriving the values given, the cathode leads are strapped.

EQUIVALENT DAMPING RESISTANCE AT 50 Mc/s DUE TO:

| Valve | All | Transit time effect | Cathode lead inductance | Valve base etc., losses |
|---|---|---|---|---|
| 6AK5<br>EF95 | 25 kilohms | 100 kilohms | 40 kilohms | 100/200 kilohms |
| EF80<br>6BX7 | 10 kilohms | 33 kilohms | 17 kilohms | 100 kilohms |
| PCC84<br>7AN7 | 30 kilohms | 100 kilohms | 50 kilohms | 200 kilohms |
| | | $\propto 1/f^2$ | | $\propto 1/f$ |

At frequencies below 200 Mc/s, the low input resistance of the earthed-grid triode circuit means that damping of the input circuit due to transit time effect can generally be ignored. However, as the transit time effect becomes more pronounced, the

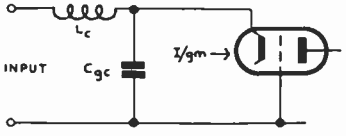



FIG. 9.6.—Earthed-grid stage, showing cathode lead inductance.

current flowing in the grid circuit may introduce a loss due to the voltage drop in the reactance of the grid-earth lead. Valves designed for this type of circuit are therefore frequently fitted with multiple grid leads, so that these may be connected in parallel to earth to minimise this effect.

Since the input and output circuits are deliberately linked in the cathode circuit, the effect of cathode lead inductance is least appreciable with the earthed grid triode. The cathode lead inductance and grid-cathode capacitance form a potential divider to the valve input, as shown in Fig. 9.6. The effect of this is, however, not serious. It is this absence of severe degeneration due to cathode lead inductance which accounts for the widespread use of the earthed grid triode at the higher frequencies. At frequencies above 200 Mc/s, in fact, the earthed-grid triode may have a higher input resistance than the comparable earthed-cathode triode in a cascode circuit.

The aerial input circuit for the earthed grid input circuit is generally of one of the four types shown in Fig. 9.7. In the circuit of Fig. 9.7 (*a*) the aerial input is introduced at a tapping on the

tuned circuit; this arrangement is obviously suitable only for use with an unbalanced feeder. The circuit of Fig. 9.7 (*b*) employs a coupling winding to terminate the feeder, and can be used with either balanced or unbalanced circuits. That of Fig. 9.7 (*c*) employs a tapping on the capacitance branch of the tuned circuit; the components $R$ and $C_3$ are necessary to provide continuity

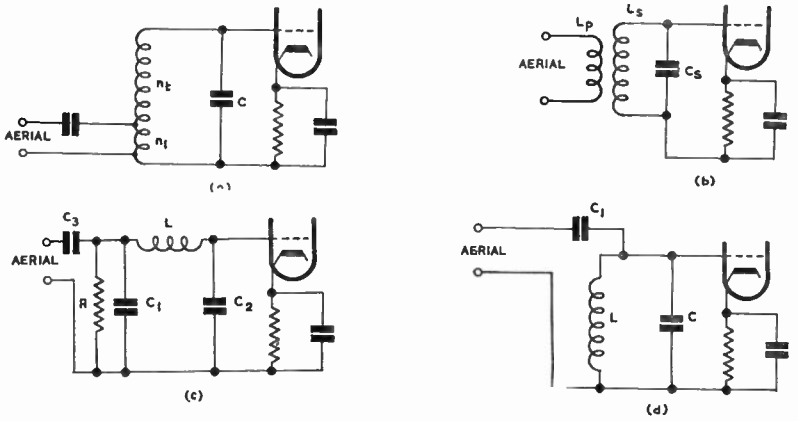



FIG. 9.7.—Four types of aerial input circuit.

from the grid of the r.f. amplifier to earth, and to exclude hum. This arrangement again is obviously suitable only for an unbalanced feeder. The circuit of Fig. 9.7 (*d*) employs a small capacitance "top-end" coupling, and is suitable for unbalanced feeder operation only.

The value to which the feeder impedance is transformed determines the tapping point in circuits (*a*) and (*c*), the coupling between the aerial coil and the tuned circuit in (*b*), and the size of the coupling capacitance in (*d*). When matched, the dynamic resistance of the tuned circuit and its $Q$ value are reduced to half the values obtained with the aerial disconnected. The voltage step-up ratio is $\frac{1}{2}(R_d/Z_0)^{\frac{1}{2}}$, referred to the aerial open circuit voltage, where $R_d$ is the dynamic resistance of the tuned circuit loaded by the valve input impedance and $Z_0$ is the characteristic impedance of the feeder. This figure does not include any allowance for losses in the feeder; these are generally of the order of 1–3 db per 100 feet of feeder, and this loss must of course be subtracted independently when computing the signal at the r.f. stage grid.

It will be of value to state here a network relationship which is of considerable use in dealing with r.f. circuits. If, in the $\pi$ section network of Fig. 9.8, the impedances $Z_1$, $Z_2$, and $Z_3$ are pure reactances, and their values are such that $Z_1+Z_2+Z_3=0$, i.e. the circuit is resonant, and the value of $R_2$ is much greater than $Z_2$, then the impedance presented by the circuit measured across $Z_1$ is given by $R_2(Z_1/Z_2)^2$.

The circuit of Fig. 9.7 (*c*) lends itself most readily to calculation.

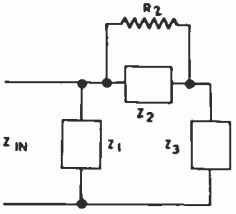



FIG. 9.8.—If $Z_1$, $Z_2$ and $Z_3$ are pure reactances, $Z_1+Z_2+Z_3=0$, and $R>1Z_2|$, then $Z_{in}=R(Z_1/Z_2)^2$.

Here $Z_1=1/j\omega C_1$, $Z_3=1/j\omega C_2$ and $Z_2=j\omega L$. The dynamic resistance of the tuned circuit may be taken as existing in parallel with $L$, and hence provided that the $Q$ value of the circuit is large, the input resistance presented to the aerial circuit at resonance is $R_d(1/\omega^2LC_1)^2$. Since $\omega L=1/\omega C_t$, where $C_t$ is the sum of $C_1$ and $C_2$ in series, the input resistance is given by $R_d(C_t/C_1)^2$. For matching, this must be equal to the characteristic impedance of the feeder $Z_0$, and hence determines the value of $C_1$ and $C_2$ for a given value of $C_t$.

The foregoing ignores the damping of the input circuit by the r.f. valve input. If the equivalent damping resistor $R_{in}$ in parallel with $C_2$ is transferred to the aerial input, the input resistance comprises two resistive components in parallel, $R_d(C_t/C_1)^2$ and $R_{in}(C_2/C_1)^2$. Knowing the values of $R_d$, $C_t$ and $R_{in}$, the values of $C_2$ and $C_1$ can be computed for correct matching. The gain from the aerial input to the grid of the r.f. amplifier is given by $C_1/C_2$. If as is usually the case for frequencies below 100 Mc/s, $C_1 \gg C_2$, $C_2$ is approximately equal to $C_t$, and hence the input resistance presented to the feeder is $R_d'(C_t/C_1)^2$, where $R_d'$ is equal to the resistance of $R_d$ and $R_{in}$ in parallel. The voltage gain is then equal to $C_t/2C_1$; i.e. $\frac{1}{2}(R_d/Z_0)^{\frac{1}{2}}$ when matched.

In the circuit of Fig. 9.7 (*a*), the correct tapping point is usually

found by experiment; for matching, this occurs when the voltage at the grid of the r.f. amplifier is maximum. It is generally not possible to calculate the correct tapping point accurately, as the turns of the inductor are usually widely spaced to minimise losses. The flux linkage between turns is thus poor, and the coil cannot be treated as a perfect autotransformer. It is however possible to calculate the tapping point on the assumption that the inductor can be treated as a perfect transformer as a first step to determining the correct tapping point. For this the input resistance is given by $R_d'(n_1/n_t)^2$, where $n_1$ is the number of turns from the

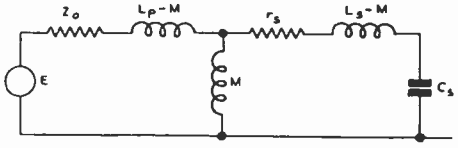



FIG. 9.9.—Equivalent circuit of Fig. 9.7 (*b*).

tapping point to earth, $n_t$ is the total number of turns, and $R_d'$ is the working dynamic resistance of the circuit, i.e. incorporating the input resistance due to the valve.

For the circuit of Fig. 9.7 (*b*), the equivalent circuit of Fig. 9.9 may be employed, where $M$ is the mutual inductance between the windings. The impedance presented to the feeder is given by $Z=j\omega L_p+\omega^2M^2/Z_s$, where $Z_s=j\omega L_s+1/j\omega C_s+r_s$; $r_s=L^2\omega^2/R_d'$. If the impedance of the coupling coil is sufficiently small to be neglected, $Z=\omega^2M^2/Z_s$; at the resonant frequency of the grid circuit, $Z=M^2R_d'/L^2$. The value of $M$ is generally not amenable to calculation, and matching is usually achieved by trial and error, occurring when the voltage at the grid of the r.f. stage is maximum. Coarse adjustment of the coupling is generally made by varying the number of turns of the coupling winding, and fine adjustment by the positioning of the coupling winding with respect to the tuned winding. With this type of circuit, it is important that the coupling winding be situated at the "earthy" end of the tuned winding; if this is not done, matching may be made considerably more difficult by stray capacitance coupling. If the reactance of the primary winding is not negligible, it can be resonated by a series capacitance, so that its effect becomes negligible.

In the circuit of Fig. 9.7 (*d*), the conditions for matching are as follows. If the grid circuit is assumed to comprise a resistance

$R_d'$ in parallel with a reactance $jX$ due to $L$ and $C$, the input impedance of the grid circuit is given by

$$R_d'X^2/(R_d'^2+X^2)+jR_d'^2X/(R_d'^2+X^2).$$

For the input impedance presented to the aerial to be purely resistive, the reactive term must be equal to the reactance of $C_1$, i.e. $1/j\omega C_1$. The input resistance presented to the feeder is then $R_d'X^2/(R_d'^2+X^2)$. For given values of $Z_0$ and $R_d'$, this determines $X$ uniquely for matching, and hence the value of $C_1$ from $1/\omega C_1=R_d'^2X/(R_d'^2+X^2)$. If $R_d'$ is much greater than $Z_0$, $X=(R_d'Z_0)^{\frac{1}{2}}=1/\omega C_1$. The voltage gain, at matching, as before is given by $\frac{1}{2}(R_d'/Z_0)^{\frac{1}{2}}$.

Where the input resistance to the valve is appreciably lower than the dynamic resistance of the tuned circuit alone, as for example with the earthed grid triode at the lower frequencies, the loss of selectivity may be troublesome. In all of the aerial input circuits discussed above, the working $Q$ value is halved at matching, and the combination of this effect with that of low valve input resistance may lead to intolerably poor aerial circuit selectivity. If, for example, second channel interference is severe, the lack of selectivity in the aerial input circuit may result in cross modulation at the r.f. stage grid. In these circumstances, the selectivity can be improved by tapping the valve input down one branch of the tuned circuit, in a manner similar to those shown in Fig. 9.7 (*a*), and (*b*) and (*c*) employed for the aerial input. This results in a raising of the $Q$ value of the input tuned circuit with consequent increase of selectivity. When this is done, of course, it will be necessary to readjust the aerial coupling to maintain the correct matching conditions. It must, however, be emphasised that this increase in selectivity can only be achieved at the expense of lower aerial input to valve gain, and higher noise factor. As a practical example we may consider the input circuit to an earthed-grid triode working at 100 Mc/s. If we take the mutual conductance of the valve as 5 mA/V, a typical figure, the characteristic impedance of the feeder as 70 ohms, the input circuit tuning capacitance as 40 pf and the undamped circuit $Q$ as 50, the dynamic resistance of the tuned circuit undamped is 2 kilohms; the $Q$ value is reduced under working conditions to 2·5, since the total resistive load in parallel with the tuned circuit under matched conditions is 100 ohms. The gain from the

equivalent aerial generator to the valve cathode is 0·85 approximately. If then the valve input is tapped down the capacitive branch of the tuned circuit as shown in Fig. 9.10, the conditions are as follows. Assuming that the total tuning capacitance $C_t$ remains constant and that $(C_3/C_t)^2=10$, the valve input resistance of 200 ohms represents a resistance of 2,000 ohms in parallel with the inductor. The undamped dynamic resistance of 2,000 is also

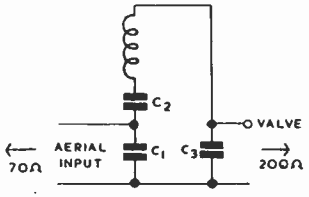



FIG. 9.10.—Tapping down input circuit with earthed-grid stage, to achieve greater selectivity.

assumed existing in parallel with the inductor, and hence the aerial circuit must be matched to a total of 1,000 ohms at this point. This requires that $(C_1/C_t)^2=1{,}000/70$. When matched, the aerial therefore presents a load of 1,000 ohms in parallel with the inductor, and hence the working dynamic resistance is equal to the resultant of the three parallel components, that due to the tuned circuit losses, 2,000 ohms, that due to the valve input 2,000 ohms and that due to the aerial, 1,000 ohms, i.e. 500 ohms total. The working $Q$ is thus raised to 12·5. The gain from the equivalent aerial generator to valve is given by $\frac{1}{2}C_1/C_3$, and from the above, this is equal to $\frac{1}{2}(10/7)^{\frac{1}{2}}=0{\cdot}6$ approximately. The gain has therefore been reduced from 0·85 to 0·6 whilst the $Q$ of the input circuit is increased from 2·5 to 12·5.

Because of the heavy damping imposed on the first tuned circuit, it is common practice in broadcast receivers for the v.h.f. band for this circuit to be pre-tuned to the centre of the band. For the U.K. broadcast band (87·5–100 Mc/s) this leads to only a slight decrease of sensitivity at the extremes of the band; for the U.S.A. broadcasting band (88–108 Mc/s) the decrease of sensitivity at the extremes of the band is more serious, and the general tendency in the U.S.A. is to employ an input circuit with variable tuning.

The tuning of the r.f. stage is, of course, determined by the method of tuning employed in the other high frequency stages of

the receiver, both variable inductance and variable capacitance tuning being commonly employed. With variable inductance tuning, the variation of inductance is generally accomplished by means of dust iron slugs which move inside the coil formers of the circuits to be tuned; these can be ganged mechanically, and this is often by means of a draw bar. Where variable capacitance tuning is employed, a pre-set dust iron slug is often employed to secure correct tracking.

For receivers working in the region of 100 Mc/s, the tuning inductors are generally wound on formers of diameter of the order of 0·25–0·5 in. The number of turns required is small, and these are generally widely spaced to minimise losses due to proximity of turns. Large diameter wire is generally employed, of the order of 18–26 s.w.g., and this is frequently silver plated to reduce losses. At these frequencies, skin effect is very pronounced and the current penetration is very small; by silver plating, the current flow is confined almost entirely to the silver. Alternatively strip conductors may be used; by employing strip, a larger surface area can be obtained for a given amount of material, with consequent higher $Q$.

The magnitude of the tuning capacitance employed varies appreciably with different receivers. With a pre-tuned r.f. stage, tuning by stray capacitance and valve input capacitance only is frequently adopted. In this case, the total tuning capacitance is of the order of 10 pf. Where variable tuning is employed, the tuning capacitance is of necessity larger, the total tuning capacitance including valve and stray capacitance being at the upper limit about 40 pf. Where the receiver is tuned by variable capacitance, the magnitude of the tuning capacitor is generally of the order of 15–25 pf. Within limits, the larger the tuning capacitance, the higher is the $Q$ value likely to be realised. This is because the dynamic resistance of the circuit is influenced largely by the valve input resistance. If a small value of capacitance is employed, the undamped dynamic resistance of the tuned circuit tends to a high value; under working conditions this may be appreciably reduced by the damping imposed by the valve, and hence the working $Q$ value may be appreciably lowered. By employing larger values of tuning capacitance, the undamped dynamic resistance tends a lower value, and hence is not so severely reduced by valve damping, with a consequent potentially higher

working $Q$ value. The upper limit obviously occurs when the dynamic resistance is appreciably less than the valve input resistance, no further potential increase of $Q$ value being then possible. In practice, the choice of tuning capacitance is governed by a great many factors, the frequency coverage required and the need to "swamp" variations of valve input capacitance generally being the primary considerations.

In the design of practical r.f. stages, it is most important that all signal and decoupling paths be kept to the minimum length possible. The reduction of signal-carrying leads to minimum length reduces circuit losses, and hence contributes materially to the achievement of high gain. If possible, all decoupling capacitors should be returned directly to cathode; this minimises losses due to cathode lead inductance discussed earlier.

The choice of variable inductance or variable capacitance tuning is governed by a number of factors; inherently, there is little to choose between the two methods. The decision is generally governed by considerations of layout.

With variable capacitance tuning by means of the usual multi-section ganged capacitor, appreciable difficulties may arise with stray capacitance and losses in the connecting leads, unless the coils can be brought close to the tuning capacitor. With variable inductance tuning, this difficulty does not generally arise, as the coils can be separated physically without detriment, since the tuning slugs, whilst ganged, can be remote. The leads between the tuning elements can be kept short, and the whole tuned circuit can generally be situated in close proximity to the valve with which it is working. Against these advantages of inductance tuning must be set the mechanical problems of accurate ganging, and the necessity for accurate tracking. Further, the fact that dust iron slugs are employed leads to higher coil losses, which in turn leads to some reduction of sensitivity. However, the loss of sensitivity due to this cause is usually very small, and may well be less than the additional losses incurred with capacitance tuning by virtue of the extra length of leads required for the latter.

The r.f. anode circuit generally comprises a parallel tuned circuit or a $\pi$ section network. With the widespread use of additive mixers, the design of the r.f. anode circuit is intimately bound up with that of the mixer, and will be considered in more detail in that section.

**Noise in R.F. Stages**

The noise output of a receiver is determined almost entirely by the conditions at the r.f. stage; the noise generated at this point is amplified equally with the signal, and hence tends to "swamp" the noise generated in later stages.

Before considering the sources of noise in the receiver itself, it must be noted that the aerial has a random noise output; this noise output has the character of thermal noise, and in fact its r.m.s. value is given by

$$E = \sqrt{4kTBR_r},$$

where $B$ is the band-width, $k$ is Boltzmann's constant $= 1{\cdot}374 \times 10^{-23}$ joules per degree centigrade, $R_r$ is the aerial input resistance, assumed constant over the range of $B$, and $T$ is the temperature of the aerial in degrees absolute ($273 + °C$).

If the aerial is terminated in a resistive load equal to $R_r$, the noise power delivered to the load is $E^2/4R_r = kTB$. This quantity is independent of the aerial itself, and is the greatest noise power output. Hence it serves as a convenient reference level for noise quantities. It also determines the absolute limit of sensitivity of the receiver because the signal power delivered by the aerial must exceed this level for a signal to noise ratio exceeding unity. The quantity $kTB$ is usually evaluated for a reference band-width. Its value, at normal temperature (20° C), is $4 \times 10^{-21}$ watts per cycle. For an f.m. system with a band-width of 200 kc/s the noise power output of the aerial is $8 \times 10^{-16}$ watts. If the r.m.s. open-circuit value of the aerial signal is $E_s$ volts, then the signal power output is $E_s^2/4R_r$; for a dipole of $R_r = 70$ ohms, $E_s$ must therefore be 0·5 microvolt for unity signal to noise ratio. It must, however, be noted that a change of radiation resistance does not necessarily mean a change of absolute sensitivity; a folded dipole, whilst having a higher resistance than a normal dipole, nevertheless has no power gain over a dipole and hence the minimum usable field strength as distinct from the minimum aerial signal output is unchanged. If an aerial is employed which has a power gain $G$ over a normal dipole, the maximum open circuit aerial signal to exceed the aerial noise output is given by $E_s = (E_{min}/G)^{\frac{1}{2}}$ where $E_{min}$ is the aerial open circuit output for unity signal to noise ratio.

In practice, the minimum aerial output required is greater than that derived above because of the noise due to the r.f. stage and signal input circuits.

The noise generated in the r.f. stage is due to five sources:

(*a*) shot noise; this arises from the fact that the cathode current of a valve is not uniform with time;

(*b*) transit time effect; not only is the valve input impedance reduced by this effect, but a noise component is introduced;

(*c*) partition noise; this arises only in tetrodes, pentodes, hexodes, etc., and is due to the fact that the distribution of cathode current between anode and screen is not uniform with time and;

(*d*) the input tuned circuit; this gives a noise output equal to that of a resistance equal to its dynamic resistance;

(*e*) the noise due to the losses in the valve base, etc.; these have a noise output; the resistive damping and noise are generally lumped with those of the tuned circuit.

The noise output due to each source can be represented as due to an equivalent generator, connected to a noise free valve. The positions of each of these generators relative to the valve electrodes is shown in Fig. 9.11 (*a*). Of necessity, the noise generators representing the sources of shot noise ($i_{sn}$) and partition noise ($i_{pn}$) are shown as constant current generators. The noise generators due to transit time noise ($E_t$) and the tuned circuit noise ($E_c$) are shown as voltage generators in the grid circuit, each in series with its appropriate resistance. Additionally, the input resistance due to feedback is shown; this resistance is, of course, noise free, but has a pronounced effect in determining the signal to noise ratio. For pentodes, this latter resistance is due almost entirely to cathode lead inductance. For earthed-cathode triodes, it is due to cathode lead inductance and Miller effect. For the earthed-grid triode, it is, of course, the input impedance of the valve; equal approximately to $1/g_m$ as shown earlier. With the earthed-grid triode, Miller effect and the effect of cathode lead inductance is generally negligible. A minor difficulty arises with the noise generator representing the noise due to transit time effect. The actual damping imposed on the input circuit is given by $R_t$; the noise due to transit time is not equal to that of a resistor $R_t$ at the ambient temperature. This difficulty is generally overcome by

postulating that the resistor $R_t$ is at a temperature $aT$, where $T$ is the ambient temperature, and $a$ is a correction factor. For oxide-coated cathodes, $a$ is 5 approximately.

The equivalent diagram of Fig. 9.11 ($a$) is somewhat inconvenient for the purposes of calculation, and to simplify the diagram it is usual to represent the noise output due to shot noise and partition noise by equivalent generators in the grid circuit. This leads to the

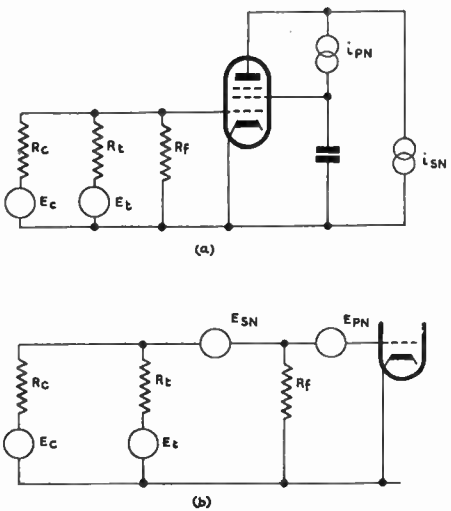



FIG. 9.11.—Equivalent generators for noise in r.f. stage.

equivalent diagram of Fig. 9.11 ($b$). The positions of the shot noise and partition noise generators are disposed to allow for the fact that the feedback due to the cathode circuit is effective in reducing shot noise, but not in reducing partition noise. The generators are now all voltage generators. The noises due to shot noise and partition noise are usually given in terms of the equivalent physical resistance having the same noise output, $R_{sn}$ and $R_{pn}$. To a good degree of approximation these are given by

$$R_{sn}=2{\cdot}5I_a/g_m(I_a+I_{sc})$$

and

$$R_{pn}=20I_{sc}I_a/g_m{}^2(I_a+I_{sc}),$$

where $I_a$ and $I_{sc}$ are the anode and screen grid currents respectively and $g_m$ is the mutual conductance of the valve.

For a triode, the resistors are given by $R_{sn}=2{\cdot}5I_a/g_m$ and $R_{pn}=0$.

For typical pentodes, $R_{pn}$ is generally of the order of from 2 to 5 times greater than $R_{sn}$; further, the feedback due to the cathode circuit is instrumental in enhancing the partition noise contribution to the total noise output relative to that of shot noise. It is this fact which leads to the abandonment of pentodes for r.f. amplifiers at the higher frequencies.

The values of $R_{sn}$ and $R_{pn}$ are frequently added together and quoted as the equivalent noise resistance of the valve; this is always permissible with triodes where $R_{pn}$ is zero, but must be treated with care where the valve is a pentode, and the input resistive component due to cathode circuit feedback is appreciable.

As a guide to the order of magnitude of the quantities involved, the following table gives approximate values of $R_{sn}$, $R_{pn}$, $R_t$, $R_f$ for typical v.h.f. receiver valves. The values of $R_t$ and $R_f$ vary inversely with the square of frequency; the values quoted are for 50 Mc/s. Additionally, the value of the valve input losses which form part of $R_c$ are given, measured at 50 Mc/s; these terms vary inversely with frequency, and hence, as will be seen from the table, can generally be neglected at frequencies above 100 Mc/s because of the much lower values of $R_t$ and $R_f$.

| Valve | Connection | $R_{pn}$ | $R_{sn}$ OHMS | $R_t$ | $R_f$ | Valve input losses OHMS |
|---|---|---|---|---|---|---|
| EF95<br>6AK5 | Pentode: twin cathode leads strapped | 1·4k | 600 | 100k | 40k | 100/200k app. |
| EF80<br>6BX7 | Pentode; twin cathode leads strapped | 600 | 400 | 33k | 17k | 100k |
| EF91<br>6AM6 | Pentode | 600 | 400 | 27k | 11k | 100k |
| PCC84 | Triode; earthed cathode. Twin cathode leads strapped | 0 | 500 | 100k | 50k | 200k |
| PCC84 | Triode: earthed grid | 0 | 500 | 100k | 160 | 200k |

In order to study the degradation of the signal to noise ratio due to the noise sources itemised above, it is necessary to consider

what occurs when the aerial is connected. By means of the tuned circuit the feeder impedance is transformed into a resistance $R_s$, having a noise output proportional to $R_s$. Thus the complete equivalent diagram is then as shown in Fig. 9.12; it is assumed that the circuit is resonant. Then assuming $E_c=E_t=E_{sn}=E_{pn}=0$, and introducing $g_s=1/R_s$, $g_c=1/R_c$, $g_t=1/R_t$, $g_f=1/R_f$, the mean

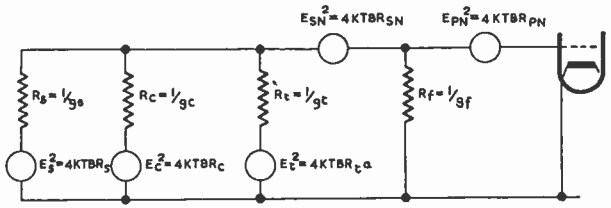



FIG. 9.12.—Equivalent noise generators in r.f. stage, including aerial circuit.

square value of the noise voltage due to the aerial at the grid of the valve is

$$\frac{4kTB}{g_s}\cdot\left(\frac{g_s}{g_s+g_c+g_t+g_f}\right)^2=\frac{4kTBg_s}{(g_s+g_c+g_t+g_f)^2}.$$

This quantity is a measure of the lowest noise power output possible, and hence forms a convenient reference level by which to measure the performance of the actual stage. If the mean square value of the voltages due to the valve and input circuit sources shown in Fig. 9.12 are $E_c'^2$, $E_t'^2$, $E_{sn}'^2$ and $E_{pn}'^2$, then since the voltages are random with respect to each other,* the total mean square noise voltage at the grid of the valve is

$$\frac{4kTBg_s+4kTBag_t+4kTBg_c+4kTBR_{sn}(g_s+g_c+g_t)^2}{(g_s+g_c+g_t+g_f)^2}+4kTBR_{pn}.$$

We may then define the noise factor $F$ of the receiver as being the ratio of this latter voltage to that due to the aerial alone. Whence

$$F=1+\frac{ag_t}{g_s}+\frac{g_c}{g_s}+\frac{R_{sn}(g_s+g_c+g_t)^2}{g_s}+\frac{R_{pn}}{g_s}\cdot(g_s+g_c+g_t+g_f)^2.$$

The noise factor is thus a direct indication of merit of the receiver in respect of internal noise. It is also, of course, a direct

* Strictly, shot noise and partition noise are not random with respect to each other; the noise voltages due to these two sources are, however, in quadrature, so that the result quoted holds.

indication of the degradation of the signal to noise ratio by the receiver, and determines by how much the actual lower limit of sensitivity of the receiver exceeds the absolute sensitivity determined as indicated previously. Whilst the noise factor is independent of band-width, it is necessary to quote the actual receiver band-width $B$ if it is desired to give the aerial open-circuit output necessary for unity signal to noise ratio; this is given by

$$E^2=16\times10^{-21}FBR_{ae},$$

where $E$ is the r.m.s. aerial open-circuit output voltage, $R_{ae}$ is the aerial resistance and $B$ is in c/s.

It must be noted that the noise factor is significant only in a receiver of high sensitivity. If, for example, a receiver at 100 Mc/s requires an input signal in the region of 100 $\mu$V for proper discriminator action, the signal to noise ratio depends only to a small degree on receiver noise, and a statement of the noise factor for such a receiver has little meaning. If, however, the receiver requires an input signal of 10 $\mu$V for proper discriminator action, and the aerial noise output is 0·5 $\mu$V, then the noise factor is of considerable importance, as in such a case the signal to noise ratio would be largely influenced by receiver noise.

Returning to the expression for the noise factor, it is instructive to expand the expression and regroup the terms as follows:

$$F=1+2R_{sn}(g_c+g_t)+2R_{pn}(g_c+g_t+g_f)+g_s(R_{sn}+R_{pn})+\frac{1}{g_s}(ag_t+g_c+R_{sn}[g_c+g_t]^2+R_{pn}[g_c+g_t+g_f]^2).$$

This expression has a minimum value for

$$g_s{}^2{}_{opt}=\frac{ag_t+g_c+R_{sn}(g_c+g_t)^2+R_{pn}(g_c+g_t+g_f)^2}{R_{sn}+R_{pn}}.$$

At this value of $g_{sopt}$,

$$F_{min}=1+2R_{sn}(g_{sopt}+g_c+g_t)+2R_{pn}(g_{sopt}+g_t+g_f).$$

Thus for given values of $g_t$, $g_c$, $g_f$, $R_{sn}$, and $R_{pn}$ there exists a certain value of $g_s$ for minimum noise factor. In general, this condition does not coincide with the condition for matching, which gives correct feeder termination and maximum gain. This latter condition is $g_s=g_c+g_t+g_f$. The choice of operating condition

adopted is determined by the importance attached to each of the three factors, gain, noise factor, and feeder termination. If the input stage is a triode, and $R_{pn}=0$, the value of $g_s$ for $F_{min}$ does not involve $g_f$; neither does the expression for $F_{min}$ itself. If then $g_f$ can be altered, it is possible to obtain minimum noise factor and simultaneous matching. This will give correct feeder termination. It does not, however, necessarily give maximum gain; this only occurs when $g_f$ is at its minimum value. Thus, with a triode r.f. stage with variable feedback damping, two conditions of operation are possible; firstly for maximum gain ($g_f$ minimum), secondly for correct feeder termination with minimum noise factor. For the latter condition:

$$g_s{}^2{}_{opt}=\{ag_t+g_c+R_{sn}(g_c+g_t)^2\}/R_{sn}=(g_c+g_t+g_f)^2,$$

whence $$F_{min}=1+2R_{sn}(g_{sopt}+g_c+g_t).$$

The value of $g_f$ can be varied in a number of ways. If it is necessary to increase $g_f$, the cathode leads can be made longer. If two cathode leads are provided, these may be strapped together, rather than separated. If $g_f$ is to be lowered, the cathode lead inductance can be reduced by employing a series capacitor, to neutralise the inductive reactance; if the capacitance is made sufficiently small, negative values of $g_f$ are obtained. If two cathode leads are provided, these may be separated. Decreasing the value of $g_f$ leads to higher aerial circuit gain.

It is not possible to secure minimum noise factor with a pentode and simultaneous matching by varying $g_f$, since the noise factor of a pentode involves $g_f$. It is, however, possible to obtain a reduction of noise factor by varying $g_f$ whilst retaining matching.

At the lower frequencies, it is generally true that the expression for $g_{sopt}$ can be simplified to

$$g_{sopt}=[ag_t+g_c/(R_{sn}+R_{pn})]^{\frac{1}{2}}$$

or to

$$g_{sopt}=[ag_t/(R_{sn}+R_{pn})]^{\frac{1}{2}}$$

ignoring $g_c$, which is often permissible. This shows that the value of $g_{sopt}$ tends to increase linearly with frequency, since $g_t$ increases with the square of frequency. The input conductance of the valve, equal to $g_t+g_f$, increases with the square of frequency for the earthed-cathode type of circuit. This is shown graphically in Fig. 9.13. Thus the ratio of $g_s$ for matching to $g_{sopt}$ becomes more nearly unity as the signal frequency increases.

In order to assist the discussion, the following table gives the values of $g_s$ at matching (i.e.$=g_f+g_t$), $g_{sopt}$, $F$ at the value of $g_s$ for matching, and $F_{min}$, for the valves the parameters of which were tabulated earlier. It is not possible to ascribe a fixed value to $g_c$; for the purposes of the table, $g_c=0$ is taken.

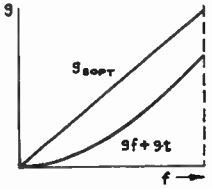



FIG. 9.13.—Approximate variation of input conductance of r.f. stage ($g_f+g_t$) and $g_{sopt}$ with frequency.

| | 50 Mc/s $g(\mu$mhos) | $F$ | 100 Mc/s $g(\mu$mhos) | $F$ | 200 Mc/s $g(\mu$mhos) | $F$ |
|---|---|---|---|---|---|---|
| 6AK5, EF95 | | | | | | |
| $g_s$ in matched condition; twin cathode leads strapped | 315 | 2·7 | 140 | 3·4 | 560 | 6·0 |
| $g_{sopt}$ | 160 | 1·8 | 340 | 2·8 | 800 | 5·9 |
| EF80, 6BX7 | | | | | | |
| $g_s$ in matched condition; twin cathode leads strapped | 93 | 3·1 | 370 | 3·9 | 1,500 | 7·5 |
| $g_{sopt}$ | 400 | 2·0 | 850 | 3·3 | 2,000 | 7·2 |
| EF91, 6AM6 | | | | | | |
| $g_s$ in matched condition | 140 | 2·9 | 560 | 4·1 | 2,240 | 9·5 |
| $g_{sopt}$ | 450 | 2·5 | 1,000 | 3·8 | 2,500 | 9·2 |
| PCC84, 7AN7 (earthed cathode) | | | | | | |
| $g_s$ in matched condition | 30* | 2·7 | 120* | 2·7 | 480* | 3·1 |
| $g_{sopt}$ | 315 | 1·3 | 630 | 1·7 | 1,260 | 2·4 |
| PCC84, 7AN7 (earthed grid) | | | | | | |
| $g_s$ in matched condition | 6,000 | 4·0 | 6,000 | 4·0 | 6,100 | 4·1 |
| $g_{sopt}$ | 315 | 1·3 | 630 | 1·7 | 1,260 | 2·4 |

The values above are calculated values, and in general the values of $F$ are somewhat lower than the values obtained in practice. They do, however, show the general trend for the noise factor to

* Ignoring Miller effect; see text.

rise with increasing frequency. With the earthed-grid triode, the degree of mismatch at $g_s = g_{sopt}$ becomes progressively less as the signal frequency is increased. Unlike the other types of r.f. circuit, the earthed-grid circuit input conductance due to feedback cannot be altered readily, to achieve matching at $g_{sopt}$.

It is of interest to note that for the earthed-cathode and earthed-grid type of circuit, the value of $g_{sopt}$ is, for the range of frequencies considered, less than the value of $g_s$ for matching for

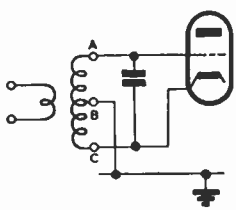



FIG. 9.14.—Circuit intermediate between earthed-cathode and earthed-grid circuits, in which matching is obtained simultaneously with minimum noise factor.

the former and greater than the value of $g_s$ for matching for the latter. This suggests that there is a form of mixed circuit intermediate between the two conditions at which $g_s = g_{sopt}$, at the matched condition. This is in fact so, and leads to the circuit of Fig. 9.14. As the frequency is increased, the optimum tapping point $B$ moves progressively from $A$ to $C$.

As noted in the table, the values of $g_s$ and $F$ for the PCC84 in the earthed-cathode condition make no allowance for Miller effect. In general it may be said that, even in the cascode circuit, Miller effect is liable to be more serious than the feedback due to cathode lead inductance, and it is common to find that the input conductance of the valve, without neutralising, exceeds the value of $g_{sopt}$. To improve the noise factor in these conditions, some form of neutralising is necessary; a number of suitable circuits are shown in Fig. 9.15. There is also the secondary advantage in these circumstances that the aerial circuit gain is increased simultaneously. This may be contrasted with the position with regard to the pentodes considered in the table. For all of these, for the range of frequencies concerned, $g_{sopt}$ exceeds the value of $g_s$ for matching. Thus improved noise factor can only be obtained at the expense of reduced aerial circuit gain, with some degree of

mismatch. At frequencies above 100 Mc/s, this loss of gain becomes very small. At 50 Mc/s, the loss of gain does not exceed 6 db for any of the examples quoted.

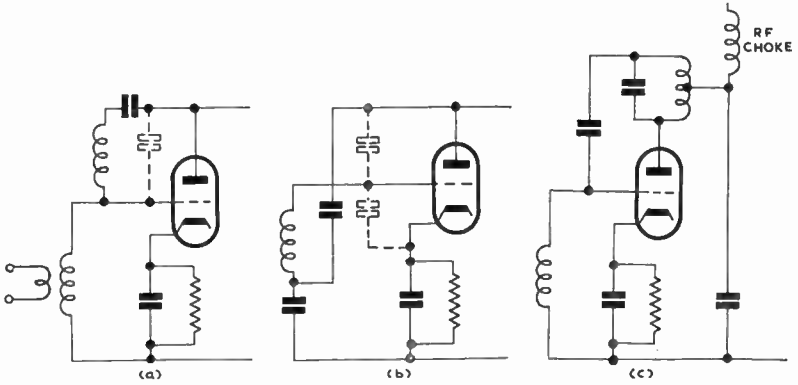



FIG. 9.15.—Three forms of neutralising circuits; that of (*a*) is suitable for fixed frequency working only.

## The Frequency Changer

Whilst a number of multiplicative type frequency changers suitable for v.h.f. working are available, the present day tendency is towards additive mixers. The reasons for this are threefold: (*a*) the additive type of frequency changer can be made to give a higher conversion conductance, values of the order of 2 mA/V being obtainable, as against 0·5–1·0 mA/V generally obtainable with multiplicative mixers; (*b*) the input impedance presented to the r.f. circuit is generally higher with additive mixers; and (*c*) the noise level in the mixer is generally lower with the additive type mixer. The additive mixers fall into two categories, the self-oscillating type and the type employing a separate oscillator. Whilst the latter has the advantage of generally higher frequency stability and usually offers the possibility of higher r.f. gain, the former is frequently employed on the grounds of economy.

Where a multiplicative mixer is employed, this may be of the heptode, triode-heptode or triode-hexode type; the circuit arrangements generally follow those commonly used at lower frequencies. The design of the oscillator section requires some care to secure optimum conversion conductance and good frequency stability; these points are considered in the next section. Typical circuits are shown in Fig. 9.16.

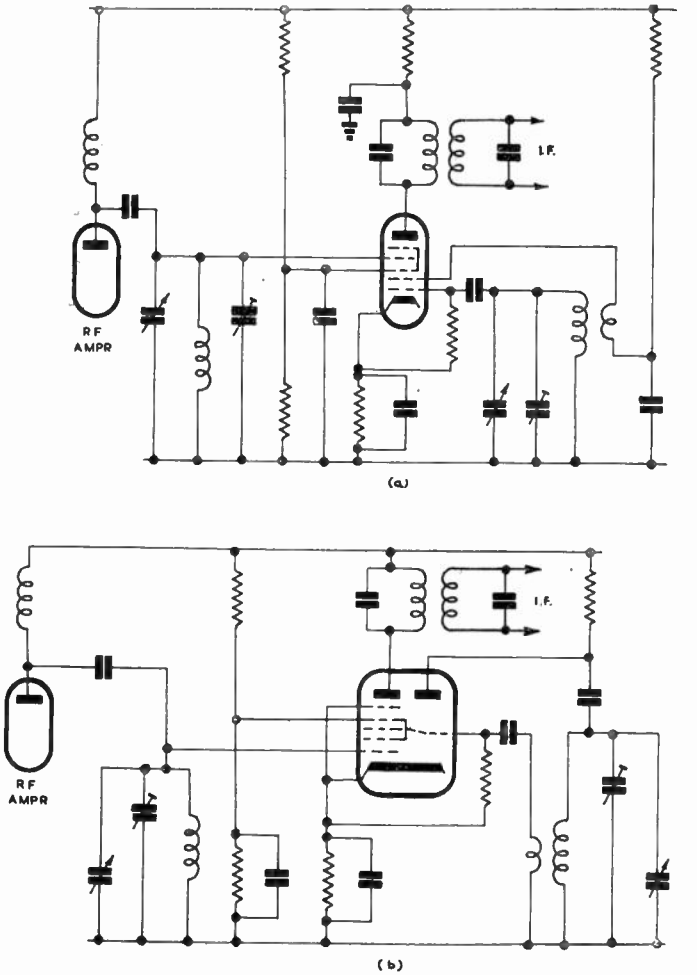



FIG. 9.16.—Heptode and triode-hexode mixer stages.

Where an additive mixer is employed, using a separate oscillator, the mixer may be either a triode or a pentode, both types being commonly met; a typical circuit is shown in Fig. 9.17. The coupling from the oscillator is somewhat critical, and is discussed in detail later. The oscillator and mixer are frequently in the same envelope, thus ensuring a compact stage.

The operating conditions of the mixer are determined by a

number of considerations. For a given set of electrode potentials, the conversion conductance rises steadily to a maximum value as the heterodyne voltage is increased; after the maximum value has been reached the conversion conductance falls again, but slowly. However, the conversion conductance does not entirely determine the gain of the mixer stage. As the standing bias applied to the valve is increased, the damping applied to the input

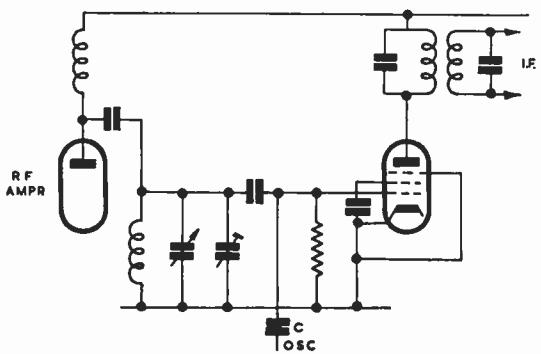



Fig. 9.17.—Additive mixer.

tuned circuit decreases, and hence increased r.f. gain may be obtained; the final operating condition is therefore usually a compromise between the values for maximum conversion conductance and minimum input damping. With triode mixers, a further consideration arises; the anode slope impedance of the valve, which effectively damps the i.f. circuit at the anode, generally increases as the heterodyne voltage at the grid is increased beyond the maximum conversion conductance figure. Thus by employing a lower value of conversion conductance, the i.f. selectivity and overall gain may be increased, and the operating conditions therefore represent a compromise between the three conflicting factors.

The choice of the size of the coupling capacitor $C$ of Fig. 9.17 is determined primarily by the amplitude of heterodyne voltage required at the mixer grid and the oscillation amplitude at the oscillator take-off point. The capacitor and the tuned circuit at the mixer grid form a potential divider, and for constant heterodyne voltage at the mixer grid, the ratio of the two arms of the divider

must be constant over the band, assuming constant oscillator output. This requirement makes this type of coupling network unsuitable for use with a pre-tuned r.f. circuit. In order to consider the matter in more detail, we shall make two assumptions. These are that the intermediate frequency is very much less than the signal frequency, and that at the oscillator frequency, the impedance of the r.f. tuned circuit is almost entirely reactive. Then the impedance of the r.f. tuned circuit is given by

$$Z=\frac{1}{j\omega_1 C_t}\frac{1}{1-\frac{1}{\omega_1^2 LC_t}}=\frac{1}{j\omega_1 C_t}\frac{\omega_1^2}{\omega_1^2-\omega_0^2},$$

where $L$ and $C_t$ are the elements of the tuned circuit resonant at the signal frequency $\omega_0/2\pi$; $\omega_1/2\pi$ is the oscillator frequency. If the intermediate signal frequency is $\omega/2\pi$, then with the first assumption made above, the expression for $Z$ simplifies to

$$Z=\pm j/2\omega C_t.$$

The sign of this reactance is positive (i.e. inductive) if the oscillator frequency is below that of the signal, and vice versa. Considering the case where the oscillator frequency is above that of the signal, this reactance is equivalent to that of a capacitance $C_{eq}$ given by

$$\omega_1 C_{eq}=2\omega C_t,$$

i.e.

$$C_{eq}=2\frac{\omega}{\omega_1}C_t.$$

If, for example, the signal frequency is 90 Mc/s, the oscillator frequency 100 Mc/s and the i.f. 10 Mc/s, $C_{eq}=0{\cdot}2C_t$. For a practical receiver, the value of $C_t$ may be in the region of 30 pf, and hence $C_{eq}$ would equal 6 pf.

The reactance of this capacitance at 100 Mc/s is 250 ohms approximately, and since this is almost certainly very much less than the dynamic resistance of the tuned circuit, confirms the validity of the second assumption made at the outset. The value of $C_{eq}$ will, of course, vary with $C_t$; where, however, the band to be covered is relatively small, the error introduced by computing $C_{eq}$ at the mid-band frequency, and assuming this value constant over the band, is small.

When the oscillator frequency is below that of the signal, the equivalent inductance is given by

$$L_{eq}\omega_1 = 1/2\omega C_t,$$

$$L_{eq} = 1/2\omega\ \omega_1 C_t,$$

and since $1/\omega_1 C_t$ may, to a first degree of approximation, be put equal to $L\omega_1$,

$$L_{eq} = \frac{L\omega_1}{2\omega},$$

i.e. the equivalent inductor for a signal frequency of 100 Mc/s and an i.f. of 10 Mc/s is equal to $5L$.

The expressions above are only approximately accurate; if it is required to know the exact value of impedance, this can of course be deduced from first principles. The values derived, however, are sufficiently accurate for practical purposes when fine adjustment of the magnitude of $C$ is generally made on test.

The necessity for the r.f. circuit to be tunable with this type of oscillator injection is most graphically illustrated by a consideration of a practical example. If the r.f. circuit is pretuned to a frequency of 98 Mc/s in a receiver designed to cover the band 88–108 Mc/s, and the i.f. is 10 Mc/s, it is obvious that as the oscillator frequency approaches 98 Mc/s, to tune the receiver to one end of the band, the heterodyne voltage will rise very appreciably. Not only will this produce a wide variation of sensitivity over the band, but may also induce very poor oscillator stability. A relatively small change in the r.f. circuit resonant frequency due, perhaps to a change of valve input capacitance, will produce very marked changes in the reactance presented at the oscillator take off point. This effect is most marked if the oscillator frequency is below that of the signal; as the oscillator frequency approaches that of the r.f. circuit, the coupling capacitance tends to produce series resonance with the effective inductance of the r.f. circuit. At the series resonant frequency itself, the oscillator is working into a very low impedance, and this may cause cessation of oscillation altogether.

An alternative oscillator injection method sometimes adopted is that of mutual inductance coupling between the r.f. and oscillator coils. In this instance, the operation of the circuit is similar to that occurring with the "top end" capacitance coupling

circuit discussed above. The e.m.f. injected in series with the r.f. circuit is given by $j\omega_1 M i_1$, where $M$ is the mutual inductance between the coils, $i_1$ is the current circulating in the oscillator circuit and $\omega_1/2\pi$ is the oscillator frequency. The e.m.f. appearing at the grid of the mixer is given by $j\omega_1 M i_1 \dfrac{1}{j\omega_1 C_t Z_{rf}}$, where $Z_{rf}=jL\omega_1+1/j\omega_1 C_t+r$. It is generally legitimate to assume that $Z_{rf}$ is purely reactive, and if $\omega \ll \omega_1$ the magnitude of the e.m.f. is equal to $\dfrac{Mi_1}{C_t 2L\omega}$, where $\omega/2\pi$ is the intermediate frequency. Since $LC_t=1/\omega_0^2$, this reduces to $i_1\omega_0^2 M/2\omega$. The oscillator voltage

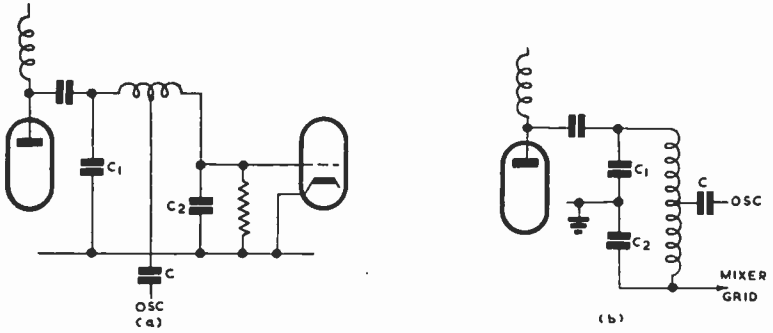



FIG. 9.18.—(*a*) Oscillator injection circuit suitable for use when r.f. stage is fixed tuned: (*b*) circuit of (*a*) re-drawn to indicate null point.

output is given approximately by $E=i_1 j\omega_1 L_1$, where $L_1$ is the oscillator inductance. Substituting for $i_1$ gives the mixer e.m.f. as $E_1\omega_0^2 M/2\omega_1 L_1\omega$, which does not vary greatly over a small range of frequency for constant $E_1$, provided that the r.f. circuit is tuned. If, however, the r.f. circuit is pre-tuned, the same difficulty arises as with the "top-end" coupling capacitance circuit; as the oscillator frequency approaches that of the r.f. circuit, $Z_{rf}$ falls rapidly, and the heterodyne voltage rises. More seriously, the impedance reflected into the oscillator circuit, $\omega^2 M^2/Z_{rf}$, also rises rapidly; the losses imposed on the oscillator may cause cessation of oscillation.

With a pre-tuned r.f. circuit, a different mode of oscillator voltage is obviously necessary, and a suitable arrangement is that of Fig. 9.18 (*a*). Here the r.f. valve is coupled to the mixer by a $\pi$ section network; the circuit is redrawn in Fig. 9.18 (*b*), to show

the existence of a null point on the inductance branch. If the local oscillator signal is injected at this point, the heterodyne voltage at the mixer grid is substantially constant. If the inductor is assumed to be perfect, i.e. there is unity coupling between turns, the impedance presented at the injection point is equal to that of the capacitors $C_1$ and $C_2$ in parallel. In practice, the coupling is likely to be well below unity, but as the reactance presented at the tapping point is generally small compared with that of the coupling capacitor, the current in each capacitor is virtually independent of the magnitude of the residual inductance in the arm.

### The Local Oscillator

The local oscillator may take any of the conventional forms of Hartley, Colpitts, tuned-anode or tuned-grid circuits. In general,

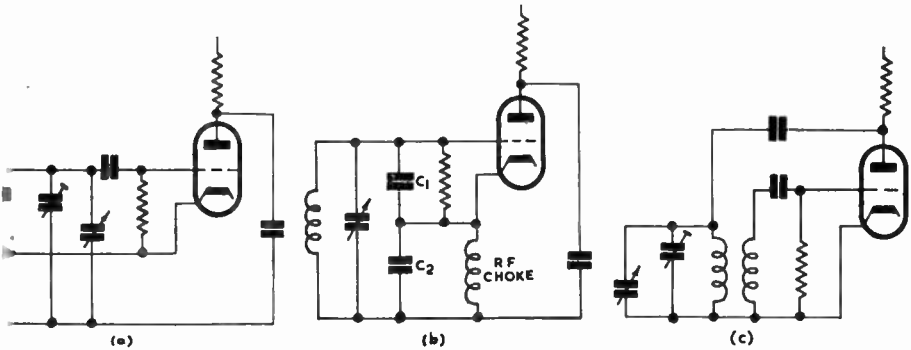



FIG. 9.19.—Three types of oscillator circuit suitable for v.h.f. working: (*a*) Hartley, earthed anode, (*b*) Colpitts, earthed anode, (*c*) tuned anode.

the most favoured types appear to be those of Fig. 9.19, in which one side of the tuning capacitor can be earthed. The Colpitts circuits have the advantage that the valve stray capacitances can be absorbed in the tapping capacitors ($C_1$ and $C_2$ of Fig. 9.19 (*b*)), but have the disadvantage that the amplitude of oscillation tends to vary more over a given band than with the other types of oscillator. At the higher frequencies, tuning by means of lumped reactances becomes progressively more difficult because of the increasing relative magnitude of strays. It is therefore common practice at these frequencies to employ resonant lines as tuning elements.

The choice of whether the oscillator frequency shall be above

or below that of the signal is governed largely by considerations of second channel protection and interference to other bands by radiation from the oscillator. For the v.h.f. bands in use for f.m. broadcasting, it would appear that the chances of second channel interference are lower if the oscillator frequency is above that of the signal. However, from the point of view of minimising oscillator interference, the lower frequency is preferable.

In general, the oscillator should be such that it is as free as possible from drift with variations of mains voltage and h.t. supply voltage, and with time. The first two sources of drift are generally not troublesome, provided that the oscillator is hard driven. Drift with time due to the effect of valve and components warming up is usually the most serious problem in oscillator design.

The principal source of such drift is the change of valve capacitance with time; the usual effect is a net increase of tuning capacitance, with a consequent fall of oscillator frequency. Serious drift due to this cause may persist for a period of the order of 10–30 minutes; usually the drift increases steadily for an initial period, and more slowly thereafter. A second source of drift is the inductor; as the temperature of this component rises, its inductance can change appreciably. The exact magnitude of drift permissible depends upon the design of the i.f. amplifier and discriminator; in general, a figure of total drift from cold of less than 30 kc/s is considered satisfactory for a receiver for the v.h.f. broadcast bands.

This initial drift may be offset in a number of ways. Firstly, the oscillator tuning capacitance can be made as large as possible, consistent with the maintenance of oscillation. As the tuning capacitance is increased, the dynamic resistance is progressively lowered, and hence the maintenance of oscillation becomes more difficult. This sets an upper limit to the magnitude of tuning capacitance permissible. Secondly, the valve connections may be tapped down the oscillator tuned circuit, with the effect that variations of valve capacitance become progressively less marked. Perhaps the best known variant of this method is that due to G. G. Gouriet, shown in Fig. 9.20. This circuit is sometimes known as the Clapp oscillator circuit, and is basically a Colpitts oscillator. In common with the first method of minimising oscillator drift, this circuit suffers from the disadvantage that as the magnitude of the tuning capacitance ($C_1$ of Fig. 9.20) is reduced relative to

the magnitude of the tapping capacitors ($C_2$ and $C_3$ of Fig. 9.20) to minimise the effect of variation of valve capacitance, oscillation becomes progressively more difficult to sustain.

The third method employs compensating capacitors having a negative temperature co-efficient, i.e. the capacitance decreases with increasing temperature. Where such a capacitance is employed, it may be operated by the rise in temperature as the receiver warms up, by heat conducted along the lead wire from a

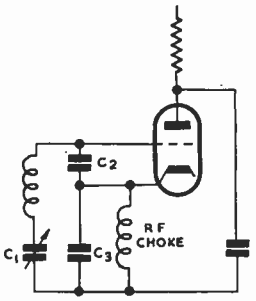



FIG. 9.20.—"Tapping down" oscillator tank circuit; Gouriet oscillator.

valve pin, or it may be heated by a resistor fed from the l.t. supply, as shown in Fig. 9.21. The former method has the advantage that the capacitor is heated relatively slowly, and can therefore be selected to minimise the oscillator drift during the whole initial period, since the total period for the capacitor to reach working temperature is comparable with that of the period of substantial valve capacitance charge. It has the disadvantage that it is prone to induce oscillator drift with changes of ambient temperature. The second method, whilst free from the last objection, has two disadvantages. Firstly, its compensating action is completed in a very short period after the receiver has been switched on, due to the low thermal inertia of the heating resistor, and hence overcompensation occurs. Where this method is adopted, the usual practice is to select the capacitor and heater power so that the final drift is held to an acceptable figure, without appreciable initial overcompensation. A second disadvantage is that the capacitor will change its value with variations of supply voltage, and consequently heater power.

The method of compensation by negative temperature co-efficient capacitors must therefore be used with caution. Provided that the drift to be corrected is reasonably small, improvement in performance can be obtained by its use. However, it is doubtful whether the performance of a fundamentally poor oscillator can be substantially improved in this way.

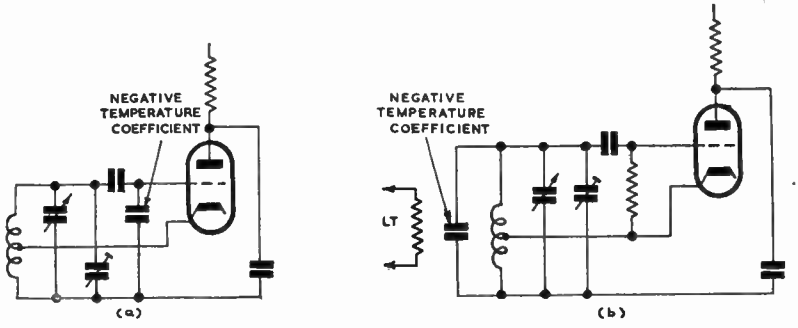



FIG. 9.21.—Negative temperature co-efficient capacitors employed to correct oscillator frequency drift: (*a*) capacitor heated by natural temperature rise, (*b*) capacitor heated by resistor.

Oscillator drift due to the heating of the tuning inductance can usually be minimised by care in siting the inductance so that its temperature does not rise appreciably. It would appear from inspection of the usual formulae for inductance of air cored coils, that the co-efficient of change of inductance per degree centigrade should be equal to the co-efficient of linear expansion. For copper, this is 16 parts in $10^6$ approximately, which is, at first sight, a commendably small value. However, in practice, the co-efficient of change of inductance is frequently very much greater than this. The discrepancy may be due to a number of causes including the way in which the coil is wound, the manner in which it is suspended, and the composition of the coil former (if any). Further, the co-efficient may suffer an additional apparent increase due to the effect of changes of coil resistance with temperature. These effects, in combination, may increase the effective co-efficient to a value in the region of 100 parts in $10^6$. Such a value would produce a drift of 5 kc/s/°C in an oscillator operating at 100 Mc/s, an unduly large value. There are a number of ways in which the co-efficient may be minimised, perhaps the most satisfactory being the employment of thin strip for the turns of the coil, these being cemented to

a former having a low co-efficient of expansion. The geometry of the coil is then largely unaffected by changes of temperature, with a corresponding reduction of the inductance temperature co-efficient. The strip can be comparatively wide, to have a large surface area, and hence low losses.

As mentioned earlier, it is possible to employ tuned line oscillators with advantage at the higher frequencies. A number of circuits utilising co-axial lines for this purpose exist, two being

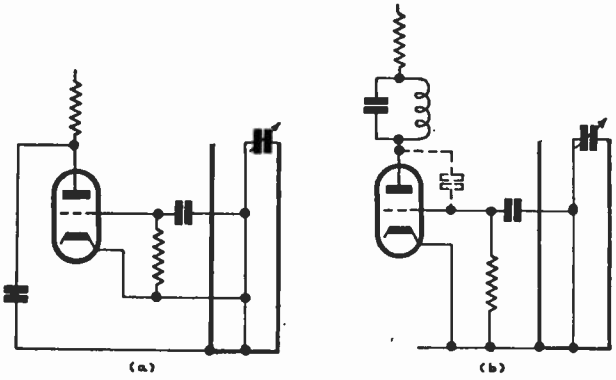



FIG. 9.22.—Oscillators with frequency controlled by tuned lines: (*a*) effective $\lambda/4$ line, (*b*) tuned anode-tuned grid circuit.

shown in Fig. 9.22. That of Fig. 9.22 (*a*) corresponds to a Hartley oscillator circuit, whilst that of Fig. 9.22 (*b*) is of the tuned anode-tuned grid (Hartley) type. In both cases the line is short circuited at one end and tuned capacitively at the other. Ignoring stray capacitance, the line in the first example behaves as a line $\lambda/4$ long, although the physical length may be very much less than this. The capacitance loading of the open end resonates with the effective inductance of the line at this point, to determine the frequency of oscillation.

## Automatic Frequency Control

Automatic frequency control is generally provided for two purposes, firstly to minimise the effect of oscillator drift, and secondly to minimise the effect of mistuning. The normal method of obtaining a.f.c. is by means of a reactance valve, the general principles of which were discussed earlier. Since there is no need

to aim for linearity of frequency shift with control voltage, the circuit can usually be made extremely simple. An example of a typical a.f.c. circuit is shown in Fig. 9.23. The triode $V_1$ *a* functions as the oscillator, whilst the triode $V_1$ *b* acts as the reactance valve. The phase shift of the oscillator voltage to the control grid of $V_1$ *b* is accomplished by the anode-grid capacitance of the valve (sometimes supplemented by a physical capacitor) in combination with a low value resistor $R_1$ of the order

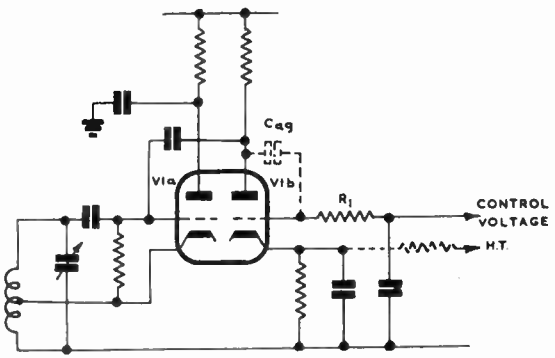



FIG. 9.23.—Double triode used as oscillator and reactance valve.

of 50–500 ohms in the grid circuit of $V_1$ *b*. The reactance valve is biased to the point where the mutual conductance changes appreciably with control voltage.

Degeneration of the control occurs because of the negative feedback applied at d.c. by the cathode resistor; this can be minimised by using a relatively low value of resistor and increasing the bias voltage by means of a bleeder resistor from the h.t. supply (shown dotted in Fig. 9.23). By employing a double triode valve for the oscillator and a.f.c. valve as shown, a compact and economical circuit arrangement is achieved. It is, of course, necessary with this type of a.f.c. circuit to employ a discriminator having zero output at the centre frequency. Due attention must also be paid to the polarity of the discriminator output, to ensure that the a.f.c. does in fact pull into tune, and not tend to increase the tuning error. It is also necessary to ensure that all the audio signal is filtered from the control supply before application to the control valve, as otherwise severe distortion may result, as the oscillator frequency follows the signal. Where a.f.c. is fitted, it is

usual to provide a means of switching it off when the receiver is initially tuned; this usually takes the form of a switch which breaks the output from the discriminator to the reactance valve, and earths the circuit to the reactance valve grid.

### Self Oscillating Mixers

The self oscillating mixer is frequently employed on grounds of economy. In general, it falls below the performance of the combination of oscillator and separate mixer, in that the requirements for oscillator stability generally conflict with those of high r.f. gain.

The most commonly adopted technique with self oscillating mixers is that of injection of the signal voltage at a null point of the oscillator circuit; this is of course the converse of the technique discussed earlier for the injection of oscillator voltage into a fixed tuned r.f. circuit. The basic form of two such circuits is shown in Fig. 9.24. In that of Fig. 9.24 (*a*) the oscillator is a Colpitts circuit. Since the control grid voltage and the screen grid voltage are in anti-phase, there is a point on the oscillator coil which is substantially at earth potential. It is thus possible to inject the signal from the r.f. stage at this point without appreciable effect on the oscillator circuit. If the capacitors $C_1$ and $C_2$ are equal, the tapping point for the signal injection is obviously the mid-point of the coil. The impedance presented to the r.f. stage is rather difficult to determine. If it is assumed that the oscillator coil has unity coupling between turns, then the impedance is a capacitance equal to $C_1+C_2$, and a parallel resistance. For the injection point to remain at earth potential at all frequencies of a band, the ratio of $C_1$ to $C_2$ must be constant, and hence both must be variable.

The circuit of Fig. 9.24 (*b*) employs also the method of null point injection. The capacitor $C_1$ is chosen to equal $C_{in}$, the input capacitance of the oscillator valve. The oscillator voltage from the anode circuit acts in series with $L_s$ to produce equal voltages across $C_1$ and $C_{in}$, but of opposite polarity. Thus the centre point of the coil is substantially at earth potential, and the signal from the r.f. stage can be injected at this point.

The input to the mixer is taken from a tapping point on the r.f. coil, to minimise the damping effect of the relatively low input impedance of the mixer. The position of the tapping point is a

compromise between the requirements of r.f. gain, maximum when the tapping point is at the anode end of the coil, and selectivity, which is at maximum when the tapping point is at the earthy end of the coil.

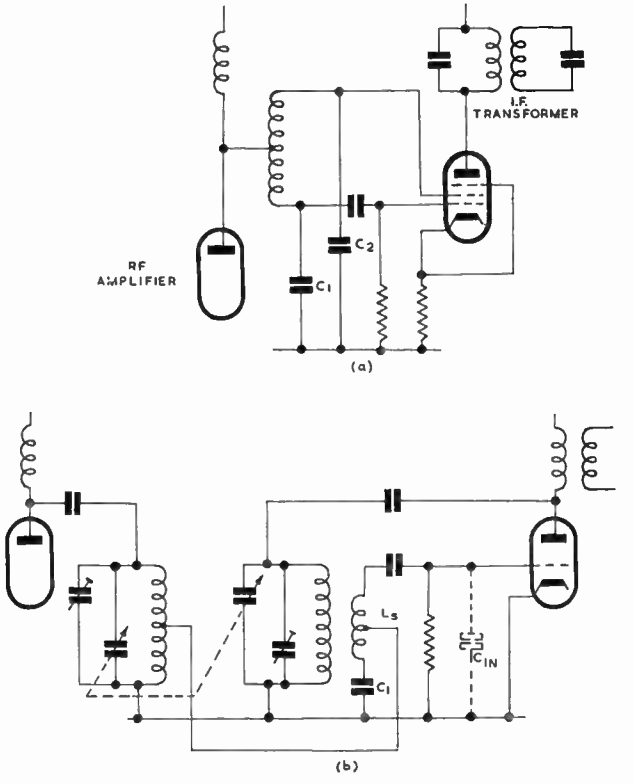



FIG. 9.24.—Self oscillating mixers, with r.f. signal injection at null point.

## Intermediate Frequency Amplifier

The choice of the intermediate frequency is governed by a number of factors. Firstly, it must be sufficiently high to secure adequate band-width; this places the minimum usable frequency in the region of 4 Mc/s. Compatible with the above requirement, the frequency must be as low as possible to ensure high stage gain with adequate selectivity. Secondly, the frequency chosen should be free from strong radiated signals, to avoid direct pick-up and amplification by the i.f. amplifier. Thirdly, the frequency should

be chosen to provide adequate second channel rejection. If the likely sources of second channel interference are within the band to be covered, second channel interference can be avoided by choosing the i.f. to be greater than half the band of frequencies to be covered.

The above may be termed the prime requirements of the intermediate frequency; in addition to these it is desirable that the frequency should be chosen so that spurious responses due to certain other causes are minimised. These comprise interference between the incoming carrier and harmonics of the i.f.; mixing of undesired signals with oscillator harmonics and mixing of signal harmonics with combinations of oscillator harmonics. For the v.h.f. broadcasting band 88–108 Mc/s an intermediate frequency of 10·7 Mc/s would appear to be standard, although frequencies in the region of 8–8·5 Mc/s have been used.

The number of i.f. stages employed is of course dependent upon the desired sensitivity and selectivity of the receiver, and the type of detector employed. If a ratio detector is employed, it is generally possible to obtain sufficient gain for satisfactory reception with one less i.f. stage than if a Foster-Seeley discriminator is used. In general, with receivers for the v.h.f. broadcast bands, the number of i.f. stages employed with a Foster-Seeley discriminator is two, and sometimes three; with a ratio detector one may be adequate to provide the necessary gain, but may not be sufficient to provide selectivity. Typical i.f. amplifier circuits are shown in the diagrams of the complete receivers given later.

The normal technique for i.f. amplifiers for broadcast receivers is to employ tuned transformers, in conjunction with high-slope straight r.f. pentodes. By using this type of valve, high stage gain can generally be obtained; the application of a.g.c. does not introduce distortion, as with a.m. receivers, except by the mistuning of the i.f. circuits which may occur. This point is discussed in more detail later.

The i.f. transformers are generally arranged to have a coupling factor ($n=KQ$) in the region of unity, and usually employ equal primary and secondary components; for these conditions, the stage gain is given by $g_m R_d n/(1+n^2)$, where $R_d$ is the dynamic resistance of either tuned circuit alone. Since $R_d$ is given by $QL\omega$, $Q$ and $L$ should be large for the highest possible stage gain. An upper limit to the value of $L$ is set by the minimum value of

tuning capacitance which can be employed. Whilst it is theoretically possible to tune by stray capacitance alone, this is undesirable, as the variation of valve strays with operating conditions can then cause very appreciable changes of the i.f. response curve. In general it may be said that the lowest minimum value of lumped capacitance is of the order of 15–20 pf; even at this figure, appreciable de-tuning may occur. Where the highest stage gain is not of prime importance, 50 pf may be taken as a good design figure.

The $Q$ value of the inductor depends upon many factors; amongst these are wire size, coil former dimensions and screen-can dimensions. By careful design $Q$ values in the region of 100 at 10 Mc/s can be realised; with a tuning capacitance in the region of 20 pf, and allowing 10 pf for strays, this represents a stage gain of about 200 using a valve with a mutual conductance of 8 mA/V. This may be taken as representative of the upper limit of gain per stage obtainable. In practice somewhat lower $Q$ values are employed; with a larger value of tuning capacitance, stage gains of 50–100 are most commonly encountered.

At a frequency near 10 Mc/s, there would appear to be little advantage in using multi-strand wire; the usual practice is to wind the coils in the form of single layer solenoids using wire sizes in the range 26–40 s.w.g. The diameter of the coil former is usually in the range 0·25–0·5 in., and the inside of the former is commonly threaded to permit tuning by means of dust iron cores.

The requirements of the i.f. amplifier in respect of selectivity depend upon the degree of distortion permissible, and the requirements of adjacent channel rejection. A complete treatment of the evaluation of this distortion has not been given, although a number of partial treatments exist. Obviously, if the i.f. passband is too narrow, significant side bands may be severely altered in amplitude, with resultant distortion; further a non-linear phase shift-frequency characteristic can also result in distortion. With a practical transformer, of course, the phase and amplitude characteristics are interdependent. As a guide, therefore, the bandwidth of a single stage employing a tuned transformer having a coupling factor in the region of unity, should be in the region of 180 kc/s between the points at which the response is 3 db down compared with that at the centre frequency, for a signal employing 75 kc/s deviation. The resultant total harmonic distortion is then

of the order of 0·25 per cent at worst. The signal will be amplitude modulated as a result of the restricted passband; this is however of minor importance, as the limiter stage will generally remove this.

The degree of adjacent channel rejection required has an important bearing upon the design of the i.f. amplifier. If the spacing between adjacent channels is 200 kc/s, it may be extremely difficult to secure adequate adjacent channel protection combined with a substantially flat response in the passband. It is impossible to give a precise figure for the degree of attenuation desirable at the adjacent channel frequency, since the degree of interference is a function of the field strengths of the wanted and interfering stations, and frequency separation. A ratio of wanted to unwanted signals of at least 30 db at the output of the i.f. amplifier would appear desirable.

In order to assist the subsequent discussion, universal curves for the amplitude and phase response error characteristics for a single tuned circuit and tuned transformers are given in Figs. 9.25 and 9.26. The phase response error is given in preference to the actual phase-frequency characteristic, since it is only the departure of the latter from linearity which is significant. The graph gives the magnitude and sense of the amount by which the phase shift departs from the value it would have if truly linear. The curves are plotted in terms of $n=KQ$, $Q$ and $x=2\Delta f/f$, where $\Delta f$ is the frequency difference of the frequency considered from the resonant frequency $f$.

With a $Q$ value of 50 and a coupling factor of unity, the response of a tuned transformer at 10 Mc/s is about 0·6 db down at 75 kc/s from resonance, about 7 db down at a frequency of 200 kc/s from resonance, and about 18 db down at 400 kc/s from resonance. Thus, with two i.f. stages, the response will be in the region of 2 db down at 75 kc/s from resonance, about 21 db down at 200 kc/s from resonance and about 54 db down at 400 kc/s from resonance (with two i.f. stages there are of course three i.f. transformers, the first being that in the mixer anode circuit). If a grid leak limiter is employed, the damping imposed on the last i.f. transformer will almost certainly result in some loss of selectivity; if a triode mixer is employed, there will be a further loss due to the damping effect of the anode slope impedance of the mixer on the primary of the first i.f. transformer.

In order to secure greater adjacent channel rejection at 200 kc/s channel spacing, high $Q$ values may be employed; to prevent undue loss in the passband, the coupling factor must be greater than unity. If $Q$ values of 100 are adopted, with a coupling factor

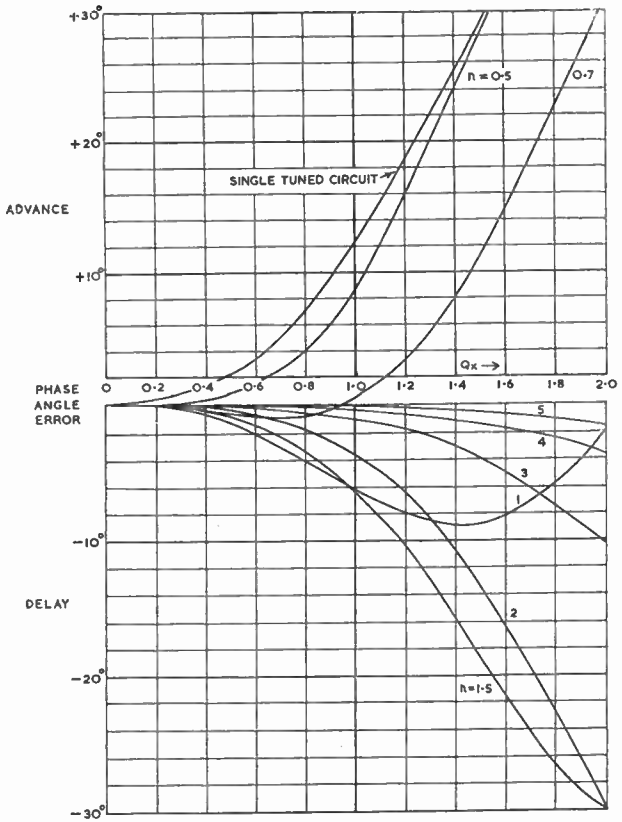



FIG. 9.25.—Universal curves for amplitude response of single tuned circuit and a pair of coupled tuned circuits.
(*From "Radio Receiver Design," by K. R. Sturley.*)

of 1·5, then with two i.f. stages the overall response rises some 2 db in the passband, the maximum occurring at a frequency of 55 kc/s from the centre frequency; at 75 kc/s the response is equal to that at the centre frequency. At 200 kc/s, the response is down by 42 db, and at 400 kc/s down by some 78 db.

In practice, the coupling factor adopted is usually between the extreme values of 1 and 1·5; however, it should be noted that in

determining the degree of adjacent channel rejection, the controlling factor is the $Q$ value; outside the passband, the difference in response between a transformer having a coupling factor of unity, and one having a coupling factor of 1·5, is some 4 db, the response of the latter being the greater. In more general terms, at a frequency well beyond the passband, the ratio of the response

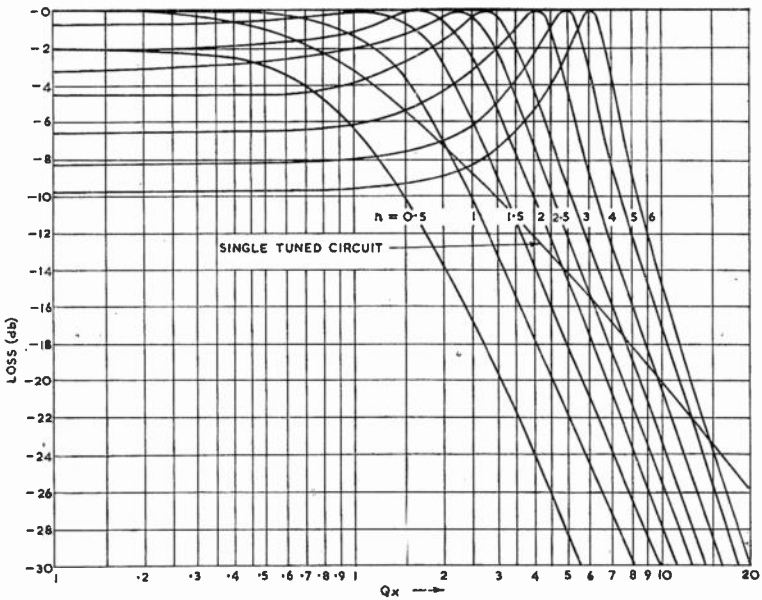



FIG. 9.26.—Universal curves for phase response of single tuned circuit and a pair of coupled tuned circuits.

(*From "Radio Receiver Design," by K. R. Sturley.*)

at resonance to the response at the frequency considered is given approximately by

$$\frac{4Q^2\Delta f^2}{(1+n^2)f_c^2},$$

where $n$=coupling factor=$KQ$.

Thus the adjacent channel selectivity is proportional to $Q^2$, and varies but slowly with $n$ in the region $1>n>1{\cdot}5$. For design purposes, therefore, the usual procedure is to select $Q$ for the degree of adjacent channel selectivity required, assuming initially $n=1$. The value of $n$ is then chosen to secure a substantially flat response in the passband. Values of $n$ greater than 1·5 are not

generally favoured, because of the variation of response in the passband, and the difficulty of adjusting and maintaining the coupling factor. The choice of a coupling factor differing from unity will, of course, alter the degree of adjacent channel rejection slightly.

It is possible to improve the response within the passband by employing different values of coupling factor through the amplifier. This has but slight effect upon the adjacent channel response as explained above. Alternatively, two transformers with a coupling factor greater than unity may be used in combination with a single tuned circuit. This latter combination has the advantage that the overall phase characteristic is made more linear in the passband. In fact with two transformers having coupling factors between 1 and 1·5, and a single tuned circuit of $Q$ identical with those employed in the transformers, the phase characteristic is substantially linear up to a frequency $\Delta f = \frac{1}{2}\frac{f}{Q}$. With this combination, the amplitude response is down by 1 db at $\Delta f = \frac{1}{2}\frac{f}{Q}$, and the adjacent channel selectivity is less than that obtained using transformers throughout. The response of a single tuned circuit at a frequency remote from resonance relative to that at resonance is down by a factor

$$\frac{2Q\,\Delta f}{f}.$$

In practice stray coupling usually makes it impossible to obtain an i.f. response curve symmetrical about the centre frequency, and hence the adjacent channel rejection is usually appreciably better on one side of the signal frequency than on the other. To minimise this asymmetry, coupling between sections of the i.f. amplifier should be kept to a minimum by screening and by the provision of adequate decoupling; ultimately, however, the degree of asymmetry is governed by the anode-grid capacitance of the valves employed, and if a high degree of symmetry is required, it may be necessary to resort to some form of neutralising.

Where a receiver is designed for both medium wave and v.h.f. working, it is common practice to employ two distinct intermediate frequencies, using the same valves in the i.f. amplifier for both. Where this is done, the two sets of i.f. transformers are

connected in series, with that for the low frequency i.f. at the "cold" end; the tuning capacitor of this transformer provides a low impedance path to earth for the v.h.f. band i.f. transformer, whilst the v.h.f. band i.f. transformer offers a low impedance at the lower frequency. Because of the need to apply a.g.c. for the medium waveband, it is usual where this technique is employed to use variable-*mu* valves in the i.f. amplifier. Typical examples of this type of i.f. amplifier are shown in the circuit diagrams of complete receivers shown later.

### Gain Control

The necessity for automatic gain control and its degree of effectiveness in an f.m. receiver depends largely upon the type of limiter and demodulator employed. If a limiter having a fixed threshold level is employed (e.g. a grid-leak limiter), a.g.c. should be applied sparingly, since its sole function is to prevent stages prior to the limiter being overloaded, whilst at the same time ensuring that the signal level applied to the limiter is sufficient to secure adequate limiting. If a dynamic limiter or ratio detector is employed, then there is no protection against long term fading, and the provision of an efficient a.g.c. circuit is essential.

The application of a.g.c. presents one major difficulty, that of change of valve input capacitance with control voltage. For a typical pentode (EF80, 6BX7) the input capacitance charge is of the order of 2 pf when the bias is increased from its normal value to that approaching cut-off. The effect may therefore be serious. If a.g.c. is applied to an r.f. stage, the de-tuning is, of course, very pronounced, since the change of capacitance is usually large in comparison with the total tuning capacitance. Additionally, the input damping decreases rapidly with a.g.c., due to the increase of the damping resistance due to cathode load inductance, which varies directly with mutual conductance. If the r.f. stage is of the fixed-tuned type, these effects may be tolerable in preference to the de-tuning which may occur in the i.f. amplifier, and may be recommended where the r.f. tuning is of this type, and the a.g.c. is used only to prevent limiter overload. It is not usually considered good practice to apply a.g.c. to the mixer as this may affect oscillator stability. If a greater degree of control is required, then it is essential that the i.f. stages should be controlled.

The effect of the charge of input capacitance in the controlled

stages may be minimised in one of three ways. Firstly, the lumped tuning capacitance may be made sufficiently large for the change of capacitance to be negligible; this, of course, leads to relatively low stage gain. Alternatively, if the valve type is suitable, the a.g.c. may be applied to the suppressor grid in addition to the control grid. As the suppressor grid bias increases, the input capacitance tends to increase, and hence opposes the change occurring due to increased control grid bias. For satisfactory operation, it is generally necessary for the voltage at the

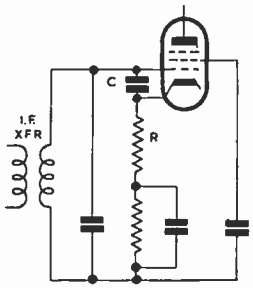



FIG. 9.27.—Correction of capacitance variation with a.g.c. voltage by cathode resistor.

suppressor grid to be of the order of ten times that at the control grid. This necessitates dividing down the a.g.c. voltage available to feed the control grid, and reduces the effectiveness of a.g.c. action; it may even necessitate amplification of the a.g.c. voltage to secure adequate control.

The third method employs a small value of undecoupled resistor in the cathode circuit of the valve, as shown in Fig. 9.27. The capacitor $C$ represents the grid cathode capacitance of the valve. When the valve is biased nearly to cut off, $C$ is at its minimum value; the working mutual conductance of the valve is low, and hence the feedback due to $R$ is negligible. At the normal working bias, $C$ increases to $C+\Delta C$, due to electronic effects within the valve. There is now a voltage $g_m RE_{gc}$ across $R$, where $E_{gc}$ is the grid-cathode voltage appearing at the valve. The total input voltage $E$ is thus $E_{gc}(1+g_m R)$. The current flowing in the grid-cathode capacitance is $E_{gc} \cdot j\omega(C+\Delta C)$, and hence the input capacitive reactance has an apparent value

$$E_{gc}(1+g_m R) \div E_{gc} \cdot j\omega(C+\Delta C) = (1+g_m R)/j\omega(C+\Delta C).$$

If, therefore, $g_m R = \Delta C/C$ the equivalent capacitance is substantially equal to its value near cut off. For the EF80, $\Delta C = 2$ pf, $C = 4$ pf, and hence $g_m R = \frac{1}{2}$. For this value at its normal bias, $g_m$ is of the order of 10mA/V (this is its mutual conductance as a triode), and hence $R$ should be of the order of 50 ohms. The effective gain of the stage is reduced by a factor $1/1+g_m R$, i.e. by about 0·6. By this means, the input capacitance is maintained substantially constant with the application of a.g.c.

### Tuning Indicators

It has already been noted at the start of this chapter that one of the theoretical assumptions upon which the improvement in signal to noise ratio is based is that the station shall be correctly tuned in. It is therefore essential that if the full advantages offered by frequency modulation are to be realised, that some means of ensuring that the signal is correctly tuned must be included in the receiver.

Quite apart from the aspect of increased noise-level, it is desirable that the receiver is tuned correctly in order to avoid the introduction of needless distortion. Even an intelligent user will find difficulty in tuning, unaided, a frequency modulation receiver with sufficient accuracy to ensure distortion-free reception. For example, the receiver may be tuned during a period when there is a relatively low level of modulation. Even if it is not tuned to the centre of the discriminator characteristic, there will be no distortion so long as the signal remains on the straight section of the discriminator curve. However, as soon as there is a deeply modulated passage, the carrier will swing over a much wider frequency band. Unless the station has been correctly tuned there will be a strong possibility that it will over-swing the straight section of the discriminator characteristic at either one end or the other.

Both on the grounds of interference suppression and distortion, it is essential that an effective tuning indicator is incorporated in all frequency modulation receivers. One source of indication is that of the voltage developed across the limiter valve's grid leak (to which the tuning indicator can be connected). In general this method is undesirable, as the top of the tuning curve will be flat over an appreciable frequency range. Under these conditions the user is merely informed that he has tuned his receiver to this

flat top—with the result that the possibility of mis-tuning is not materially reduced. There is the further possibility that with slight misalignment the maximum i.f. response may occur at some point other than the mid-point of the receiver passband.

As the most precise indication of the mid-point of the receiver characteristic is the zero voltage frequency of the discriminator curve, it follows that the control voltage for the tuning indicator should be derived from this point. Perhaps the simplest way of indicating the zero point is to connect a centre zero metre across

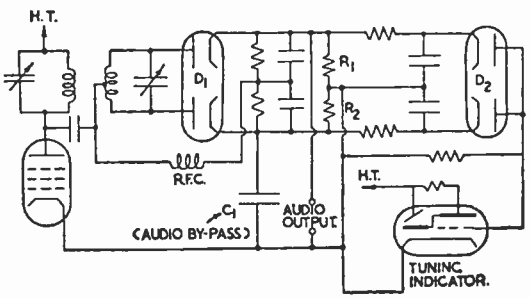



FIG. 9.28.—The above circuit arrangement gives a very sharp indication of the frequency at which the discriminator gives zero voltage output.
(*By courtesy of the British Institute of Radio Engineers.*)

the discriminator output terminals. As, however, there is a marked objection to including anything which savours of the laboratory in equipment intended for domestic use, the cathode ray magic eye type of tuning indicator is a preferable choice. It has the added advantage that in a properly designed circuit it is capable of giving an even more accurate indication of the true mid-frequency.

Fig. 9.28 shows one of a number of possible arrangements which permit the discriminator output voltages to be used as the feed for a cathode ray type of tuning indicator. It will be noted that although the condenser $C_1$ effectively earths one end of the discriminator output for audio signal voltages, that the earthing of the tap between $R_1$ and $R_2$ results in balanced d.c. voltages being applied to the two cathodes of the double diode valve $D_2$. As the tuning of the station is approaching the discriminator pass-band, the voltage applied to one of the diode cathodes will be positive, while the voltage applied to the other will be

negative. This negative voltage will be passed on via the diode to the grid of the tuning indicator, with the result that it will commence to close. As the signal frequency approaches the discriminator mid-frequency, the voltage output falls, with the result that the magic eye will again open. At resonance there will be zero output voltage and the tuning indicator aperture will be a maximum. Once past the mid-point of the discriminator characteristic, the voltage applied to the second diode will become negative with the result that the magic eye will again start to close. The behaviour of the tuning indicator as the receiver is

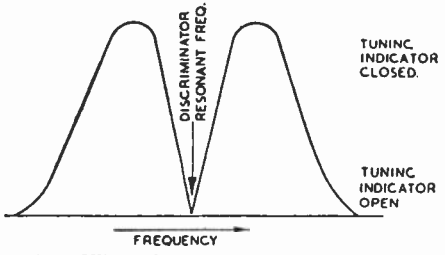



FIG. 9.29.—When the tuning indicator circuit shown in Fig. 9.28 is employed, the magic eye aperture will first be reduced and then, at resonance, fully opened before again being reduced as the station is tuned out.

(*By courtesy of the British Institute of Radio Engineers.*)

tuned through a station is indicated in Fig. 9.29. It will be noted that there is a sharp indication of the true discriminator resonant frequency.

With the circuit in Fig. 9.28 it is necessary to use an indicator with a sharp cut-off if sufficient sensitivity is to be available for accurate tuning. In receivers designed for the reception of both amplitude and frequency modulated signals, where the same magic eye has to be used, this imposes an undesirable limitation, as the indicator operated from the A.V.C. control voltage line normally has to be of the remote cut-off type.

It may therefore be more convenient to use the circuit shown in Fig. 9.30. In this circuit a double triode having a sharp cut-off is used in place of the two diodes. Fig. 9.31 shows the voltages developed across the two resistances $R_1$ and $R_2$; these voltages are applied to the grids of the two triodes. It will be noted that while one grid is positive the other is negative by an equal amount. At first sight it might appear that any effect produced on the

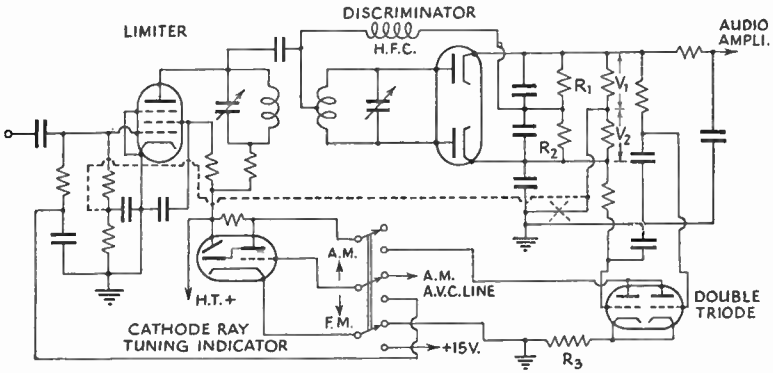



FIG. 9.30.—By using a double triode in place of the double diode shown in Fig. 9.28 it is possible to use the remote cut-off type of tuning indicator for both frequency and amplitude modulation reception.

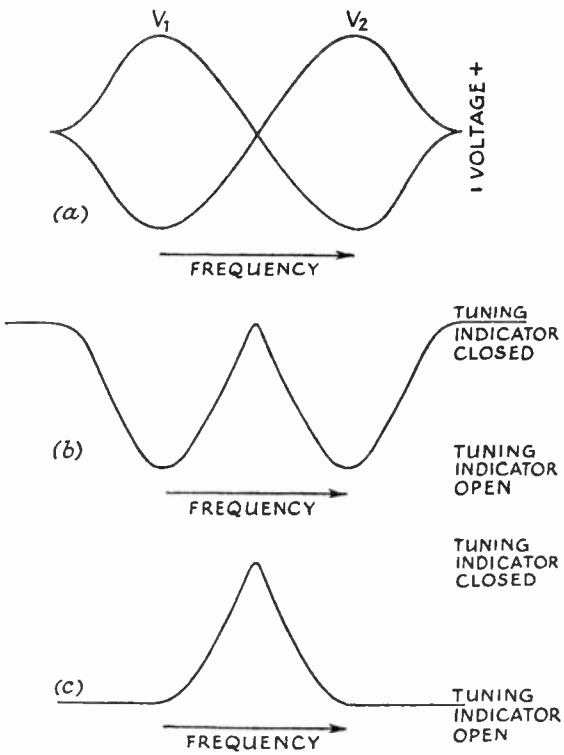



FIG. 9.31.—Diagram (*a*) shows the voltages $V_1$ and $V_2$ developed across the two resistances $R_1$ and $R_2$. With the circuit as shown in Fig. 9.30 the tuning indicator aperture will vary as shown in (*b*), while if the circuit is modified as illustrated by the dotted line it will vary as (*c*).

tuning indicator by one triode would be immediately cancelled by the other. However, the inclusion of the resistance $R_3$ in the common cathode line to both triodes, coupled with the fact that they are being operated in the region of anode current cut-off, causes the negative voltage from the discriminator to result in a negligible anode current, while the corresponding positive voltage applied to the other grid produces an abrupt rise in anode current. The two anode currents result in the same general overall tuning characteristic as that illustrated in Fig. 9.29, with the exception that the tuning indicator will now be closed at the discriminator zero voltage frequency.

This type of indicator is extremely sensitive. By employing a part of the limiter grid voltage it can be arranged that the magic eye can be open when no station is being received and closed only when a signal is correctly tuned. The circuit variation necessary to obtain this result is indicated in Fig. 9.30 by the dotted lines.

If the tuning indicator used includes a remote cut-off triode within the same envelope, provision (as is indicated in Fig. 9.30) must be made for the biasing back of the unwanted triode. As it is connected to the focusing electrode internally the slightest conduction will render the indicator circuit insensitive.

### Squelch Circuits

It has earlier been indicated that as frequency modulation is offered as a noise-free service, it is most desirable to include some form of inter-station noise suppression. The most normal form for this to take is that of a pre-audio amplifier stage which is biased into cut-off when the received signal strength falls below a given level. Such an arrangement is known as a squelch circuit. While it is not normally included in f.m. broadcast receivers, it is frequently included in communication receivers. The squelch circuit may take the form of a simple arrangement by which the audio amplifier stage is suppressed when no voltage is developed across the first limiter grid leak, or it may be combined with a more ambitious arrangement under which the interference occurring while the signal is being received is amplified and rectified, and then used to suppress the audio signal. This latter arrangement is employed in Motorola frequency modulation communication receivers, a typical circuit being shown in Fig. 9.32.

The noise amplifier selects noise in the band above the audio frequency range, and is arranged to discriminate against signals in the audio range. The noise-level is only high so long as no carrier is being received. As soon as one has been tuned in it will suppress the noise. It follows, therefore, that when the set is not tuned to a station the amplified noise will cause the squelch rectifier to produce a substantial d.c. voltage across the resistance

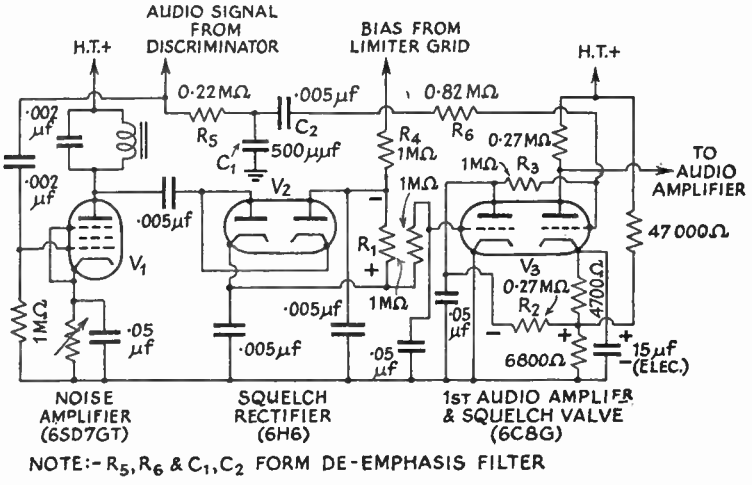



FIG. 9.32.—A typical squelch circuit as used in frequency modulation communication receivers.

$R_1$. This voltage will be of the polarity indicated on the circuit diagram. The application of this positive voltage on the grid of the first triode section of the audio and squelch amplifier valve causes the triode to draw a substantial current. This results in a voltage drop across the resistance $R_2$. It will be noted that this voltage drop has the effect of lowering the voltage on the grid of the second triode section, with respect to its cathode. This bias is transmitted to the grid of the second triode section via the grid resistance $R_3$.

As the second triode of the valve $V_3$ also acts as the first audio amplifier stage, the application of a substantial bias to its grid drives it into the region of anode current cut-off. It thus follows that the advent of a high noise-level results in this valve "squelching" or suppressing the audio signal. There will thus be

no audio signal output until a transmission has been tuned in, and until the transmission has suppressed or "quietened" the noise.

In order to ensure that when a signal is being received the first triode section of the valve $V_3$ is completely cut off (thus allowing the second section to function as a normal amplifier), its grid circuit is returned via $R_4$ to the limiter grid circuit. In this way as soon as a signal has been tuned in there will be a negative voltage applied via this resistance. This voltage is ample to back-off the first section of the squelch valve.

### Oscillator Squelch Circuit

In trans-receivers of the "Walkie-Talkie" variety, filament-type valves operated from a common low-tension battery have to be

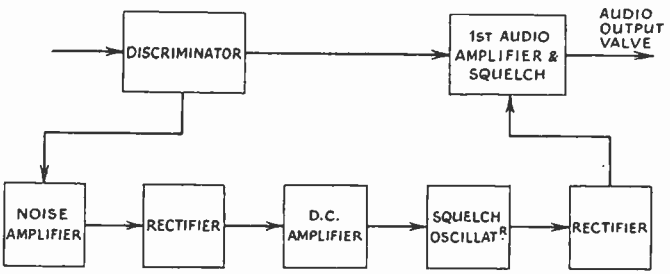



FIG. 9.33.—A block diagram of a squelch circuit employed in "Walkie-Talkie" receivers, in which all the valve cathodes are connected to a common filament battery.

employed. The squelch circuit indicated in the above section cannot therefore be used, as it requires the cathodes of the squelch valve to be at other than earth potential. An alternative circuit is therefore incorporated. Fig. 9.33 shows a block diagram of such an alternative circuit. It will be noted that, as previously, the noise is amplified and rectified. The output from the rectifier circuit is, however, fed to a d.c. amplifier. This d.c. amplifier has a common anode circuit with an oscillator valve, and is so arranged that when the rectifier develops a bias voltage across the grid circuit of the d.c. amplifier the amplifier's anode voltage rises. This permits the oscillator to function. The output from this oscillator is rectified and used to bias the audio amplifier and squelch valve into cut-off.

## Typical Frequency Modulation Receivers

In order to illustrate the design features discussed earlier, details of two receiver units are given. These are the Stromberg-Carlson model SR-401 feeder unit, and the Zenith model K725 receiver. Details of these receivers are as follows:

Stromberg-Carlson SR-401

Ranges F.M. 87–109 Mc/s
A.M. 535–1650 kc/s

Sensitivity F.M. 5 microvolts input signal provides 30 db quieting
,, A.M. 5 microvolts signal produces 0·25 volt output

Aerial Input Impedance F.M. 70–300 ohms

Intermediate frequencies F.M. 10·7 Mc/s (see Fig. 9.34 for selectivity curve)
A.M. 455 kc/s

The circuit diagram is given in Fig. 9.35.

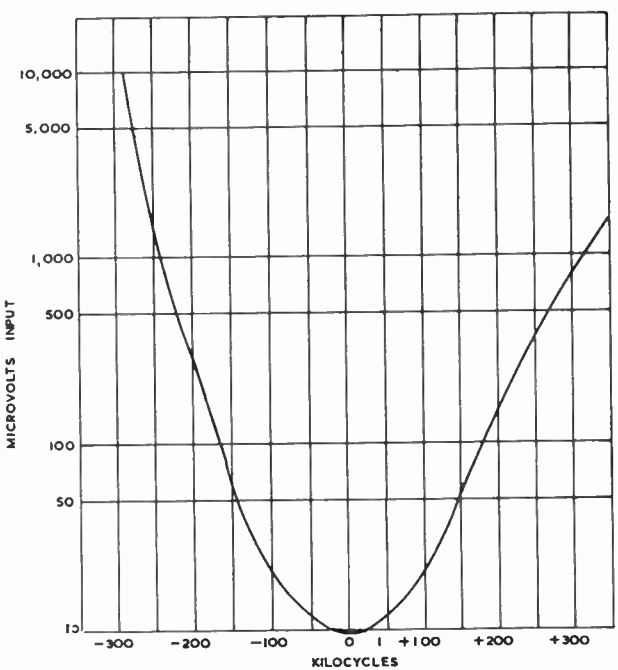



FIG. 9.34.—I.F. selectivity curve of Stromberg-Carlson SR-401 tuner unit intermediate frequency 10·7 Mc/s.

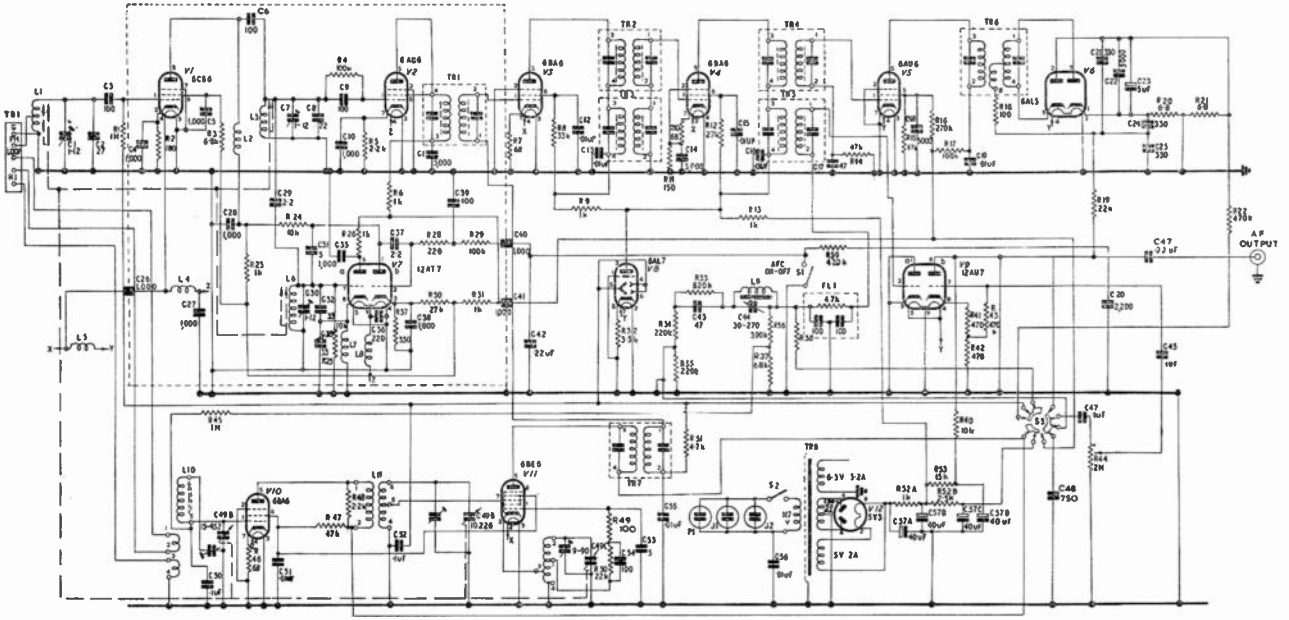


FIG. 9.35.—Circuit diagram of Stromberg-Carlson SR-401 tuner unit.

### Stromberg-Carlson Model SR-401

This feeder unit is unusual in that it incorporates a ratio detector and a limiter stage. The latter ensures efficient a.m. rejection at high signal levels, whilst the inherent a.m. rejection properties of the ratio detector are effective at lower input signal levels.

The receiver employs separate r.f. amplifiers and frequency changers for the two bands; the i.f. amplifier valves are common to both bands, separate i.f. transformers for each being employed, connected in series in the anode and grid circuits of each valve. At the output of the i.f. chain, separate detectors are, of course, employed.

The valve types and functions are as follows:

| | |
|---|---|
| 6 BA6 | a.m. r.f. amplifier |
| 6 BE6 | a.m. frequency charger |
| 6 CB6 | f.m. r.f. amplifier |
| 6 AU6 | f.m. frequency charger |
| 12 AT7 | f.m. oscillator and variable reactance valve |
| 6 BA6 | 1st i.f. amplifier |
| 6 BA6 | 2nd i.f. amplifier |
| 6 AU6 | f.m. limiter |
| 6 AL5 | ratio detector |
| 12 AU7 | a.m. detector and a.f. amplifier |
| 6 AL7 | tuning indicator |
| 5 Y3GT | mains rectifier |

### Zenith Model K725

This is a relatively simple seven-valve receiver covering both medium wave a.m. and v.h.f. f.m. bands; the circuit diagram is given in Fig. 9.36. Economy of valves is achieved by employing the same valves throughout in both bands; this, of course, requires careful arrangement of layout. In the f.m. receiver condition the detector is a Foster-Seeley discriminator preceeded by a limiter; the limiter stage is not, of course, employed on the a.m. band. Details of the receiver are as follows:

| | | |
|---|---|---|
| Aerial input impedance | 300 ohms | |
| R.F. stage gain | × 10 | F.M. |
| Mixer stage gain | × 9 | |
| I.F. gain | × 1100 | |

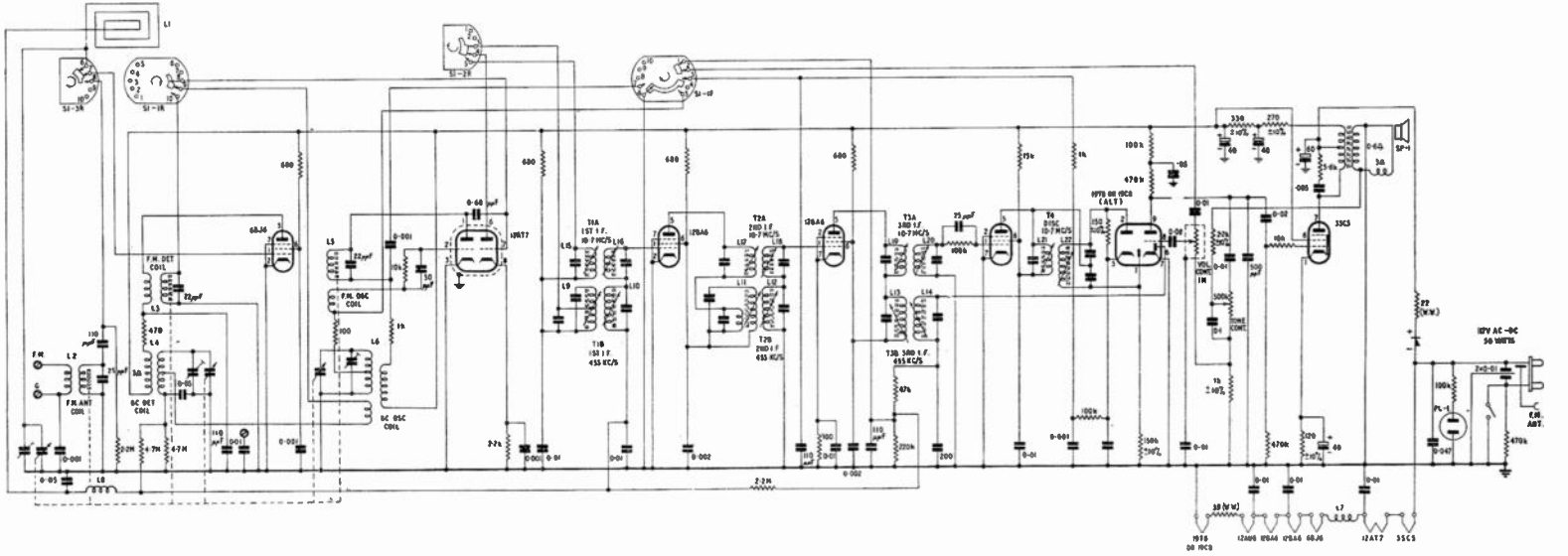


FIG. 9.36.—Circuit diagram of Zenith model K725 receiver.

| | |
|---|---|
| Input signal for limiting to commence | 10 microvolts |
| Input signal for 30 db quieting | 11 microvolts |
| I.F. band-width | 200 kc/s (6 db down)<br>900 kc/s (60 db down) |
| I.F. rejection at 98 Mc/s | 80 db |
| A.M. rejection | 23 db at 11 microvolts input<br>29 db at 110 microvolts input<br>33 db at 1100 microvolts input |

The valve types and functions are as follows:

| | |
|---|---|
| 6 BJ6 | r.f. amplifier |
| 12 AT7 | Mixer and oscillator |
| 12 BA6 | 1st i.f. amplifier |
| 12 BA6 | 2nd i.f. amplifier |
| 12 AU6 | f.m. limiter |
| 19T8 | f.m. discriminator, a.m. detector, a.f. amplifier |
| 35C5 | Output valve |

H.T. is supplied by a half-wave rectifier, fed directly from the mains supply.

## SELECTED REFERENCES


Carson, J. R., and Fry, T. C., Variable-frequency Electrical Circuit Theory, *B.S.T.J.*, 1937, 16, p. 513.

Hobbs, Marvin, F.M. Receivers, *Electronics*, August 1940.

Wheeler, H. A., Two Signal Cross Modulation in a Frequency Modulation Receiver, *Proc. I.R.E.*, December 1940.

Levy, M. L., F.M., A.M. Engineering Data, *F.M.*, March 1941.

Olney, B., Improved Speaker System for F.M., *F.M.*, April 1941.

Rice, H. E., Factory Alignment Equipment for Frequency Modulation Receivers, *Proc. I.R.E.*, October 1941.

Foster, D. E., and Rankin, J. A., Intermediate Frequency Values for Frequency Modulated Wave Receivers, *Proc. I.R.E.*, October 1941.

Thomas, K. E., The Development of an F.M. Police Receiver, *R.C.A. Review*, October, 1941.

Wunderlich, N. E., Data on Motorola Emergency F.M., *F.M.*, December 1941.

Wheeler, H. A., Common Channel Interference between two Frequency Modulated Signals, *Proc. I.R.E.*, January 1942.

RODGERS, J. A., Tuning Indicators and Circuits for Frequency Modulation Receivers, *Proc. I.R.E.*, March 1943.

PRESSMAN, LOUIS, F.M. Receiver Design, *Communications*, July 1943.

GARDINER, P. C., and MAYNARD, J. E., Aids in the Design of Intermediate Frequency Systems, *Proc. I.R.E.*, November 1944.

BEERS, G. L., A Frequency Dividing Locked-in Oscillator Frequency Modulation Receiver, *Proc. I.R.E.*, December 1944.

PARKER, W. H., The Design of an Intermediate Frequency System for Frequency Modulation Receivers, *Proc. I.R.E.*, December 1944.

NOBLE, D. E., Details of the SCR-300 F.M. Walkie-Talkie, *Electronics*, page 204, June 1945.

JAFFE, D. L., A Theoretical and Experimental Investigation of Tuned Circuit Distortion in Frequency-modulated System, *Proc. I.R.E.*, 1945, Vol. 33, p. 318.

SMITH, D. B., and BRADLEY, W. E., The Theory of Impulse Noise in Ideal Frequency Modulation Receivers, *Proc. I.R.E.*, October 1946.

STUMPERS, F. L. H. M., Distortion of Frequency Modulated Signals in Electrical Networks, *Communication News*, 1948, Vol. 9, p. 82.

ASSADOURIUN, F., Distortion of a Frequency-modulated Signal by small Loss and Phase variations, *Proc. I.R.E.*, 1952, Vol. 40, p. 172.

MEDHURST, R. G., Harmonic Distortion of Frequency-modulated Waves by Linear Networks, *Journal I.E.E.*, May 1954.

*Chapter Ten*

# MEASUREMENTS ON FREQUENCY MODULATION EQUIPMENT

As with any other branch of electrical engineering, it is necessary to refer the indicating instruments normally used for practical work to some absolute standard. With one exception all measurements associated with frequency modulation equipment are based on well-established techniques. This one measurement—that of the carrier's frequency deviation—has been introduced with the advent of this further branch of engineering.

Although conventional methods may be employed to indicate the approximate frequency deviation, such methods are not capable of giving a precise answer. A fairly accurate idea may be obtained by employing a discriminator, the output of which has been previously calibrated by slowly varying the carrier frequency and recording the output voltage readings. Under dynamic conditions the discriminator output can be measured on a peak reading voltmeter. The voltage reading thus obtained may then be compared with that obtained by the static calibration. Although this method has simplicity to recommend it, the accuracy will only be in the order of some 2 to 5 per cent, after allowance has been made for the inaccuracies and failings of the various circuits involved. While this may be quite satisfactory for the majority of practical purposes, it can hardly be considered as a standard against which to calibrate instruments such as signals generators and deviation monitors for transmitting stations. It therefore follows that in cases where an absolutely precise measurement is required, this and other similar methods must be abandoned in favour of a more basic type of measurement.

## The Bessel Zero Method of Measuring Frequency Deviation

The only absolute method of measuring frequency deviation has as its basis the side-band spectrum which is produced when a wave is modulated in frequency. Referring to Fig. 2.4,

it will be noted that at certain modulation indices the amplitudes of carrier and side bands fall to zero. The first of the modulation indices at which the carrier component's amplitude falls to zero is seen from Fig. 2.9 to be 2·40, the second 5·52, and the third 8·65, and so on. A table of the first twelve values of modulation index at which the carrier amplitude becomes zero is given below.

TABLE 16

*Deviation ratios at which the carrier component has zero amplitude*

| Zero points | Modulation index | Zero points | Modulation index |
|---|---|---|---|
| 1st | 2·4048 | 7th | 21·2116 |
| 2nd | 5·5201 | 8th | 24·353 |
| 3rd | 8·6537 | 9th | 27·4935 |
| 4th | 11·7915 | 10th | 30·6346 |
| 5th | 14·9309 | 11th | 33·7758 |
| 6th | 18·0711 | 12th | 40·0584 |

(It will be noted that the first zero is separated from the second by an amount equal to 3·115, and that this difference approaches $\pi$ or 3·1416 at the higher modulation indices.)

In the same way as the carrier component amplitude passes through a succession of zero points, the first pair of side bands fall to zero at modulation indices of 3·84, 7·01, 10·17, 13·32, and so on. It follows therefore that it is only necessary to arrange for a device which will indicate the exact point at which either the carrier or any selected pair of side bands falls to zero, in order to determine to a very high order of accuracy modulation indices of 2·40, 5·52, 8·65, 11·79, etc., or 3·84, 7·01, 10·17, 13·32, etc.

At first sight it might seem that when employing this method of measurement the number of calibration points is limited: in practice, however, this is not the case. Let it be supposed that it is necessary to calibrate the "frequency deviation" dial on a frequency modulation signal generator, and that calibration points are required every 1,000 cycles from 1,000 to 10,000 cycles, and after that at every 5,000 cycles. For simplicity of operation, it is by far the most convenient to work from the modulation indices at which the carrier amplitude falls to zero. Taking the first of these deviation ratios—2·40408—it follows that the modulating frequency which will result in the carrier component being

precisely zero with a frequency swing of exactly ±1,000 cycles will be $\frac{1,000}{2{\cdot}4048} = 415{\cdot}83$ cycles.

The carrier component will again be zero with a swing of precisely ±2,000 cycles when the modulating frequency is $2 \times 415{\cdot}83 = 831{\cdot}66$ cycles. The table on page 411 gives modulating

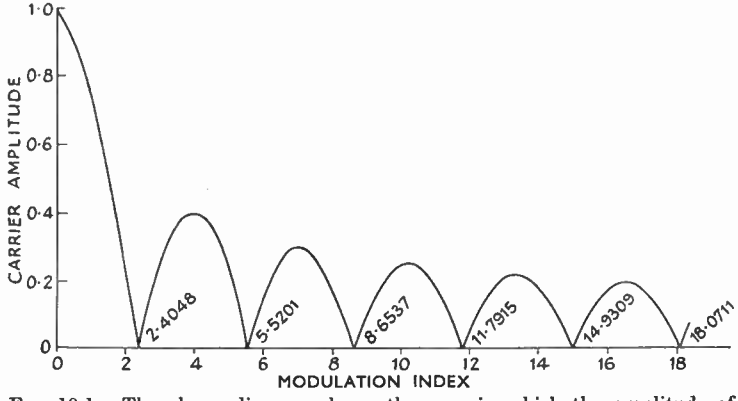



FIG. 10.1.—The above diagram shows the way in which the amplitude of the carrier component of a frequency modulated wave alters as the modulation index is varied.

frequencies which result in a number of frequency swings which will be found useful for calibration purposes.

It should be noted that the chance of error will be smallest when the earliest practical carrier suppression point is used. It should be added that the accuracy of the reading obtained is entirely dependent upon the accuracy to which the frequency of the modulating signal is known. The accuracy will also be affected by the means used to determine the point at which the carrier falls to zero.

## The Panoramic Monitor

By far the most convincing means of indicating the deviation ratio at which carrier and side band components fall to zero is that of the panoramic monitor. Fig. 10.2 shows a block circuit diagram of such a monitor, the use of which for this purpose was first proposed by Pieracci. The frequency modulated signal, the swing of which it is desired to determine, is fed into a mixer stage. Here it heterodynes with the output from an oscillator

TABLE 17

*Modulating frequencies corresponding to various useful frequency modulation deviations*

| First carrier suppression point (mod. index 2·4048) | | Second carrier suppression point (mod. index 5·5201) | | Fourth carrier suppression point (mod. index 11·7915) | |
|---|---|---|---|---|---|
| Freq. swing in kc/s | Mod. freq. in c/s | Freq. swing in kc/s | Mod. freq. in c/s | Freq. swing in kc/s | Mod. freq. in c/s |
| 1 | 415·8 | 5 | 961·1 | 40 | 3,392 |
| 2 | 813·7 | 10 | 1,920 | 45 | 3,816 |
| 3 | 1,247 | 15 | 2,880 | 50 | 4,240 |
| 4 | 1,663 | 20 | 3,840 | 55 | 4,664 |
| 5 | 2,079 | 25 | 4,800 | 60 | 5,088 |
| 6 | 2,594 | 30 | 5,760 | 65 | 5,512 |
| 7 | 2,911 | 35 | 6,720 | 70 | 5,936 |
| 8 | 3,326 | 40 | 7,680 | 75 | 6,360 |
| 9 | 3,742 | 45 | 8,640 | 80 | 6,784 |
| 10 | 4,158 | 50 | 9,601 | 85 | 7,208 |
| 15 | 6,237 | 55 | 10,561 | 90 | 7,732 |
| 20 | 8,316 | 60 | 11,521 | 95 | 8,156 |
| 25 | 10,395 | 65 | 12,481 | 100 | 8,480 |

which is frequency modulated by the time base wave-form producing the horizontal sweep on a cathode ray display tube. It therefore follows that the i.f. signal emerging from the mixer stage will be modulated in frequency. As a result the whole spectrum of the incoming frequency modulated signal is caused

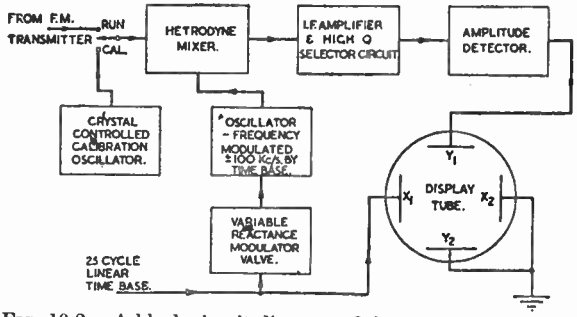



FIG. 10.2.—A block circuit diagram of the panoramic monitor.

*(By courtesy of the British Institute of Radio Engineers.)*

to successively pass through the admittance frequency of the narrow band selector circuit included in the i.f. amplifier. This in effect causes the signal's side band spectrum to be scanned by the narrow passband amplifier. As a result each side band component will pass in succession "in front of" the selector circuit. In this way the voltages resulting from the individual side bands will be displayed across the cathode ray tube screen in graphical form.


FIG. 10.3.—A typical panoramic monitor for the display of frequency modulation side band spectrum.

(*By courtesy of Panoramic Radio Corporation.*)

If, as is shown in Fig. 10.2, the monitor is used to check the outgoing signals from a frequency modulated transmitter, a crystal-controlled calibration oscillator having the frequency which has been assigned to the station is arranged so that it can be switched into circuit in order to facilitate adjustment to the monitor's oscillator frequency. Such adjustments may be necessary in order to make the assigned carrier frequency correspond exactly with the zero mark on the display screen.

So long as the carrier is unmodulated the display on the screen will be simply that of the single component representing the carrier alone. This condition is illustrated in the photograph shown in Fig. 10.4. As the carrier is frequency modulated the various side bands begin to appear on the screen. If the modulation index is increased until it is 2·4 it will be found that the amplitude

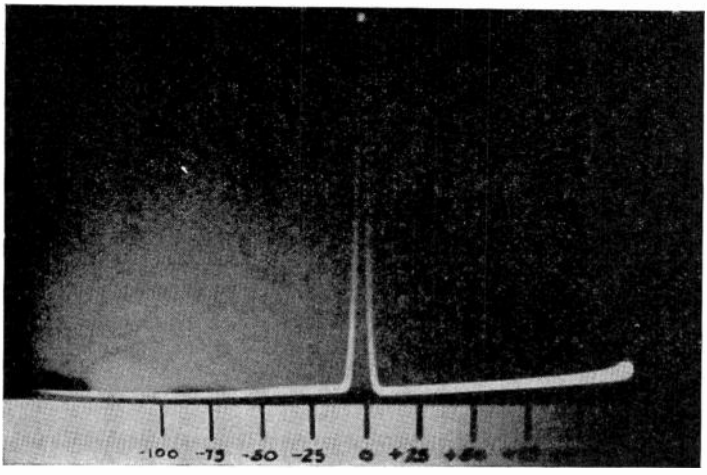



FIG. 10.4.—No frequency modulation—carrier component only.
*(By courtesy of I.R.E.)*

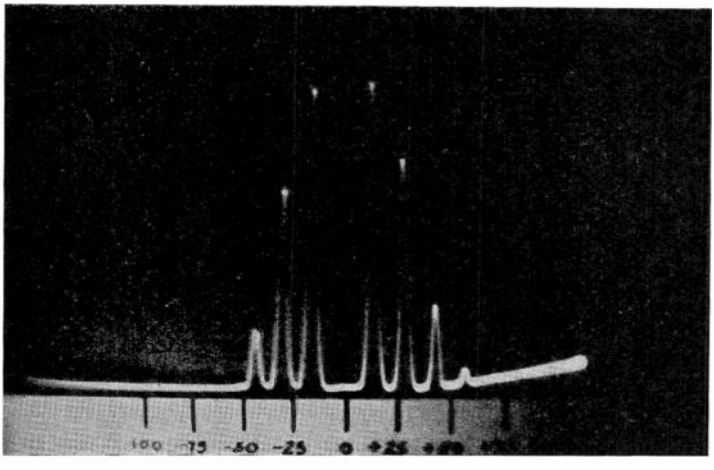



FIG. 10.5.—Modulation index 2·4—carrier component amplitude is zero.
*(By courtesy of I.R.E.)*

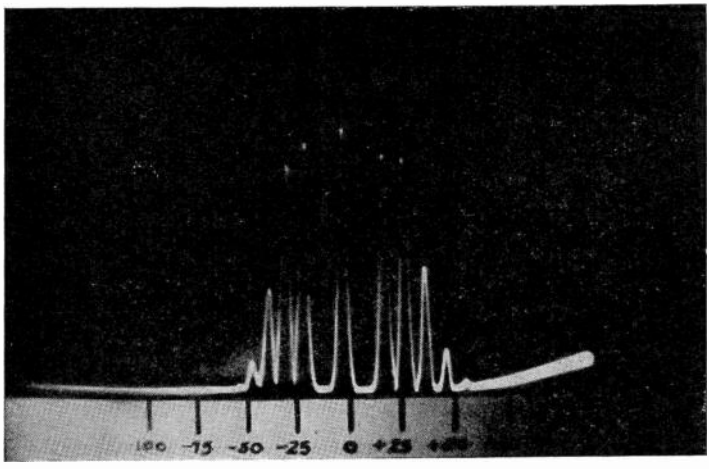



FIG. 10.6.—Modulation index 3·84—amplitude of the first pair of side bands is zero.

*(By courtesy of I.R.E.)*

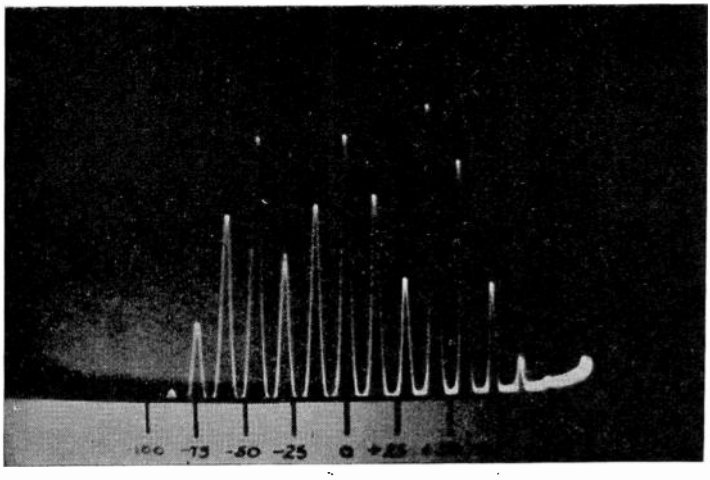



FIG. 10.7.—Modulation index approximately 4·8.

*(By courtesy of I.R.E.)*

of the carrier component will fall to zero. This condition is shown in Fig. 10.5. As the modulation index is still further increased a point will be reached at which, as shown in Fig. 10.6, the amplitude of the first pair of side bands falls to zero. The modulation index at this point is 3·8. If the modulation index continues to be increased the number of side bands is multiplied until, as shown in Fig. 10.8, they are so numerous that it is very difficult to

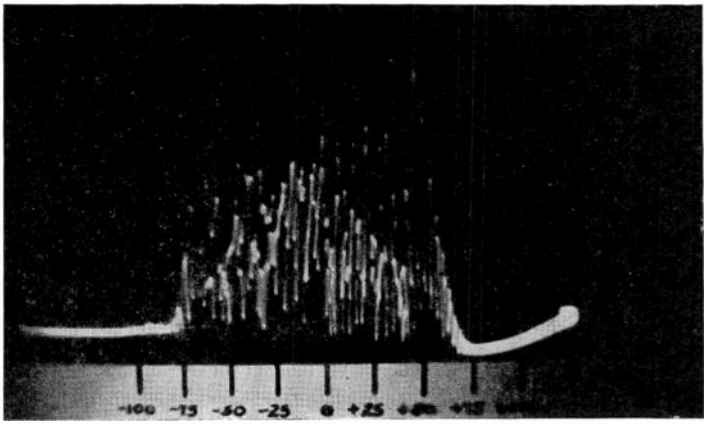



FIG. 10.8.—Modulation index approximately 24—this display gives an idea that obtained on normal programme material.
(*By courtesy of I.R.E.*)

derive much useful information from the resulting display. This last diagram also gives an idea of the type of display which may be expected from the average programme material.

## The Single Frequency Method of Measuring Frequency Deviation

As an alternative to the panoramic monitor, the single frequency method proposed by M. G. Crosby will normally be found more accurate. The general arrangement is indicated in Fig. 10.9. The oscillator, mixer, and i.f. amplifier can for convenience be those incorporated in an amplitude modulation receiver, provided that this receiver includes a "C.W." or beat oscillator, and is capable of tuning over the frequency modulation band. It is first necessary to tune the receiver to the carrier of the frequency modulated signal, while that signal is unmodulated. The receiver beat oscillator frequency should then be so set that it produces

a low audio frequency note in the headphones or loud-speaker. It is immaterial whether the beat-note is produced between the "C.W." oscillator and the incoming carrier or the i.f. signal. In order to ensure that all other signals are eliminated, an audio frequency filter should be arranged to pass the audio beat-note frequency only.

If now the carrier is frequency modulated with, say, a 1,000-cycle signal, the amplitude of which is gradually increased, it will be found that the carrier component's amplitude, as indicated by

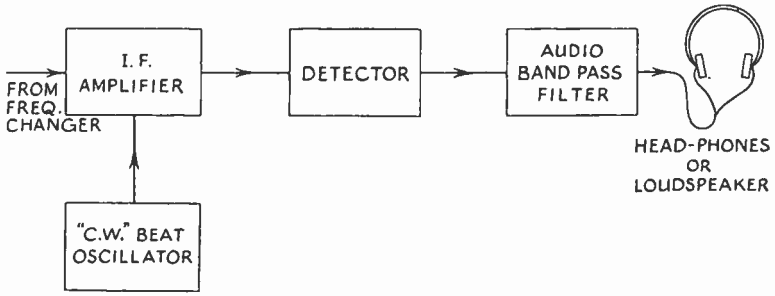



FIG. 10.9.—In the single frequency method of indicating the Bessel Zero points, the carrier is heterodyned with an oscillator of the type used to assist in the separation of continuous wave signals in communication receivers. The audio beat-note thus produced is selected by means of an audio band-pass filter. When this beat-note falls to zero the carrier amplitude is at a null point.

the volume of the audio beat-note, will fall until it finally reaches a null point. The modulation index at this point will be 2·4048 and the actual swing ±2,405 cycles. A further increase in the deviation ratio applied to the carrier causes the audio note to rise again to a maximum, after which it falls until a second zero point is reached. The modulation index at this point is 5·5201.

The band-pass filter included in the audio output circuit does not have to meet any very rigid requirements. The selectivity obtainable from a single-tuned audio circuit, using a paper condenser and an iron-cored inductance, is quite sufficient. It has even been found possible to observe the null points with the aid of a poor pair of headphones having a pronounced natural resonance. If a high modulating frequency of, say, 10,000 to 15,000 cycles is used, the selectivity of the average amplitude modulation receiver's i.f. channel will be sufficient to reduce the side band amplitudes to a satisfactorily low level. A high selectivity audio band-pass filter will, however, be necessary if

satisfactory results are to be obtained with low modulating frequencies in the order of 1,000 cycles or less.

It may be found that the frequency of the carrier component shifts as the modulation is applied. When this occurs the modulation must be raised slowly and the receiver carefully retuned to follow the shift. A small amount of shift may occur with a modulator which in all other respects gives negligible distortion.

### The Quieting Signal

The measurement of the quieting signal is of importance in that it indicates in a single figure that a receiver has sufficient gain, that excessive regeneration is absent, and that the amplifier stages are correctly aligned. In short, it is an overall figure of merit for the whole set. It should, however, be noted that it is only applicable to high sensitivity frequency modulation receivers.

The measurement actually consists of determining the input voltage at which the incoming carrier commences to suppress the interference. This voltage will vary from receiver to receiver, but for any given frequency modulation system it provides a means of laying down a figure of merit by which the standard of both design and manufacture of sets produced by different makers may be assessed. For example, the standard for a receiver operating in the 70 to 80 Mc/s band, and designed to receive a frequency deviation of plus and minus 15 kc/s, would be about 0·5 microvolts. If this figure is greatly exceeded it means either that the receiver has been poorly designed in the first place, and is as a result picking up an excessive noise voltage, or that its alignment is imperfect. Unbalance of the discriminator will also result in high quieting signals.

If it is noticed that an increase in signal input results in an *increased* noise voltage at very low inputs, this indicates that regeneration is present and is also contributing noise. A high quieting signal may also be due to a high local interference level due to electrical disturbances resulting from such causes as fluorescent lighting.

By definition the quieting signal is that signal which is necessary to reduce the output noise, at the loud-speaker, by 20 db or, sometimes, 30 db. The measurement has to be made in a well-screened room in which extraneous noises are at a low level, and may conveniently be used as a factory acceptance figure for high sensitivity

frequency modulation communication receivers. The actual measurement is made by first observing the noise voltage developed across the loud-speaker without any carrier input to the aerial. This condition may for convenience be obtained by connecting the signal generator to the receiver aerial input terminals, and temporarily setting the attenuator between the two lowest taps. The volume control should be adjusted to give a convenient reading on the output meter. A signal should next be applied to the aerial sockets; this signal should be increased until the noise voltage has fallen to one-tenth or one-thirtieth approximately of its former value. The signal input at this level is known as the quieting signal for 20 db and 30 db quieting respectively.

## SELECTED REFERENCES

*British Association Report*, 1915, pages 29–32. A complete set of Bessel Function Tables.

CROSBY, MURRAY G., A Method of Measuring Frequency Deviations, *R.C.A. Review*, April 1940.

PIERACCI, ROGER J., A Frequency Modulation Monitoring System, *Proc. I.R.E.*, August 1940.

RICE, H. E., Factory Alignment Equipment for Frequency Modulation Receivers, *Proc. I.R.E.*, October 1941.

BUSSARD, E. J. H., and MICHEL, T. J., A Wide-Band High Frequency Sweep Generator, *Electronics*, May 1942.

HILL, D. M., and CROSBY, MURRAY G., Design of F.M. Signal Generator, *Electronics*, November 1946.

*Chapter Eleven*

# PRACTICAL USES OF FREQUENCY MODULATED SIGNALS

THE largest single use to which frequency modulation has as yet been put is the provision of low interference high-fidelity broadcasting services. This being so, it was decided at the outset that this book should present the accepted technique employed for this particular purpose, and that in order to avoid confusion the various other applications should be grouped together for discussion in the last chapter.

It should be noted that the applications discussed are those in which frequency modulation is technically the correct choice. It should also be noted at this point that frequency modulated signals are sometimes employed for reasons entirely apart from the improvement in signal to noise ratio. Examples of such uses are radio altimeters, various radar applications (i.f.f., etc.), circuit alignment oscillators, and panoramic monitors. As the use of frequency modulated signals is in these cases dictated by entirely different technical considerations, these applications are not considered.

## Frequency Modulation Broadcasting

The medium- and long-wave broadcast bands—which up to the introduction of frequency modulation were the only bands extensively employed for the transmission of material intended for entertainment—suffer from various fundamental shortcomings. Such services have as their object the provision of music and speech which will give pleasure to the listener. It therefore follows that distortion and interference must both be maintained at the lowest possible level. This being so, any band on which ionospheric reflections occur will, due to the resultant distortion, be unsuitable for broadcasting services. A survey of the radio-frequency spectrum shows that such reflections have a maximum severity over the band from approximately 1·5 to 30 Mc/s.

It will be apparent, therefore, that local broadcast services can only be operated below 1·5 Mc/s and above 30 Mc/s. In the early

days when there was very little known about the band above 30 Mc/s, attention was confined to that below 1·5 Mc/s. It was, however, found that during the hours of daylight communication was confined to the area covered by the surface ray although, after dark, there was a considerable reflection from the ionised layers with the result that signals were transmitted over considerable distances. This increased night range could not be utilised as the mutual interference between the surface and reflected waves resulted in considerable fading and distortion.

The useful range of a medium-wave broadcast station is therefore quite a short distance—some 100 or so miles—while its interference range is that experienced at night as a result of reflection from the ionised layers. This range may extend up to some 500 or more miles. This unfortunate phenomenon meant that it was frequently impossible to have more than one station on a particular frequency in any given continent. In congested areas such as Europe, where many different languages are spoken, this results in a most serious limitation. In order to provide each country with a bare minimum of broadcast stations, the maximum channel width which could be allowed was 9 kc/s.

This meant that the quality of reproduction obtained was very severely restricted, it being necessary to limit the receiver's upper frequency response to some 3,000 to 5,000 c/s in order to secure adequate adjacent channel selectivity. It therefore follows that so long as broadcasting is confined to the band below 1·5 Mc/s, it will not be possible to realise the full entertainment value which could otherwise be provided. Even after the band above 30 Mc/s had been opened up—as a result of improvements in equipment—it was still not found possible to utilise it widely owing to the high interference level produced by automobile ignition systems.

The advent of frequency modulation, however, makes it possible to provide the ideal broadcast service. Frequency modulation permits a very high fidelity reproduction standard, low interference levels, and coverage without any large areas in which the transmission is not of a high enough standard to provide entertainment, but has a sufficiently high field strength to mar reception from other stations. Compared with the interference area produced by a medium-wave broadcast station, that produced by a frequency modulated station is negligible.

By using frequency modulation broadcasting it is therefore possible to provide a very large number of local programmes, each transmitter having almost the same useful range as the old medium-wave stations. The fringe of this useful range is, however, sharply defined, and it is possible to operate another station carrying a different programme within 200 to 300 miles without mutual interference.

### Frequency Modulated Radio Telephones

There is a wide demand for communications between mobile units and fixed control points, and also between one mobile unit and another. The need for this type of communication is felt by police, fire services, tram and bus companies, railway, gas and electric supply undertakings, to name but a few.

In America frequency modulation has been widely employed for such services, and the advantages it shows have been proved beyond all doubt. For example, the Connecticut State Police operate such a system. Where previously they could only employ 15-watt amplitude modulation mobile transmitters, the replacement frequency modulated equipment gave an output of 25 watts. Due to the greater transmission efficiency obtained with frequency modulation, this increase in power output was possible without increasing the power drawn from the battery.

In car-to-car tests carried out in New York one car was parked while the other was driven slowly away. With the amplitude modulation equipment it was not found possible to go more than nine or ten city blocks (up to half a mile) before the signal was lost in the general noise-level. When similar tests were carried out with the f.m. system, it was found possible to maintain communication for a distance of approximately five miles.

In two-way communication tests between a central control station, two patrol cars set out together, one fitted with an amplitude modulation and the other with frequency modulation equipment. At a distance of 7 miles from the central station the amplitude modulation system had lost contact due to the heavy noise conditions. The second car drove into New York, a distance of 45 miles, and was able to maintain two-way contact over the entire trip, despite the fact that it travelled through some of the heaviest ignition interference areas in the city.

The above results were obtained at a transmission frequency

of some 39 Mc/s. While they clearly demonstrate the advantages resulting from the use of frequency modulation, such very great improvements may not always be obtained. The range will naturally vary with the height of the central station aerial and the noise conditions existing at both the central station and the mobile unit. The ranges obtained will normally vary between 10 miles under heavy noise conditions and severe intervening terrain, to some 30 to 35 miles under low noise conditions and favourable terrain. Favourable terrain, in the latter case, would assume flat country, or in the case of hilly country, that the mobile units were located on high ground. Transmitter power of between 25 and 50 watts is assumed.

Central station transmitter powers of 50 to 250 watts are the most common. The 50-watt stations are normally used when reliable two-way communication of up to some 20 miles is desired. The usual aerial is a half-wave co-axial dipole. Directional aerials are rarely used because the central station is normally near the middle of the area to be covered, and uniform transmission and reception in all directions is generally desired. The average central station has an aerial height of about 100 feet above the surrounding terrain (or buildings in the case of a town).

For police or public utility applications a much greater range is usually required, and the use of 250-watt central stations is common. In addition, every attempt is made to locate the central station in a rural area or on a hill-top where noise is at a minimum. This procedure allows the central station receiver to make full use of its inherent sensitivity, so that the talk-back range from mobile unit to central station will more nearly equal the range of the central station. The high power of the central station is offset to some extent by the high noise conditions under which mobile units are normally forced to operate. The central station transmitter is usually remotely controlled by means of wire lines or, in some cases, by an auxiliary radio control circuit. Two-way ranges up to 60 miles are commonly reported for such installations, but conservative estimates for system design purposes usually average 40 to 45 miles. The installations reporting consistent ranges up to 60 miles are using central station aerials 250 to 300 feet above the general level of the surrounding country.

Some installations embodying 3-kW. central stations have been made, with ranges to mobile units of 100 to 120 miles. The

talk-back range is, of course, not increased by raising the central station power.

The advantages shown by frequency modulation over amplitude modulation are also amply borne out in tests made between aircraft and ground stations. In typical tests of this type, using a 4-watt frequency modulated transmitter in an aircraft, it was found to be possible to obtain a fairly reliable talk-back range of between 150 and 175 miles. Up to the threshold of improvement reception was almost perfect: at this point, however, it suddenly became impossible. The transmitter used in these tests had a 3-kc/s maximum audio frequency signal and a peak deviation of 12 kc/s. One of the most interesting points of comparison was that the signals from an amplitude modulated transmitter of the same power became gradually worse until at about two-thirds the maximum range of the frequency modulated transmitter, reception became so unreliable and fading so bad that regular communication had to be regarded as impossible.

### Frequency Shift Radio Telegraph Systems

The advantages of frequency modulation are not confined to the transmission of speech and music. When employed on a radio-telegraph circuit, as much as 20 db signal to noise improvement can be expected by changing over to frequency shift transmission. To take a practical example: during the war a mobile 400-watt frequency shift transmitter on the beachhead in France transmitted press traffic direct to the United States at a rate of 500 words a minute—over a million words a month. In former days a 50-kilowatt transmitter would have had trouble in maintaining the circuit.

Several types of carrier shift equipment are being used. One commercial transmitting equipment takes energy from a crystal oscillator and beats it against an extremely stable self-excited 200-kc/s oscillator. The frequency of this self-excited oscillator is shifted or modulated in frequency by the signal which is to be transmitted, being increased in frequency on mark and decreased on space. The resultant beat signal is selected and forms the outgoing carrier.

The signal which it is desired to transmit is usually in the form of a square wave. It is first filtered to eliminate frequencies higher than the third harmonic of the highest keying frequency required.

The filter used for this purpose must be so designed that the phase relations of the harmonics up to the third are not altered with respect to the fundamental. In practice it is possible to design filters introducing less than 1·5 per cent distortion. This filtering stage is included in order that the band-width transmitted may be a minimum.

The normal amplitude modulated teletype signal has a fundamental of 23 cycles and a third harmonic of one-third the fundamental's amplitude. If the carrier amplitude is keyed "make and break" by this signal, the ideal band-width would be twice the third harmonic frequency or, say, 138 cycles. In practice, however, such a narrow band would never be attained, as the power amplifier stages of the transmitter tend to square the keying signal. The best possible transmitter adjustment requires a band-width of approximately 1,200 cycles. Only side-bands greater than 40 db below the unmodulated carrier level are considered in this value.

If on the other hand the carrier is frequency shifted by 850 cycles (by the same teletype signal) the side bands of the emitted signal would occupy a band-width of 1,100 cycles (see Chapter Two). Were the carrier shift reduced to 250 cycles with the same signal, the band-width would be only 480 cycles. Thus frequency shift telegraph transmission can result in a smaller band-width than the normal amplitude modulated make-and-break keying of the carrier.

A typical commercial system, that of Press Wireless, has adopted a frequency shift of 850 cycles as standard. It is claimed that this gives the best compromise between signal to noise level and band-width. However, the frequency shift is varied between 400 cycles and 1,200 cycles for special services. For example, when, as is discussed in a later section, it is used for high-speed facsimile and photograph services, a 1,200 cycles frequency shift is used. In frequency shift transmission the signal varies symmetrically about the assigned frequency. A frequency shift of 1,200 cycles would be a shift of from 600 cycles above the assigned frequency to 600 cycles below the same frequency.

### Frequency Shift Receivers

The receiving systems of the various commercial companies using frequency shift transmission are similar in principle, but

differ in circuit details. For example, Press Wireless use an amplitude modulated communication receiver which delivers an audio beat-note to a band-pass filter. This beat-note shifts in frequency about a mean of 2,550 cycles in accordance with the transmitter's frequency variations. The band passed by the filter must be wide enough to pass not only the two frequencies between which the audio beat-note shifts, but also all frequency modulation side bands which are 10 per cent or more of the carrier amplitude. The passband must also be wide enough to tolerate possible transmitter or receiver drift.

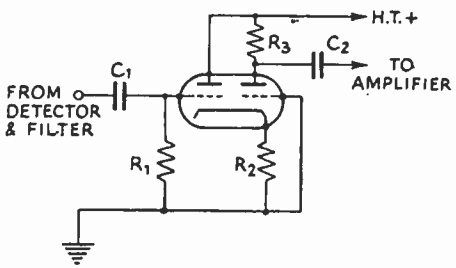



FIG. 11.1—A transient-free limiter in which the first section of the double triode limits the negative peaks, and the second section, cathode coupled to the first, limits the positive peaks.

From the filter, the signal goes to a limiter. As the "carrier" is in the audio frequency range it is comparable in frequency with the intelligence it carries. The requirements of this limiter are therefore somewhat more stringent than those of the normal "i.f." limiter used in frequency modulation broadcast receivers. Thus, unlike the broadcast receiver limiter, where transients only need to be short as compared to the intelligence, the transients of the limiter in question must be extremely short compared with both the carrier and the intelligence frequencies.

The outline circuit of a suitable limiter is shown in Fig. 11.1. The valve used is a dual high-*mu* triode. Considering, firstly, the effect of a signal applied through the condenser $C_1$ to the grid of the first triode section. As small negative potentials will cut off the valve, it follows that the voltage across $R_2$ due to the current in the first triode section is zero during most of the negative half-cycle. The input resistance of the limiter is only $R_1$. As the grid swings positive with respect to ground, the current

drawn by the valve increases, so increasing the voltage across $R_2$. $R_2$ is sufficiently large to ensure that at no time will the grid voltage exceed the cathode voltage. Thus the grid never goes positive with respect to its cathode; as no grid current flows, no charge which would subsequently leak off through $R_1$ is developed across $C_1$. It follows that as there is no time constant effect involving $R_1$ and $C_1$, the circuit is instantaneous in its action and no transient effects can result.

As the voltage across $R_2$ increases due to the positive swing of the first grid, the second triode passes into the region of cut-off. This second triode is essentially a cathode drive stage excited by the first triode. The gain of the second triode is low because its plate resistance $R_3$ is small. It cuts off at about the same positive swing of the first triode grid, as does the first triode for negative swings of its own grid. Therefore, the action of the limiter is symmetrical about the zero axis, and is both transientless and instantaneous for any abrupt level or frequency change. This circuit gives about 30 db of limiting. Two limiter stages separated by a class A amplifier supply the 60 db of limiting normally necessary.

### Discriminators for Teletype, Telephoto, and Facsimile

In many commercial types of receiver the limiter is followed by a single-ended slope circuit. No attempt is made to eliminate noise side bands outside the deviation spectrum. When such a discriminator is employed it must be of the extended range type, i.e. it must be linear far beyond the deviation band so as not to discriminate against noise components, otherwise undesired amplitude modulation will result in noise.

Experiments have shown a definite advantage in the use of a double slope circuit type of discriminator filter. Its use is therefore recommended in all terminal equipment for teletype, telephoto, and facsimile. Two forms of such a frequency discriminating filter are shown in Fig. 11.2. The input impedance of both is a constant over the working band and is equal to $R$. The values of $L$ and $C$ are given by $LC = \frac{1}{\omega^2}$ and $\frac{L}{C} = 2R^2$, where $\omega$ is $2\pi$ times the cross-over frequency. The output characteristic of the circuit shown in Fig. 11.2 ($a$), although it does not give a linear response over as wide a frequency range as does that of the circuit shown

in Fig. 11.2 (*b*), delivers a higher output voltage and is perfectly symmetrical.

Following the discriminator there is a detector from which the signal is fed to a low-pass filter in order to eliminate noise caused by phase modulation of the signal at frequencies higher than the

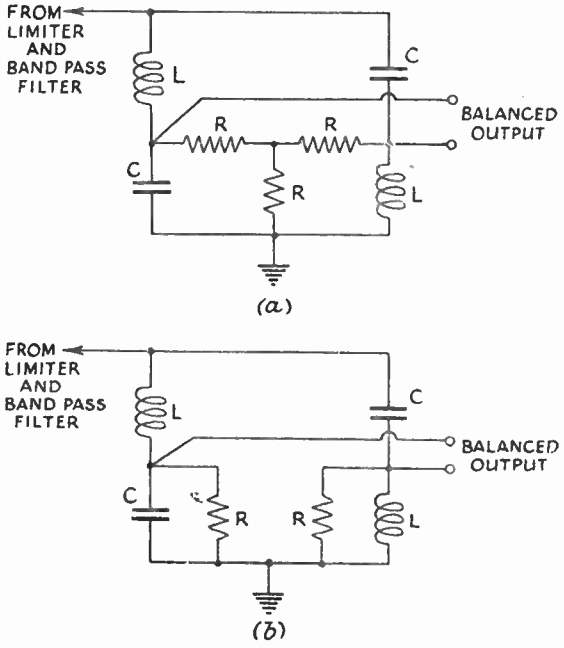



FIG. 11.2—Two forms of discriminator. The discriminator for frequency shift reception must be linear over a band far wider than the deviation limits, so as to avoid amplitude modulation in the output being produced by noise in the input.

desired intelligence frequency. This filter passes frequencies up to the third harmonic, which for teletype signals may be 70 to 100 c/s; for high-speed Morse, 250 c/s and for telephoto and facsimile (as discussed later), some 600 c/s.

## Sub-Carrier Frequency Modulation Systems

Prior to 1939 long-distance radio-telephoto services operated on the constant frequency variable depth method. As it is possible to obtain a substantial improvement by the use of frequency modulation in one field it is natural to expect that this improvement could be obtained in all similar fields. This has, in fact,

been found to be the case. By May 1939 Cable and Wireless, Ltd., and R.C.A. Communications, Inc., acting co-operatively, made available for public service a sub-carrier system of frequency modulation. This system made possible the following improvements over the earlier method:

1. The speed of operation was increased threefold. Facsimile matter was sent at 60 revolutions per minute with a line advance of 120 lines per inch.
2. Linear amplitude recording was obtained and resulted in much improved detail, fidelity of tonal values, and elimination of the screen pattern associated with the older CFVD method.
3. Usable pictures were obtained through very much poorer signal conditions.
4. Streaks caused by multipath or selective fading were minimised.

The advantages in signal to noise ratio normally associated with frequency modulation of a radio-frequency carrier are equally applicable in the case of a frequency modulated carrier in the audio frequency range. If such a frequency modulated note is itself used as the modulating signal for an amplitude modulated transmission, the process is known as sub-carrier frequency modulation. When receiving such signals the audio note is first demodulated in the normal manner, after which it is passed through a limiter stage and then fed to a frequency demodulation filter.

In the sub-carrier system just referred to, black corresponded to a frequency of 1,600 cycles and white to a frequency of 2,000 cycles. The use of this narrow band of frequencies permitted as much noise as possible to be removed by the audio band-pass filters preceding the frequency discriminating filter. It also ensured that any undesired harmonics of the sub-carrier which were inserted by selective fading were filtered out before the signal was applied to the limiter.

It is desirable that the lowest sub-carrier frequency possible should be used in order that the components of the transmitted carrier may be confined to the smallest practical band-width. The lowest frequency limit is determined by the smallest picture detail to be transmitted. It must be such that the narrowest vertical

line or space will contain at least one full cycle of the sub-carrier. Theoretically, it should contain even more cycles, but it has been found in practice that generally one complete cycle will suffice.

### Picture Transmission by the Frequency Shift Method

The sub-carrier method of transmission for long-distance telephoto circuits has now been largely displaced by the frequency


FIG. 11.3.—A commercial radiophoto picture transmitted from London to New York by the CFVD method.
*(By courtesy of International Newsphoto.)*

shift method already outlined in connection with telegraph transmissions. This method requires less transmission band-width than the sub-carrier system. When receiving the frequency shift telephoto signals it is normal to demodulate the received signal

after limiting, by beating it against a fixed oscillator, after which the resultant is fed to a frequency discriminating filter. There are two methods normally used to maintain the stability of the beat oscillator frequency. Firstly, it may be crystal-controlled, and, secondly, its frequency may be determined from the mean

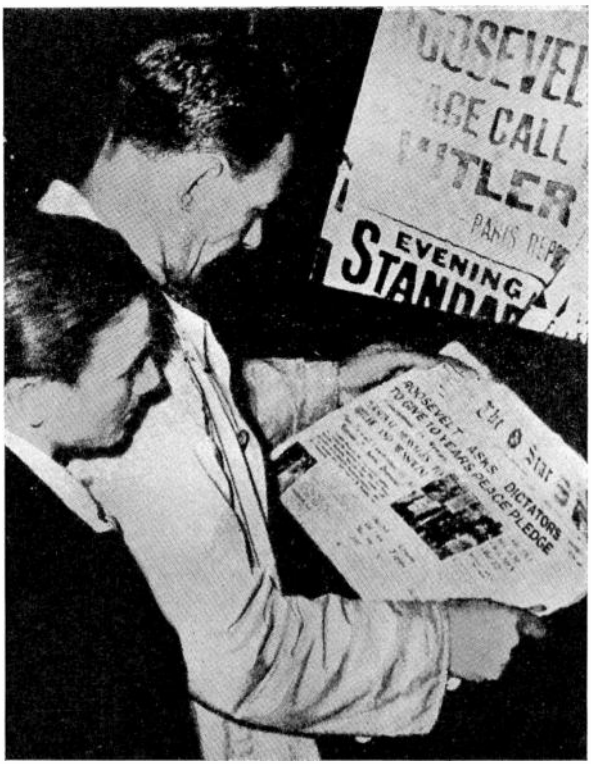



FIG. 11.4.—The same commercial radiophoto picture transmitted by the sub-carrier frequency modulation system.
(*By courtesy of International Newsphoto.*)

of the frequency deviations of the received signal. In this latter case it has been found possible to maintain the zero point accurate to within some $\pm 100$ cycles.

## Combined Frequency and Amplitude Modulation Transmission

There are sometimes cases where both frequency and amplitude modulation may be used in a single transmission with advantage.

One such example occurs when it is desired to obtain the advantages of frequency shift transmission for long-distance telegraph circuits, without having to replace all the receiving equipment. In one such case a keyed "make and break" amplitude modulated transmission was frequency modulated to the extent of 400 cycles peak to peak, and at the same time amplitude modulated by a 400-cycle note. A most marked reduction in fading was obtained in this way.

### Phase Modulated Signals

Phase modulation has been used for at least one mobile radio-telephone system.* A study of the latter part of Chapter Two shows, however, that for all such systems a direct relationship between the carrier frequency variations and the initial audio signal utilises the frequency band allocated to the transmitter in the most efficient way.

If the carrier's phase variations are proportional to the initial audio signal, it follows that a given angle of phase modulation will result in a greater frequency deviation at the higher audio frequencies than the lower. It follows, therefore, that the depth of modulation will have to be such that the maximum frequency deviation is not exceeded in the higher audio frequency region. As, however, the mean amplitude of the audio signal is substantially uniform over the restricted audio band used for radio-telephone channels, it follows that the use of phase modulation must result in a reduction in the general deviation amplitude of the signal. For radio-telephone, telephoto, and facsimile services phase modulation therefore results in a definitely inferior signal to noise ratio to that obtained with frequency modulation.

In the case of high-fidelity broadcasting the audio frequency range is greatly extended and the average amplitude of the higher audio frequencies is considerably smaller than that of the lower frequencies. It follows, therefore, that the use of the frequency modulation relationship up to some 2,000 to 3,000 cycles and beyond that of phase modulation will result in the most efficient use of the available band-width. In practice this modified form of phase modulation is known as pre-emphasised frequency modulation, and is discussed under this heading elsewhere in the book.

* "P.M. Communication System for Chicago Surface Lines" (Beverly Dudley), *Electronics*, January 1944.

## SELECTED REFERENCES

MATHES, R. E., and WHITAKER, J. N., Radio Facsimile by Sub-Carrier Frequency Modulation, *R.C.A. Review*, October 1939.

GUY, R. F., and MORRIS, R. M., N.B.C. Frequency Modulation Field Test, *R.C.A. Review*, October 1940.

WARNER, S. E., Two-Way Police F.M. Performance, *F.M.*, January 1941.

NEITZERT, CARL, and MURNANE, JOHN E., New Two-Way F.M. Plan for Jersey, *F.M.*, May 1942.

Pennsylvania Turnpike U.H.F. Traffic Control System, *Electronics*, May 1942.

TIBBS, C. E., Future Applications of F.M. (F.M. Television Synchronising Pulses), *Wireless World*, June 1943.

BLISS, W. H., The Use of Sub-Carrier Modulation in Communication Systems, *Proc. I.R.E.*, August 1943.

DUDLEY, BEVERLY, P.M. Communication System for Chicago Surface Lines, *Electronics*, January 1944.

DILLON, PAUL, A 337 Mc/s. F.M. Studio to Station Link, *Electronics*, March 1944.

SPRAGUE, R. M., Frequency Shift Radio Telegraph and Teletype System, *Electronics*, November 1944.

MIESSNER, B. F., Frequency Modulation Phonograph Pickup, *Electronics*, November 1944.

F.M. Carrier Telephony for 230 kV. Lines, *Electronics*, December 1944.

DELORAINE, E. M., and LABIN, E., Pulse-Time Modulation, *Electrical Communications*, Vol. 22, No. 2, 1944. Also *Electronics*, January 1945.

SUCKLING, E. E., A Stabilized Narrow-Band Frequency Modulation System for Duplex Working, *Proc. I.R.E.*, January 1945.

BUDELMAN, F. T., F.M. Carrier Communication Equipment, *F.M. and Television*, January 1945.

BEATTY, W. A., Proposals for Television and Broadcasting Transmission Systems (Discussion on Pulse-Time Transmission), *Journal Brit. I.R.E.*, April 1945.

# INDEX